



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

MÁSTER EN BIG DATA: TECNOLOGÍA Y ANALÍTICA
AVANZADA

GOBIERNO DE PROCESOS CLOUD

Autor: Alicia Ocón Galilea

Director: Katherin Salazar Sepúlveda

Madrid

Julio 2021

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

GOBIERNO DE PROCESOS CLOUD

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2020/21 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

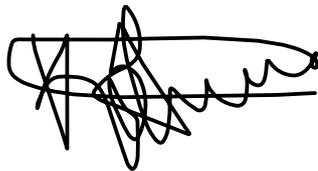


Fdo.: Alicia Ocón Galilea

Fecha: 09/07/2021

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Katherin Salazar Sepúlveda

Fecha: 09/07/2021

Vº Bº del Coordinador de Proyectos

Fdo.: Carlos Morrás Ruiz-Falcó

Fecha://

AUTORIZACIÓN PARA LA DIGITALIZACIÓN, DEPÓSITO Y DIVULGACIÓN EN RED DE PROYECTOS FIN DE GRADO, FIN DE MÁSTER, TESIS O MEMORIAS DE BACHILLERATO

1º. Declaración de la autoría y acreditación de la misma.

El autor D. Alicia Ocón Galilea, DECLARA ser el titular de los derechos de propiedad intelectual de la obra: GOBIERNO DE PROCESOS CLOUD, que ésta es una obra original, y que ostenta la condición de autor en el sentido que otorga la Ley de Propiedad Intelectual.

2º. Objeto y fines de la cesión.

Con el fin de dar la máxima difusión a la obra citada a través del Repositorio institucional de la Universidad, el autor **CEDE** a la Universidad Pontificia Comillas, de forma gratuita y no exclusiva, por el máximo plazo legal y con ámbito universal, los derechos de digitalización, de archivo, de reproducción, de distribución y de comunicación pública, incluido el derecho de puesta a disposición electrónica, tal y como se describen en la Ley de Propiedad Intelectual. El derecho de transformación se cede a los únicos efectos de lo dispuesto en la letra a) del apartado siguiente.

3º. Condiciones de la cesión y acceso

Sin perjuicio de la titularidad de la obra, que sigue correspondiendo a su autor, la cesión de derechos contemplada en esta licencia habilita para:

- a) Transformarla con el fin de adaptarla a cualquier tecnología que permita incorporarla a internet y hacerla accesible; incorporar metadatos para realizar el registro de la obra e incorporar “marcas de agua” o cualquier otro sistema de seguridad o de protección.
- b) Reproducirla en un soporte digital para su incorporación a una base de datos electrónica, incluyendo el derecho de reproducir y almacenar la obra en servidores, a los efectos de garantizar su seguridad, conservación y preservar el formato.
- c) Comunicarla, por defecto, a través de un archivo institucional abierto, accesible de modo libre y gratuito a través de internet.
- d) Cualquier otra forma de acceso (restringido, embargado, cerrado) deberá solicitarse expresamente y obedecer a causas justificadas.
- e) Asignar por defecto a estos trabajos una licencia Creative Commons.
- f) Asignar por defecto a estos trabajos un HANDLE (URL *persistente*).

4º. Derechos del autor.

El autor, en tanto que titular de una obra tiene derecho a:

- a) Que la Universidad identifique claramente su nombre como autor de la misma
- b) Comunicar y dar publicidad a la obra en la versión que ceda y en otras posteriores a través de cualquier medio.
- c) Solicitar la retirada de la obra del repositorio por causa justificada.
- d) Recibir notificación fehaciente de cualquier reclamación que puedan formular terceras personas en relación con la obra y, en particular, de reclamaciones relativas a los derechos de propiedad intelectual sobre ella.

5º. Deberes del autor.

- El autor se compromete a:
 - a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.
 - b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.
 - c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e intereses a causa de la cesión.
 - d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción de derechos derivada de las obras objeto de la cesión.

6º. Fines y funcionamiento del Repositorio Institucional.

La obra se pondrá a disposición de los usuarios para que hagan de ella un uso justo y respetuoso con los derechos del autor, según lo permitido por la legislación aplicable, y con fines de estudio, investigación, o cualquier otro fin lícito. Con dicha finalidad, la Universidad asume los siguientes deberes y se reserva las siguientes facultades:

- La Universidad informará a los usuarios del archivo sobre los usos permitidos, y no garantiza ni asume responsabilidad alguna por otras formas en que los usuarios hagan un uso posterior de las obras no conforme con la legislación vigente. El uso posterior, más allá de la copia privada, requerirá que se cite la fuente y se reconozca la autoría, que no se obtenga beneficio comercial, y que no se realicen obras derivadas.
- La Universidad no revisará el contenido de las obras, que en todo caso permanecerá bajo la responsabilidad exclusiva del autor y no estará obligada a ejercitar acciones legales en nombre del autor en el supuesto de infracciones a derechos de propiedad intelectual derivados del depósito y archivo de las obras. El autor renuncia a cualquier reclamación frente a la Universidad por las formas no ajustadas a la legislación vigente en que los usuarios hagan uso de las obras.
- La Universidad adoptará las medidas necesarias para la preservación de la obra en un futuro.
- La Universidad se reserva la facultad de retirar la obra, previa notificación al autor, en supuestos suficientemente justificados, o en caso de reclamaciones de terceros.

Madrid, a 09 de Julio de 2021

ACEPTA



Fdo. Alicia Ocón Galilea

Motivos para solicitar el acceso restringido, cerrado o embargado del trabajo en el Repositorio Institucional:



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

MÁSTER EN BIG DATA: TECNOLOGÍA Y ANALÍTICA
AVANZADA

GOBIERNO DE PROCESOS CLOUD

Autor: Alicia Ocón Galilea

Director: Katherin Salazar Sepúlveda

Madrid

Junio 2018

GOBIERNO DE PROCESOS CLOUD

Autor: Ocón Galilea, Alicia.

Director: Salazar Sepúlveda, Katherin.

Entidad Colaboradora: Kabel Sistemas de Información S.L.

RESUMEN DEL PROYECTO

El presente proyecto se realiza en la empresa Kabel y presenta una solución integral en la nube de Microsoft basada en la integración de documentos Excel en SQL Server, mediante un proceso ETL con Azure Data Factory, y la monitorización de la información en Power BI, desarrollando una estrategia de gobierno del ciclo de vida de los datos.

Palabras clave: Nube, Gobierno del Dato, Proceso ETL, Inteligencia de Negocios

1. Introducción

La información se ha convertido en un activo clave para garantizar una adecuada toma de decisiones, además de para mejorar las operaciones de negocio y adquirir una ventaja competitiva frente a las demás organizaciones [1].

Debido al crecimiento que ha experimentado la compañía en los últimos años, ya no resulta posible llevar un registro manual del gran volumen de información referente a las capacidades técnicas de los consultores. Como resultado, dicha herramienta, basada en varios conjuntos de hojas de cálculo en Excel, ha quedado obsoleta.

Es por eso por lo que la asignación actual de recursos se realiza con datos fragmentados, imprecisos y desactualizados. Asimismo, al no disponer de procesos de validación, se incurre en varios problemas, encontrándose entre ellos la falta de definición de una estrategia de gobierno que permita profundizar en el tratamiento de los datos, garantizando su calidad, accesibilidad o trazabilidad [2].

2. Definición del Proyecto

Para dar solución a la problemática existente, se propone la implantación de una nueva herramienta que posibilite la integración de datos de varias fuentes, facilite el mantenimiento de la información mediante un proceso automatizado y favorezca la extracción de conocimiento. Todo esto con el objetivo de tomar decisiones más ajustadas a la realidad, en base a la experiencia interna de la compañía.

El proyecto se despliega en una nube pública, que permite almacenar la información de forma centralizada, disponer de la capacidad de cómputo necesaria para su análisis, monitorizar datos en tiempo real o compartirlos en un entorno colaborativo [3].

En febrero de 2021, la consultora Gartner posiciona a Microsoft como líder en plataformas de análisis de datos e inteligencia de negocios por decimocuarto año consecutivo [4]. Este reconocimiento, junto con su fácil integración con otros productos y el asesoramiento de personas ampliamente formadas en esta tecnología dentro de la compañía Kabel, han sido los motivos que han llevado a la elección de Azure como entorno de desarrollo para el proyecto.

3. Descripción del modelo

En la ilustración 1 se incluyen todas las fases de la gestión del ciclo de vida de los datos tratadas en el proyecto, que son descritas brevemente a continuación.



Ilustración 1 – Etapas del proyecto

- **Exploración de datos.** En primer lugar, se estudian los datos relevantes para la migración, que se encuentran recopilados en varios ficheros de Excel y compartidos en un repositorio de SharePoint.
- **Almacenamiento de datos.** El sistema actual de almacenamiento es una base de datos relacional de SQL Server que recoge y centraliza la información procedente de las aplicaciones internas de la compañía. Gracias a SQL Server Management Studio, se adapta el modelo actual de la base de datos para albergar el nuevo contenido.
- **Integración de datos.** Tras identificar el origen y el destino, se realiza un proceso ETL con Azure Data Factory que permite integrar la información de los ficheros de Excel en la base de datos de SQL Server.
- **Mantenimiento de datos.** Paralelamente, se participa en el diseño de una API e interfaz de usuario para la implantación de una nueva aplicación de uso interno, que aporte las herramientas básicas necesarias para el mantenimiento de la información y los ciclos de aprobación y actualización periódica.
- **Análisis de datos.** Se extiende el modelo en Power BI Desktop mediante consultas SQL a la base de datos. A continuación, se crean las medidas y columnas calculadas necesarias utilizando expresiones de análisis de datos, DAX, para confeccionar un conjunto de cuadros de mando que permitan visualizar e interpretar la información.
- **Acceso a datos.** Por último, se organiza el contenido en un área de trabajo y se despliega la aplicación como espacio de consumo de la información y medio de distribución de los informes a un público más amplio, a través de Power BI Service, administrando los permisos de acceso acordes al perfil de cada usuario.

4. Resultados

En la ilustración 2, se muestran los puntos abordados en el proyecto, teniendo en cuenta todo el ciclo de vida del sistema con el objetivo de solucionar la problemática actual. Abarca desde la integración y centralización de la información, pasando por la creación de una herramienta para su mantenimiento, hasta el uso de técnicas de inteligencia de negocios para convertir los datos en conocimiento.

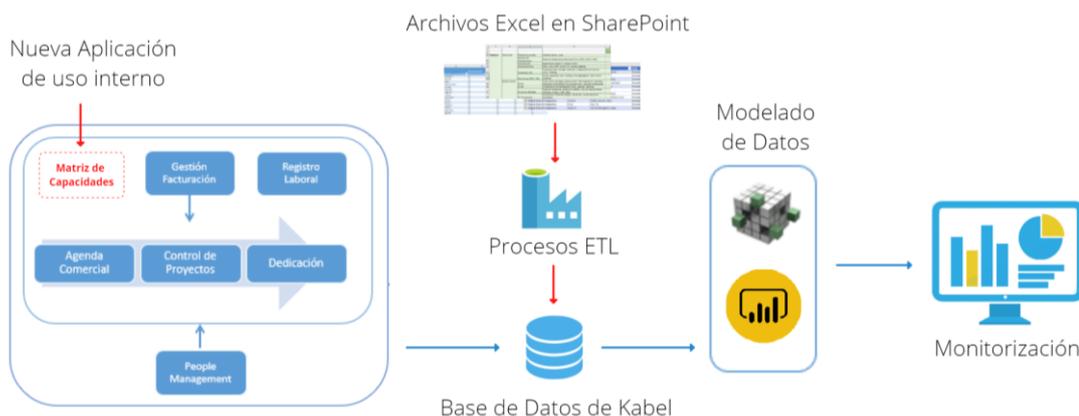


Ilustración 2 - Diagrama de resultados

5. Conclusiones

Como resultado, se ha aportado a la compañía una nueva herramienta para la toma de decisiones en cuanto a la asignación de recursos y a la gestión de equipos, optimizando así el proceso de selección del candidato que mejor se ajuste a las capacidades técnicas demandadas para cada proyecto.

Para conseguir dicho objetivo, se ha prestado, en primer lugar, gran atención al diseño de procesos de validación que aseguren la calidad del dato. Igualmente, se ha almacenado el histórico de cambios para posibilitar su trazabilidad. Además, se ha considerado la seguridad de los datos mediante la definición de roles de acceso. En conclusión, se ha contribuido a la democratización del dato, poniendo la información al alcance de todos los usuarios. Todo ello a través de una aplicación que representa el punto de partida para proponer nuevas iniciativas que enriquezcan este proyecto en el futuro.

6. Referencias

- [1] Fawcett, T. "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking". July 2013. O'Reilly Media.
- [2] Dama International. "DAMA-DMBOK: Guía del Conocimiento Para la Gestión de Datos". Julio 2020. Technics Publications.
- [3] Babcock, C. "Management Strategies for the Cloud Revolution. How Cloud Computing Is Transforming Business and Why You Can't Afford to Be Left Behind". May 2010. McGraw-Hill.
- [4] Sun, J. "Magic Quadrant for Analytics and Business Intelligence Platforms". February 2021. Gartner.

GOBIERNO DE PROCESOS CLOUD

Author: Ocón Galilea, Alicia.

Supervisor: Salazar Sepúlveda, Katherin.

Collaborating Entity: Kabel Sistemas de Información S.L.

ABSTRACT

This project is carried out within the Kabel company and presents a comprehensive solution on the Microsoft cloud based on integrating Excel documents in SQL Server, through an ETL process with Azure Data Factory, and monitoring information in Power BI, developing a data lifecycle governance strategy.

Keywords: Cloud, Data Governance, ETL Process, Business Intelligence

1. Introduction

Information has become a key asset in order to ensure proper decision-making, as well as to improve business operations and gain a competitive advantage over other organizations [1].

Due to the growth experienced by the company over the last years, it is no longer possible to keep a manual record of the large volume of information regarding the technical capabilities of consultants. Therefore, this tool, based on several sets of spreadsheets in Excel, has become obsolete.

That is why today's resource allocation is being carried out with fragmented, inaccurate, and outdated data. Likewise, if there are no validation processes, several problems arise, among them the lack of definition of a government strategy that allows focusing on the data treatment, guaranteeing its quality, accessibility or lineage [2].

2. Project definition

In order to solve the existing problem, we propose the implementation of a new tool that enables the integration of data from various sources. It also facilitates information maintenance through an automated process and improves the extraction of knowledge. All of this is aimed at making decisions more adjusted to reality, based on the company's internal experience.

The project is conducted in a public cloud, which allows the central storage of information, the necessary computing capacity for its analysis, the monitoring of data in real time or the sharing in a collaborative environment [3].

In February 2021, the consulting firm Gartner positioned Microsoft as a leader in data analysis and business intelligence platforms for the fourteenth consecutive year [4]. This recognition, together with its easy integration with other products and the mentoring of people widely trained in this technology within the Kabel company, have been the reasons that have led to choosing Azure as the development environment for the project.

3. Description of the model

Illustration 1 includes all the phases of the data lifecycle management treated in the project, briefly described below.



Illustration 1 – Project stages

- **Exploring data.** First of all, we examine data relevant for migration, which is collected in several Excel files and shared in a SharePoint repository.
- **Storing data.** The current storage system is a relational SQL Server database that collects and centralizes information from the company's internal applications. Thanks to SQL Server Management Studio, we are able to adapt the current database model to accommodate the new content.
- **Integrating data.** After identifying the source and destination, an ETL process that allows the integration of the information from the Excel files into the SQL Server database is carried out through Azure Data Factory.
- **Maintaining data.** At the same time, we participate in the design of an API and user interface for the implementation of a new application for internal use, which provides the basic tools necessary for information maintenance and cycles of validation and periodic updates.
- **Analyzing data.** We extend the model within Power BI Desktop using SQL queries to the database. Then, we create the necessary measures and calculated columns using DAX, in order to build a set of dashboards that make it possible to display an interpret information.
- **Accessing data.** Finally, we organize the content in a workspace and we deploy the application as an environment for information consumption and a means of distributing the reports to a wider audience, through Power BI Service, managing the adequate access permissions to each user's profile.

4. Outcomes

Illustration 2 shows the points addressed in the project, taking into account the entire system lifecycle for purposes of solving the current problem. It goes from the information integration and centralization, through the creation of a tool for its maintenance, to the use of business intelligence techniques that convert data into knowledge.

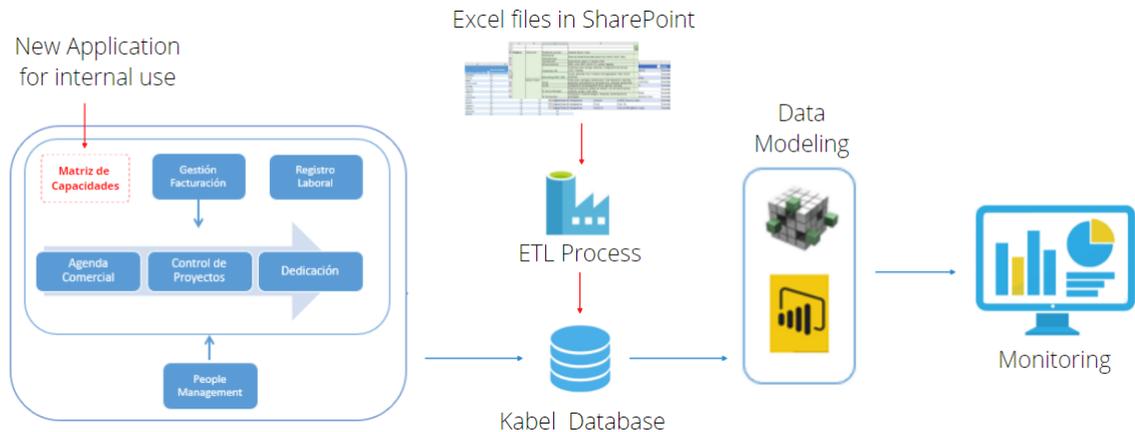


Illustration 2 – Results diagram

5. Conclusions

As a result, we have provided the company with a tool for decision-making regarding allocation of resources and team management, thus optimizing the candidate selection process that best fits the technical capabilities required for each project.

In order to achieve this goal, great attention has been paid, firstly, to the design of validation processes that ensure the data quality. Furthermore, the change log has been stored to enable its lineage. In addition, data security has been considered by defining access roles. In conclusion, a significant contribution has been made in terms of data democratization, making the information available to all users. All this through an application that represents the starting point to suggest initiatives that would eventually lead to future enhancement of this project.

6. References

- [1] Fawcett, T. “Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking”. July 2013. O’Reilly Media.
- [2] Dama International. “DAMA-DMBOK: Guía del Conocimiento Para la Gestión de Datos”. Julio 2020. Technics Publications.
- [3] Babcock, C. “Management Strategies for the Cloud Revolution. How Cloud Computing Is Transforming Business and Why You Can’t Afford to Be Left Behind”. May 2010. McGraw-Hill.
- [4] Sun, J. “Magic Quadrant for Analytics and Business Intelligence Platforms”. February 2021. Gartner.

Índice de la memoria

Capítulo 1. Introducción	6
1.1 Definición del problema.....	6
1.2 Motivación del proyecto.....	7
1.3 Estructura de la memoria.....	8
Capítulo 2. Marco teórico.....	10
2.1 Gobierno del dato	10
2.1.1 Definición de gobierno del dato	10
2.1.2 Ámbitos de actuación	11
2.1.3 Data Management Association (DAMA).....	13
2.1.4 Ventajas y retos	16
2.2 Computación en la nube	17
2.2.1 Definición de cloud.....	18
2.2.2 Tipos de cloud.....	20
2.2.3 Principales proveedores.....	23
2.3 Soluciones tecnológicas actuales.....	26
2.3.1 Herramientas de gestión de bases de datos.....	26
2.3.2 Herramientas de integración de datos	28
2.3.3 Herramientas de inteligencia de negocios	29
Capítulo 3. Definición del Proyecto	30
3.1 Alcance del proyecto	30
3.2 Objetivos	32
3.3 Especificaciones funcionales.....	33
3.4 Entorno de desarrollo	34
3.5 Recursos empleados	36
3.5.1 SQL Server Management Studio.....	36
3.5.2 Azure Data Factory	37
3.5.3 Power BI.....	39
3.6 Metodología.....	41
3.7 Planificación.....	42

Capítulo 4. Diseño e implantación de la solución	45
4.1 Exploración de datos	46
4.2 Almacenamiento de datos	49
4.3 Integración de datos.....	55
4.4 Mantenimiento de datos	57
4.4.1 API.....	58
4.4.2 UI.....	64
4.5 Análisis de datos.....	69
4.5.1 Modelo de datos	70
4.5.2 Creación de informes	73
4.6 Acceso a datos	80
Capítulo 5. Análisis de Resultados	82
Capítulo 6. Conclusiones y Trabajos Futuros	84
Capítulo 7. Bibliografía	87
ANEXO	89

Índice de figuras

Figura 1. Metas y principios del gobierno del dato [6]	11
Figura 2. Áreas de conocimiento del gobierno del dato según DAMA [6].....	14
Figura 3. Tipos de cloud según el modelo de desarrollo	21
Figura 4. Tipos de cloud según el modelo de servicio	22
Figura 5. Cuadrante mágico de Gartner sobre servicios de plataforma e infraestructura en la nube [16].....	24
Figura 6. Tendencias de búsqueda en Google sobre sistemas de gestión de bases de datos [17]	27
Figura 7. Tendencias de búsqueda en Google sobre herramientas de integración de datos [17]	28
Figura 8. Tendencias de búsqueda en Google sobre plataformas de análisis e inteligencia de negocios [17]	29
Figura 9. Ciclo de vida del dato considerado en el proyecto.....	30
Figura 10. Cuadrante mágico de Gartner sobre plataformas de análisis e inteligencia de negocios [18]	35
Figura 11. Comparación entre Power BI Desktop y Power BI Service [21].....	40
Figura 12. Diagrama de Gantt	43
Figura 13. Etapas del proyecto	45
Figura 14. Biblioteca de documentos en SharePoint.....	46
Figura 15. Contenido de la hoja Estructura del fichero Matriz Competencias - Infra	47
Figura 16. Contenido de la hoja Matriz del fichero Matriz Competencias - Infra	48
Figura 17. Contenido del fichero Matriz Competencias - Data.....	49
Figura 18. Modelo relacional diseñado en SQL Server.....	53
Figura 19. Procesos ETL con Azure Data Factory	55
Figura 20. Formulario de mantenimiento de capacidades.....	65
Figura 21. Formulario de validación de cambios	66

Figura 22. Formulario de actualizaciones por lotes.....	67
Figura 23. Formulario de peticiones pendientes.....	68
Figura 24. Herramientas de Power BI Desktop.....	69
Figura 25. Modelo de datos diseñado en Power BI.....	72
Figura 26. Panel matriz global.....	74
Figura 27. Panel creación de equipos	75
Figura 28. Panel detalles de consultor	77
Figura 29. Panel staffing avanzado	78
Figura 30. Linaje de los datos en el espacio de trabajo de Power BI	81
Figura 31. Diagrama de resultados	83

Índice de tablas

Tabla 1. Transformaciones que se pueden utilizar en el flujo de datos en ADF.....	39
Tabla 2. Asignación de tareas según el perfil de consultor	42
Tabla 3. Tablas incorporadas al modelo de datos en SQL Server.....	50
Tabla 4. Modificación de tablas en SQL Server.....	54
Tabla 5. Valores del método Skills (PUT)	60
Tabla 6. Valores del método Skills (PATCH).....	60
Tabla 7. Valores del método SkillsValidations (POST).....	62
Tabla 8. Valores del método SkillUpdates (PUT).....	63
Tabla 9. Tablas incorporadas al modelo de datos en Power BI.....	71

Capítulo 1. INTRODUCCIÓN

El proyecto se desarrolla en la empresa Kabel Sistemas de Información, una consultora informática que pone a disposición de otras compañías soluciones para la transformación de sus negocios a través de la tecnología, y que ha experimentado un gran crecimiento en los últimos años.

Uno de los pilares fundamentales de cualquier empresa de consultoría es la gestión de proyectos, donde la información se ha convertido en un activo clave para garantizar una adecuada toma de decisiones, además de mejorar las operaciones de negocio y adquirir una ventaja competitiva frente a las demás organizaciones [1].

Para generar valor a partir de la información es necesario recopilar, integrar y analizar datos de fuentes muy diversas para transformarlos en conocimiento. Gracias a la nube, es posible almacenar la información de forma centralizada, disponer de la capacidad de cómputo necesaria para analizarla, monitorizar datos en tiempo real o compartirla en un entorno colaborativo, entre otros [2].

Además, definir un modelo de gobierno sobre todo el ciclo de vida del dato resulta indispensable para garantizar la calidad de la información, asegurar la protección de los datos sensibles, facilitar su accesibilidad o permitir su trazabilidad [3].

1.1 DEFINICIÓN DEL PROBLEMA

El recurso utilizado actualmente en la compañía para recoger la información referente a las capacidades técnicas de los consultores se basa en varios conjuntos de hojas Excel, cuyo contenido se muestra en el apartado 4.1. Estos ficheros se encuentran alojados en un repositorio de SharePoint, una herramienta de almacenamiento en la nube que permite compartir la información y garantizar un acceso seguro entre los usuarios de la organización.

INTRODUCCIÓN

Debido al gran desarrollo que ha experimentado la empresa en los últimos años el volumen de consultores que tiene que controlar cada área resulta inmanejable, por lo que el sistema actual ha quedado obsoleto.

Esta herramienta ya no permite gestionar y actualizar la información de forma sencilla, ni ofrece la funcionalidad deseada. La toma de decisiones a la hora de organizar los equipos se realiza con información que no es fiable ya que los datos son imprecisos, están desactualizados y no se encuentran organizados adecuadamente.

Además, al realizar esta acción de forma manual y no disponer de un proceso automatizado de validación de la información se incurre en varios problemas e inconvenientes, encontrándose entre ellos la falta de definición de una estrategia de gobierno del dato que permita profundizar en su tratamiento, garantizando su calidad y trazabilidad.

Hay que tener en cuenta que una mala administración de la información a la hora de gestionar los proyectos de la compañía puede traer consigo varias repercusiones negativas, como pueden ser costes innecesarios, pérdidas de rentabilidad, incumplimiento de los plazos previstos o equipos de trabajo mal estructurados, entre otras.

1.2 MOTIVACIÓN DEL PROYECTO

Para una gestión correcta se requiere conocer, por un lado, el grado de conocimiento con el que cuentan los consultores en las diversas tecnologías y herramientas y, por otro, el uso que hacen de ellas en su día a día, para poder organizar los equipos y analizar si la capacitación de los técnicos se ajusta a la demanda actual de los clientes.

Ante los problemas descritos en el apartado 1.1, se propone como solución una nueva herramienta que permita llevar un control y gestión de la información relacionada con las capacidades técnicas de cada consultor conforme a las necesidades actuales de la empresa. La herramienta tiene tanto que acoger, de forma eficiente, la información actual, como prever el volumen creciente de datos que permita a la dirección poder tomar decisiones más ajustadas a la realidad, en base a la experiencia interna de la compañía.

Además, se ha detectado que los datos referentes a la asignación de consultores a proyectos se encuentran almacenados en la base de datos de la empresa, pero no se obtiene ninguna utilidad de esta información a la hora de tomar decisiones.

Por ello, en este proyecto también se abordará la implantación de otra herramienta de uso interno que permita obtener información sobre los consultores de una manera rápida y clara, la cual hará referencia a su grado de ocupación en los distintos proyectos a los que se encuentra asignado.

1.3 ESTRUCTURA DE LA MEMORIA

El informe se ha dividido en 7 capítulos, cuyo contenido se presenta a continuación:

El capítulo 1 ofrece una breve introducción y define el contexto del proyecto, en base al cual se explican las motivaciones que justifican su realización.

El capítulo 2 presenta una amplia base teórica, donde se exponen los conceptos necesarios para la comprensión del proyecto. En primer lugar, se introduce una presentación de los fundamentos del gobierno del dato. En segundo lugar, se aclara el término de *cloud computing*, definiendo los diferentes modelos de despliegue y niveles de servicio, además de las ventajas y retos de esta tecnología emergente, y se exponen los principales proveedores que existen el mercado. En tercer lugar, se clasifican las principales herramientas que permitirán abordar las diferentes cuestiones planteadas a lo largo del proyecto.

El capítulo 3 se centra en la definición del alcance del proyecto, que tiene como finalidad expresar los objetivos que se persigue alcanzar y establecer los requerimientos. A continuación, se describe el entorno de desarrollo y se realiza un análisis de las herramientas y recursos a emplear. Para finalizar, se explica cómo se va a resolver cada una de las cuestiones planteadas y qué técnicas y procedimientos se utilizarán, además de incluir una planificación de las tareas a realizar.

INTRODUCCIÓN

El capítulo 4 profundiza en el diseño e implantación de la solución. Se muestran en detalle los procedimientos seguidos en cada una de las fases del proyecto; desde la exploración de datos, pasando por la adaptación del modelo relacional para acoger en la base de datos de la compañía la información integrada a través de procesos de extracción, transformación y carga del dato, hasta el diseño del modelo analítico de datos para monitorizar la información. Además, se plantea la solución para el mantenimiento de la información presentando, a nivel de diseño, la API y la interfaz de usuario.

El capítulo 5 presenta un análisis de la solución propuesta, destacando los resultados más relevantes del proyecto y evaluando sus aportaciones. Se discuten los resultados obtenidos, para contrastarlos con los objetivos.

El capítulo 6 resume la conclusión sobre la solución propuesta y expone sus limitaciones proponiendo nuevas vías de investigación.

El capítulo 7 incluye una lista de las referencias bibliográficas de las fuentes o recursos en los que se basa el trabajo.

Capítulo 2. MARCO TEÓRICO

En este capítulo se documentan las consideraciones teóricas en las que se sustenta el proyecto. El gobierno del dato y las tecnologías *cloud* son conceptos necesarios para entender el planteamiento del problema, además se revisan las herramientas más relevantes que existen actualmente en el mercado para abordar la implementación de una solución adecuada a cada una de las cuestiones que se plantean en el proyecto.

2.1 GOBIERNO DEL DATO

La transformación digital evidencia la creciente importancia de contar con una estrategia de gobierno del dato adecuada, dado que los procesos con los que se obtienen, almacenan y gestionan los datos de una compañía resultan clave para garantizar un uso eficaz y eficiente de la información.

2.1.1 DEFINICIÓN DE GOBIERNO DEL DATO

El gobierno del dato es el conjunto de personas, procesos y tecnología que permiten a una compañía capturar, mantener y gestionar los datos de su actividad de forma fiable y eficiente. Esto a su vez permite asegurar la veracidad, integridad, confidencialidad y disponibilidad de los datos y conseguir que estos ayuden a la compañía a tomar las decisiones oportunas en el momento adecuado [4].

La definición de un modelo adecuado de gobierno ayuda a convertir el dato en el activo más valioso, situándolo en el centro de todas las decisiones estratégicas, focalizándose en cómo se toman las decisiones acerca de los datos y cómo se espera que las personas y los procesos se comporten en relación a ellos. Como cada organización es diferente, será necesario adaptar la estrategia a cada caso particular, pero manteniendo los objetivos comunes de definir quién puede emprender determinadas acciones, sobre qué datos, en qué situaciones y mediante qué métodos.

2.1.2 ÁMBITOS DE ACTUACIÓN

El gobierno del dato se basa en tres principios fundamentales: procesos (modelo operativo), personas (modelo organizativo) y tecnología (modelo de servicio y tecnologías). Una buena estrategia de gobierno tendrá en cuenta todos los ámbitos de actuación, desde los procesos de negocio, pasando por la tecnología utilizada, hasta la adopción de una cultura del dato que afectará en todos los niveles a la compañía [5].

En la figura 1 aparecen representadas las metas y principios en los que se basan estos tres puntos clave:

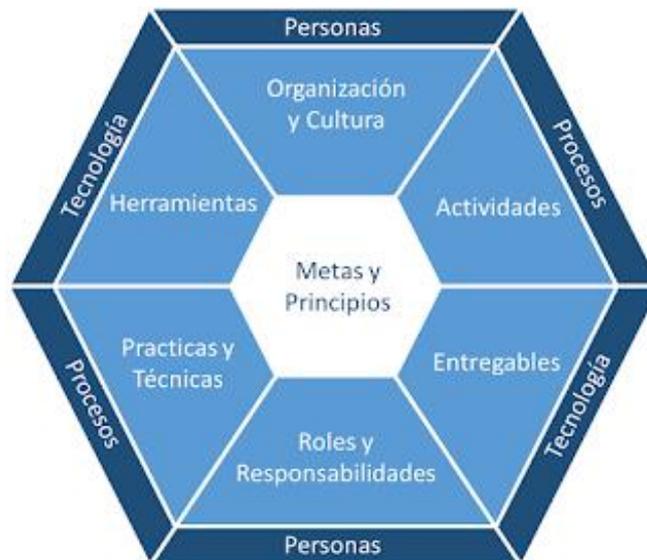


Figura 1. Metas y principios del gobierno del dato [6]

Estos principios se encuentran estrechamente relacionados entre sí, ya que las decisiones tomadas para cada uno de ellos, acabará afectando ulteriormente a las demás. Un ejemplo claro es que, a la hora de definir los procesos, delimitaremos parte de la organización y con ello también los roles del personal o su formación, lo cual será decisivo para el uso que los mismos hagan de la tecnología. A continuación, se explican detalladamente los tres modelos:

2.1.2.1 Modelo operativo

Un factor clave dentro del modelo operativo es el desconocimiento de los datos, es decir, las empresas no son conscientes de qué datos tienen y carecen de procedimientos y/o políticas para gestionarlos. A esto se suman la escasa automatización y el uso de modelos de gobierno basados en herramientas obsoletas que necesitarían pasar por un proceso de actualización para poder adaptarse a las nuevas necesidades. En resumen, sería necesaria una metodología que facilitase la operativa y el catálogo de servicios del dato.

En este sentido, sería esencial determinar ciertos procedimientos para el almacenamiento, categorización y protección de los datos, este último especialmente frente a posibles casos de ataques, robos u otros incidentes. Más en concreto, el gobierno del dato debería abarcar la gestión de riesgos, plan de recuperación en caso de catástrofe, seguridad, privacidad, integridad, respuesta ante incidentes, gestión de acceso y asunción de responsabilidades [7]. Asimismo, es indispensable una regularización del proceder por parte del personal autorizado (véase el apartado 2.1.2.2) junto con ciertos procedimientos y controles de auditoría que permitan garantizar las normas de gobierno.

2.1.2.2 Modelo organizativo

Las personas han de tenerse también en cuenta en la estrategia de gobierno, puesto que es esencial realizar un cambio de cultura en la cual los datos sean un activo transversal a toda la organización, y no sólo dependan de un único trabajador o de un solo departamento. El gobierno del dato ha de integrarse en los procesos y comportamientos diarios de manera que no se considere una disciplina independiente.

También es básico disponer de los empleados adecuados con la formación necesaria, además de definir las responsabilidades asignadas a personas o áreas dentro de la compañía. Dependiendo de la organización, un mismo individuo puede desempeñar varios roles y, aunque este no sea el caso, siempre será útil definir cada uno de ellos para facilitar una organización óptima y que, eventualmente, estos puedan ser asignados a otro trabajador. Los tres principales se enumeran a continuación [8].

- *Data steward* (administrador de datos): es el encargado de orquestar las necesidades y las acciones técnicas, auditando la calidad de los datos y el uso correcto de ellos, así como los aspectos relacionados con la confidencialidad y metadatos.
- *Data owner* (propietario de datos): es quien tiene la autoridad para decidir sobre el acceso y el uso de los datos. Se preocupa por el acceso adecuado a los datos, es el encargado del direccionamiento estratégico para gobernar el valor y uso del dato, y de corregir la información en caso necesario.
- *Data custodian* (custodio de datos): es el responsable del entorno técnico y la estructura de la base de datos, es decir, se preocupa por quién puede acceder a los datos, administra servidores, copias de seguridad o redes.

2.1.2.3 Modelo de servicios y tecnología

Es importante establecer un ciclo de vida del dato gobernado diseñando una arquitectura de datos que muestre el esquema de componentes y estándares para lograr una gestión eficiente de datos. De esta manera, se podrán definir las etapas por las que pasará el dato, los procesos y tecnologías que hay que tener en cuenta en cada una de esas etapas y prestar especial atención a los metadatos para poder entender y automatizar los procesos.

Las principales herramientas tecnológicas para un gobierno del dato adecuado comprenderán un diccionario de datos, que permita establecer un repositorio y lenguaje único que evite confusiones, y la existencia de más de una definición para un mismo indicador. También se contará con soluciones para medir, mejorar y asegurar la calidad de los datos y para implantar medidas que garanticen la privacidad de los datos teniendo en cuenta el cumplimiento del Reglamento General de Protección de Datos y otras normativas [9].

2.1.3 DATA MANAGEMENT ASSOCIATION (DAMA)

Con el objetivo de definir los principios del Gobierno del Dato, el presente trabajo se basa en las directrices marcadas por DAMA (Data Management Association). Esta es una organización sin ánimo de lucro con base en Washington, cuya misión es sensibilizar y formar a los trabajadores dentro de la industria y la profesión de la gestión de datos.

Entre los fundamentos de su misión, también se encuentran el avance en la concepción y las prácticas de la gestión de información y de datos. Respecto a su visión, consiste en representar un recurso esencial para aquellos dedicados a este campo. Entre sus objetivos, cabe mencionar los siguientes [3]:

- ayudar a los usuarios a mejorar sus conocimientos y habilidades en esta profesión,
- tener influencia sobre las prácticas, la educación y la certificación en este campo,
- apoyar a los miembros de DAMA y a sus organizaciones a cubrir sus necesidades y
- establecer alianzas con unos principios similares para fortalecer así esta profesión.

Dentro de la asociación DAMA, se define brevemente qué es DAMA España. En concreto, es el capítulo español, activo desde marzo de 2019, que sirve como referente para trabajadores y organizaciones y tiene la misma misión que la organización internacional, solo que dentro del contexto español. Entre sus objetivos, se encuentran promover la formación en territorio nacional, tanto en centros u organismos educativos, como en empresas, y también la oferta de recursos y de programas de formación. Finalmente, fomentar las alianzas con otras organizaciones y establecer el DMBok como marco de referencia para el gobierno del dato en España [10].



Figura 2. Áreas de conocimiento del gobierno del dato según DAMA [6]

En la figura 2 se encuentran las diez áreas de conocimiento del gobierno del dato que establece DAMA y que aparecen desarrolladas brevemente a continuación [3].

- La arquitectura abarca la estructura general de los mismos y sus recursos, definida en base a las propias necesidades de la empresa.
- El modelado y diseño consisten en las tareas de análisis, diseño, construcción, pruebas y mantenimiento de los procesos a través de los que se cargan los datos en el sistema de almacenamiento. El objetivo es desarrollar soluciones para satisfacer las necesidades de la empresa.
- El almacenamiento y la operación se relacionan directamente con la protección, la disponibilidad y el rendimiento de los datos, así como con los mecanismos de despliegue y la administración de procesos de carga.
- La seguridad se basa en la definición de los requisitos de seguridad y privacidad que permiten controlar el acceso, la confidencialidad y la integridad de la empresa.
- La integración y la interoperabilidad sirve para analizar si los datos cumplen con los requisitos de seguridad previamente mencionados y si estos son apropiados para su uso previsto. Aquí, se deben tener en cuenta también los procedimientos para compartir, replicar y consolidar datos ya existentes y para limpiar, transformar, integrar y enriquecer otros nuevos.
- La documentación y el contenido se utilizan para almacenar, proteger, indexar y permitir el acceso a datos que se encuentran en fuentes no estructuradas y hacer que estos datos estén disponibles para la integración e interoperabilidad con datos estructurados.
- Las referencias y datos maestros se concentran en la reducción de redundancias y tienen como objetivo mejorar la calidad de los datos y promover su reutilización.
- Las bodegas de datos y la inteligencia de negocios se centran en facilitar el acceso a los datos con el fin de que puedan ser utilizados para el análisis y la redacción de informes, convirtiéndose en una herramienta clave a la hora de tomar decisiones.

- Los metadatos describen características de los datos y se encuentran almacenados en repositorios que tienen que ser supervisados y que facilitan la gestión de todo el sistema.
- La calidad debe medirse y evaluarse a través de métricas que permitan mejorar la adecuación de dichos datos para su uso.

2.1.4 VENTAJAS Y RETOS

La mayoría de las empresas ya tienen alguna forma de gobierno del dato instaurada en su organización, ya sea individualmente o en un departamento de negocio, aunque no esté institucionalizada de manera integral. Por lo tanto, la introducción sistemática del gobierno del dato es a menudo una adaptación de reglas informales a un control formal.

Este gobierno del dato normalizado formal se implementa, normalmente, una vez dicha empresa ha alcanzado un tamaño en el que las tareas transversales entre departamentos ya no se pueden implementar de manera eficiente [11]. Por ello, este modelo es necesario para numerosas tareas o proyectos, ya que aporta muchos beneficios como:

- Una mayor seguridad de los datos tanto internos como externos, mediante el monitoreo y revisión de políticas de privacidad.
- Aumento de la calidad de los datos, gracias a la definición de reglas, monitorización y corrección de datos.
- Mayor eficiencia del proceso, a través de la automatización y la reducción de los largos procesos de coordinación.
- Capacidad de incrementar la escalabilidad del panorama de IT, a nivel técnico, comercial y organizacional mediante reglas claras para cambiar procesos y datos.
- Optimización del coste de la gestión de datos, al aumentar los mecanismos de control.
- Comunicación clara y transparente a través de la estandarización. Este es el requisito previo para las iniciativas centradas en datos en toda la empresa.

- Una mejor democratización de datos que permita al mayor número posible de empleados acceder a la mayor cantidad posible de información, aportando así transparencia y confianza.

A pesar de estas ventajas, muchas empresas temen implementar programas de gobierno del dato, ya sea por complejidad o incertidumbre. La implementación de este tipo de proyectos no es de ninguna manera una tarea trivial y enfrenta numerosos obstáculos, entre los que destacan:

- La organización, principalmente porque se deben nombrar nuevos roles y responsabilidades, es decir, es necesario reformar la política actual de la empresa.
- La comunicación y aceptación, no solo por parte de los empleados, sino también de los accionistas. Se debe tener en cuenta que es necesario convencer a los accionistas de la organización de la necesidad de un gobierno del dato y destinar una parte del presupuesto a implementarlo. Por otro lado, los encargados del proyecto deben estar informados sobre el nuevo panorama y tienen que ser capaces de comprender tanto los aspectos técnicos y comerciales de los cambios, como la nueva visión conceptual de la empresa.
- La estandarización y flexibilidad, puesto que es necesario mantener un equilibrio adecuado entre la flexibilidad de adaptarse a nuevos cambios y los estándares del gobierno del dato de acuerdo con los requisitos comerciales de cada empresa.

2.2 COMPUTACIÓN EN LA NUBE

El término computación en la nube tiene como núcleo los siguientes elementos: los servicios de computación se brindan a través de internet, bajo demanda y desde una ubicación remota. Para una organización, esto significaría que, por una tarifa fija o variable basada en el uso, o incluso posiblemente gratis, podría contratar a un proveedor los recursos necesarios tanto de infraestructura, como de plataforma o aplicaciones. Estos se ofrecerían como servicios a través de internet, de forma que los usuarios ya no necesitarían tener un conocimiento o control sobre la infraestructura tecnológica.

Además, esto permite que las tareas informáticas y la información estén disponibles en cualquier momento, en cualquier lugar y desde cualquier dispositivo, siempre que haya acceso a internet, sin importar dónde se aloje la información ni dónde se realice su procesamiento [12].

2.2.1 DEFINICIÓN DE *CLOUD*

El término *cloud* o nube, según su traducción al castellano, se utiliza para referirse a una entidad que no es física, sino a un recurso computacional que se encuentran accesible a los usuarios a través de internet. El *cloud* está formado por una red interconectada de servidores remotos que funcionan como un único ecosistema.

Aunque el uso principal de la nube es el almacenamiento de datos, también se utiliza para administrar los datos almacenados, para ejecutar aplicaciones que necesiten una gran capacidad de procesamiento o para entregar todo tipo de contenido o servicio [13].

La computación en la nube se caracteriza principalmente por:

- La accesibilidad a través de internet, que permite el acceso desde cualquier dispositivo conectado a la red, sin importar el lugar en el que se encuentre el usuario.
- El autoservicio bajo demanda, ya que permite al cliente gestionar y alterar los recursos de los que dispone, sin involucrar al proveedor de servicios. De esta forma, se optimizan el tiempo y los costes empleados.
- La escalabilidad y rápida elasticidad, que posibilita la ejecución rentable de altas cargas de trabajo que requieren una gran cantidad de servidores, pero solo durante un período breve de tiempo.
- La medición y comunicación del servicio, para disponer de una monitorización constante del uso de los recursos.
- La puesta en común de recursos o *Resource pooling*, con el objetivo de proveer a múltiples clientes con servicios provisionales y escalables. De esta forma, los servicios se pueden ajustar a las necesidades del cliente sin que este perciba los cambios.

Las características mencionadas hacen que la tecnología *cloud* presente una gran variedad de ventajas, siendo las más notables la reducción de costes, ya que se ahorra en *hardware*, *software*, soporte y seguridad; y la alta flexibilidad para escalar los recursos de manera fácil y rápida.

Por tanto, se puede decir que la computación en la nube es un gran avance tecnológico que está ayudando a aumentar la productividad y la eficiencia tanto en el sector privado como en el público. Los diferentes beneficios que aporta a la sociedad son:

- Nuevos productos y servicios, que han aparecido gracias a la reducción de costes, ya que antes no eran rentables.
- Aumento de la productividad, permitiendo la posibilidad de trabajar desde cualquier lugar de forma *online*, aumentando así la flexibilidad de los empleados para trabajar a distancia.
- Trabajo colaborativo, que proviene de la opción que ofrece la computación en la nube para que diferentes personas trabajen sobre el mismo documento o herramienta a la vez.
- Innovación, como ya se ha mencionado anteriormente, existe una gran reducción de costes en las empresas, lo que les permite diversificar sus esfuerzos e invertir en innovación.
- Reducción de la obsolescencia tecnológica, permitiendo a todos los usuarios el acceso a la última tecnología.
- Aumento de la eficiencia, ya que, en caso de aparición de errores o fallos, los proveedores *cloud* ofrecen soporte de manera rápida y eficiente al conocer completamente la infraestructura.

A pesar del gran número de ventajas que aporta la computación en la nube, es necesario conocer también los principales retos o riesgos que conlleva su uso, para poder realizar una valoración completa de específicos ante los que se enfrenta este paradigma y así poder valorar como un conjunto este recurso.

A continuación, se enumeran varios de estos retos:

- Seguridad y privacidad es uno de los aspectos más críticos desde el punto de vista de las empresas, debido a que se produce una transmisión de información confidencial al *cloud* y con ello una pérdida de control sobre la misma.
- Disponibilidad para mantener los niveles de servicio acordados en todo momento. Esta nueva tecnología tiene que ser capaz de asegurar el servicio durante todos los procesos del cliente.
- Estandarización para que exista una estructura de datos estándar que facilite la migración de datos y aplicaciones y la integración con otros subsistemas de negocio. Todavía existe una gran variedad de formas de trabajar en entornos *cloud*, que dificulta enormemente el cambio de proveedor.
- Adaptabilidad que permite adecuar las diferentes aplicaciones usadas por el cliente a la nube y trabajar de forma fluida y paralelamente con otras aplicaciones, maximizando así el rendimiento. También ha de tenerse en cuenta que, en caso de aparecer nuevas versiones, estas deben ser lo más estándares posibles para que el cambio se note lo mínimo posible.

2.2.2 TIPOS DE CLOUD

En función del parámetro elegido existen varias formas de clasificar las diferentes soluciones de computación en la nube. En este trabajo se van a exponer las dos principales, en función del modelo de desarrollo y en función del modelo de servicio [14].

2.2.2.1 Clasificación según el modelo de desarrollo

La primera forma de clasificar los distintos tipos de *cloud* se basa en el modelo de desarrollo para la computación en la nube, teniendo en cuenta que influye principalmente en la propiedad del servicio, es decir, en a quién pertenezcan los recursos empleados y la información almacenada en la nube. De esta manera, la privacidad es el parámetro más importante de la clasificación. Las opciones disponibles en el mercado se presentan en la figura 3.



Figura 3. Tipos de cloud según el modelo de desarrollo

- Nube pública: es un entorno operado por un proveedor de servicios al cual tienen acceso diferentes clientes. Se paga en función del uso, por lo que tiene la ventaja de no necesitar una gran inversión de capital, de reducir los gastos del departamento de IT de la empresa y de ser un servicio flexible y escalable. Las desventajas más destacables de esta elección son la falta de personalización, un desafío mayor en el gobierno del dato y la posibilidad de aparición de periodos de inactividad.
- Nube privada: al contrario que en el caso anterior, el entorno *cloud* pasa a ser utilizado por una sola empresa, la cual tiene control completo sobre su estructura y configuración. De esta forma se puede adaptar totalmente a las necesidades de la empresa, manteniendo una alta seguridad y productividad. El inconveniente principal es el aumento de los costes, tanto iniciales, como operativos.
- Nube híbrida: por último, esta opción permite tener lo mejor de la nube pública y de la privada, mediante la combinación de ambos entornos. Por lo tanto, estamos ante una solución económica, flexible y altamente personalizable. En contra, cabe destacar la disminución de la seguridad, debido al mayor recorrido que tiene que atravesar la información y la necesidad de que exista compatibilidad entre las diferentes herramientas.

Finalmente, cabe mencionar que existe otro modelo menos extendido, la nube comunitaria, en la cual varias organizaciones con un mismo objetivo y que comparten las mismas preocupaciones realizan un esfuerzo colaborativo y comparten la misma infraestructura.

2.2.2.2 Clasificación según el modelo de servicio

La otra forma de clasificar los distintos tipos ofertados de computación en la nube es en base al servicio ofrecido. Dentro de esta categoría, existen tres grandes grupos que se conocen por sus siglas en inglés, IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and SaaS (Software as a Service), y que se presentan en la figura 4, clasificados según los servicios proporcionados, así como según el nivel de control sobre la información.

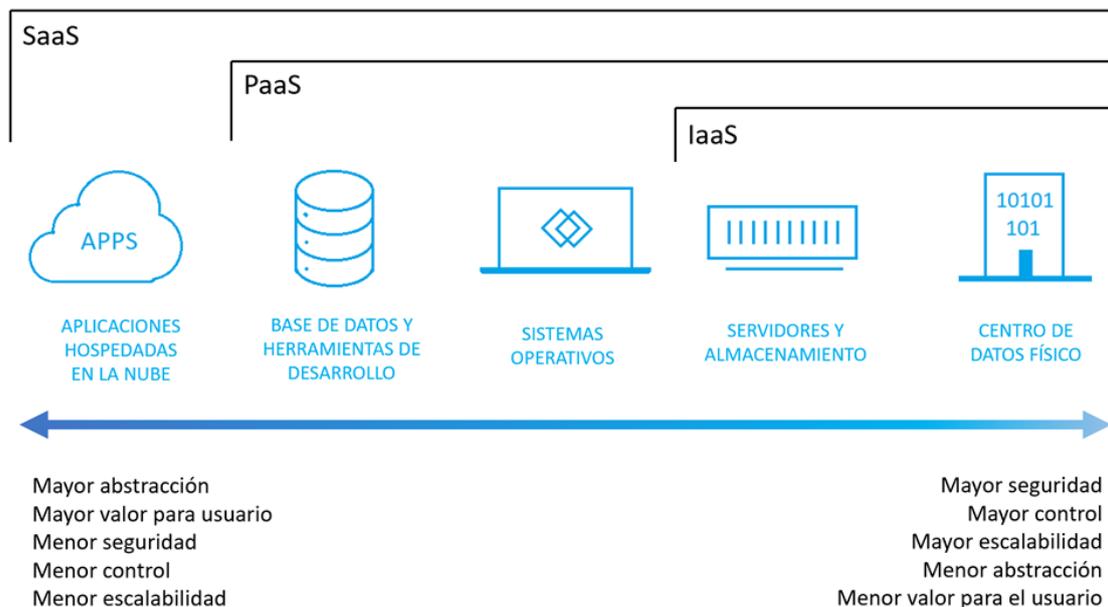


Figura 4. Tipos de cloud según el modelo de servicio

- SaaS, o software como servicio: consiste en ofrecer productos finales, es decir, las aplicaciones las administra un proveedor externo y el cliente las ejecuta desde un navegador *web* sin necesidad de descargas o instalaciones. Debido al modelo de entrega *web*, SaaS elimina la necesidad de instalar y ejecutar aplicaciones en computadoras personales. Con esta solución, es fácil para las empresas optimizar su mantenimiento y soporte, porque todo puede ser administrado por los proveedores.

Los tipos populares de ofertas de SaaS incluyen correo electrónico y colaboración, gestión de relaciones con los clientes y aplicaciones relacionadas con la atención médica. Algunas grandes compañías que tradicionalmente no se consideran proveedores de *software* han comenzado a construir SaaS como una fuente adicional de ingresos para obtener una ventaja competitiva.

- PaaS, o plataforma como servicio: se utiliza para aplicaciones y otros desarrollos, al tiempo que proporciona componentes de nube al *software*. Esto quiere decir que ofrece las herramientas para desarrollar o personalizar aplicaciones. PaaS hace que el desarrollo, las pruebas y la implementación de aplicaciones sea rápido, simple y rentable. Con esta tecnología, las operaciones empresariales o un proveedor externo pueden administrar sistemas operativos, virtualización, servidores, almacenamiento, redes y el propio *software* PaaS.
- IaaS, o infraestructura como servicio: es un modelo de autoservicio para acceder, monitorear y administrar infraestructuras de centros de datos remotos. En lugar de tener que comprar hardware directamente, los usuarios adquieren almacenamiento y recursos computacionales que usarán para desarrollar su propio *software*. Los proveedores aún administran la virtualización, los servidores, los discos duros, el almacenamiento y las redes. Lo que los usuarios obtienen con IaaS es una infraestructura sobre la cual pueden instalar cualquier plataforma requerida, sin necesidad de comprar servidores, espacio de almacenamiento y equipos de red.

2.2.3 PRINCIPALES PROVEEDORES

Hasta ahora se han clasificado los tipos de *cloud* en función de dos parámetros, pero hay que tener en cuenta que las soluciones tecnológicas actuales están evolucionando rápidamente y una tecnología puede pertenecer a diferentes grupos a la vez.

Además de seleccionar el tipo de implementación y la clase de servicio deseado, también es necesario plantear la cuestión sobre qué proveedor de servicios elegir, ya que en los últimos años han aparecido multitud de empresas que ofrecen servicios en la nube [15].

Una de las mejores formas de comparar las diferentes soluciones tecnológicas que se encuentran actualmente en el mercado es utilizar el cuadrante Mágico de Gartner, desarrollado por la consultora Gartner, tras realizar un estudio exhaustivo de los diferentes competidores para un mercado específico.

Para entenderlo es necesario saber que el eje X representa la integridad de visión, que hace referencia a la innovación del proveedor y a su visión de cómo se desarrollará el mercado, mientras que el eje Y indica la capacidad de ejecución, es decir, la habilidad del proveedor para llevar a cabo su visión. De esta manera, el cuadrante queda dividido en cuatro secciones, clasificando cada una de las empresas como líderes, retadores, visionarios o jugadores de nicho.

En la figura 5 se evalúan los proveedores de servicios en la nube que ofrecen plataforma e infraestructura como servicio, donde se observa claramente como Amazon, Microsoft y Google, siguiendo este orden, son los líderes indiscutibles en el mercado.



Figura 5. Cuadrante mágico de Gartner sobre servicios de plataforma e infraestructura en la nube [16]

2.2.3.1 Amazon Web Services

En el año 2006, como pionera en la nueva tecnología *cloud*, Amazon decidió poner a disposición de los clientes sus servidores, para que estos pudieran aprovechar los recursos de la empresa. A esta nueva línea de negocio la llamó Amazon Web Services y en ella se ofrece actualmente tanto PaaS como IaaS, permitiendo a los usuarios multitud de opciones entre las que destacan la potencia informática y la cantidad de almacenamiento. Como AWS cuenta con 81 zonas de disponibilidad repartidas en 25 regiones geográficas de todo el mundo, tiene un alcance mayor y permite a los clientes mantener su información en su zona.

También cuenta con servicios adicionales incorporables a aplicaciones existentes o utilizables de forma independiente. Estos servicios no suelen estar a disposición de los usuarios finales, sino que se usan para desarrollar un *software* más adaptado a estos usuarios.

Una de las partes esenciales de AWS es Amazon EC2, con el que se puede disponer de equipos virtuales que permitan alojar aplicaciones sin tener que adquirir equipos dedicados. Basándose en la virtualización, pone a disposición de los usuarios diferentes sistemas operativos adaptables, en los que se añaden a las características básicas de la computación en la nube, la recuperación de datos y la separación de procesos en instancias. Como ejemplo se encuentran las instancias de uso general, las de almacenamiento o las de GPU.

2.2.3.2 Google Cloud Platform

La empresa Google utiliza Google Cloud Platform para ofrecer también servicios de PaaS e IaaS. Esta plataforma de computación en la nube puede ser usada tanto para desarrollar aplicaciones, como para alojarlas mediante el uso de un gran número de programas que son capaces de crear, desde simples páginas *web*, hasta complejas aplicaciones.

Google ha dividido su plataforma en un conjunto de servicios modulares para facilitar su uso y con ello ofrecer multitud de servicios, entre los cuales destacan App Engine y Compute Engine como PaaS e IaaS respectivamente; Cloud Storage y Cloud Datastore como almacenamiento; y servicios Big Data y APIs de predicción y traducción.

Gracias a las herramientas de las que dispone la solución PaaS de Google (App Engine), el usuario puede llegar a formar un SaaS que se aloje en la misma plataforma. También es capaz de soportar diferentes lenguajes de programación y *frameworks*. Esta plataforma libera al usuario de la preocupación por la administración, configuración, balanceo, etc. de sus servidores o bases de datos.

2.2.3.3 Microsoft Azure

Por último, Microsoft ha desarrollado Microsoft Azure, la cual es una IaaS que usa sus propios centros de datos. Azure trabaja como una capa en la nube sobre Windows Server, los servidores de Microsoft. Esta capa es la que permite escalar los recursos y administrar la información de la aplicación *web* del usuario. Otra de las grandes ventajas de esta plataforma es el sistema de copias de seguridad automático que, en caso de pérdida, permite recuperar la información, ya que esta se encuentra cifrada en los servidores.

Además, Windows Azure se divide en una serie de componentes que, al igual que los ya comentados, proporcionan almacenamiento, bases de datos, capacidad de cómputo, mercado de aplicaciones y diferentes servicios para las aplicaciones.

2.3 SOLUCIONES TECNOLÓGICAS ACTUALES

En este apartado se realiza un análisis de los recursos disponibles que se encuentran actualmente en el mercado para plantear y resolver cada una de las tareas necesarias del proyecto.

2.3.1 HERRAMIENTAS DE GESTIÓN DE BASES DE DATOS

Las herramientas de gestión de bases de datos se pueden definir como un *software* para crear y administrar los datos que se encuentran almacenados en las mismas. De esta forma, se proporciona a los usuarios y programadores un proceso definido para la recuperación, gestión, actualización y creación de datos.

El *software* de administración de bases de datos también mantiene los datos protegidos y seguros. Estas herramientas ayudan a reducir la redundancia de datos y a mantener su eficiencia. Según el uso y los requisitos del proyecto, habrá que elegir una herramienta de *software* que disponga de las características necesarias para obtener el resultado deseado.

En este proyecto, se va a gestionar una base de datos alojada en SQL Server. Aunque el servidor en el que se encuentra es de Microsoft, no es necesario usar una herramienta del mismo proveedor para su gestión, por lo que existen multitud de opciones. Las más utilizadas son MySQL Workbench, desarrollado por Oracle, DBeaver, con un *software* de código abierto y las principales herramientas de los tres grandes proveedores estudiados en el apartado 2.2.3, que son Microsoft SQL Server Management Studio, Amazon RDS y Google Cloud SQL.

Para comparar la popularidad de las diferentes herramientas se ha utilizado Google Trends, una herramienta de uso gratuito proporcionada por Google que muestra de forma gráfica la frecuencia de los términos de búsqueda seleccionados en un determinado periodo de tiempo para una región del mundo concreta. La figura 6 muestra la evolución de las tendencias de búsqueda en España en los últimos doce meses de tres de las aplicaciones más utilizadas para administrar una base de datos. De aquí se puede observar que MySQL Workbench es la herramienta que cuenta con una mayor acogida, seguida de DBeaver y de SQL Server Management Studio.

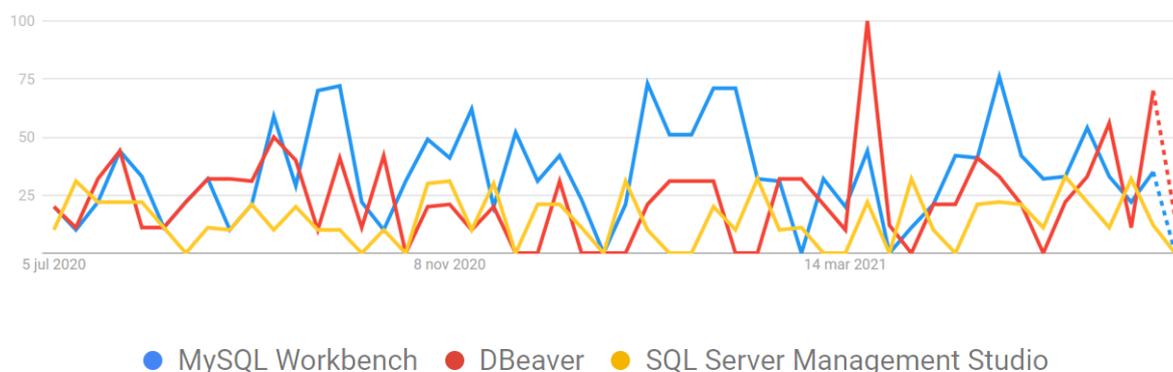


Figura 6. Tendencias de búsqueda en Google sobre sistemas de gestión de bases de datos [17]

2.3.2 HERRAMIENTAS DE INTEGRACIÓN DE DATOS

Por otro lugar, será necesario realizar una migración de datos de varios conjuntos de hojas de Excel a una base de datos de Microsoft SQL Server. Para ello, se recurre a una herramienta de integración de datos, cuya utilidad es llevar datos provenientes de diferentes fuentes de información a un solo destino, realizando al mismo tiempo un mapeo, transformación y limpieza de estos datos.

Este proceso también se conoce como ETL (por sus siglas en inglés *Extract, Transform and Load*) y hace referencia a las tareas de extracción, transformación y carga de datos, realizadas en este mismo orden. Se trata de un proceso fundamental para facilitar el movimiento de los datos, permitiendo recopilar información completa, fiable, correcta y actualizada.

Para realizar esta operación, disponemos de una gran oferta de aplicaciones, algunos ejemplos son Talend Data Integrator, Informatica o Pentaho, aunque también existen herramientas propias de los tres principales proveedores de servicios en la nube, como son Azure Data Factory, Microsoft SQL Server Integration Services, AWS Glue o Google Cloud Data Fusion. En la figura 7, se presenta una comparación en Google Trends del número de búsquedas de varias de estas herramientas en España, teniendo en cuenta los últimos doce meses. Aquí se puede observar que no hay una preferencia clara por una de estas soluciones, ya que las tres analizadas presentan una tendencia de búsqueda muy similar.

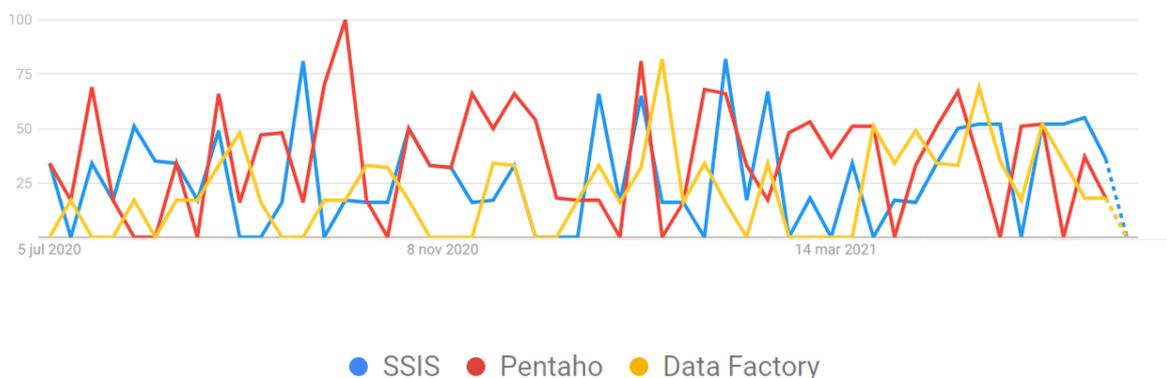


Figura 7. Tendencias de búsqueda en Google sobre herramientas de integración de datos [17]

2.3.3 HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS

Para cumplir con los objetivos del proyecto, también se utilizará una herramienta de inteligencia de negocios, o BI del inglés *Business Intelligence*. Estas herramientas se utilizan para recopilar, procesar, analizar y visualizar grandes volúmenes de datos pasados, actuales y futuros, con el fin de generar información empresarial utilizable, crear informes interactivos y simplificar los procesos de toma de decisiones.

Utilizar un *software* de este tipo aporta muchas ventajas, entre las que destacan reunir todos los datos importantes en un solo lugar, dar autoservicio permitiendo a los empleados acceder a la información sin necesidad de incluir al departamento de IT y extraer valor de la información para ayudar en la toma de decisiones con datos actualizados en tiempo real. Los beneficios de la inteligencia de negocios pueden ser alcanzados tanto en pequeñas como en grandes empresas y tienen como objetivo final convertir los datos de la empresa en conocimiento para obtener una ventaja competitiva frente a sus competidores.

Igual que para el resto de las soluciones, existe una gran variedad de oferta en el mercado de este tipo de herramientas. Sin embargo, existen tres opciones que cuentan con una mayor acogida que el resto, las cuales son Power BI, Tableau y Qlik. La figura 8 muestra la frecuencia de búsqueda de estas tres herramientas en España en los últimos doce meses, mostrando un claro predominio de Power BI frente a las demás.

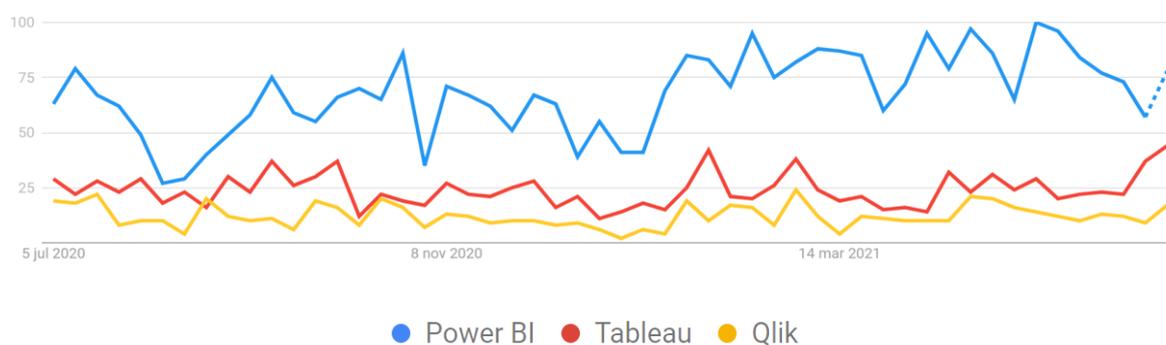


Figura 8. Tendencias de búsqueda en Google sobre plataformas de análisis e inteligencia de negocios [17]

Capítulo 3. DEFINICIÓN DEL PROYECTO

En este capítulo abordaremos las tareas previas al desarrollo del trabajo. El alcance del proyecto ayudará a definir la estrategia a seguir para alcanzar los objetivos, y las especificaciones, a ampliar en detalle cómo se van a cumplir los requerimientos establecidos. A continuación, se justificará el entorno de desarrollo utilizado y las herramientas empleadas para cumplir los objetivos. Para finalizar, se planteará la metodología de trabajo, necesaria para organizar la planificación.

3.1 ALCANCE DEL PROYECTO

El alcance del proyecto incluye todas las fases de la gestión del ciclo vida de los datos, desde su recopilación, hasta su almacenamiento, procesamiento y análisis, haciendo énfasis en los distintos aspectos relacionados con los datos. Estos últimos son, por ejemplo, el diseño de la arquitectura, los procesos que experimenta el dato, su forma de almacenaje o cómo serán puestos en valor a través de visualizaciones.



Figura 9. Ciclo de vida del dato considerado en el proyecto

En la figura 9 se muestran todas las fases que se tendrán en cuenta en el proyecto y que se describen brevemente a continuación.

DEFINICIÓN DEL PROYECTO

- 1) **Exploración:** en esta primera etapa se identifican los datos, que serán el punto de partida del proyecto. Se utilizará información interna de la empresa, que se encuentra recopilada en varios archivos Excel, y que recogen información sobre las capacidades técnicas de los trabajadores de la empresa.
- 2) **Almacenamiento:** se adaptará la arquitectura de la base de datos para integrar el contenido de los archivos Excel, además de los datos provenientes de la aplicación desarrollada para el mantenimiento y actualización de la información. Habrá que ajustar el modelo de datos actual, para integrar adecuadamente la nueva información volcada a la base de datos con los datos ya existentes procedentes de otras aplicaciones internas.
- 3) **Integración:** para solucionar el problema de consolidación de la información, se migrará la información existente en los archivos Excel a la base de datos actual de la compañía mediante procesos de extracción, transformación y carga.
- 4) **Mantenimiento:** para los procesos de validación y consolidación de los datos se implementará una aplicación que permita la interacción con el usuario en las tareas de actualización de la información. Como se ha tratado en el apartado 2.1, una buena estrategia de gobierno tendrá en cuenta todo el ciclo de vida de los datos. Gracias a esta aplicación, se pretende asegurar la calidad de los datos, ayudar en el proceso de linaje o trazabilidad mediante el almacenamiento del histórico, controlar el acceso o automatizar los procesos de actualización, entre otros.
- 5) **Análisis:** se utilizarán técnicas y herramientas de visualización que permitan comunicar gráficamente los resultados del análisis para una interpretación efectiva de la información por parte de cualquier usuario de la compañía. La información tratada será aquella referente a las capacidades técnicas de los consultores, que proviene de los datos de los archivos Excel, previamente integrados en la base de datos de la compañía. Adicionalmente, se analizará la información relacionada con la asignación de estos consultores a proyectos de clientes, cuya información se encuentra actualmente en la base de datos, pero no se extrae conocimiento de ella mediante técnicas de visualización.

- 6) **Acceso:** se creará un espacio de trabajo y una aplicación que permita compartir la información con los empleados de la compañía, administrando de forma adecuada los permisos de acceso requeridos según el rol de cada usuario.

3.2 OBJETIVOS

El objetivo principal del proyecto es la implantación de una solución que permita almacenar la información de forma centralizada, que facilite el mantenimiento de los datos de forma automática y sencilla y que favorezca la extracción de conocimiento para la toma de decisiones. Además, el trabajo se puede delimitar según los siguientes objetivos parciales, basados en la consecución de este objetivo principal.

- Dar solución a la problemática actual de introducción de la información de forma manual en diferentes formatos de archivos Excel, donde el margen de error y la falta de exactitud son elevados, además de la dificultad existente para mantener la información actualizada.
- Centralizar la información en la base de datos de la compañía mediante la integración de los datos a través de procesos de extracción, transformación y carga.
- Diseñar una aplicación de gestión interna de la empresa que aporte los instrumentos necesarios para el mantenimiento de la información y los ciclos de aprobación y actualización periódica requeridos.
- Implementar una herramienta de inteligencia de negocios, que permita agrupar, filtrar y visualizar la información referente a las capacidades técnicas de los consultores y a su grado de ocupación, para optimizar la gestión de proyectos a la hora de asignar los consultores a tareas y formar equipos, facilitando la tarea de selección del candidato que mejor encaja con las habilidades demandadas para cada proyecto.
- Crear un espacio de trabajo para compartir la información en cualquier dispositivo con los usuarios de la compañía, definiendo las pautas de acceso a los datos confidenciales.

3.3 ESPECIFICACIONES FUNCIONALES

Se inicia una primera fase de toma de requisitos, describiendo la funcionalidad esperada por parte de los principales usuarios de la herramienta. Todo ello para posteriormente abordar un análisis técnico que permita identificar las tecnologías a utilizar, con el fin de que la información recogida se integre con el resto de información, que actualmente se mantiene para otras aplicaciones de gestión interna.

Es importante definir los roles de los usuarios y sus responsabilidades para definir una estrategia de gobierno conforme a lo establecido en el apartado 2.1, por lo que el análisis funcional comienza mediante una recopilación de requisitos con los siguientes actores.

- Director de Producción & Staff: identificado como el principal actor de negocio y que asumirá el rol de *Customer Owner*.
- Kabel IT Area Manager: encargado del mantenimiento de las aplicaciones internas y con un amplio conocimiento del resto de aplicaciones.
- Area managers: encargados actualmente del mantenimiento de la información y que aportarán la experiencia adquirida y casos de uso identificados.

En cuanto a la definición de la funcionalidad de la aplicación, la dividiremos en dos ámbitos de actuación.

El primero hace referencia a las capacidades técnicas de los consultores en las diferentes tecnologías ofrecidas como servicio en la empresa, se denomina matriz de capacidades, y debe contar con los siguientes requisitos de alto nivel.

- Se traspasará toda la información actual a la nueva base de datos.
- Se creará una aplicación que permita la gestión y actualización de la información.
- Cada consultor será responsable de mantener su propia información.
- Los cambios deberán ser validados por su responsable.
- El consultor deberá revisar la información, como mínimo una vez al mes, al menos en algunas de las secciones.

DEFINICIÓN DEL PROYECTO

El segundo está relacionado con la asignación de los consultores a los proyectos de los clientes, se denomina *staffing*, y debe incluir los siguientes requerimientos de alto nivel.

- Permitir la monitorización del *staffing* de toda la compañía.
- Mostrar la asignación de horas por semana asociada a cada tarea de cada proyecto para cada consultor.
- Permitir la visualización de las asignaciones ya previstas de los proyectos en curso, filtrados por diferentes categorías, como área o cliente.
- Cruzar la información de *staffing* con la matriz de capacidades, para poder seleccionar siempre el candidato que mejor encaje con las habilidades técnicas demandadas.

3.4 ENTORNO DE DESARROLLO

El proyecto se desarrollará en un entorno *cloud* público, que evita la necesidad de una gran inversión inicial y de configurar y mantener servidores físicos, aunque se tendrá un menor control sobre la infraestructura.

En el apartado 2.2.3, se definen las tres soluciones de tecnología *cloud* más populares en la actualidad, entre las que se tienen en cuenta Amazon Web Services, Google Cloud Platform y Microsoft Azure como los proveedores candidatos para llevar a cabo las diferentes tareas definidas en el apartado 3.1. Esto se debe a que, tanto si se trata de soluciones IaaS, SaaS o PaaS, todos ellos ofrecen una gran variedad de productos y servicios.

Aunque en la empresa en la que se realiza el proyecto se utilizan recursos de todos los proveedores mencionados, mayoritariamente se ofrecen a otras empresas soluciones tecnológicas basadas en Azure, la plataforma *cloud* de Microsoft. Por lo tanto, la elección de Azure como el entorno *cloud* para conseguir cada uno de los objetivos planteados en el apartado 4.2, permitirá disponer de una amplia gama de productos para resolver cada una de las fases del proyecto y contar con el conocimiento de personas ampliamente formadas en esta tecnología para solventar los problemas que puedan ir surgiendo.

DEFINICIÓN DEL PROYECTO

Una de las consideraciones esenciales a la hora de tomar esta decisión ha sido la elección de la herramienta para el análisis y monitorización de la información. Como se puede observar en la figura 10, en febrero de 2021 la consultora Gartner reconoce a Microsoft como líder en plataformas de análisis de datos e inteligencia de negocios, patrón que se ha repetido por decimocuarto año consecutivo. Microsoft Power BI no sólo se sitúa al frente, sino que además lo hace con una ventaja notable respecto a sus competidores.



Figura 10. Cuadrante mágico de Gartner sobre plataformas de análisis e inteligencia de negocios [18]

Este reconocimiento, junto con su fácil integración con otros productos y el asesoramiento de personas ampliamente formadas en esta tecnología dentro de la compañía, han sido factores clave para elegir Azure como el entorno de desarrollo en la nube más adecuado para resolver todas las cuestiones que se vayan planteando.

3.5 RECURSOS EMPLEADOS

Tras determinar el entorno *cloud* sobre el que se va a desarrollar el caso de estudio, el siguiente paso es seleccionar las herramientas adecuadas entre las soluciones tecnológicas actuales que se plantearon en el apartado 2.3, así como los lenguajes de programación para cumplir con las especificaciones funcionales definidas en el apartado 3.3.

Una vez se ha aclarado la utilización de Microsoft Power BI como la plataforma de inteligencia de negocios a utilizar, la elección del resto de herramientas se basará en soluciones que se integren sin conflictos con dicha herramienta y que trabajen juntas para proporcionar una solución de BI completa en Azure.

3.5.1 SQL SERVER MANAGEMENT STUDIO

Para gestionar la base de datos alojada en Microsoft SQL Server, se empleará la herramienta de gestión y administración de base de datos SQL Server Management Studio (SSMS), ya que fue diseñada por Microsoft para administrar cualquier infraestructura de SQL, siendo completamente funcional para trabajar en esta base de datos [19].

SSMS es un sistema de gestión de bases de datos relacional para almacenar y gestionar toda la información proveniente de las aplicaciones de gestión interna de la empresa. Combina varias herramientas gráficas para acceder, configurar, administrar y desarrollar todos los componentes de SQL Server.

El lenguaje usado en esta herramienta es T-SQL (Transact-SQL). Este proporciona un poderoso conjunto de funciones para ejecutar comandos tanto de administración como de consulta. Es un lenguaje muy útil durante la manipulación de datos, ya que los trabajos ejecutados se procesan con una sobrecarga mínima cuando se transfieren dentro de la aplicación, minimizando de esta forma el tráfico en el servidor.

SSMS es muy popular y ampliamente utilizado por los desarrolladores y administradores de bases de datos debido a que es gratis, fácil de instalar, incluye el acceso a una alta variedad de componentes y existe una alta experiencia entre sus usuarios.

3.5.2 AZURE DATA FACTORY

Para los procesos de extracción, transformación y carga de datos se utilizará Azure Data Factory (ADF) como herramienta de integración de datos basada en la nube. Esta permite mover datos tanto estructurados como no estructurados, usando una interfaz de usuario sin código que favorece la creación intuitiva y una supervisión y administración desde un único panel. Aunque no proporciona tanta flexibilidad como un lenguaje completo, sigue permitiendo bastante flexibilidad para diseñar cualquier tipo de flujo de trabajo.

Esta solución proporciona un servicio de integración de datos sin servidor totalmente administrado. Esto significa que no hay necesidad de preocuparse por dónde se ejecutan las canalizaciones, administrar servidores o configurar redes. De forma que el cliente solo paga por lo que usa y permite escalar según sea necesario. Además, cuenta con más de noventa conectores permitiendo la ingesta desde muy diversos orígenes, sin mantenimiento y sin aumento de costes [20].

En ADF encontramos los siguientes recursos o componentes:

- Conjunto de datos: estructuras de los almacenes de datos, que hacen referencia a los datos que se desea utilizar como entrada o salida, admitiendo una gran variedad de tipos.
- Actividades: representan cada uno de los pasos del procesamiento de una canalización y admiten tres tipos de actividades, que son de movimiento de datos, de transformación de datos y de control.
- Canalización: definida como una agrupación lógica de actividades para realizar una actividad de trabajo.
- Servicios vinculados: definen la información necesaria para que ADF se pueda conectar a los recursos externos como origen de datos.
- Flujo de datos: permite realizar transformaciones de datos sin código y representarlas como un grafo dirigido ejecutable como una actividad a través de un pipeline.

DEFINICIÓN DEL PROYECTO

La tabla 1 resume las transformaciones disponibles en la versión actual de la herramienta, que se pueden agregar a un flujo de datos.

Tipo	Icono	Transformación	Descripción
Dataset		Source	Origen del flujo de datos
		Sink	Destino del flujo de datos
Multiple Inputs / Outputs		Join	Combinar datos de dos flujos en un mismo flujo en función de una condición
		Conditional Split	Enrutar las filas de datos a diferentes flujos en función de condiciones
		Exists	Filtrado de filas que comprueba si los datos existen en otro flujo
		Union	Combinar varios flujos de datos en uno
		Lookup	Buscar datos de otro origen en una secuencia de flujo de datos
Schema Modifier		Derived Column	Generar nuevas columnas en el flujo o modificar los campos existentes
		Select	Cambiar el nombre de las columnas, quitarlas o reordenarlas
		Aggregate	Agregaciones de columnas de flujos de datos
		Surrogate Key	Agregar un valor de clave incremental a cada fila de datos
		Pivot	Transformar valores de filas en columnas individuales
		Unpivot	Transformar valores de columnas en filas individuales
		Window	Crear agregaciones basadas en ventanas de d columnas basadas en ventanas en el flujo
	Rank	Generar una clasificación ordenada según las condiciones de ordenación	

DEFINICIÓN DEL PROYECTO

Formatters		Flatten	Desenrollar los valores dentro de matrices en filas individuales
		Parse	Tomar valores de matriz incluidos en estructuras jerárquicas y expandirlos en filas
Row Modifier		Filter	Filtrado de filas en un flujo en función de una condición
		Sort	Ordenar las filas entrantes de la secuencia de datos actual
		Alter Row	Establecer directivas de update, insert, upsert y delete en filas

Tabla 1. Transformaciones que se pueden utilizar en el flujo de datos en ADF

3.5.3 POWER BI

Para el análisis de la información se utilizará Power BI, la herramienta líder en inteligencia de negocios basada en la nube de Microsoft que permite integrar datos de diferentes fuentes y representar la información a través de elementos visuales, con el fin de interpretarlos y compartir instantáneas claras y útiles de lo que está sucediendo en la empresa para poder tomar decisiones basadas en datos [21].

Power BI es una colección de servicios de *software*, aplicaciones y conectores que funcionan conjuntamente para convertir datos de diferentes orígenes en información valiosa y compartirlos con los usuarios. De esta manera, está formado por dos componentes principales:

- Power BI Desktop: es la aplicación gratuita de escritorio para vincular o importar datos externos, transformarlos y limpiarlos, y crear colecciones de objetos visuales en informes, que proporcionan representaciones visuales de los mismos.
- Power BI Service: también conocido como Power BI en línea es el servicio *online* (SaaS) con una funcionalidad similar a la aplicación de escritorio y que permite publicar y editar informes y compartirlos con los usuarios.

DEFINICIÓN DEL PROYECTO

También existe la solución de Power BI Mobile, la aplicación móvil de Power BI, disponible para dispositivos Windows, iOS y Android, que permiten interactuar con los informes y compartirlos desde la aplicación en cualquier momento y desde cualquier lugar.

En la figura 11 se muestra un diagrama de Venn, donde se comparan los dos componentes principales de Power BI. Asimismo, se muestran las características únicas de cada uno y las tareas en común que se pueden realizar tanto en Power BI Desktop, como en Power BI Service.



Figura 11. Comparación entre Power BI Desktop y Power BI Service [21]

Mientras que Power BI Desktop es una herramienta completa de creación de informes y modelado de datos con conexión a orígenes de datos muy diversos, Power BI Service es un servicio basado en la nube que admite la edición de informes de forma limitada y cuenta con algunos conectores, aunque su funcionalidad principal es publicar informes y colaborar con otros usuarios.

El flujo de trabajo habitual en Power BI, en el que los tres elementos coinciden para ofrecer una solución integral, comenzaría con la conexión a los orígenes de datos en Power BI Desktop. Aquí se combinan en un modelo de datos y se diseñan los informes, y en Power BI Service se publican, de forma que los usuarios puedan interactuar con ellos a través de este servicio o a través de la aplicación móvil de Power BI Mobile.

DEFINICIÓN DEL PROYECTO

Los lenguajes de programación principales utilizados son M y DAX (por sus siglas en inglés *Data Analysis Expressions*) y se pueden usar para filtrar, administrar y visualizar datos. Ambos lenguajes son independientes, con lógicas y códigos subyacentes totalmente diferentes. M se puede considerar como un lenguaje de fórmulas de consulta utilizado para preparar los datos antes de cargarlos en el modelo de Power BI, mientras que DAX es un lenguaje de cálculo que se puede utilizar para un análisis de datos en profundidad durante la fase de visualización. Además, se pueden utilizar otros lenguajes de programación como Python y R para transformar los datos, o utilizar consultas SQL para importar datos.

A continuación, se presentan alguno de los motivos por los que Power BI es una excelente alternativa a la hora de elegir una herramienta de inteligencia de negocios.

- Integración con otras herramientas, contando con multitud de fuentes de datos, tanto en la nube como *On Premise*.
- Posibilidades infinitas de visualización, con 32 elementos visuales integrados y con acceso a un mercado dinámico de objetos visuales personalizados.
- Asequibilidad y accesibilidad, ofreciendo una buena relación calidad precio e incluso una versión gratuita, además de una gran capacidad de colaboración.
- Herramienta líder en el mercado, en pleno crecimiento y con actualizaciones constantes.

3.6 METODOLOGÍA

El presente trabajo forma parte de un plan más ambicioso que se está desarrollando como un proyecto interno de la empresa, con diferentes ámbitos que requieren de diversos perfiles con conocimientos y capacidades distintas.

En la tabla 2 se realiza un análisis de las tareas con el fin de identificar el perfil más adecuado para su consecución. Se observa que la mayoría de las tareas corresponden a un consultor de datos, sin embargo, son necesarios cuatro roles de consultores diferentes dentro de la compañía para cubrir el proyecto en su totalidad.

DEFINICIÓN DEL PROYECTO

Grupo	Tarea	Perfil
BBDD	Adaptación del modelo de datos	Consultor Datos
ETL	Proceso de ingesta desde Excel	Consultor Datos
API	Desarrollo de API REST	Consultor .NET
UI	Desarrollo de la interfaz de usuario	Consultor Power Apps
BI	Preparación modelo de datos, cálculos, informes y aplicación	Consultor Datos
DEVOPS	Ciclo de vida de desarrollos, IaC, docker, despliegue kubernetes	Consultor DevOps

Tabla 2. Asignación de tareas según el perfil de consultor

En la memoria se asumen principalmente las tareas correspondientes a un consultor de datos, mientras que el resto se tratan a un nivel más general, abordándolas desde el punto de vista del diseño, pero sin profundizar en su implementación, ya que, como se aclaró en el apartado 3.1, quedarán fuera del alcance de este proyecto.

Una vez definido el proyecto y conformado el equipo con la experiencia y competencias requeridas, asignando los roles y responsabilidades de cada uno a los distintos perfiles se pueden paralelizar prácticamente todas las tareas. No obstante, dentro de las tareas asignadas al rol de consultor de datos, se seguirá una metodología en cascada, según la cual las actividades se desarrollarán secuencialmente, de manera que el inicio de cada etapa tendrá lugar tras la finalización de la anterior.

3.7 PLANIFICACIÓN

En la figura 12 se muestra la planificación de las etapas propuestas para llevar a cabo el diseño e implementación de la herramienta, definiendo cada una de las actividades principales correspondientes a las tareas que son responsabilidad de un consultor de datos y que se tratarán con mayor detalle en este proyecto.

DEFINICIÓN DEL PROYECTO

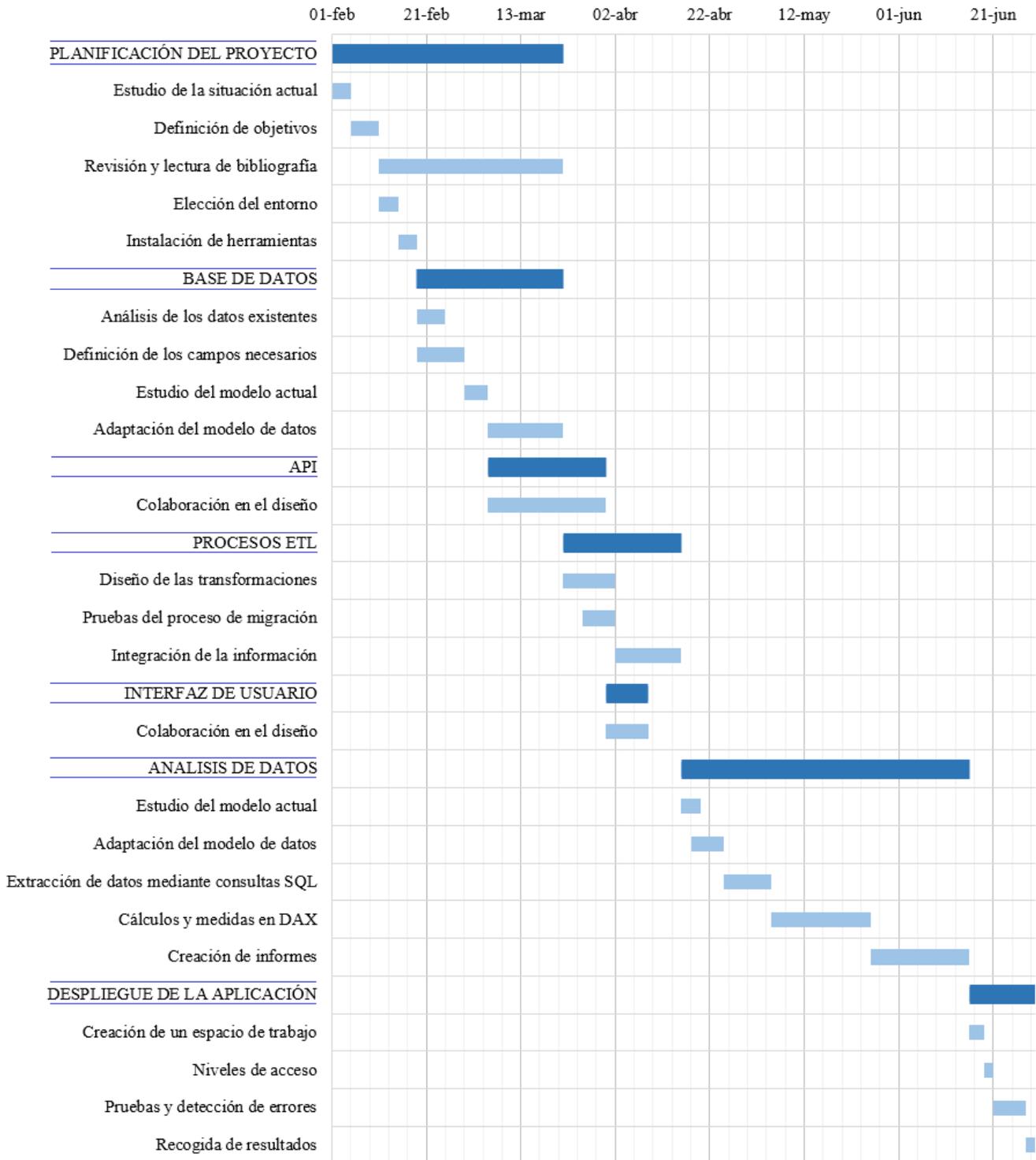


Figura 12. Diagrama de Gantt

DEFINICIÓN DEL PROYECTO

Como se puede observar, el proyecto consta de siete partes principales y tiene una duración estimada de cinco meses.

La primera parte consiste en la planificación del proyecto, donde se debe examinar la problemática actual para establecer los objetivos y definir los requerimientos, además de estudiar las tecnologías pertinentes. A continuación, se ejecutan las tareas necesarias para el almacenamiento y migración de la información en la base de datos de la compañía. Paralelamente, se trabaja en equipo con otros consultores de la empresa para diseñar la solución de una API e interfaz de usuario que permita la creación de una nueva herramienta de uso interno para actualizar y mantener la información integrada en la base de datos. Finalmente, se realizan las tareas que permiten analizar la información y extraer conocimiento, además de compartirlo con los usuarios para detectar posibles mejoras y analizar los resultados.

Capítulo 4. DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

En este capítulo se realiza una descripción detallada de todas las tareas que se han abordado en el proyecto, y que se encuentran resumidas en la figura 13.



Figura 13. Etapas del proyecto

En primer lugar, se exploran los datos en bruto contenidos en un conjunto de archivos Excel compartidos en SharePoint.

A continuación, gracias a la herramienta Microsoft SQL Server Management Studio, se adapta el modelo actual de la base de datos de manera que permita acoger la información proveniente de los archivos previamente examinados.

Tras identificar la fuente y el destino de la información, se realizan los procesos de extracción, transformación y carga del dato con ayuda de la herramienta Azure Data Factory.

Paralelamente, se participa en el diseño de una aplicación que aporte las herramientas básicas necesarias para el mantenimiento de la información y los ciclos de aprobación y actualización periódica.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

Posteriormente, se incorpora la información a analizar en Power BI Desktop mediante consultas SQL con Power Query Editor y se ajusta el modelo de datos existente, para la creación de informes utilizando expresiones de análisis de datos DAX.

Finalmente, se crea un espacio de trabajo y una aplicación con Power BI Service para compartir el conocimiento con los usuarios de la compañía.

4.1 EXPLORACIÓN DE DATOS

En primer lugar, se estudian los datos existentes para determinar aquellos que son relevantes para el proyecto. Se trata de información estructurada contenida en un conjunto de archivos Excel, recopilados en un repositorio de SharePoint.

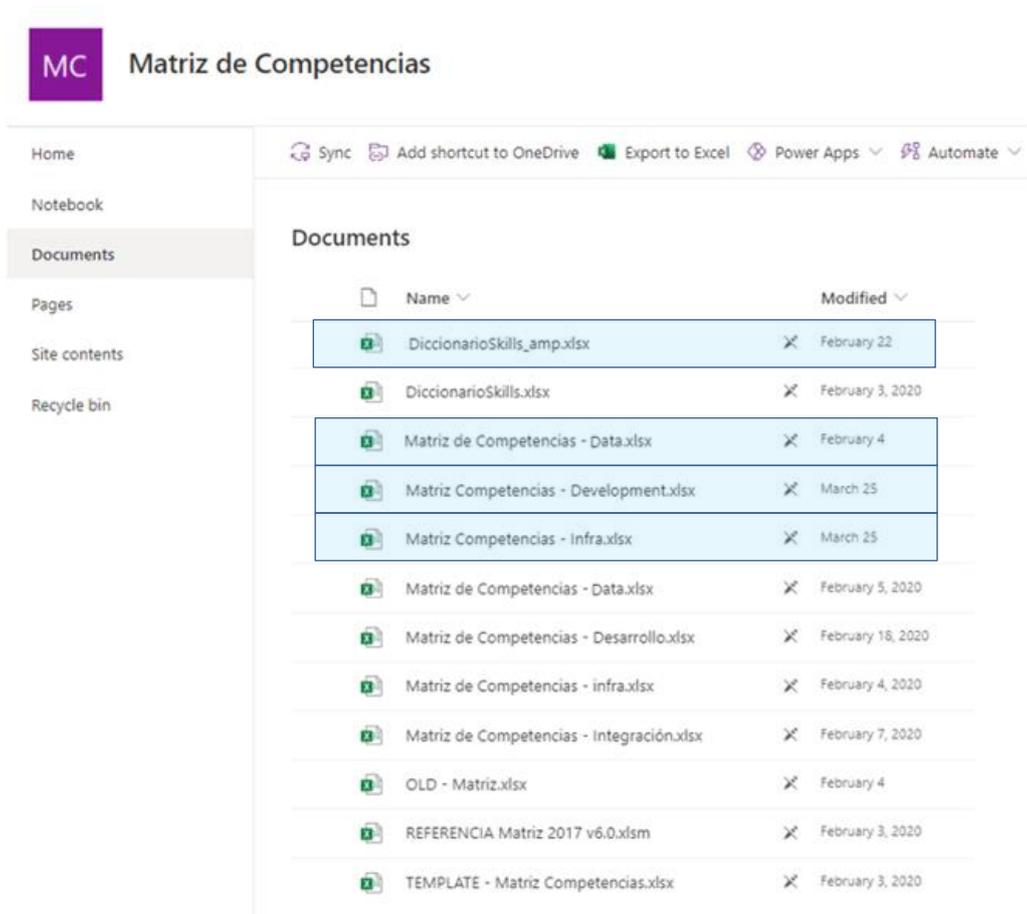


Figura 14. Biblioteca de documentos en SharePoint

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

En la figura 14 se muestra el contenido de esta biblioteca de documentos, identificando cuatro archivos indicados en azul de donde se extraerá la información de interés. El resto de los ficheros son en su mayoría copias antiguas, que se han ido utilizando para actualizar la información a lo largo de los años, por lo que no se consideran útiles para el proyecto por tratarse de información redundante y desactualizada.

El fichero *DiccionarioSkills_amp.xlsx* contiene una hoja de cálculo con los datos organizados en forma de tabla y que incluye todas las *skills* o habilidades técnicas, ordenadas siguiendo una jerarquía formada por área, familia, ámbito y tecnologías. Para mayor detalle, el contenido completo de este archivo se encuentra en el Anexo.

El conocimiento indicado por los consultores para cada una de las habilidades técnicas se recoge en tres archivos diferentes, de forma que la información se encuentra clasificada según las áreas de *Digital Development*, *Digital Platforms* y *Digital Data & Integration*. Estas tres áreas conforman la estructura de *Technology & Operations*, que engloba a todos los consultores dedicados a proyectos de clientes y representa el 85% del total de empleados de la empresa.

El fichero *Matriz Competencias - Development.xlsx* contiene la información de los consultores del área *Digital Development* y el fichero *Matriz Competencias - Infra.xlsx* del área *Digital Platforms*. Ambos tienen la misma estructura, con dos hojas de cálculo.

	A	B	C	D	E	F	G
1	AREA	CODIGO	NOMBRE	SEDE	SUPERVISOR	FTE	POSICION
2	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C4
3	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C5
4	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C3
5	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C4
6	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C1
7	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C6
8	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C5
9	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C2
10	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C2
11	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C3
12	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C3
13	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C3
14	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C2
15	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C3
16	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C2
17	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C5
18	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C2
19	Digital Data & Integration	00000000	Administración General de	Foronda	Administración General de	1	C4

Figura 15. Contenido de la hoja Estructura del fichero Matriz Competencias - Infra

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

La primer hoja denominada Estructura es común para los dos archivos y contiene, en forma de tabla, un registro de los consultores de todas las áreas indicando a cuál pertenecen, su responsable o su nivel dentro de la compañía, entre otros. La figura 15 muestra la estructura de esta hoja de cálculo.

	A	B	C	D	DS	DT	DU	DV	DW
1					Digital Platforms				
2									
89	Platform	Datacenter	Plataforma servidor	Windows Server, Linux,	4	4	4	3	4
90			Servicios de infraestructura	Roles de infraestructura básicos(AD DS y GPOS, DHCP, DNS)	6	2	2	5	5
91			Virtualización	Hypervisores (Hyper-V, Vmware, KVM).	3	1	1	3	3
92			Almacenamiento	RAID, LUNS, MPIO, iSCSI o FC, zoning, mapping	3	1	1	1	1
93			Clustering y HA	Clustering (roles, storage, networks), configuración de Quorum, CSVs, Teaming.	4	1	1	2	2
94			Networking (WIFI, LAN)	VLANs, Spanning Tree, Trunking o link aggregation, SSID, WLAN Channels.	3	3	3	2	2
95		System Center	SCCM	Colecciones, packages, publicaciones, Task Sequences, reporting, despliegue automatizado de actualizaciones, upgrades desatendida	2	1	1	4	4
96			SCOM	Configuración de Management Packs, agentes, reporting	7	1	1	1	1
97			SC Service Manager	Fujos de incidencias, gestion de cambios, ciclo de vida de activos, runbooks, grupos, colas, listas	6	1	1	2	2
98			SC Orchestrator	Connections, Runbook designer, Runbooks, monitorización de actividades	6	1	1	1	1
99			SC Data Protection	Protection Groups, configuración de almacenamiento, agentes, monitorización, backup- Recovery, reporting	3	1	1	1	1
100		Cloud	SC VMM	Management Servers, Networking, Hosts Fabric, VM templates, SDNI (Software Defined Network infrastructure), Storage Spaces Direct	3	1	1	1	1
101			Azure IaaS	Arquitectura híbrida, Arquitectura Hub & Spoke, Compute, Almacenamiento (Blob storage, Azure files, Premium storage), Networking (VNETs, LB, DNS, Traffic Manager, VPN Gateway, Application Gateway).	6	4	1	1	1
102			Azure PaaS	Hybrid Operations services (Azure AD, AD connect, Automation), Webapps, logicapps.	4	4	1	1	1
103	ASR		Azure Backup Server, Diseño e implantación de Disaster Recovery	4	4	1	1	1	

Figura 16. Contenido de la hoja Matriz del fichero Matriz Competencias - Infra

La segunda hoja denominada Matriz y representada parcialmente en la figura 16 tiene una estructura de matriz. En las filas aparecen las tecnologías jerarquizadas por área, familia y ámbito y en las columnas los consultores clasificados por área. En cada celda se recoge un valor del 1 al 6 para cada consultor que haya indicado conocimientos sobre esas tecnologías.

El fichero *Matriz Competencias - Data.xlsx* contiene la información de los consultores del área de *Digital Data & Integration*, y tiene una estructura diferente, como se muestra en la figura 17. Contiene una sola hoja con forma de tabla. En las filas se encuentran los consultores y en las columnas las habilidades técnicas, con un sistema de puntuación de C1 a C6. Además de las habilidades técnicas, esta hoja contiene columnas con información sobre las certificaciones obtenidas y previstas, o la experiencia profesional en distintos sectores.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

G	H	I	J	K	L	M
Escriba su usuario Kabel	Azure Ecosystem - Almacenamiento/BBDD Azure	Azure Ecosystem - Data Catalog	Azure Ecosystem - Data Factory	Azure Ecosystem - Data Lake Analytics / SQL DW	Azure Ecosystem - HDInsight	Big Data Processing - Arquitectura Kappa
Administración	C2	C1	C1	C1	C1	C1
Almacenamiento	C3	C1	C2	C2	C1	C1
Análisis	C2	C2	C2	C2	C2	C1
Arquitectura	C3	C2	C2	C2	C1	C1
Automatización	C4	C3	C2	C2	C2	C2
Big Data	C3	C2	C2	C1	C1	C1
Big Data Analytics	C1	C1	C2	C1	C1	C1
Big Data Processing	C2	C1	C2	C2	C1	C1
Big Data Storage	C5	C4	C4	C4	C4	C4
Big Data Visualization	C4	C2	C3	C5	C2	C1
Big Data Warehouse	C4	C3	C4	C3	C4	C4
Big Data Analytics	C3	C2	C3	C3	C1	C1
Big Data Processing	C4	C4	C4	C5	C5	C5
Big Data Storage	C3	C1	C4	C4	C1	C2
Big Data Visualization	C4	C1	C1	C1	C1	C1
Big Data Warehouse	C6	C3	C6	C3	C3	C3
Big Data Analytics	C4	C1	C3	C1	C1	C1
Big Data Processing	C2	C1	C2	C2	C1	C1
Big Data Storage	C4	C4	C4	C4	C3	C2
Big Data Visualization	C2	C1	C1	C1	C1	C1
Big Data Warehouse	C1	C1	C2	C3	C1	C1
Big Data Analytics	C3	C1	C2	C4	C1	C1
Big Data Processing	C3	C1	C1	C3	C1	C1

Figura 17. Contenido del fichero *Matriz Competencias - Data*

4.2 ALMACENAMIENTO DE DATOS

Una vez identificado el origen de los datos, es necesario considerar el destino de los mismos. El sistema de almacenamiento que se va a utilizar es una base de datos de SQL Server denominada *KabelManagement*, se trata de la base de datos relacional que recoge la información procedente de las aplicaciones internas de la compañía.

Antes de realizar los procesos de extracción, transformación y carga de los datos, se estudia el modelo actual de la base de datos y se determinan los cambios necesarios en su arquitectura. Se extenderá el modelo de datos creando las tablas, campos y claves necesarias mostradas en la tabla 3 para incorporar la información referente a las capacidades técnicas de los consultores, gestionar peticiones de cambios y validaciones, y establecer las relaciones necesarias con las tablas ya existentes en la base de datos.

Tabla	Descripción	Campos
SkillFamilies	Relación de familias de habilidades	<u>SkillFamilyId</u> Name Enabled Position

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

SkillScopes	Relación de ámbitos en los que se agrupan las tecnologías	<u>SkillScopeId</u> Name Enabled Position SkillFamilyId
SkillTechnologies	Relación de tecnologías	<u>SkillTechnologyId</u> Name Enabled Position SkillScopeId
SkillTechnologiesUsers	Mantiene el nivel de capacidad o conocimiento, informado y validado de cada consultor para cada una de las tecnologías	<u>UserId</u> <u>SkillChangeId</u> SkillTechnologyId Level UpdatedOn
SkillTechnologiesDepartments	Mantiene la relación entre cada una de las tecnologías y los departamentos o áreas de la empresa en los que tiene aplicación	<u>SkillTechnologyId</u> <u>SkillChangeId</u>
SkillChanges	Recoge las solicitudes de cambios pendientes de informar o pendientes de validar por su responsable. También hará las veces de histórico de todos los cambios solicitados	<u>SkillChangeId</u> CreatedOn CreatedBy SkillTechnologyId UserId Status Level UpdatedOn ValidationStatus ValidatedBy ValidatedOn ValidationNotes
SkillChangesStates	Relación de posibles estados para las peticiones de cambios	<u>SkillChangeStateId</u> Name
SkillValidationStates	Relación de posibles estados del proceso de validación	<u>SkillValidationStateId</u> Name

Tabla 3. Tablas incorporadas al modelo de datos en SQL Server

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

La tabla *SkillFamilies* recoge las familias de tecnologías, *SkillScopes* recopila los ámbitos en relación con la familia a la que pertenecen y *SkillTechnologies* reúne las capacidades técnicas en relación con el ámbito al que pertenecen.

Por cada cambio que el consultor realiza en la información, su responsable debe crear un proceso de validación. Hasta que esta validación no sea efectiva, se seguirá mostrando y utilizando para los informes los datos de capacitación existentes antes del cambio. La tabla *SkillChanges* mantiene la relación de cambios y el estado de aprobación, y será utilizada como histórico de los cambios realizados sobre la información del nivel de capacitación de cada consultor.

En la tabla *SkillChangesStates* se especifican los siguientes valores posibles para los estados de las peticiones de cambios.

- Pendiente de informar
- Pendiente de validar
- Cerrado

En la tabla *SkillValidationStates* se definen los siguientes valores posibles para los estados del proceso de validación.

- Pendiente
- Validado
- Rechazado

Adicionalmente, se crean las tablas intermedias *SkillTechnologiesDepartments* y *SkillTechnologiesUsers*, que permiten establecer relaciones muchos a muchos entre dos tablas y de esta manera conectar las nuevas tablas creadas con el modelo de la base de datos.

La tabla *SkillTechnologiesUsers* relaciona el nivel de capacitación en una determinada tecnología con los usuarios de la tabla *Users* ya existente en el modelo de datos.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

La tabla *SkillTechnologiesDepartments* relaciona las tecnologías con los departamentos de la tabla *Departments* del modelo actual. Una novedad destacable introducida respecto a los datos que se encontraban en los archivos de Excel ha sido un cambio en el nivel de agregación, dado que antes una familia se consideraba directamente relacionada con un departamento o área concreta, y en esta nueva implementación se permite que una misma tecnología tenga aplicación en varias áreas o departamentos.

Además de crear las relaciones convenientes entre las nuevas tablas introducidas en el modelo, se requiere establecer las siguientes relaciones entre estas tablas y aquellas del modelo relacional.

- Relación de clave foránea desde la tabla *SkillChanges*, en los campos *CreatedBy*, *UserId* y *ValidatedBy* hacia la tabla *Users* por su clave primaria *UserId*.
- Relación de clave foránea desde la tabla *SkillTechnologiesUsers*, en el campo *UserId*, hacia la tabla *Users* por su clave primaria *UserId*.
- Relación de clave foránea desde la tabla *SkillTechnologiesDepartments*, en el campo *DepartmentId*, hacia la tabla *Departments* por su clave primaria *DepartmentId*.

En la figura 18 se muestra de forma gráfica, y con mayor detalle, el modelo relacional de datos descrito, así como su relación con el modelo existente. También se indican las claves primarias y foráneas utilizadas para conectar las tablas, el tipo de dato que contiene cada campo y las relaciones de cardinalidad entre las tablas.

En azul, aparecen las tablas nuevas que se incorporan en la arquitectura de datos para albergar y mantener la información relativa a las capacidades técnicas de los consultores y, en verde, las tablas ya existentes en el modelo.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

Aunque no se precisa realizar cambios en la estructura de las tablas existentes en la base de datos actual, sí que será necesario introducir los datos necesarios en las tablas del modelo *Applications*, *ApplicationRoles* y *UserRoles* para registrar la aplicación y definir los perfiles de acceso para la herramienta que se va a implementar, como se muestra en la tabla 4.

La tabla *Applications* guarda un registro de las aplicaciones internas de la empresa que vuelcan la información en la base de datos. La tabla *ApplicationRoles* contiene los roles definidos para cada aplicación. La tabla *UserRoles* relaciona los usuarios de la compañía a través de su *UserId* con su rol para una determinada aplicación.

Tabla	Campos y Valores
Applications	ApplicationId: MTCA Name; Matriz de Capacidades
ApplicationRoles	RoleId: 1 ApplicationID: MTCA Name: Usuario de aplicación
	RoleId: 8 ApplicationID: MTCA Name: Director de área
	RoleId: 128 ApplicationID: MTCA Name: Administrador
UserRoles	UserId: [Todos los managers de área y arquitectos de workload] ApplicationId: MTCA RoleId: 8
	UserId: [Todos los usuarios administradores] ApplicationId: MTCA RoleId: 128
	La aplicación asignará permiso de acceso a nivel de usuario (RoleId: 1) si no existe una asignación específica de rol para el usuario actual.

Tabla 4. Modificación de tablas en SQL Server

La creación de las nuevas tablas y campos, así como las relaciones, claves primarias y foráneas, resultado de la definición de este nuevo modelo de datos, se han realizado utilizando el lenguaje Transact-SQL, que permite interactuar con la base de datos SQL Server gracias a la herramienta Microsoft SQL Server Management Studio (véase apartado 3.5.1).

4.3 INTEGRACIÓN DE DATOS

Tras identificar el origen y el destino de los datos, se implementa los procesos de extracción, transformación y carga de datos necesarios para migrar la información desde el repositorio de SharePoint donde se encuentran los ficheros Excel a las tablas previamente creadas en la base de datos SQL Server. Estos procesos se realizan siguiendo el flujo que se muestra en la figura 19 con la herramienta Azure Data Factory (véase apartado 3.5.2).

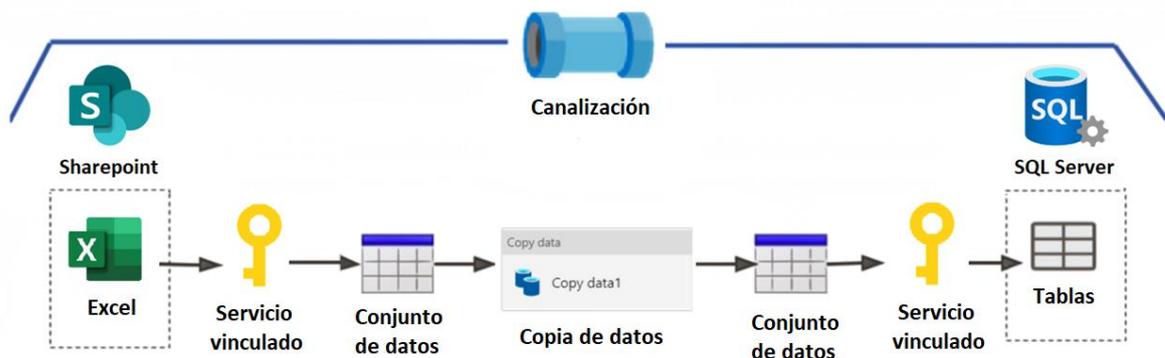


Figura 19. Procesos ETL con Azure Data Factory

Hay que establecer un servicio vinculado con las credenciales de la conexión para conectarse a la fuente de origen de los datos de SharePoint con tipo de formato de los datos en Excel, y al destino de la base de datos de SQL Server, además de crear los conjuntos de datos de entrada y de salida utilizando estos servicios vinculados. A continuación, hay que diseñar las actividades ETL y encadenarlas en una canalización que copie y transforme los datos desde el origen hasta el destino mediante un flujo de datos de asignación.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

El fichero *DiccionarioSkills_amp.xlsx* se utiliza como origen para cargar los datos en las tablas *SkillFamilies*, *SkillScopes*, *SkillTechnologies* y *SkillTechnologiesDepartments*. Para esta última tabla también se utilizará la tabla *Departments* de SQL Server como origen y de esta manera asociar los departamentos de la base de datos con las áreas del fichero Excel. Hasta que no se hayan establecido las nuevas reglas en el nivel de agregación entre familias de tecnologías y áreas mencionadas en el capítulo 4.2, se adoptará la relación entre ambas de acuerdo con el contenido del archivo Excel.

Además, hay que tener en consideración que parte de la información recogida en los archivos Excel de origen ya se encuentra en la base de datos de destino, por lo que la hoja denominada Estructura de los ficheros *Matriz Competencias - Development.xlsx* y *Matriz Competencias - Infra.xlsx*, que contienen información desfasada de consultores que ya no se encuentran en la empresa, no se incluye en la migración.

Las hojas denominadas Matriz de los ficheros *Matriz Competencias - Development.xlsx* y *Matriz Competencias - Infra.xlsx*, junto al fichero *Matriz Competencias - Data.xlsx* se emplean para obtener la información del nivel de capacidad de cada consultor para cada tecnología.

En cuanto a los procesos de transformación, primero hay que modificar el esquema de los ficheros *Matriz Competencias - Development.xlsx* y *Matriz Competencias - Infra.xlsx* pivotando las columnas en filas para convertirlos en una tabla, además de otros procesos de transformación y limpieza como seleccionar las columnas de interés o eliminar las filas con valores nulos. A continuación, hay que tratar el fichero *Matriz Competencias - Data.xlsx*, realizando operaciones como editar y filtrar los nombres de las columnas deseadas, convertir los valores de C1-C6 a 1-6 o eliminar las filas con valores nulos.

Una vez los tres archivos presentan la misma estructura, se unen para recopilar la información de los consultores de todas las áreas en una tabla única. También hay que utilizar la tabla *Users* de SQL Server como origen, localizando el registro correspondiente de cada usuario para asociar los registros de capacitación creados con ese elemento que ya se encuentra actualizado en la base de datos.

Para cada registro de nivel de capacitación que se añade en la tabla *SkillTechnologiesUsers*, se crea un registro relacionado en la tabla *SkillChanges*, así se considera el cambio en estado Cerrado, asignando las fechas de generación y validación, además de un usuario de creación y validación asociado al sistema, que permita identificar que ese registro proviene del proceso de importación inicial.

Este proceso se ha ejecutado realizando una primera carga de prueba en el entorno de desarrollo de la base de datos, y antes de la puesta en producción de la solución ha sido ejecutado de nuevo para cargar los valores sobre las tablas del entorno de producción.

Tras la carga en el entorno de producción se han deshabilitado los accesos de escritura a las hojas Excel actuales para evitar que se sigan realizando cambios sobre ellas. A partir de entonces, la información migrada a la base de datos se mantendrá a través de la nueva aplicación de uso interno.

4.4 MANTENIMIENTO DE DATOS

En este apartado se aborda el diseño de una herramienta de uso interno que permita interactuar con el usuario para los procesos de validación y consolidación de los datos, garantizar la calidad y fiabilidad de la información y ayudar en el proceso de linaje manteniendo el historial de todos los cambios.

Para el entorno de ejecución de la API se hará uso de la tecnología de contenedores Docker, realizando su despliegue en un clúster de Kubernetes, mientras que, para la interacción con el usuario en las tareas de mantenimiento de la información, se utilizarán herramientas sencillas que no impliquen la generación de interfaces de usuario costosas de mantener posteriormente y que simplifiquen en gran medida su rápida implementación como Microsoft Power Apps. Estas tareas sólo se abordarán a nivel de diseño y su implementación quedará fuera del alcance de este proyecto.

4.4.1 API

Para alojar toda la lógica de la aplicación, especialmente la relativa a las consultas y operaciones a realizar en cada una de las tablas para cada una de las operaciones realizadas desde la interfaz, se implementará una API REST basada en el modelo API First.

Existen un conjunto de operaciones necesarias sobre la información ya existente actualmente en la base de datos, para lo que se recomienda implementar una API global con los siguientes métodos.

1. **Users (GET)**. Permite obtener la relación de usuarios ordenada alfabéticamente por “apellidos, nombre”. Puede admitir como parámetro de entrada el ID de un departamento y la lista resultante será filtrada. Deberá tener en consideración el rol del usuario actual para la aplicación MTCA.
2. **Departments (GET)**. Permite obtener la relación de departamentos ordenada alfabéticamente.

Además, la API skills a implementar debe facilitar al usuario la funcionalidad necesaria, utilizando para ello la publicación de los siguientes métodos.

3. **Families (GET)**. Permite obtener la relación de familias completa, sin filtros, y ordenada alfabéticamente,
4. **Scopes (GET)**. Permite obtener la relación de ámbitos ordenada alfabéticamente. Puede admitir como parámetro de entrada el ID de una familia y la lista resultante será filtrada.
5. **Technologies (GET)**. Permite obtener la relación de tecnologías ordenada alfabéticamente. Puede admitir como parámetro de entrada el ID de un ámbito y la lista resultante será filtrada.
6. **Skills (GET)**. Obtendrá la relación de niveles de capacidad en tecnologías para usuarios a partir de la información existente en la tabla *SkillChanges*.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

Se podrán incluir parámetros en la URL para aplicar filtros en los resultados con el objetivo de obtener únicamente los relacionados con un consultor (*ser*), área (*department*), familia (*family*), ámbito (*scope*), pendientes de validar (*pendingvalidation*) y/o pendientes de informar (*pendinginform*). Además, permitirá especificar un único ID (*SkillChangeId*) para obtener la información de un registro concreto.

A los usuarios con nivel de acceso Usuario de aplicación solo se les permitirá consultar registros que coincidan con su *UserId*, aplicándose de forma automática ese filtro a los resultados consultados.

El método devolverá todos los valores existentes en la tabla para cada registro y obtendrá los nombres de los elementos relacionados haciendo uso de sus claves foráneas. También realizará una consulta cruzada con la tabla *SkillTechnologiesUsers* para incluir una columna que permita diferenciar aquellos registros que son los valores en curso utilizados en la matriz de capacidades.

Ejemplos de peticiones:

- GET <https://internalit.kabel.es/api/skills?user=123&scope=15>
Para obtener los registros de un usuario y para un ámbito de tecnologías
- GET <https://internalit.kabel.es/api/skills?department=5&family=2>
Para obtener todos los registros de un departamento y familia de tecnologías.
- GET <https://internalit.kabel.es/api/skills?user=123&pendingvalidation=true>
Para obtener todos los registros pendientes de validar de un usuario.
- GET <https://internalit.kabel.es/api/skills?user=123&pendinginform=true>
Para obtener todos los registros pendientes de informar de un usuario.
- GET <https://internalit.kabel.es/api/skills/334455>
Para obtener un registro específico por su ID.

7. **Skills (PUT)**. Permite insertar un nuevo registro en la tabla *SkillChanges*.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

Los valores del nuevo registro que se deben incluir en el cuerpo del mensaje, así como los valores que se asignarán de forma automática se muestran en la tabla 5.

Valores a definir	Valores asignados automáticamente
SkillTechnologyId UserId Level	SkillChangeId: Autonumérico CreatedOn: Fecha y hora actual CreatedBy: ID del usuario que solicita la creación Status: Pendiente de validar UpdatedOn: Fecha y hora actual ValidationStatus: Pendiente

Tabla 5. Valores del método Skills (PUT)

A los usuarios con nivel de acceso Usuario de aplicación solo se les permitirá añadir registros que coincidan con su *UserId*, devolviendo un error del tipo 403 – *Forbidden* si no se cumple esta condición.

Ejemplo de petición:

```
PUT https://internalit.kabel.es/api/skills
{
  "UserId": 1234
  "SkillTechnologyId: 665
  "Level": 5
}
```

8. **Skills (PATCH).** Permite modificar un registro en la tabla *SkillChanges*.

Los valores del nuevo registro que se deben incluir en el cuerpo del mensaje, así como los valores que se asignarán de forma automática, se muestran en la tabla 6.

Valores a definir	Valores asignados automáticamente
Level	Status: Pendiente de validar UpdatedOn: Fecha y hora actual ValidationStatus: Pendiente

Tabla 6. Valores del método Skills (PATCH)

La URL de la petición debe indicar el ID del registro a modificar.

A los usuarios con nivel de acceso Usuario de aplicación solo se les permitirá modificar registros que coincidan con su *UserId*, devolviendo un error del tipo 403 – *Forbidden* si no se cumple esta condición.

Solo permitirá modificar los registros, para cualquier nivel de acceso del usuario, que se encuentren en los estados Pendiente de informar o Pendiente de validar. Si no se cumple esta condición se devolverá un error del tipo 405 - *Not Allowed*, indicando en el texto descriptivo el motivo.

Ejemplo de petición:

```
PATCH https://internalit.kabel.es/api/skills/33445
{
  "Level": 5
}
```

9. **Skills (DELETE).** Permite eliminar un registro específico de la tabla *SkillChanges*.

A los usuarios con nivel de acceso Usuario de aplicación solo se les permitirá eliminar registros que coincidan con su *UserId*, devolviendo un error del tipo 403 – *Forbidden* si no se cumple esta condición y si el registro está en estado *Pendiente de validar*.

No permitirá el borrado, para ningún nivel de acceso, si el registro está en estado Cerrado y si el registro en curso en la tabla *SkillTechnologiesUsers* está haciendo uso del *SkillChangeId* que se solicita borrar, devolviendo un error del tipo 405 - *Not Allowed*, indicando en el texto descriptivo el motivo.

Ejemplo de petición:

```
DELETE https://internalit.kabel.es/api/skills/334455
```

10. **SkillsValidations (POST)**. Permite realizar cambios en el estado de validación de los registros de la tabla *SkillChanges*. Sólo podrá ser invocado por managers y administradores.

Los valores de validación, que se deben incluir en el cuerpo del mensaje, así como los valores que se asignarán de forma automática, se muestran en la tabla 7.

Valores a definir	Valores asignados automáticamente
SkillTechnologyId ValidationStatus ValidationNotes	Status: Cerrado si el valor de ValidationsStatus es Validado, Pendiente de Validar para el resto de valores UpdatedOn: Fecha y hora actual ValidatedBy: ID del usuario que solicita la validación ValidatedOn: Fecha y hora actual

Tabla 7. Valores del método SkillsValidations (POST)

Si el campo *ValidationStatus* tiene el valor Rechazado, se realizarán las siguientes acciones adicionales:

- Se asignará el valor Pendiente de informar al campo *Status*.
- Deberá indicarse un valor en el campo *ValidationNotes*. Si no se cumple esta condición devolverá un error 405 – *Not Allowed* indicando el motivo del error.

Si el nuevo valor del campo *ValidationStatus* es Validado, se realizarán las siguientes acciones adicionales:

- Se asignará el valor Cerrado al campo *Status*.
- Se creará un registro en la tabla *SkillTechnologiesUsers*, o si ya existe se modificará el registro existente actualizando los campos *SkillTechnologyId* y *Level* con los valores validados de la tabla *SkillChanges*, y el campo *UpdatedOn* con la fecha y hora actual.

Ejemplo de petición:

```
POST https://internalit.kabel.es/api/skills/validations/334455
[
  {
    "SkillTechnologyId": 12345
    "ValidationStatus": 3
    "ValidationNotes": "Revisa el level, creo que lo has puesto muy alto"
  },
  {
    "SkillTechnologyId": 667
    "ValidationStatus": 2
  }
]
```

11. **SkillUpdates (PUT)**. Permite la creación de peticiones de cambios sobre los *skills* o capacidades de forma masiva, insertando varios registros en la tabla *SkillChanges*. Sólo podrá ser invocado por managers y administradores.

Los valores de petición de actualización que se deben incluir en el cuerpo del mensaje, así como los valores que se asignarán de forma automática se muestran en la tabla 8.

Valores a definir	Valores asignados automáticamente
Technologies Users	SkillChangeId: Autonumérico CreatedOn: Fecha y hora actual CreatedBy: ID del usuario que solicita la creación Status: Pendiente de informar UpdatedOn: Fecha y hora actual ValidationStatus: Pendiente

Tabla 8. Valores del método SkillUpdates (PUT)

Technologies será una colección de IDs de tecnologías y *Users* una colección de IDs de usuarios, de forma que se iterará sobre una colección, y para cada uno de los registros existentes en la otra colección se creará un registro en la tabla *SkillChanges*.

Ejemplo de petición:

```
PUT https://internalit.kabel.es/api/skills/updates
{
  "Technologies": [665, 145, 112]
  "Users": [44, 76, 143, 5]
}
```

Todas las APIs requerirán que el usuario que las invoca esté autenticado con credenciales de *tenant* de Azure Active Directory de Kabel. Se implementará un proceso de autenticación basada en roles, haciendo uso de las tablas existentes en el modelo de datos para validar que el usuario está activo en la tabla *Users*. Adicionalmente, se validará en cada uno de los métodos si el usuario tiene permitido su invocación, mediante un sistema de autorización basado en el rol asignado al usuario, y en algunos casos se filtrará la información resultante dependiendo del rol del usuario.

4.4.2 UI

Se requiere la creación de varios formularios que permitan a cada usuario revisar y editar el nivel de capacitación almacenado actualmente. Siempre que un consultor realice un cambio en la información, su responsable deberá crear un proceso de validación.

Además, se debe poder generar una tarea solicitando al consultor que revise determinadas capacidades a través de un formulario filtrado de toda la información que actualmente visualiza, centrada en los elementos solicitados. De igual forma, tras realizar el cambio, este deberá ser validado por su responsable.

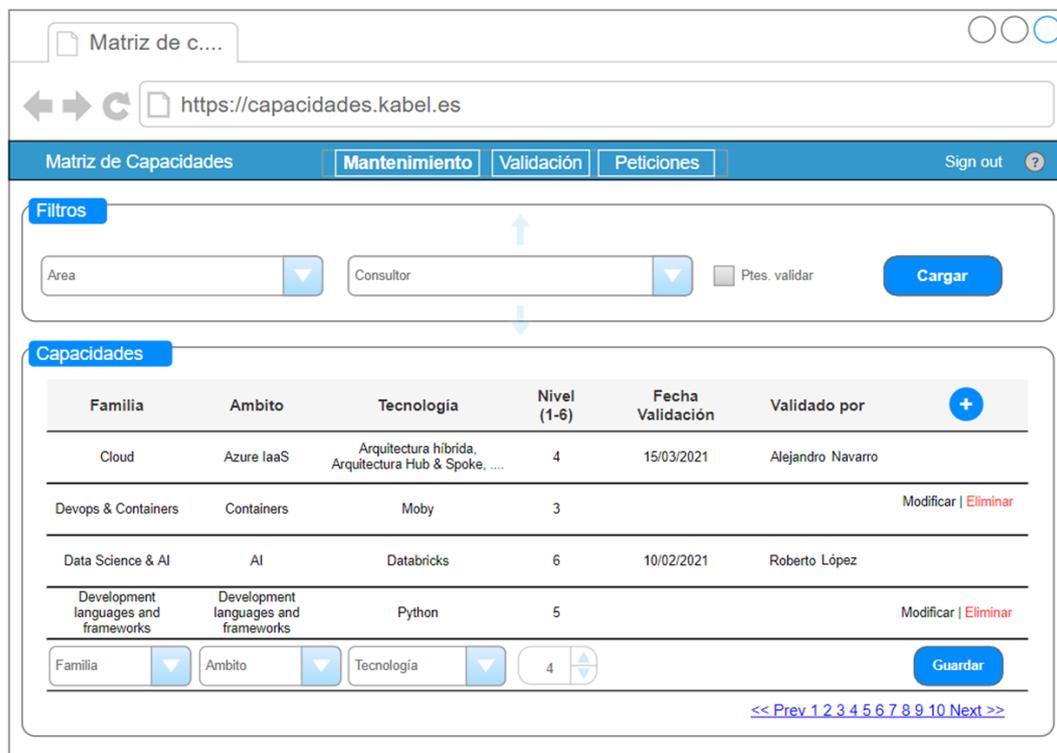
4.4.2.1 Mantenimiento de capacidades

Atendiendo al nivel de acceso del usuario, se le permitirá filtrar y buscar los registros a mostrar por área y/o consultor, utilizando controles de formulario de tipo menú desplegable.

Además, se permitirá aplicar un filtro para que se muestren únicamente los registros pendientes de validar y, mediante un botón, el usuario solicitará la carga de registros según los filtros.

La interfaz permitirá modificar el nivel o eliminar cada uno de los registros mostrados. También se permitirá añadir nuevos registros, teniendo que seleccionar un valor para los campos de familia, ámbito, tecnología y nivel.

La figura 20 muestra una aproximación gráfica del aspecto del formulario de la interfaz de usuario descrita, donde aparecen los registros correspondientes a las capacidades existentes para el usuario seleccionado, con la estructura de columnas mostrada.



Familia	Ambito	Tecnología	Nivel (1-6)	Fecha Validación	Validado por	
Cloud	Azure IaaS	Arquitectura híbrida, Arquitectura Hub & Spoke,	4	15/03/2021	Alejandro Navarro	+
Devops & Containers	Containers	Moby	3			Modificar Eliminar
Data Science & AI	AI	Databricks	6	10/02/2021	Roberto López	
Development languages and frameworks	Development languages and frameworks	Python	5			Modificar Eliminar

Figura 20. Formulario de mantenimiento de capacidades

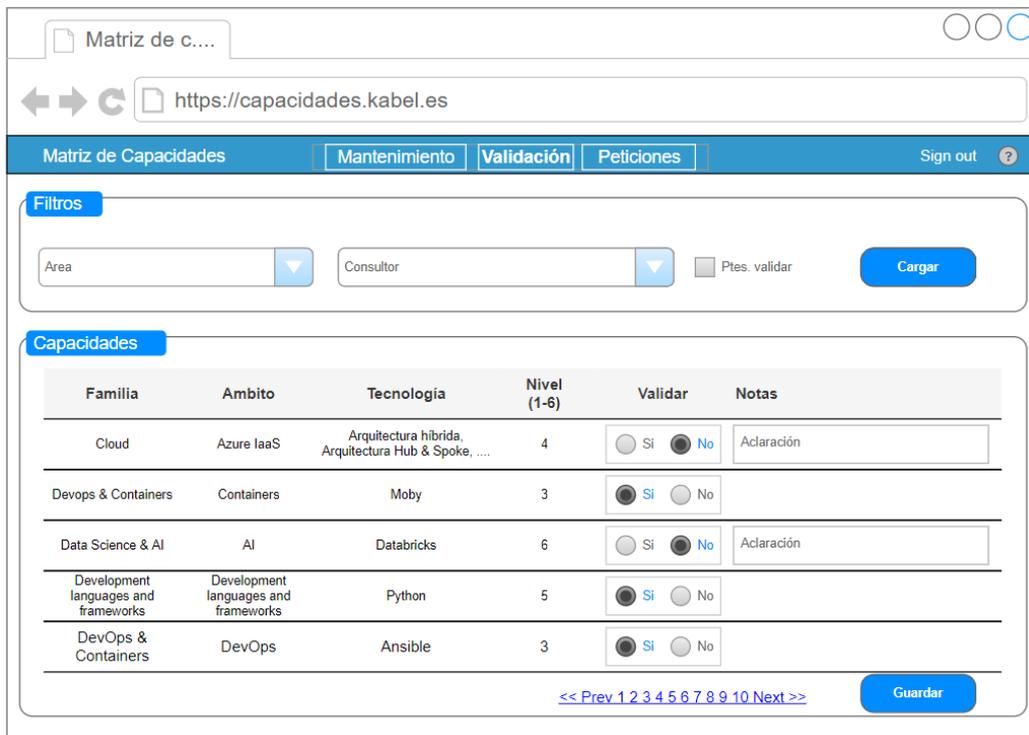
4.4.2.2 Validación de cambios en capacidades

Para la selección de filtros sobre la información a mostrar se utilizarán los mismos criterios que en el formulario anterior, pero este no será accesible a los consultores.

En un breve formulario se permitirá indicar si los registros seleccionados son validados o rechazados, pudiendo indicar un texto explicativo en caso de rechazo.

En caso de validación, el sistema registrará la fecha y usuario que valida de forma automática y, en caso de rechazo, el sistema enviará una notificación al consultor indicándole que tiene elementos pendientes de revisar.

La figura 21 muestra una aproximación gráfica del aspecto del formulario de la interfaz de usuario descrita, donde aparecen los registros correspondientes a las capacidades existentes para el usuario seleccionado, con la estructura de columnas mostrada.



The screenshot shows a web browser window with the URL <https://capacidades.kabel.es>. The page title is 'Matriz de Capacidades'. There are navigation tabs for 'Mantenimiento', 'Validación', and 'Peticiones', along with a 'Sign out' link. Below the tabs is a 'Filtros' section with dropdown menus for 'Area' and 'Consultor', a checkbox for 'Ptes. validar', and a 'Cargar' button. The main content area is titled 'Capacidades' and contains a table with the following data:

Familia	Ambito	Tecnología	Nivel (1-6)	Validar	Notas
Cloud	Azure IaaS	Arquitectura híbrida, Arquitectura Hub & Spoke,	4	<input type="radio"/> Sí <input checked="" type="radio"/> No	Aclaración
Devops & Containers	Containers	Moby	3	<input checked="" type="radio"/> Sí <input type="radio"/> No	
Data Science & AI	AI	Databricks	6	<input type="radio"/> Sí <input checked="" type="radio"/> No	Aclaración
Development languages and frameworks	Development languages and frameworks	Python	5	<input checked="" type="radio"/> Sí <input type="radio"/> No	
DevOps & Containers	DevOps	Ansible	3	<input checked="" type="radio"/> Sí <input type="radio"/> No	

At the bottom of the table, there are navigation links: '<< Prev 1 2 3 4 5 6 7 8 9 10 Next >>' and a 'Guardar' button.

Figura 21. Formulario de validación de cambios

4.4.2.3 Solicitud de actualizaciones por lotes

La finalidad de este formulario es la de poder crear peticiones para que los consultores actualicen o informen de su capacidad o conocimiento sobre un conjunto de tecnologías.

En la parte derecha, el formulario mostrará unos controles de área y consultor para seleccionar los consultores que deseamos añadir a la lista, permitiendo la selección múltiple. De igual forma, en la parte izquierda, mediante los controles de familia, ámbito y tecnología, permitirá hacer una selección múltiple de tecnologías y añadirlas a la lista de tecnologías de la petición. Las listas de familia, ámbito y tecnología son dependientes y sus registros mutuamente excluyentes, por lo que al seleccionar un valor en una lista las siguientes deben filtrar su contenido. Mediante un botón de Crear Petición se almacenará la solicitud y se notificará a los consultores seleccionados que tienen pendiente actualizar capacidades.

La figura 22 muestra una aproximación gráfica del aspecto del formulario de la interfaz de usuario descrita.

Matriz de Capacidades

Mantenimiento Validación Peticiones Sign out ?

Tecnologías

Familia Ambito

Tecnología +

Familia	Ambito	Tecnología
Cloud	Azure IaaS	Arquitectura híbrida, Arquitectura Hub & Spoke, ...
Devops & Containers	Containers	Moby
Data Science & AI	AI	Databricks

<< Prev 1 2 3 4 5 6 7 8 9 10 Next >>

Consultores

Area

Consultor +

Área	Consultor
Digital Platform	Natalia Gómez
Data & Integration	Raul Heras

<< Prev 1 2 3 4 5 6 7 8 9 10 Next >>

Crear petición

Figura 22. Formulario de actualizaciones por lotes

4.4.2.4 Seguimiento de peticiones pendientes

Este formulario será utilizado por los consultores para actualizar la información de capacidades requerida, así como por el resto de los usuarios para realizar un seguimiento de las peticiones de actualización pendientes de completar.

Mediante un conjunto de controles de consultor, familia y ámbito se permitirá seleccionar los registros a mostrar. Se tendrá en consideración el nivel de acceso del usuario, de manera que los consultores únicamente puedan visualizar sus propios registros. También se permitirá aplicar un filtro para que se muestren únicamente los registros pendientes de informar. En los registros pendientes de informar o pendientes de validar se le permitirá al consultor especificar el nivel de capacitación para esa tecnología.

La figura 23 muestra una aproximación gráfica del aspecto del formulario de la interfaz de usuario descrita.

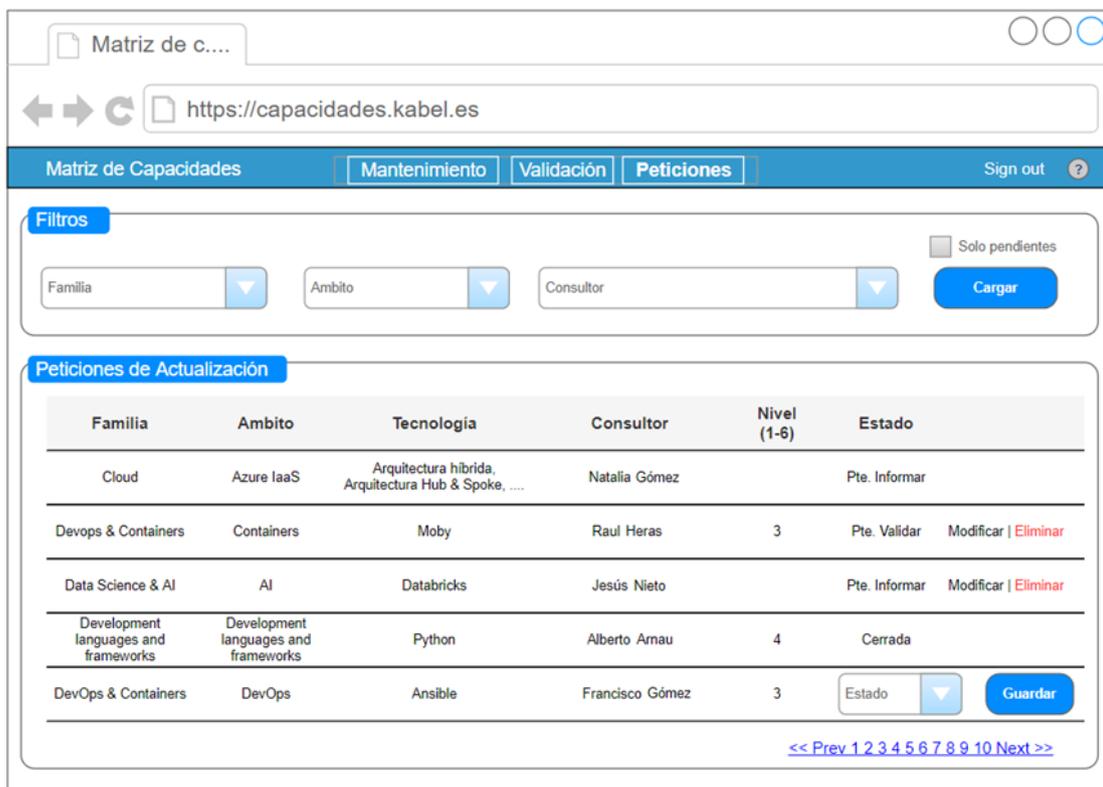


Figura 23. Formulario de peticiones pendientes

4.5 ANÁLISIS DE DATOS

Las herramientas de inteligencia de negocios son una pieza clave tanto para entender los datos y ayudar en la toma de decisiones, como para su posterior medición, seguimiento y control. Gracias a Power BI Desktop (véase apartado 3.5.3) se transforman los datos en conocimiento a través de informes analíticos siguiendo los pasos que se muestran en la figura 24.

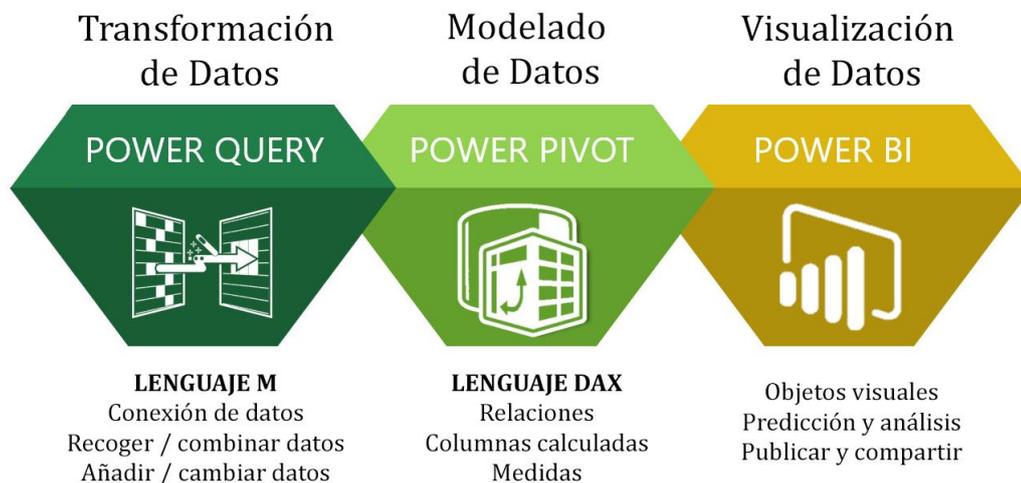


Figura 24. Herramientas de Power BI Desktop

En primer lugar, se extraen los datos de la base de datos de SQL Server mediante consultas SQL en Power Query, una herramienta de conexión para transformar y combinar datos de varias fuentes. Estos datos son los datos que inicialmente se encontraban en varias hojas Excel y fueron integrados en dicha base de datos mediante procesos ETL.

En segundo lugar, se define un modelo en Power Pivot, una herramienta de modelado de datos para integrar estos datos con el resto de información que actualmente mantiene para otros informes. Además, se crean las medidas y columnas calculadas necesarias para representar la información.

En último lugar, mediante Power View, una herramienta de visualización de datos, se crean los objetos visuales, se agrupan en paneles y en informes, y se publican en Power BI Service.

4.5.1 MODELO DE DATOS

El nuevo origen de la información serán las tablas de la base de datos de SQL Server, cuyos datos provienen de las aplicaciones internas de la compañía. Se trata de una base de datos OLTP, con una estructura altamente normalizada, diseñada para la introducción y actualización de datos destinada a la administración de las operaciones transaccionales, pero que no será la más apropiada para el análisis de datos.

A través de un entorno gráfico de extracción, transformación y carga se integra la información referente a las capacidades técnicas de los consultores de la base de datos en el modelo de datos analítico utilizado actualmente en Power BI. Se trata de un modelo OLAP, desnormalizado y más cercano a un *data warehouse*, utilizado en inteligencia de negocios, ya que se encuentra optimizado para realizar consultas e informes.

Las nuevas tablas se obtienen mediante consultas SQL a la base de datos de SQL Server. Las tablas *DimSkills*, *SkillUsers* y *SkillDepartments* tendrán como origen de datos las tablas integradas en la base y mantenidas gracias a la aplicación Matriz de Capacidades, mientras que la tabla *Staffing* proviene de información recogida por otras aplicaciones internas. Estas se muestran con más detalle en la tabla 9.

Tabla	Descripción	Campos
DimSkills	Relación de familias y ámbitos en los que se agrupan las tecnologías	SkillFamily SkillScope SkillTechnology SkillTechnologyId
SkillUsers	Relación del nivel de capacidad de cada consultor para cada una de las tecnologías	UserId Level SkillTechnologyId

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

SkillDepartments	Relación entre cada una de las tecnologías y las áreas o departamentos en los que tiene aplicación	SkillDepartmentId Name SkillTechnologyId
Staffing	Relación de consultores, clientes y proyectos con la asignación de horas y coste por semana	UserId UserName CustomerId CustomerName ProjectId ProjectArea ProjectName StaffingDate StaffingRemainingCost StaffingRemainingHours

Tabla 9. Tablas incorporadas al modelo de datos en Power BI

Además de las tablas descritas anteriormente, se utilizan las de *Users*, *Area*, *Dedication* y *Date* del modelo actual para la confección de los informes.

La información referente a los consultores se encuentra en la tabla *Users* y contiene información sobre el nombre y apellidos de los consultores, su email, su nivel de consultor dentro de la empresa o el departamento, entre otros.

La tabla *Users* se encuentra relacionada con la tabla *Area*, y será de utilidad para obtener el nombre del responsable de cada consultor a partir del área a la que pertenece.

La tabla *Dedication* permite extraer el historial de proyectos de cada consultor y sus proyectos en curso, mostrando si se encuentra o no disponible.

La tabla *Date* es una tabla de calendario para establecer relaciones entre tablas en base a fechas de forma que permite trabajar con funciones de inteligencia de tiempo

En la figura 25 se representa el modelo adaptado. Aquí se muestran las relaciones establecidas entre las nuevas tablas y aquellas del modelo analítico existente, además de los campos empleados para conectarlas. Las tablas representadas en blanco son aquellas que ya se encontraban en el modelo de Power BI, mientras que las nuevas aparecen en azul.

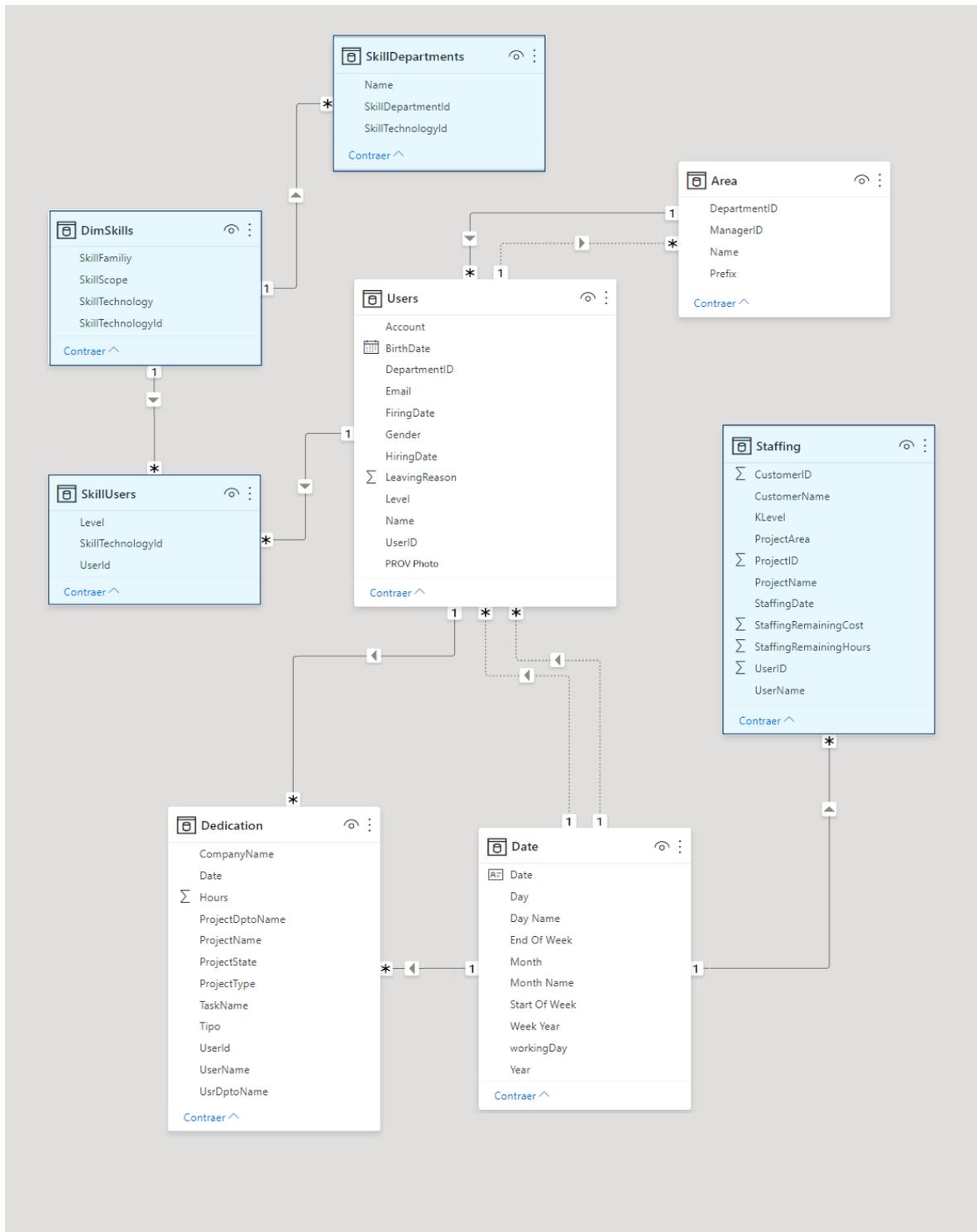


Figura 25. Modelo de datos diseñado en Power BI

4.5.2 CREACIÓN DE INFORMES

Sobre la adaptación del modelo, se han confeccionado un conjunto de cuadros de mando para la visualización y el análisis de los datos. Se han diseñado un total de cuatro paneles, los tres primeros forman parte del mismo informe, cuya función es representar la información relacionada con las capacidades técnicas de los consultores desde una visión general de la compañía hasta el detalle a nivel de consultor, y el último panel forma parte de otro informe para monitorizar el grado de ocupación de estos usuarios en el futuro.

Para representar la información ha sido necesario generar varias columnas calculadas en Power Query, y en los casos en los que el resultado se obtiene accediendo a datos en diferentes filas de la misma tabla o de diferentes tablas, estas columnas calculadas se han obtenido basándose en expresiones DAX. Además, este lenguaje de análisis de datos se ha utilizado en gran profundidad para crear numerosas medidas y tablas calculadas, empezando por expresiones sencillas hasta alcanzar en algunos casos un nivel avanzado.

Todos los informes siguen un mismo patrón con un fondo gris claro, y dos bandas horizontales en la parte superior. En la primera se sitúan el logo de la compañía y el título, y en la segunda se ubican los filtros en fila. Además, a la izquierda aparece una franja vertical donde se añaden más funcionalidades, como otras maneras de filtrar la información o indicadores clave de rendimiento. Para los colores y estilos de texto se utiliza un tema personalizado con los colores corporativos de la empresa.

A continuación, se muestra cada uno de los paneles creados, describiendo los objetos visuales y mostrando su funcionalidad, pero sin entrar en detalle en las tablas, campos o medidas utilizadas.

4.5.2.1 Matriz global

La figura 26 muestra el panel con los datos a nivel global de todos los consultores de la compañía, para un primer contacto con la información referente al conocimiento de los consultores sobre las diferentes herramientas.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

Esta información también se muestra de forma detallada en una matriz, donde en las columnas se representa la jerarquía de ámbito, familia y tecnología, y en las filas los nombres de los consultores, recogiendo la información por cada consultor para cada una de las tecnologías, con un sistema de puntuación del 1 al 6 según sus conocimientos de la herramienta. A nivel de consultor, se puede acceder a su ficha de capacidades situando el cursor sobre su nombre y seleccionando la opción obtener detalles, navegando directamente al panel detalles del consultor.

4.5.2.2 Creación de equipos

La figura 27 muestra un panel más enfocado a analizar los detalles de los equipos, ofreciendo una visión más consolidada de toda la información relacionada con el grupo seleccionado.



Figura 27. Panel creación de equipos

Como en el panel anterior, en la banda superior se permite filtrar la información por familia, ámbito y tecnología utilizando menús desplegables de selección múltiple.

En el margen izquierdo se puede filtrar por nivel de consultor y área a la que pertenece. Además, encima de la tabla aparece un filtro de nivel de capacidad en las diferentes tecnologías a través de un objeto visual de segmentación de datos con control deslizante.

En la parte central se muestra una tabla similar a la visualización principal de la matriz global, con las tecnologías representadas en las columnas, los consultores en las filas, y en cada una de las celdas una puntuación del 1 al 6 según sus conocimientos de la herramienta. De forma análoga, seleccionando un consultor, se puede navegar a su ficha de capacidades.

En el margen derecho se agrupan una serie de tarjetas y gráficos. Las tarjetas muestran el tamaño y el nivel medio del equipo. Un gráfico de columnas muestra la distribución de consultores por niveles y, al pasar el cursor por el gráfico, se muestra otra tarjeta de varias filas con los consultores que pertenecen al nivel indicado. También se muestra la distribución de conocimientos general utiliza un gráfico de anillos, pudiendo obtener los detalles en forma de tablas de consultores y tecnologías para cada nivel de conocimiento. Para finalizar, se representa un gráfico de radar con la puntuación máxima y media de equipo para cada familia de tecnologías.

4.5.2.3 Detalles de consultor

La figura 28 muestra el panel destinado a analizar en detalle la información referente al consultor seleccionado, mostrando su información de contacto, ocupación en proyectos, disponibilidad o conocimiento sobre las distintas tecnologías, entre otros.

En el margen izquierdo, se puede filtrar la información por departamentos y elegir el consultor deseado. También se ha incorporado un gráfico de radar que representa el nivel medio de capacidad agrupadas por familia, ámbito o tecnologías por las que se puede navegar gracias al uso de botones, además aparece un botón para restablecer los valores por defecto del panel.

Una vez seleccionado un consultor, en la banda superior aparece su nombre y su nivel. Además, se puede elegir el rango de nivel de capacidad que se quiere visualizar.

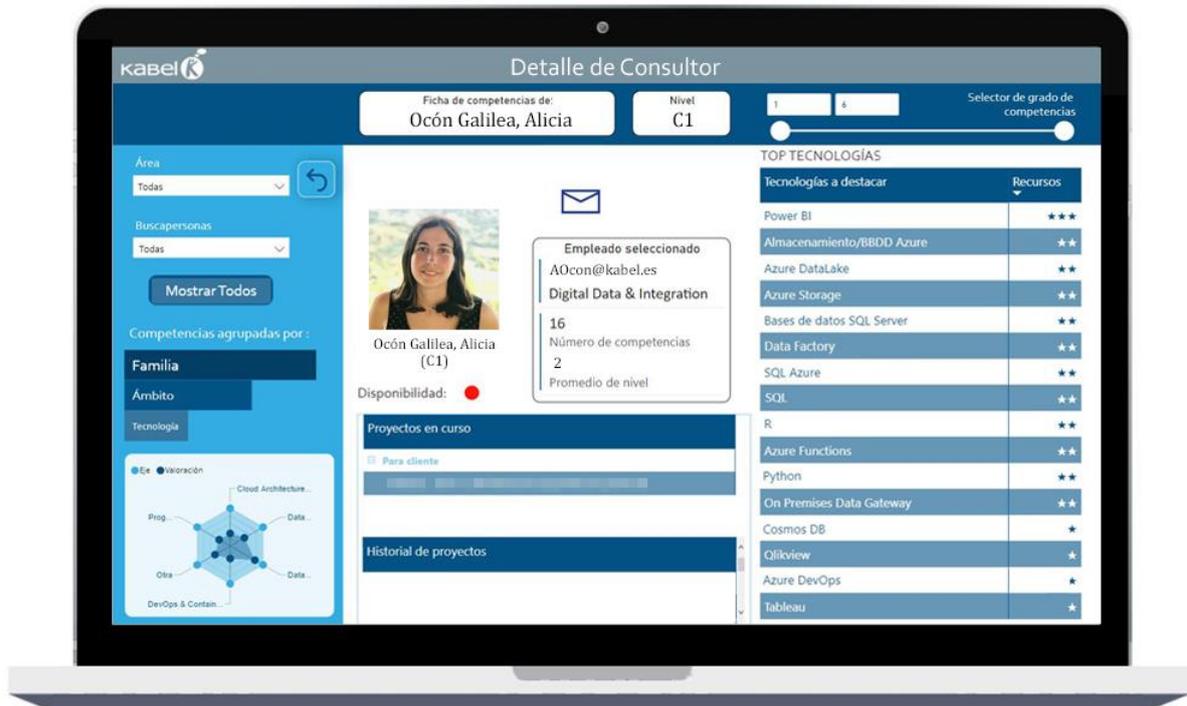


Figura 28. Panel detalles de consultor

La parte central incluye la información de contacto del consultor como el correo electrónico, el área, el número de competencias y nivel promedio de capacidad. Además, a través de hipervínculos, se muestra su foto de perfil y pinchando en el icono de correo se accede al servicio de email de Outlook que permite enviarle un correo electrónico directamente. Además, un círculo de color verde o rojo indica si el consultor se encuentra disponible o sin asignación a proyectos. En la parte inferior hay una tabla donde se muestran sus proyectos en curso y su historial de los previos. Por último, en la parte izquierda aparece un *ranking* de las tecnologías que conoce, representando, con un valor de una a seis estrellas, sus conocimientos de la herramienta.

4.5.2.4 Staffing

La figura 29 representa el panel que servirá como herramienta que permita realizar un seguimiento de la asignación semanal de los consultores de la compañía a los proyectos.

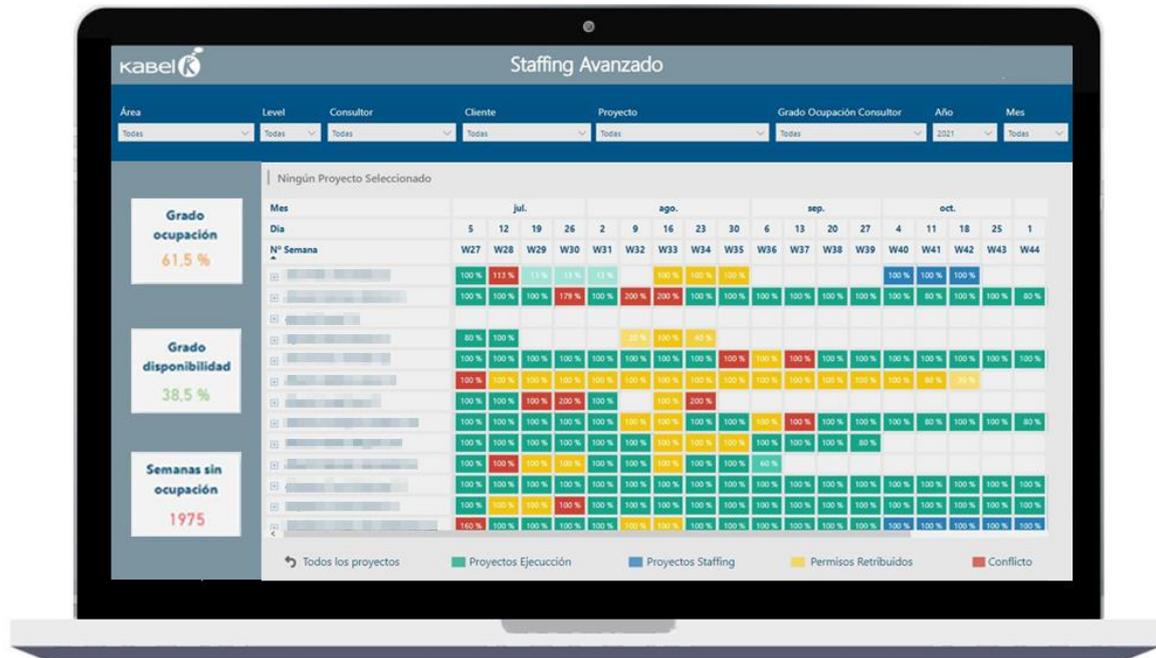


Figura 29. Panel staffing avanzado

En la banda superior se permite filtrar la información por campos relacionados con el consultor, como el área a la que pertenece, su nivel dentro de la compañía, su grado de ocupación medio y su nombre. En cuanto a las tareas, se puede filtrar por cliente y proyecto y, con relación al tiempo, también se añaden los filtros de año y mes.

La matriz incluye el porcentaje de horas laborales en las que los usuarios están asignados a proyectos de clientes. En las filas, la información se encuentra anidada a nivel de consultor, de forma que se puede desplegar para visualizar los proyectos a los que se encuentra asignado cada empleado. En las columnas se representa el eje temporal agrupando la información de forma jerárquica por meses y semanas, donde la semana se encuentra representada tanto por el número de semana respecto al calendario anual, como por el primer día de la semana en formato numérico.

Si se despliega la matriz a nivel de fila y se selecciona uno de los proyectos asociados al consultor, en la parte superior de la matriz, donde actualmente se lee “Ningún Proyecto Seleccionado”, aparecerá el nombre del proyecto con un *link* que nos permite consultarlo directamente en la aplicación.

DISEÑO E IMPLANTACIÓN DE LA SOLUCIÓN

Se ha establecido un código de cuatro colores según el tipo de proyecto: el color verde corresponde a proyectos en ejecución, el amarillo a ausencias y vacaciones, el azul a proyectos de *staffing*, y el rojo se utiliza para representar conflicto, resultado de una combinación de varios colores o de una ocupación mayor del 100%. Dentro de cada una de estas categorías de colores, se ha establecido una gama con tres tonalidades de más claras a más oscuras que indican si esta ocupación es menor del 30%, se encuentra entre el 30% y el 75% o es superior al 75%. Esta clasificación se ha utilizado como formato condicional en las celdas para hacer más interpretables los valores de la matriz.

En la parte inferior se sitúa la leyenda que muestra el código de colores de la matriz explicado. Gracias a botones con acciones, se puede interactuar con ella seleccionando cualquiera de los colores para filtrar la tabla, mientras que la flecha que indica “Todos los proyectos” permite restablecer los valores predeterminados.

En el margen izquierdo, se muestra una serie de indicadores clave de rendimiento, que muestran una rápida instantánea para ayudar a controlar la ocupación de los consultores a futuro, desde la semana actual, hasta el final de año. Los resultados de estos indicadores interaccionan con todos valores filtrados en la parte superior del panel, además de interactuar con las filas y columnas de la matriz, pudiendo seleccionar tanto un consultor, como una semana o un mes para visualizar la información actualizada.

El grado de ocupación hace referencia a la ocupación media en porcentaje para todos los consultores y semanas representadas en la tabla.

El grado de disponibilidad tiene como objetivo calcular la disponibilidad media en porcentaje para todos los consultores y semanas representadas en la tabla, es lo contrario al grado de ocupación, pero teniendo en cuenta que el grado de ocupación puede ser mayor del 100% pero el grado de disponibilidad no podrá ser inferior al 0%.

Las semanas sin ocupación muestran el número de semanas totales que aparecen en la tabla para todos los consultores y semanas del año con una ocupación del 0%.

4.6 ACCESO A DATOS

Una vez se ha testado la herramienta prestando especial atención a que no haya errores en la estructura que puedan llevar a que la información mostrada sea incorrecta y realizar algunas validaciones básicas de la información representada, se procede al despliegue de la aplicación.

En primer lugar, se organiza el contenido creado en un área de trabajo con Power BI Service, un espacio donde compartir contenido y colaborar con otros compañeros. Hay que configurar el espacio de trabajo otorgando a los usuarios los permisos correspondientes dentro de las funciones de visor, colaborador, miembro y administrador. El perfil de visor proporciona acceso solo de lectura a los elementos del espacio de trabajo. El colaborador puede, además, editar los elementos del espacio de trabajo, publicar informes, programar actualizaciones o modificar puertas de enlace a los orígenes de datos. El rol de miembro puede realizar las funciones anteriores y agregar a otros usuarios al rol de visor, colaborador o miembro, mientras que los usuarios administradores pueden, además, agregar y eliminar todos los usuarios, incluidos otros administradores.

En este caso solo se ha identificado un usuario como administrador y los demás con rol de miembro, permitiendo que todos los managers de área y arquitectos de *workload*, así como los consultores que se encuentran asignados a proyectos internos, cuenten con pleno acceso a esta información y puedan realizar cualquier cambio o aprobación. Por defecto, el usuario con rol de administrador será el que reciba las notificaciones de error informando de los problemas que se produzcan en el área de trabajo, por lo que también se añaden a esa lista los consultores encargados de mantener la herramienta.

En la figura 30 se muestra una vista de la trazabilidad para ayudar a entender de donde proceden los datos cubriendo todos los objetos del espacio de trabajo, incluidos los flujos de datos, desde la conexión con la fuente de datos externa, pasando por el conjunto de datos con la última hora de actualización, hasta los informes.

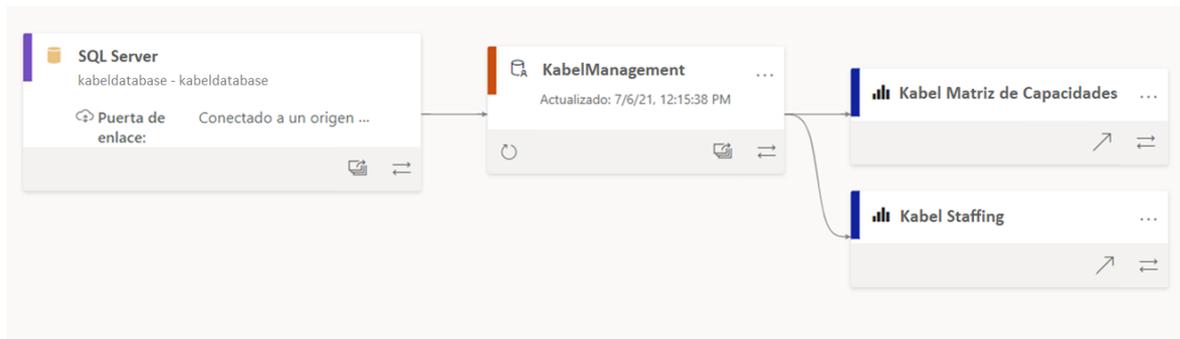


Figura 30. Linaje de los datos en el espacio de trabajo de Power BI

Para asegurar que los datos presentados en los informes sean recientes, se realiza una programación automática. De esta manera siempre que se actualicen los datos, Power BI consultará el origen de datos subyacente, para cargar los de origen en el conjunto de datos, y actualizar las visualizaciones de los paneles. A la hora de definir la frecuencia con la que se actualizará el modelo, como en este caso el origen de datos cambia de forma constante, se configuran ocho franjas de tiempo diarias, el máximo permitido si el conjunto de datos se encuentra en una capacidad compartida.

La colección de paneles, informes y conjunto de datos almacenados en el área de trabajo se comparte con otros usuarios de la compañía a través de la creación y publicación de una aplicación. Mientras que el espacio de trabajo se utiliza como un entorno de desarrollo y colaboración para revisar que la información que será compartida en la aplicación sea correcta, una aplicación está enfocada como un entorno de usuario final, actuando como una versión distribuida de solo lectura del contenido del espacio de trabajo.

La aplicación será el espacio de consumo de la información y el medio para distribuir los informes a un público más amplio, al que tendrán acceso el equipo comercial y los cargos directivos de la empresa, de manera que podrán interactuar con el contenido creado a través de Power BI Service o en una de las aplicaciones móviles disponibles.

Capítulo 5. ANÁLISIS DE RESULTADOS

En este capítulo se destacan los resultados más relevantes del proyecto, revisando que cada uno de los objetivos definidos en el apartado 3.2 se hayan logrado con éxito.

En primer lugar, se ha conseguido dar solución a la problemática actual que existía en la compañía para recopilar la información referente a las capacidades técnicas de los consultores en las diferentes tecnologías, de forma que ya no será necesario llevar un registro manual de la información en hojas de Excel.

En segundo lugar, se ha centralizado la información en la base de datos actual de la compañía, integrando los datos con el resto de información que actualmente mantiene para otras aplicaciones de gestión interna. Esto ha hecho posible organizar y ordenar los datos, facilitando su consulta y permitiendo un control de los mismos de una manera más sencilla, a medida que el volumen de consultores de la compañía crece. Además, se ha evitado la duplicidad de los datos, ya que, como se vio en el apartado 4.1, una misma hoja de cálculo aparece en varios archivos Excel pertenecientes a departamentos diferentes. Asimismo, cada año se realizaba una copia de todos los archivos, conservando las versiones desactualizadas de años anteriores.

En tercer lugar, se ha colaborado en el planteamiento de una solución más eficaz para mantener la información, participando en el diseño de una herramienta de gestión interna con otros perfiles de consultores y obteniendo una visión transversal de todos los departamentos de la compañía. Esta herramienta permite interaccionar con los usuarios para revisar y actualizar la información almacenada en la base de datos, generando las tareas necesarias para los procesos de validación y consolidación de la información a través de formularios. Todo el histórico de cambios se recoge en la base de datos, permitiendo un mayor entendimiento de la información, desde su origen, pasando por todas las modificaciones, hasta su validación.

En último lugar, se han transformado los datos en conocimiento mediante el análisis y monitorización de la información en paneles de visualización. La información tratada ha sido aquella referente a la variedad de soluciones tecnológicas ofrecidas como servicio, el nivel de conocimiento de los consultores en estas herramientas y la asignación de estos recursos a los proyectos de la compañía. Presentar estos datos de una manera más atractiva facilita su comprensión y permite observar patrones o conexiones que antes pasaban desapercibidos. Además, se facilita el acceso a los datos mediante la creación de una aplicación que permite compartir información actualizada en tiempo real de forma remota y accesible desde cualquier dispositivo.

En la figura 31 se presentan de forma gráfica los puntos abordados en el proyecto para lograr los objetivos planteados. La solución proyectada ha tenido en cuenta todo el ciclo de vida del sistema, desde la integración y centralización de los datos, hasta la creación de una aplicación para garantizar su correcta gestión, y el uso de técnicas de inteligencia de negocios para extraer valor de los datos.

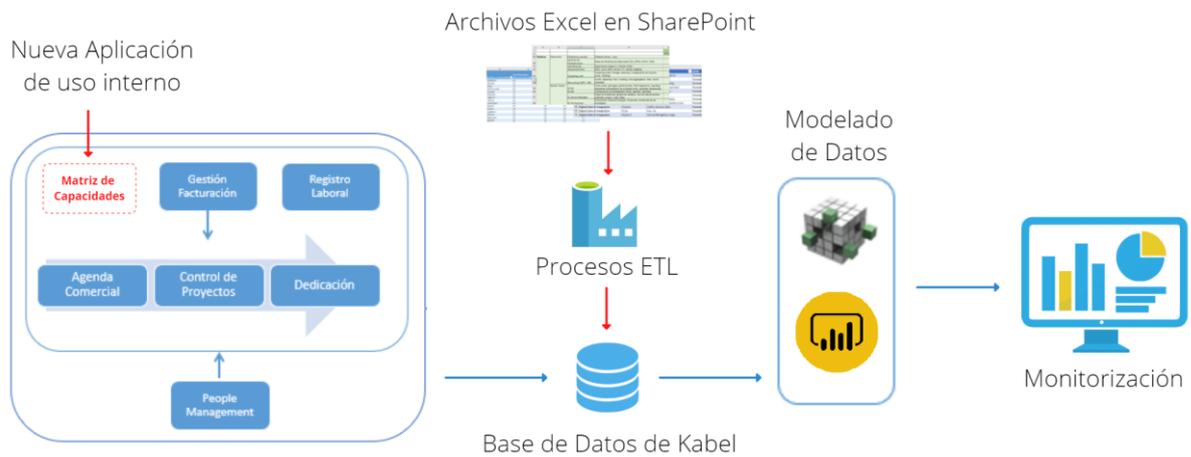


Figura 31. Diagrama de resultados

Capítulo 6. CONCLUSIONES Y TRABAJOS FUTUROS

Como conclusión, se puede afirmar, como se ha revisado en el capítulo 5, que tanto el objetivo principal como los objetivos parciales planteados para este proyecto se han alcanzado de forma exitosa. Debido a la complejidad del proyecto y el requerimiento de perfiles muy diversos, se ha trabajado conforme al alcance del proyecto (véase apartado 3.1), asumiendo las tareas propias de un consultor de datos y participando paralelamente a un nivel más de diseño en el resto de las tareas.

Las tareas abordadas han comprendido la exploración de la herramienta actual para gestionar la información referente a los consultores y el nivel de conocimiento en las distintas tecnologías, su integración en la base de datos de la compañía mediante procesos de extracción, transformación y carga del dato, la colaboración en el diseño de una aplicación interna para el mantenimiento de la información y el análisis mediante paneles de visualización. El proyecto se ha llevado a cabo gracias a tecnologías *cloud*, como SQL Server Management Studio, Azure Data Factory y Power BI, además de la utilización de lenguajes de programación como Transact-SQL y DAX, cumpliendo con los requisitos funcionales definidos en el apartado 3.3.

Como resultado, se ha compartido con la compañía una aplicación que guíe en la toma de decisiones basadas en datos, ayudando a conocer mejor la compañía y a sus trabajadores, y permitiendo una mejor asignación de los recursos y una mayor eficiencia operativa, aumentando la rentabilidad de la empresa. Esta aplicación será el punto de partida para formular nuevas preguntas y enriquecer los paneles con más información y, aunque se ha testado la aplicación, seguro que surgen algunos errores que haya que depurar y se identifican áreas de mejora.

A continuación, se añaden algunas líneas de investigación como pautas para posibles ampliaciones del trabajo en el futuro.

CONCLUSIONES Y TRABAJOS FUTUROS

Actualmente, la información sobre la matriz de capacidades y *staffing* se encuentran en informes diferentes, pudiendo entenderse como dos herramientas separadas. Se podría replantear la solución actual para conseguir una mejor integración entre estas dos partes, facilitando el proceso de selección del candidato que mejor se ajuste a las capacidades técnicas demandadas para esa posición. De esta manera, permitirá una mayor visibilidad para analizar si la capacitación de los técnicos se ajusta a la actual demanda de los clientes. A partir de ello, se podrán definir planes de formación de forma óptima y focalizados en las tecnologías en las que se cuente con un menor grado de capacitación global.

Una sugerencia es ampliar la funcionalidad de la aplicación incorporando nueva información que aporte valor añadido. En la empresa se fomenta el conocimiento y desarrollo profesional, contando con un modelo de formación flexible basado en la consecución de objetivos, que incluyen la obtención de certificaciones y la realización de planes de capacitación para facilitar el aprendizaje en una tecnología o área tecnológica concreta. Como se vio en el apartado 4.1, parte de esta información se encuentra en los ficheros Excel, fuente de las que se han extraído los datos. Sería interesante tener estos datos en cuenta para incluir los objetivos de capacitación y certificaciones de los consultores, proporcionando una nueva herramienta de gestión que permita llevar un registro de los objetivos.

Otro punto a considerar es la estrategia de gobierno del dato. Sigue siendo un reto garantizar una correcta gestión y gobierno de los datos cumpliendo los puntos aclarados en el apartado 2.1, a pesar de que se ha prestado gran atención al diseño de los procesos de validación para asegurar la calidad del dato, se ha tenido en cuenta la trazabilidad proyectando una solución para almacenar el histórico, se ha pensado en la seguridad registrando la aplicación y definiendo los roles de acceso en la base de datos y se ha contribuido a la democratización del dato, poniendo los datos al alcance de los usuarios gracias a la publicación de una aplicación. Por ello, se propone el empleo de una herramienta como Anjana Data que ofrezca una solución integral de gobierno del dato controlando todos los procesos diseñados en la nube, y que brinda funcionalidades como un modelo de gobierno flexible en base a roles configurables, un repositorio centralizado de metadatos con un glosario de negocio y un catálogo de datos o una herramienta de mapeo del linaje de los datos, entre otros [22].

CONCLUSIONES Y TRABAJOS FUTUROS

Por los motivos expuestos en el apartado 3.4, el proyecto se ha llevado a cabo utilizando soluciones de Microsoft Azure, pero, como se vio en el apartado 2.2, existen otros proveedores de servicios en la nube, destacando Amazon Web Services y Google Cloud Platform como sus competidores más fuertes. Por lo que una propuesta interesante sería la elección de otro entorno *cloud* para el despliegue de la aplicación, adquiriendo un conocimiento más amplio de las diferentes soluciones líderes en el mercado en la actualidad, evaluando y contrastando las fortalezas y debilidades de cada una de ellas.

Capítulo 7. BIBLIOGRAFÍA

- [1] Fawcett, T. “Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking”. July 2013. O’Reilly Media.
- [2] Babcock, C. “Management Strategies for the Cloud Revolution. How Cloud Computing Is Transforming Business and Why You Can’t Afford to Be Left Behind”. May 2010. McGraw-Hill.
- [3] Dama International. “DAMA-DMBOK: Guía del Conocimiento Para la Gestión de Datos”. Julio 2020. Technics Publications.
- [4] Villalobos Fernández, E. “Auditoría Interna del Gobierno del Dato”. Febrero 2020. https://auditoresinternos.es/uploads/media_items/lfp-f%C3%A1brica-dato-web-022020.original.pdf
- [5] Bonet López, M. “El gobierno del dato ante la transformación digital. Ciclo El Gobierno del Dato y Transformación Digital”. Enero 2021. <https://ifgeekthen.everis.com/es/el-gobierno-del-dato-ante-la-transformaci%C3%B3n-digital-ciclo-el-gobierno-del-dato-y-transformaci%C3%B3n>
- [6] Parra, J. W. “¿Qué es el DMBOK?”. Mayo 2018. <http://technologyasstrategy.blogspot.com/2018/05/que-es-el-dmbok.html>
- [7] López Pajares, R. “El camino hacia un programa de Gobierno del Dato. Ciclo El Gobierno del Dato y Transformación Digital”. Febrero 2021. <https://ifgeekthen.everis.com/es/el-camino-hacia-un-programa-de-gobierno-del-dato>
- [8] Constaín Rengifo, S. C. “Guía Técnica de Información - Gobierno del dato”. Octubre 2019.
- [9] Howard, D. “Cuaderno del programa de gobierno de datos. Un curso práctico para los principales líderes de datos”. Agosto 2017.
- [10] Dama España. <https://www.damaspain.org/que-es-dama-espana/>
- [11] Ladley, J. “Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program”. November 2019.
- [12] Buyya, R. “Cloud Computing: Principles and Paradigms”. January 2011. Wiley.
- [13] Alta van der Merwe, M. C. “Secure Cloud Computing. Benefits, Risks and Controls”. August 2011. IEEE.

BIBLIOGRAFÍA

- [14] Mell, P. “The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology”. September 2011. NIST Publications.
- [15] Dutta, P. “Comparative Study of Cloud Services Offered by Amazon, Microsoft and Google”. April 2019. International Journal of Trend in Scientific Research and Development.
- [16] Bala, R. “Magic Quadrant for Cloud Infrastructure and Platform Services”. September 2020. Gartner.
- [17] Google Trends. <https://trends.google.es/trends/?geo=ES>
- [18] Sun, J. “Magic Quadrant for Analytics and Business Intelligence Platforms”. February 2021. Gartner.
- [19] Microsoft. <https://docs.microsoft.com/es-es/sql/ssms/sql-server-management-studio-ssms?view=sql-server-ver15>
- [20] Microsoft. <https://docs.microsoft.com/es-es/azure/data-factory/introduction>
- [21] Microsoft. <https://docs.microsoft.com/es-es/power-bi/>
- [22] Anjana Data. <https://anjanadata.com/compania/sobre-anjana/>

ANEXO

Área	Familia	Ámbito	Tecnología
Data & Integration	Azure, Big Data Processing, and Data Integracion	Big Data Processing	Ecosistema Spark
Data & Integration	Azure, Big Data Processing, and Data Integracion	Big Data Processing	Arquitectura Kappa
Data & Integration	Azure, Big Data Processing, and Data Integracion	Big Data Processing	Arquitectura Lambda
Data & Integration	Azure, Big Data Processing, and Data Integracion	Big Data Processing	Soluciones Big Data Streaming
Data & Integration	Azure, Big Data Processing, and Data Integracion	Big Data Processing	Tecnologías Big Data
Data & Integration	Azure, Big Data Processing, and Data Integracion	Big Data Processing	NoSQL Clave-Valor
Data & Integration	Azure, Big Data Processing, and Data Integracion	Big Data Processing	Bases de datos columnares
Data & Integration	Azure, Big Data Processing, and Data Integracion	Big Data Processing	Diseño Arquitectura Docker
Data & Integration	Azure, Big Data Processing, and Data Integracion	Azure Ecosystem	Almacenamiento/BBDD Azure
Data & Integration	Azure, Big Data Processing, and Data Integracion	Azure Ecosystem	Data Factory
Data & Integration	Azure, Big Data Processing, and Data Integracion	Azure Ecosystem	HDInsight
Data & Integration	Azure, Big Data Processing, and Data Integracion	Azure Ecosystem	Data Catalog
Data & Integration	Azure, Big Data Processing, and Data Integracion	Azure Ecosystem	Data Lake Analytics / SQL DW
Data & Integration	Azure, Big Data Processing, and Data Integracion	ETL	DataStage
Data & Integration	Azure, Big Data Processing, and Data Integracion	ETL	Power Center

Data & Integration	Azure, Big Data Processing, and Data Integracion	ETL	Talend
Data & Integration	Data Science & AI	Data Science	Deep Learning - Tensor Flow
Data & Integration	Data Science & AI	Data Science	Data Science - Machine Learning - Statistical Analysis
Data & Integration	Data Science & AI	Data Science	Aprendizaje Supervisado
Data & Integration	Data Science & AI	Data Science	Aprendizaje No Supervisado
Data & Integration	Data Science & AI	IA	Databricks
Data & Integration	Data Science & AI	IA	Azure Machine Learning
Data & Integration	Data Visualization	Data Visualization	Power BI
Data & Integration	Data Visualization	Data Visualization	Elastic Search
Data & Integration	Data Visualization	Data Visualization	D3.js
Data & Integration	Data Visualization	Data Visualization	Tableau
Data & Integration	Data Visualization	Data Visualization	Angular JS, Node JS, Bootstrap
Data & Integration	Cloud Architecture & DevOps	Containers	Azure Functions
Data & Integration	Cloud Architecture & DevOps	Containers	Docker
Data & Integration	Cloud Architecture & DevOps	Containers	Moby
Data & Integration	Cloud Architecture & DevOps	Containers	Kubernetes (AKS)
Data & Integration	Cloud Architecture & DevOps	Cloud Architecture & DevOps	Amazon Cloud Platform
Data & Integration	Cloud Architecture & DevOps	Cloud Architecture & DevOps	Google Platform
Data & Integration	Cloud Architecture & DevOps	Cloud Architecture & DevOps	DevOps approach
Data & Integration	Cloud Architecture & DevOps	DevOps	Azure DevOps
Data & Integration	Cloud Architecture & DevOps	DevOps	Jenkins
Data & Integration	Cloud Architecture & DevOps	DevOps	Ansible
Data & Integration	Cloud Architecture & DevOps	DevOps	Terraform
Data & Integration	Programing Languages	Programing technologies and DB Languages:	SQL
Data & Integration	Programing Languages	Programing technologies and DB Languages:	Java
Data & Integration	Programing Languages	Programing technologies and DB Languages:	.NET Core

Data & Integration	Programing Languages	Programing technologies and DB Languages:	R
Data & Integration	Programing Languages	Programing technologies and DB Languages:	Python
Data & Integration	Programing Languages	Programing technologies and DB Languages:	C/C++
Data & Integration	Data Integration & Store	On Premise Integration	Integración SAP
Data & Integration	Data Integration & Store	On Premise Integration	Integración Oracle
Data & Integration	Data Integration & Store	On Premise Integration	Bases de datos SQL Server
Data & Integration	Data Integration & Store	On Premise Integration	Servidor BizTalk
Data & Integration	Data Integration & Store	On Premise Integration	MuleSoft
Data & Integration	Data Integration & Store	On Premise Integration	On Premises Data Gateway
Data & Integration	Data Integration & Store	On Premise Integration	Conexiones híbridas
Data & Integration	Data Integration & Store	Cloud Storage Azure Ecosystem	Cosmos DB
Data & Integration	Data Integration & Store	Cloud Storage Azure Ecosystem	Data Factory
Data & Integration	Data Integration & Store	Cloud Storage Azure Ecosystem	SQL Azure
Data & Integration	Data Integration & Store	Cloud Storage Azure Ecosystem	Data Catalog
Data & Integration	Data Integration & Store	Cloud Storage Azure Ecosystem	Azure DataLake
Data & Integration	Data Integration & Store	Cloud Storage Azure Ecosystem	Azure Storage
Data & Integration	Data Integration & Store	Cloud Storage Azure Ecosystem	Amazon S3
Data & Integration	Data Integration & Store	Cloud Storage Azure Ecosystem	Amazon DynamoDB
Data & Integration	Internet of Things	Ingestion & Provisioning	Azure IoT Edge
Data & Integration	Internet of Things	Ingestion & Provisioning	Módulos IoT Edge
Data & Integration	Internet of Things	Ingestion & Provisioning	Event Hub
Data & Integration	Internet of Things	Ingestion & Provisioning	IoT Hub
Data & Integration	Internet of Things	Ingestion & Provisioning	Arduino/Raspberry/MXChip
Data & Integration	Internet of Things	Ingestion & Provisioning	Azure IoT Device SDK
Data & Integration	Internet of Things	Ingestion & Provisioning	Azure IoT Central
Data & Integration	Internet of Things	Ingestion & Provisioning	Azure IoT Solutions Accelerators
Data & Integration	Internet of Things	Data Analytics	Time Series Insights

Data & Integration	Internet of Things	Data Analytics	Azure Stream Analytics
Data & Integration	Cloud Integration	Azure Ecosystem	Logic Apps
Data & Integration	Cloud Integration	Azure Ecosystem	Azure Functions
Data & Integration	Cloud Integration	Azure Ecosystem	Messaging services
Data & Integration	Cloud Integration	Azure Ecosystem	Key Vault usage
Data & Integration	Cloud Integration	Azure Ecosystem	API Management
Data & Integration	Cloud Integration	Amazon Ecosystem	AWS Lambda
Data & Integration	Cloud Integration	Amazon Ecosystem	AWS Step Functions
Data & Integration	Cloud Integration	Amazon Ecosystem	AWS CloudWatch Events
Data & Integration	Cloud Integration	Amazon Ecosystem	Amazon API Gateway
Data & Integration	Cloud Integration	Amazon Ecosystem	Amazon SNS and Amazon SQS
Data & Integration	DevOps & Containers	Containers	Docker
Data & Integration	DevOps & Containers	Containers	Moby
Data & Integration	DevOps & Containers	DevOps	Azure DevOps
Data & Integration	DevOps & Containers	DevOps	Jenkins
Data & Integration	DevOps & Containers	DevOps	Ansible
Data & Integration	DevOps & Containers	DevOps	Terraform
Platform	Datacenter	Plataforma servidor	Windows Server, Linux
Platform	Datacenter	Virtualización	Hipervisores
Platform	Datacenter	Almacenamiento	Almacenamiento Datacenter
Platform	Datacenter	Clustering	Clustering
Platform	Datacenter	Networking (WIFI, LAN)	VLANs
Platform	System Center	SCCM	Secuencia de tareas
Platform	System Center	SCOM	MGMT Packs
Platform	System Center	SC Service Manager	Gestión de incidencias y cambios
Platform	System Center	SC Orchestrator	Orquestación
Platform	System Center	SC Data Protection	Grupos de protección
Platform	System Center	SC VMM	Servidores de gestión
Platform	Cloud	Azure IaaS	Azure IaaS

Platform	Cloud	Azure PaaS	Azure PaaS
Platform	Cloud	DevOps	DevOps en Infraestructura
Platform	Cloud	Google Cloud (GCP)	IaaS, PaaS Google Cloud
Platform	Cloud	AWS	IaaS, PaaS AWS
Platform	Colaboración	O365	Mensajería
Platform	Seguridad	Dispositivos gestionados	Gestión de permisos
Platform	Seguridad	Dispositivos gestionados	Plataforma Servidor
Platform	Seguridad	Dispositivos gestionados	Plataforma Cliente
Platform	Seguridad	Dispositivos no gestionados	Dispositivos móviles
Platform	Seguridad	Dispositivos no gestionados	Dispositivos IoT
Platform	Seguridad	Identidades y accesos	Seguridades de accesos privilegiados
Platform	Seguridad	Identidades y accesos	Identidades de terceros
Platform	Seguridad	Identidades y accesos	Protección de identidades
Platform	Seguridad	Identidades y accesos	Public Key Infrastructure
Platform	Seguridad	Seguridad de la información	Protección y clasificación de información
Platform	Seguridad	Seguridad de la información	Proteccion Info BBDD
Platform	Seguridad	Cloud	Gobierno, gestión y seguridad cloud
Platform	Seguridad	Cloud	SOC Azure
Platform	Datacenter	Plataforma servidor	Windows Server, Linux
Platform	Datacenter	Servicios de infraestructura	Roles de infraestructura básicos
Platform	Datacenter	Virtualización	Hipervisores
Platform	Datacenter	Almacenamiento	Almacenamiento Datacenter
Platform	Datacenter	Clustering y HA	Clustering
Platform	Datacenter	Networking (WIFI, LAN)	VLANs
Platform	System Center	SCCM	Secuencia de tareas
Platform	System Center	SCOM	MGMT Packs
Platform	System Center	SC Service Manager	Gestión de incidencias y cambios
Platform	System Center	SC Orchestrator	Orquestación
Platform	System Center	SC Data Protection	Grupos de protección

Platform	System Center	SC VMM	Servidores de gestión
Platform	Cloud	Azure IaaS	Azure IaaS
Platform	Cloud	Azure PaaS	Azure PaaS
Platform	Cloud	ASR	Azure Backup Server
Platform	Cloud	Google Cloud (GCP)	Google Cloud (GCP) IaaS
Platform	Cloud	Google Cloud (GCP)	Google Cloud (GCP) PaaS
Platform	Cloud	AWS	AWS IaaS
Platform	Cloud	AWS	AWS PaaS
Platform	DevOps & Containers	Containers	Docker
Platform	DevOps & Containers	Containers	Kubernetes
Platform	DevOps & Containers	DevOps	Azure DevOps
Platform	DevOps & Containers	DevOps	Jenkins
Platform	DevOps & Containers	DevOps	Plat Ansible
Platform	DevOps & Containers	DevOps	Terraform
Platform	Colaboración	O365	Mensajería
Platform	Colaboración	O365	Microsoft Teams
Platform	Seguridad Perimetral	Next Generation Firewalls	Network filtering
Platform	Seguridad Perimetral	Networking	VLANs
Platform	Seguridad Dispositivos gestionados	Sistemas de gestión de permisos	Gestión de permisos
Platform	Seguridad Dispositivos gestionados	Plataforma servidor	Plataforma Servidor
Platform	Seguridad Dispositivos gestionados	Plataforma cliente	Plataforma Cliente
Platform	Seguridad de dispositivos no gestionados	Dispositivos móviles	Dispositivos móviles
Platform	Seguridad de dispositivos no gestionados	Dispositivos IoT	Dispositivos IoT
Platform	Seguridad de identidades y accesos	Seguridad de accesos privilegiados	Seguridades de accesos privilegiados
Platform	Seguridad de identidades y accesos	Entornos colaborativos con identidades de terceros	Identidades de terceros
Platform	Seguridad de identidades y accesos	Protección de identidades	Protección de identidades

Platform	Seguridad de identidades y accesos	PKI y gestión de certificados	Public Key Infrastructure
Platform	Seguridad de la información	Protección y clasificación de la información	Protección y clasificación de información
Platform	Seguridad de la información	Protección de información en BBDD	Proteccion Info BBDD
Platform	Seguridad Cloud	Servicios CLOUD de gobierno, gestión y seguridad	Gobierno, gestión y seguridad cloud
Platform	Seguridad Cloud	SOC Azure (Centro de operaciones de seguridad)	SOC Azure
Digital Development	Cloud Apps	Front-end technologies	Angular Front-end
Digital Development	Cloud Apps	Front-end technologies	Vue
Digital Development	Cloud Apps	Front-end technologies	Svelte
Digital Development	Cloud Apps	Front-end technologies	Yeoman
Digital Development	Cloud Apps	Front-end technologies	Bootstrap Front-end
Digital Development	Cloud Apps	Front-end technologies	NodeJS
Digital Development	Cloud Apps	Front-end technologies	PWA
Digital Development	Cloud Apps	Front-end technologies	npm
Digital Development	Cloud Apps	Front-end technologies	Gulp
Digital Development	Cloud Apps	Front-end technologies	Webpack
Digital Development	Cloud Apps	Office 365 & Project Ecosystem	Microsoft Graph
Digital Development	Cloud Apps	Office 365 & Project Ecosystem	PowerApps
Digital Development	Cloud Apps	Office 365 & Project Ecosystem	Office UI Fabric
Digital Development	Cloud Apps	Office 365 & Project Ecosystem	Teams AddIns
Digital Development	Cloud Apps	Office 365 & Project Ecosystem	Project Server / Online
Digital Development	Cloud Apps	Office 365 & Project Ecosystem	Outlook AddIns
Digital Development	Cloud Apps	Office 365 & Project Ecosystem	Power Automate / Flow
Digital Development	Cloud Apps	Office 365 & Project Ecosystem	SharePoint / SPFX
Digital Development	Cloud Apps	Azure Ecosystem	Web Apps
Digital Development	Cloud Apps	Azure Ecosystem	API Apps

Digital Development	Cloud Apps	Azure Ecosystem	App insights
Digital Development	Cloud Apps	Azure Ecosystem	DevTest Labs
Digital Development	Cloud Apps	Azure Ecosystem	Azure SQL
Digital Development	Cloud Apps	Azure Ecosystem	File Storage
Digital Development	Cloud Apps	Azure Ecosystem	MongoDB
Digital Development	Cloud Apps	Azure Ecosystem	Dev Cosmos DB
Digital Development	Cloud Apps	Azure Ecosystem	Cognitive Services
Digital Development	Cloud Apps	Google Ecosystem	Firebase
Digital Development	Cloud Apps	Google Ecosystem	App Engine
Digital Development	Cloud Apps	Google Ecosystem	Cloud Storage
Digital Development	Cloud Apps	Google Ecosystem	Cloud SQL
Digital Development	Cloud Apps	Google Ecosystem	Apigee
Digital Development	Cloud Apps	Google Ecosystem	Cloud Build
Digital Development	Cloud Apps	Google Ecosystem	Container Registry
Digital Development	Cloud Apps	Amazon Ecosystem	Dev AWS Lambda
Digital Development	Cloud Apps	Amazon Ecosystem	Dev Amazon API Gateway
Digital Development	Cloud Apps	Amazon Ecosystem	Dev Amazon DynamoDB
Digital Development	Cloud Apps	Amazon Ecosystem	Dev AWS CloudWatch Events
Digital Development	Cloud Apps	Amazon Ecosystem	CodeStar / CodeBuild / CodePipeline
Digital Development	Development languages and frameworks	Development languages and frameworks	Dev SQL
Digital Development	Development languages and frameworks	Development languages and frameworks	Dev Java
Digital Development	Development languages and frameworks	Development languages and frameworks	.NET Core
Digital Development	Development languages and frameworks	Development languages and frameworks	.NET Framework
Digital Development	Development languages and frameworks	Development languages and frameworks	Dev Python
Digital Development	Development languages and frameworks	Development languages and frameworks	PHP

Digital Development	Development languages and frameworks	Development languages and frameworks	Ruby
Digital Development	Development languages and frameworks	Development languages and frameworks	JavaScript
Digital Development	Development languages and frameworks	Development languages and frameworks	TypeScript
Digital Development	Development languages and frameworks	Development languages and frameworks	Dev C/C++
Digital Development	Development languages and frameworks	Development languages and frameworks	PowerShell
Digital Development	Desktop Apps	Microsoft	WPF
Digital Development	Desktop Apps	Microsoft	UWP
Digital Development	CMS	SiteCore	Sitecore Architecture best practices
Digital Development	CMS	SiteCore	Onion/Hex
Digital Development	CMS	SiteCore	DMS and campaign automation
Digital Development	CMS	SiteCore	Layouts, Renderings, Templates and Components
Digital Development	CMS	SiteCore	CMS, marketing automation, web analytics, personalization, business intelligence, and CRM concepts
Digital Development	CMS	Wordpress	Core Backend
Digital Development	CMS	Wordpress	Wordpress MySQL
Digital Development	CMS	Wordpress	Templating
Digital Development	CMS	Wordpress	Themes
Digital Development	CMS	Wordpress	CMS Features
Digital Development	CMS	Wordpress	Plugins
Digital Development	CMS	Other CMS	Joomla
Digital Development	CMS	Other CMS	Drupal
Digital Development	Bots	Microsoft Bot Framework	Bot Framework 3.0
Digital Development	Bots	Microsoft Bot Framework	Bot Framework 4.0
Digital Development	Mobile Apps	Mobile Development	Xamarin

Digital Development	Mobile Apps	Mobile Development	Native iOS
Digital Development	Mobile Apps	Mobile Development	Native Android
Digital Development	Cloud Integration	Azure Ecosystem	Logic Apps Dev
Digital Development	Cloud Integration	Azure Ecosystem	Dev Azure Functions
Digital Development	Cloud Integration	Azure Ecosystem	Dev Messaging services
Digital Development	Cloud Integration	Azure Ecosystem	Dev Key Vault usage
Digital Development	Cloud Integration	Azure Ecosystem	Dev API Management
Digital Development	DevOps & Containers	Containers	Dev Docker
Digital Development	DevOps & Containers	Containers	Dev Moby
Digital Development	DevOps & Containers	DevOps	Dev Azure DevOps
Digital Development	DevOps & Containers	DevOps	Dev Jenkins
Digital Development	DevOps & Containers	DevOps	Dev Ansible
Digital Development	DevOps & Containers	DevOps	Dev Terraform