



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Proyecto profesional de Master Data Management aplicado
al ámbito empresarial multinacional

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Criado Beltrán, Juan Carlos

Tutor/a: Rubio Moreno, Alicia

Cotutor/a externo: VIDAL TELLOLS, JAVIER

CURSO ACADÉMICO: 2023/2024



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Proyecto profesional de Master Data Management aplicado al ámbito empresarial multinacional

TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

Autor: Juan Carlos Criado Beltrán

Tutor: Alicia Rubio Moreno
Javier Vidal Tellols

Curso 2023-2024

Agradecimientos

A mi hermano Borja y a mis padres Mar y Juan Carlos por apoyarme durante todos estos años de carrera, siendo un pilar fundamental en mi vida y gracias a los cuales he llegado a ser la persona que soy hoy en día.

A mis amigos por todos los buenos momentos vividos y por alegrar los momentos no tan buenos.

A la empresa Infoverity S.L. y a mis compañeros Javier Vidal Tellols y Vicente LLorca Pastor por su ayuda en el proyecto.

A Alicia Rubio Moreno por ser la tutora de este proyecto.

Resum

El Treball de Fi de Grau presentat s'enfoca en una àrea que ha guanyat rellevància en l'àmbit empresarial i tecnològic: la Gestió de Dades Mestres . Aquest projecte, que s'ha desenvolupat en aliança amb l'empresa Infoverity S.L., té un caràcter dual. D'una banda, aprofundeix en els aspectes teòrics i conceptuals del MDM, proporcionant una anàlisi exhaustiva dels seus fonaments. Aquesta fase del treball busca establir un marc teòric sòlid que servisca com a punt de referència en la disciplina, al mateix temps que introdueix les potencialitats inherents a l'adopció de tecnologies emergents com el núvol i el Programari com a Servei en l'àmbit del MDM. Cal recalcar la intenció divulgativa d'aquest projecte, volent establir les bases per a futurs projectes i donant llum a un camp de la informàtica que hui dia continua sent d'un cert nínxol. Aquest sòlid pilar teòric estableix l'escenari per a una segona part més aplicada del TFG. En aquesta fase, es duu a terme la descripció detallada d'un projecte específic en MDM que es troba a nivell multinacional. L'estudi abasta el cicle de vida complet d'aquest projecte, des de l'etapa inicial de disseny d'un *Proof of Concept*, passant per les fases de desenvolupament i ajust, fins a culminar en la implementació i desplegament efectiu de les solucions concebudes. Aquesta narrativa no sols serveix per a contextualitzar i aplicar el coneixement teòric prèviament exposat, sinó que també es configura com un valuós cas d'estudi que llança llum sobre els desafiaments pràctics, metodològics i tecnològics que les organitzacions poden enfrontar en gestionar dades en un entorn complex i globalitzat.

Paraules clau: Gestió de Dades Mestres (MDM), Programari com a Servei (SaaS), Qualitat de Dades, Integració de Dades, Registre Daurat, Prova de Concepte (PoC), Arquitectura de Dades, Processos Empresarials, Solucions en el Núvol, Informàtica

Resumen

El Trabajo de Fin de Grado presentado se enfoca en un área que ha ganado relevancia en el ámbito empresarial y tecnológico: la Gestión de Datos Maestros. Este proyecto, que se ha desarrollado en alianza con la empresa Infoverity S.L., tiene un carácter dual. Por un lado, ahonda en los aspectos teóricos y conceptuales del MDM, proporcionando un análisis exhaustivo de sus fundamentos. Esta fase del trabajo busca establecer un marco teórico sólido que sirva como punto de referencia en la disciplina, al tiempo que introduce las potencialidades inherentes a la adopción de tecnologías emergentes como la nube y el Software como Servicio en el ámbito del MDM. Cabe recalcar la intención divulgativa de este proyecto, queriendo sentar las bases para futuros proyectos y dando luz a un campo de la informática que a día de hoy continúa siendo de cierto nicho.

Este sólido pilar teórico establece el escenario para una segunda parte más aplicada del TFG. En esta fase, se lleva a cabo la descripción detallada de un proyecto específico en MDM que se encuentra a nivel multinacional. El estudio abarca el ciclo de vida completo de dicho proyecto, desde la etapa inicial de diseño de un *Proof of Concept*, pasando por las fases de desarrollo y ajuste, hasta culminar en la implementación y despliegue efectivo de las soluciones concebidas.

Esta narrativa no solo sirve para contextualizar y aplicar el conocimiento teórico previamente expuesto, sino que también se configura como un valioso caso de estudio que arroja luz sobre los desafíos prácticos, metodológicos y tecnológicos que las organizaciones pueden enfrentar al gestionar datos en un entorno complejo y globalizado.

Palabras clave: Gestión de Datos Maestros (MDM), Software como Servicio (SaaS), Calidad de Datos, Integración de Datos, Registro Dorado, Prueba de Concepto (PoC), Arquitectura de Datos, Procesos Empresariales, Soluciones en la Nube, Informática

Abstract

The presented thesis focuses on an area that has gained prominence in the business and technological realms: Master Data Management. Developed in partnership with Infoverity S.L., this project has a dual nature. On one hand, it delves into the theoretical and conceptual aspects of MDM, providing a thorough analysis of its fundamentals. This phase of the work aims to establish a solid theoretical framework to serve as a reference point in the field, while introducing the inherent potential of adopting emerging technologies such as cloud computing and Software as a Service in the context of MDM. It is important to highlight the project's educational intent, aiming to lay the groundwork for future projects and shed light on an area of computer science that remains somewhat niche to this day.

This robust theoretical foundation sets the stage for a more applied second part of the thesis. In this phase, a detailed description of a specific, multinational-level MDM project is provided. The study covers the complete lifecycle of said project, from the initial design of a 'proof of concept', through the development and adjustment phases, and culminating in the effective implementation and deployment of the conceived solutions. This narrative not only serves to contextualize and apply the previously presented theoretical knowledge but also acts as a valuable case study that illuminates the practical, methodological, and technological challenges organizations may face when managing data in a complex and globalized environment.

Key words: Master Data Management (MDM), Software as a Service (SaaS), Data Quality, Data Integration, Golden Record, Proof of Concept (PoC), Data Architecture, Business Processes, Cloud Solutions, Informática

Índice general

Índice general	VII
Índice de figuras	XI

1	Introducción	1
1.1	Motivación	2
1.2	Objetivos	3
1.3	Impacto esperado	4
1.4	Estructura de la memoria	4
1.5	Colaboraciones	5
2	Estado del arte	7
2.1	<i>Master Data Management</i>	8
2.1.1	Introducción a MDM	8
2.1.2	Contexto histórico	9
2.1.3	Definición de MDM	10
2.1.4	Componentes principales de MDM	10
2.1.5	Tecnologías y herramientas en MDM	11
2.1.6	Aplicaciones y casos de uso	12
2.1.7	Retos y limitaciones de MDM	14
2.1.8	Tendencias y evolución de MDM	15
2.2	<i>Software as a Service</i> y MDM en la Nube	16
2.2.1	Introducción a SaaS y a MDM en la nube	16
2.2.2	Evolución del software tradicional a SaaS	16
2.2.3	Soluciones Cloud para MDM	17
2.2.4	Comparación de soluciones On Premise y Cloud	18
2.3	Gobernanza de Datos	20
2.3.1	Definición y alcance	20
2.3.2	Retos y oportunidades	21
2.3.3	Importancia en MDM	22
2.4	Calidad de Datos	23
2.4.1	Introducción a la Calidad de Datos	23
2.4.2	Parámetros de la Calidad de Datos	24
2.4.3	Impacto en MDM	25
2.5	Ciclo de vida de un Proyecto de MDM	27
2.5.1	Obtención y análisis de los datos fuentes	27
2.5.2	Modelo de Datos en MDM	27
2.5.3	Integración y Transformación de Datos	28
2.5.4	Elaboración de Reglas de Match and Merge	28
2.5.5	Planificación de Procesos de Datos	29
2.5.6	Pruebas y Análisis de Rendimiento	29
2.6	<i>DAMA Data Management Body of Knowledge</i>	30

2.6.1	Visión General: Una Introducción al Cuerpo de Conocimientos de DAMA sobre Gestión de Datos	30
2.6.2	Relación con MDM: Complementariedad entre DAMA y Estrategias de MDM	31
2.7	Contexto empresarial: Infoverity S.L.	32
2.7.1	Historia de la empresa	33
2.7.2	Servicios ofrecidos	33
2.7.3	Reconocimientos e Hitos	35
2.8	Crítica al estado del arte	36
2.9	Propuesta	37
3	Análisis del problema	39
3.1	Especificación de requisitos	39
3.2	Análisis de riesgos	40
3.2.1	Impacto de los Requisitos	40
3.2.2	Criterios de aceptación del cliente	41
3.3	Análisis legal	41
3.3.1	Regulación General de Protección de Datos (GDPR)	41
3.3.2	Sarbanes-Oxley Act (SOX)	42
3.3.3	Impacto en los Requisitos	42
3.4	Solución propuesta	42
3.5	Plan de trabajo	43
4	Diseño de la solución	45
4.0.1	Arquitectura cloud y tecnología utilizada	45
4.0.2	Definiciones técnicas	46
4.0.3	Diseño detallado de la solución	47
5	Desarrollo de la solución propuesta	49
5.1	Introducción al entorno	49
5.1.1	Vista Inicial de la <i>sandbox</i>	49
5.1.2	Customer 360	50
5.1.3	Consola de Business 360	50
5.1.4	Application Integration	51
5.1.5	Repositorio Personal	51
5.2	Desarrollo de Actualización Automática de Campos	52
5.2.1	Creación de una <i>Business Entity</i>	52
5.2.2	Creación del <i>CAI Process</i>	53
5.2.3	Publicación del Proceso	54
5.2.4	Creación de la <i>External Validation</i>	54
5.3	Desarrollo de Verificación Manual de Actualizaciones	55
5.3.1	Creación de la <i>Business Entity</i>	55
5.3.2	Creación de la Componente de Búsqueda	56
5.3.3	Creación de la Página de Creación de Objetos	56
5.3.4	Creación de la Página de Vista y Edición de Objetos	56
5.3.5	Creación del Business Event	57
5.3.6	Creación del <i>Workflow</i>	57
5.4	Problemas y Dificultades	58
6	Implantación, Pruebas y Resultados	59
6.1	Introducción y Metodología Usada	59
6.2	Resultados en Actualización Automática de Campos	60

6.2.1	Resultados	60
6.2.2	Motivos del Fracaso	60
6.2.3	Posibles Implementaciones Futuras	60
6.3	Resultados en Verificación Manual de Actualizaciones	61
6.3.1	Resultados	61
6.3.2	Pruebas	61
6.3.3	Impacto del Éxito en el Proyecto	61
7	Conclusiones	63
	Bibliografía	65

Apéndices

.1	Objetivos de Desarrollo Sostenible	69
----	----------------------------------------------	----

Índice de figuras

5.1	Vista inicial de la <i>sandbox</i>	49
5.2	Vista inicial de Customer 360	50
5.3	Vista inicial de la Consola de Business 360	51
5.4	Vista inicial de Application Integration	51
5.5	Vista del repositorio personal	52
5.6	Vista de la <i>Business Entity</i> creada	53
5.7	Vista del <i>CAI Process</i> creado	53
5.8	Vista del <i>CAI Process</i> publicado	54
5.9	Vista de la <i>External Validation</i> creada	55
5.10	Vista de la BE con el <i>Field Group</i>	56
5.11	Vista de la componente de búsqueda creada	56
5.12	Vista de la página de vista y edición	57
5.13	Vista del evento creado	57
5.14	Vista del <i>Workflow</i>	58

CAPÍTULO 1

Introducción

En la era de la digitalización masiva, el manejo de datos ha pasado a ser una de las cuestiones más críticas para las organizaciones modernas. El flujo constante de datos desde una variedad de fuentes ha generado tanto oportunidades como desafíos para las empresas que buscan utilizar estos datos para impulsar sus operaciones y estrategias. Si bien el uso eficiente de los datos puede brindar una ventaja competitiva, la falta de una gestión adecuada puede resultar en ineficiencias operativas, riesgos de seguridad y la pérdida de oportunidades de negocio.

Dentro de este complejo panorama, la Gestión de Datos Maestros (MDM, por sus siglas en inglés) ha surgido como una estrategia esencial para el manejo eficiente de datos. MDM se encarga de la creación y gestión de un único punto de referencia para todos los datos críticos de la empresa, permitiendo una visión más clara y una mejor utilización de los datos disponibles. La importancia de MDM se ha visto amplificadas por el rápido desarrollo y adopción de tecnologías en la nube, dando origen a soluciones de MDM como Servicio (MDM SaaS), que buscan combinar la flexibilidad de la computación en la nube con las robustas capacidades de MDM.

La correcta implantación de estas tecnologías y estrategias requiere un enfoque multidisciplinario que abarque desde la Gobernanza de Datos (*Data Governance*) hasta la gestión de la calidad de datos (*Data Quality*). A medida que las organizaciones buscan adaptarse a este cambiante entorno, también deben considerar marcos de trabajo y mejores prácticas en la gestión de datos, como los propuestos por DAMA (*Data Management Association*) en su *DAMA Data Management Body of Knowledge*.

El propósito de este TFG es llevar a cabo un estudio transversal de los diversos aspectos involucrados en la gestión de datos maestros y sus aplicaciones en entornos empresariales. Buscamos proporcionar un análisis exhaustivo que aborde tanto las dimensiones teóricas como las aplicaciones prácticas de temas como MDM, MDM SaaS, Gobernanza de Datos, y demás. La finalidad es equipar al ingeniero informático medio con los conocimientos y habilidades necesarios para entender y aplicar estos conceptos en un contexto real, preparándolo para abordar los desafíos asociados con la gestión de datos en el siglo XXI.

Dadas las restricciones de este estudio, el enfoque principal será en la teorización de los conceptos y la presentación de *proof of concept*(PoC) en la sección

práctica, que servirá como una base sólida para futuras investigaciones y aplicaciones prácticas.

Al final de este trabajo, se espera que el lector no solo esté familiarizado con las terminologías y conceptos clave sino que también comprenda las complejidades y consideraciones necesarias para llevar a cabo un proyecto exitoso de gestión de datos maestros.

1.1 Motivación

La elección del tema para el presente Trabajo de Fin de Grado, se inscribe en un marco de relevancia tanto académica como profesional, así como en la detección de un vacío bibliográfico en el ámbito lingüístico de habla hispana. La importancia del tema en cuestión, vinculado a la Gestión de Datos Maestros y tecnologías afines, puede contextualizarse a través de diversos ejes.

En primer lugar, la concordancia con el perfil profesional del estudiante configura un factor crucial en la selección del tema. La temática seleccionada muestra una alineación significativa con los conocimientos y habilidades adquiridos a lo largo de la formación universitaria, en particular en la rama de Tecnologías de la Información y Comunicación en la Escuela Técnica Superior de Ingeniería Informática de la Universitat Politècnica de València. Este alineamiento no solo aporta coherencia al perfil académico sino que, adicionalmente, ofrece la posibilidad de un enriquecimiento mutuo entre las competencias profesionales y el desarrollo del TFG.

En el ámbito académico, se detecta una afinidad sustancial entre los contenidos del TFG y los temas abordados en la formación universitaria del estudiante. La profundización en campos como la Gestión de Datos Maestros y las soluciones basadas en *Software as a Service* representa una extensión lógica de los conocimientos académicos, otorgando al trabajo una relevancia que trasciende el mero cumplimiento de un requisito formativo.

Una tercera consideración de importancia es la limitada disponibilidad de recursos bibliográficos en lengua española sobre el tema. Gran parte de la literatura especializada y documentación técnica se encuentra predominantemente en inglés. Esta escasez plantea un desafío y una oportunidad para contribuir al corpus de conocimiento en español en este campo especializado, extendiendo así la accesibilidad a esta temática en el entorno hispanohablante.

Por último, cabe subrayar el carácter emergente y especializado del campo de la Gestión de Datos Maestros. A pesar de su relevancia creciente en el entorno empresarial actual, la materia todavía se percibe como un ámbito de nicho que no ha alcanzado una visibilidad generalizada entre los profesionales de la ingeniería informática. En este contexto, el presente TFG se presenta como una herramienta potencialmente valiosa para incrementar la conciencia sobre la importancia de un adecuado manejo de datos en las organizaciones.

En resumen, la motivación para la realización de este TFG reside en una confluencia de factores académicos, profesionales y de divulgación, que en conjunto

perfilan un panorama propicio para una investigación exhaustiva y aplicada en el ámbito de la Gestión de Datos Maestros y tecnologías relacionadas.

1.2 Objetivos

El objetivo general de este Trabajo de Fin de Grado es llevar a cabo un estudio transversal en el ámbito de la Gestión de Datos Maestros, prestando especial atención a su implementación en entornos empresariales y a las tecnologías emergentes que facilitan su despliegue, como *MDM Software as a Service* (SaaS). Dada la creciente importancia de una gestión eficiente de datos, este trabajo pretende servir como recurso integral tanto para la comunidad académica como para los profesionales del sector. Para lograr este fin, se establecen los siguientes objetivos específicos:

1. **Divulgación en Español:** Uno de los propósitos clave es realizar una labor de documentación exhaustiva y divulgación en idioma español. Este objetivo responde a la escasez de recursos académicos y técnicos en esta lengua, a pesar de la relevancia del tema.
2. **Comprensión Multidisciplinaria:** Abordar de forma íntegra la complejidad que envuelve a la Gestión de Datos Maestros, desde la *Data Governance* hasta la *Data Quality*, proporcionando así una visión multidisciplinaria que permita comprender mejor los retos y oportunidades del campo.
3. **Implementación de *Proof of Concept*:** Realizar varios PoC con el objetivo de añadir funcionalidades al software por defecto. Este ejercicio práctico tiene como fin ilustrar cómo los conceptos teóricos se materializan en aplicaciones prácticas, permitiendo así una mejor comprensión de las funcionalidades y limitaciones de las tecnologías existentes.
4. **Contribución a la Formación Académica:** Al centrarse en un área directamente relacionada con la rama de Tecnologías de la Información y la Comunicación en la ETSINF, este trabajo busca contribuir al cuerpo de conocimiento académico y servir como recurso para futuros estudiantes y profesionales en la materia.
5. **Fomento de la Conciencia Profesional:** Con este trabajo se busca aumentar el nivel de conciencia sobre la importancia del MDM en el entorno empresarial actual, un campo de nicho que a menudo es ignorado o desconocido para la comunidad de ingenieros informáticos en general.
6. **Base para Futuras Investigaciones:** Al ser un estudio transversal, el TFG tiene el objetivo de sentar las bases para investigaciones futuras, estableciendo un marco teórico y práctico desde el cual se puedan explorar aspectos más específicos o avanzados del MDM y las tecnologías asociadas.

De esta manera, se espera que este Trabajo de Fin de Grado sea un compendio riguroso y a la vez accesible que permita avanzar en la comprensión y aplicación efectiva de la Gestión de Datos Maestros y las tecnologías asociadas.

1.3 Impacto esperado

El impacto esperado del presente Trabajo de Fin de Grado radica en múltiples dimensiones, tanto para el mundo académico como para el entorno empresarial. Desde un punto de vista académico, este trabajo busca llenar un vacío existente en la literatura en español acerca de *Master Data Management* (MDM), *Software as a Service*, y temas relacionados. La escasez de recursos en este idioma ha representado una barrera para el acceso a la información especializada, limitando el desarrollo profesional y académico en contextos de habla hispana.

En el ámbito empresarial, los beneficios se manifiestan de forma más pragmática. Para los profesionales de la tecnología de la información, la implementación de los *proof of concept* propuestos en este trabajo no solo añadirá funcionalidades al software existente sino que también servirá como un recurso de fácil consulta para futuras implementaciones y desarrollos. Para los responsables de la toma de decisiones en las empresas, la consolidación de datos a través de MDM y SaaS ofrecerá una visión más clara y unificada de la información crítica, facilitando estrategias más efectivas y decisiones más informadas.

De manera transversal, el TFG también se alinea con los Objetivos de Desarrollo Sostenible (ODS), específicamente en el ODS 9¹: Industria, Innovación e Infraestructura. Al promover una gestión de datos más eficiente y segura, se contribuye indirectamente a la creación de infraestructuras resilientes y a la adopción de tecnologías innovadoras que pueden tener un impacto positivo en el desarrollo sostenible de las organizaciones.

Por tanto, el impacto de este TFG es multidimensional, beneficiando a una variedad de usuarios y contribuyendo tanto a la especialización académica como a la eficiencia operacional en contextos empresariales.

1.4 Estructura de la memoria

En este Trabajo de Fin de Grado (TFG), el lector encontrará un profundo análisis sobre el *Master Data Management* y su intersección con *Software as a Service*, gobernanza y calidad de datos, entre otros temas. El Capítulo 2, Estado del arte, ofrece una completa introducción al MDM, su contexto histórico y definición, antes de sumergirse en los componentes técnicos y aplicaciones prácticas del mismo. Este capítulo también aborda cómo el SaaS está impactando en MDM, y explora temas críticos como la gobernanza y calidad de datos, haciendo incluso una crítica al estado del arte actual y presentando propuestas para mejorar.

A continuación, en el Capítulo 3, Análisis del problema, se detallan los requisitos del proyecto, se identifican riesgos potenciales y se analizan consideraciones legales como GDPR y SOX. Este capítulo sentará las bases para la solución propuesta, que se detalla en el Capítulo 4, Diseño de la solución. Aquí se define la arquitectura y las tecnologías utilizadas, así como un diseño detallado de cómo se implementará la solución.

¹<https://www.un.org/sustainabledevelopment/es/infrastructure/>

El desarrollo práctico de la solución se aborda en el Capítulo 5. Este capítulo se enfoca en el entorno de desarrollo, la creación y publicación de componentes, y las dificultades encontradas durante el desarrollo. Luego, en el Capítulo 6, Implantación, Pruebas y Resultados, se describen los métodos de prueba utilizados y los resultados obtenidos, identificando tanto los éxitos como los fracasos del proyecto.

Finalmente, el Capítulo 7, Conclusiones, resume los hallazgos más importantes, los impactos del proyecto y las futuras líneas de investigación o desarrollo.

Es de interés notar que el documento cuenta con un glosario o diccionario de términos específicos del área, ubicado al final del TFG, para que el lector pueda consultar el significado de cualquier término técnico con el que no esté familiarizado.

Esta estructura tiene como objetivo ofrecer un enfoque equilibrado y exhaustivo del tema, yendo desde la teoría hasta la implementación práctica, y abordando tanto los éxitos como los desafíos en el camino.

1.5 Colaboraciones

El presente Trabajo de Fin de Grado ha sido una iniciativa individual en lo que respecta al desarrollo teórico y la labor de documentación. Sin embargo, cabe destacar la colaboración en la fase práctica del proyecto, específicamente en la elaboración y ejecución de los *Proof of Concept*.

En este sentido, se ha contado con la participación activa de Javier Vidal Tello, cotutor del TFG y compañero de departamento; y de Vicente Llorca Pastor, compañero de departamento, ambos en la empresa Infoverity S.L. Su colaboración ha sido instrumental para la definición, desarrollo y evaluación de los PoC presentados. Dicha colaboración ha enriquecido el aspecto práctico del trabajo, proporcionando una perspectiva técnica adicional y consolidando la aplicación de los conceptos teóricos expuestos.

Es relevante señalar que la colaboración se ha centrado exclusivamente en la parte práctica del TFG, sin influir en el desarrollo teórico o en la redacción de la memoria. La inclusión de esta información busca garantizar la transparencia en el proceso de desarrollo del TFG y proporcionar un contexto más completo para futuros lectores interesados en las distintas fases del proyecto.

CAPÍTULO 2

Estado del arte

La gestión de datos ha dejado de ser una actividad puramente administrativa para convertirse en un diferenciador estratégico en el ámbito empresarial. Ya no se trata simplemente de recolectar y almacenar información; ahora es un catalizador para una toma de decisiones más informada, una operatividad más eficiente y una mayor satisfacción del cliente. Dentro de esta metamorfosis, el *Master Data Management* ha surgido como uno de los desafíos más fundamentales para las organizaciones actuales. La importancia de una estrategia eficaz de MDM radica en su capacidad para evitar la fragmentación, inexactitud e inconsistencia de los datos, factores que pueden tener graves repercusiones en diversos ámbitos de la empresa.

No menos importante es el papel que están jugando las tecnologías en la nube como el *Software as a Service*. Estas tecnologías están redefiniendo la forma en que las organizaciones implementan software y gestionan datos, al eliminar la necesidad de infraestructuras costosas y mantenimiento in situ. Esta convergencia de MDM y SaaS ha tenido una fuerte influencia en las ofertas actuales del mercado, incluyendo las suites de soluciones en la nube ofrecidas por Informatica MDM. Estos paquetes de soluciones en la nube han establecido nuevas mejores prácticas y planteado desafíos críticos en la gestión de datos y la implementación de software, los cuales merecen un análisis profundo.

Este capítulo abarca un estudio comprensivo del estado actual y evolutivo del arte tecnológico en estas áreas clave. No sólo examina cómo han avanzado individualmente el MDM y el SaaS, sino también cómo estos se integran y complementan con áreas como la Gobernanza de Datos, la Calidad de Datos, y la Generación de Pruebas de Concepto. Este entramado tecnológico conforma un ecosistema complejo y en constante evolución que se examinará desde múltiples perspectivas, incluyendo la revisión de tecnologías pasadas y presentes que ofrecen funcionalidades similares o complementarias.

El MDM ha evolucionado para incluir una gama más amplia de funciones que van más allá del simple almacenamiento y recuperación de datos. Ahora abarca aspectos críticos como la calidad, la seguridad y la gobernanza de los datos. El SaaS, por su parte, ha revolucionado el acceso y uso del software en las empresas, proporcionando una mayor flexibilidad y escalabilidad. La combinación de ambos ofrece una perspectiva apasionante sobre cómo las soluciones de gestión

de datos y tecnologías en la nube están convergiendo para resolver problemas cada vez más complejos en la gestión de datos y servicios empresariales.

La elección de centrarnos en las soluciones cloud de Informática MDM y en la sinergia entre las tecnologías MDM y SaaS se deriva de un análisis cuidadoso. Este análisis tiene en cuenta la madurez y aplicabilidad de estas tecnologías para satisfacer los objetivos específicos de este TFG, así como la disponibilidad de soporte y documentación.

Para añadir una capa más de complejidad, debemos considerar que las tecnologías no operan en un vacío. Están influidas por una variedad de factores externos como las regulaciones gubernamentales, las expectativas del cliente y las dinámicas de mercado. Por tanto, este capítulo también abordará cómo los desarrollos en MDM y SaaS están siendo moldeados por, y a su vez están moldeando, estos elementos externos.

En resumen, el capítulo servirá como una base sólida y multifacética para el resto del TFG, integrando de forma coherente los diversos elementos que componen el ecosistema de MDM y SaaS. Este enfoque no sólo permitirá un entendimiento más completo y contextualizado de la tecnología actual sino que también justificará la selección de ciertas tecnologías y enfoques para los desarrollos que se abordarán más adelante en este trabajo. Este fundamento teórico es crucial para entender cómo la Gobernanza de Datos, la Calidad de Datos y las Pruebas de Concepto se enlazan intrínsecamente con los conceptos centrales de MDM y SaaS, lo cual será explorado en profundidad en las secciones subsecuentes.

2.1 *Master Data Management*

2.1.1. Introducción a MDM

La gestión de datos ha pasado de ser una actividad auxiliar a ser un pilar crítico en la estrategia empresarial moderna. Esta transformación se ha visto impulsada por la creciente complejidad de los sistemas informáticos y la explosión en la cantidad y variedad de datos disponibles. En este escenario complejo, la necesidad de una gestión efectiva de los datos maestros o *Master Data Management* ha emergido como una prioridad empresarial. MDM no es simplemente una tecnología o una técnica, sino una estrategia clave que abarca la organización, gobernanza y administración de los datos fundamentales en un negocio.

Los datos maestros son aquellos datos que son esenciales para las operaciones en una empresa. Estos datos generalmente son referenciados por varios sistemas de negocios y aplicaciones. Ejemplos típicos incluyen información de clientes, productos, empleados y proveedores. La falta de una gestión eficiente de estos datos cruciales puede resultar en una variedad de problemas, desde decisiones empresariales erróneas hasta ineficiencias operativas y desafíos regulatorios. En la era actual, donde los datos se han convertido en un activo de gran valor, tener una *versión única de la verdad* se ha vuelto indispensable para la eficiencia operativa, la toma de decisiones efectiva y la agilidad.

2.1.2. Contexto histórico

La evolución de la gestión de datos maestros está estrechamente ligada a los avances tecnológicos y cambios en la gestión de la información a lo largo del tiempo. Para entender el surgimiento y la importancia actual de MDM, es crucial explorar su contexto histórico, que revela las necesidades empresariales y tecnológicas que han impulsado su desarrollo.

Originalmente, el concepto de gestión de datos maestros no era una preocupación relevante. En los primeros días de la computación, las organizaciones dependían típicamente de un único recurso informático que albergaba todas las aplicaciones y conjuntos de datos asociados. En este escenario, los archivos de datos planos eran la norma y la redundancia de datos no se consideraba un problema significativo.

Sin embargo, con la introducción de la informática de grupo de trabajo en la década de 1980 y el auge de los sistemas de computadoras de escritorio con acceso a sus propias aplicaciones, comenzó una nueva era de distribución de la gestión de la información. Simultáneamente, la arquitectura de datos evolucionó para incorporar el modelo relacional, dando lugar a la reingeniería de muchas aplicaciones existentes para explotar la eficiencia y el rendimiento de los nuevos Sistemas de Gestión de Bases de Datos Relacionales. Este cambio en la distribución del control administrativo condujo a una variación en cómo se definían los conceptos y objetos empresariales comúnmente utilizados.

Este período marcó el surgimiento de la primera necesidad de MDM: la racionalización de las definiciones y significados de conceptos empresariales comunes. La distribución del control administrativo permitió que cada línea de negocio dictara sus propios procesos y limitaciones, lo que llevó a una confusión de conceptos y a métodos creativos para implementar esos conceptos.

Desde mediados de la década de 1990, la tendencia ha oscilado de nuevo hacia la centralización de la computación. Iniciativas organizacionales para implementar sistemas como Planificación de Recursos Empresariales, Gestión de Relaciones con el Cliente y Gestión de la Cadena de Suministro condujeron al desarrollo de infraestructuras técnicas como almacenes de datos. Aunque la centralización de la información para el análisis y la presentación de informes ofrecía grandes promesas, también introdujo un nuevo conjunto de desafíos. Específicamente, las diferencias en los informes alertaron a los usuarios empresariales sobre la confiabilidad de los datos, lo que llevó a un aumento en la desconfianza hacia los sistemas centralizados.

Este recorrido histórico pone de manifiesto que el MDM no es un concepto aislado, sino una respuesta a las cambiantes necesidades y desafíos empresariales y tecnológicos a lo largo del tiempo. Su relevancia actual puede entenderse mejor al reconocer las diversas etapas de su evolución, marcadas por una serie de impulsores como la necesidad de calidad de datos, consistencia y una visión unificada de la información empresarial.

Con este marco histórico, los siguientes apartados abordarán aspectos técnicos y estratégicos más detallados del MDM, incluida su relación con las tecnologías emergentes y los desafíos actuales y futuros que enfrenta.

2.1.3. Definición de MDM

Master Data Management es más que una tecnología o una técnica; es un conjunto integral de las mejores prácticas de gestión de datos que abarca desde la creación de registros maestros únicos para cada entidad crítica de una empresa, como personas, lugares o cosas, hasta la orquestación de partes interesadas clave, participantes y clientes empresariales. Estos registros maestros se generan a partir de una variedad de fuentes de datos internas y externas, así como de aplicaciones[8]. El proceso incluye unificar, reconciliar y enriquecer los datos, convirtiéndolos en una fuente de información consistente y confiable. Simultáneamente, MDM implementa políticas, procedimientos, servicios e infraestructura con el objetivo de apoyar la captura, integración y uso compartido posterior de estos datos maestros, que son precisos, oportunos, consistentes y completos.

Desde una perspectiva de negocio, MDM tiene como objetivo principal apoyar las necesidades de la organización proporcionando acceso a vistas coherentes de las entidades de datos maestros únicamente identificables en toda la infraestructura de aplicaciones operacionales. Los datos maestros son, por lo tanto, aquellas entidades, relaciones y atributos que son fundamentales para una empresa y son la base de los procesos empresariales clave y los sistemas de aplicación[7].

Es importante destacar que no todos los datos tienen la misma relevancia en el ecosistema de MDM. MDM sostiene que algunas entidades (entidades maestras) son más importantes que otras porque están ampliamente distribuidas en toda la empresa y además residen y se mantienen en múltiples sistemas y silos de aplicaciones. Estas entidades maestras son críticas para múltiples procesos empresariales y operacionales. Aunque estas entidades y los dominios de datos asociados pueden constituir solo un pequeño porcentaje del modelo de datos de una empresa, su importancia desde la perspectiva de la calidad de los datos es desproporcionadamente alta .

En resumen, MDM se presenta como una estrategia esencial que va más allá de la mera tecnología, involucrando diversas prácticas y metodologías para gestionar datos que son críticos para el funcionamiento efectivo de una empresa.

2.1.4. Componentes principales de MDM

Entender *Master Data Management* requiere desglosar sus componentes esenciales. Estos elementos facilitan la formación y mantenimiento de un registro maestro, que es la *versión dorada de la verdad* sobre entidades críticas como clientes, productos, y ubicaciones[4]. Esta sección ofrece una descripción detallada de los componentes clave, incluidos los tipos y fuentes de datos, y la arquitectura del sistema MDM. Para términos adicionales, consulte el glosario al final del documento.

- **Modelo de Datos Maestros:** Es un modelo centralizado que representa todas las entidades y sus relaciones. Debe ser flexible para adaptarse a las fuentes de datos existentes y retroalimentar dichas fuentes.

- **Arquitectura del Sistema MDM:** Es un marco orientado a servicios que se enfoca en el ciclo de vida de los datos. Incluye métodos para soportar actividades como crear, acceder, actualizar y retirar registros maestros.
- **Capa de Servicios de MDM:** Se centra en los componentes técnicos para soportar el ciclo de vida de los datos. Contiene servicios para la sincronización, control de acceso, integración, consolidación y acceso.
- **Tipos de Datos:** Incluye varios tipos de datos como datos de cliente, producto y ubicación, que son críticos para funciones y procesos empresariales.
- **Fuentes de Datos:** Son las bases de datos, sistemas y aplicaciones de donde se extraen los datos para crear el registro maestro. Pueden ser internas o externas.
- **Metadatos:** Contiene información que ayuda en la gestión efectiva de los datos maestros. Los metadatos pueden ser gestionados en un registro de metadatos.
- **Políticas y Procedimientos:** Son las normas y reglas que guían la captura, almacenamiento, gestión y compartición de los datos maestros.
- **Gestión del Ciclo de Vida de Datos:** Enfoque estructurado para la gestión completa del ciclo de vida de los datos maestros, desde su creación hasta su retiro.

Comprender estos componentes es esencial para implementar y administrar una solución MDM efectiva. Estos trabajan conjuntamente para asegurar que los datos maestros sean precisos, consistentes y estén disponibles cuando se necesitan.

2.1.5. Tecnologías y herramientas en MDM

Dentro del entorno de la gestión de datos maestros, la elección de la tecnología y herramientas adecuadas es crucial para el éxito de cualquier proyecto de MDM. Entre las diversas soluciones disponibles, Informatica MDM¹ destaca como una de las más potentes y versátiles, siendo multidominio. Esta plataforma proporciona una flexibilidad sin precedentes en el modelado de datos, permitiendo a las organizaciones adaptar la herramienta a sus necesidades específicas, ya sea para gestionar datos de clientes, productos o cualquier otra entidad crítica. Además, su capacidad para realizar un *matching* y consolidación avanzados garantiza la creación de registros maestros únicos y confiables. Sin olvidar sus robustas características de gobernanza y calidad de datos, Informatica MDM se integra de manera eficiente con diversas fuentes de datos, lo que lo convierte en una solución integral para las necesidades de MDM de cualquier organización.

Aunque Informatica MDM es nuestra herramienta de enfoque, es prudente mencionar brevemente otras alternativas como *SAP Master Data Governance*, *IBM*

¹<https://www.informatica.com/es/products/master-data-management/multidomain-mdm.html>

InfoSphere MDM y Talend, que también ofrecen robustas soluciones para la gestión de datos maestros y podrían ser más adecuadas dependiendo del contexto empresarial específico.

Para concluir, es relevante señalar que la parte práctica de este Trabajo de Fin de Grado se ha llevado a cabo utilizando Informatica MDM Cloud. Esta versión en la nube de la plataforma será discutida en más detalle en el apartado dedicado a SaaS y MDM en la nube, destacando sus ventajas y capacidades específicas en el entorno de la nube.

2.1.6. Aplicaciones y casos de uso

La Gestión de Datos Maestros ha demostrado ser una herramienta invaluable en una amplia gama de aplicaciones y contextos industriales. Los beneficios que aporta son especialmente notables en organizaciones complejas con múltiples sistemas y fuentes de datos que necesitan sincronizarse y armonizarse. En este apartado, exploraremos varias aplicaciones y casos de uso de MDM para ilustrar su versatilidad y capacidad para resolver problemas complejos de datos.

Gestión de Relaciones con Clientes (CRM)

Uno de los usos más comunes de MDM es en la gestión de relaciones con clientes. Las organizaciones a menudo tienen datos de clientes dispersos en varios sistemas, como ventas, marketing y soporte al cliente. MDM permite consolidar esta información en un único registro dorado, lo que facilita una visión de 360 grados del cliente. Este enfoque mejora la calidad del servicio, personaliza la experiencia del cliente y permite estrategias de marketing más efectivas.

Cadena de Suministro y Logística

En el entorno de la cadena de suministro, la coherencia en la información del producto y del proveedor es crucial. Un sistema MDM eficaz puede consolidar datos de múltiples fuentes, lo que permite una mejor gestión del inventario, una planificación más eficiente y un cumplimiento más estricto de los estándares y regulaciones. Así, la organización puede responder de manera más ágil a las fluctuaciones del mercado y las demandas del consumidor.

Cumplimiento y Gobernanza

Las organizaciones están bajo presión constante para cumplir con un creciente cuerpo de regulaciones y estándares, como el RGPD en la Unión Europea o la Ley Sarbanes-Oxley en los Estados Unidos. MDM ayuda en el cumplimiento al permitir la trazabilidad y la auditoría de los datos, garantizando que la información sea precisa, consistente y accesible solo para aquellos con los permisos adecuados.

Atención Médica

En el sector de la salud, MDM puede ser una herramienta valiosa para consolidar registros médicos electrónicos provenientes de diversas fuentes. Esto no solo mejora la calidad de la atención al paciente al proporcionar una visión completa de sus historiales médicos, sino que también facilita la investigación y el análisis de datos para mejorar las prácticas médicas en general.

Servicios Financieros

Las instituciones financieras manejan una cantidad enorme de datos sensibles que deben ser precisos para tomar decisiones informadas en actividades como la evaluación de riesgos de crédito, el cumplimiento normativo y la detección de fraudes. MDM permite a estas organizaciones mantener registros coherentes de clientes y transacciones, lo cual es crucial para la toma de decisiones y el cumplimiento normativo.

Retail y E-commerce

El retail y el comercio electrónico son otros sectores que se benefician significativamente de MDM. Al mantener datos maestros consistentes sobre productos, proveedores y clientes, estas empresas pueden ofrecer una experiencia de compra más personalizada, gestionar inventarios de manera más eficiente y realizar análisis de datos más precisos.

Ejemplos de implantaciones exitosas

Uno de los indicadores más fiables del éxito y la eficacia de cualquier tecnología es su tasa de adopción en distintos sectores y su impacto tangible en las operaciones comerciales o institucionales. En el caso del MDM, numerosas organizaciones, tanto en el ámbito comercial como en el no comercial, han reportado notables mejoras tras su implementación.

- **Ejemplos Comerciales:** Varias corporaciones multinacionales y entidades financieras han implementado soluciones MDM para optimizar sus operaciones y mejorar la toma de decisiones. Entre ellas, cabe destacar a Banco Santander, que ha utilizado MDM para mejorar la gestión de los datos de sus clientes y así ofrecer un servicio más personalizado. BMW, por su parte, ha aprovechado el MDM para centralizar la información de sus productos, mejorando tanto la eficiencia en la producción como la experiencia del cliente. Santalucía Seguros ha utilizado MDM para mejorar la gestión de pólizas y reclamaciones, lo que ha agilizado significativamente sus procesos internos. Además, el Departamento de Cultura y Turismo de Abu Dhabi ha empleado MDM para consolidar y gestionar datos turísticos de múltiples fuentes[5]
- **Ejemplos No Comerciales:** En el ámbito no comercial, instituciones académicas y de investigación también han encontrado valor en la implementación de MDM. Un caso de estudio en el Instituto XYZ examinó el impacto

de la implementación de la gestión de riesgos de seguridad de la información y los controles de seguridad en la madurez de la ciberseguridad. Los resultados mostraron una mejora significativa en la madurez de la ciberseguridad del Instituto XYZ tras la implementación de MDM[10].

En resumen, la gestión de datos maestros es una disciplina que encuentra aplicabilidad en una diversidad de sectores debido a su capacidad para resolver problemas fundamentales relacionados con la calidad, coherencia y gobernanza de los datos. Ya sea mejorando las relaciones con los clientes, optimizando las cadenas de suministro o garantizando el cumplimiento normativo, MDM se presenta como una solución integral para la gestión de datos en organizaciones modernas.

2.1.7. Retos y limitaciones de MDM

La implementación de una solución de MDM no es una tarea trivial y viene acompañada de varios retos y limitaciones, muchos de los cuales no son técnicos, sino organizacionales. Uno de los primeros obstáculos radica en la preparación organizativa para acoger una nueva infraestructura técnica para la colaboración. Tradicionalmente, las empresas operan en silos verticales, con cada línea de negocio manejando sus propios conjuntos de datos. Este enfoque crea barreras significativas para la colaboración horizontal entre diferentes unidades de negocio y resulta en islas de información aisladas unas de otras. La transición a un modelo más integrado inevitablemente desencadenará resistencias a nivel organizativo. Estos desafíos organizacionales se discutirán más a fondo en el capítulo sobre MDM SaaS y Cloud.

Otro desafío crucial es la aceptación e integración empresarial de ida y vuelta. El éxito de cualquier programa de MDM depende de la adopción a nivel empresarial y de cómo se integra con las operaciones y procesos empresariales existentes. Este factor de aceptación empresarial también se verá profundizado en capítulos posteriores.

En lo que respecta a la calidad de los datos, cualquier implementación de MDM debe asegurar un alto grado de calidad de los datos. Los datos de baja calidad no solo hacen ineficaz el sistema de MDM sino que también pueden llevar a decisiones empresariales erróneas. Las estrategias para abordar los desafíos relacionados con la calidad de los datos serán tema de un capítulo dedicado a *Data Quality*.

El gobierno de datos emerge como uno de los factores críticos para el éxito en el crecimiento de un programa de MDM. Implica la definición de políticas y procedimientos para la supervisión de los datos maestros. La transferencia de la responsabilidad y la rendición de cuentas desde las líneas de negocio hacia una estructura organizativa central plantea desafíos adicionales. Este aspecto será el foco de un capítulo completo sobre *Data Governance*.

Finalmente, la integración tecnológica es también un desafío en sí mismo. Muchas veces, las organizaciones intentan adaptar sus procesos alrededor de una tecnología específica, en lugar de integrar la tecnología para apoyar sus procesos. Además, el entorno tecnológico en sí es un ente en constante cambio, lo que re-

quiere que las soluciones de MDM sean lo suficientemente flexibles como para adaptarse a cambios tecnológicos futuros.

En resumen, los retos y limitaciones en MDM son multidimensionales y van más allá de los problemas técnicos. Abordar estos desafíos requiere un enfoque holístico que tenga en cuenta tanto las complejidades organizativas como tecnológicas. Aunque no se propondrán soluciones específicas en esta sección, los capítulos siguientes intentarán abordar estos desafíos de manera más detallada.

2.1.8. Tendencias y evolución de MDM

El campo del *Master Data Management* está en una etapa dinámica, con avances y evoluciones constantes para adaptarse a un entorno empresarial en rápida transformación. En esta sección, exploraremos algunas de las tendencias clave que están perfilando el futuro del MDM, sin entrar en demasiados detalles técnicos, ya que cada uno de estos temas será abordado en profundidad en capítulos posteriores.

Una de las tendencias más prominentes en el ámbito del MDM es la adopción creciente de soluciones basadas en la nube y *Software as a Service*. Esta tendencia no solo facilita una implementación más rápida y flexible sino que también ofrece una solución a los desafíos organizacionales, como la aceptación empresarial y la colaboración interdepartamental. Estos sistemas en la nube pueden funcionar como un catalizador para el cambio organizacional, ayudando a las empresas a superar las barreras tradicionales entre diferentes unidades de negocio.

La gobernanza de datos es otra área que está ganando mucho impulso. Con el incremento de la complejidad y la diversidad de los datos, la necesidad de políticas robustas y estructuras de gobernanza se ha vuelto más crítica que nunca. La gobernanza de datos efectiva puede resolver varios de los desafíos identificados previamente, como la calidad de los datos y la aceptación organizacional, y es esencial para cualquier estrategia de MDM exitosa.

En el contexto de la calidad de los datos, la evolución de técnicas más sofisticadas y herramientas especializadas está cambiando la forma en que las organizaciones abordan este desafío crítico. Las metodologías modernas no solo buscan corregir datos incorrectos o incompletos sino que también se esfuerzan por asegurar que los datos sean útiles y relevantes para las operaciones y decisiones empresariales.

Se espera que esta exploración de las tendencias y evolución en el campo del MDM despierte interés y curiosidad para entender cómo cada uno de estos temas contribuye a resolver los retos y limitaciones en este dominio. Las implicaciones de estas tendencias son vastas y tienen el potencial de revolucionar la forma en que las organizaciones gestionan sus datos. Como preámbulo a los capítulos que siguen, estos temas serán desglosados en detalle, proporcionando una visión completa de cómo las soluciones de MDM están adaptándose para satisfacer las necesidades de un mundo empresarial cada vez más complejo y dinámico.

2.2 *Software as a Service* y MDM en la Nube

2.2.1. Introducción a SaaS y a MDM en la nube

La gestión de datos maestros en el entorno de la computación en la nube y las soluciones de software como servicio presenta una revolución en la manera en que las empresas abordan la gestión de su información crítica. Este cambio es tan significativo como la transición de las bases de datos físicas a las soluciones en línea, pues plantea desafíos y oportunidades que transforman el paradigma de la gestión de datos. En un mundo cada vez más digitalizado, las organizaciones buscan formas más ágiles, escalables y rentables para gestionar sus datos. Aquí es donde entran en juego las soluciones de MDM basadas en la nube, ofreciendo la promesa de flexibilidad, eficiencia y colaboración mejorada.

La computación en la nube ha extendido las arquitecturas de TI empresariales más allá de las redes corporativas, impulsando la adopción de aplicaciones empresariales llave en mano ofrecidas como soluciones SaaS. Estas aplicaciones van desde la gestión de relaciones con los clientes (CRM) hasta la gestión de recursos humanos y el procesamiento de nóminas. Sin embargo, en el contexto de las soluciones de MDM, no todas las empresas están completamente a bordo con la transición a la nube. Algunas perciben riesgos asociados a legalidades o políticas corporativas que limitan su capacidad para desplegar soluciones de MDM en la nube.

Este dilema destaca la importancia de considerar cómo el middleware de integración en la nube encaja con arquitecturas híbridas de nube que abarcan servicios en las instalaciones y en la nube. Conforme se desarrolla esta tendencia, la integración de soluciones MDM en las instalaciones con aplicaciones SaaS se convierte en un punto crítico para lo que se conoce como la *empresa extendida*. [2]

La siguiente sección del capítulo se dedicará a explorar cómo ha evolucionado el concepto de software tradicional hacia SaaS, y cómo esto afecta a las soluciones de MDM en la nube. La comprensión de estas tendencias y su impacto en la gestión de datos maestros es crucial para cualquier organización que busque aprovechar al máximo las ventajas que ofrecen estas tecnologías emergentes.

2.2.2. Evolución del software tradicional a SaaS

El cambio de software tradicionalmente alojado en servidores locales a aplicaciones ofrecidas como un servicio a través de la nube (SaaS, por sus siglas en inglés, *Software as a Service*) ha sido una de las transformaciones más significativas en la arquitectura de TI empresarial en la última década.

¿Qué es SaaS?

SaaS es un modelo de servicio en la nube que ofrece una aplicación completa y administrada centralmente a través de la Internet. A diferencia del software tradicional, que se instala y ejecuta en servidores locales, SaaS es accesible a través de un navegador web, lo que elimina la necesidad de instalar, mantener y actua-

lizar el software a nivel local. El proveedor de SaaS se encarga de todo, desde la infraestructura y el software hasta las actualizaciones de seguridad, permitiendo que las empresas se centren más en sus problemas comerciales y menos en la administración de la infraestructura.

Diferencias entre SaaS, PaaS e IaaS

- **IaaS (Infrastructure as a Service):** Ofrece una infraestructura virtualizada, como servidores y almacenamiento, a través de la nube. Los usuarios tienen el control sobre los sistemas operativos y las aplicaciones, pero no gestionan el hardware subyacente.
- **PaaS (Platform as a Service):** Proporciona una plataforma y entorno que permite a los desarrolladores crear, alojar y desplegar aplicaciones. PaaS se encuentra en una capa por encima de IaaS y ofrece servicios adicionales como bases de datos, sistemas de colas y frameworks de desarrollo.
- **SaaS (Software as a Service):** Es la capa más alta, ofreciendo una aplicación completa como un servicio. En SaaS, todo, desde la infraestructura hasta el software, es responsabilidad del proveedor.

¿Por qué SaaS en el campo de MDM?

En el contexto de la Gestión de Datos Maestros, muchas organizaciones están optando por soluciones SaaS debido a la simplicidad y eficacia que ofrecen. Mientras que IaaS y PaaS requieren un grado significativo de gestión y personalización, lo que puede desviar los recursos de las tareas centrales del negocio, SaaS permite un acceso más fácil y rápido a capacidades críticas de MDM sin la carga de la gestión de la infraestructura y la plataforma.

No obstante, también es cierto que algunas organizaciones tienen reservas sobre la adopción de soluciones MDM en la nube debido a preocupaciones sobre la seguridad, la privacidad y la conformidad legal. Estas cuestiones se explorarán más detalladamente en la siguiente sección.

En resumen, la evolución hacia SaaS ha cambiado fundamentalmente la forma en que las organizaciones abordan la TI, permitiéndoles ser más ágiles y centrarse en lo que mejor saben hacer, mientras aprovechan las ventajas de la computación en la nube.

2.2.3. Soluciones Cloud para MDM

La Nube en MDM y SaaS

La gestión de datos maestros en la nube se ha convertido en una solución emergente en el mundo empresarial, alineada con la creciente adopción de tecnologías cloud para una variedad de aplicaciones empresariales. La nube ofrece una arquitectura flexible que puede escalar rápidamente para satisfacer las necesidades cambiantes de las empresas en el ámbito de la gestión de datos. Al integrar

MDM con modelos de Software como Servicio, las empresas obtienen una doble ventaja: la escalabilidad y la flexibilidad inherentes a las soluciones en la nube, y el modelo de entrega de software ágil y eficiente que caracteriza al SaaS.

Dentro de esta arquitectura, el término *multitenant* se vuelve relevante. En un entorno multitenant, una única instancia del software sirve a múltiples inquilinos o grupos de usuarios, permitiendo un uso más eficiente de los recursos y ofreciendo a cada inquilino la posibilidad de personalizar la aplicación según sus necesidades específicas. Este es un enfoque comúnmente asociado con SaaS MDM y ofrece la mayor flexibilidad y agilidad[6].

Informatica MDM Cloud y Customer 360

En este contexto, Informatica MDM Cloud² emerge como una solución robusta y escalable para la gestión de datos maestros. Esta plataforma ha sido diseñada para abordar las diversas preocupaciones en torno a la calidad de los datos, la gobernabilidad, la seguridad, y la arquitectura subyacente. Con una base en microservicios, la solución puede escalar fácilmente para satisfacer los requisitos de cualquier tamaño de empresa y puede adaptarse rápidamente a los cambios en las necesidades del negocio.

Para la parte práctica de este Trabajo Fin de Grado, se ha empleado concretamente Customer 360 dentro de Informatica MDM Cloud. Esta herramienta específica permite una visión 360 grados del cliente, facilitando así una gestión de datos centrada en el cliente que se detallará más adelante.

Consideraciones Finales

Al decidir entre una solución en la nube o una solución local (on-premises), las organizaciones deben considerar diversos factores como el control, la seguridad, la escalabilidad, el cumplimiento y el coste[9]. Sin embargo, para las empresas que ya están cómodas con las soluciones SaaS, adoptar una oferta de MDM SaaS en lugar de una solución local es una transición lógica.

2.2.4. Comparación de soluciones On Premise y Cloud

En el ámbito de la gestión de datos maestros o *Master Data Management*, la elección entre implementaciones en la nube o on-premise es crucial. La nube ha ganado prominencia en los últimos años, en gran medida debido a su integración con modelos de Software como Servicio, que ofrecen soluciones flexibles y escalables sin requerir una gran inversión inicial en hardware. Ahora, veamos las ventajas y desventajas de cada enfoque.

MDM en la Nube

- **Ventajas**

²<https://www.informatica.com/products/master-data-management.html>

1. **Sin inversión inicial:** No requiere una gran inversión de capital en hardware y servidores.
2. **Escalabilidad:** Se puede escalar fácilmente según las necesidades y el presupuesto de la organización.
3. **Recursos de TI:** Reduce la carga sobre los recursos de TI al externalizar la infraestructura.

■ Desventajas

1. **Acceso limitado:** Acceso limitado a servidores e infraestructura.
2. **Dependencia de la conectividad:** Requiere una conexión de red estable.
3. **Costos variables:** Los costos pueden aumentar con un mayor uso y demanda de ancho de banda.

MDM On-Premise

■ Ventajas

1. **Bajos costos de ancho de banda:** Al estar localizado en las instalaciones, los costos de ancho de banda son generalmente más bajos.
2. **Mayor seguridad:** Ofrece una mayor seguridad y privacidad al permitir personalizaciones más detalladas.
3. **Control sobre el hardware:** Mayor control sobre el hardware del servidor.

■ Desventajas

1. **Inversión inicial:** Requiere una gran inversión de capital para adquirir servidores y establecer infraestructura.
2. **Responsabilidad de cumplimiento:** La organización asume toda la responsabilidad del cumplimiento normativo.
3. **Recursos de TI:** Necesita más recursos de TI para administrar y mantener la infraestructura.
4. **Limita la escalabilidad:** Las infraestructuras on-premise son más difíciles de escalar en comparación con las soluciones en la nube.

La elección entre MDM en la nube y on-premise dependerá de varios factores, como los requisitos de seguridad, la disponibilidad de recursos de TI y el presupuesto. Para las empresas que buscan minimizar los costos iniciales y prefieren un modelo de suscripción, las soluciones en la nube son ideales. Por otro lado, las organizaciones que tienen estrictas regulaciones de seguridad y privacidad pueden encontrar más adecuado un modelo on-premise.

2.3 Gobernanza de Datos

2.3.1. Definición y alcance

Introducción al Concepto de Gobernanza de Datos

La Gobernanza de Datos es un conjunto de prácticas y políticas diseñadas para garantizar la gestión eficaz de los activos de información de una organización. No es simplemente una tarea técnica, sino que es una función estratégica que asegura que los datos en una empresa se manejan de manera coherente y segura, cumpliendo con todas las normativas y políticas empresariales. La gobernanza de datos está destinada a garantizar que los datos cumplan con las expectativas de diversos fines empresariales, dentro del contexto de la administración de datos, propiedad, cumplimiento, privacidad, seguridad, y riesgos asociados a los datos. Resumiendo, la gobernanza de datos es el ejercicio de autoridad y control (planificación, monitoreo y aplicación) sobre la gestión de activos de datos.[3]

Importancia de la Gobernanza de Datos en el Contexto General de la Gestión de Datos

En un entorno empresarial cada vez más impulsado por los datos, la gobernanza de datos se ha vuelto fundamental para asegurar que los datos sean confiables, seguros, y utilizables. Más que un lujo o una opción, se ha convertido en una necesidad. Su importancia radica en cómo habilita a las organizaciones a tomar decisiones informadas basadas en datos de alta calidad. Afecta directamente a la eficacia de las estrategias de gestión de datos y se entrelaza con otros elementos clave como la calidad de datos y la gestión de datos maestros.

Alcance de la Gobernanza de Datos

La gobernanza de datos cubre un rango amplio de aspectos, algunos de los cuales son:

- **Calidad de Datos:** Asegura que los datos sean precisos, consistentes y accionables.
- **Seguridad de Datos:** Protege los datos contra accesos no autorizados y violaciones de seguridad.
- **Privacidad de Datos:** Asegura que los datos personales sean manejados respetando todas las leyes y regulaciones aplicables.
- **Cumplimiento:** Incluye el seguimiento de las regulaciones y leyes, como GDPR en Europa o Sarbanes-Oxley en los Estados Unidos.
- **Propiedad y Administración de Datos:** Establece quién es responsable de qué datos dentro de la organización.

Cada uno de estos aspectos se puede desglosar en un conjunto de políticas de información que reflejan las necesidades y expectativas empresariales. Estas políticas, a su vez, se materializan en un conjunto de reglas o restricciones asociadas con las definiciones, formatos, y usos de los elementos de datos subyacentes.

En resumen, la gobernanza de datos se trata de instituir un conjunto de políticas, procedimientos y responsabilidades que guían la gestión efectiva de los datos de una organización. Su alcance es amplio y su impacto es profundo, especialmente en la era moderna donde el no cumplimiento con los requisitos externos de informes puede resultar en sanciones severas.

Este marco de gobernanza de datos sienta las bases para comprender los intrincados detalles de la gestión de datos maestros que se discutirán en las secciones subsiguientes.

2.3.2. Retos y oportunidades

Retos:

- **Riesgos Empresariales y Financieros:** El incumplimiento de las políticas de información puede poner en riesgo los objetivos financieros de la empresa. Por ejemplo, errores en informes financieros pueden resultar en una pérdida de valor de la empresa.
- **Informe Regulatorio:** La necesidad de generar documentos y reportes que demuestren el cumplimiento de diversas regulaciones introduce riesgos en la precisión y auditoría de los informes.
- **Conocimiento de la Entidad:** Mantener un conocimiento detallado de las partes con las que la empresa hace negocios es crucial para entender y reducir tanto riesgos empresariales como regulatorios.
- **Protección de la Información:** El cumplimiento de regulaciones como HIPAA y Graham-Leach-Bliley requiere que las organizaciones protejan la información individual para limitar las violaciones de datos.
- **Limitación del Uso:** Las políticas de gobernanza también deben abordar cómo se utilizan, comparten y acceden a los conjuntos de datos.
- **Consenso para la Coordinación y Colaboración:** La transición hacia el uso de datos maestros implica la necesidad de un acuerdo entre divisiones y unidades de negocio, lo que puede ser un reto en entornos empresariales.
- **Semántica:** La consolidación de datos en un ambiente de MDM requiere la resolución de variaciones sutiles en definiciones, formatos y usos de elementos de datos.

Oportunidades:

- **Mitigación de Riesgos:** Un programa de gobernanza efectivo puede ayudar a identificar y mitigar riesgos asociados con el incumplimiento de políticas de información, mejorando así la estabilidad y el control financiero.

- **Cumplimiento Asegurado:** Un régimen sólido de gobernanza de datos asegura que la empresa pueda generar informes precisos y auditables, satisfaciendo así las demandas de cumplimiento.
- **Mejora de la Calidad de los Datos:** La gobernanza de datos permite un enfoque sistemático para mejorar la calidad de los datos, lo que a su vez mejora la toma de decisiones.
- **Protección Mejorada:** Implementar políticas y procedimientos rigurosos para la protección de datos puede limitar significativamente las violaciones de datos y proteger la información personal.
- **Uso Óptimo de Datos:** Una buena gobernanza puede asegurar que los datos se utilicen de manera que estén alineados con las políticas de la empresa, lo que puede abrir nuevas vías para la utilización de datos.
- **Colaboración Eficiente:** La gobernanza efectiva puede facilitar la colaboración y el intercambio de información entre diferentes divisiones o unidades de negocio.
- **Claridad Semántica:** La gobernanza permite la armonización de los elementos de datos, reduciendo los riesgos asociados con la variabilidad semántica.

En conjunto, la gobernanza de datos ofrece una estructura que no sólo mitiga riesgos sino que también aprovecha los datos como un activo estratégico, ofreciendo varias oportunidades para mejorar la operación y la toma de decisiones en la organización.

2.3.3. Importancia en MDM

La gobernanza de datos es una pieza clave para el éxito de cualquier estrategia de *Master Data Management*. Este papel no es puramente teórico sino práctico y de vital importancia en la planificación e implementación de objetivos de MDM. Aquí se subrayan algunas razones específicas por las que la gobernanza de datos es indispensable en el contexto de MDM.

Establecimiento de Prioridades y Toma de Decisiones

En una estrategia MDM multi-dominio, la gobernanza de datos ayuda a establecer prioridades, hacer selecciones de dominio y tomar decisiones sobre la propiedad de los datos. Estas decisiones son cruciales para alinear los datos maestros con los objetivos empresariales y las operaciones del día a día.

Resolución de Conflictos y Problemas Funcionales

Durante las fases de planificación e implementación de MDM, es probable que surjan numerosos conflictos y problemas relacionados con el uso, la definición y el alineamiento de los datos maestros. La capacidad para resolver muchos

de estos problemas interfuncionales supera el alcance del MDM PMO (Project Management Office). Aquí es donde una sólida gobernanza de datos se convierte en una necesidad, proporcionando un proceso autorizado para abordar diversas necesidades de resolución de problemas que surgirán.

Camino Crítico y Dependencia

Si no existe ya un proceso de gobernanza de datos, la estrategia y el plan de ejecución para MDM deberán identificar la gobernanza de datos como una dependencia crítica. En este escenario, el propio programa de MDM deberá convertirse en la fuerza impulsora para la implementación de un proceso de gobernanza de datos.

Coordinación y Colaboración

La gobernanza de datos facilita la colaboración y el intercambio de información no solo dentro del PMO de MDM sino también entre diferentes áreas de proceso empresarial. Esto se correlaciona estrechamente con el papel que desempeñan los administradores de datos, quienes deben estar comprometidos en las áreas operativas y de proceso donde se necesitan.

En resumen, la gobernanza de datos no es un complemento opcional sino una necesidad imperante para garantizar que el MDM sea efectivo, eficiente y alineado con las metas empresariales. Su importancia radica en su capacidad para servir como el mecanismo regulador y autoritativo que permite a la empresa manejar sus activos de datos de manera más efectiva.

2.4 Calidad de Datos

2.4.1. Introducción a la Calidad de Datos

La calidad de datos es un concepto que va más allá de la mera precisión o exactitud de los datos en sí. Se trata de un marco integral que evalúa los datos en términos de su fiabilidad, coherencia, relevancia y utilidad dentro de un contexto empresarial específico. En términos simples, la calidad de datos se refiere a qué tan confiables, coherentes y útiles son los datos para los propósitos para los cuales se recolectan y utilizan.

La idea de que los datos deben ser de alta calidad no es solo un objetivo en sí mismo, sino también una expectativa intrínseca en cualquier sistema de información empresarial. Esta expectativa se convierte en un pilar fundamental en entornos que dependen fuertemente de datos para su funcionamiento, desde la toma de decisiones estratégicas hasta las operaciones cotidianas.

La calidad de datos es especialmente crítica en entornos de negocios modernos, que se caracterizan por la interconexión, la rapidez y la complejidad. En estos entornos, los datos a menudo se recopilan de múltiples fuentes, en diferentes formatos y estructuras, y se utilizan para una variedad de propósitos. La calidad

deficiente de los datos en uno de estos aspectos puede tener efectos de amplio alcance, afectando la eficiencia, la rentabilidad y la reputación de una organización.

La confianza en los datos es un factor clave en la calidad de datos. Los usuarios finales, desde ejecutivos hasta analistas de datos, deben confiar en que los datos son lo suficientemente buenos para respaldar sus decisiones y acciones. Esta confianza se establece y mantiene a través de una serie de mecanismos, incluidos los procesos de validación, verificación y supervisión.

Es vital comprender que la calidad de datos no es un concepto estático; es dinámico y debe adaptarse a los cambios en los entornos empresariales, las tecnologías y las necesidades de los usuarios. Por lo tanto, es fundamental para las organizaciones adoptar un enfoque proactivo para gestionar y mejorar continuamente la calidad de sus datos.

En resumen, la calidad de datos se centra en la fiabilidad, utilidad, coherencia y precisión de los datos dentro de un contexto específico, y es un factor crítico para el éxito de cualquier iniciativa empresarial basada en datos.

2.4.2. Parámetros de la Calidad de Datos

A continuación, se detallan algunas de las dimensiones clave:

Unicidad

La unicidad se refiere a que cada entidad dentro de un conjunto de datos debe existir de manera única, sin duplicaciones. Este parámetro garantiza que cada registro tenga un identificador único que facilite su localización y gestión. Para asegurar la unicidad, se pueden implementar procesos de análisis de duplicados y resolución de identidades.

Precisión

La precisión se centra en qué tan bien los datos representan los objetos o conceptos del mundo real a los que se supone deben modelar. Puede implicar cotejar los datos con fuentes de información corroborativas o de referencia. La precisión es desafiante de monitorear, ya que el mundo real puede cambiar con el tiempo, afectando la precisión de los datos.

Consistencia

La consistencia se trata de que los valores de los datos en un conjunto sean coherentes con los valores en otro conjunto. La consistencia no solo es importante dentro de un único conjunto de datos, sino también entre diferentes conjuntos de datos que pueden existir en diferentes departamentos o aplicaciones dentro de una organización. Las reglas de consistencia se pueden establecer a varios niveles: a nivel de registro, entre registros y a lo largo del tiempo (consistencia temporal).

Compleitud

La completitud se refiere a la presencia de valores no nulos en campos de datos específicos. En otras palabras, todos los elementos de datos que son necesarios para un proceso o decisión particular deben estar presentes. La completitud puede ser obligatoria, opcional bajo ciertas condiciones, o incluso inaplicable en ciertos contextos.

Actualidad (Timeliness)

La actualidad se refiere al tiempo que transcurre entre cuando se espera que la información esté disponible y cuando realmente está lista para ser utilizada. Este parámetro es crucial para sistemas en tiempo real o para procesos empresariales que dependen de información actualizada.

Vigencia (Currency)

Este aspecto está relacionado con la frecuencia con la que se actualizan los datos y cuán actualizados están. Si la información cambia rápidamente en el mundo real, los datos deben actualizarse con una frecuencia igualmente alta para mantener su relevancia y utilidad.

Cumplimiento de Formato (Format Compliance) Este parámetro se centra en si los valores de los elementos de datos se almacenan, intercambian y presentan en un formato que sea coherente con las expectativas del dominio de valor. Las herramientas de análisis y estandarización se pueden usar para validar los datos contra formatos y patrones definidos.

Integridad Referencial

La integridad referencial implica que cada identificador único asignado a una entidad dentro de un entorno debe referirse efectivamente a una entidad existente en ese entorno. Este parámetro se asegura de que todas las claves extranjeras en una base de datos tengan valores que correspondan a una clave principal en otra tabla.

Cada uno de estos parámetros es una dimensión que ayuda a evaluar y asegurar la calidad de los datos. Juntos, forman un marco integral para medir, validar y monitorear la calidad de los datos en cualquier organización.

2.4.3. Impacto en MDM

La calidad de datos es un pilar crítico para el éxito de cualquier iniciativa de MDM. Su impacto en MDM es multifacético y abarca desde la identificación de objetos de datos maestros hasta la agregación de datos y su presentación en una forma única y coherente.

Identificación de Objetos de Datos Maestros

El análisis de perfil de datos, una funcionalidad avanzada en las herramientas modernas de calidad de datos, permite la identificación de objetos de datos maestros (como cliente o producto) en sus diferentes representaciones en toda la empresa. Esto es vital para la consolidación de registros y la eliminación de duplicados, lo que a su vez facilita una representación única y precisa de cada objeto.

Resolución de Anomalías

Las herramientas de calidad de datos son indispensables para descubrir y resolver anomalías que degradan el valor de los activos informativos de la empresa. La habilidad para identificar datos duplicados o inconsistentes y resolver estas anomalías es crucial para establecer una estrategia efectiva dentro de un marco de MDM.

Agregación de Datos y Autoridad Central

Las tecnologías modernas de integración de datos permiten la agregación transparente de datos para presentar una representación única y fácilmente identificable a través de una autoridad central. En ausencia de una estrategia de integración estandarizada, las iniciativas de MDM enfrentarían importantes obstáculos, incluida la necesidad de modernizar todas las aplicaciones de producción existentes.

Acceso y Interacción con la Autoridad Central

Una gestión de datos maestros efectiva también exige que las aplicaciones puedan interactuar con esta autoridad central de forma eficaz. Este nivel de interacción solo es posible si la calidad de los datos se mantiene alta, lo que permite una identificación precisa y una agregación efectiva de datos.

Relevancia para los Negocios

El impacto de la calidad de datos en MDM no es solo técnico sino también empresarial. Una representación precisa y confiable de datos maestros es fundamental para la toma de decisiones empresariales y para asegurar la coherencia en operaciones multifuncionales.

La convergencia entre las herramientas de calidad de datos y las de integración de datos demuestra la importancia intrínseca de la calidad de datos en la implementación exitosa de MDM. Las organizaciones que no invierten suficientemente en asegurar la calidad de datos en sus iniciativas de MDM están destinadas a enfrentar problemas de incoherencia de datos, falta de confianza en los datos y, en última instancia, fallos en la implementación de MDM.

2.5 Ciclo de vida de un Proyecto de MDM

En el presente apartado se pretende ejemplificar un caso de uso real para ilustrar cómo se integran todos los componentes previamente comentados, con el objetivo de aclarar posibles dudas debido a la gran carga teórica y de demostrar cómo se desarrolla un proyecto de Master Data Management en el ámbito empresarial internacional.

El caso de uso descrito corresponde a un proyecto real llevado a cabo durante el desempeño laboral del autor de este Trabajo Fin de Grado, lo que le confiere actualidad y relevancia. Esta experiencia proporciona una visión práctica y concreta de la implementación de soluciones MDM, destacando los retos y las estrategias empleadas para superarlos.

2.5.1. Obtención y análisis de los datos fuentes

El presente proyecto de ejemplo utiliza tres fuentes de datos claramente diferenciadas, denominadas A, B y C para mayor claridad. Las fuentes A y B están compuestas por archivos Parquet almacenados en Azure Blob Storage. La fuente C, en cambio, es una colección de bases de datos almacenadas en un servidor SQL. Para poder trabajar con estas fuentes, el primer paso es establecer las conexiones dentro de la plataforma de Informática Cloud, que es donde se ha llevado a cabo este proyecto real.

Perfilado de los Datos de las Fuentes

Una vez establecidas las conexiones, se procede al perfilado de los datos de las fuentes. Para ello, se utiliza la herramienta de Data Profiling, que recolecta de forma autónoma los datos de las bases de datos de las fuentes y los presenta para su posterior análisis.

Análisis de Calidad de Datos de los Perfilados Obtenidos

Una vez realizados los perfilados, se procede a analizar los datos atributo por atributo. La herramienta de Data Profiling genera informes automáticos que muestran la calidad de los datos de las fuentes, independientemente del tamaño de estas o del número de atributos. En esta fase, se analizan aspectos como la relevancia de los campos, claves primarias, porcentaje de atributos nulos o vacíos, formato de los atributos, y la frecuencia de aparición de valores específicos, entre otros.

2.5.2. Modelo de Datos en MDM

Una vez realizado el perfilado y el análisis de los datos fuentes, se procede a elaborar el modelo de datos a implementar en MDM utilizando Informática Cloud. Gracias al análisis previo, es posible establecer equivalencias entre los

campos de las diferentes fuentes para determinar la futura agrupación de los atributos. Esto permite crear un modelo de datos unificado y óptimo para la fusión de las tres fuentes, el cual será utilizado para gestionar los datos maestros o golden record.

Por ejemplo, se ha establecido un campo denominado `OrgName` en el modelo de datos definitivo. Este campo se alimentará del campo `_Name` de la fuente A, del campo `Organization` de la fuente B y del campo `CompName` de la fuente C. Este proceso de unificación asegura que los datos se integren de manera coherente y precisa, facilitando así la gestión centralizada de los datos maestros.

2.5.3. Integración y Transformación de Datos

A continuación, se establecen una serie de componentes que ayudan a la estandarización de los datos de las fuentes, permitiendo que los datos ingeridos sean óptimos para los procesos posteriores y fomentando así una mayor calidad de datos.

Elaboración de Reglas de Integración de Datos

A través de la herramienta de Data Quality, se establecen una serie de reglas para determinar automáticamente si los campos son válidos o no una vez ingeridos desde las fuentes. El objetivo es descartar cualquier dato que no cumpla con los estándares definidos por el cliente o los requisitos funcionales del modelo de datos. Algunos ejemplos de estas reglas incluyen la comprobación de la longitud del atributo de teléfono, la ausencia de espacios en los correos electrónicos, y la verificación de que no haya atributos vacíos.

Elaboración de Mapeos y Transformaciones de Datos

Una vez rechazados los datos erróneos o irrelevantes, se procede a la integración de mapeos para realizar transformaciones sobre los datos. Estas transformaciones aseguran que los datos se muestren correctamente una vez consolidados en el Golden Record. Para ello, se emplea la herramienta de Data Integration para realizar un mapeo por cada fuente. Estos mapeos, una vez automatizados, se encargan de suministrar los datos transformados a Customer 360 para su futura visualización o consolidación como Golden Record. Ejemplos de transformaciones incluyen la estandarización de los atributos a mayúsculas o minúsculas, la extracción de prefijos de país en los campos de identificación fiscal, y la normalización de caracteres regionales como la "ñ.º "ç", entre otros.

2.5.4. Elaboración de Reglas de Match and Merge

Posteriormente, se elaboran las reglas de match and merge, encargadas de detectar y fusionar posibles registros idénticos entre las diferentes fuentes de datos. Estas reglas no se limitan a fusionar registros con los mismos valores en sus campos, sino que también buscan posibles coincidencias en los atributos para llevar a cabo estas fusiones.

Por ejemplo, las coincidencias exactas pueden incluir el mismo número de identificación fiscal o nombre. En cuanto a los registros no idénticos, se pueden considerar casos como calles con y sin prefijos como "Vía", "Calle", "Avenida", etc., o teléfonos con y sin prefijo nacional, entre otros casos específicos.

Cabe destacar que el presente modelo de datos presenta una jerarquía. En caso de discrepancia, los datos de la fuente A tendrán un mayor índice de supervivencia que los datos de las fuentes B y C, y los datos de la fuente B tendrán mayor prioridad que los de la fuente C.

2.5.5. Planificación de Procesos de Datos

A continuación, es necesario llevar a cabo una serie de planificaciones tanto para la ingesta de datos como para su egreso y transformación en el golden record.

Planificación de Ingestas Iniciales y Diarias de Datos de las Fuentes

En este caso particular, la ingesta inicial de datos ya se ha realizado en el ambiente de desarrollo; sin embargo, será necesario trasladarla al ambiente de producción, replicándola para cuando el cliente necesite iniciar el servicio. Además, este proyecto tiene la particularidad de que el cliente prefiere una ingesta de datos diaria automatizada siempre a la misma hora, en lugar de una ingesta a demanda. Por lo tanto, será necesario programar un job para ello. En el caso concreto del proyecto, se ha optado por realizar dicha ingesta a las 7:30 de la mañana, de lunes a viernes, lo que añadirá nuevos datos a procesar por todas las herramientas mencionadas en los apartados anteriores.

Planificación del Egreso de los Datos y del Golden Record

Una vez que los datos han sido ingeridos, transformados y procesados por las distintas herramientas, se procede a llevar a cabo el egreso de los datos maestros o golden record. Estos datos son el resultado de todos los procesos previos y representan la versión más verídica y auténtica de los datos. Estos datos ya no serán modificados, ya que se consideran absolutos, y el cliente podrá acceder a ellos a través de Customer 360 o exportándolos en un archivo CSV.

2.5.6. Pruebas y Análisis de Rendimiento

Por último, como en prácticamente cualquier proceso de software, se realiza un proceso de pruebas y testeo de los componentes para asegurar su correcto funcionamiento. Además, se llevan a cabo pruebas de rendimiento para analizar los tiempos de respuesta y verificar si se adaptan a los requisitos establecidos por el cliente.

Con esto, se concluye el ciclo de vida de un proyecto de MDM en el ámbito empresarial internacional. Al finalizar este proceso, el cliente contará con una versión unificada de todos los datos actuales y futuros, conocida como Golden

Record, todo realizado de forma automatizada y sin requerir un elevado soporte o intervención. Además de la integración y desarrollo en la plataforma de Informática Cloud, se presenta también una amplia documentación que incluye explicaciones sobre las reglas de calidad de datos, mapeos, match and merge, pruebas, entre otros aspectos.

2.6 DAMA Data Management Body of Knowledge

No suele ser habitual el estudio de una fuente directamente en un proyecto de este calibre, pero es necesario discutir la importancia del DAMA³ *Data Management Body of Knowledge* (DAMA DMBoK) en el mundo de la ingeniería de datos actual. Este recurso integral actúa como un compendio de mejores prácticas, terminología estandarizada y directrices que sirven como un marco de trabajo para cualquier profesional en el ámbito de la gestión y la ingeniería de datos. En un entorno empresarial que se está volviendo cada vez más dependiente de la información y del análisis de datos para la toma de decisiones estratégicas, la relevancia del DAMA DMBoK es más aparente que nunca.

En una era donde los datos son considerados como recursos de valor incalculable, las organizaciones se enfrentan al desafío de manejar cantidades inmensas de datos de manera efectiva y eficiente. Esto incluye desde la recopilación, almacenamiento y procesamiento hasta la análisis y la visualización de datos. Aquí es donde el DAMA DMBoK se convierte en una herramienta invaluable, ofreciendo una hoja de ruta que guía a las empresas en la gestión de sus activos de datos de una forma que es tanto sostenible como escalable.

El cuerpo de conocimiento proporcionado por DAMA DMBoK aborda una amplia gama de temas, incluyendo la calidad de datos, la gobernanza de datos, la arquitectura de datos, la seguridad de datos y mucho más. Al proporcionar un marco de trabajo que es a la vez riguroso y flexible, permite que las organizaciones adapten las mejores prácticas a sus necesidades específicas, fomentando así un enfoque más coherente y eficaz hacia la gestión de datos.

Por todo ello, examinar y comprender los preceptos y directrices contenidos en el DAMA DMBoK no es simplemente un ejercicio académico, sino una necesidad práctica para cualquier empresa que busque maximizar el valor de sus activos de datos en un entorno empresarial cada vez más competitivo y orientado a datos.

2.6.1. Visión General: Una Introducción al Cuerpo de Conocimientos de DAMA sobre Gestión de Datos

El DAMA *Data Management Body of Knowledge* sirve como un faro en el mundo complejo y en constante evolución de la gestión de datos. Este recurso brinda un marco integral que cubre los principales aspectos y disciplinas involucradas en la gestión de datos, ofreciendo un mapa conceptual y metodológico que los

³<https://www.dama.org/cpages/home>

profesionales pueden seguir para diseñar, implementar y mantener estrategias efectivas de gestión de datos.

El DAMA DMBok se organiza en áreas de conocimiento específicas, que van desde la gobernanza de datos y la calidad de datos hasta la arquitectura, el diseño y la construcción de sistemas de datos. También se abordan temas como la seguridad de datos, la privacidad y el cumplimiento, entre otros. Cada una de estas áreas de conocimiento se desglosa en subtemas, proporcionando una visión detallada de las prácticas recomendadas, los enfoques y las tecnologías relevantes.

Este cuerpo de conocimientos no es un marco de trabajo estático; se actualiza y evoluciona para reflejar las tendencias emergentes, los avances tecnológicos y las cambiantes necesidades del negocio. Esto lo convierte en un recurso vivo que se adapta a las realidades del paisaje de datos en rápido movimiento.

A través de su enfoque estructurado pero amplio, el DAMA DMBok ofrece a los profesionales de la gestión de datos las herramientas que necesitan para enfrentar los desafíos contemporáneos en la gestión de activos de datos. No solo establece un lenguaje común que facilita la comunicación efectiva entre los diversos actores involucrados, sino que también proporciona una serie de métricas y directrices que pueden ayudar a medir el rendimiento y la eficacia de las iniciativas de gestión de datos.

En resumen, el DAMA DMBok actúa como una fuente única de sabiduría en el ámbito de la gestión de datos, ofreciendo una visión panorámica que es invaluable tanto para los novatos como para los expertos en el campo.

2.6.2. Relación con MDM: Complementariedad entre DAMA y Estrategias de MDM

DAMA DMBok y las estrategias de Gestión de Datos Maestros no son entidades aisladas, sino que se complementan de forma significativa en el ecosistema de la gestión de datos. Mientras que el DAMA DMBok proporciona un marco amplio y multidisciplinario para la gestión de datos en general, las estrategias y técnicas de MDM se centran más en la gestión y consolidación de datos maestros específicos que son críticos para las operaciones del negocio. Aquí es donde los caminos de ambos se cruzan y se enriquecen mutuamente.

- **Lenguaje Común:** Uno de los principales aportes de DAMA es que proporciona un vocabulario y una terminología comunes que pueden ser utilizados en diversas disciplinas de la gestión de datos, incluido el MDM. Este lenguaje común es fundamental para que los equipos de MDM colaboren de manera efectiva con otros actores en la organización.
- **Mejores Prácticas y Directrices:** DAMA sirve como una fuente confiable de mejores prácticas que pueden ser incorporadas en la planificación e implementación de una estrategia de MDM. Estas mejores prácticas pueden incluir aspectos como la gobernanza de datos, la calidad de datos y la privacidad, que son vitales en el contexto de MDM.

- **Gobernanza de Datos:** Ambos marcos destacan la importancia de una sólida gobernanza de datos. En una estrategia de MDM, la gobernanza de datos se vuelve aún más crucial, dado que los datos maestros suelen ser compartidos y utilizados por múltiples sistemas y unidades de negocio. DAMA ofrece una estructura para implementar esta gobernanza de una manera más organizada y efectiva.
- **Métricas y KPIs:** DAMA DMBok propone un conjunto de métricas y KPIs que pueden ser adaptados y aplicados en el entorno de MDM para evaluar el rendimiento, la calidad y el valor empresarial de los datos maestros.
- **Adaptabilidad y Flexibilidad:** El DAMA DMBok, al ser un marco adaptable, permite a las organizaciones seleccionar aquellos componentes que son más relevantes para sus necesidades específicas de MDM, desde la calidad de datos hasta la seguridad y el cumplimiento.
- **Interoperabilidad:** Las directrices de DAMA pueden ayudar a garantizar que las iniciativas de MDM sean compatibles con otros sistemas y procesos de gestión de datos dentro de la organización, mejorando así la interoperabilidad y la cohesión entre los diferentes departamentos y funciones.

En conjunto, la relación entre DAMA y MDM no es simplemente de coexistencia, sino de cooperación estratégica. Al integrar los principios y directrices de DAMA en las estrategias de MDM, las organizaciones tienen la oportunidad de potenciar la eficacia de su gestión de datos maestros, alineándola con un marco más amplio y robusto para la gestión de datos en general.

2.7 Contexto empresarial: Infoverity S.L.

Este Trabajo de Fin de Grado (TFG) se ha llevado a cabo en el marco de Infoverity⁴, una empresa que se erige como líder y pionera en el sector de MDM.

Infoverity se distingue en el mercado por su enfoque integral y especializado en la gestión de datos. Con una rica trayectoria en la implementación de soluciones MDM de alto impacto, la empresa no solo se adhiere a las mejores prácticas del sector, sino que también contribuye activamente a la evolución de los estándares y metodologías relacionadas con la gestión de datos. Su compromiso con la innovación y la calidad lo demuestran una amplia cartera de clientes satisfechos que abarcan diversas industrias y geografías.

La empresa ofrece una gama de servicios que incluyen la gobernanza de datos, la calidad de datos, la integración de datos y, por supuesto, las soluciones de MDM. La capacidad de Infoverity para ofrecer servicios de extremo a extremo en el ciclo de vida de los datos la convierte en un socio estratégico para las organizaciones que buscan transformar su gestión de datos en una ventaja competitiva.

En este contexto, el presente TFG no solo busca entender las complejidades y desafíos del MDM, sino también cómo las prácticas y soluciones implementadas por Infoverity pueden servir como un modelo ejemplar para otras empresas en el

⁴<https://www.infoverity.com/>

campo. El compromiso de Infoverity con la excelencia y la innovación en MDM proporciona un escenario ideal para la investigación y el análisis profundo de las mejores prácticas y tendencias emergentes en este ámbito crucial de la ingeniería de datos.

La colaboración con Infoverity para la realización de este TFG ofrece una perspectiva única sobre los retos prácticos y las soluciones innovadoras en el mundo real de la gestión de datos maestros. Este contexto empresarial no solo enriquece el alcance académico del trabajo, sino que también lo sitúa firmemente en las aplicaciones y desafíos empresariales contemporáneos en el campo del MDM.

2.7.1. Historia de la empresa

Infoverity fue fundada en el año 2011 con el objetivo principal de resolver problemas empresariales originados por la proliferación y diversificación de datos en grandes corporaciones. Desde sus inicios, la empresa ha experimentado un crecimiento notable y ha adquirido un prestigio global en el campo de los servicios profesionales centrados en la gestión de datos.

La empresa ha recibido múltiples reconocimientos que avalan su excelencia y compromiso en el ámbito de la gestión de datos. Infoverity está destacada en la Guía del Mercado de Proveedores de Servicios Externos para MDM de Gartner, y ha sido incluida en listas de relevancia empresarial como la Inc. 5000 y la Columbus Fast 50. Además, ha sido nombrada como uno de los mejores lugares para trabajar en el sector de TI tanto por Computerworld de IDG como por Columbus Business First.

Con respecto a su alcance geográfico, la sede global de Infoverity se encuentra en Dublin, Ohio, en los Estados Unidos. Sin embargo, su presencia internacional no se limita a América del Norte. La empresa cuenta con equipos de servicio y soporte adicionales situados en diferentes partes del mundo, incluyendo Valencia, España, que funciona como la sede para la región de EMEA (Europa, Oriente Medio y África), así como en Alemania y Rusia.

A lo largo de su trayectoria, Infoverity ha ampliado su portfolio de servicios para incluir no solo la gestión de datos maestros sino también la gestión de la información de producto (PIM), la gobernanza de datos y analítica, y servicios y soluciones de alojamiento gestionado. Estos servicios han resultado ser de valor inestimable para clientes de diversos sectores, incluidos el comercio minorista, bienes de consumo, manufactura, finanzas y atención sanitaria.

La evolución de Infoverity demuestra una adaptabilidad y visión de futuro notables, posicionándola como un actor clave y un referente en el sector de la gestión de datos.

2.7.2. Servicios ofrecidos

Expertos en MDM y PIM

Infoverity ofrece una completa gama de soluciones de gestión de la información, con un enfoque en lograr rendimientos tangibles a partir de una óptima

gestión de datos. Los principales servicios y grupos de expertos que forman su portafolio son:

Estrategia y Hoja de Ruta de MDM

Este servicio se centra en la planificación y ejecución de estrategias de gestión de datos maestros. La filosofía de Infoverity en este ámbito es que la planificación es esencial, pero la ejecución es primordial. Su objetivo es establecer un portafolio de tácticas que mejoren continuamente la calidad, el contexto y la disponibilidad de la información de manera eficaz y rentable.

Grupo de Soluciones al Cliente

Este equipo aborda estrategias y su ejecución para mejorar la experiencia del cliente, reducir el riesgo crediticio y aumentar los retornos de los clientes mediante análisis. El grupo está altamente cualificado para abordar temas que son importantes para los proyectos de sus clientes.

Grupo de Soluciones de Producto

Este grupo planifica y ejecuta soluciones para la optimización de las cadenas de suministro, mejora de la eficiencia en la comercialización de productos, incremento de la velocidad de los productos y habilitación del comercio omnicanal.

Gobernanza de Datos y Analítica

Este servicio se centra en proporcionar un marco para la gobernanza de datos y ofrecer análisis que puedan ayudar a las empresas a tomar decisiones informadas y estratégicas.

Grupo de Servicios Operacionales

Este equipo proporciona servicios de soporte de aplicaciones de ITSM y/o alojamiento de infraestructuras centrados en la gestión de solicitudes, incidentes, problemas y cambios. El modelo de Infoverity ofrece la flexibilidad para adaptar el soporte a las necesidades específicas de cada cliente, ya sea para el mantenimiento de la producción o para un equipo completo de soporte de aplicaciones.

Soluciones Comerciales a Largo Plazo

Además de estos servicios, Infoverity también ofrece soluciones comerciales que ayudan a las empresas a tomar decisiones de alta velocidad con más confianza. Estas soluciones ofrecen un contexto vital para los responsables empresariales en áreas como:

- Marketing

- Operaciones de Ventas
- Cumplimiento y Regulación
- Comercialización
- Seguridad de la Información y Auditoría
- Gestión de Crédito
- Gestión de Compras y Proveedores
- Gestión de la Cadena de Suministro e Inventario
- Diseño de Producto
- Comercio Electrónico

En resumen, Infoverity es un socio integral que ofrece una amplia gama de servicios y soluciones especializadas para abordar los desafíos más complejos de la gestión de datos en grandes empresas.

2.7.3. Reconocimientos e Hitos

Experiencia en la Gestión de Datos

Infoverity ha demostrado constantemente su capacidad para ofrecer soluciones de gestión de datos a largo plazo, escalables y confiables en una amplia variedad de industrias. Desde organizaciones sin fines de lucro hasta el sector minorista, ha establecido arquitecturas de información empresarial confiables y personalizadas.

Casos de Estudio Destacados

- **Fabricante de Ropa de Marca:** Identificación y corrección de fallos en la gestión del flujo de datos, junto con la implementación de un sistema de documentación adecuado y un plan de acción de tres fases.
- **Holding Financiero:** Construcción de un sistema de automatización centralizado que mejora la visibilidad del riesgo crediticio del cliente.
- **Distribuidor de Componentes Electrónicos:** Implementación de una plataforma PIM modernizada que mejora la experiencia del cliente mediante el uso de contenido de producto enriquecido.
- **Minorista Nacional de Joyería:** Eliminación de silos de datos de productos, mejora de la experiencia digital en todos los canales web y automatización del flujo de productos al mercado.

Asociaciones Estratégicas

Se ha asociado con líderes tecnológicos mundiales, lo que permite una integración, limpieza, gestión y administración de información de primera clase en toda la empresa. Algunos de sus socios más destacados incluyen:

- **Adobe:** Líder mundial en soluciones de medios y marketing digitales.
- **Akeneo:** Líder mundial en soluciones de Gestión de Información de Productos (PIM).
- **BigID:** Ofrece visibilidad y control de datos a través de todo el panorama de datos.
- **Boomi:** Facilita las experiencias integradas para maximizar el valor de sus datos.

Sectores de Experiencia

El equipo está compuesto por expertos que no solo son líderes en sus campos, sino que también se integran a la perfección con las organizaciones de nuestros clientes. Han ofrecido soluciones transformadoras en diversos sectores, incluyendo:

- **Servicios Financieros:** Bancos comerciales, minoristas e inversiones.
- **Atención Médica:** Proveedores de atención clínica y productos médico-quirúrgicos.
- **Retail:** Desde moda y deportes hasta bienes de lujo.
- **Seguros:** Incluyendo seguros de propiedad y de vida.
- **Manufactura Industrial:** Desde controles industriales hasta defensa y aeroespacial.
- **Bienes de Consumo Empaquetados:** Desde bienes de consumo hasta alimentos y bebidas.

Estos logros y asociaciones subrayan nuestro compromiso continuo con la excelencia y la innovación, demostrando por qué Infoverity es un líder confiable en soluciones de gestión de información.

2.8 Crítica al estado del arte

El análisis de los trabajos previos presentados en la Escuela Técnica Superior de Ingeniería Informática (ETSINF) revela algunas lagunas significativas y oportunidades para la contribución al cuerpo de conocimiento en el ámbito de *Master Data Management* y *Software as a Service*. Aunque se ha cubierto terreno valioso en estas áreas, los enfoques anteriores han tendido a centrarse principalmente

en aspectos técnicos como la instalación y utilización de software específico como Informatica MDM. Esta focalización ha dejado un espacio considerable para la exploración y desarrollo en áreas menos tratadas que son igualmente críticas, particularmente en lo que se refiere al marco teórico y transversal de la gestión de datos y tecnologías en la nube.

Una de las lagunas más evidentes es el tratamiento superficial o directamente la omisión del análisis sobre la integración de MDM y SaaS, un área que está ganando relevancia en el mundo empresarial y tecnológico actual. Además, se observa una escasez en el abordaje teórico de cómo estas tecnologías se cruzan con otros componentes críticos de la gestión de datos, como la gobernanza y calidad de datos. Este tipo de análisis teórico es crucial para cualquier organización que busque comprender integralmente las implicancias de adoptar soluciones MDM en un entorno en la nube.

Otro aspecto a destacar es que, hasta la fecha, no se ha presentado ningún Trabajo de Fin de Grado (TFG) que aborde específicamente el uso de MDM en un entorno SaaS, ni que examine las capacidades y limitaciones de los servicios en la nube de Informatica. Este punto es particularmente relevante dada la evolución hacia soluciones en la nube y la creciente importancia de la agilidad y escalabilidad en la gestión de datos.

Es imperativo subrayar que el enfoque previo en la instalación y uso de programas y servicios puede considerarse limitado si no se combina con una sólida base teórica que permita contextualizar esos aspectos técnicos en el ámbito empresarial más amplio. Para implementar soluciones eficaces, se necesita más que una comprensión técnica; es fundamental entender cómo se enmarcan estas tecnologías en la estrategia de datos de una organización, cómo interactúan con las políticas de gobernanza y qué tipo de impacto pueden tener en la eficiencia operacional y la toma de decisiones.

En resumen, este TFG busca llenar el vacío existente en la literatura y en los trabajos anteriores de la ETSINF, al ofrecer una investigación que no solo se centra en el aspecto técnico de la instalación y uso de MDM y SaaS, sino que también profundiza en el contexto teórico y transversal de estas tecnologías. Este enfoque integral permitirá una mejor comprensión de cómo las soluciones de MDM, especialmente en un entorno SaaS, pueden ser implementadas de manera más eficaz y estratégica en una organización. Este trabajo es pionero en la ETSINF en su enfoque sobre MDM SaaS y los servicios en la nube de Informatica, y espera sentar un precedente para futuras investigaciones en esta área emergente y crítica.

2.9 Propuesta

El presente Trabajo de Fin de Grado se propone llenar un vacío en la literatura académica y técnica al explorar la interacción entre *Master Data Management* y *Software as a Service* en un entorno empresarial. Si bien existen estudios anteriores que se centran en aspectos individuales de MDM o SaaS, la sinergia entre estas dos áreas tecnológicas es un campo que aún requiere investigación exhaustiva.

Este trabajo se distingue por su enfoque integral que abarca tanto la teoría como la aplicación práctica. Además de un estudio teórico exhaustivo sobre MDM

y SaaS, se realizarán pruebas de concepto para investigar la aplicabilidad y escalabilidad de soluciones específicas en la nube, particularmente las ofrecidas por Informatica Cloud. Esta dualidad de enfoque permite un examen completo de cómo las políticas y estrategias de gestión de datos pueden integrarse eficazmente con soluciones tecnológicas en un entorno en la nube.

La originalidad del trabajo radica en su intento de ofrecer una perspectiva holística. No sólo se abordan los aspectos teóricos de MDM y la estrategia detrás de SaaS, sino que también se examinan sus aplicaciones prácticas en casos de uso empresariales específicos. El objetivo es proporcionar a las organizaciones una guía práctica para alinear sus políticas de gobernanza de datos con las soluciones tecnológicas disponibles, especialmente en un entorno en la nube.

En resumen, el presente Trabajo de Fin de Grado busca explorar un ámbito que ha recibido atención limitada tanto en la academia como en la industria. Mediante la combinación de una revisión teórica y un enfoque práctico, se aspira a contribuir de manera significativa al cuerpo de conocimiento en la gestión de datos maestros y las soluciones en la nube.

CAPÍTULO 3

Análisis del problema

El Trabajo de Fin de Grado abarca una parte teórica, en la cual se ha llevado a cabo un exhaustivo estudio sobre las diversas áreas de la Informática Cloud y la Gestión de Datos Maestros. Para la parte práctica de este TFG, se ha optado por realizar un proyecto multinacional en un entorno de Informática Cloud. Este enfoque permite aplicar los conceptos teóricos en un contexto real y solventar problemas que se podrían encontrar en un escenario profesional. Al utilizar un entorno de Informática Cloud, se busca aprovechar la escalabilidad, flexibilidad y capacidad de integración que estas tecnologías ofrecen, lo cual es especialmente relevante en el contexto multinacional del proyecto. Este enfoque también facilita la colaboración con el cliente y otros stakeholders, permitiendo una implementación más ágil y eficiente de las soluciones.

Esta práctica aplicada no solo aporta un valor inmediato en términos de solución de problemas reales, sino que también ofrece la oportunidad de validar teorías y métodos estudiados en la fase teórica del TFG. De este modo, se busca que el proyecto sirva como un puente eficaz entre la teoría y la práctica, proporcionando perspectivas útiles tanto para futuras investigaciones como para aplicaciones en el mundo real.

3.1 Especificación de requisitos

Se ha establecido una colaboración con el cliente para definir los requisitos del sistema. A fin de que estos requisitos sean accesibles a un público amplio, incluso a aquellos sin un conocimiento especializado en MDM o en Informática Cloud, se ha redactado la especificación de requisitos de una forma más intuitiva y sin emplear lenguaje técnico.

A continuación, se detallan los requisitos especificados:

1. **Actualización Automática de Campos:** El primer requisito es el desarrollo de un sistema que permita la actualización automática de cada uno de los campos en las tablas de datos finales. El sistema debe ser capaz de comparar el mismo atributo proveniente de diversas fuentes y seleccionar el más actualizado en función del tiempo. Así, si un atributo *A* de una fuente *X* ha sido actualizado más recientemente que un atributo *B* de una fuente *Y*, el sistema elegirá el atributo *A*, sin tener en cuenta su contenido.

2. **Verificación Manual de Actualizaciones:** El segundo requisito es la incorporación de un sistema de verificación manual que permita a los administradores de la empresa cliente revisar las actualizaciones realizadas en los datos finales. Este sistema no será obligatorio para todas las actualizaciones, sino solo para aquellas que los administradores consideren necesarias. A través de este sistema, los administradores podrán decidir si los cambios efectuados deben persistir en los datos finales o si deben ser revertidos.

Estos requisitos serán la base para los Pruebas de Concepto que se desarrollarán en los capítulos siguientes del TFG y determinarán las soluciones a implementar.

Con estos requisitos definidos, se procederá a realizar un análisis de riesgos, una revisión de la protección de datos, y finalmente se expondrá la solución propuesta y el plan de trabajo. Cada uno de estos aspectos se tratará en detalle en las siguientes secciones de este capítulo.

3.2 Análisis de riesgos

El objetivo de este apartado es evaluar los riesgos asociados con los requisitos especificados para el proyecto. Se analizará tanto el impacto positivo como el negativo que cada requisito podría tener en el desarrollo y operación del sistema. Además, se discutirán los criterios de aceptación del cliente, que sirven como indicadores de éxito o fracaso en la implementación de cada requisito. Este análisis es crucial para la planificación y toma de decisiones efectiva en las fases posteriores del proyecto.

3.2.1. Impacto de los Requisitos

Actualización Automática de Campos

- **Impacto Positivo:** Este sistema podría mejorar significativamente la eficiencia y la fiabilidad de los datos en el entorno de la empresa. Una actualización automática permitiría una mejor toma de decisiones basada en datos más recientes y precisos.
- **Impacto Negativo:** Existe el riesgo de que la automatización introduzca errores o inconsistencias si no se implementa correctamente. Además, si la lógica de selección de atributos no es robusta, podrían priorizarse datos erróneos o irrelevantes.

Verificación Manual de Actualizaciones

- **Impacto Positivo:** Este sistema proporciona un control de calidad adicional, lo que podría resultar en una mayor precisión y confiabilidad de los datos finales.

- **Impacto Negativo:** La verificación manual podría ralentizar el flujo de trabajo y requerir más recursos humanos para su gestión, lo que aumenta los costos operativos. Además, si el sistema no se usa adecuadamente, podrían ocurrir errores humanos que deriven en la aprobación de datos erróneos.

3.2.2. Criterios de aceptación del cliente

Los criterios de aceptación para ambos requisitos son fundamentalmente binarios y se basan en la funcionalidad:

- **Actualización Automática de Campos:** El criterio principal será si el sistema es capaz de actualizar automáticamente los campos y si selecciona el atributo más reciente entre múltiples fuentes de manera efectiva y confiable.
- **Verificación Manual de Actualizaciones:** El éxito de este requisito se medirá en función de si los administradores son capaces de revisar y aprobar o revertir las actualizaciones de manera eficiente.

Estos criterios de aceptación son directos y se alinean con las expectativas del cliente de tener sistemas funcionales que aborden sus necesidades específicas.

3.3 Análisis legal

El análisis legal es una parte fundamental en cualquier proyecto de tecnología de la información, especialmente cuando se trata de la gestión de datos a gran escala en entornos multinacionales. Este proyecto no es una excepción y, por lo tanto, se deben considerar diversas leyes y regulaciones que podrían tener un impacto en los requisitos establecidos para el desarrollo del sistema. A continuación, se describen algunas de las consideraciones más relevantes:

3.3.1. Regulación General de Protección de Datos (GDPR)

Si el proyecto se lleva a cabo en un entorno que implica el procesamiento de datos personales de ciudadanos de la Unión Europea, el cumplimiento del Reglamento General de Protección de Datos¹ (GDPR) es obligatorio. Los requisitos establecidos en este proyecto, como la actualización automática de campos y la verificación manual de actualizaciones, deben diseñarse teniendo en cuenta la privacidad desde su concepción. Es esencial garantizar que los datos se manejen de forma segura y que se obtenga el consentimiento adecuado de los individuos antes de cualquier procesamiento de datos.

¹<https://gdpr-info.eu/>

3.3.2. Sarbanes-Oxley Act (SOX)

Si el proyecto tiene implicaciones en los Estados Unidos, especialmente para empresas cotizadas en bolsa, la Ley Sarbanes-Oxley² (SOX) puede tener un impacto significativo en los requisitos. Este reglamento se centra en la integridad de los informes financieros y requiere controles internos efectivos, que también podrían extenderse a los sistemas de TI que manejan datos financieros. El sistema de verificación manual, por ejemplo, deberá estar diseñado para cumplir con estos controles internos.

3.3.3. Impacto en los Requisitos

Ambas leyes, GDPR y SOX, tienen un enfoque significativo en la trazabilidad y la responsabilidad. Esto podría requerir ajustes en los requisitos del proyecto para incluir capacidades de auditoría más robustas, lo que permitiría rastrear todas las actualizaciones y verificaciones de datos. Los requisitos también podrían necesitar ser ajustados para incluir procesos de consentimiento y revocación del consentimiento, lo que podría afectar tanto al diseño como a la implementación del sistema.

En resumen, las leyes y regulaciones como GDPR en Europa y Sarbanes-Oxley en los Estados Unidos pueden tener un impacto sustancial en cómo se diseñan e implementan los sistemas de gestión de datos. Por lo tanto, es crucial que este proyecto se lleve a cabo con una comprensión completa de estas leyes y regulaciones para garantizar el pleno cumplimiento y evitar cualquier riesgo legal o financiero.

3.4 Solución propuesta

Con base en los requisitos específicos proporcionados por el cliente, se propone llevar a cabo una serie de pruebas de concepto (*Proof of Concept*, PoC) para validar la viabilidad y eficacia de las soluciones que se desarrollarán en este Trabajo de Fin de Grado. Estos PoC servirán como una fase preliminar crucial que nos permitirá evaluar los métodos, tecnologías y enfoques que serán más efectivos para satisfacer los requisitos del cliente en un entorno real.

Los PoC se diseñarán para abordar específicamente los dos requisitos principales:

- Actualización Automática de Campos: Se creará un PoC para probar diferentes métodos y algoritmos que permitan la actualización automática de los campos de las tablas de datos finales, asegurando que el atributo más reciente se seleccione efectivamente entre múltiples fuentes.
- Verificación Manual de Actualizaciones: Un segundo PoC se enfocará en implementar y probar un sistema que permita a los administradores de la

²<https://sarbanes-oxley-act.com/>

empresa cliente revisar manualmente las actualizaciones realizadas, para decidir si deben persistir o ser revertidas.

Es importante destacar que los PoC no representan la solución final pero sí una etapa esencial para asegurar que los métodos y tecnologías elegidas sean los adecuados para el proyecto. Posteriormente, las soluciones validadas en la fase de PoC serán desarrolladas e implementadas en un entorno de Informática Cloud, aprovechando sus capacidades para la integración de datos, la calidad de datos y otras funcionalidades de gestión de datos maestros.

En resumen, la solución propuesta busca no solo cumplir con los requisitos del cliente sino también asegurar que las tecnologías y métodos empleados sean robustos, escalables y en consonancia con las mejores prácticas del sector.

3.5 Plan de trabajo

En el contexto de la parte práctica del Trabajo de Fin de Grado, el desarrollo se ha ejecutado de forma autónoma, subrayando la importancia de una planificación cuidadosa y una administración efectiva del tiempo y los recursos. Aunque el trabajo ha sido principalmente autónomo, se ha mantenido una comunicación continua con compañeros de departamento, compañeros del proyecto y clientes para asegurar una ejecución cohesiva y alineada con los objetivos generales.

El plan de trabajo ha estado estructurado en torno a objetivos específicos y entregables, cada uno de los cuales ha estado asociado a distintas fases del proyecto. Los entregables, bien definidos pero adaptables, han servido como puntos de referencia para evaluar el progreso del proyecto. Estos han incluido desde elementos tangibles como *Proof of Concept* hasta revisiones de desarrollo, implementación y documentación que respalden los desarrollos realizados.

Las reuniones periódicas con compañeros de departamento, compañeros del proyecto y clientes han sido una parte integral para el éxito de esta fase práctica. Estas reuniones han ofrecido una plataforma para discutir los avances recientes, abordar problemas que puedan haber surgido y revisar cualquier riesgo percibido en la implementación de los requisitos. Estas discusiones han sido vitales para realizar ajustes en el plan de trabajo, replantear estrategias y tomar decisiones fundamentadas para el avance del proyecto.

Fuera de estas reuniones estructuradas, se ha mantenido un flujo constante de comunicación a través de herramientas de gestión de proyectos, chats empresariales y correo electrónico. Este canal abierto ha permitido una coordinación más ágil, permitiendo abordar cambios y desafíos de forma proactiva y en tiempo real.

En resumen, el plan de trabajo para la parte práctica del TFG ha sido un ejercicio de autogestión, complementado con colaboración e interacción estratégicas con compañeros para asegurar que los objetivos del proyecto se cumplan de manera efectiva y eficiente.

CAPÍTULO 4

Diseño de la solución

En este capítulo, abordamos el diseño de la solución para cumplir con los requisitos especificados en el análisis del problema. Una parte esencial de cualquier diseño es la tecnología que se utilizará para implementarlo, especialmente en un campo tan especializado como la Informática Cloud y la Gestión de Datos Maestros. Es por ello que dedicaremos una sección a explicar la tecnología y la arquitectura cloud utilizadas en este proyecto, con un enfoque particular en Informática Cloud, las instancias de *sandbox* de la empresa en esta plataforma, más específicamente en las soluciones Customer360 y Business 360.

Además, es crucial entender ciertos términos y conceptos que son específicos de Informática MDM para seguir el hilo del diseño detallado que se presentará. Aunque se ha llevado a cabo un estudio exhaustivo en el marco teórico del Trabajo de Fin de Grado, hay definiciones que no se han incluido en dicha sección debido a su especificidad con respecto a la tecnología que estamos utilizando.

En la última sección de este capítulo, nos adentraremos en el diseño detallado de la solución. Este segmento es de vital importancia ya que desglosará los elementos técnicos y operacionales necesarios para llevar a cabo los requisitos del cliente. Aquí, detallaremos cómo cada componente interactuará, cuáles son los flujos de datos y cómo se gestionarán los procesos, desde la actualización automática de campos hasta la verificación manual de las actualizaciones. En resumen, esta sección será la columna vertebral técnica sobre la cual se desarrollarán e implementarán los *Proof of Concept* en Informática Cloud.

Por lo tanto, este capítulo se estructurará en tres partes fundamentales: la tecnología y arquitectura cloud utilizadas, las definiciones específicas que son cruciales para entender este proyecto y, finalmente, el diseño detallado que servirá de hoja de ruta para la implementación de la solución. Con esto, buscamos no solo ofrecer una solución viable y eficiente sino también proporcionar una comprensión completa de las consideraciones técnicas involucradas.

4.0.1. Arquitectura cloud y tecnología utilizada

Informatica Cloud

Aunque Informática Cloud ha sido un tema discutido en la sección teórica de este Trabajo de Fin de Grado, es vital reintroducirlo brevemente en este capítulo

para asegurar la completa comprensión de las implicaciones técnicas y arquitectónicas que se abordarán. Informatica Cloud es una plataforma robusta que ofrece soluciones de integración de datos, calidad de datos y gestión de datos maestros en un entorno cloud. Para los fines de este proyecto, se ha optado por utilizar Customer 360 como la solución de MDM debido a su capacidad para proporcionar una visión 360 grados de los datos del cliente, facilitando así la toma de decisiones informadas.

Customer360

Customer 360 (C360) es mucho más que un simple sistema de gestión de datos del cliente; es una solución integral que actúa como el núcleo del ecosistema de Informatica Cloud. Abarca un amplio rango de funcionalidades y se integra perfectamente con otras soluciones especializadas como Business 360, Product 360 (PIM), Supplier 360, Finance 360, Reference 360, Data as a Service, y *Data Governance*. Estas soluciones son componentes críticos que trabajan en conjunto para proporcionar un entorno cohesivo y potente para la gestión de datos maestros. Al hacerlo, ayudan a las empresas a mantener la integridad, la coherencia y la calidad de sus datos.

Business360

Business 360 (B360) no es solo una extensión de C360, sino que es una solución por sí misma, diseñada para abordar las complejidades inherentes a los datos empresariales. Actúa como el nexo entre los datos brutos y la información útil y accionable que se presenta en Customer 360. B360 es crucial para establecer y administrar las políticas, reglas, y lógicas de negocio que estructuran los datos. Además, es el catalizador que convierte los datos brutos en información procesable, ofreciendo una única fuente de verdad que puede ser utilizada para la toma de decisiones estratégicas.

Sandbox empresarial

El entorno de *sandbox* ofrecido por Informatica Cloud es un recurso invaluable para el desarrollo y las pruebas. Este entorno viene cargado con todas las soluciones que forman parte del ecosistema de Informatica Cloud, incluidos C360 y B360. Este *sandbox* ha sido fundamental para llevar a cabo tanto la investigación teórica del TFG como las pruebas de concepto que validan las funcionalidades y requisitos especificados. Es esencial resaltar que este entorno de pruebas ha permitido una simulación cercana a un entorno de producción, lo que a su vez ha garantizado que los resultados sean lo más fiables y aplicables posible.

4.0.2. Definiciones técnicas

Para garantizar un entendimiento claro y completo del diseño de la solución, es crucial definir algunos términos técnicos que son específicos de Informatica

MDM y que no se han abordado en el marco teórico del TFG. A continuación, se presentan estas definiciones en orden alfabético:

- **Base Object:** Objeto base en el que se almacenan los registros maestros después del proceso de concordancia y fusión.
- **Business Entities:** Entidades de negocio que representan los datos maestros en un formato que puede ser consumido y entendido por aplicaciones de negocio.
- **Business Events:** Eventos que se activan en función de ciertas condiciones o cambios en los datos.
- **CAI Process:** Proceso de integración de aplicaciones compuestas, que permite la orquestación de tareas.
- **External Validation:** Proceso de validación que se lleva a cabo fuera del sistema MDM, generalmente a través de servicios web o aplicaciones externas.
- **Last Update Date (LUD):** Campo que registra la última vez que se actualizó un registro particular. Por defecto se incluye uno por cada RowObject, indicando cuándo se ha actualizado.
- **Match:** Proceso de identificación de registros que representan la misma entidad real.
- **Merge:** Proceso de fusión de varios registros coincidentes en un solo registro maestro.
- **RowIdObject:** Identificador único para un objeto base dentro del sistema MDM.
- **RowObject:** Referencia al objeto real dentro del sistema MDM, usualmente vinculado a través del 'RowIdObject'.
- **Survivorship:** Reglas y lógica aplicadas para determinar qué datos sobreviven durante el proceso de fusión de registros.
- **XREF:** Referencia cruzada que mantiene la relación entre los registros de origen y los registros maestros.

4.0.3. Diseño detallado de la solución

La fase de diseño detallado es crucial para asegurar que los requisitos establecidos se materialicen en una solución efectiva y fiable. En este apartado, se abordarán los enfoques específicos que se han elegido para los Proofs of Concept en relación con los requisitos del cliente.

Actualización Automática de Campos

Para la actualización automática de campos, el enfoque ha sido emplear los campos Last Update Date para determinar cuál de los atributos ha sido modificado más recientemente. Este LUD se utiliza en los procesos de survivorship y merge para los registros que coinciden (match). Sin embargo, un desafío que se ha encontrado es que el LUD está asociado al RowObject completo, y no a los atributos individuales. Esto hace difícil determinar cuándo un atributo específico ha sido modificado. Para solucionar este problema, se ha añadido un atributo adicional junto a cada atributo original que registra el LUD específico de ese atributo. Luego, mediante una *External Validation* a un *CAI Process*, se revisa cuál de los atributos de ese campo es el más reciente, y ese valor se devuelve al Base Object.

Verificación Manual de Actualizaciones

Para el requisito de verificación manual de actualizaciones, se ha elegido incorporar un Business Event que permitirá verificar la validez de la actualización que se reciba. Este evento se activará únicamente cuando el administrador marque un campo especial que será invisible para los clientes. Este campo sirve como la condición de activación del Business Event, permitiendo al administrador decidir si una actualización particular necesita ser revisada manualmente o no. Una vez marcado, el evento se activa y da lugar a una verificación manual donde el administrador puede aprobar o revertir la actualización, asegurando que se refleje en el Base Object según corresponda.

Este diseño detallado proporciona una estructura sólida sobre la que construir los PoC, garantizando que los requisitos del cliente sean satisfechos de una manera efectiva y eficiente.

CAPÍTULO 5

Desarrollo de la solución propuesta

5.1 Introducción al entorno

En este capítulo, se proporcionará una descripción detallada del entorno de desarrollo utilizado para la realización de este proyecto. El entorno se basa principalmente en la plataforma de Informatica Cloud, con un enfoque en las soluciones Business 360 y Customer 360. A continuación, se exploran las diferentes partes de este entorno.

5.1.1. Vista Inicial de la *sandbox*

Al acceder a la plataforma de Informatica Cloud mediante el inicio de sesión único (Single Sign-On, SSO) de la empresa, nos encontramos con una interfaz de usuario desde donde podemos elegir a cuál de las diversas soluciones de Informatica Cloud queremos acceder. Para los propósitos de este proyecto, nos centraremos exclusivamente en B360 y C360. (Ver Figura: 5.1)

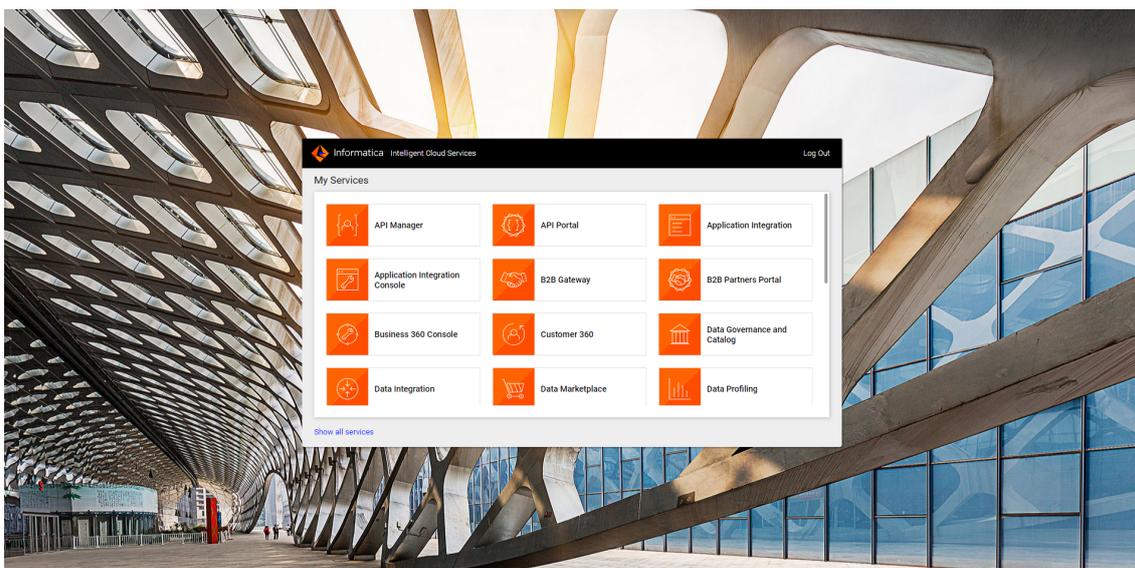


Figura 5.1: Vista inicial de la *sandbox*

5.1.2. Customer 360

Una vez seleccionada la opción de Customer 360, la plataforma ofrece un menú conceptual con diversas funcionalidades. Entre ellas, se incluyen opciones para la creación de nuevos objetos, búsqueda dentro de las Business Entities, gestión de flujos de trabajo, generación de informes, y manipulación de jerarquías, entre otros. Estas funcionalidades nos permitirán desarrollar y gestionar los datos y procesos del proyecto. (Ver Figura: 5.2)

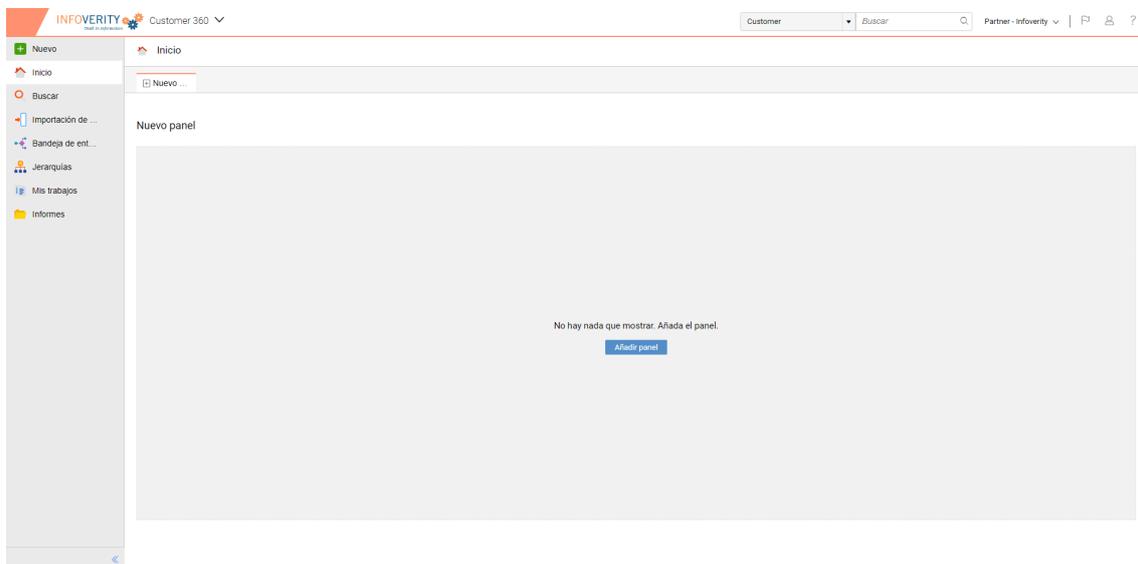


Figura 5.2: Vista inicial de Customer 360

5.1.3. Consola de Business 360

La Consola de Business 360 actúa como el centro neurálgico para la administración y la creación de todos los componentes que serán visualizados en Customer 360. Desde aquí, es posible llevar a cabo la creación de Business Entities, eventos, páginas, trabajos, entre otros elementos. Esta consola será fundamental para el desarrollo y la administración del proyecto. (Ver Figura: 5.3)

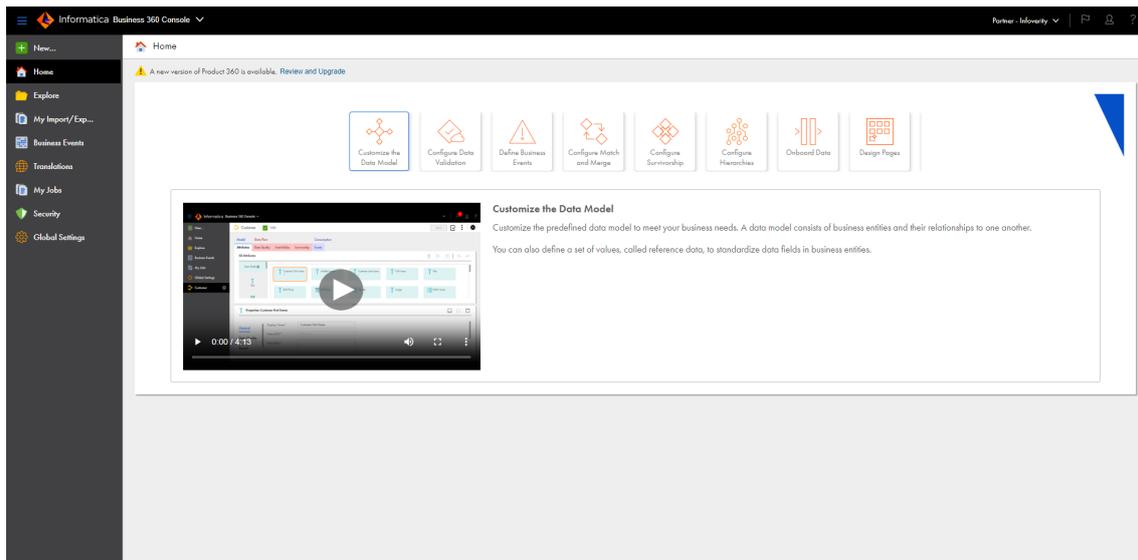


Figura 5.3: Vista inicial de la Consola de Business 360

5.1.4. Application Integration

Aunque la plataforma de Application Integration ofrece una gran variedad de funcionalidades, para los fines de este proyecto, nuestra atención se centrará en la capacidad de crear CAI (Cloud Application Integration) Processes. Esta funcionalidad es esencial para la integración y la automatización de diversos procesos en el proyecto. (Ver Figura: 5.4)

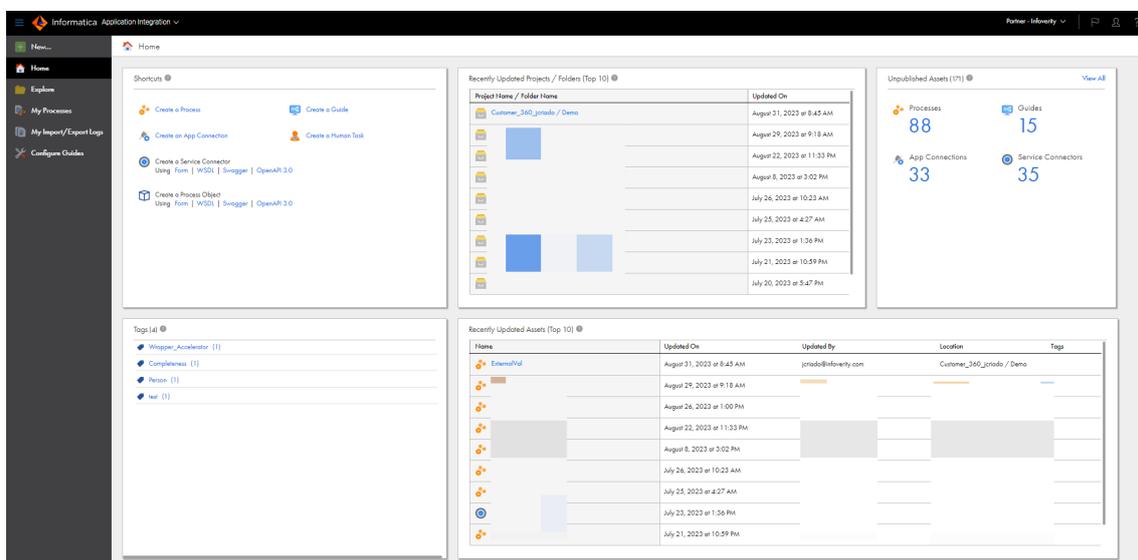


Figura 5.4: Vista inicial de Application Integration

5.1.5. Repositorio Personal

Dentro de la consola de Business 360, se ofrece la posibilidad de crear un repositorio personal o específico para cada proyecto. Este repositorio facilita el acceso rápido a los componentes clave del proyecto, como las Business Entities, páginas,

CAI Processes y otros componentes esenciales. Este espacio será utilizado para organizar y almacenar todos los elementos desarrollados durante el proyecto. (Ver Figura: 5.5)

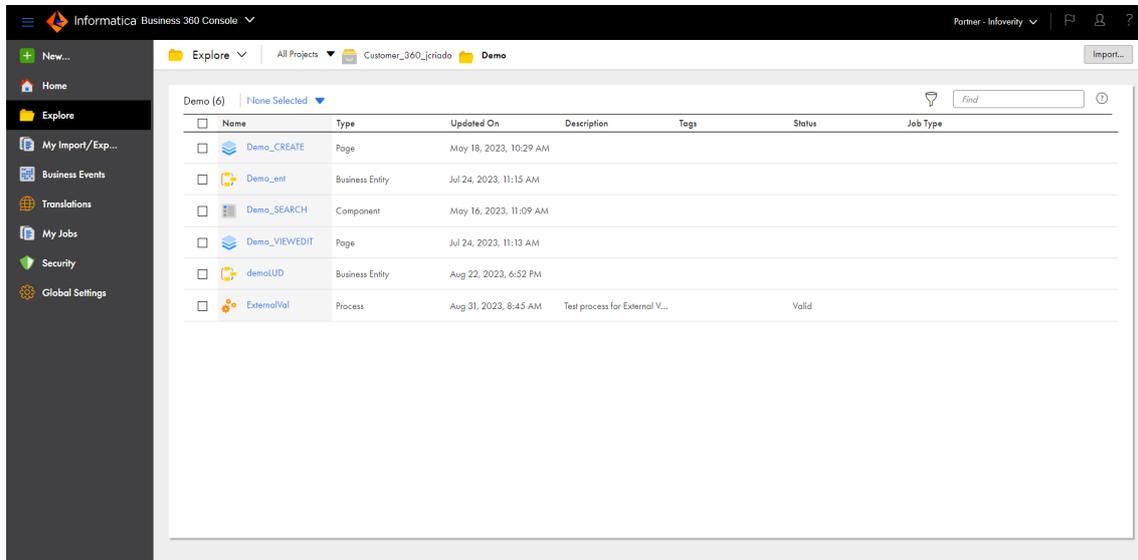


Figura 5.5: Vista del repositorio personal

Con esta introducción, se espera que el lector tenga una comprensión clara del entorno en el cual se ha desarrollado la solución propuesta para este proyecto. En las siguientes secciones, se detallará el desarrollo específico de cada uno de los requisitos.

5.2 Desarrollo de Actualización Automática de Campos

En esta sección, se detallará el desarrollo de la funcionalidad para la Actualización Automática de Campos, uno de los requisitos clave del proyecto. El objetivo es mantener actualizados los campos de datos en la *Business Entity* (BE) correspondiente, tomando en cuenta las últimas actualizaciones de datos (Last Update Date, LUD) para cada atributo.

5.2.1 Creación de una *Business Entity*

El primer paso consiste en la creación de una *Business Entity* a través de la consola de Business 360. A esta BE se le añaden varios campos para almacenar los atributos de interés, como Name, Surname, Phone, etc. Posteriormente, se agregan campos adicionales para almacenar la información del Last Update Date para cada uno de estos atributos. De este modo, cada atributo tendrá su correspondiente campo LUD, formando parejas como Name y NameLUD, Surname y SurnameLUD, Phone y PhoneLUD, etc. (Ver Figura: 5.6)

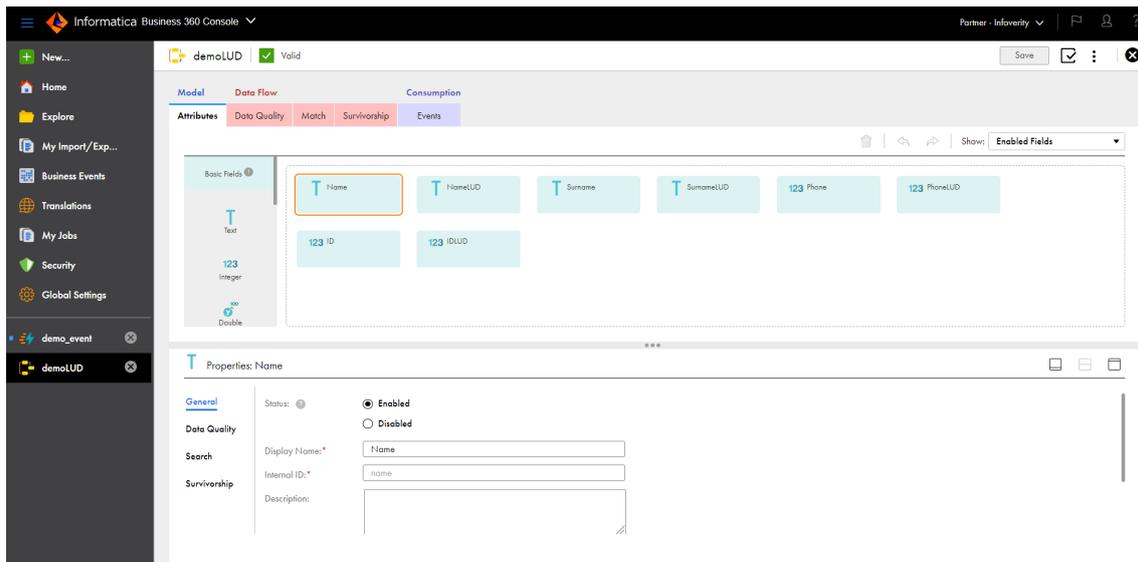


Figura 5.6: Vista de la *Business Entity* creada

5.2.2. Creación del *CAI Process*

El siguiente paso es la creación de un *CAI Process* en Application Integration. Inicialmente, se definen los campos de entrada (Input Fields) como param1 y param2, y un campo de salida (Output Field) que será una casilla de verificación (checkbox) denominada isValid, inicializada a True.

La lógica del proceso incluye una decisión que compara los valores LUD de los atributos para determinar cuál es el más reciente. El registro con el LUD más reciente tendrá isValid marcado como True, mientras que los registros más antiguos tendrán isValid marcado como False. (Ver Figura: 5.7)

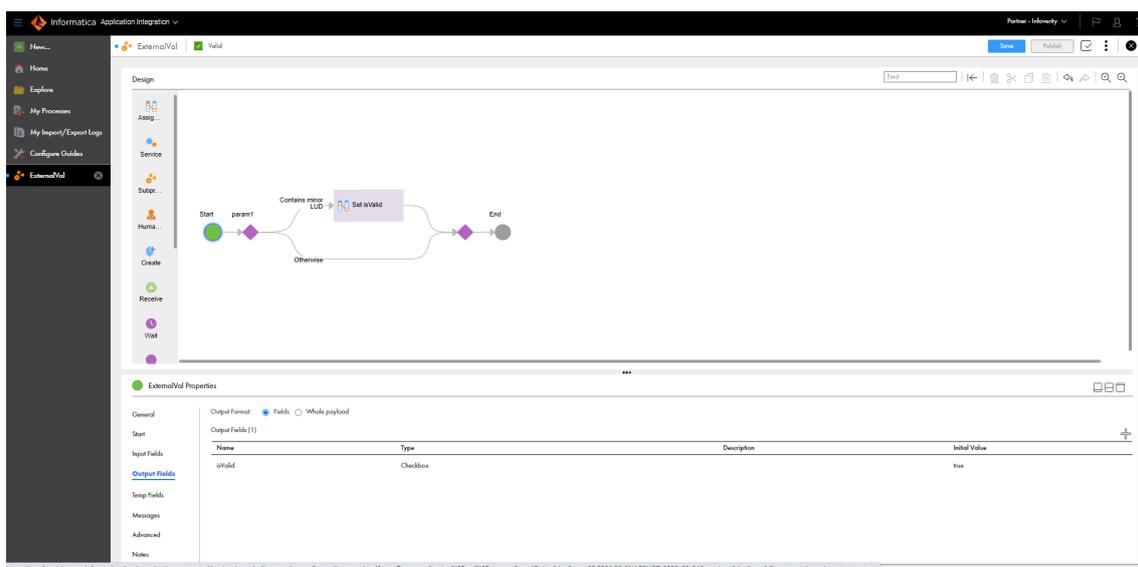


Figura 5.7: Vista del *CAI Process* creado

5.2.3. Publicación del Proceso

Una vez creado y probado el *CAI Process*, se procede a su publicación. Para ello, se pulsa el botón de Publish y se copia la Service URL generada, que será utilizada en el siguiente paso para la configuración de la validación externa. (Ver Figura: 5.8)

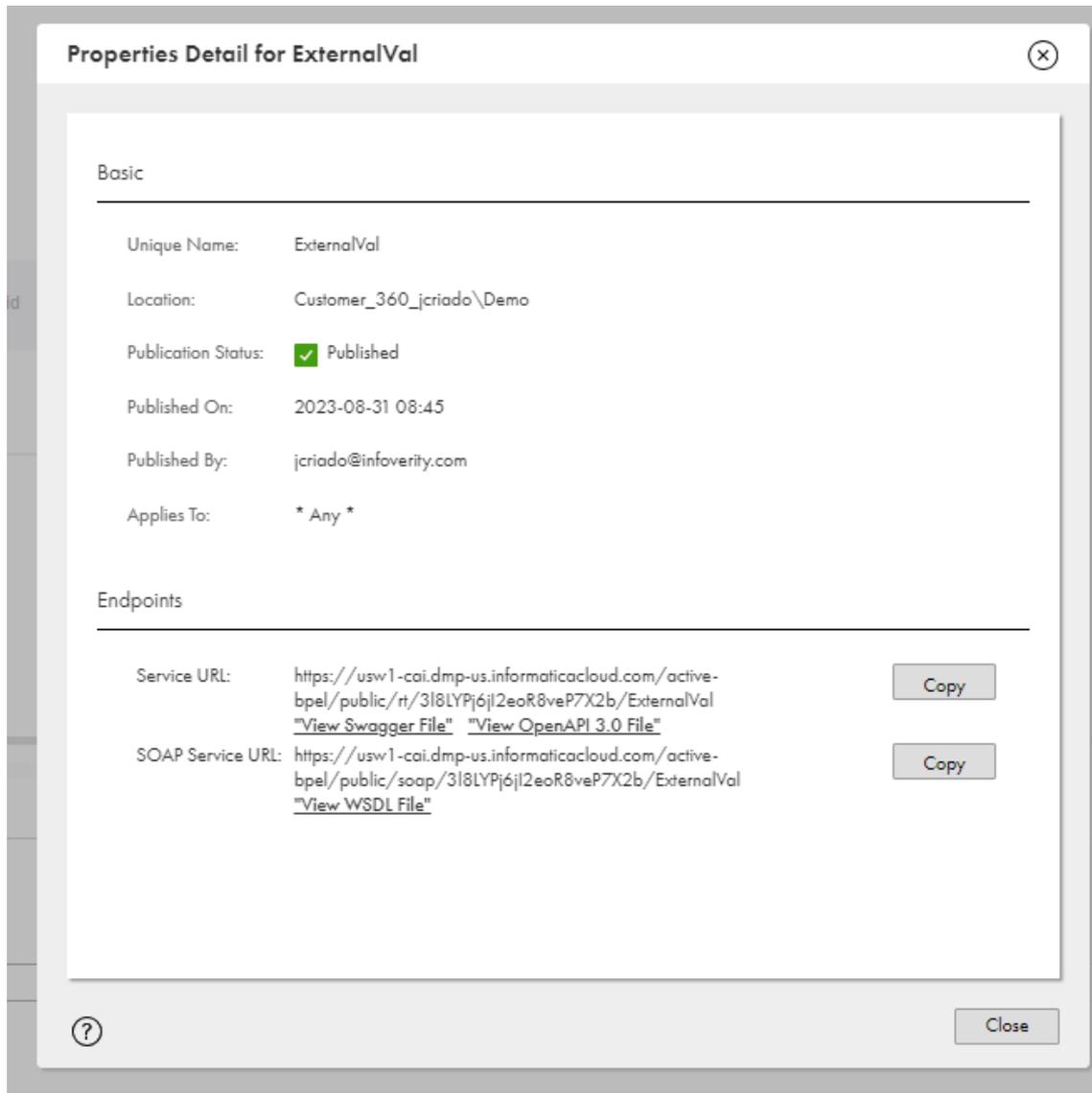


Figura 5.8: Vista del CAI Process publicado

5.2.4. Creación de la External Validation

El último paso consiste en la configuración de la validación externa. Para ello, en la consola de B360, se seleccionan los campos originales y se accede a la opción Data Quality. Ahí se crea una nueva regla de asociación básica seleccionando External Validation como función. Se pega el Process URL copiado en el paso anterior y se configura el mensaje de error para el caso en el que isValid sea False. (Ver Figura: 5.9)

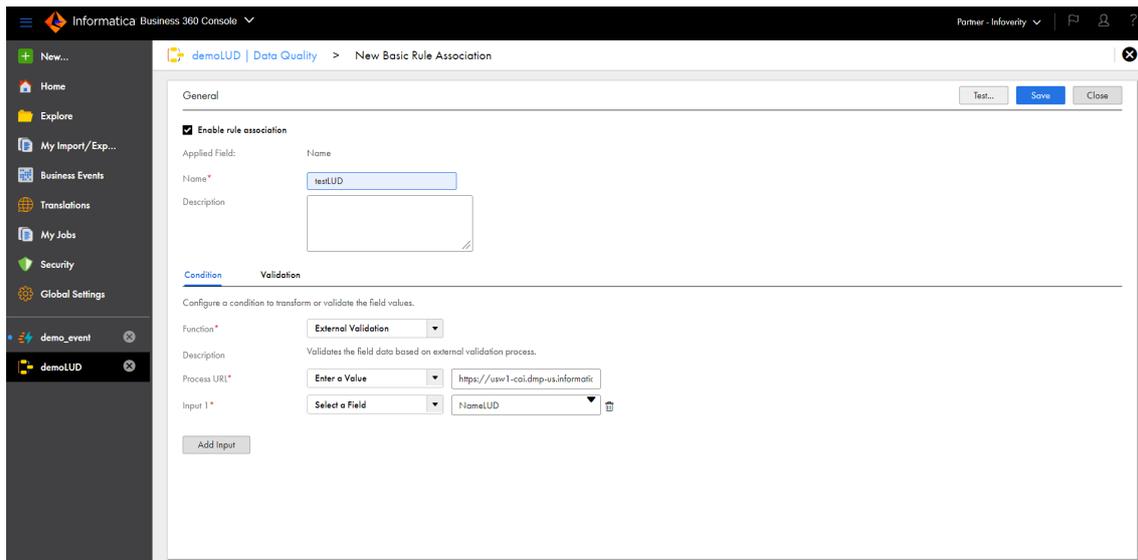


Figura 5.9: Vista de la *External Validation* creada

Con la implementación de estos pasos, se logra la actualización automática de los atributos en la *Business Entity*. Los atributos que pasen la validación externa y devuelvan `isValid=True` serán actualizados, mientras que los que devuelvan `isValid=False` no serán tomados en cuenta. Este desarrollo garantiza que sólo los datos más recientes sean considerados, cumpliendo así con los requisitos del proyecto.

5.3 Desarrollo de Verificación Manual de Actualizaciones

Esta sección describe el procedimiento detallado para desarrollar la funcionalidad de Verificación Manual de Actualizaciones, que forma parte esencial de los requisitos del proyecto.

5.3.1. Creación de la *Business Entity*

El primer paso en el desarrollo es la creación de una *Business Entity* que contendrá todos los campos necesarios. Además de los campos visibles para los usuarios, se añade un campo especial que sólo será accesible para los administradores. Este campo, implementado como un *Field Group*, contendrá un subcampo obligatorio denominado `On`, que servirá como criterio de unicidad para determinar si una actualización requiere verificación manual o no. (Ver Figura 5.10)

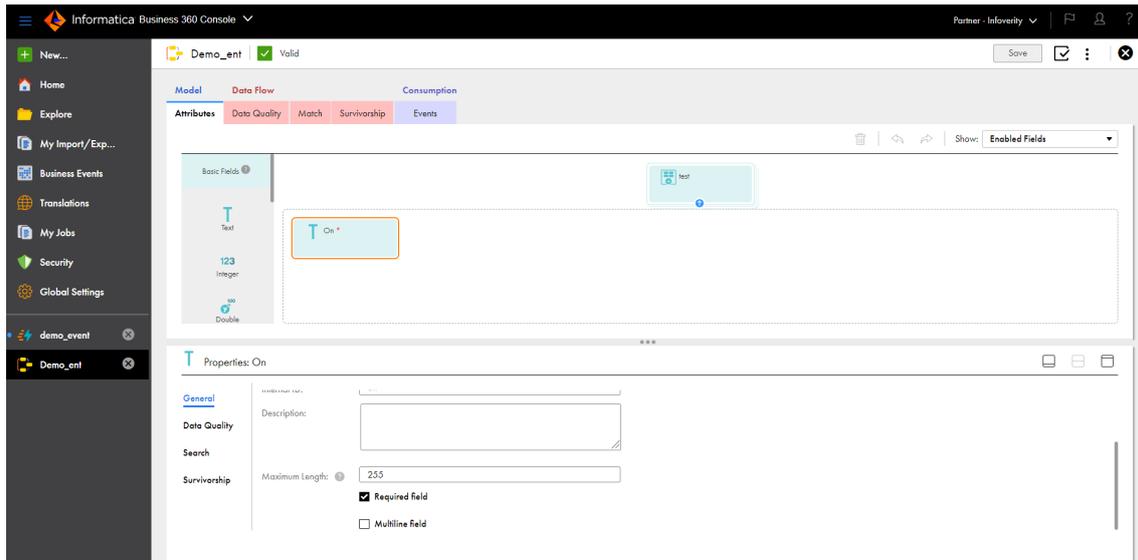


Figura 5.10: Vista de la BE con el *Field Group*

5.3.2. Creación de la Componente de Búsqueda

Se desarrolla una componente de búsqueda que se utilizará para previsualizar los resultados después de realizar una búsqueda en la BE, pero antes de acceder a un objeto específico. (Ver Figura: 5.11)

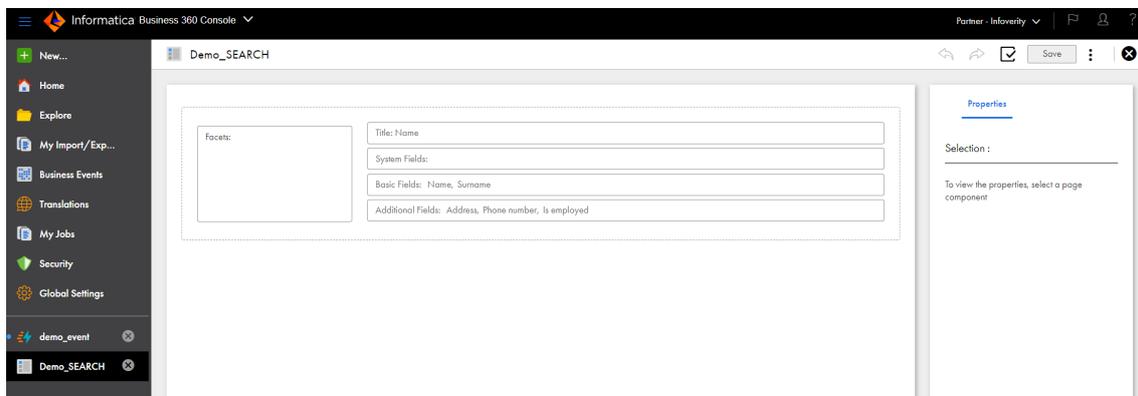


Figura 5.11: Vista de la componente de búsqueda creada

5.3.3. Creación de la Página de Creación de Objetos

Esta página será la interfaz que aparecerá cuando se quiera crear un nuevo objeto dentro de la BE. Se incluyen en la página todos los campos relevantes, excepto el *Field Group* que contiene el campo *On*.

5.3.4. Creación de la Página de Vista y Edición de Objetos

Similar a la página de creación, esta página será la interfaz para la visualización y edición de un objeto existente dentro de la BE. De nuevo, se omitirá el *Field Group* que contiene el campo *On*. (Ver Figura: 5.12)

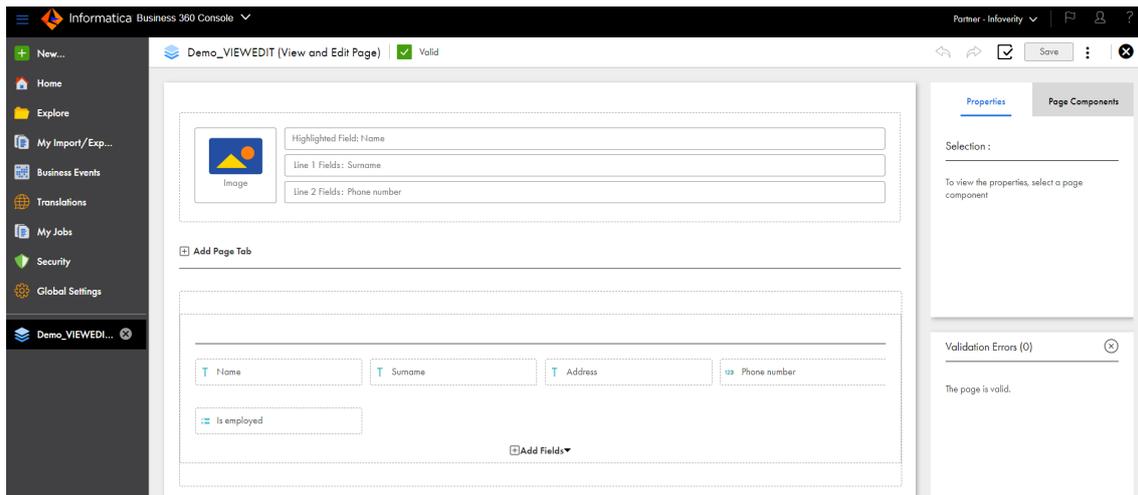


Figura 5.12: Vista de la página de vista y edición

5.3.5. Creación del Business Event

El siguiente paso es la creación del Business Event que será responsable de generar la tarea de verificación manual. Se elige la opción *User Triggered event*, se rellenan los campos correspondientes, se especifican los roles que estarán involucrados y se añade el escenario de acción, seleccionando la BE y en *Actions* se elige *Manual Merge*. (Ver Figura: 5.13)

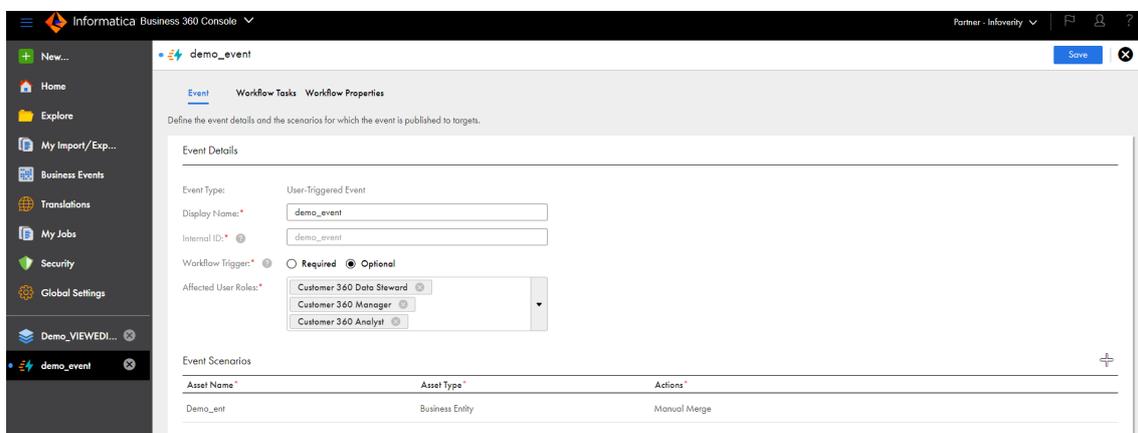


Figura 5.13: Vista del evento creado

5.3.6. Creación del Workflow

Finalmente, todavía dentro de la creación del evento, se configura el *Workflow*. En este ejemplo, se opta por *MDM One Step Approval*, ya que sólo se requiere una comprobación. Se seleccionan los roles que estarán a cargo de la revisión y se definen las propiedades del *Workflow*, como el nombre, prioridad y el período durante el cual las actualizaciones pueden ser validadas. (Ver Figura: 5.14)

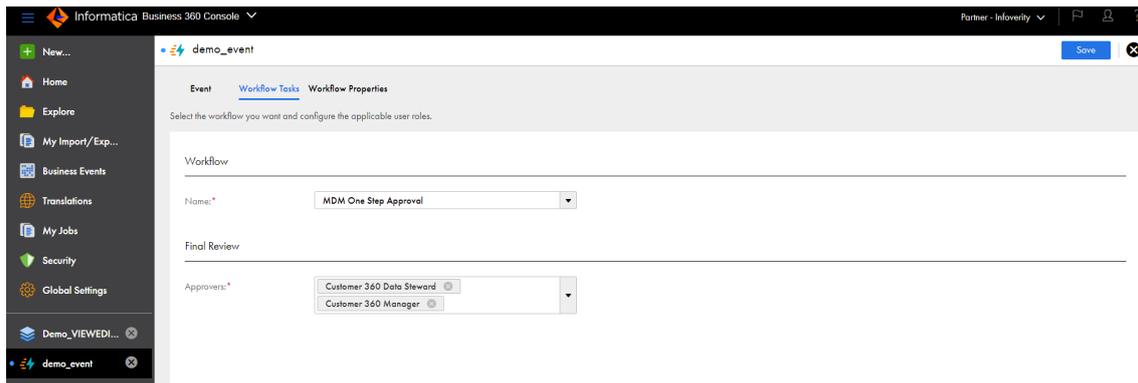


Figura 5.14: Vista del *Workflow*

Con la implementación de estos elementos, el requisito de Verificación Manual de Actualizaciones queda plenamente funcional. Cada vez que se realice una actualización, se generará una tarea de revisión manual que deberá ser completada por un usuario con uno de los roles asignados para este fin.

5.4 Problemas y Dificultades

El desarrollo del proyecto no estuvo exento de desafíos y obstáculos. A continuación, se describen algunas de las principales dificultades que surgieron a lo largo del proyecto en forma de una lista de ítems:

- **Inversión de Tiempo en Investigación:** Uno de los principales problemas enfrentados fue la inversión de tiempo adicional necesaria para investigar funcionalidades que no vienen implementadas *out of the box*. Es crucial recordar que el Trabajo de Fin de Grado (TFG) se lleva a cabo en un contexto de proyecto empresarial, donde el tiempo es un recurso valioso.
- **Caída del Servicio de Informatica Cloud:** Otra considerable complicación fue la caída del servicio de Informatica Cloud, que se extendió durante varias semanas. Este incidente retrasó conmesurablemente el desarrollo de las funcionalidades previstas y, por extensión, el progreso del TFG.
- **Bug Interno en la *sandbox*:** Para añadir más desafíos al proyecto, poco después de que el servicio de Informatica Cloud se restaurara, se encontró un bug interno que afectaba el funcionamiento de las Business Entities en el Customer 360. Este bug impactó significativamente el desarrollo del requisito de verificación manual.

A lo largo del proyecto, se enfrentaron diversas dificultades que afectaron tanto los tiempos de entrega como la calidad del trabajo final. Estos problemas subrayan la importancia de contar con estrategias de contingencia y una planificación robusta en proyectos de desarrollo de software, especialmente en un entorno empresarial donde los plazos son cruciales.

CAPÍTULO 6

Implantación, Pruebas y Resultados

La fase de implantación, pruebas y resultados constituye un elemento esencial en el ciclo de vida del desarrollo del proyecto. En esta etapa se procede no solo a la implementación de la solución en el entorno de destino, sino que también se somete a un riguroso conjunto de pruebas para garantizar que cumple con los requisitos técnicos y empresariales establecidos.

6.1 Introducción y Metodología Usada

Para la validación de la solución propuesta, se han aplicado diversas modalidades de pruebas. Inicialmente, se efectuaron pruebas dinámicas de integración con el objetivo de verificar la correcta interacción entre los distintos componentes de la solución. Posteriormente, se realizaron pruebas funcionales para confirmar el adecuado rendimiento de cada una de las funcionalidades incorporadas. Por último, se ejecutaron pruebas de aceptación centradas en la validación de los requisitos empresariales.

Adicionalmente, se ha contado con la revisión y validación por parte de colegas de proyecto y usuarios finales. El retroalimentación obtenida mediante este proceso colaborativo ha resultado invaluable para el ajuste y refinamiento de la solución, asegurando que esta esté en sintonía con las expectativas y requerimientos del cliente.

Esta metodología de validación de múltiples niveles tiene como finalidad la detección y corrección de cualquier defecto o imperfección previo al despliegue de la solución en el entorno de producción, lo que contribuye a una implementación más eficaz y exitosa.

6.2 Resultados en Actualización Automática de Campos

El requisito de Actualización Automática de Campos se diseñó para mejorar la eficiencia en la gestión de datos dentro del marco de C360 y B360. No obstante, durante el proceso de validación y pruebas, se identificaron ciertas limitaciones inherentes a las herramientas y tecnologías empleadas, lo cual constituye una oportunidad para futuras investigaciones y mejoras en el sistema.

6.2.1. Resultados

A pesar de contar con un diseño robusto y bien fundamentado, la implementación no logró operar como se esperaba. Esto no debe interpretarse como un fracaso, sino más bien como una valiosa fuente de aprendizaje que marca un camino para futuras investigaciones.

6.2.2. Motivos del Fracaso

- La función de *external validation* empleada se encuentra limitada a los atributos de tipo texto en el ámbito de la calidad de datos. Este aspecto restringe su aplicabilidad en la comparación de los LUD (*Last Updated Date*) de cada uno de los atributos.
- El proceso CAI no admite la recepción de múltiples valores de referencias cruzadas (xref) para su comparación, limitándose a trabajar con pares de valores.

6.2.3. Posibles Implementaciones Futuras

- Una opción sería convertir todos los campos relevantes a formato de texto. Sin embargo, esto podría suponer un coste significativo y reducir la flexibilidad del sistema, haciéndolo una opción poco viable a priori.
- Otra alternativa sería la integración con otros tipos de servicios como REST, SOAP, o alguna otra API que permita una mayor flexibilidad en el manejo de los datos y sus atributos.

La presente evaluación de la Actualización Automática de Campos, más que un punto final, representa una etapa en el continuo proceso de optimización y mejora de las prácticas de gestión de datos empresariales. Cada hallazgo, ya sea positivo o negativo, contribuye al cuerpo de conocimiento en este requisito, brindando información invaluable para sucesivas fases de desarrollo para una futura implantación exitosa.

6.3 Resultados en Verificación Manual de Actualizaciones

Esta sección tiene como objetivo examinar los resultados obtenidos de la funcionalidad de Verificación Manual de Actualizaciones, que fue diseñada para reforzar el control administrativo sobre las actualizaciones de datos. Contrario a los retos enfrentados en la Actualización Automática de Campos, esta funcionalidad ha demostrado un alto nivel de eficiencia y utilidad práctica.

6.3.1. Resultados

El diseño fue sólidamente construido y las pruebas subsiguientes confirmaron su eficacia en la operación. Por lo tanto, puede afirmarse que la implementación de esta funcionalidad ha sido un éxito, satisfaciendo tanto los requisitos técnicos como las expectativas de los usuarios.

6.3.2. Pruebas

Las pruebas de validación se llevaron a cabo con el involucramiento de distintos miembros del proyecto y clientes. La retroalimentación ha sido positiva, destacando el valor añadido que la funcionalidad de Verificación Manual ofrece al sistema de gestión de datos.

6.3.3. Impacto del Éxito en el Proyecto

El éxito de esta implementación augura un impacto significativamente positivo para el proyecto en general. Es especialmente relevante para los clientes finales, quienes serán los administradores del sistema. Esta funcionalidad permite un mayor control sobre las actualizaciones de datos, lo que contribuye a la mejora de la calidad y seguridad de la información almacenada en la base de datos.

En resumen, la Verificación Manual de Actualizaciones representa una mejora tangible en el marco general del sistema de gestión de datos, contribuyendo a un mayor control y calidad en la administración de la información.

CAPÍTULO 7

Conclusiones

Este Trabajo de Fin de Grado (TFG) se propuso realizar un estudio transversal en el ámbito de la Gestión de Datos Maestros, prestando especial atención a su implementación en entornos empresariales y a las tecnologías emergentes que facilitan su despliegue. Se ha abordado la falta de recursos académicos y técnicos en español sobre este tema, contribuyendo así a su divulgación en este idioma.

Se ha logrado proporcionar una comprensión multidisciplinaria del campo de MDM, desde la *Data Governance* hasta la *Data Quality*, cubriendo una amplia gama de temas que son fundamentales para entender los retos y oportunidades que ofrece este campo. Este trabajo ha servido también como recurso para futuros estudiantes y profesionales, aportando al cuerpo de conocimiento académico en la materia.

En cuanto a la implementación de pruebas de concepto, aunque se enfrentaron desafíos significativos como la caída del servicio de Informática Cloud y ciertas limitaciones de funcionalidades, estos obstáculos han sido instructivos. Han mostrado cómo los conceptos teóricos pueden materializarse en aplicaciones prácticas, permitiendo una mejor comprensión de las funcionalidades y limitaciones de las tecnologías existentes.

Además, este TFG ha fomentado el nivel de conciencia sobre la importancia del MDM en el entorno empresarial, un campo que a menudo es ignorado o desconocido para la comunidad de ingenieros informáticos en general. Al ser un estudio transversal, este trabajo también ha sentado las bases para futuras investigaciones en el campo del MDM.

Sobre la realización del proyecto, se han encontrado varios problemas y desafíos, tales como la caída de servicios y limitaciones técnicas, que se han abordado mediante ajustes en la planificación y la búsqueda de soluciones alternativas. Los errores cometidos y las dificultades encontradas han servido como aprendizaje tanto profesional como personal, enseñando la importancia de la adaptabilidad y resiliencia en el desarrollo de soluciones tecnológicas.

Finalmente, este proyecto ha sido una oportunidad invaluable para profundizar en el entendimiento de diferentes tecnologías, muchas de las cuales no se habían cubierto durante la carrera. Ha permitido integrar conocimientos de diferentes disciplinas para resolver un problema complejo, haciendo un uso adecuado de la tecnología de vanguardia.

Relación del trabajo desarrollado con los estudios cursados

El Trabajo de Fin de Grado ha sido una oportunidad excepcional para aplicar los conocimientos adquiridos durante mi formación académica en un contexto práctico y laboral real con la empresa Infoverity S.L. Las áreas del conocimiento de gestión de proyectos, SaaS, arquitecturas cloud, diseño de bases de datos y estudio de requisitos, entre otras, se han abordado en asignaturas obligatorias generales de la carrera. Específicamente, las asignaturas de la rama de TIC, como *Gestión de Bases de Datos*, *Integración de Aplicaciones*, *Sistemas y Servicios en Red* y *Desarrollo Centrado en el Usuario*, han sido fundamentales para el desarrollo de este TFG debido a las similitudes y coincidencias de los campos con el desarrollo de un proyecto en MDM.

En términos de competencias transversales, las más relevantes han sido:

- **CT-01. Comprensión e integración:** El proyecto permitió la síntesis y aplicación efectiva de conocimientos especializados, así como su integración en campos adicionales relacionados con la gestión de datos.
- **CT-02. Aplicación y pensamiento práctico:** La naturaleza práctica del TFG implicó una aplicación directa de teoría en la resolución de problemas específicos, así como en el diseño e implementación de soluciones efectivas.
- **CT-03. Análisis y resolución de problemas:** El desarrollo del TFG exigió la identificación, análisis y resolución de diversos desafíos, demostrando una efectiva capacidad para abordar problemas complejos.
- **CT-08. Comunicación efectiva:** La redacción académica del TFG, complementada por presentaciones y discusiones con asesores y colaboradores, requirió un alto grado de competencia en comunicación tanto oral como escrita.
- **CT-12. Planificación y gestión del tiempo:** La coordinación y ejecución exitosa del proyecto dentro de un marco de tiempo predeterminado puso de relieve las habilidades en planificación y en la gestión efectiva del tiempo.
- **CT-13. Instrumental específica:** El empleo de tecnologías y herramientas actualizadas en la implementación del proyecto denota la habilidad para emplear técnicas, habilidades y herramientas relevantes en el campo de la Gestión de Datos Maestros.

En conclusión, este Trabajo de Fin de Grado ha logrado cumplir la mayoría de los objetivos planteados, ofreciendo una contribución significativa tanto a la comunidad académica como profesional en el campo de la Gestión de Datos Maestros.

Bibliografía

- [1] Alex Berson and Lawrence Dubov. *Master Data Management and Data Governance*. McGraw-Hill, Second edition, 2011.
- [2] W.-J. Chen, B. Eshwar, R. Rajendiran. *Master Data Management for SaaS Applications*. IBM Corporation International Technical Support Organization, 2014.
- [3] Earley S, Henderson D, Data Management Association. *Dama-Dmbok : Data Management Body of Knowledge*. Technics Publications, Bradley Beach, New Jersey, 2nd edition, 2017.
- [4] Informatica: *What is Master Data Management?* Consultado en <https://www.informatica.com/resources/articles/what-is-master-data-management.html>.
- [5] Informatica: *Customer Success Stories* Consultado en <https://www.informatica.com/about-us/customers/customer-success-stories.html>.
- [6] Informatica: *MDM SaaS: Five Things You Need to Know*. Consultado en <https://www.informatica.com/blogs/mdm-saas-five-things-you-need-to-know.html>.
- [7] Mike Kavis. *Architecting the Cloud: Design Decisions for Cloud Computing Service Models (SaaS, PaaS, and IaaS)*. Wiley, 1st edition, 2014.
- [8] David Loshin. *Master Data Management [electronic Resource]*. Elsevier/Morgan Kaufmann, 1st edition, 2009.
- [9] Rajnil Thakur. *Cloud vs On-Premise MDM Deployment*. *Scalefusion Blog*, 22/09, 2022. Consultado en <https://blog.scalefusion.com/cloud-vs-on-premise-mdm-deployment/>.
- [10] Joko Wibowo, E., Kalamullah Ramli. *Impact of Implementation of Information Security Risk Management and Security Controls on Cyber Security Maturity (A Case Study at Data Management Applications of XYZ Institute)*. *Jurnal Sistem Informasi (Journal of Information System)*, 18(2), 1–17, 2022.

Glosario y definiciones

- **B360: Business 360** – Solución Cloud de administración.
- **Base Object** – Objeto donde se almacenan registros maestros post-concordancia y fusión.
- **Business Entities** – Representaciones de datos maestros para aplicaciones de negocio.
- **Business Events** – Eventos disparados por condiciones o cambios en los datos.
- **C360: Customer 360** – Solución Cloud de MDM SaaS.
- **CAI Process** – Proceso para la integración de aplicaciones compuestas.
- **DAMA DMBoK** – Libro de conocimientos de gestión de datos de DAMA.
- **Data Governance** – Gestión y control de la calidad y disponibilidad de los datos.
- **Data Quality** – Medida de la precisión y coherencia de los datos.
- **External Validation** – Validación realizada fuera del sistema MDM.
- **Informatica** – Empresa especializada en MDM.
- **LUD: Last Update Date** – Fecha de la última actualización de un registro.
- **Match** – Identificación de registros que representan la misma entidad.
- **MDM: Master Data Management** – Gestión de datos maestros.
- **Merge** – Fusión de registros coincidentes en un registro maestro.
- **PoC: Proof of Concept** – Demostración de la viabilidad de una idea o concepto.
- **RowIdObject** – Identificador único para un objeto base en MDM.
- **RowObject** – Referencia al objeto real en MDM, usualmente vinculado al 'RowIdObject'.
- **SaaS: Software as a Service** – Software ofrecido como servicio en la nube.

- **Survivorship** – Reglas para determinar qué datos sobreviven en una fusión de registros.
- **XREF** – Referencia cruzada entre registros de origen y registros maestros.

.1 Objetivos de Desarrollo Sostenible



ANEXO

OBJETIVOS DE DESARROLLO SOSTENIBLE

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.				X
ODS 3. Salud y bienestar.				X
ODS 4. Educación de calidad.				X
ODS 5. Igualdad de género.				X
ODS 6. Agua limpia y saneamiento.				X
ODS 7. Energía asequible y no contaminante.				X
ODS 8. Trabajo decente y crecimiento económico.				X
ODS 9. Industria, innovación e infraestructuras.	X			
ODS 10. Reducción de las desigualdades.				X
ODS 11. Ciudades y comunidades sostenibles.				X
ODS 12. Producción y consumo responsables.				X
ODS 13. Acción por el clima.				X
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz, justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos.				X

Reflexión sobre la relación del TFG/TFM con los ODS y con el/los ODS más relacionados.

El presente TFG sobre Master Data Management (MDM) se sitúa en un punto de confluencia entre la tecnología de la información y la sostenibilidad, proporcionando una plataforma única desde la cual abordar algunos de los desafíos más significativos identificados en los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas. Al centrarse en el ODS 9: Industria, Innovación e Infraestructura, este trabajo no solo resalta la importancia de una gestión eficiente de los datos para el crecimiento y desarrollo de la industria moderna, sino que también subraya cómo la innovación en este campo puede servir como catalizador para alcanzar una infraestructura sostenible y resiliente.

La gestión efectiva de los datos maestros es fundamental para cualquier organización que busque optimizar sus operaciones y facilitar la toma de decisiones basada en datos. En este contexto, el MDM puede jugar un papel crucial en la promoción de prácticas de negocio sostenibles. Por ejemplo, al garantizar la precisión, coherencia y disponibilidad de los datos esenciales, las empresas pueden mejorar significativamente su eficiencia operativa, reduciendo así su huella ambiental a través de la minimización del desperdicio de recursos y la optimización del consumo energético.

Además, la implementación de soluciones de MDM avanzadas fomenta la innovación al proporcionar un marco sólido sobre el cual se pueden desarrollar nuevas tecnologías y aplicaciones. Esto es especialmente relevante en el contexto del ODS 9, ya que el estímulo a la innovación tecnológica es un componente clave para construir infraestructuras resilientes que puedan resistir a los desafíos del mañana. Por ejemplo, datos gestionados de manera eficaz pueden facilitar el desarrollo de ciudades inteligentes, las cuales utilizan tecnologías de información y comunicación para mejorar la calidad de los servicios urbanos, reducir costos y consumos de recursos, y conectar y mejorar la vida de sus ciudadanos.

Este TFG también destaca cómo el MDM puede ser instrumental en el fomento de una industria inclusiva y sostenible. Al mejorar la calidad y accesibilidad de los datos, las organizaciones pueden no solo aumentar su competitividad sino también asegurar que sus prácticas de negocio sean inclusivas, promoviendo la igualdad de oportunidades y fomentando un entorno de trabajo equitativo. Esto, a su vez, puede contribuir a la creación de empleo y al fomento de una economía más robusta y resiliente.

En términos de infraestructura, la gestión eficiente de los datos maestros puede apoyar la planificación y ejecución de proyectos de infraestructura más sostenibles. Al disponer de datos fiables y precisos, los planificadores y diseñadores pueden tomar decisiones más informadas que no solo busquen cumplir con los objetivos a corto plazo, sino que también consideren la sostenibilidad y resiliencia a largo plazo. Esto es crucial para el



desarrollo de infraestructuras que no solo sirvan a las necesidades actuales, si no que también sean capaces de adaptarse a las cambiantes condiciones ambientales y sociales.

Por último, es importante reconocer que la contribución de este TFG a los ODS va más allá del ámbito técnico. Al enfocarse en el MDM desde una perspectiva de sostenibilidad, este trabajo fomenta una cultura de responsabilidad y conciencia sobre la importancia de los datos en la consecución de un futuro más sostenible. Promueve, por tanto, una reflexión sobre cómo la tecnología de la información puede y debe ser diseñada y utilizada de manera que contribuya positivamente a los desafíos globales, marcando un camino hacia un desarrollo que beneficie tanto a la sociedad actual como a las generaciones futuras.