



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería de Sistemas

Aplicación cloud para la carga de información crediticia de clientes en el área comercial de una entidad financiera

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el Título Profesional de Ingeniera de Sistemas

AUTOR

Sttefany Hilda REGALADO VALENZUELA

ASESOR

Víctor Hugo BUSTAMANTE OLIVERA

Lima, Perú

2022



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Regalado, S. (2022). *Aplicación cloud para la carga de información crediticia de clientes en el área comercial de una entidad financiera*. [Trabajo de suficiencia profesional de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Sistemas]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	STTEFANY HILDA REGALADO VALENZUELA
Tipo de documento de identidad	DNI
Número de documento de identidad	72528525
URL de ORCID	https://orcid.org/0000-0003-3284-6783
Datos de asesor	
Nombres y apellidos	Víctor Hugo Bustamante Olivera
Tipo de documento de identidad	DNI
Número de documento de identidad	25655590
URL de ORCID	https://orcid.org/0000-0001-8805-9629
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Fany Yexenia Sobero Rodríguez
Tipo de documento	DNI
Número de documento de identidad	20120467
Miembro del jurado 1	
Nombres y apellidos	Pablo Jesús Romero Naupari
Tipo de documento	DNI
Número de documento de identidad	06182185
Datos de investigación	
Línea de investigación	Data Science
Grupo de investigación	ITDATA
Agencia de financiamiento	Financiamiento Propio
Ubicación geográfica de la investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: Los Olivos

	Avenida Naranjal 1456 interior 501, Urb. Parque Naranjal. Universidad Nacional Mayor de San Marcos Latitud: -11.9759376 Longitud: -77.0853817
Año o rango de años en que se realizó la investigación	2021
URL de disciplinas OCDE	2.02.04 -- Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
Escuela Profesional de Ingeniería de Sistemas

Acta Virtual de Sustentación
del Trabajo de Suficiencia Profesional

Siendo las 19:56 horas del día 10 de enero del año 2022, se reunieron virtualmente los docentes designados como Miembros del Jurado del Trabajo de Suficiencia Profesional, presidido por la Ing. Sobero Rodríguez Fany Yexenia (Presidente), Lic. Pablo Jesús Romero Naupari (Miembro) y el Lic. Bustamante Olivera Víctor Hugo (Miembro Asesor), usando la plataforma Meet (<https://meet.google.com/jjy-yahj-fza>), para la sustentación virtual del Trabajo de Suficiencia Profesional intitulado: **“APLICACIÓN CLOUD PARA LA CARGA DE INFORMACIÓN CREDITICIA DE CLIENTES EN EL ÁREA COMERCIAL DE UNA ENTIDAD FINANCIERA”**, por la Bachiller **Regalado Valenzuela Sttefany Hilda**; para obtener el Título Profesional de Ingeniera de Sistemas.

Acto seguido de la exposición del Trabajo de Suficiencia Profesional, la Presidente invitó a la Bachiller a dar las respuestas a las preguntas establecidas por los miembros del Jurado.

La Bachiller en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez a las observaciones y preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, la Bachiller obtuvo la nota de **19 DIECINUEVE**.

A continuación la Presidente de Jurados la Ing. Sobero Rodríguez Fany Yexenia, declara a la Bachiller **Ingeniera de Sistemas**.

Siendo las 20:50 horas, se levantó la sesión.

Presidente

Ing. Sobero Rodríguez Fany Yexenia

Miembro

Lic. Romero Naupari Pablo Jesús

Miembro Asesor

Lic. Bustamante Olivera Víctor Hugo



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, DECANA DE AMÉRICA
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Profesional de Ingeniería de Sistemas

INFORME DE EVALUACIÓN DE ORIGINALIDAD

1. Facultad de Ingeniería de Sistemas e Informática
2. Escuela Profesional de Ingeniería de Sistemas
3. Autoridad académica que emite el informe de originalidad
Director de la EPIS
4. Apellidos y Nombres de la autoridad académica
Dra. Luzmila Elisa Pro Concepción
5. Operador del programa informático de similitudes
Lic. Ulises Roman Concha
6. Documento evaluado
Título de pregrado: "Aplicación cloud para la carga de información crediticia de clientes en el área comercial de una entidad financiera"
7. Autor del documento
Bach. Regalado Valenzuela, Sttefany Hilda
8. Fecha de recepción del documento 06/12/2021
9. Fecha de aplicación del programa informático de similitudes 02/12/2021
10. Software utilizado
 - Turnitin
11. Configuración del programa detector de similitudes
 - Excluye textos entrecomillados
 - Excluye bibliografía
 - Excluye cadenas menores a 40 palabras
12. Porcentaje de similitudes según programa detector de similitudes **5 (cinco) %**
13. Fuentes originales de las similitudes encontradas
Se adjunta en el anexo 1
14. Observaciones
15. Calificación de originalidad
 - Documento cumple criterios de originalidad, sin observaciones ☒
 - Documento cumple criterios de originalidad, con observaciones ☐
 - Documento no cumple criterios de originalidad ☐
16. Fecha de informe 03/06/2021



Firmado digitalmente por PRO
CONCEPCIÓN Luzmila Elisa FAU
20148092282 soft
Motivo: Soy el autor del documento
Fecha: 24.03.2022 16:19:10 -05:00

Firma de evaluador
Dra. Luzmila E. Pró Concepción
Directora (e) de la EPIS



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, DECANA DE AMÉRICA
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Profesional de Ingeniería de Sistemas

ANEXO 1

Fuentes originales de las similitudes encontradas

1. cybertesis.unmsm.edu.pe: 2%
2. hdl.handle.net: 1%
3. vicentegarciadiez.com: 1%
4. azure.microsoft.com: 1%
5. www.bnamericas.com: <1%
6. docs.microsoft.com: <1%



UNMSM

Firmado digitalmente por PRO
CONCEPCION Luzmila Elisa FAU
20148092282 soft
Motivo: Soy el autor del documento
Fecha: 24.03.2022 16:19:10 -05:00

Firma de evaluador

Dra. Luzmila E. Pró Concepción

Director (e) de la EPIS

DEDICATORIA

El presente trabajo está dedicado principalmente a Dios quien es hacedor de todas las cosas, y hoy nos permite seguir creciendo profesionalmente. Y a mis queridos familiares, padre, madre, hermana, abuela y novio, quiénes son sinónimo de amor y apoyo incondicional.

AGRADECIMIENTOS

A mis padres Hilda Valenzuela Flores y Godofredo Mauro Regalado Osorio:

Mi madre que debe sentirse orgullosa de verme lograr mis metas y sueños desde lo alto del cielo, le dedico este logro. A mi padre que nunca deja de creer en mí y me brinda su apoyo incondicional. A ambos les doy las gracias por inculcarme con amor todos los valores que me hacen ser la profesional que soy ahora.

A mi hermana Evelyn Rosa Regalado Valenzuela:

Por el apoyo, la empatía y comprensión brindados durante esta etapa.

A mi abuela Zenobia Flores Ramírez:

Por el apoyo y el ánimo brindados durante esta etapa.

A mi novio Jhoan Leonardo Calderón Hernández:

Por su amor, paciencia, comprensión, ánimo y conocimientos compartidos para lograr la realización del informe.

A mis amigos, profesores y compañeros de trabajo:

Por el apoyo y ánimo para cumplir con este objetivo.

A mi asesor Víctor Bustamante:

Por las sugerencias y recomendaciones brindadas durante la elaboración de este trabajo.

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA
PROFESIONAL DE INGENIERÍA DE SISTEMAS

**Aplicación cloud para la carga de información crediticia de clientes en el área comercial
de una entidad financiera**

Autor: Regalado Valenzuela, Sttefany Hilda

Asesor: Bustamante Olivera, Víctor Hugo

Título: Trabajo de Suficiencia Profesional para optar el Título Profesional de
Ingeniero de Sistemas

Fecha: Noviembre del 2021

RESUMEN

La aplicación cloud basado en una solución Big Data que realiza un proceso de carga de información crediticia de clientes en una base de datos no relacional como es CosmosDB, esta información se dispone mediante un API de negocio, que a su vez consume esta data y proporcionar una vista única al área de ventas, de esa manera les facilita la evaluación de venta de créditos o demás productos financieros, ya que tendrían una forma fácil de acceder a esa información, con 0 costo.

Con la implementación del presente proyecto se logró contribuir en la colocación y oferta de productos minoristas en el área comercial de la entidad financiera. Asimismo, se está generando ahorro en los costos de operación por la solicitud de dicha información.

El presente trabajo de suficiencia profesional se detalla mi participación en el proyecto de desarrollo de este aplicativo cloud para una entidad financiera del Perú.

Este proyecto fue desarrollado en una plataforma cloud (Microsoft Azure), y desplegado sobre un clúster de Azure Databricks para el procesamiento de la información y la ingesta en CosmosDB.

Palabras claves: Big Data, Cosmos DB, aplicación, cloud, Azure Databricks entidad financiera.

MAJOR NATIONAL UNIVERSITY OF SAN MARCOS
FACULTY OF SYSTEMS AND COMPUTER ENGINEERING
PROFESSIONAL SCHOOL OF SYSTEMS ENGINEERING

**Cloud application for uploading customer credit information in the commercial area
of a financial institution**

Author: Regalado Valenzuela, Sttefany Hilda
Advisor: Bustamante Olivera, Víctor Hugo
Title: Work of Professional Sufficiency to opt for the Professional Title of
Systems Engineer
Date: November 2021

ABSTRACT

The cloud application based on a Big Data solution that performs a process of loading customer credit information in a non-relational database such as CosmosDB, this information is available through a business API, which in turn consumes this data and provides A single view of the sales area, thus making it easier for them to evaluate the sale of credits or other financial products, since they would have an easy way to access that information, with 0 cost.

With the implementation of this project, it was possible to contribute to the placement and offer of retail products in the commercial area of the financial institution. Likewise, savings in operating costs are being generated by requesting such information.

The present work of professional sufficiency details my participation in the development project of this cloud application for a financial entity in Peru.

This project was developed on a cloud platform (Microsoft Azure) and deployed on an Azure Databricks cluster for information processing and ingest in CosmosDB.

Keywords: Big Data, Cosmos DB, application, cloud, Azure Databricks financial institution.

TABLA DE CONTENIDOS

RESUMEN.....	vi
ABSTRACT.....	vii
INDICE DE FIGURAS.....	x
INDICE DE TABLAS.....	xii
INTRODUCCION.....	1
CAPITULO I TRAYECTORIA PROFESIONAL.....	3
CAPÍTULO II CONTEXTO EN EL QUE SE DESARROLLÓ LA EXPERIENCIA.....	7
2.1. Empresa – actividad que realiza	7
2.2. Misión	7
2.3. Visión.....	7
2.4. Organización de la empresa	7
2.5. Área, cargo y funciones desempeñadas	8
2.6. Experiencia profesional realizada en la organización.....	9
CAPITULO III ACTIVIDADES DESARROLLADAS	10
3.1. Situación problemática.....	10
3.1.1. Definición del problema	10
3.2. Solución.....	10
3.2.1. Objetivos	10
3.2.2. Alcance.....	11
3.2.3. Etapas y metodología	12
3.2.4. Fundamentos utilizados	19
3.2.5. Implementación de las áreas, procesos, sistemas y buenas practicas ...	29

3.3.	Evaluación	53
3.3.1.	Evaluación económica	53
3.3.2.	Evaluación de los beneficios.....	59
CAPITULO IV REFLEXION CRITICA DE LA EXPERIENCIA		60
4.1.	Aportes/ lecciones aprendidas/ en el qué se puede mejorar	60
4.1.1.	Aportes	60
4.1.2.	Lecciones aprendidas.....	61
CAPITULO V CONCLUSIONES Y RECOMENDACIONES		62
5.1.	Conclusiones	62
5.2.	Recomendaciones	63
GLOSARIO		65
ANEXOS.....		67
ANEXO 1:	Repositorio del proyecto de carga RCC	67
ANEXO 2:	Azure Data Factory – Pipeline que invoca proceso Databricks.....	68
ANEXO 3:	Azure Data Factory - Pipeline que realiza el copiado al ADLS	70
ANEXO 4:	PowerShell script	71
ANEXO 5:	Script del archivo BAT.....	72
ANEXO 6:	Script del Job mainframe.....	72

INDICE DE FIGURAS

Figura 1. Organigrama Data & Analytics	8
Figura 2. Nueva estructura organizacional	13
Figura 3. Roadmap anual	14
Figura 4. Ciclo de vida del Scrum.....	20
Figura 5. Big Data 4Vs	21
Figura 6. Proceso ETL.....	22
Figura 7. Azure Overview	23
Figura 8. Ecosistema de Databricks en Microsoft Azure	24
Figura 9. Modelo NoSQL	25
Figura 10. Funcionalidades del Data Factory	26
Figura 11. Registro de una épica en Jira	30
Figura 12. Pipeline del Data Factory.....	35
Figura 13. Ruta en Azure Data Lake Storage	35
Figura 14. Delta Tables Bronze, Silver, Golden.....	36
Figura 15. Pipeline que ejecuta JAR de aplicación	37
Figura 16. Diagrama de la malla automática.....	38
Figura 17. Archivo plano con información RCC	41
Figura 18. Archivo plano de control	41
Figura 19. Arquitectura de Solución.....	42
Figura 20. Job de despliegue en Jenkins.....	45
Figura 21. Creación de Pull Request en BitBucket.....	47

Figura 22. Creación de MVP	48
Figura 23. Estadios del MVP	49
Figura 24. Casos de pruebas	50

INDICE DE TABLAS

Tabla 1 Experiencia profesional	3
Tabla 2 Formación académica profesional	5
Tabla 3 Cursos	5
Tabla 4 Idiomas	5
Tabla 5 Conocimientos	6
Tabla 6 Logros por iteración	15
Tabla 7 Análisis funcional	31
Tabla 8 Análisis técnico	32
Tabla 9 Estructura de archivo plano origen de la información RCC	39
Tabla 10 Estructura del archivo de control	41
Tabla 11 Actividades de desarrollo	43
Tabla 12 MVPs desarrollados en el proyecto	49
Tabla 13 Test Cases	51
Tabla 14 Gastos tecnológicos	53
Tabla 15 Activos	54
Tabla 16 Gastos administrativos	55
Tabla 17 Ahorro de costos proyectado a 5 años	56
Tabla 18 Flujo de caja	57
Tabla 19 VAN y TIR	58

INTRODUCCION

En el presente informe profesional se describirá la implementación de un modelo y proceso de carga que disponibilizará información de las deudas de los clientes (RCC¹) en el sistema financiero proporcionado por la SBS para contribuir y asegurar la colocación y oferta de productos minoristas en el área comercial de la entidad financiera. Todo esto se originó debido a la necesidad de contar y acceder a una única vista de la información RCC de los clientes en el portal del área comercial. El área del banco Salesforce es responsable de proporcionar la información de deudas de los clientes al área comercial, sin embargo, el proceso de solicitud es engorroso y costoso monetariamente dado que, para la generación de esta información, el área a cargo (Datalake²) requiere ejecutar diversos procesos complejos y costosos a nivel de procesamiento ya que se maneja volúmenes altos de data. De esta manera, nace la necesidad de buscar una nueva y única fuente de información para mejorar la experiencia de la actividad comercial y facilitar la oferta de productos minoristas generando mayores ventas mensuales en la entidad financiera y ahorro monetario en la adquisición de la información crediticia de clientes.

En el CAPÍTULO I – TRAYECTORIA PROFESIONAL, se describe mi formación educativa y mi experiencia profesional. Asimismo, se precisan los logros que he realizado.

En el CAPITULO II - CONTEXTO EN EL QUE SE DESARROLLÓ LA EXPERIENCIA, se describe la empresa donde se desarrolló el presente proyecto, su organización, visión, misión, las funciones que desempeñé durante la ejecución del proyecto y los logros obtenidos.

En el CAPÍTULO III - ACTIVIDADES DESARROLLADAS, se define la situación problemática, la solución, los objetivos y todo el aporte que realicé durante la ejecución del proyecto.

En el CAPÍTULO IV - REFLEXIÓN CRÍTICA DE LA EXPERIENCIA, realiza una reflexión

¹ RCC: ver el apartado GLOSARIO para más detalle.

² Datalake: ver el apartado GLOSARIO para más detalle.

sobre lo comentado en capítulos anteriores.

El CAPÍTULO V - CONCLUSIONES Y RECOMENDACIONES describe las conclusiones y recomendaciones realizadas en base al presente trabajo de suficiencia profesional, además se incluyen algunos anexos con información adicional.

CAPITULO I

TRAYECTORIA PROFESIONAL

La autora actualmente tiene el grado de Bachiller en la carrera de Ingeniería de Sistemas de la Universidad Nacional Mayor de San Marcos. Presenta habilidades enfocadas en el análisis, diseño y desarrollo de procesamiento de datos mediante soluciones de Big Data en entorno Cloud Computing, creando propuestas de valor de alto impacto para el negocio con una experiencia de 5 años en el sector financiero.

Asimismo, se caracteriza por ser responsable, observadora, analítica, organizada y autodidacta, cuenta con habilidades interpersonales y blandas para el trabajo en equipo y resolución de problemas.

A continuación, se detalla la experiencia profesional:

Tabla 1

Experiencia profesional

EXPERIENCIA PROFESIONAL	
Empresa	Entidad financiera peruana
Cargo	Sub-Gerente Adjunto de Análisis de Aplicaciones
Área	Gerencia Data y Analytics
Fecha	Septiembre 2020 – Enero 2021
Funciones	• Colíder Técnico. • Desarrollo y gestión de aplicaciones Scala usando Spark Databricks, Kafka, para soluciones Big Data Batch y Streaming. • Migración de procesos de ingesta bajo un modelo de datos en Cassandra IaaS hacia recursos PaaS como Azure Databricks, Data Factory, CosmosDB, Blob Storage, SQL Database, etc. • Desarrollo de Scripts PowerShell y BATs en servidores Onpremise para despliegue de servicios Azure cumpliendo lineamientos de seguridad. • Desarrollo y mantenimiento de microservicios en Docker. • Integración y despliegue continuo con BitBucket y Jenkins. • Referente técnico para flujos de certificación y pases a producción, además de realizar revisiones de código y aprobaciones de pull request. • Migración de un microservicio ad-hoc desarrollado en el área hacia la nueva plataforma Kafka Confluent en todas las soluciones streaming que atribuimos.

Logros	• Lideró la migración del microservicio Kafka Mirror a la nueva plataforma Kafka Confluent para 4 soluciones streaming, responsable de la gestión y desarrollo. La finalidad de esta migración es la mejora de la experiencia de los clientes y eficiencia en la arquitectura de las soluciones Real Time. • Lideró la migración de las soluciones Batch de DataStax hacia Azure Databricks, CosmosDB, SQL Database, Data Factory y Azure Datalake Storage.
Cargo	Sub-Gerente Adjunto de Análisis de Aplicaciones
Área	Ingeniería y Desarrollo de TI
Fecha	Abril 2019 – Agosto 2020
Funciones	• Analista Experta del aplicativo mainframe CONNEX en la Tribu Canales Alternativos. • Responsable de la estabilidad del aplicativo mainframe CONNEX, Switch transaccional de los canales ATM y POS. • Análisis y resolución de incidencias en Producción. • Análisis e implementación de iniciativas/mejoras para el aplicativo mainframe CONNEX. • Configuración y gestión de cambios implementados a los cajeros mediante el aplicativo Connex en ambiente productivo. • Atención de requerimientos e implementación de optimizaciones del aplicativo Connex bajo la metodología Ágil KANBAN aplicado en la Tribu Canales Alternativos.
Logros	• Lideró la prueba de contingencia general en el Site de Chorrillos y La Molina, se realizó la activación del aplicativo CONNEX. • Lideró la carga masiva de cuentas corrientes y afiliación en el aplicativo Connex para el abono del AFP (2020).
Cargo	Ingeniero de Aplicaciones
Área	Ingeniería y Desarrollo de TI
Fecha	Abril 2017 – Marzo 2019
Funciones	• Desarrollo y mantenimiento de aplicaciones basadas en plataforma distribuida, enfocado en tecnologías como Visual C#, SQL Server y herramientas de entorno Linux / AIX, bajo estándares definidos en el banco. • Desarrollo y mantenimiento aplicaciones basadas en plataforma mainframe z/OS, enfocado en lenguajes de programación JCL y COBOL. • Elaboración de documento técnico, evaluación de factibilidad y solución técnica. • Elaboración de diagramas de las rutinas o malla de Jobs. • Responsable del desarrollo y seguimiento de todo el ciclo de vida de Software.

Logros	Lideró el proyecto de visualización grafica de mallas scheduler de la rutina critica en todos los ambientes a través de una aplicación web desarrollada en visual C#.
---------------	---

Nota: Elaboración propia

Tabla 2

Formación académica profesional

FORMACION ACADEMICA PROFESIONAL		
Formación	Institución	Periodo
Bachiller en Ingeniería de Sistemas (R.R. 05/06/2017)	Universidad Nacional Mayor de San Marcos – Facultad de Ingeniería de Sistemas e informática – Escuela Académico Profesional de Ingeniería de Sistemas	2012 - 2016

Nota: Elaboración propia

Tabla 3

Cursos

CURSOS			
Curso	Institución	Certificado(s)	Año
Curso de Especialización Big Data	Big Data Academy	• Big Data Engineer Professional • Big Data Architect Fundamentals	2019
Curso de Especialización en Spark	Big Data Academy	• Spark Professional	2020
Programa de Especialización de análisis de datos con Python	Data Science Research Perú	• Python Professional	2020

Nota: Elaboración propia

Tabla 4

Idiomas

IDIOMAS		
Curso	Institución	Periodo
Ingles nivel avanzado	ICPNA	2009 – 2012
Methodology Learning Teaching	ICPNA	2012 – 2013

Nota: Elaboración propia

Tabla 5**Conocimientos**

	CONOCIMIENTOS
Lenguajes de programación	Scala, Python, Java, C#
Base de datos	Oracle, MySQL, Azure SQL Database, CosmosDB
Gestores de bases de datos	SQL Server, Toad for Oracle, PLSQL Developer
Metodologías ágiles	SCRUM, KANBAN
Herramientas de data processing	Apache Spark (Batch y Streaming), Apache Kafka, Ecosistema Hadoop (HDFS, Hive, Sqoop, MapReduce, HBase), Integration Services (SSIS)
Herramientas de integración y despliegue continuo	Jenkins, groovy, ansible, Jira, Bitbucket, Artifactory and Fortify
Plataformas Cloud	Microsoft Azure (Blob Storage, ADLS, KeyVault, Active Directory, Databricks, DataFactory)
Herramienta de visualización de datos	Power BI

Nota: Elaboración propia

CAPÍTULO II

CONTEXTO EN EL QUE SE DESARROLLÓ LA EXPERIENCIA

2.1. Empresa – actividad que realiza

La entidad bancaria líder del holding financiero más importante del Perú. Ofrece una gran gama de productos y servicios financieros para diversos segmentos como personas, pymes y empresas. Su división de banca principalmente es por mayorista y minorista, donde la primera división abarca las operaciones de banca negocios, corporativa e institucional en donde ofrecen servicios como leasing, financiamiento corporativo, comercio exterior y servicios para empresas y en la segunda división ofrecen productos como créditos hipotecarios, tarjetas de crédito, productos de ahorro y productos para pymes.

La entidad bancaria tiene 130 años de presencia en el Perú y su red de más de 8340 puntos de contacto sirve a sus más de 13 millones de clientes.

2.2. Misión

Tiene como propósito transformar los planes de sus clientes en realidad ofreciendo productos y servicios financieros.

2.3. Visión

Ser la empresa peruana que brinda la mejor experiencia a los clientes de manera simple, cercana y oportuna, y ser referentes en la gestión empresarial potenciando el liderazgo histórico y transformador de la industria financiera en el Perú.

2.4. Organización de la empresa

A continuación, se presenta el organigrama del área Data & Analytics.

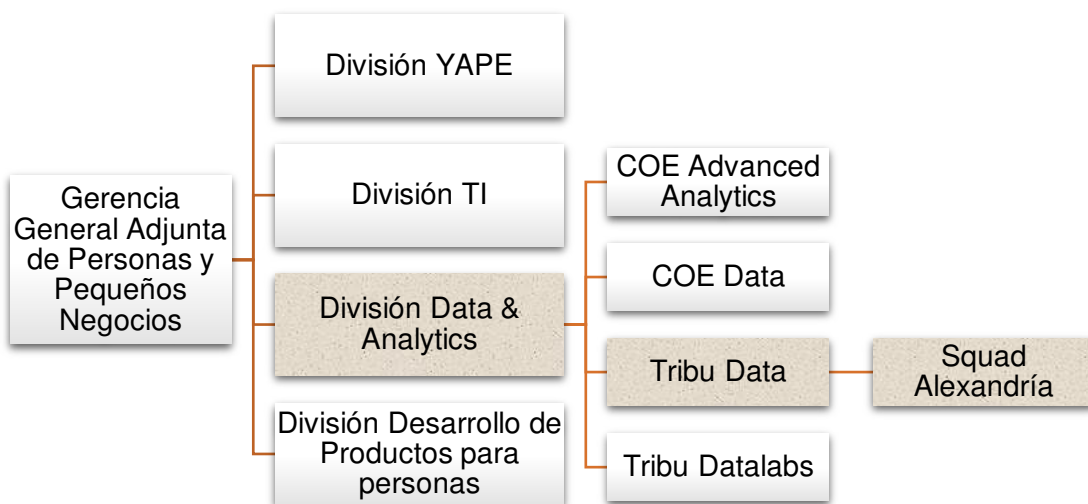


Figura 1. Organigrama Data & Analytics

Fuente: Elaboración propia

2.5. Área, cargo y funciones desempeñadas

La autora de este informe de experiencia profesional se desempeña como Data Engineer en el Squad Alexandria en el área GCIA DE DIVISIÓN DATA & ANALYTICS.

Las funciones que presenta la autora están alineadas a lo siguiente:

- Colíder Técnico.
- Desarrollo y gestión de aplicaciones Scala usando Spark Databricks, Kafka, para soluciones Big Data Batch y Streaming.
- Desarrollo de Scripts PowerShell y BATs en servidores Onpremise para despliegue de servicios Azure configurando y cumpliendo protocolos de seguridad.
- Desarrollo y mantenimiento de microservicios en Docker.
- Integración y despliegue continuo con BitBucket y Jenkins.
- Referente técnico para flujos de certificación y pases a producción, además de realizar revisiones de código y aprobaciones de pull request.

2.6. Experiencia profesional realizada en la organización

La autora es miembro del Squad Alexandria y presento las siguientes experiencias:

- Lideró la migración del microservicio Kafka Mirror a la nueva plataforma Kafka Confluent para 4 soluciones streaming, responsable de la gestión y desarrollo. La finalidad de esta migración es la mejora de la experiencia de los clientes y eficiencia en la arquitectura de las soluciones Real Time.
- Lideró la migración de procesos de ingesta bajo un modelo de datos en Cassandra IaaS hacia recursos PaaS como Azure Databricks, Data Factory, CosmosDB, Blob Storage, SQL Database, etc.
- Lideró la prueba de contingencia general en el Site de Chorrillos y La Molina, se realizó la activación del aplicativo CONNEX.
- Lideró la carga masiva de cuentas corrientes y afiliación en el aplicativo Connex para el abono del AFP (2020).
- Lideró el proyecto de visualización grafica de mallas scheduler de la rutina critica en todos los ambientes a través de una aplicación web desarrollada en visual C#.

CAPITULO III

ACTIVIDADES DESARROLLADAS

3.1. Situación problemática

El área comercial de una entidad financiera actualmente ha tenido la dificultad de no contar con un aplicativo que les permita visualizar y acceder fácilmente a la información crediticia de los clientes imposibilitando la colocación de productos financieros a los clientes.

3.1.1. Definición del problema

El problema no contar y poder acceder a una única vista de la información de deudas de los clientes en el portal del área comercial. El área del banco Salesforce es responsable de proporcionar la información de deudas de los clientes al área comercial, sin embargo, el proceso de solicitud es engorroso y costoso monetariamente dado que, para la generación de esta información, el área a cargo (Datalake) requiere ejecutar diversos procesos complejos y costosos a nivel de procesamiento ya que se maneja volúmenes altos de data. De esta manera, nace la necesidad de buscar una nueva y única fuente de información para mejorar la experiencia de la actividad comercial y facilitar la oferta de productos minoristas generando mayores ventas mensuales en la entidad financiera.

3.2. Solución

3.2.1. Objetivos

3.2.1.1. Objetivo general

Implementar un aplicativo de carga de la información crediticia de los clientes (Personas Naturales) en el sistema financiero utilizando una solución Big Data con el fin de generar ahorros monetarios por cada

operación de solicitud de información deudora de clientes en el área de ventas de una entidad bancaria.

3.2.1.2. Objetivos específicos

- Analizar y diseñar una solución Big Data en entorno Cloud (Azure).
- Desarrollar una aplicación en lenguaje de programación Scala, que será desplegado sobre Azure Databricks que realizará la carga de información crediticia de clientes en la base de datos CosmosDB, asimismo, desarrollar pipelines en el recurso Data Factory para orquestar cada subproceso del flujo de carga.
- Realizar pruebas y validar el modelo con data real de los momentos de vida de los colaboradores de la entidad bancaria y grupo de clientes.
- Despliegue del aplicativo en Producción.

3.2.2. Alcance

Alcance Funcional:

- Implementar una solución Big Data en entorno Cloud (Azure) para el procesamiento y carga de información crediticia de clientes (RCC).
- El aplicativo debe contener una lógica de encriptación de aquellos campos que tienen información sensible del cliente.
- El aplicativo debe manejar un control y validación de registros que pueden llegar de manera incorrecta, por lo tanto, el aplicativo tener una lógica implementada para filtrar los registros rechazados que no cumplen con el formato establecido a inicio del proyecto.
- Implementar dos Jobs mainframe que serán programados para ejecutarse con frecuencia mensual.

Alcance no funcional:

- Desarrollar una aplicación en lenguaje Scala que será desplegada sobre Azure Databricks.
- Configurar y desplegar el contenedor y base de datos en el recurso CosmosDB, mediante el cual persistirá la data.
- Desarrollar pipelines en el recurso Data Factory para orquestar cada subproceso del flujo de carga.
- Dentro del desarrollo se debe contemplar la transformación de la data que a través de procesos ETL serán cargadas en las capas de modelamiento de datos como Bronze, Silver, Golden, donde en la última capa se tendrá la data limpia y lista para ser cargada a CosmosDB.
- Desarrollar scripts PowerShell y BATs que a su vez ejecutaran los pipelines de los procesos de copiado de archivos de Onpremise a Cloud (Azure Data Lake Storage) y el proceso de carga hacia CosmosDB.
- La aplicación debe notificar al PO (Product Owner) y al líder técnico del Squad cuando se haya finalizado la carga mediante un correo.

No cubierto:

- En el presente proyecto no está abarcando el desarrollo del API de negocio que consumirá la información cargada en CosmosDB para exponerla como única vista al área de ventas de la entidad financiera, es decir, el alcance es acotado solo al proceso de carga/ingesta de la información crediticia de clientes.

3.2.3. Etapas y metodología

3.2.3.1. Metodología aplicada en el desarrollo del proyecto

La metodología SCRUM nos ofrece diversas estrategias y buenas prácticas que permiten la integración, transparencia y colaboración entre miembros del equipo para la entrega continua y generación de valor a la empresa.

Asimismo, la empresa donde se desarrolló el presente trabajo optó por realizar una transformación organizacional, digital y cultural debido a que surgieron necesidades en las cuales se requería ser más flexibles y ágiles para adaptarse a la demanda cambiante del cliente, de igual modo, poder mejorar el Time to Market³, y lograr incrementar la productividad de los trabajadores. De esa manera nacieron las tribus, squads y chapters.

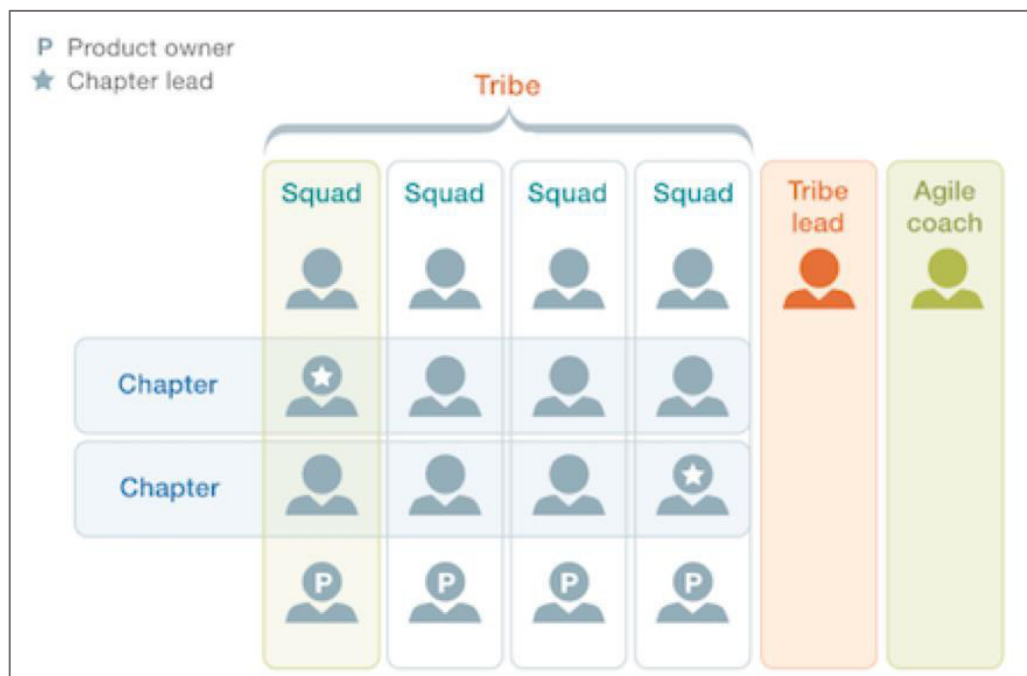


Figura 2. Nueva estructura organizacional

Fuente: (El futuro de los equipos, 2018)

“Una tribu es un conjunto de equipos con misiones interconectadas.” (El futuro de los equipos, 2018). Asimismo, mediante ceremonias ágiles se definen determinados objetivos y resultados claves (OKR – Objectives and Key Results) trimestralmente.

Los equipos que conforman la tribu son llamados Squads. “Un Squad tiene objetivos específicos y está conformado por expertos multidisciplinares” (El

³ Time To Market: ver el apartado GLOSARIO para más detalle.

futuro de los equipos, 2018) como desarrolladores, analistas de calidad, product owner, líder técnico y Agile Couch.

Por otro lado, con respecto al termino Chapter, se entiende que es un equipo conformado por miembros de la misma disciplina o especialidad y en donde tienen como finalidad el desarrollo y crecimiento profesional mediante el intercambio de conocimientos y programación para impartir capacitaciones, las cuales son organizados y liderados por el Chapter Lead.

Cada trimestre tiene una duración de 8 iteraciones y cada sprint o iteración dura 10 días laborables.

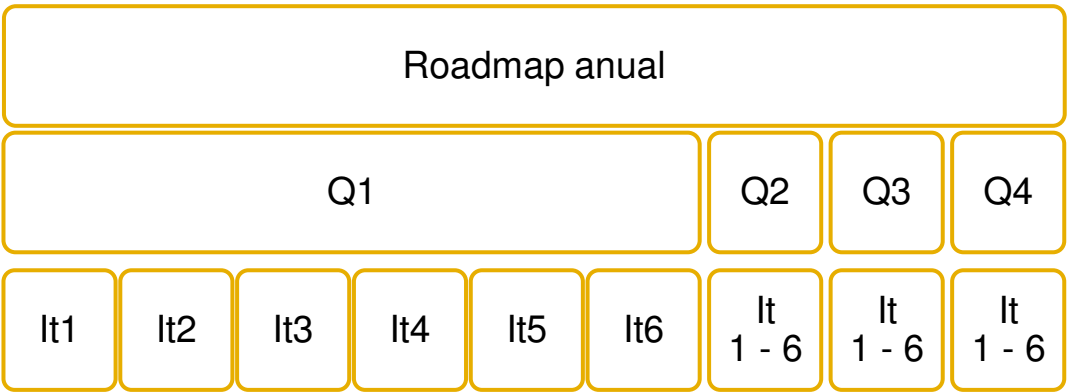


Figura 3. Roadmap⁴ anual

Fuente: Elaboración propia

El desarrollo e implementación del presente proyecto tuvo una duración de 6 iteraciones. A continuación, se describirán los logros por cada iteración.

⁴ Roadmap: ver el apartado GLOSARIO para más detalle.

Tabla 6**Logros por iteración**

Iteración 1 (01 abr - 14 abr)	
	Levantamiento de informacion para la definicion del modelado de datos. Desarrollo de malla automática mediante Jobs mainframe de ejecución Desarrollo de scripts bat y powershell para ejecución de pipelines Desarrollo de pipeline de copiado de archivo fuente en Azure Data Factory
Logro 1	Definición del modelado y desarrollo de malla automatica terminado
Iteración 2 (15 abr – 28 abr)	
	Certificación y Pase a producción de malla automática y Jobs mainframe Definicion de Seguridad para el uso de un algoritmo de encriptacion Desarrollo de la aplicación para la primera carga al stage Bronze Desarrollo de logica de encriptacion para campos sensibles Desarrollo del proceso de carga al stage Silver aplicando filtros y limpieza de datos
Logro 2	Certificación y Pase a producción de malla automática y Jobs mainframe
Iteración 3 (29 abr - 12 may)	
	Certificación y Pase a Producción de pipelines en Data Factory Desarrollo del proceso de carga al stage Silver aplicando filtros y limpieza de datos Desarrollo del proceso de carga al stage Golden aplicando transformaciones
Logro 3	Certificación y Pase a Producción de pipelines Data Factory
Iteración 4 (13 may - 26 may)	
	Desarrollo del proceso de carga al stage Golden aplicando transformaciones Desarrollo del proceso de carga a CosmosDB Pruebas unitarias por cada carga a las capas Bronze, Silver, Golden y CosmosDB Pruebas integrales end-to-end de las cargas a Bronze, Silver, Golden y CosmosDB
Logro 4	Avance del desarrollo a un 80%

Iteración 5 (27 may - 9 jun)

Pruebas ejecutando flujo completo desde la malla automática
Correcciones de vulnerabilidades en reportes de Fortify para el desarrollo de la aplicación
Correcciones de calidad de código según reporte de SonarQube para el desarrollo de la aplicación

Logro 5 Pruebas integrales en desarrollo completado.

Iteración 6 (10 jun - 23 jun)

Certificación y Pase a Producción de proceso de carga RCC
Coordinación con equipo de soporte para monitoreo de la aplicación
Reuniones con el equipo de Salesforce para validar resultados

Logro 6 Certificación y Pase a Producción exitoso de proceso de carga RCC

Nota: Elaboración propia

A continuación, se listarán los roles involucrados:

- **Product Owner:** es el rol conocedor del negocio, el que toma las decisiones asegurándose que se entregue valor a la empresa. Se encarga de interpretar y comprender los pedidos del área usuaria para ordenar por prioridad y colocarlo en el Backlog del Squad.
- **Agile Couch:** es el rol encargado de organizar las ceremonias para cada iteración y asegurarse que se está empleando correctamente la agilidad y conocimientos impartidos.
- **Team members:** son los especialistas (Data Engineers, Data Modelers, Data Governant, QA, etc.) directamente involucrados en el desarrollo de los requerimientos priorizados en el Squad.

Al inicio de cada trimestre se lleva a cabo una ceremonia llamada Big Room Planning⁵ donde se definen los objetivos e iniciativas que se priorizarán en el backlog de cada Squad que conforma la Tribu durante ese trimestre.

⁵⁵⁵ Big Room Planning: ver el apartado GLOSARIO para más detalle.

Cada Squad planifica las historias de usuario comprometidas en el Backlog, se evalúan también la existencia de dependencias y/o posibles impedimentos o riesgos.

La siguiente ceremonia es el sprint planning con frecuencia quincenal, tiene como objetivo priorizar las historias de usuario comprometidas que den valor al negocio. Durante el sprint, también se realizaba reuniones diarias donde cada team member sincroniza y comenta en caso tengan algún impedimento, asimismo, comentar los logros obtenidos el día anterior y los objetivos a lograr en el día. Al finalizar el sprint se realiza el sprint review, mediante esta ceremonia el squad presenta los logros obtenidos durante el sprint, esta exposición va dirigida a los stakeholders y a la líder de la tribu. Posterior a ello, se lleva a cabo la ceremonia de retrospectiva, en el cual cada team member del squad realiza una auto inspección para generar planes de acción sobre sus oportunidades de mejorar.

3.2.3.2. Etapas del desarrollo e implementación del aplicativo

Se tienen las siguientes etapas de implementación de la aplicación cloud para la carga de información crediticia de clientes:



1. Planificación:

- Se realiza la evaluación y revisión del alcance.
- Se elabora un plan de trabajo.

- Se realiza la estimación del esfuerzo de los desarrolladores y analistas de calidad en las actividades correspondientes.

2. Análisis y diseño:

- Levantamiento de información sobre la data que se va a cargar, se debe validar y tener en claro todas las transformaciones que realizara el aplicativo.
- Aprobación de la arquitectura de solución.
- Aprobación del uso adecuado de los lineamientos de seguridad.

3. Desarrollo de aplicación:

- Construcción del aplicativo cumpliendo estándares y lineamientos de la entidad bancaria.
- Realización de pruebas unitarias y pruebas integrales.
- Revisión de pares del código.
- Elaboración de documentación para el congelamiento⁶ y pase a producción.

4. Certificación:

- Reunión de transferencia entre los involucrados desarrolladores y analistas de calidad.
- La analista de calidad elabora sus casos de prueba.
- La analista de calidad ejecuta los casos de prueba.

5. Implementación:

- Instalación, despliegue y puesta en producción del aplicativo.
- Capacitación y transferencia al equipo de Soporte sobre estos nuevos cambios para que puedan monitorear que todo se ejecute correctamente.

⁶ Congelamiento: ver el apartado GLOSARIO para más detalle.

3.2.4. Fundamentos utilizados

3.2.4.1. Metodología “Scrum”

Según Bruno, Diogo, Eurico, Filipe y Paulo (2015), “la metodología Scrum contribuye en la identificación de problemas, cambios o imprevistos a lo largo del proceso de desarrollo de software. Su principal objetivo es el enfoque de valor al cliente”.

De acuerdo a (Antunes, Santos, Lopes, Fidalgo, & Alves, 2015) la metodología Scrum está estructurada en varios ciclos, en cada uno de estos ciclos se estima lograr un conjunto de tareas que conducen a un aumento funcional del producto final.

Por lo que en mi opinión esta metodología imparte muchas técnicas y conocimientos para optimizar el flujo de trabajo para la entrega de un producto que aporte valor al cliente eliminando cuellos de botella y así poder llegar a la fecha comprometida.

El ciclo de desarrollo Scrum tal como se ilustra en la Figura 4, inicia con la priorización de las tareas o historias de usuario a desarrollarse (Product Backlog).

Al comienzo de cada Sprint se realiza una reunión de planificación (Sprint Planning), donde se seleccionan las historias de usuario del Product Backlog, para ser priorizadas a implementarse durante el Sprint, por ende, son incluidas en el Sprint Backlog.

Todos los días se realiza una reunión (Sprint Daily Meeting) con una duración de no más de 15 minutos para que todo el equipo exponga la situación en la que se encuentra su trabajo, lo avanzado desde la última reunión, lo que va a desea lograr durante el día, y si hay impedimentos para su progreso.

Al final del Sprint se realiza otra reunión (revisión) con el Product Owner y stakeholders donde se exponen los avances y logros alcanzados durante el sprint. Después, el equipo realiza una retrospectiva (Scrum Sprint Retrospective) para identificar los problemas que han surgido durante el Sprint, resolviéndolos y evitándolos en el siguiente Sprint. Una vez finalizada la retrospectiva, se inicia de nuevo el proceso procediendo a la planificación del próximo Sprint. Este ciclo continuará hasta que se hayan desarrollado todas las historias de usuario planificadas.

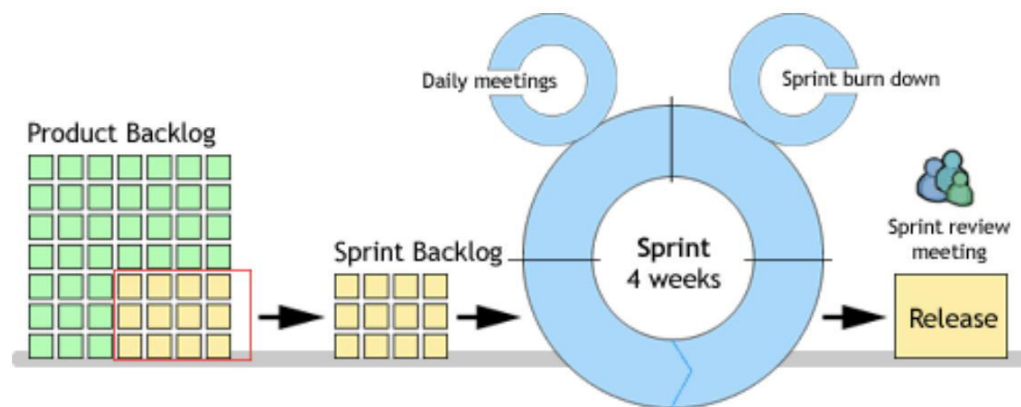


Figura 4. Ciclo de vida del Scrum

Fuente: (PEREYRA, 2010)

3.2.4.2. Big Data

El uso generalizado de tecnologías novedosas ha dado lugar a la generación de enormes volúmenes de datos. Por ejemplo, los trabajos de MapReduce (MR) se ejecutan continuamente en los clústeres de Google, procesando más de 20 petabytes de datos por día y Facebook genera más de 10 petabytes de datos mensualmente. (Bala, Boussaid, & Alimazighi, 2017)

Dichos datos se vuelven tan grandes y, por lo tanto, será difícil seguir utilizando las herramientas de procesamiento tradicionales. Big data se caracteriza por las cuatro "Vs" donde "Volumen" se refiere a la gran cantidad de datos que van más allá de las unidades habituales, "Velocidad"

a la velocidad con la que se generan y procesan estos datos, "Variedad" a la diversidad de formatos y estructuras, y "Veracidad" a la precisión y confiabilidad de los datos. (Bala, Boussaid, & Alimazighi, 2017)



Figura 5. Big Data 4Vs

Fuente: (QASource Blog, 2017)

3.2.4.3. ETL (Extract-Transform-Load)

El proceso ETL está dedicado a la actualización de sistemas de almacenamientos de datos (Data Warehouse). El proceso de extracción-transformación-carga trabaja con grandes volúmenes de datos que se procesan con fines de limpieza, filtrado y estandarización. (Bala, Boussaid, & Alimazighi, 2017)

El sistema ETL tradicional no puede manejar un gran volumen de datos (en la escala de terabytes y petabytes). Bajo ese contexto, se dispone de nuevas tecnologías como Cloud Computing, MapReduce (MR) y base de datos NoSQL para poder trabajar con la paralelización / distribución del procesamiento de datos en clústeres y abordar la complejidad de Big Data. (Bala, Boussaid, & Alimazighi, 2017)

Extracción de datos de origen relevantes y complicados de acceder, aquellos que se transformarán con fines de limpieza, estandarización y conformidad y luego se cargarán en un sistema de almacenamiento donde pueda persistir la data y proporcionar información relevante para aplicaciones de análisis y toma de decisiones. (Bala, Boussaid, & Alimazighi, 2017)

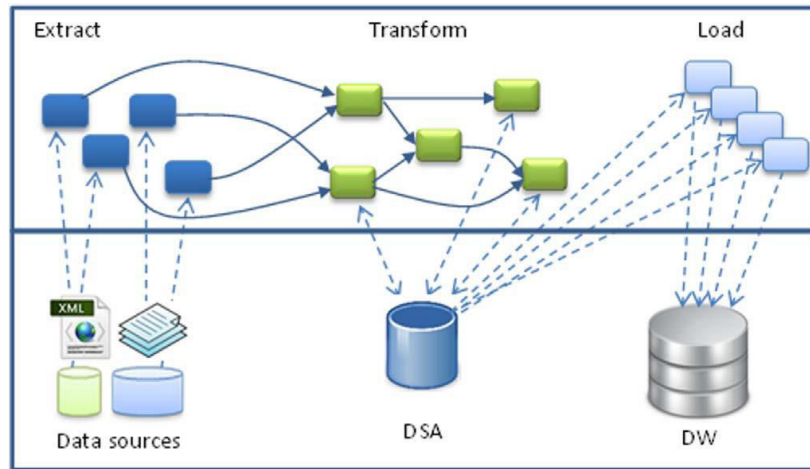


Figura 6. Proceso ETL

Fuente: (Bala, Boussaid, & Alimazighi, 2017)

3.2.4.4. Cloud Computing

Según (Microsoft, 2021), “Es proveer servicios informáticos a través de Internet, tales como servidores, servicios de almacenamiento, bases de datos, redes, software, Big Data, IoT (Internet de las cosas) e inteligencia artificial. La computación en la nube ofrece una innovación más rápida, recursos flexibles y economías de escala.”

En términos simples, se define a Cloud Computing como una tecnología mediante el cual se puede acceder remotamente a un pool de recursos y servicios que proveen a través de la conectividad a internet.

Dentro de toda la gama de recursos y servicios que ofrece uno de los proveedores de nube, en este proyecto he considerado a Microsoft Azure,

se detallarán aquellos servicios que son importantes resaltar para la comprensión de la solución del presente proyecto.

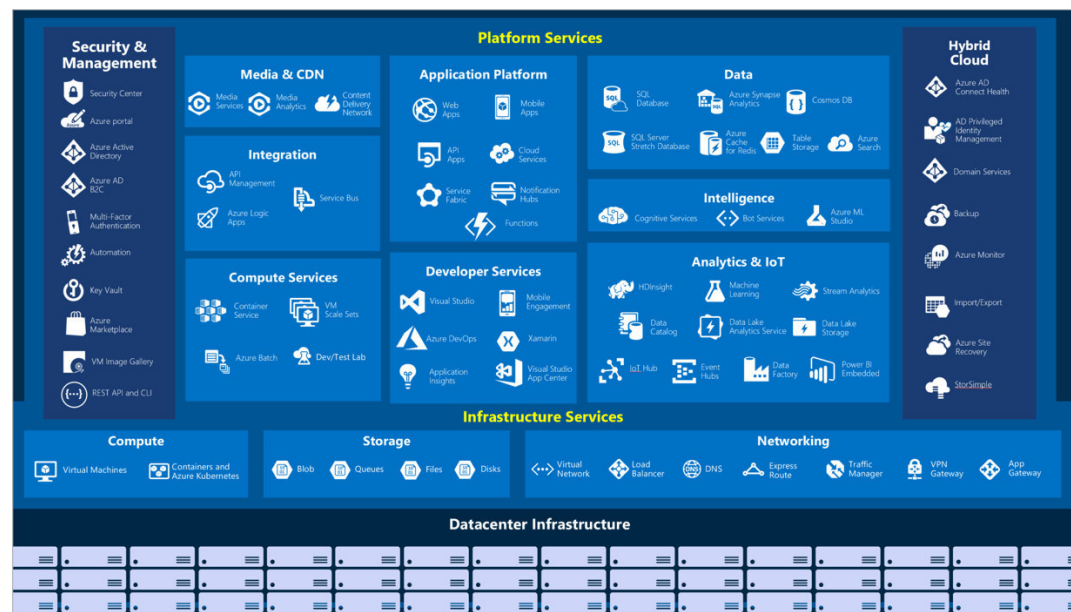


Figura 7. Azure Overview

Fuente: (Microsoft, 2021)

a) Azure Databricks

Azure Databricks es una plataforma de análisis de datos optimizada que proporciona las últimas versiones de Apache Spark y permite la integración con librerías de código abierto.

Se define Apache Spark como un motor de procesamiento de datos a gran escala. Básicamente, es un clúster Open Source de computación distribuida pensado para ejecutar consultas de análisis de datos.

Considero que Azure Databricks es un recurso de gran ayuda para el desarrollo de proyectos colaborativos dado que brinda un espacio de trabajo interactivo llamado notebook, el cual es útil para desarrollar procesos y prototipos de transformación y análisis de datos. En base a

mi experiencia, esta herramienta fue de mucha utilidad al momento de realizar el debug de algunas clases y funciones principales de mi proyecto, lo interesante es que se puede elegir el lenguaje de programación que desees, Scala, Python, SQL, etc.

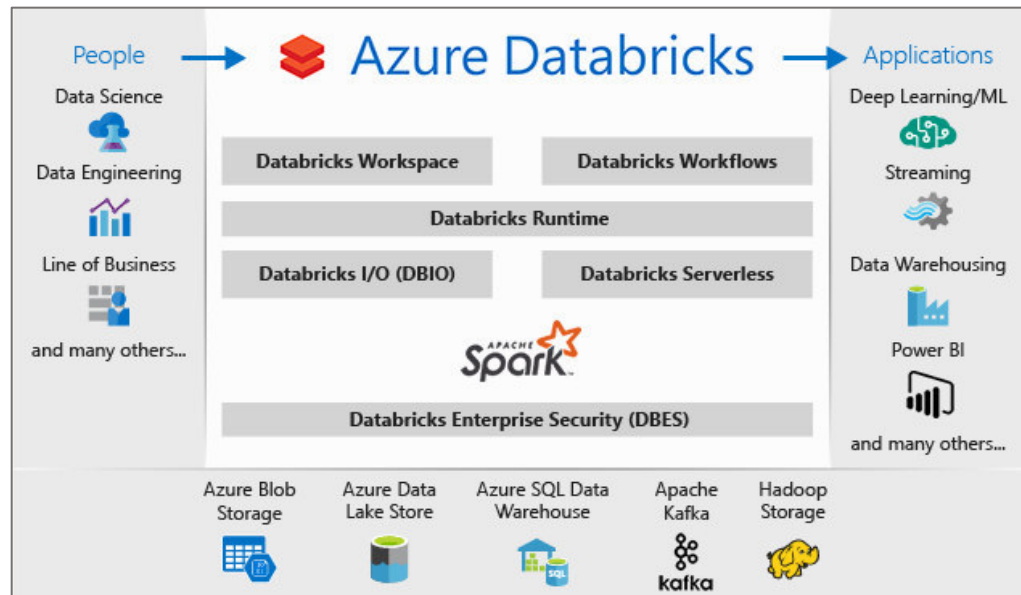


Figura 8. Ecosistema de Databricks en Microsoft Azure

Fuente: (Fernández, 2021)

b) Azure Cosmos DB

Azure Cosmos DB es un servicio de base de datos NoSQL, es decir no relacional. Asimismo, es escalable y permite la distribución y replicación de datos de manera global sobre todas las regiones que maneja Azure. (Microsoft Azure, 2021)

Con respecto a la disponibilidad del servicio es 99,999 % garantizado, respaldado con acuerdos de nivel de servicio, escalabilidad automática e inmediata.

Alguna de las características principales:

- Velocidad a gran escala mediante operaciones de expansión, con una elasticidad inmediata e ilimitada, por ello, tanto la lectura como escrituras son rápidas y se replican en cualquier parte del mundo.
- Desarrollo de aplicaciones rápido y flexible con SDK para los lenguajes más populares, una API Core (SQL) nativa, API para MongoDB, Cassandra, Graph y Gremlin.
- Base de datos sin servidor totalmente administrada y rentable con escalabilidad automática instantánea que responde a las necesidades de la aplicación.

Cabe mencionar que Cosmos DB trabaja con un modelo no relacional en el cual se trabaja con datos semiestructurados, justamente para el presente proyecto, se realiza el almacenamiento de la información crediticia de clientes en formato Json, que es un tipo de dato semiestructurado.

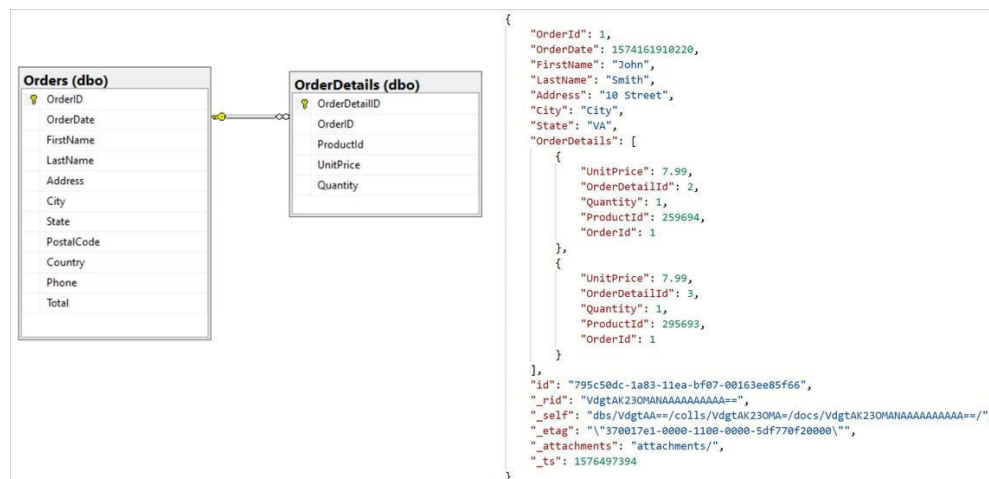


Figura 9. Modelo NoSQL

Fuente: (Microsoft Azure, 2021)

Otro concepto importante que se debe resaltar es RU (Request Units o unidades de solicitud) indica el nivel de performance de los recursos computacionales como el CPU, IOPS (operaciones de lectura y escritura) y memoria, necesarios para realizar las operaciones de base de datos compatibles con Azure Cosmos DB.

c) Azure Data Factory

Es un servicio de transformación e integración de datos que permite fácilmente la creación de procesos ETL y ELT sin necesidad de desarrollar algún código de programación en un entorno intuitivo. Es totalmente administrado, contiene más de 90 conectores integrados que no necesitan mantenimiento, y sin costo adicional. (Microsoft Azure, 2021)



Figura 10. Funcionalidades del Data Factory

Fuente: (Microsoft Azure, 2021)

Para no tener que desplegar un clúster en Databricks, el recurso Azure Data Factory ofrece un servicio de integración de datos y ETL basado

en la nube que permite crear flujos de trabajo orientados movimiento o transferencia de datos y la transformación de estos a escala.

Con Azure Data Factory, puede crear y programar flujos de trabajo mediante Pipelines que pueden ingerir datos de distintos almacenes de datos.

Para el uso de este recurso, se definen los conectores según la necesidad. Con relación al uso que le he dado específicamente para mi proyecto, es el copiado de información de un ambiente Onpremise hacia otro recurso llamado Azure Data Lake Storage, y también la ejecución de un JAR, con el fin de orquestar la aplicación de carga que es desplegada en un clúster Databricks.

d) Azure Data Lake Storage

De acuerdo a la documentación de (Microsoft Azure, 2021), “Azure Data Lake Storage Gen2 proporciona la semántica del sistema de archivos, la seguridad de nivel de archivo y la escala. Asimismo, ofrece un almacenamiento por niveles de bajo costo, con funcionalidades de alta disponibilidad y recuperación ante desastres”.

A continuación, se describen las características:

- Acceso compatible con Hadoop: Data Lake Storage Gen2 permite administrar y obtener acceso a los datos igual al sistema de archivos distribuido de Hadoop (HDFS).

- Seguridad: Data Lake Storage Gen2 es compatible con las listas de control de acceso (ACL), es decir, permite establecer los permisos de los datos almacenados en un nivel de directorio o de archivo.
- Rendimiento: Azure Data Lake Storage organiza los datos almacenados en una jerarquía de directorios y subdirectorios muy similar a un sistema de archivos para facilitar la navegación. Como resultado, el procesamiento de datos requiere menos recursos informáticos, lo que reduce el tiempo y el costo.
- Rentabilidad: Data Lake Storage Gen2 ofrece transacciones y capacidad de almacenamiento a bajo costo.
- Redundancia de datos: Data Lake Storage Gen2 proporciona redundancia de datos en un solo centro de datos con almacenamiento con redundancia local (LRS), o en una región secundaria mediante la opción de almacenamiento con redundancia geográfica (GRS). Esta característica garantiza que los datos estén siempre disponibles y protegidos si se produce una catástrofe.

e) Azure Key Vault

Azure Key Vault es el servicio de la nube de Microsoft que sirve para proteger y almacenar claves criptográficas y secretos que utilizan aplicaciones y servicios en la nube.

En mi opinión Azure Key Vault nos ayuda a cerrar brechas de seguridad al momento de desarrollar, gestionar y configurar aplicaciones en la nube.

Alguna de las funciones que ofrece este recurso:

- Gestión de secretos y monitoreo del acceso seguro de contraseñas, certificados y tokens de manera segura y confiable.
- Gestión de claves criptográficas para el cifrado de los datos que las aplicaciones y servicios en la nube manejan.
- Gestión de certificados TLS y SSL, tanto públicos como privados. Para ser utilizados de una manera segura entre Azure y los recursos contratados en tu suscripción.

3.2.5. Implementación de las áreas, procesos, sistemas y buenas practicas

En este punto se describen las actividades desarrolladas para la implementación de una solución Big Data que permitirá la carga de información crediticia de clientes mediante metodologías y tecnologías presentadas anteriormente.

A continuación, se detalla cada una de las fases correspondientes, que permitieron construir una aplicación eficiente y sostenible en el tiempo.

3.2.5.1. Fase 1: Planificación

Al culminar la ceremonia Big Room Planning, se definieron las historias de usuarios que se desarrollaran a lo largo del Q (trimestre), una de ellas consiste en el desarrollo de una aplicación cloud para la carga de información crediticia de clientes en una base de datos no relacional para que sea disponibilizada y consumida por el área comercial de la entidad financiera.

Para poder mapear y listar las actividades a realizar para implementar este proyecto, se tomó como input el alcance funcional recopilado por el Product Owner, de presentarse dudas, es en esta etapa donde nos reunimos con el usuario para llegar a acuerdos finales para el cierre del alcance del proyecto mediante un acta de reunión.

En esta etapa se planificaron y estimaron las actividades a realizar para la implementación. Así como el registro de dicha historia de usuario en la herramienta Jira⁷.



Figura 11. Registro de una épica en Jira

Fuente: Intranet de entidad financiera

⁷ Jira: ver el apartado GLOSARIO para más detalle.

3.2.5.2. Fase 2: Análisis y Diseño

a) Análisis Funcional

Tabla 7

Análisis funcional

Proceso de Negocio	Funcionalidad	Funciones Principales
Proveer información de la SBS (reporte de deudas) de los productos financieros de los clientes (Personas Naturales) para ser consumido a través de un API para el uso del área comercial.	Proveer información de clientes mediante el desarrollo de una aplicación cloud que permita la carga de datos a una base de datos no relacional.	• Desarrollo de un proceso que transmite los archivos orígenes con información de los clientes (archivos .txt) desde un ambiente Onpremise a la nube (Azure Data Lake Storage). • Desarrollo de una solución Big Data Batch en lenguaje Scala y Azure Databricks que permita la carga de los archivos orígenes a una base de datos no relacional como CosmosDB.

Nota: Elaboración propia

La propuesta de solución es revisada en conjunto con el líder técnico y de ser muy compleja, se programan reuniones con el área de arquitectura para llegar a un acuerdo de la definición de la solución siguiendo los lineamientos y buenas prácticas. De igual forma, para asegurar que la arquitectura de solución no presente riesgos de seguridad o vulnerabilidad, se requiere la conformidad del analista de seguridad. Asimismo, se elabora un documento de análisis y diseño en donde se especifica el alcance funcional y técnico del proyecto, tanto el Product Owner como el Líder Técnico deben brindar su conformidad.

b) Análisis Técnico

Tabla 8

Análisis técnico

Funciones Principales	Funcionalidades
Copiado de archivos orígenes con información de los clientes al Azure Data Lake Storage.	• Desarrollo de un pipeline en Azure Data Factory que realiza el proceso de upload de archivos orígenes con información de los clientes (archivos .txt) desde un servidor Onpremise hacia Azure Data Lake Storage. • Desarrollo de un script Powershell para la invocación del pipeline que realiza el copiado de los archivos orígenes. • Desarrollo de un archivo BAT que ejecutara el script Powershell.

	• Desarrollo de un Job que se ejecuta desde una plataforma mainframe, este se conectara al servidor Onpremise y ejecutara el archivo BAT.
Desarrollo de una aplicación Databricks que cargue los archivos orígenes a una base de datos Azure Cosmos DB.	• Desarrollo de una solución Big Data Batch en lenguaje Scala que realice transformaciones en ciertos campos del archivo origen obtenido desde Azure Data Lake Storage, el cual va a ser modelado en capas de stage (bronze, silver y golden). • Desarrollo de lógica de encriptación para aquellos campos con información sensible de los clientes. • Desarrollo de ingesta de la capa golden en donde la data se encuentra totalmente limpia y lista para cargarse a la base de datos no relacional como CosmosDB. • Desarrollar de lógica de validación de los registros del archivo origen, de no cumplir con el formato establecido en la documentación de mapeo de datos, estos registros se cargarán en una tabla delta que se almacenara en el Azure Data Lake Storage en formato parquet. • Desarrollo de un pipeline en Azure Data Factory que ejecuta el JAR desplegado en Azure Databricks para la carga a CosmosDB.

	• Desarrollo de un script Powershell para la invocación del pipeline que realiza la carga a CosmosDB. • Desarrollo de un archivo BAT que ejecutara el script Powershell. • Desarrollo de un Job que se ejecuta desde una plataforma mainframe, este se conectara al servidor Onpremise y ejecutara el archivo BAT.
--	--

Nota: Elaboración propia

c) Descripción del proceso

- El equipo encargado en la generación del archivo origen para ejecutar el proceso de ingesta, con la información de los clientes, así como el archivo de control, serán disponibilizados en una ruta compartida de un servidor Onpremise.
 - H_DEUDORSBSAGG.txt
 - H_DEUDORSBSAGG_CTRL.txt
- Mediante el siguiente proceso (Job ejecutado desde un scheduler en el ambiente Mainframe). Este proceso tiene como objetivo realizar el copiado de los archivos desde el servidor Onpremise al Azure Data Lake Storage.
 - @P4SHT10: Ejecuta el archivo SHCL_UPLOAD_RCC.BAT ubicado en el servidor Onpremise, el cual ejecuta el script Powershell SHCL_UPLOAD_RCC.ps1, este script al ejecutarse establece conectividad al recurso Data Factory en la suscripción de Azure, luego ejecuta el pipeline que realiza el copiado de archivos

H_DEUDORSBSAGG.txt y H_DEUDORSBSAGG_CTRL.txt del servidor de Alexandria al Azure Data Lake Storage.

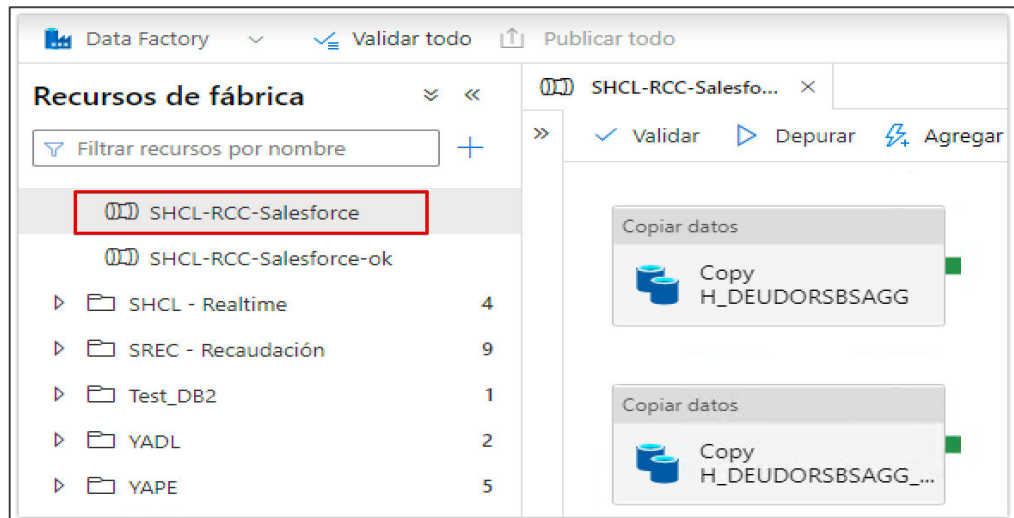


Figura 12. Pipeline del Data Factory

Fuente: Intranet de entidad financiera

A continuación, se muestra la ruta destino, donde son copiados los archivos, dentro del recurso Azure Data Lake Storage.

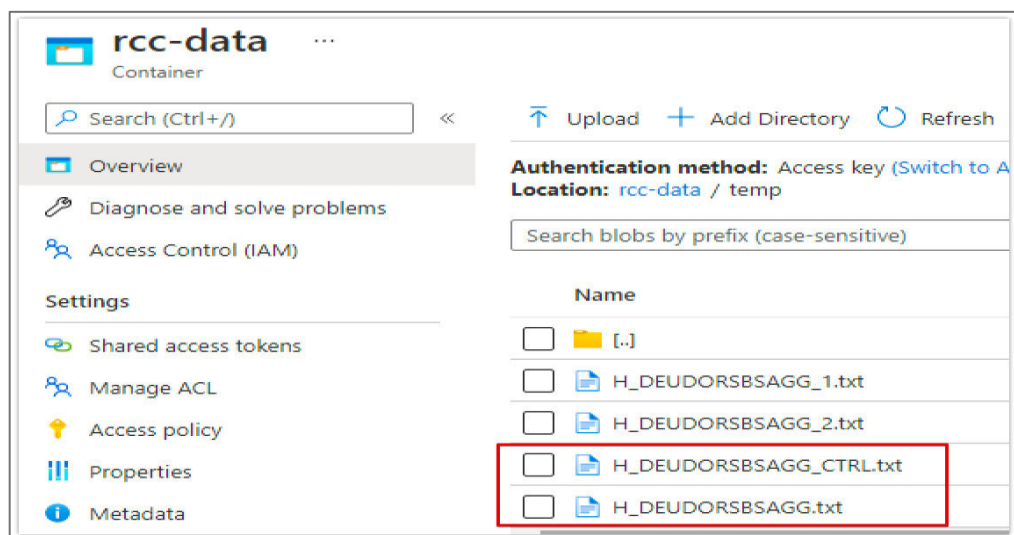


Figura 13. Ruta en Azure Data Lake Storage

Fuente: Intranet de entidad financiera

- Los datos serán transformados y validados mediante capas bronze, silver y golden, se generarán tablas delta que serán ubicados en la ruta dentro del Azure Data Lake Storage.

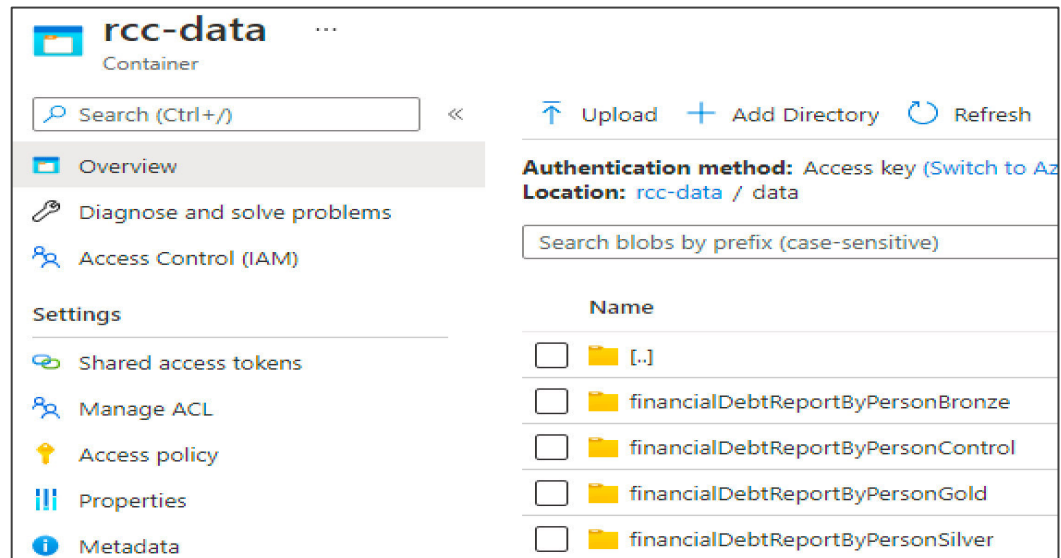


Figura 14. Delta Tables Bronze, Silver, Golden

Fuente: Intranet de entidad financiera

- El proceso de carga validara la data de todos los campos del archivo origen y de no estar correctos tal como se ha acordado con el usuario, serán registrados en una tabla de registros rechazados también ubicado en una ruta dentro del recurso Azure Data Lake Storage.
- Ejecución de proceso de ingesta.
 - @P4SH011: Ejecuta el SHCL_ETL_RCC.BAT ubicado en el servidor Onpremise, el cual ejecuta el script Powershell SHCL_ETL_RCC.ps1, este script al ejecutarse establece conectividad al recurso Data Factory en la suscripción de Azure, luego ejecuta el pipeline que realiza la carga de datos a las capas bronze, silver y golden.

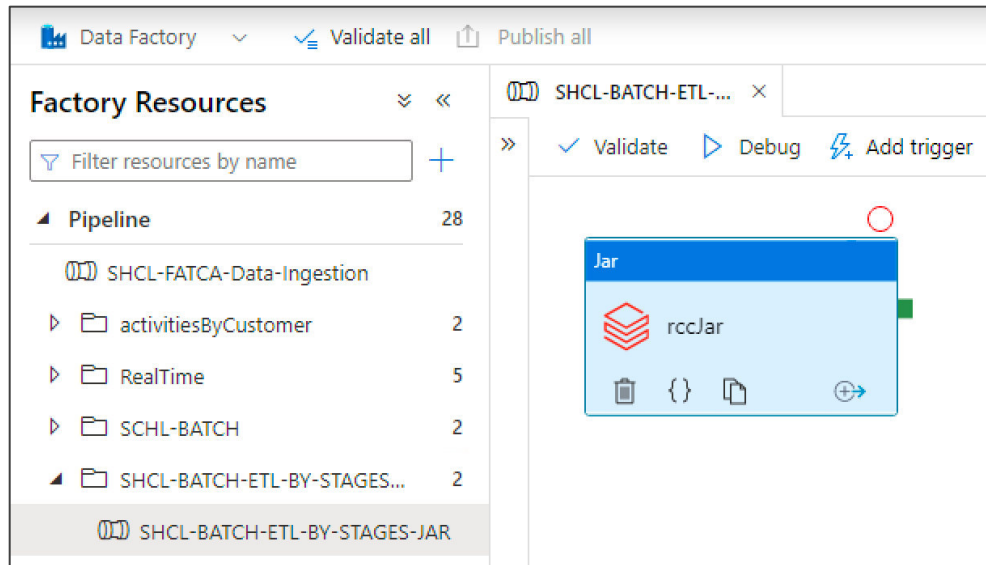


Figura 15. Pipeline que ejecuta JAR de aplicación

Fuente: Intranet de entidad financiera

- @P4SH012: Ejecuta el SHCL_ETL_RCC_UPDATERUMAX.BAT ubicado en el servidor Onpremise, el cual ejecuta el script Powershell SHCL_ETL_RCC_UPDATERUMAX.ps1, proceso que realiza la subida de RUs ⁸necesarios para realizar la carga en CosmosDB.
- @P4SH013: Ejecuta el SHCL_ETL_RCC_COSMOS.BAT ubicado en el servidor Onpremise, el cual ejecuta el script Powershell SHCL_ETL_RCC_COSMOS.ps1, proceso de carga del stage golden a CosmosDB.
- @P4SH014: Ejecuta el SHCL_ETL_RCC_UPDATERUMIN.BAT ubicado en el servidor Onpremise, el cual ejecuta el script Powershell SHCL_ETL_RCC_UPDATERUMIN.ps1, proceso que realiza la bajada de RUs.

⁸ RU: ver el apartado GLOSARIO para más detalle.

- En caso se produzca algún error durante el procesamiento entre las capas (bronze, silver, golden) o incluso en la ingesta a CosmosDB, se desarrollaron lógicas de reprocesos mediante los siguientes Jobs:
 - @P8SH033 → Ejecuta el SHCL_ETL_RCC_SILVER.BAT, proceso que realiza el reproceso de la carga a la capa silver y golden.
 - @P8SH034 → Ejecuta el SHCL_ETL_RCC_GOLD.BAT, proceso que realiza la carga a la capa golden.
- La carga de datos se realiza de manera mensual, la data está versionada por corte mensual del estado financiero de cada persona con código SBS.
- El proceso cargara alrededor de 60 millones de registros provenientes del archivo origen.

d) Diagrama de la malla automática

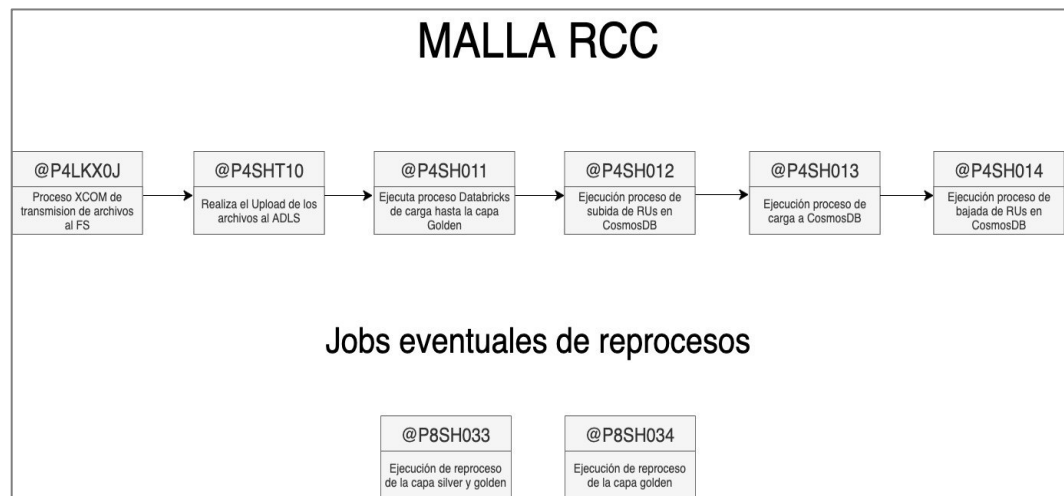


Figura 16. Diagrama de la malla automática

Fuente: Intranet de entidad financiera

e) Descripción de los archivos orígenes

- H_DEUDORSBSAGG.TXT

Tabla 9

Estructura de archivo plano origen de la información RCC

ID	Nombre	Descripción	Longitud	Tipo de dato fuente	¿Histórico?
1	CODSBS	Código SBS.	10	Texto	NO
2	CODCLAVEPARTYEMPSTESMAFINANCIERO	Código de la empresa del sistema financiero en la cual el cliente posee una deuda.	30	Texto	NO
3	CODSUBFAMILIABPSBS	Código Subgrupo Familia BP SBS.	2	Número	NO
4	CODIDENTIFICACIONPARTY	Identificador de la persona construido uniendo el número de documento de identidad y el tipo de documento.	30	Texto	NO
5	TIPPARTYIDENTIFICACION	Código del tipo de documento de identidad. '1'= Libreta electoral C DNI, '2'= Carné de identidad, '3'= Carné de extranjería, '4'= Pasaporte, '5'= Libreta tributaria, '6'= Registro único contribuyente RUC, '7'= Identificador ficticio.	20	Texto	NO
6	MTODEUDAMNSOL	Monto de la Deuda que se mantiene en Moneda Nacional expresada en Nuevos Soles	37	Número	NO
7	MTODEUDAMESOL	Monto de la Deuda que se mantiene en Moneda Extranjera expresada en Nuevos Soles	37	Número	NO
8	MINCONDICIONDIAMOROSIDAD	Mínimo de días de mora durante el mes reportado.	5	Número	NO

8	MINCONDICIONDIAMOROSIDAD	Mínimo de días de mora durante el mes reportado a nivel de entidad financiera, tipo de familia SBS, grupo de situación de crédito, familia SBS y subtipo de Familia SBS.	5	Número	NO
9	MAXCONDICIONDIAMOROSIDAD	Máximo de días de mora durante el mes reportado.	5	Número	NO
10	CODAGRUPACIONSITUACIONCREDISBS	Códigos SBS de agrupación de la situación de crédito.	2	Texto	NO
11	DESAGRUPACIONSITUACIONCREDISBS	Descripción del grupo de situación de crédito de la SBS.	60	Texto	NO
12	TIPFAMILIABPSBS	Tipo Familia BP SBS	2	Número	NO
13	DESTIPFAMILIABPSBS	Descripción tipo familia SBS	120	Texto	NO
14	CODFAMILIABPSBS	Código familia SBS	2	Número	NO
15	DEFAMILIABPSBS	Descripción Familia SBS	120	Texto	NO
16	DESSUBFAMILIABPSBS	Descripción Subfamilia SBS	120	Texto	NO
17	DESEMPISTEMAFINANCIERODEUDA	Nombre de la Empresa del Sistema Financiero donde la persona tiene deuda	120	Texto	NO
18	DEMPISTEMAFINANCIEROREPORTADODEUDORSBS	Número de entidades donde la persona tiene deuda.	2	Número	NO
19	PCTDEUDACLASIFRIESGONORMAL	Porcentaje de deuda con calificación normal	30	Número	NO
20	PCTDEUDACLASIFRIESGOCPP	Porcentaje de deuda con calificación de problema potencial.	30	Número	NO
21	PCTDEUDACLASIFRIESGODEFICIENTE	Porcentaje de deuda con calificación de deficiente	30	Número	NO
22	PCTDEUDACLASIFRIESGODUDOSO	Porcentaje de deuda con calificación dudoso.	30	Número	NO
23	PCTDEUDACLASIFRIESGOPERDIDA	Porcentaje de deuda con calificación de perdida.	30	Número	NO
24	FECTUALIZACIONTABLA	Fecha de proceso, o Identifica la fecha de actualización de la tabla.	20	Fecha Hora	NO
25	FECDIA	Identifica la fecha a la que pertenece la información.	10	Fecha	NO

Nota: Elaboración propia

Muestra:

H_DEUDORSBSAGG - Notepad							
File	Edit	Format	View	Help			
000007215500061	10000000072155	1	154.81	0.00	0	0	1 Vigente
000007647900366	19000000076479	1	0.00	1095317.73	9959	9959	5 Vencido
000007647900366	43000000076479	1	40001.00	3142365.26	9959	9959	. Sin agrupación
000009063000006	11000000090630	1	0.00	2146.94	0	0	1 Vigente
000011091000006	46000000110910	1	0.00	122270.40	0	0	. Sin agrupación
000011091000145	91000000110910	1	7785.36	0.00	0	0	. Sin agrupación
000011252600002	0 000000112526	1	3411.42	0.00	0	0	. Sin agrupación
000011252600002	84000000112526	1	14.94	0.00	0	0	. Sin agrupación
000011252600073	0 000000112526	1	1500.00	0.00	82	82	. Sin agrupación
000011295900142	91000000112959	1	1200.00	0.00	0	0	. Sin agrupación
000013087600480	30000000130876	1	1380.89	5705.51	8114	8818	. Sin agrupación
000017602000072	30000000176020	1	9459.43	0.00	2838	2838	. Sin agrupación

Figura 17. Archivo plano con información RCC

Fuente: Intranet de entidad financiera

- H_DEUDORSBSAGG_CTRLTXT

Tabla 10

Estructura del archivo de control

ID	Nombre	Descripción	Longitud	Tipo de dato fuente
1	TOTALINGRESOS	Total de registros	15	Número
2	FECHAPROCESO	Fecha de carga de archivo H_DEUDORSBSAGG	8	Texto
3	HORAPROCESO	Hora de carga del archivo H_DEUDORSBSAGG	8	Texto

Nota: Elaboración propia

Muestra:

temp/H_DEUDORSBSAGG_CTRL.txt ...	
Blob	
Save Discard Download Refresh Delete	
Overview Versions Edit Generate SAS	
1	000000000000322021012212:32:37

Figura 18. Archivo plano de control

Fuente: Intranet de entidad financiera

f) Diagrama del proceso técnico

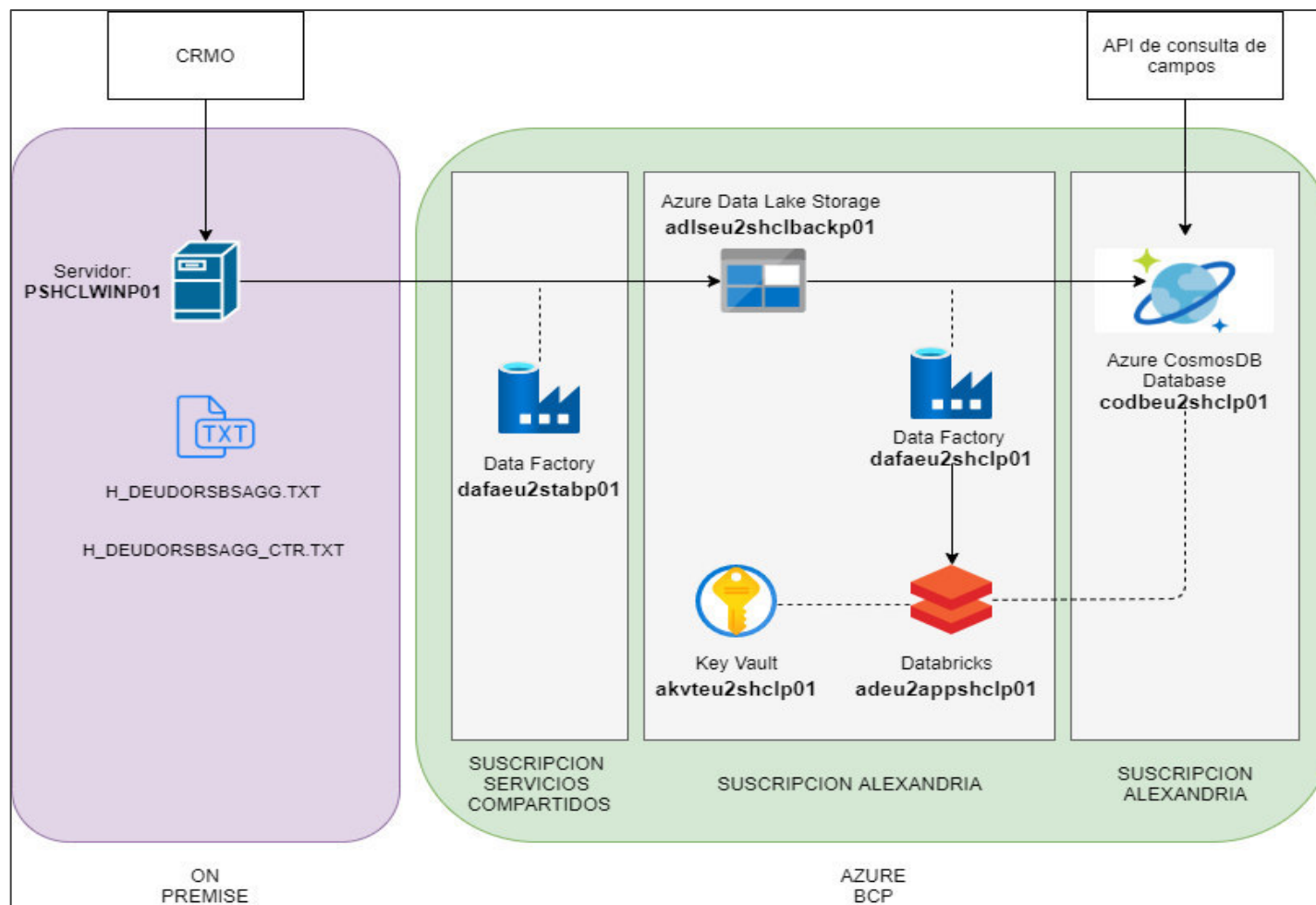


Figura 19. Arquitectura de Solución

Fuente: Elaboración propia

Fase 3: Desarrollo de aplicación

A continuación, se describe el desarrollo del aplicativo y los entregables que son evidencia de la implementación de las áreas de procesos del proyecto.

Los artefactos elaborados se encuentran en la sección Anexos.

Tabla 11

Actividades de desarrollo

Épica	Historia de Usuario	Actividades	Entregables
E.6. Implementar un modelo ingesta de datos en cosmos DB de RCC (Reporte Crediticio de Clientes)	E.6.1 Desarrollar aplicación Databricks de ingesta	• Carga de archivos planos por posición. • Desarrollo de encriptación de campos sensibles y carga a la capa bronze. • Desarrollo de proceso para guardado en silver (validaciones, filtros, limpieza de data). • Modelamiento para CosmosDB. • Desarrollo de proceso para guardado en CosmosDB (Golden) • Pruebas integrales end-to-end (cargas bronze, silver, golden y tablas logs)	Revisar el apartado ANEXO 1 para mayor detalle.
	E.6.2 Crear pipelines en Data Factory	• Creación de pipeline para la ejecución de la aplicación Databricks de ingesta. • Creación de pipeline para la ejecución del proceso de copiado de los archivos orígenes	Revisar el apartado ANEXO 2 y 3 para mayor detalle.

		desde un ambiente Onpremise al Azure Data Lake Storage.	
	E.6.3 Crear archivos ejecutables para la invocación del flujo de la aplicación	• Desarrollo de scripts PowerShell para la invocación de los pipelines creados en el recurso Data Factory. • Desarrollo de un archivo BAT que ejecutara el script PowerShell.	Revisar el apartado ANEXO 4 y 5 para mayor detalle.
	E.6.4 Crear malla automática mediante Jobs mainframe de ejecución	• Desarrollo de un Job que se ejecuta desde una plataforma mainframe, este se conectara al servidor Onpremise y ejecutara el archivo BAT.	Revisar el apartado ANEXO 6 para mayor detalle.

Nota: Elaboración propia

3.2.5.3. Fase 3: Integración y pruebas

En este punto, luego de haber concluido con la construcción del aplicativo y pruebas unitarias de cada una de las historias de usuario desarrolladas, se procede con la integración de los componentes y realización de pruebas integrales.

- a) Ejecución de Jobs para el despliegue del clúster Databricks versión final mediante Jenkins⁹. En esta instancia, se realiza el despliegue para validar todo el proceso de carga de información a CosmosDB.

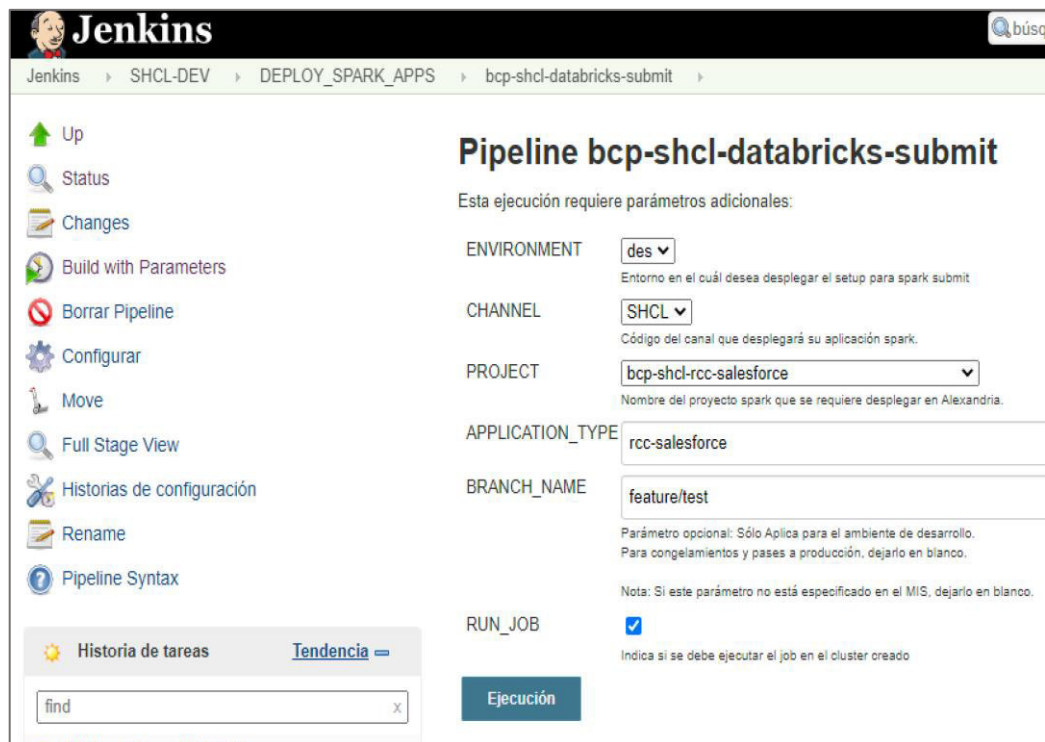


Figura 20. Job de despliegue en Jenkins

Fuente: Intranet de entidad financiera

⁹ Jenkins: ver el apartado GLOSARIO para más detalle.

- b) Ejecución de Jobs mainframe para probar el proceso end-to-end. En esta instancia se realiza la prueba integral de la aplicación, en cada uno de sus procesos. De salir todo conforme, se procede con la siguiente fase, que es documentar mediante un manual de instalación todas las actividades realizadas en el ambiente de desarrollo para lograr el correcto funcionamiento de la aplicación construida.

3.2.5.4. Fase 4: Preparación de ambiente

En esta fase, se tiene como objetivo realizar una serie de actividades para gestionar un pase de implementación de todos los cambios realizados en desarrollo al ambiente de Certificación, a esta gestión se le conoce como la fase congelamiento. A continuación, se detallará cada una de las actividades a realizar:

- a) BitBucket¹⁰: mediante esta herramienta, que permite la administración e integración del código de manera colaborativa entre los miembros del equipo de desarrollo, se trabajaron los cambios en un repositorio correspondiente al aplicativo, y bajo una rama. Una vez realiza la prueba integral exitosa, se solicita el Pull Request, para promover los cambios de la rama de trabajo a la rama develop, la cual se le considera como la rama oficial ya que es la versión que pasara a producción.

¹⁰ BitBucket: ver el apartado GLOSARIO para más detalle.

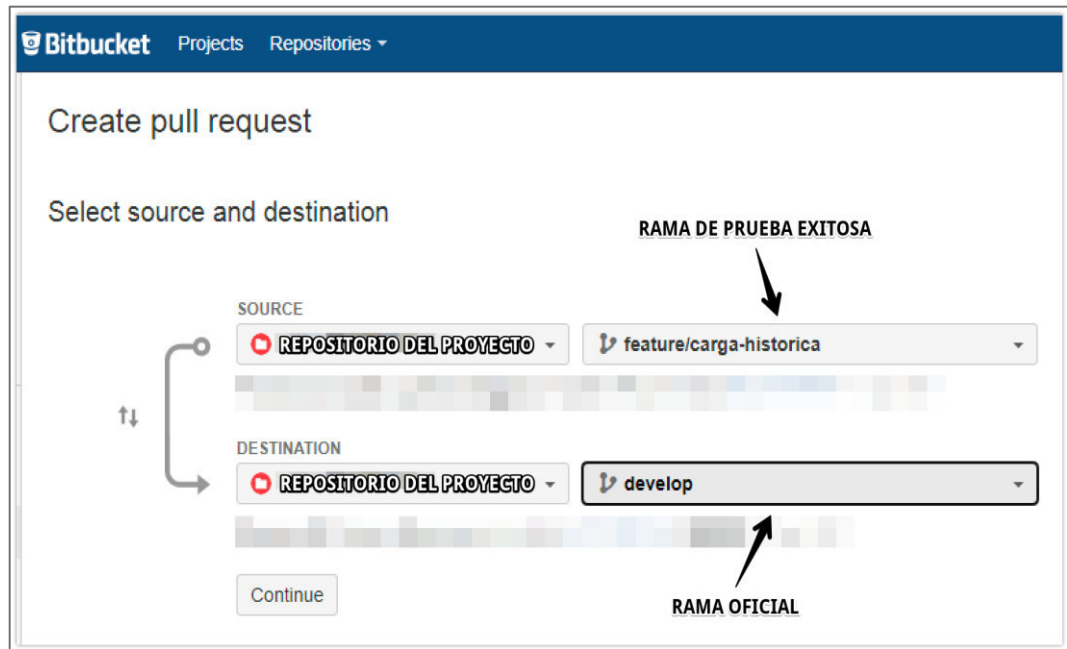


Figura 21. Creación de Pull Request en BitBucket

Fuente: Intranet de entidad financiera

- b) Jira: mediante esta herramienta, que permite la administración de tareas de un proyecto, así como también la gestión operativa de los proyectos. En este punto, se creará un documento en línea, en la empresa lo llamamos MVP¹¹, y alude a un manual de instalación que será implementado por un Agile OPS¹². Este manual o documento debe estar redactado al mínimo detalle, ya que el operador accionara en base al documento al pie de la letra. De tener alguna omisión y presentar ambigüedades, será rechazado, lo que conllevará reiniciar al punto inicial.

¹¹ MVP: ver el apartado GLOSARIO para más detalle.

¹² Agile OPS: ver el apartado GLOSARIO para más detalle.

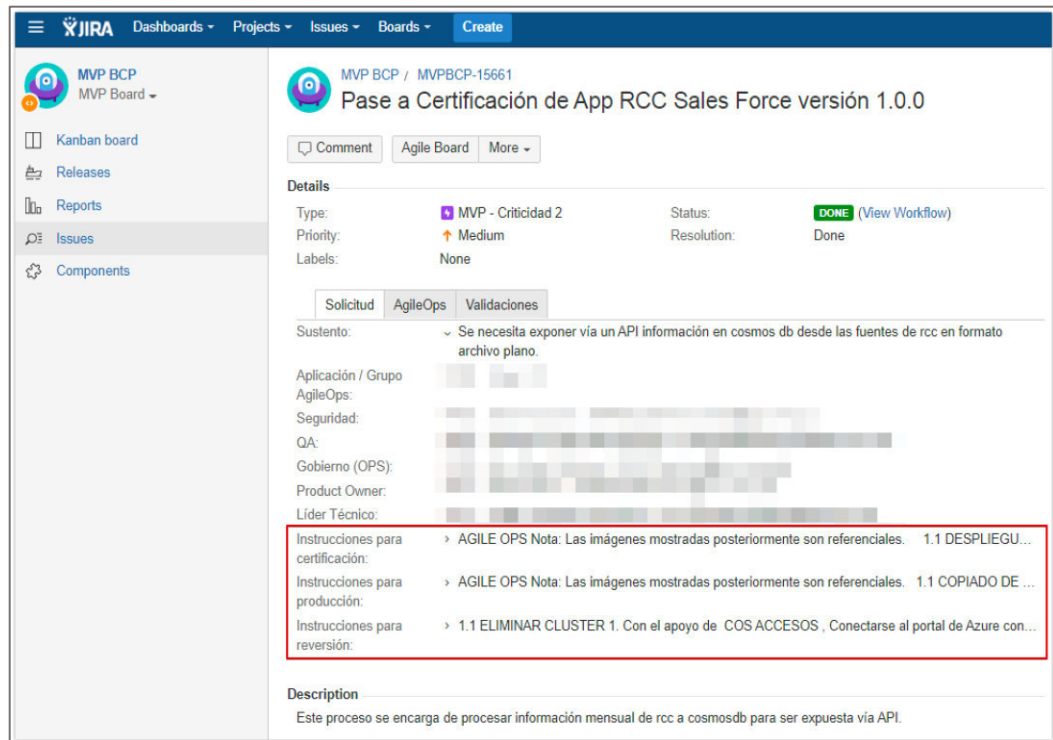


Figura 22. Creación de MVP

Fuente: Intranet de entidad financiera

El MVP tiene 3 secciones importantes tal como se visualiza en la figura 22, certificación, producción y reversión. Durante el congelamiento, el operador (Agile OPS) implementa solo la sección de certificación.

- c) Jira: durante la etapa de redacción y edición, el MVP se encuentra en el estadio “Registro de Solicitud”, una vez culminada la documentación del MVP, se procede promoviéndolo a los siguientes carriles hasta el estadio “Ejecutar Congelamiento”, de esa manera, el operador (Agile OPS) puede implementar el MVP en el ambiente de certificación.

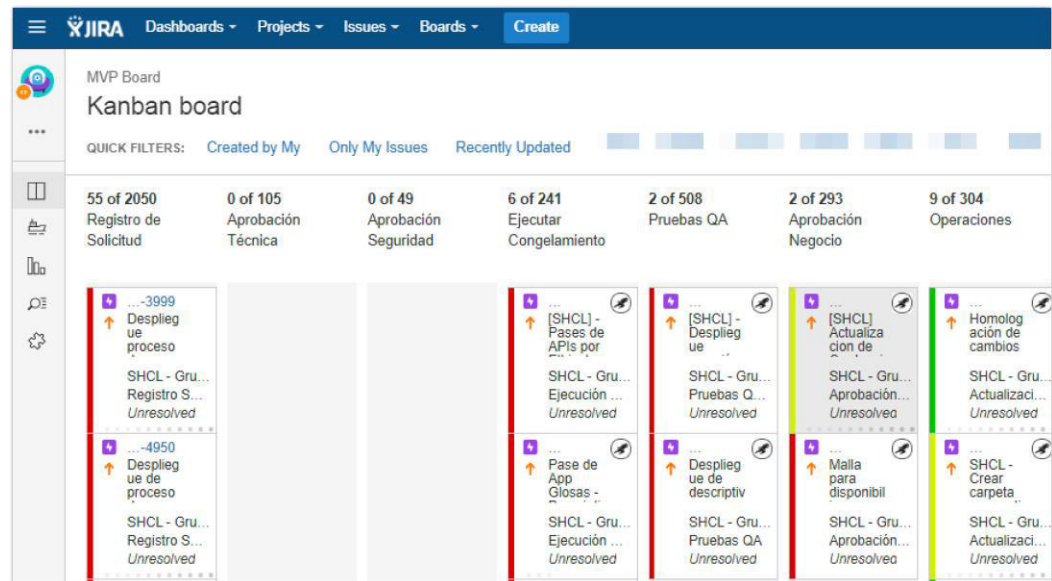


Figura 23. Estadios del MVP

Fuente: Intranet de entidad financiera

Para el presente proyecto se documentaron los siguientes MVPs:

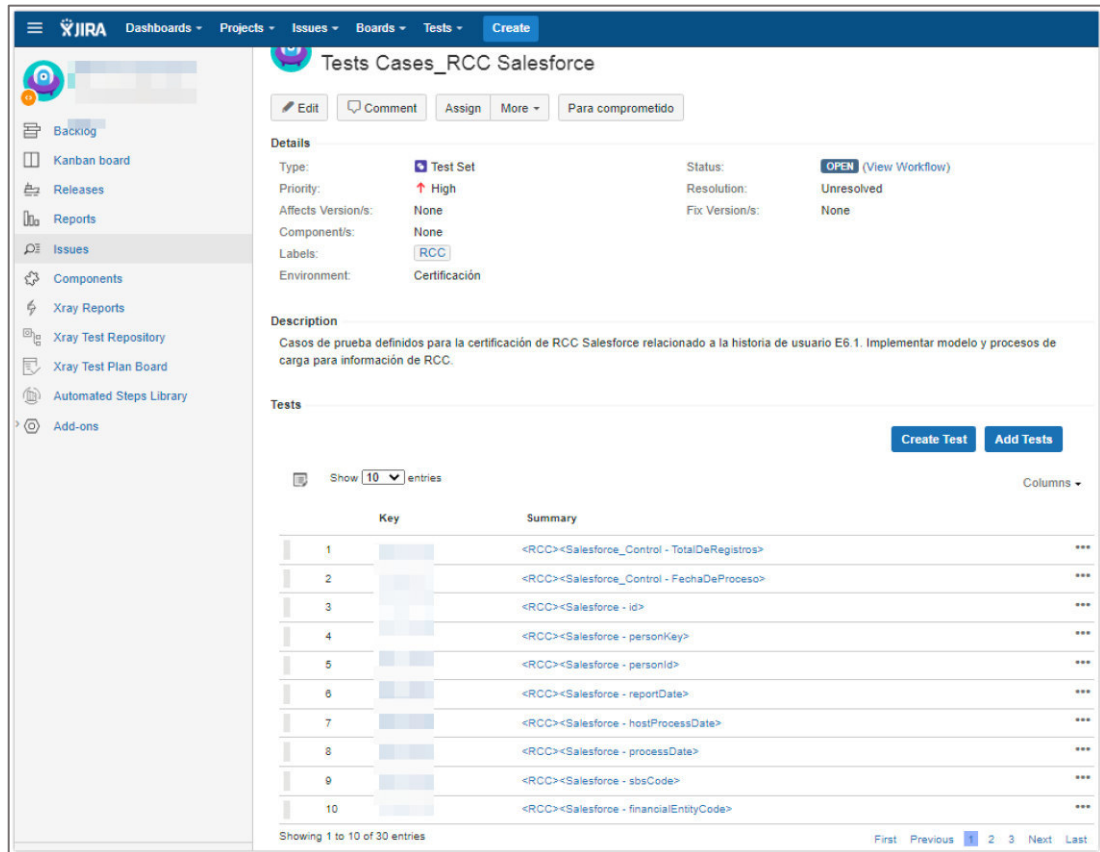
Tabla 12
MVPs desarrollados en el proyecto

Nombre del MVP	Código de MVP
Implementación de malla scheduler	MVP-13525
Despliegue de pipelines en el recurso de Data Factory	MVP-14184 MVP-13916
Despliegue de aplicación RCC Salesforce	MVP-15661

Nota: Elaboración propia

3.2.5.5. Fase 5: Pruebas en Certificación

En esta fase, el analista de calidad elaboró sus casos de pruebas, que son registrados en el Jira y posteriormente, se procedió a realizar la certificación.



The screenshot displays the Jira interface for a project named 'Tests Cases_RCC Salesforce'. The left sidebar shows navigation options like Backlog, Kanban board, Releases, Reports, Issues, Components, Xray Reports, Xray Test Repository, Xray Test Plan Board, Automated Steps Library, and Add-ons. The main content area shows the details of a Test Set. The 'Details' section includes fields for Type (Test Set), Priority (High), Status (OPEN), Resolution (Unresolved), and Environment (Certificación). The 'Description' section contains text about test cases for RCC Salesforce. Below the description is a table of test cases with columns 'Key' and 'Summary'. The table lists 10 test cases, each with a key and a summary. The 'Summary' column contains test case identifiers and their descriptions. The table is paginated, showing 1 to 10 of 30 entries.

Key	Summary
1	<RCC><Salesforce_Control - TotalDeRegistros> ***
2	<RCC><Salesforce_Control - FechaDeProceso> ***
3	<RCC><Salesforce - id> ***
4	<RCC><Salesforce - personKey> ***
5	<RCC><Salesforce - personId> ***
6	<RCC><Salesforce - reportDate> ***
7	<RCC><Salesforce - hostProcessDate> ***
8	<RCC><Salesforce - processDate> ***
9	<RCC><Salesforce - sbsCode> ***
10	<RCC><Salesforce - financiaEntityCode> ***

Figura 24. Casos de pruebas

Fuente: Intranet de entidad financiera

El autor del informe brindó el soporte a las pruebas realizadas en este proceso, una vez terminada las verificaciones y validaciones de esta etapa, el analista de calidad documenta y adjunta evidencias de todas las pruebas realizadas en certificación. En la tabla se muestra un resumen de las pruebas ejecutadas durante el proceso de certificación.

Tabla 13
Test Cases

Test Case	Actividad	Resultado esperado	Test Status
Archivo de Control <TotalDeRegistros> Transformación: “000000000011364” rellena en “0” al inicio. Tipo de dato: Número entero, cantidad total de registros grabados en el archivo.	Ingresar 15 caracteres numéricos válidos para la cantidad.	Validar la trazabilidad de la inserción de registros en la tabla log y la coincidencia con la cantidad de registros en la tabla financialDebtReportByPerson	PASS
	Ingresar 15 caracteres no válidos 000000000122b bb 000000000.123 65	Validar que se carguen los registros en la tabla financialDebtReportByPerson, dado que el archivo, ya que un error en el archivo de control no detiene el proceso de carga.	PASS
Archivo de Control <FechaDeProceso> Transformación: Aplicar formato de fecha. Tipo de dato: Alfanumérico, fecha del sistema en la se creó el archivo en el formato AAAAMMDD.	Ingresar 8 caracteres numéricos válidos para una fecha. 20210205	Validar la trazabilidad de la inserción de registros en la tabla log.	PASS
	Ingresar 8 caracteres no válidos para una fecha. bbbbbbbb 00000000 !"#\$\$%&/()	Validar que se carguen los registros en la tabla financialDebtReportByPerson, dado que el archivo, ya que un error en el archivo de control no detiene el proceso de carga.	PASS
Campo Id hashado	Realizar el proceso de carga de data RCC Salesforce.	Validar en la tabla financialDebtReportByPerson de Cosmos que cargue el campo id hashado.	PASS

Nota: Elaboración propia

3.2.5.6. Fase 6: Despliegue del aplicativo a Producción

Con el aplicativo probado en certificación, se procedió a realizar la gestión para llevar a cabo la implementación en la entidad financiera. En esta etapa se realizaron las siguientes actividades:

- a) Registrar el MVP en el calendario de pases a producción.
- b) Se elabora un Plan de Pase a Producción, donde se resumen los puntos más importantes a implementarse, la fecha y hora de cada actividad como el pase a producción, la ratificación, la reversión y la ratificación de la reversión.
- c) Se gestionan las conformidades tanto del Agile OPS como del OPS de gobierno para realizarse el pase a producción.
- d) Una vez registrado el MVP en el calendario de pases y aprobado por el Agile OPS y el OPS de gobierno, el líder técnico del Squad procede a promover el MVP al carril de “Ejecución de pase a Producción”, mediante la herramienta del Jira.
- e) El autor del proyecto además de realizar las gestiones del pase a producción realiza el acompañamiento durante el pase.
- f) Una vez culminada la ratificación total exitosa, el autor del proyecto prepara una capacitación para la transferencia al equipo de soporte sobre los nuevos cambios que se han implementado durante el pase, con el fin de documentar el troubleshooting¹³ en caso haya algún problema en el ambiente de producción al momento de realizar el monitoreo.

¹³ Troubleshooting: ver el apartado GLOSARIO para más detalle

3.3. Evaluación

3.3.1. Evaluación económica

La evaluación económica se basa en su mayoría de datos supuestos. A continuación, se detallarán los puntos a incluir:

- Gastos tecnológicos
- Gastos administrativos
- Flujo de caja
- VAN, TIR

3.3.1.1. Gastos tecnológicos

En la siguiente tabla se detallan tanto los costos mensuales como los anuales referente a los costos por el uso de tecnologías cloud en Azure.

Tabla 14
Gastos tecnológicos

Gastos tecnológicos		
Servicios de Azure	Costo mensual	Costo anual resultante
Azure Databricks	S/. 32	S/. 384
Azure Cosmos DB	S/. 4,000	S/. 48,000
Data Factory	S/. 25	S/. 300
Storage Account	S/. 520	S/. 6,240
Key Vault	S/. 8	S/. 96
Support	S/. 450	S/. 5,400
Total		S/. 60,420

Nota: Elaboración propia

Asimismo, se detalla los activos del proyecto.

Tabla 15

Activos

Activos				
Hardware	Cantidad	Valor de Adquisición	Valor útil	Depreciación
Servidor	1	S/. 80,000	5	S/. 20,000
Total		S/. 80,000		

Nota: Elaboración propia

3.3.1.2. Gastos administrativos

En la siguiente tabla se detallan los costos mensuales del proyecto, como la inversión en el personal requerido que participo durante todo el desarrollo del proyecto (4 meses), asimismo, se considera otro personal de trabajo como el analista de soporte, el cual es el encargado de velar por la estabilidad de todos los proyectos (aplicativos desarrollados) del Squad.

Tabla 16

Gastos administrativos

Gastos administrativos				
Capital Humano	Cantidad	Costo mensual	Meses del proyecto	Costo resultante
Product Owner	1	S/. 7,000	4	S/. 28,000
Ingeniero de Datos	1	S/. 8,000	4	S/. 32,000
Líder Técnico	1	S/. 6,000	4	S/. 24,000
Analista de Calidad	1	S/. 4,000	4	S/. 16,000
Analista de soporte	1	S/. 3,000	12	S/. 36,000
Otros				
Luz	1	S/. 800	4	S/. 3,200
Agua	1	S/. 400	4	S/. 1,600
Red	1	S/. 500	4	S/. 2,000
Mantenimiento del local	1	S/. 450	4	S/. 1,800
Total				S/. 144,600

Nota: Elaboración propia

3.3.1.3. Flujo de caja

En la tabla 16, se detalla el ahorro del costo de operación por la necesidad que tiene el área comercial de solicitar la información del cliente para una evaluación crediticia, este ahorro finalmente beneficia a la entidad financiera aportando más utilidades en una vida proyectada de 5 años.

Tabla 17

Ahorro de costos proyectado a 5 años

Año	1	2	3	4	5
Costo por operación	200	200	200	200	200
Cantidad (incremento 20%)	960	1,152	1,382	1,659	1,991

Nota: Elaboración propia

En la siguiente tabla se detalla el flujo de caja.

Tabla 18

Flujo de caja

AÑO	Inversión inicial	1	2	3	4	5
Ahorro en costos		192,000	230,400	276,480	331,776	398,131
Costo anual Azure		-60,420	-60,420	-60,420	-60,420	-60,420
Utilidad bruta		131,580	169,980	216,060	271,356	337,711
Gastos administrativos		-144,600	-144,600	-144,600	-144,600	-144,600
Depreciación		0	-16,000	-16,000	-16,000	-16,000
Utilidad antes de impuestos		-13,020	9,380	55,460	110,756	177,111
Impuestos		3,841	-2,767	-16,361	-32,673	-52,248
Utilidad después de impuestos		-9,179	6,613	39,099	78,083	124,863
Depreciación		0	16,000	16,000	16,000	16,000
Inversión bienes	-80,000					
FCN (Flujo caja neto)	-80,000	-9,179	22,613	55,099	94,083	140,863

Nota: Elaboración propia

3.3.1.4. Análisis del VAN y del TIR

El valor actual neto (VAN) es un indicador financiero para el análisis de rentabilidad de un proyecto de inversión. El VAN obtenido para el presente proyecto es de S/. 123,466.00. Este resultado indica que el proyecto tiene una rentabilidad positiva, es decir, se están generando utilidades en una vida proyectada de 5 años.

Por otro lado, el indicador complemento del VAN vendría a ser la tasa interna de retorno (TIR), la cual denota si el proyecto puede ser considerado como apto mediante un análisis y evaluación de la factibilidad del proyecto. En la tabla 18, se muestra que el TIR es mayor que la tasa de descuento, lo cual se demuestra que el proyecto es viable.

Tabla 19
VAN y TIR

Tasa de descuento	15%
VAN	S/. 123,466.00
TIR	38%

Nota: Elaboración propia

3.3.2. Evaluación de los beneficios

La implementación de la aplicación cloud para la carga de información crediticia de clientes para el consumo de un API de negocio que proporcione una única vista al área comercial de una entidad financiera, generando de esta manera los siguientes beneficios:

- Permitir al área de ventas consultar las deudas de los clientes en el sistema financiero (Información de RCC) de una manera rápida y con 0 costos por realizar esta operación.
- Permita una carga de información eficiente mediante tecnologías de Big Data y recursos computacionales escalables en el tiempo, con respaldo en todas las regiones del mundo y con precios muy asequibles y atractivos.
- Tenga cero costos de operación por la petición de información crediticia de los clientes, generando así, un ahorro potencial por los costos de petición de esta operación a las áreas Data Lake o Data Warehouse de la entidad financiera, por ello, se considera que el proyecto tiene una excelente relación de costo/beneficio.

CAPITULO IV

REFLEXION CRITICA DE LA EXPERIENCIA

4.1. Aportes/ lecciones aprendidas/ en el qué se puede mejorar

4.1.1. Aportes

La participación del autor del presente informe en la implementación de un aplicativo cloud de carga de información bajo tecnologías Big Data fue como el rol de Data Engineer. Los aportes fueron los siguientes:

- Definir el diseño y arquitectura de la solución basándose en experiencias previas, cumpliendo con los lineamientos y estándares de desarrollo y seguridad del banco, asimismo, que satisfaga la necesidad del cliente para este caso de uso.
- Participar en las reuniones con el Data Modeler¹⁴ para la definición del modelo de datos tanto del archivo plano que tiene la información crediticia de los clientes como de la tabla (collection) de la base de datos CosmosDB.
- Desarrollar las transformaciones y limpieza de la data en las diferentes capas para procesarla y almacenarla de manera óptima y eficiente.
- Realizar pruebas unitarias e integrales, asimismo la documentación técnica de análisis y diseño de solución.
- Realizar la documentación de las instrucciones para la instalación y despliegue de la aplicación en el ambiente de certificación.
- Dar soporte al proceso de certificación y subsanar observaciones impartidas por el analista de calidad.
- Dar soporte al pase y post-pase a producción de la aplicación.
- Capacitar al equipo de soporte del Squad sobre la aplicación y documentar una guía de operaciones para accionar ante cualquier posible incidente.

¹⁴ Data Modeler: ver el apartado GLOSARIO para más detalle.

4.1.2. Lecciones aprendidas

- La delimitación del alcance no se formalizó debidamente, lo que generó que el usuario solicite un cambio en el alcance, relacionado al archivo plano, el cual es el input para el procesamiento del aplicativo de carga.
- Dependencia con otro Squad que se encargaría de generar el archivo plano que requerimos para la carga, el cual presentó retrasos. Por lo que se tuvo que optar por crear data dummy¹⁵ para las pruebas iniciales unitarias.
- El punto anterior conllevó a tener un inconveniente al momento de realizar las pruebas integrales con el archivo original proporcionado por el Squad CRMO¹⁶, debido a que se encontraron observaciones por tema de calidad de datos en dicho archivo.
- Antes que inicie el desarrollo, se debió resolver dudas sobre el modelo de datos proporcionado por el Data Modeler en conjunto con el analista de calidad dado que hubo reglas de transformación de data que se entendió de manera diferente por el Data Engineer y el analista de calidad, esto conllevó a destinar más tiempo en las pruebas de certificación por tener que revisar el mapeo nuevamente y realizar el ajuste respectivo en el desarrollo.
- Considero como oportunidad de mejora, poder optimizar los Jobs de despliegue continuo en la herramienta Jenkins, dado que la memoria que se usa es compartida y esto origina que se crucen las ejecuciones por usar la misma instancia, generando un encolamiento y demora en los despliegues y pruebas unitarias durante la etapa de desarrollo.

¹⁵ Dummy: ver el apartado GLOSARIO para más detalle.

¹⁶ CRMO: ver el apartado GLOSARIO para más detalle.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Se logró realizar la definición diseño y arquitectura de la solución cumpliendo con los lineamientos y estándares de desarrollo y seguridad del banco, asimismo, que satisfaga la necesidad del cliente para este caso de uso.
- Se logró elaborar el documento de análisis y diseño de la solución, en el cual se describe toda la especificación del alcance técnico y funcional necesaria para el inicio del desarrollo.
- La ejecución de las pruebas unitarias e integrales tuvieron un resultado exitoso, sin embargo, por ambigüedades en el mapeo de datos se tuvo que revisar y redefinir algunas reglas de transformación en conjunto con el Data Modeler. Afortunadamente, se subsanaron las observaciones en un par de días y se pudo continuar con el flujo en la misma iteración comprometida para realizar las pruebas.
- Se logró desplegar la aplicación cloud de carga de información en el ambiente de producción de manera correcta, brindando un gran valor y satisfacción al usuario y a la entidad financiera debido a que las operaciones que realiza el área de ventas conllevan a un proceso engorroso y costoso, lo que este aplicativo está brindando es un ahorro monetario considerable en la consulta de información crediticia de los clientes.

5.2. Recomendaciones

Debido a la importancia que tiene este proyecto para la empresa financiera con relación a los beneficios y aporte valor que genera, en base a los resultados obtenidos, se detallan las siguientes recomendaciones:

- Optimizar los Jobs de despliegue continuo en la herramienta Jenkins, dado que la memoria que se usa es compartida y esto origina que se crucen las ejecuciones por usar la misma instancia, generando un encolamiento y demora en los despliegues y pruebas unitarias durante la etapa de desarrollo.
- Revisar la factibilidad de poder usar otra metodología ágil como Kanban en lugar de Scrum dado que Kanban admite cambios de alcance durante la iteración siempre que haya capacidad disponible para abordarlos y para el Squad, específicamente para el presente proyecto, hubo muchas casuísticas de cambio de alcance de redefiniciones en el modelado de datos, lo que generaba realizar varios ajustes en el desarrollo. Por lo tanto, considero que esta metodología ágil se adapta a la forma de trabajo del Squad.
- Implementar una solución Big Data dado que la aplicación trabaja con grandes volúmenes de datos, estos provienen de fuentes difíciles de acceder que son cargados a una base de datos no relacional en formato semiestructurado como el JSON, con el objetivo de poder disponibilizar la información crediticia de clientes en una base de datos fácil de consumir.
- Agilizar el proceso de revisión de pares tanto del código fuente como de la documentación para la instalación de los cambios a los ambientes de certificación y producción, esto mejoraría si se planifican horas de un miembro del equipo con más experiencia para que destine su tiempo en la revisión de pares sin impactar sus requerimientos en curso.

FUENTES DE INFORMACION

- Antunes, B., Santos, D., Lopes, E., Fidalgo, F., & Alves, P. (2015). *Blisstrail: an agile project business case study*.
- Bala, M., Boussaid, O., & Alimazighi, Z. (2017). *A Fine-Grained Distribution Approach for ETL Processes in Big Data Environments*.
- *El futuro de los equipos*. (2018). Obtenido de <https://institutodeliderazgo.com/blog/si-el-entorno-ha-cambiado-la-estructura-y-el-funcionamiento-de-equipos-y-personas-ha-de-hacerlo-tambien>
- Fernández, Ó. (27 de 10 de 2021). *Databricks: Introducción a Spark en la nube*. Obtenido de <https://aprenderbigdata.com/databricks/>
- Microsoft. (2021). *What is cloud computing?* Retrieved from Azure Learning Path: <https://docs.microsoft.com/en-us/learn/modules/intro-to-azure-fundamentals/what-is-cloud-computing?ns-enrollment-type=LearningPath&ns-enrollment-id=learn.az-900-describe-cloud-concepts>
- Microsoft Azure. (2021). *Introducción a Azure Data Lake Storage Gen2*. Obtenido de Microsoft Docs: <https://docs.microsoft.com/es-es/azure/storage/blobs/data-lake-storage-introduction>
- Microsoft Azure. (2021). *¿Qué es Azure Data Factory?* Obtenido de Azure Learning Path: <https://docs.microsoft.com/es-es/azure/data-factory/introduction>
- Microsoft Azure. (2021). *Welcome to Azure Cosmos DB*. Obtenido de Azure Learning Path: <https://docs.microsoft.com/en-us/azure/cosmos-db/introduction>
- PEREYRA, D. A. (2010). Implementación de la metodología ágil SCRUM.
- QASource Blog. (2017). *QACOURSE*. Obtenido de <https://blog.qasource.com/big-data-testing-challenges-and-solutions-technocast-spring-2017>

GLOSARIO

- **RCC:** (Reporte Crediticio Consolidado): El Reporte de Deudas SBS muestra, de manera mensual, la relación de los créditos contratados con las empresas del sistema financiero, así como la calificación asignada por dichas entidades, en base al comportamiento de pago de los clientes (personas naturales).
- **Datalake:** Área especializada y habilitadora del banco, responsable en la generación de información de clientes de toda índole a través de procesos de ETL.
- **Time To Market:** Indicador de tiempo de consumo de un bien/producto desde su realización.
- **Roadmap:** Plan o cronograma de actividades para el desarrollo de tecnologías, en agilidad se maneja bajo hitos y en Sprint.
- **Big Room Planning:** Ceremonia donde se definen los objetivos e iniciativas que se priorizaran en el backlog de cada Squad que conforma la Tribu durante ese trimestre.
- **Congelamiento:** Proceso que indica que tu desarrollo finalizo y por ende ya no se pueden hacer más cambios, correcciones o mejoras. Es el pase de los cambios del ambiente desarrollo al ambiente de certificación.
- **Jira:** Herramienta DevOps que facilita el seguimiento de objetivos y promueve una entrega de valor iterativa e incremental al negocio.
- **RU:** Request Units o unidades de solicitud, indica el nivel de performance de los recursos computacionales como el CPU, IOPS (operaciones de lectura y escritura) y memoria, necesarios para realizar las operaciones de base de datos compatibles con Azure Cosmos DB.
- **Jenkins:** Herramienta DevOps que facilita la automatización de los procesos de construcción y despliegue en los desarrollos.
- **BitBucket:** Herramienta DevOps que facilita el control de versiones de los proyectos.
- **MVP:** Documento que se elabora en línea (desde Jira), tiene como finalidad indicar el paso a paso de la instalación, configuración, despliegue, ejecución o cualquier acción que se requiera realizar para que tu desarrollo pueda promoverse a los ambientes de certificación y producción.
- **Agile OPS:** Recurso de IBM encargado de implementar el paso a paso documentado en el MVP, durante el congelamiento y pase a producción.

- **Troubleshooting:** Guía de actividades para resolver un incidente.
- **Data Modeler:** Rol encargado de realizar el modelado de los datos y entendimiento de definición de campos y transformaciones de estos.
- **Dummy:** Hace referencia al archivo de prueba, que es preparado de manera manual para usarlo como input de las aplicaciones o proyectos en el Squad, se elabora según las necesidades de cada proyecto.
- **CRMO:** Squad responsable de proporcionar los archivos orígenes (archivo plano con información crediticia y archivo de control) para que la aplicación del presente proyecto pueda consumir y realizar la ingesta.

ANEXOS

Descripción de las unidades de programación

ANEXO 1: Repositorio del proyecto de carga RCC

Nombre del Elemento o Programa	Repositorio de Proyecto de carga RCC
Descripción del cambio	Proceso Databricks de carga RCC
Entradas	- Archivo plano con información crediticia de cliente - Archivos de control
Lógica de programación (pseudocódigo)	<pre>Definir variable de proceso con valor 3 Validar parámetro de proceso Leer parámetro de proceso Si parámetro de proceso es mayor igual que 3 Iniciar proceso de carga de archivo de control. Iniciar proceso de carga de archivo de origen RCC. Iniciar proceso de encriptación de campos DAC. Iniciar proceso de carga en tablas delta bronze Si parámetro de proceso es mayor igual que 2 Leer tabla delta bronce Iniciar proceso de limpieza y validación de registros Iniciar proceso de encriptación de campos DAC. Iniciar proceso de carga en tablas delta Silver Si parámetro de proceso es mayor igual que 1 Leer tabla delta silver Iniciar proceso de cálculo de totales Iniciar proceso de agrupamiento formato JSON Iniciar proceso de carga en tablas delta Gold Iniciar proceso de carga en base de datos Cosmos DB</pre>
Salidas	- Carga de datos en la colección de CosmosDB - Registro de registros con error en tabla delta

Interfaces usadas	Azure KeyVault: Azure Data Lake Storage Cosmos DB Azure Databricks
Alertas y Mensajes de Retorno	Envío de correos que notifican el inicio de la ejecución del Job Databricks y también informa si terminó correctamente o falló.

ANEXO 2: Azure Data Factory – Pipeline que invoca proceso Databricks

Nombre del Elemento o Programa	BATCH-ETL-BY-STAGES																
Descripción del cambio	Pipeline que invoca el proceso Databricks de carga RCC.																
Entradas	- Clúster Id - Job Id - Stage Id																
Lógica de programación (pseudocódigo)	Web KeyVault Web StartDatabricks Web StartJob General Settings User properties URL * databricks-token-aphsc-des?api-version=7.0 Parameters Variables Settings Output New Delete <table><thead><tr><th>☐</th><th>Name</th><th>Type</th><th>Default value</th></tr></thead><tbody><tr><td>☐</td><td>stageId</td><td>String</td><td>Value</td></tr><tr><td>☐</td><td>clusterId</td><td>String</td><td>Value</td></tr><tr><td>☐</td><td>jobId</td><td>String</td><td>Value</td></tr></tbody></table>	☐	Name	Type	Default value	☐	stageId	String	Value	☐	clusterId	String	Value	☐	jobId	String	Value
☐	Name	Type	Default value														
☐	stageId	String	Value														
☐	clusterId	String	Value														
☐	jobId	String	Value														



General **Settings** User properties

URL *

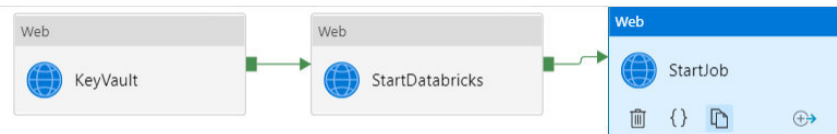
Method *

Headers * + New Delete

Name	Value
Authorization	@concat('Bearer ',activity('KeyVault').o...

Body

@concat('{\"cluster_id\":\",pipeline().parameters.clusterId,\"'})



General **Settings** User properties

URL *

Method *

Headers * + New Delete

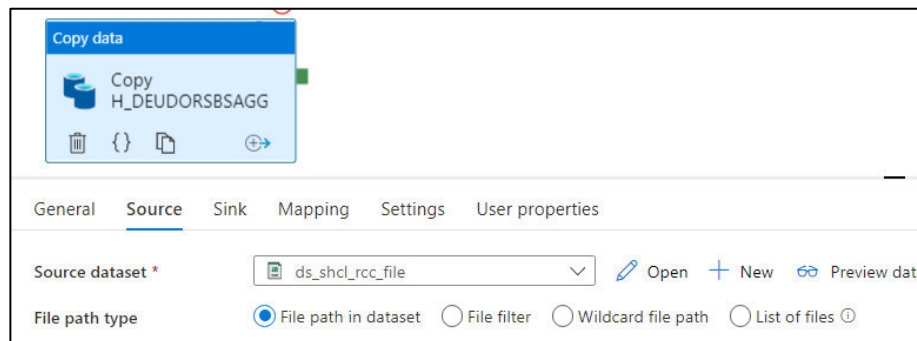
Name	Value
Authorization	@concat('Bearer ',activity('KeyVault').o...

Body

@concat('{\"job_id\":\",pipeline().parameters.jobId,\"jar_params\":\",\",pipeline().parameters.stageId,\"'})

ANEXO 3: Azure Data Factory - Pipeline que realiza el copiado al ADLS

Nombre del Elemento o Programa	Pipeline-RCC-Salesforce
Descripción del cambio	Pipeline que realiza el proceso de copiado de los archivos H_DEUDORSBSAGG.txt y H_DEUDORSBSAGG_CTRL.txt del servidor de Alexandria al ADLS.
Entradas	- Archivos (plano y de control)
Lógica de programación (pseudocódigo)	



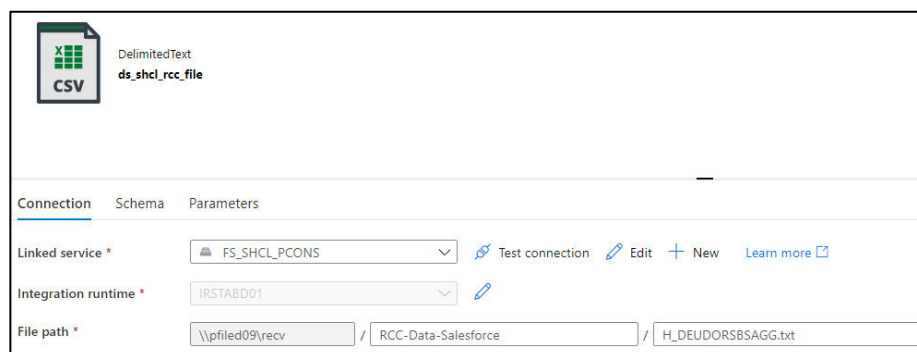
Copy data

Copy H_DEUDORSBSAGG

General **Source** Sink Mapping Settings User properties

Source dataset * ds_shcl_rcc_file [Open](#) [New](#) [Preview data](#)

File path type ☒ File path in dataset ☐ File filter ☐ Wildcard file path ☐ List of files ⓘ



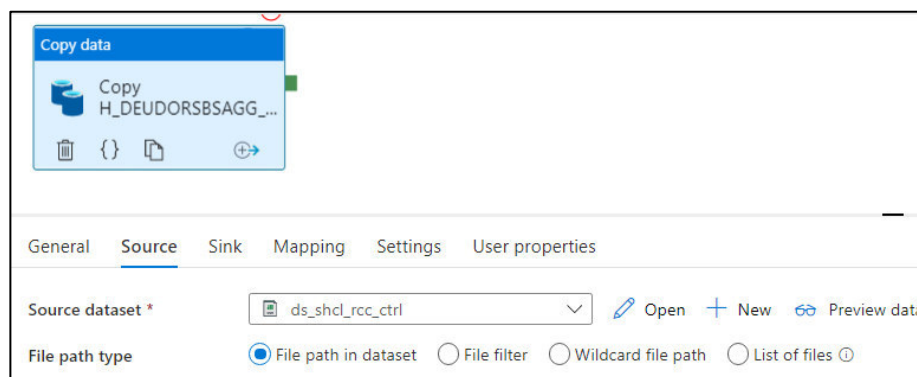
DelimitedText
ds_shcl_rcc_file

Connection Schema Parameters

Linked service * FS_SHCL_PCONS [Test connection](#) [Edit](#) [New](#) [Learn more](#)

Integration runtime * IRSTABD01 [Edit](#)

File path * \\pfiled09\recv / RCC-Data-Salesforce / H_DEUDORSBSAGG.txt



Copy data

Copy H_DEUDORSBSAGG_...

General **Source** Sink Mapping Settings User properties

Source dataset * ds_shcl_rcc_ctrl [Open](#) [New](#) [Preview data](#)

File path type ☒ File path in dataset ☐ File filter ☐ Wildcard file path ☐ List of files ⓘ

 DelimitedText ds_shcl_rcc_ctrl	
Connection Schema Parameters	
Linked service *	FS_SHCL_PCONS Test connection Edit + New Learn more
Integration runtime *	IRSTABD01
File path *	\\pfiled09\recv / RCC-Data-Salesforce / H_DEUDORSBSAGG_CTRL.txt
Salidas	Azure Data Lake Storage - Container: rcc-data - Ruta: data/temp

ANEXO 4: PowerShell script

Nombre del Elemento o Programa	PowerShell_UPLOAD_RCC.ps1
Descripción del cambio	Archivo Powershell que invoca al Pipeline-RCC-Salesforce ubicado en el Data Factory, con el fin de ejecutar el proceso Databricks para la carga de RCC.
Entradas	- Archivo de configuración JSON
Lógica de programación (pseudocódigo)	<pre> ##Load config File - Alexandria 's DataFactory configuration \$configFilePath = - join("D:\MASTER\config\\${(\$configfile.ToLower()).json}") Write-Host ("Loading Alexandria's DataFactory configuration for environment..." Out-String) \$config=Get-Content -Path \$configFilePath -Raw -ErrorAction Stop ConvertFrom-Json ##Get parameters \$Subscription = \$config.subscriptionId \$CertificateEnvironment = \$config.certificateEnvironment \$ApplicationId = \$config.applicationId \$TenantId = \$config.tenantId \$ResourceGroupName = \$config.resourceGroupName \$DataFactoryName = \$config.dataFactoryName \$PipelineName = \$config.pipelineUploadRCC ##Get certificate thumbprint \$certificate = Get-Childitem -path Cert:\LocalMachine\My -Force where {\$_.subject -like "CN=\${(\$CertificateEnvironment)}" } \$CertificateThumbprint = \$certificate.Thumbprint ##Login to Azure \$state=loginAzure ##Run data load Pipeline \$state=runPipeline exit \$state </pre>
Salidas	- Se realiza la carga RCC a CosmosDB.

ANEXO 5: Script del archivo BAT

Nombre del Elemento o Programa	BAT_RCC_INGESTION.BAT
Descripción del cambio	Archivo BAT que invoca el archivo PowerShell, con el fin de ejecutar el proceso Databricks de carga RCC.
Entradas	
Lógica de programación (pseudocódigo)	<pre> ::Powershell script SET dataFactoryScript=D:\MASTER\scripts\PowerShell_UPLOAD_RCC.ps1 ::Command to execute DataFactory powershell script SET pscommand="try { & %dataFactoryScript% -configfile:%configFile% -logfile:%pathlog%; \$msj= -join('[End] Se ejecuto powershell con Código de terminación RC=', \$LASTEXITCODE); Write-Output \$msj; exit \$LASTEXITCODE } catch { Write-Output \$_.Exception.Message; \$state=100; \$msj= -join('[End] RC=', \$state); Write-Output \$msj; exit \$state }" </pre>

ANEXO 6: Script del Job mainframe

Nombre del Elemento o Programa	@P4SH001 (Job mainframe)
Descripción del cambio	Job que realiza la ejecución del archivo BAT
Entradas	
Lógica de programación (pseudocódigo)	<pre> AGENT PSHCLWINP01 CMDNAME - D:\MASTER\ BAT_RCC_INGESTION.BAT </pre>