

CARRERA DE INGENIERIA INFORMATICA

CUCEI UDG

Languis

# Integrantes

|  |  |
| --- | --- |
| CÓDIGO DEL ALUMNO | NOMBRE COMPLETO DEL ALUMNO |
| 210404002 | Álvaro Josué García González |

# Colaboración con otras carreras

|  |  |  |  |
| --- | --- | --- | --- |
| Ingeniería en Computación | Ingeniería en Comunicaciones y Electrónica | Ingeniería Biomédica | Otra carrera |
| No | No | No | N/a |

# Módulos desarrollados en este proyecto

|  |  |  |
| --- | --- | --- |
| Gestión de la Tecnología de la Información | Sistemas Robustos, paralelos y distribuidos | Cómputo flexible (Softcomputing) |
| Sí | Sí | Sí |

# Asesores del proyecto modular

|  |  |
| --- | --- |
| NOMBRE DEL PRIMER ASESOR | NOMBRE DEL SEGUNDO ASESOR |
| Jorge Fausto Hernández Andrade | Ivonne del Rocío López |

# Bitácora de revisiones

|  |  |  |  |
| --- | --- | --- | --- |
| No. de revisión | Fecha de revisión  (dd/mm/yyyy) | Observaciones | Firma del asesor |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Tabla de contenido

[Integrantes 0](#_Toc507507119)

[Colaboración con otras carreras 0](#_Toc507507120)

[Módulos desarrollados en este proyecto 0](#_Toc507507121)

[Asesores del proyecto modular 0](#_Toc507507122)

[Bitácora de revisiones 2](#_Toc507507123)

[Agradecimientos 6](#_Toc507507124)

[Resumen del proyecto 8](#_Toc507507125)

[Objetivo 8](#_Toc507507126)

[Antecedentes 8](#_Toc507507127)

[Justificación 8](#_Toc507507128)

[Descripción 9](#_Toc507507129)

[Módulo 1: Gestión de la información 9](#_Toc507507130)

[Módulo 2: Sistemas distribuidos, paralelos o robustos 9](#_Toc507507131)

[Módulo 3: Softcomputing 9](#_Toc507507132)

[Clasificación del proyecto 9](#_Toc507507133)

[Impacto social 9](#_Toc507507134)

[Tipo de compañía 10](#_Toc507507135)

[¿Cómo se clasifican las sociedades mercantiles? 10](#_Toc507507136)

[Organigrama 11](#_Toc507507137)

[Puestos y salarios 12](#_Toc507507138)

[Protección de datos de carácter personal 13](#_Toc507507139)

[Aviso de privacidad 14](#_Toc507507140)

[Legislación y jurisdicción aplicable 18](#_Toc507507141)

[Impacto social 19](#_Toc507507142)

[Emprendurismo 20](#_Toc507507143)

[Organismos financiadores 20](#_Toc507507144)

[Públicos 20](#_Toc507507145)

[Privados 22](#_Toc507507146)

[Introducción 23](#_Toc507507147)

[Contexto de la problemática detectada 23](#_Toc507507148)

[Marco de referencia 23](#_Toc507507149)

[Descripción del problema detectado 24](#_Toc507507150)

[Contenido 24](#_Toc507507151)

[Objetivos 24](#_Toc507507152)

[Restricciones 25](#_Toc507507153)

[Tabla comparativa con otros sistemas encontrados 25](#_Toc507507154)

[Módulo por diseñar 25](#_Toc507507155)

[Módulo I y II: Gestión de la información 25](#_Toc507507156)

[Módulo III: Sistemas distribuidos, paralelos o robustos 26](#_Toc507507157)

[Módulo IV: Softcomputing 26](#_Toc507507158)

[Metodología para resolver el problema detectado 26](#_Toc507507159)

[Tipo de investigación 26](#_Toc507507160)

[Investigación cuantitativa 26](#_Toc507507161)

[Definición del tamaño y la muestra de la investigación 26](#_Toc507507162)

[Consulta y acceso a repositorios de datos 27](#_Toc507507163)

[Estándares usados 27](#_Toc507507164)

[Diseño de la propuesta 28](#_Toc507507165)

[Arquitectura 30](#_Toc507507166)

[Características 30](#_Toc507507167)

[Modelado de la base de datos 31](#_Toc507507168)

[Diseño conceptual 33](#_Toc507507169)

[Primer caso 34](#_Toc507507170)

[Segundo caso 38](#_Toc507507171)

[Tercer caso 41](#_Toc507507172)

[Dependencias con otros elementos de software o hardware de terceros 45](#_Toc507507173)

[Cronograma de actividades 45](#_Toc507507174)

[Iteración I 46](#_Toc507507175)

[Iteración II 46](#_Toc507507176)

[Iteración III 46](#_Toc507507177)

[Iteración IV 46](#_Toc507507178)

[Iteración V 47](#_Toc507507179)

[Iteración VI 47](#_Toc507507180)

[Iteración VII 48](#_Toc507507181)

[Iteración VIII 48](#_Toc507507182)

[Iteración IX 49](#_Toc507507183)

[Iteración X 49](#_Toc507507184)

[Resultados obtenidos 49](#_Toc507507185)

[Análisis de resultados 49](#_Toc507507186)

[Capturas de Pantalla 50](#_Toc507507187)

[Gestión de la tecnología de la información 50](#_Toc507507188)

[Sistemas robustos, paralelos y distribuidos 53](#_Toc507507189)

[Cómputo flexible 57](#_Toc507507190)

[Conclusiones 59](#_Toc507507191)

[Gestión de la tecnología de la información 59](#_Toc507507192)

[Sistemas robustos, paralelos y distribuidos 59](#_Toc507507193)

[Cómputo flexible 60](#_Toc507507194)

[Referencias bibliográficas 62](#_Toc507507195)

[Referencia de figuras 63](#_Toc507507196)

# Agradecimientos

Desarrollar los tres módulos requeridos por mi carrera para un único proyecto modular ha sido uno de los mayores desafíos que he enfrentado durante toda mi estadía en la Benemérita Universidad de Guadalajara.

Este trabajo presentado es el resultado final de arduos meses de constante dedicación, aprendizaje e investigación a un alto grado de dificultad y complejidad y al que después de un análisis objetivo muestra inmediatamente que la magnitud de este mismo no habría sido posible sin el apoyo de personas e instituciones que han simplificado y facilitado en su total posibilidad el desarrollo satisfactorio en el que ha resultado.

Es por ello por lo que me complace agradecer a todos aquellos que directa e indirectamente han sido parte de este trabajo.

A ti, Ivonne, por haber sido una grandísima líder técnica durante todo el desarrollo de este proyecto y por haber confiado en mí desde el primer día en que era capaz de liderar un proyecto prácticamente desde cero y sin conocimiento real sobre todo lo que implicaba el desarrollo de este mismo.

La asesoría técnica y el coaching técnico, y emocional inclusive, para presentaciones con los clientes no podrían haber sido lo que fueron sin ti, sin lugar a duda.

Agradezco tu exigencia a un nivel top para que lográramos que este proyecto fuese capaz de cumplir cada uno de los estándares y requisitos propios de la IBM y el cliente. Y sobre todo a tu absoluta disposición para resolver cada una de las inquietudes que surgieron conforme avanzaba el desarrollo.

A ti, Dr. Fausto, por aceptarme para realizar este proyecto modular bajo tu total dirección. Tu apoyo y confianza en mi trabajo y tu absoluta capacidad para guiar mis ideas han sido, sin duda alguna, un aporte intangible e invaluable, no solamente para el desarrollo de este documento, sino también en mi formación como investigador.

Las ideas propias, siempre enmarcadas por tu orientación y rigurosidad, han sido la clave del buen trabajo que hemos realizado. Agradezco infinitamente el haberme facilitado siempre los medios suficientes para llevar a cabo todas las actividades propuestas durante el desarrollo de este proyecto.

Quiero hacer mención especial de ustedes, mis amigos y compañeros de generación, Daniel y Uriel, porque al haber sido compañeros míos supieron cómo ayudarme y asesorarme con dudas puntuales y guías que me ofrecieron desde el enfoque que solamente estudiantes que realizaron sus proyectos a la par podrían ofrecer.

Durante la carrera fuimos compañeros en incontables ocasiones y sin duda eso provocó que indirectamente supiésemos como ayudarnos mutuamente y eso es valiosísimo. Gracias totales por eso, compadres.

Finalmente, quiero agradecerte a ti, Juan José, porque al ser mi First line manager dentro de la IBM me has dado todas las facilidades para lograr realizar este documento en base a un desarrollo interno para la compañía.

Estoy muy agradecido porque creo que este desarrollo ha sido uno de los mejores de mi recién iniciada carrera como desarrollador y la oportunidad de poder presentar mi opera prima para la universidad es lo máximo.

Ustedes han sido los máximos involucrados y sin su apoyo y exigencia esto no habría resultado en lo que ha resultado, que en mí no tan humilde opinión, es un fantástico trabajo.

Estoy seguro de que tendremos la oportunidad de volver a trabajar juntos en algún otro momento y no podría estar más contento ante tal inminente hecho.

Gracias por tanto a todos.

# Resumen del proyecto

## Objetivo

Desarrollar un sistema donde sea posible dar un seguimiento puntual y preciso a la renovación de cotizaciones para todo tipo de cliente (Pymes y Grandes empresas) que posee una compañía, aunque está dirigido principalmente para Pymes con la capacidad para alertar de renovaciones a los clientes y al mismo tiempo tener en tiempo real el acceso a la información relacionada con las cotizaciones de los servicios de software ofrecidos (SaaS).

## Antecedentes

Durante mucho tiempo, si alguna empresa o usuario deseaba obtener algún servicio provisto por la compañía, IBM, era necesario contactar a un representante de ventas (Bussiness Parnert - BP), el cual a través de un sistema llamado Software Quote and Order (SQO) generaba cada una de las cotizaciones. Una vez capturada la cotización por el BP esta se guardaba directamente en un mainframe de SAP.

Una vez la cotización se encuentra en el mainframe de SAP, se envía la información correspondiente hacía una base de datos Db2 donde se almacenará para su posterior uso en la compañía.

A partir del año 2016 fue lanzado un Marketplace donde la compañía ofrecía sus SaaS a quien estuviese interesado en ello, para esto fue utilizado el mismo sistema, SQO, que ya era utilizado por los BPs anteriormente.

Los personales encargados de trabajar con los datos relacionados a las cotizaciones sufrían el problema de que la información no estaba actualizada, ya que al ser una cantidad masiva podían transcurrir horas para que se actualizarán los datos. Aunado al hecho de la curva de aprendizaje porque el sistema donde se obtiene esa información tiene una interfaz de usuario poco amigable y es considerablemente lenta.

## Justificación

Ante la necesidad de obtener la información relacionada con las cotizaciones a través del sistema SQO para accesar, visualizar y manipular se decidió implementar el desarrollo de una solución capaz de comunicarse en tiempo real entre diferentes sistemas de bases de datos como lo son db2 y Cloudant, un servicio de bases de datos no relacional (jerárquico), en el cuál es posible implementar nuevas tecnologías web (tal como ficheros JSON), donde se podrá ser capaz de mover cantidades masivas de datos.

La solución deberá ser capaz de comunicarse en tiempo real entre diferentes bases de datos, que sea fácil de accesar, de visualizar y de manipular el estatus de las cotizaciones donde siempre esté disponible, con una interfaz de usuario amigable.

Debido a que existe una cantidad masiva de información dentro de las cotizaciones que Db2 recolecta de SAP se utilizarán diversos algoritmos para emplear minería de datos, donde se comparará la información existente entre ambas bases de datos y se procesará para que sea posible obtener cotizaciones que cumplan con las características que se necesitan ya que disponemos de 10 minutos para revisar 100 terabytes en ese tiempo por lo que es necesario optimizar el proceso de búsqueda y detección de campos que han sido cambiados para otorgar la cotización más reciente.

## Descripción

### Módulo 1: Gestión de la información

Se utilizan dos bases de datos: una base de datos relacional en DB2 y una base de datos no relacional (jerárquica) en Cloudant.

### Módulo 2: Sistemas distribuidos, paralelos o robustos

La base de datos de Cloudant vive en la nube, se utilizan tecnologías web como archivos JSON y WEB APIs para la conexión de la cual hace uso el sistema para hacer la conexión entre ambas bases de datos.

### Módulo 3: Softcomputing

Se utiliza un mainframe que viene de sistemas en SAP, de donde DB2 obtiene los datos. A partir de ahí la solución debe obtener de la base de datos en DB2, donde se utilizan múltiples tablas, las cuales contienen un volumen masivo de datos para manipular y procesar a través de diversos algoritmos de minería de datos donde se extraerá toda la información que impliqué una relación de uniformidad en los datos para ser enviados a la base de datos de Cloudant.

# Clasificación del proyecto

La clasificación del proyecto entra dentro del área de desarrollo tecnológico y el reto a solventar es el de automatización.

# Impacto social

Glez Enterprise S.A. de C.V., también conocida como Glezcorp, es una compañía de tecnología y consultoría con sede en Guadalajara, Jalisco, México. Glezcorp fabrica y comercializa software especializados para computadoras, y ofrece servicios de infraestructura y consultoría en una amplia gama de áreas relacionadas con la informática. Fue fundada en noviembre de 2017 por Álvaro González. En la figura 1 se muestra el logo de la compañía.



Figura 1: Logo de la compañía

El logo de Glezcorp tiende desde pequeños círculos que tienden a lograr un conjunto mayor de los mismos representando la evolución que Glezcorp como compañía ha tenido y tendrá eventualmente. Evoca a una evolución gradual y atemporal de forma constante y continúa logrando un ciclo continuo e infinito. El color rojo, tal como el primer color que los recién nacidos pueden ver, simboliza una nueva creación.

Dentro de los proyectos más sobresalientes de la compañía se encuentra el proyecto Languis.

## Tipo de compañía

Según los lineamientos para la formación de sociedades mercantiles en México actuales al mes de noviembre de 2017, una sociedad mercantil es reconocida por la ley como una entidad propia y distinta a sus miembros. Por lo tanto, cuenta con un patrimonio común y con la participación de socios, tanto en sus ganancias como en sus pérdidas. Es un ente colectivo con un interés o propósito en común.

Específicamente, una sociedad mercantil:

* Comparte el propósito de obtener ganancias a través de actos lucrativos comerciales (lo cual lo diferencia de una sociedad civil),
* Comparte la contribución a las pérdidas por riesgo de negocio entre todos los socios,
* Necesita que cada socio aporte una contribución patrimonial, ya sea de capital, bienes o trabajo,
* Requiere de una transacción comercial de una compraventa.

Formar una sociedad mercantil ofrece los beneficios de poder reunir una mayor cantidad de capital para desarrollar negocios, además de gozar del beneficio de la responsabilidad limitada.

¿Cómo se clasifican las sociedades mercantiles?

Para clasificar las sociedades se toman en cuenta los elementos y el tipo de capital.

Existen tres elementos:

* El personal: formado por socios que aportan y reúnen sus esfuerzos (bienes, capital o trabajo).
* El patrimonial: constituido por bienes que juntos forman un capital social.
* El formal: compuesto por reglas que rigen la transformación del contrato, justificando así el origen de la sociedad como ente.

Y dos tipos de capital:

* Capital fijo: el que no se puede modificar, salvo por un cambio de los estatutos.
* Capital variable (C.V.): aquel que es flexible conforme al avance de la sociedad mercantil*.*

Por lo mencionado anteriormente se fundó la compañía como una sociedad anónima (S.A.) ya que este tipo de sociedad se caracteriza porque su capital está representado por acciones nominativas; sus socios están obligados al monto de sus acciones, ya sea en efectivo o en especie; y los administradores tienen responsabilidad ilimitada.

El mínimo de capital es de cincuenta mil pesos. Las reservas representan 5% de las utilidades anuales hasta reunir 20% del capital social fijo. Existe un mínimo de dos socios y no tiene limitaciones en el número de socios. Su duración puede ser de 1 a 99 años. Estas sociedades pueden ser cerradas o abiertas, es decir, pueden hacer oferta pública de acciones.

Una vez definido esto a continuación se muestra el organigrama de la empresa.

## Organigrama

Puesto que se trata de una empresa pequeña se tiene el organigrama mostrado en la figura 2:

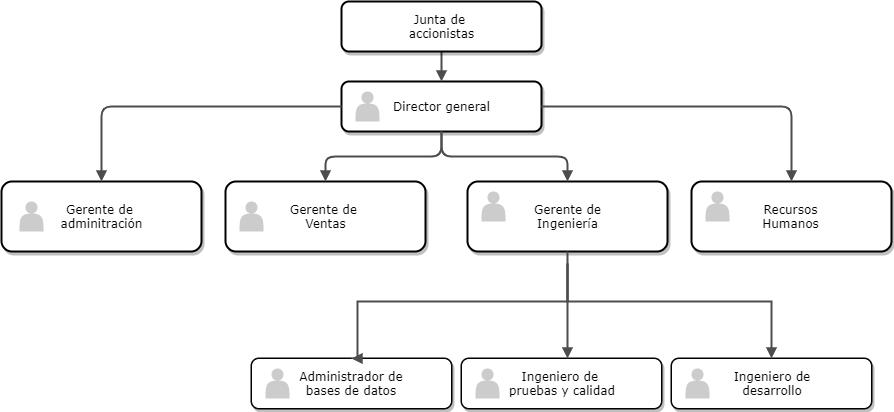


Figura 2: Organigrama de la compañía

De acuerdo con lo expresado en el organigrama la empresa cuenta inicialmente con los siguientes cargos:

**Junta de accionistas**: es la principal fuente de ingresos de la compañía, los inversores se encargan de tomar decisiones sobre la dirección a la cual va el enfoque del negocio. Son la principal autoridad dentro de la empresa.

**Director general**: es el encargado principal y la cara de la empresa ante los clientes de mayor impacto. Es el mayor responsable de cumplir con la misión y visión junto con el resto de los empleados.

**Gerente de administración**: su principal función es la de encargarse de los temas administrativos de la empresa, tal como el manejo responsable del presupuesto destinado semestralmente, consolidar los números de la empresa y gestar la correcta administración de todos los recursos disponibles.

**Gerente de ventas**: es el principal encargado de todos los temas de negocios, representa a la empresa ante los clientes.

**Gerente de ingeniería**: es el principal encargado de gestionar y crear los todos los proyectos ante los clientes. Es el analista del negocio y es capaz de lograr concretar una idea abstracta de lo solicitado por el cliente en una solución integral.

**Administrador de bases de datos**: encargado de gestionar las bases de datos de la compañía y principal analista en bases de datos en todos los proyectos.

**Ingeniero de pruebas y calidad**: encargado de verificar y cumplir los estándares a los cuales la compañía se compromete y de probar cada uno de los proyectos antes de ser enviados a producción.

**Ingeniero de desarrollo:** es la principal mano de obra, desarrolla todas las soluciones en base a sus conocimientos.

**Recursos Humanos**: principal puesto encargado de ser un intermediario entre la empresa y los empleados.

## Puestos y salarios

En base al trabajo realizado para la compañía se establece que la cuantía de los salarios, a partir del 1ro de julio de 2017, será conforme a la tabla 1 de posiciones y salarios:

|  |  |
| --- | --- |
| Posición | Salario mensual |
| Socio/Accionista | Información no disponible |
| Director general | $ 49,999.99 MXN |
| Gerente de administración | $ 35,000.00 MXN |
| Gerente de ventas | $ 35,000.00 MXN |
| Gerente de ingeniería | $ 35,000.00 MXN |
| Recursos Humanos | $ 22,000.00 MXN |
| Administrador de bases de datos | $ 29,999.99 MXN |
| Ingeniero de pruebas y calidad | $ 24,999.99 MXN |
| Ingeniero de desarrollo | $ 24,999.99 MXN |

Tabla 1: Posiciones y salarios

Los pagos se realizarán el día 14 de cada mes, con excepción de aquellos días en que coincida con un día inhábil, en ese caso se pagará un día hábil inmediatamente anterior.

Se busca mantener un salario adecuado para los empleados para que estos se sientan valorados por la empresa, es por ello por lo que de forma adicional a su remuneración mensual se les ofrece una serie de prestaciones adicionales a las obligadas por la Ley Federal del Trabajo las cuales consisten en:

* Fondo de ahorro
* Vales de despensa
* Bono de productividad
* Servicio de comedor
* Un día adicional de vacaciones por cada 30 días laborados
* 30 días de aguinaldo
* Seguro de gastos médicos mayores
* Seguro de vida
* Plan de carrera

La importancia de mantener a los empleados felices es que gracias a ello podrán mantener un rendimiento superior a la media, lo cual es de vital importancia para la empresa. De esa misma forma los empleados se comprometen más antes la compañía de forma que se puede ofrecer una mayor confianza y seguridad ante los clientes que trabajan con información sensible.

Con respecto a lo anterior, Glezcorp cuenta con una regla interna de protección de datos de carácter personal y un aviso público de confidencialidad con el fin de ofrecer un mejor servicio y cumplir con las normas establecidas por la Ley Federal de Protección de Datos Personales.

A continuación, se muestra el contenido de ambos documentos.

## Protección de datos de carácter personal

Los derechos de propiedad intelectual de Glezcorp, su código fuente, diseño, estructura de navegación, bases de datos y los distintos elementos contenidos en el mismo son titularidad de Glezcorp a quien corresponde el ejercicio exclusivo de los derechos de explotación de los mismos en cualquier forma y, en especial, los derechos de reproducción, distribución, comunicación pública y transformación.

Se autoriza la visualización, impresión ni descarga parcial o total del contenido del software solo y exclusivamente si concurren las siguientes condiciones: 1) Que sea compatible con los fines del proyecto. 2) Que se realice con el exclusivo ánimo de obtener la información contenida para su uso personal y privado. Se prohíbe expresamente su utilización con fines comerciales o para su distribución, comunicación pública, transformación o des compilación. 3) Dicho documento deberá siempre incluir referencia de que es propiedad de Glezcorp. 4) Que ninguno de los contenidos relacionados en este software sea utilizados, copiados o distribuidos separadamente del texto o resto de imágenes que lo acompañan. [1]

De la misma forma, y dado que los datos personales son sensibles se tiene un aviso de privacidad con el cual nuestros clientes pueden estar seguros de que su información no será mal utilizada.

## Aviso de privacidad

De conformidad con lo establecido en la Ley Federal de Protección de Datos Personales en Posesión de los Particulares, **Glez Enterprise Sociedad Anónima de Capital Variable** (en adelante Glezcorp), pone a su disposición el siguiente aviso de privacidad.

**Glezcorp**, es responsable del uso y protección de sus datos personales, en este sentido y atendiendo las obligaciones legales establecidas en la Ley Federal de Protección de Datos Personales en Posesión de los Particulares, a través de este instrumento se informa a los titulares de los datos, la información que de ellos se recaba y los fines que se le darán a dicha información.

Glez Enterprise S.A. de C.V. (Glezcorp), con domicilio en:

Avenida Infinito, número 934;

Colonia Providencia, código postal 90821

Guadalajara, Jalisco, México,

y al respecto le informa lo siguiente:

Los datos personales que recabamos de usted, los utilizaremos para las siguientes finalidades que son necesarias para el servicio que solicita: La recolección de los datos es con la finalidad de ofrecer un mejor servicio al momento de hacer uso de nuestras tecnologías. Personalmente, esto es en presencia de la persona que otorga sus datos personales (titular de los datos personales), por ejemplo: a través de formatos impresos o entrevistas directas.

De manera adicional, utilizaremos su información personal para las siguientes finalidades que no son necesarias para el servicio solicitado, pero que nos permiten y facilitan brindarle una mejor atención: origen étnico, estado de salud, información genética, creencias religiosas, preferencia sexual.

Los datos personales que recabamos de usted, los utilizaremos para las siguientes finalidades que son necesarias para el servicio que solicita:

Nombre,

Teléfono,

Domicilio,

Correo electrónico.

En caso de que no desee que sus datos personales sean tratados para estos fines adicionales, usted puede presentar desde este momento un escrito en esta sucursal, manifestando lo anterior. Solicite el formato correspondiente a través de la página web: [www.glezcorp.com/ARCO](http://www.glezcorp.com/ARCO).

La negativa para el uso de sus datos personales para estas finalidades no podrá ser un motivo para que le neguemos los servicios y productos que solicita o contrata con nosotros.

Para llevar a cabo las finalidades descritas en el presente aviso de privacidad, utilizaremos datos personales de identificación, contacto, laborales y patrimoniales.

Le informamos que para las transferencias indicadas con un asterisco (\*) requerimos obtener su consentimiento. Si usted no manifiesta su negativa para dichas transferencias, entenderemos que nos lo ha otorgado.

No autorizo que mis datos personales sean compartidos con los siguientes terceros:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Firma

Usted tiene derecho a conocer qué datos personales tenemos de usted, para qué los utilizamos y las condiciones del uso que les damos (Acceso). Asimismo, es su derecho solicitar la corrección de su información personal en caso de que esté desactualizada, sea inexacta o incompleta (Rectificación); que la eliminemos de nuestros registros o bases de datos cuando considere que la misma no está siendo utilizada conforme a los principios, deberes y obligaciones previstas en la normativa (Cancelación); así como oponerse al uso de sus datos personales para fines específicos (Oposición). Estos derechos se conocen como derechos ARCO.

Para el ejercicio de cualquiera de los derechos ARCO, usted deberá presentar la solicitud respectiva en [Describir los medios].

El procedimiento y requisitos para el ejercicio de estos derechos es el siguiente:

Nombre,

Teléfono,

Carta de expresión de motivos,

Correo electrónico.

Los datos de contacto de la persona o departamento que dará trámite a las solicitudes para el ejercicio de los derechos ARCO, así como atender cualquier duda que pudiera tener respecto al tratamiento de su información son los siguientes:

Departamento de Privacidad, con domicilio en Avenida Infinito, número 934, Colonia Providencia, código postal 90821, Guadalajara, Jalisco, México, correo electrónico [atencion@glezcorp.com](mailto:atencion@glezcorp.com) y número telefónico 01 800 000 0000.

Usted puede revocar el consentimiento que, en su caso, nos haya otorgado para el tratamiento de sus datos personales. Sin embargo, es importante que tenga en cuenta que no en todos los casos podremos atender su solicitud o concluir el uso de forma inmediata, ya que es posible que por alguna obligación legal requiramos seguir tratando sus datos personales. Asimismo, usted deberá considerar que, para ciertos fines, la revocación de su consentimiento implicará que no le podamos seguir prestando el servicio que nos solicitó, o la conclusión de su relación con nosotros.

Para revocar su consentimiento deberá presentar su solicitud a través del sitio web: [www.glezcorp.com/ARCO](http://www.glezcorp.com/ARCO).

El procedimiento y requisitos para la revocación del consentimiento es el siguiente:

1. Llenar un formulario con los siguientes datos:

Nombre,

Teléfono,

Carta de expresión de motivos,

Correo electrónico.

2. En un lapso no menos a tres días hábiles recibirá una respuesta por parte del área administrativa que será la encargada de llevar su situación.

Con objeto de que usted pueda limitar el uso y divulgación de su información personal, le ofrecemos los siguientes medios:

Su inscripción en el Registro Público para Evitar Publicidad, que está a cargo de la Procuraduría Federal del Consumidor, con la finalidad de que sus datos personales no sean utilizados para recibir publicidad o promociones de empresas de bienes o servicios. Para más información sobre este registro, usted puede consultar el portal de Internet de la PROFECO, o bien ponerse en contacto directo con ésta.

Su registro en el listado de exclusión “Exclusión ARCO”, a fin de que sus datos personales no sean tratados para fines mercadotécnicos, publicitarios o de prospección comercial por nuestra parte. Para más información llamar al número telefónico 01 800 000 0000, enviar un correo electrónico a la siguiente dirección [atencion@glezcorp.com](mailto:atencion@glezcorp.com), o bien, consultar nuestra página de Internet [www.glezcorp.com/ARCO](http://www.glezcorp.com/ARCO).

Le informamos que en nuestra página de Internet utilizamos cookies, web beacons y otras tecnologías a través de las cuales es posible monitorear su comportamiento como usuario de Internet, brindarle un mejor servicio y experiencia de usuario al navegar en nuestra página, así como ofrecerle nuevos productos y servicios basados en sus preferencias.

Los datos personales que obtenemos de estas tecnologías de rastreo son los siguientes: horario de navegación, tiempo de navegación en nuestra página de Internet, secciones consultadas, y páginas de Internet accedidas previo a la nuestra.

Asimismo, le informamos que sus datos personales que se obtienen a través de estas tecnologías los compartiremos con las siguientes personas, empresas, organizaciones o autoridades distintas a nosotros, para los siguientes fines:

* Destinatario de los datos personales Finalidad
* Google Análisis de visitas en el sitio web

Estas tecnologías podrán deshabilitarse siguiendo los siguientes pasos: 1. Acceder a nuestra página de Internet, sección “Términos y condiciones del sitio”, subsección “Cookies”; 2. Dar clic en la subsección “Cookies”; 3. Leer el mensaje de advertencia sobre la deshabilitación de cookies, y 4. Dar clic en la leyenda de activar el mecanismo de deshabilitación de cookies.

Para más información sobre el uso de estas tecnologías, puede consultar el sitio de Internet [www.glezcorp.com/Privacidad](http://www.glezcorp.com/Privacidad).

El presente aviso de privacidad puede sufrir modificaciones, cambios o actualizaciones derivadas de nuevos requerimientos legales; de nuestras propias necesidades por los productos o servicios que ofrecemos; de nuestras prácticas de privacidad; de cambios en nuestro modelo de negocio, o por otras causas.

Nos comprometemos a mantenerlo informado sobre los cambios que pueda sufrir el presente aviso de privacidad, a través de [www.glezcorp.com/AvisoPrivacidad](http://www.glezcorp.com/AvisoPrivacidad).

El procedimiento a través del cual se llevarán a cabo las notificaciones sobre cambios o actualizaciones al presente aviso de privacidad es el siguiente: www.glezcorp.com/AvisoPrivacidad.

Le informamos que es de nuestro especial interés cuidar la información personal de los menores de edad y personas en estado de interdicción y capacidades diferentes en términos de ley, a través del establecimiento de medidas específicas, como son:

La obtención del consentimiento de los padres y tutores por medio de un correo electrónico, para el tratamiento de los datos personales de este grupo de personas.

La implementación y mantenimiento de medidas de seguridad más estrictas a efecto de asegurar la confidencialidad de los menores y este grupo de personas.

Si usted considera que su derecho a la protección de sus datos personales ha sido lesionado por alguna conducta u omisión de nuestra parte, o presume alguna violación a las disposiciones previstas en la Ley Federal de Protección de Datos Personales en Posesión de los Particulares, su Reglamento y demás ordenamientos aplicables, podrá interponer su inconformidad o denuncia ante el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI). Para más información, le sugerimos visitar su página oficial de Internet [www.inai.org.mx](http://www.inai.org.mx).

Otorgo mi consentimiento para que mis datos personales sean tratados conforme a lo señalado en el presente aviso de privacidad.

1. La identidad y domicilio del responsable;
2. Las finalidades del tratamiento, distinguiendo las que dieron origen y son necesarias para la relación jurídica entre titular y responsable, de las que no lo son;
3. Los mecanismos para que el titular pueda manifestar previamente su negativa para el tratamiento de sus datos personales respecto de aquellas finalidades secundarias o accesorias, y
4. Los mecanismos para que el titular conozca el aviso de privacidad integral.

--------------------------

Nombre y firma autógrafa del titular

[2]

## Legislación y jurisdicción aplicable

Como parte de la protección de la creación del contenido desarrollado se sigue un procedimiento el cual consta de:

* **Asesoría legal** esto como principio básico ya que ofrece la mejor garantía para mayores posibilidades de éxito, y sobre todo para evitar errores y perder el tiempo.
* **Evaluar la situación** ¿cuál es la intención que se busca ante la situación?
* **Solución amistosa** debido a que en la mayor parte de los casos de plagio y usos ilícitos sedan por alguna confusión o error no malintencionado, de esa forma se evita un proceso prolongado.
* **Recabar pruebas de autoría** haciendo uso de los registros de propiedad intelectual ante el IMPI
* **Asegurar la prueba del presunto uso ilícito** esta parte corresponde al abogado externo que lleve el caso de forma que evidenciar el delito
* **Enviar carta de incidencia** Una notificación para informar a quien está utilizando sin permiso los desarrollos de la empresa
* **Emprender acciones legales** si después del envío de la incidencia no se aplica ningún cambio se puede comenzar la acción legal en contra, siempre recibiendo asesoría legal [3]

Como parte de la jurisdicción aplicable se mencionan en la tabla 2 de etapas y documentos aplicados para la compañía y desarrollo de software, a continuación, las normas y estándares utilizados en el desarrollo de proyectos por parte de la empresa.

|  |  |  |
| --- | --- | --- |
| Etapa | Documento | Uso |
| Análisis de requerimientos | ISO 12207 | Se utiliza con la finalidad de evaluar los procesos de desarrollo y mantenimiento de sistemas. [4] |
| Diseño y desarrollo de software | ISO 9001, ISO 27001, Manifiesto Ágil | Centrarse en el proceso y satisfacción del cliente en lugar de en procedimientos, de tal forma se proporciona la infraestructura, procedimientos, procesos y recursos necesarios para controlar y mejorar el rendimiento y conducir a la empresa hacia la eficiencia, servicio al cliente y excelencia en el producto. [5]  De igual forma al manejar información sensible sobre nuestros clientes está se vuelve nuestra columna vertebral por lo que necesita ser protegida ante cualquier amenaza. [6]  Enfocar el diseño del software hacia la simplicidad, de tal forma que sea posible simplificar el trabajo realizado por todos los miembros del equipo. [7] |
| Mantenimiento de software | ISO 14764 | Se debe mantener un estándar para aclarar los requerimientos para el proceso de mantenimiento al ser un proceso primario en el ciclo de vida del mismo. [8] |
| Requerimientos de software | IEEE 830 | Mantener y redactar especificaciones de software que indiquen exactamente lo solicitado por el cliente y que permitan a su vez entender clara y concisamente al desarrollador lo que el cliente necesita. [9] |
| Pruebas de software | ISO 29119 | Mantener un proceso constante de pruebas de tal manera que el software ofrecido a los clientes sea de la más alta calidad. [10] |
| Seguridad, privacidad y legalidad | ASCA IBM | Consiste en entender e interpretar las políticas y prácticas dentro de IBM para cumplir con los procesos internos de negocio. Esto incluye la reducción de riesgos sobre el tratamiento de información o datos confidenciales, sensibles y/o personales. |

Tabla 2: Etapas y documentos

## Impacto social

Al ser un Glezcorp una empresa de desarrollo de software y consultoría nuestros servicios se ofrecen al público en general con cierto énfasis al sector de la pequeñas y medianas empresa (Pymes).

En especial sentido ofrecemos la solución desarrollada para el proyecto Languis debido a que esta herramienta de software facilitará el trabajo de extracción, transformación y carga de datos (ETL por sus siglas en inglés) en cantidades masivas de datos.

El proyecto Languis sirve como ópera prima para presentar al público en general la capacidad de desarrollo de software a la medida, demostrando la capacidad de innovación en nuevas tecnologías y soluciones integrales.

# Emprendurismo

## Organismos financiadores

### Públicos

#### Inadem [11]

El propósito de este Módulo es presentar a los micro, pequeños y medianos empresarios (MIPYMES), elementos que les permitan conocer de manera sintética, qué es una empresa integradora, sus objetivos, funciones y beneficios, así como las etapas para configurar un proyecto de estas características, los requisitos de constitución y los apoyos institucionales, entre otros aspectos.

El desarrollo regional debe sustentarse en estrategias que permitan aprovechar de manera equilibrada las vocaciones productivas de las diferentes regiones del país, con el propósito de generar ventajas competitivas sostenibles.

En este sentido, las formas de organización interempresarial se constituyen como una estrategia eficaz para organizar y priorizar los factores de la producción, desarrollar actividades complementarias o de soporte y, lo más importante, realizar una planeación estratégica regional de mediano y largo plazos en donde se involucren a todos los agentes económicos, para el cumplimiento de las metas.

La adopción de alternativas de cooperación empresarial es el resultado de la mayor competencia que representa la globalización de los mercados. Por ello, existe un renovado interés en la articulación productiva, particularmente, entre las micro, pequeñas y medianas empresas (MIPYME´s) con un sólo objetivo:

**Fortalecerse para producir y comercializar sus productos y/o servicios en mercados de alta competitividad.**

##### Datos de contacto:

Lic. Hugo Hernández  
Director de Articulación Estratégica.  
52296100 ext. 32065

##### Objetivos

* Elevar el poder de negociación de las micro, pequeñas y medianas empresas en los mercados de proveeduría, comercialización, financieros y tecnológicos, entre otros.
* Consolidar su presencia en el mercado interno e incrementar su participación en  
  el de exportación.
* Fomentar la especialización de las empresas asociadas en productos y procesos que cuenten con ventajas comparativas.

##### Beneficios

* Elevar la competitividad de las empresas asociadas.
* Crea economías de escala para sus asociados, derivadas de una mayor capacidad de negociación para comprar, producir y vender.
* Se facilita al acceso a servicios especializados a bajo costo, ya que éstos se pagan de manera común entre los asociados.
* Evita la duplicidad de inversiones, al estar en condiciones de adquirir a nombre y cuenta de sus asociados y de forma común activos fijos, incluso de renovar o innovar la maquinaria y el equipo.
* Propicia la especialización de los asociados en determinados procesos y productos con ventajas comparativas.
* Favorece la concurrencia a mercados más amplios al consolidar la producción de sus asociados.
* Los empresarios se dedican a producir, mientras la empresa integradora se encarga de realizar gestiones y actividades especializadas que requieren los asociados.
* Mantiene la individualidad de los empresarios en las decisiones internas de sus empresas.
* Esta forma de organización empresarial tiene flexibilidad para adaptarse a cualquier actividad económica.

##### Requisitos

Presentar en la Secretaría de Economía la siguiente documentación:

* Proyecto de acta constitutiva conforme a la Ley General de Sociedades Mercantiles y al Artículo 4o. del Decreto que promueve la organización de Empresas Integradoras publicado en el Diario Oficial de la Federación el 7 de mayo de 1993 y sus modificaciones del 30 de mayo de 1995 (Ver anexo en la Caja de Herramientas, en la sección Normatividad).
* Proyecto de factibilidad económica – financiera que sustente la integración.

##### Apoyos

###### Apoyos Fiscales

La empresa integradora deberá acogerse al Régimen General de Ley sin límite de ingresos.

###### Apoyos Financieros

La Secretaría de Economía a través del INADEM, apoyará proyectos productivos integrales para empresas dadas de alta en este esquema.

###### Apoyos a la Exportación

Cuando los socios tienen una vocación exportadora, las empresas integradoras también tienen facilidades para acceder a programas para la promoción de las exportaciones.

###### Limitantes

* No pueden ser socios de una empresa integradoras, aquellas personas físicas o morales que no estén formalmente constituidas.
* La empresa integradora actúa exclusivamente por cuenta y orden de sus socios; por tanto, sólo es su representante ante clientes y proveedores.
* No puede realizar ninguna parte de proceso productivo que involucre la actividad de sus asociadas.
* Los activos que se adquieran para el proceso productivo de las socias, no pueden ser propiedad de la integradora, ya que son los socios los que realizan la actividad de producir y, por tanto, son los dueños de la maquinaria y el equipo.
* No se permite constituir empresas integradoras con empresarios que realicen prácticas desleales o tengan actividades monopólicas o que su actividad los conduzca a estas situaciones.

### Privados

#### Kickstarter

Kickstarter es una plataforma que provee a artistas, músicos, cineastas, diseñadores y otros creadores con los recursos y el apoyo necesarios para llevar sus ideas a la realidad. Hasta la fecha, decenas de miles de proyectos creativos de todos los tamaños han podido ser realizados gracias al apoyo de la comunidad de Kickstarter.

Cada proyecto de Kickstarter debe incluir lo siguiente:

* La página del proyecto con un vídeo y una descripción que explique claramente la historia detrás de tu proyecto
* Las recompensas que recibirán los patrocinadores cuando finalice el proyecto
* Actualizaciones sobre el proceso creativo a medida que tu proyecto se hace realidad

Si tu proyecto es […] algo más complejo es necesario que incluyas más detalles, para que la gente vea que sabes lo que estás haciendo. En ambos casos, la página de proyecto debe de proveer respuestas a unas cuantas preguntas básicas:

##### ¿Quién eres?

Preséntate. Presenta a tu equipo y trabajos similares que realizaste en el pasado

##### ¿Qué quieres crear?

¡Cuántos más detalles, mejor! Incluye bocetos, muestras, prototipos y contagia tu entusiasmo a los patrocinadores.

##### ¿Cómo surgió la idea para este proyecto?

Cuenta cómo comenzó todo y hasta dónde has llegado en este momento. De esta forma, los patrocinadores pueden entender qué tipo de trabajo ofreces y cómo lo manejas.

##### ¿Qué es el plan para tu proyecto, y en qué margen de tiempo?

Establece un cronograma claro y específico para los patrocinadores.

##### ¿Qué presupuesto tienes?

Un simple desglose demuestra que has pensado bien las cosas y que tienes un plan razonable, es decir, eres una persona fiable que usará los fondos de los patrocinadores de una forma adecuada.

##### ¿Por qué este proyecto?

Transmite tu entusiasmo sobre el proyecto a la gente, hazle sentir tu compromiso de llevarlo a cabo.

A continuación, se presenta una introducción al sistema desarrollado para este proyecto.

# Introducción

## Contexto de la problemática detectada

### Marco de referencia

“Hasta hace unos años, si un equipo de desarrolladores de una empresa estaba buscando una mejor forma de crear tecnologías para cubrir las cambiantes necesidades del negocio, lo que hacía era buscar en la web, en páginas de distintos vendedores.” [12]

Durante años IBM solamente era capaz de ofrecer software a través de sus representantes de ventas, los cuales, durante ese tiempo, se convirtieron en la herramienta fundamental sobre la que dependía plenamente IBM.

En 2014 comenzó la transición de ese modelo de negocios a un modelo ya consolidado para la época tal como lo es la nube. En abril de ese año se presentó el Cloud Marketplace.

Una tienda central para ofrecer servicios propios de la IBM y de terceros de una forma en que el usuario final tenga un acceso directo sin la necesidad de pasar por un representante de ventas. Esto agiliza el proceso de venta de productos y comienza una nueva forma de negocios para la IBM tal como SaaS (Software como un servicio, por sus siglas en inglés).

“Los clientes pueden registrase en Cloud Marketplace y descubrir, probar y experimentar online cientos de servicios cloud abiertos, escalables y seguros de IBM y de empresas asociadas. El contenido se presenta de forma dinámica, según las funciones y las páginas de servicios ofrecen un acceso sencillo e intuitivo para aquellos que estén interesados en categorías concretas como pueden ser Start-ups, Movilidad o Juegos, entre otras.

Hasta ahora, si un equipo de desarrolladores de una empresa estaba buscando una mejor forma de crear tecnologías para cubrir las cambiantes necesidades del negocio, lo que hacía era buscar en la web, en páginas de distintos vendedores. Ahora, con el mercado cloud de IBM, podrá evaluar, probar en el momento y adquirir las aplicaciones o bien contactar con IBM para realizar una compra.” [12]

# Descripción del problema detectado

## Contenido

La infraestructura usada por el Cloud Marketplace de IBM era la misma que utilizaban los representantes de ventas, por lo que para la compañía se volvió muy costoso en términos de desarrollo, mantenimiento y soporte. Esto en gran parte al tener que manejar volúmenes masivos de información a través de cuatro instancias:

* **ODS** (Operational Data Store): Es una data warehouse o repositorio de información en donde confluyen datos operacionales de distintas fuentes, como SAP, Sistemas de clientes, Precios, etc., que a su vez se conecta con otras aplicaciones para que sea consumida esa información. Es considerado una fuente confiable ya que cuenta con certificación ASCA, una herramienta de conformidad utilizada por la IBM. [13]
* **ESB** (Enterprise Service Bus): es un middleware que envía y transforma datos entre aplicaciones
* **DST** (Dedicated Service Tools): es un ambiente de servidores entorno de servidor
* **CMC** (Call Management Center): es una aplicación integral para la gestión de pedidos que ofrece soluciones a los desafíos que se encuentran en los típicos escenarios empresariales. Este es el front end donde usualmente los representantes de ventas gestionan sus cotizaciones. [14]

Contar con cuatro instancias provocaba para la compañía un riesgo mayor ya que eran demasiados puntos de fallo, creando cierta inestabilidad para el Cloud Marketplace, el cual por sus características y la necesidad de permanecer disponible todo el tiempo para así ser capaz de ofrecer sus servicios a básicamente todo el mundo, es algo inaceptable.

Las ventajas de la creación de un nuevo sistema son básicamente la posibilidad de mantener una conexión directamente entre los datos del mainframe de SAP y la base de datos en Cloudant, al ser una sola instancia la que se encargue de todas las actividades se vuelve mucho más sencillo su desarrollo, mantenimiento y soporte, lo cual implicaría una disminución de recursos para la compañía y un aumento gradual en la entrega de los datos hacia el Cloud Marketplace.

Al ser una solución completamente nueva se ofrece la infinita posibilidad del uso de nuevas tecnologías.

El uso anterior de las cuatro instancias por parte de IBM era por cuestiones de seguridad, la cual era y es la máxima prioridad, por lo que al hacer uso de DataStage se mantienen los protocolos y prioridades utilizando un certificado de seguridad de la capa de transporte (TLS por sus siglas en inglés).

## Objetivos

Los objetivos principales por conseguir en la solución diseñada son:

* Conectar ODS con Cloud manteniendo la transmisión segura (para IBM es muy importante la seguridad de los datos por regla de conformidad, zonas de colores)
* Mantener los datos siempre disponibles en el Marketplace y transfiere los datos de forma inmediata para ser consumidos por los clientes
* Que la aplicación sea auto recuperable, es decir, que cuando falle, minimice la dependencia de recursos, es decir que cuando falle la aplicación será capaz de reenviar los datos pendientes.

Paralelamente a los objetivos a conseguir se encuentran las siguientes restricciones para el desarrollo de la solución.

## Restricciones

* Cloudant no es capaz de acceder directamente a la base de datos de DB2. DB2 debe ser quien envíe la información a Cloudant.
* Todos los datos obtenidos como resultado de la consulta deben ser recibidos por la API de Cloudant a través de un solo fichero JSON.
* El sistema debe ser capaz de ejecutarse desde cualquier fecha especificada a través de un parámetro.
* El archivo JSON debe ser capaz de partirse en lotes de máximo 10,000 registros para mantener una transmisión ligera mediante la API de Cloudant.
* El archivo JSON debe cumplir con las especificaciones de formato provistas por el cliente.
* Uso de tecnología de IBM.

## Tabla comparativa con otros sistemas encontrados

No fue necesario consultar otros sistemas debido a que esta es una solución completamente nueva, diseñada específicamente para funcionar como una herramienta de ETL (Extract, Transform and Load, por sus siglas en inglés) mediante el uso de dos bases de datos SQL y no-SQL.

## Módulo por diseñar

### **Módulo I y II:** Gestión de la información

Se hace uso de un almacén de datos en el sistema de administración de bases de datos propio de IBM: DB2 donde se almacenan cerca de 100 millones de registros pertinentes a las cotizaciones, los cuales serán parte de la información a enviar a la segunda base de datos utilizada en el proyecto la cual reside en Cloudant, un servicio de base de datos (DBaaS por sus siglas en inglés) la cual es una base de datos no relacional que almacena datos en forma de documentos JSON.

La base de datos de Cloudant debe mantenerse actualizada en tiempo real y estar disponible 24/7, por lo que es indispensable mantener un flujo contante de datos.

### **Módulo III:** Sistemas distribuidos, paralelos o robustos

El proyecto está alojado en un servidor con capacidad capaz de proveer el servicio y la ejecución repetitiva de la solución.

Se genera un archivo que lista (mejor conocido como Log file) donde se registren todos los eventos y las acciones realizadas durante el proceso de ejecución, acciones tales como: finalizaciones exitosas, errores y excepciones, de tal forma se puede mantener un registro de los eventos ocurridos para prevención de errores o mantenimiento.

La solución desarrollada hace uso de tecnologías web para poder transferir cantidades masivas de datos, las misma son una WEB API donde el proyecto se comunica para enviar los múltiples ficheros JSON, que a su vez los insertará en la base de datos de Cloudant.

### **Módulo IV:** Softcomputing

Se utiliza un mainframe que viene de sistemas en SAP, de donde DB2 obtiene los datos, a partir del cual se genera un almacén de datos.

Al contar con un volumen masivo de datos es menester hacer uso de algoritmos de minería de datos capaces de:

Hacer comparaciones en tiempos mínimos y óptimos

Permitir detectar modificaciones entre el almacén de datos y la tabla base utilizada en el proyecto

Mantener uniformidad en los datos

Procesar los datos

Con estos mismos algoritmos se busca prevenir enviar datos ya existentes en Cloudant y evitar su duplicación.

# Metodología para resolver el problema detectado

## Tipo de investigación

### Investigación cuantitativa

La necesidad de hacer uso de una investigación cuantitativa radica en el hecho de adquirir conocimientos fundamentales y la elección del modelo más adecuado que se capaz de obtener la realidad de una manera imparcial, ya que se obtienen y analizan datos a través de los conceptos y variables. [15]

## Definición del tamaño y la muestra de la investigación

El tamaño de la muestra es de 385 para una población de donde el nivel de confianza es de y el margen de error es de

Para obtener el resultado del tamaño de la muestra fue aplicada la siguiente fórmula matemática [16]:

En donde representa al valor del tamaño de la población, al valor del margen de error y es la cantidad de desviaciones estándar que una proporción dada se aleja de la media, para ello fue utilizada la tabla 3:

|  |  |
| --- | --- |
| Nivel de confianza deseado | Puntuación Z |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Tabla 3: Valor de Z

## Consulta y acceso a repositorios de datos

Los datos son obtenidos de múltiples bases de datos que residen dentro del almacén de datos de ODS Interfaces, donde la fuente de estas múltiples bases de datos son mainframes de SAP que a su vez son datos generados por los mismos representantes de ventas, es por ello por lo que absolutamente todos los datos son propiedad de la IBM.

Los factores discriminatorios para la elección de los datos son aquellos que pertenecen a los distintos requerimientos, que en este caso son los siguientes:

* Clientes
* Encabezados de las cotizaciones
* Detalles de las cotizaciones
* Distribuidores
* Vendedores
* Contactos

## Estándares usados

Las cuentas dentro de la IBM generalmente desarrollan un acuerdo de seguridad basado sobre en las normas y estándar del ISO 27000 y se mantienen documentos de seguridad granulares bajo CSD e ITCS104 que cubren los sistemas operativos específicos y la seguridad de las aplicaciones. El DPE debería poder dirigirse a la cuenta de seguridad focal.

**IBM**, al ser una compañía con presencia mundial desarrollas sus propios estándares debido a que debe cumplir integralmente en todos y cada uno de los países con presencia, por lo que no existe uno solo en específico que no haya sido utilizados, de tal manera Glezcorp imitó las políticas de su principal cliente para llevar a cabo un exitoso proyecto Languis.

Aunado a esto, para la realización de la solución del proyecto Languis se utilizó la metodología de trabajo ágil (Agile), en donde el requerimiento principal se dividió en diversos features donde se definen paso a paso las fases a seguir.

Una feature dura tres iteraciones, una iteración dura 10 días en donde se trabajó con diversas historias de usuario (user stories) con la finalidad de trabajar con pequeños módulos para lograr un cometido final.

Cada historia de usuario se documenta sobre lo que se hizo para llegar al criterio de aceptación de la misma de forma que se expliquen los problemas que existieron, las soluciones y la investigación realizada si es que fue necesaria para completar la historia de usuario, al igual que el código fuente que fue desarrollado durante la iteración.

Todo lo que se trabajó se hizo dentro de cuatro ambientes de desarrollo, los cuales son: desarrollo (Development), pruebas (FVT – Functional Verification Testing), pruebas de usuario (UAT – User Acceptance Testing) y producción (Production), de esta forma se garantiza entregar un producto final probado y eliminando las posibilidades de errores, y a pesar de que estos pudiesen aparecer una vez en producción, las probabilidades a que sucederán serían menores si no existiera la etapa de pruebas.

Como parte de la seguridad requerida por IBM, al trabajar con distintos tipos de información de carácter confidencial y sensibles se llevan a cabo prácticas que sean capaces de cumplir inclusive con el reglamento general de protección de datos [17] (GDPR por sus siglas en inglés) de la Unión Europa, esto debido a que IBM es una compañía global y como tal Glezcorp debe cumplir con los estándares de seguridad requeridos por sus clientes en la totalidad de su presencia, de tal forma que las políticas más importantes en la rama de seguridad se cumplen, tales políticas son:

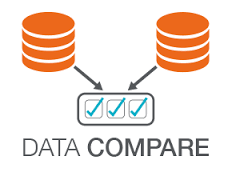
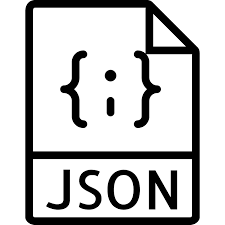
* Derecho al olvido
* Reducir las brechas de riesgos o fugas de información sensible

Glezcorp también se compromete con las políticas internas de seguridad de la propia IBM.

# Diseño de la propuesta

La propuesta para la solución de este proyecto es utilizar el software especializado para ETL (Extract, Transform and Load) propio de IBM: DataStage en su versión 11.3 la cual integra la función de datos jerárquicos, necesaria para la transferencia de los datos a través de uno o varios archivos JSON en conexión con una API.

Se presenta un diagrama explicativo de la solución en la figura 3 a continuación:



Productos

Contactos

Clientes

Cotizaciones

Figura 3: Diseño de la solución

Como se puede apreciar, los distintos datos relacionados a:

* Cotizaciones,
* Productos,
* Clientes y
* Contactos

Son almacenados en distintas tablas del mainframe de SAP, por lo que se obtienen a través de una única consulta compleja.

Una vez obtenidos estos datos, pasan por la etapa de la data compare donde haciendo uso de la herramienta de IBM: DataStage 11.3, se utilizan los diferentes algoritmos de minería de datos para obtener todos los registros que coincidan con diferentes especificaciones dadas por el cliente y que pertenezca a un cierto rango de fechas, que de igual forma es especificada por el cliente.

Una vez procesada toda la información la misma herramienta de DataStage es capaz de transformar esos datos, crudos como tal, en un archivo JSON.

Dependiendo de la cantidad de datos en el archivo JSON se toma la decisión de dividir en lotes si es que el mismo excede los 10,000 registros, de tal forma se logra el cometido de hacer transmisiones ligeras y siempre enviando todo a través de la API de Cloudant en una sola transmisión.

## Arquitectura

Existe un almacén de datos dentro de la instancia la cual mantiene toda la información, desde aquí se obtiene los datos a través de los algoritmos de minería de datos.

Se hace una consulta completa a 17 tablas las cuales proporcionaran todos los elementos requeridos para el Cloud Marketplace.

* Base de Datos: DB2
* Tablas:
* Cotizaciones web
* Configuración de cotizaciones web
* Elementos de cotizaciones web
* Dimensiones de contactos
* Dimensiones de productos
* Código de descripción
* Maestro de código de referencia de país
* Descripción de códigos
* Clientes
* Identificador de clientes
* Cotizaciones
* Elementos de cotizaciones
* Orden de compra de elemento
* Usuarios
* Identificador de información de producto
* Unidad de paquete de carga
* Unidad de paquete de carga de producto

Se obtienen los datos deseados haciendo una consulta completa (Complex Query) a todas las tablas, de tal forma se obtiene un solo conjunto de datos para de esa forma hacer un tratamiento de los mismos de forma más simple.

## Características

La solución creada es capaz de:

* Mantener un registro del tiempo de la última ejecución de la solución.
* Dividir en lotes los archivo JSON cuando este contenga más de 10,000 registros.
* Notificar al momento de finalización si la ejecución de la solución fue exitosa o no.
* Utilizar una API para enviar los datos que serán añadidos a la base de datos de Cloudant.
* Intentar una conexión al menos cinco veces en caso de que a conexión con la API de Cloudant no se encuentre disponible antes de terminar la ejecución

## Modelado de la base de datos

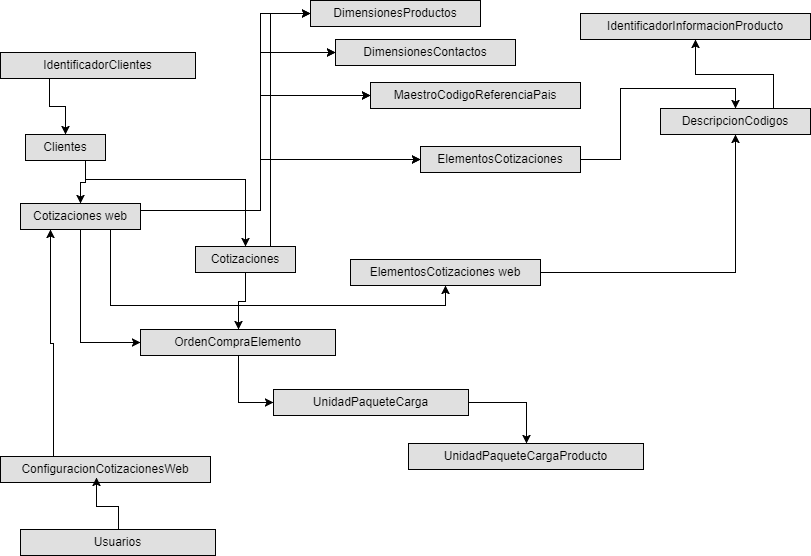


Figura 4: Diagrama entidad relación DB2

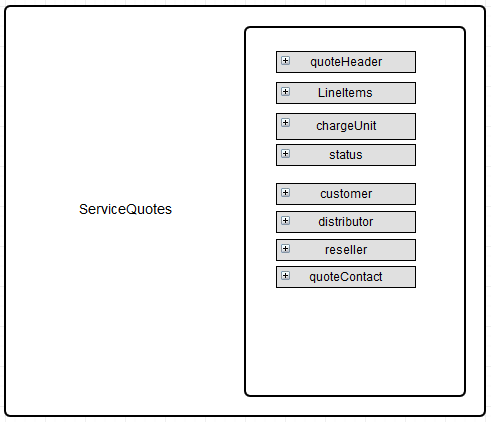


Figura 5: Documento No-SQL Cloudant

## Diseño conceptual

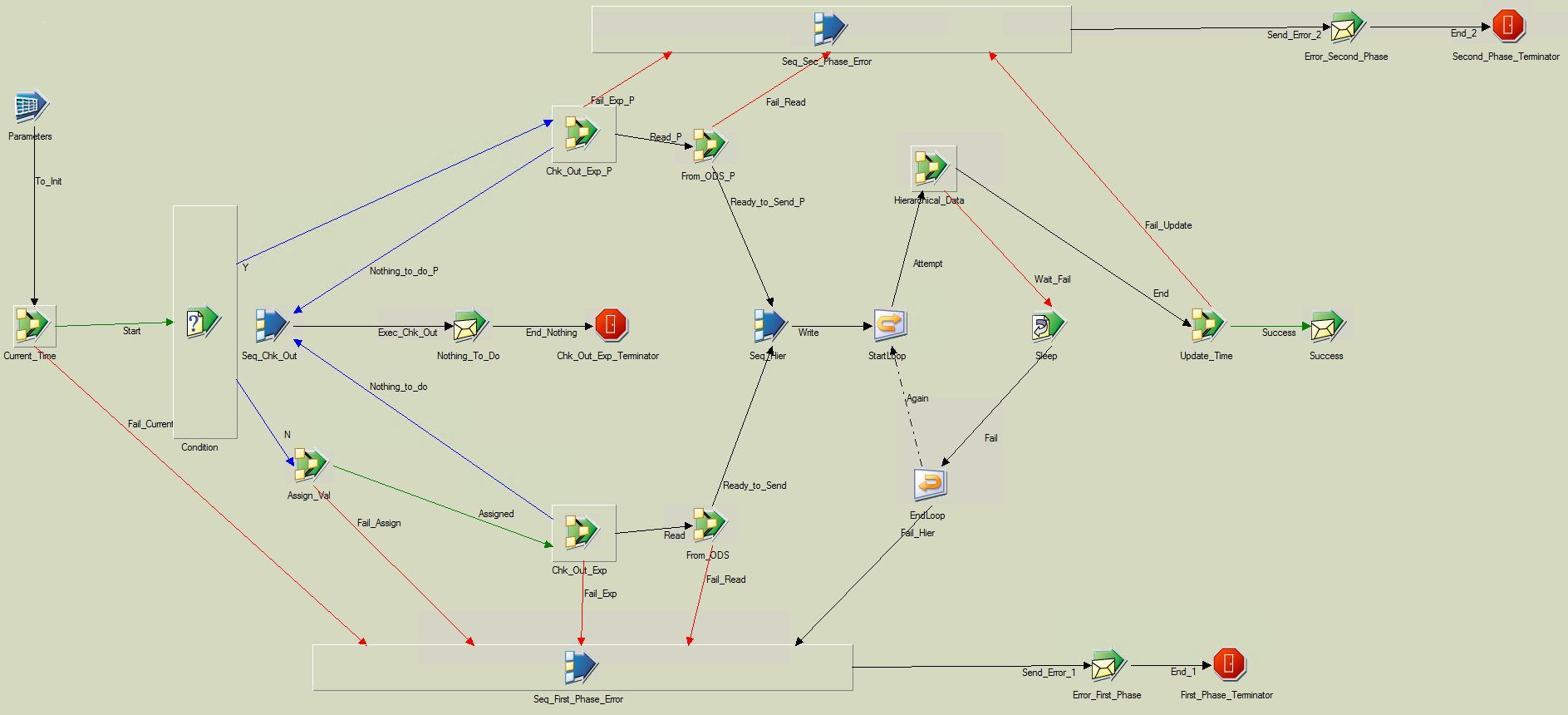


Figura 6: Diagrama de flujo Checkout Experience

A continuación, se describe el sistema mediante diagramas de casos de uso, colaboración y estados de forma que se pueda comprender, a groso modo, el funcionamiento.

### Primer caso

En este primer caso se muestra uno de los principales casos de uso: cuando el administrador del sistema ejecuta el sistema a partir de una fecha específica, en donde, el sistema a su vez obtendrá los datos desde esa fecha especificada y comparará con los datos ya cargados en la tabla base, de esa forma se obtendrán todas las diferencias, las cuales pueden ser datos nuevos o actualizados.

Una vez actualizada la tabla base se procede a convertir esos datos actualizados a texto plano, durante este proceso de conversión de datos se aplica un algoritmo para permitir al sistema dividir los datos en lotes, que a su vez serán convertidos a uno (o varios, dependiendo de la cantidad de datos a enviar) archivo(s) JSON listo(s) para ser enviado(s) a la base de datos de Cloudant a través de una API.

Realizada esta conexión con la API y enviado los datos en uno o más archivos JSON el cliente es capaz de consumir esos datos desde Cloudant mientras que el sistema actualiza una tabla de registros la reciente ejecución exitosa y finalmente, notifica, a través de correo electrónico, a una lista de distribución que ha finalizado el proceso exitosamente y en el mismo correo envía el registro de todo lo realizado durante el proceso.

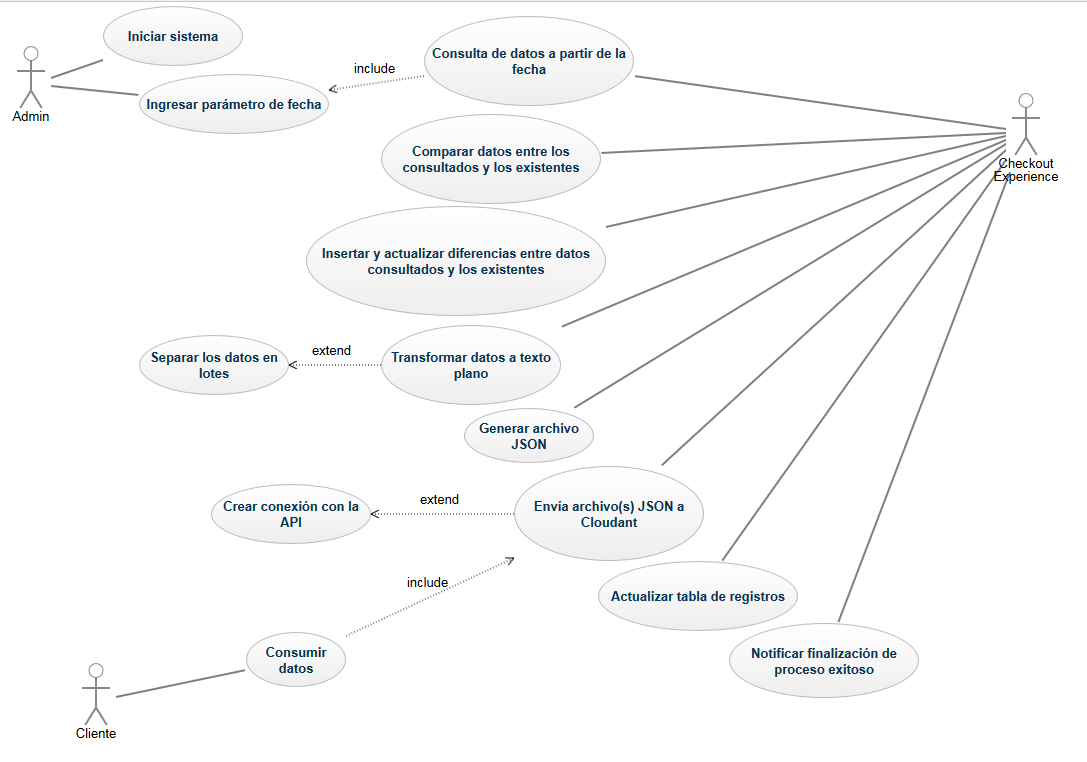


Figura 7: Diagrama de caso de uso

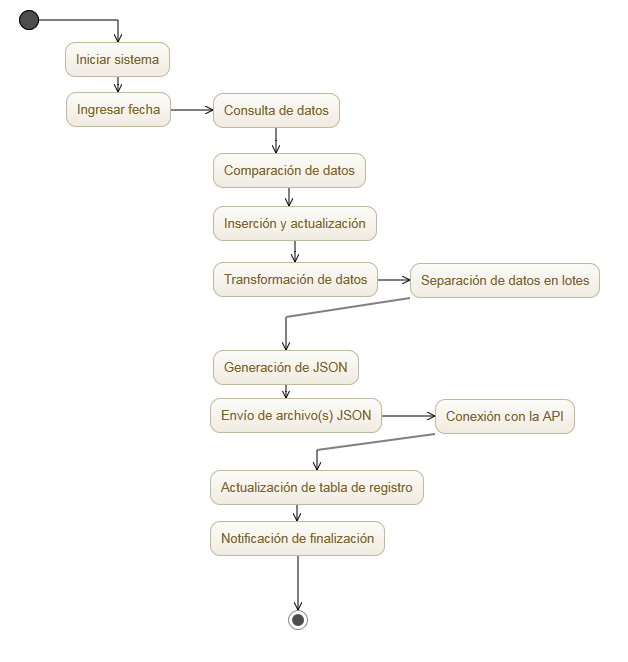


Figura 8: Diagrama de actividades

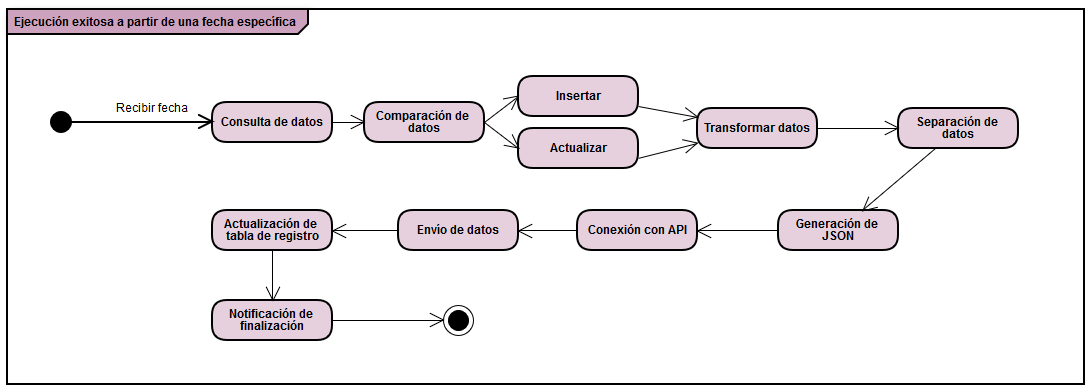


Figura 9: Diagrama de estados

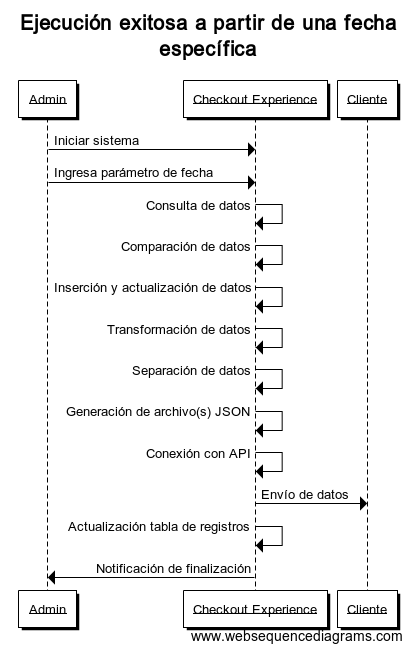


Figura 10: Diagrama de secuencia

### Segundo caso

La idea es que el sistema se ejecute no solamente cuando un administrador del sistema lo ejecute manualmente, sino, que corra automáticamente cada cinco minutos, para mantener actualizada la base de datos en, lo más cercano posible, tiempo real.

Para este caso se muestra una situación que sucederá por periodos específicos como fines de semana o días feriados porque no habrá datos nuevos que enviar a Cloudant debido a la inactividad que puede darse debido a lo anterior.

Por lo que en el caso de que no haya nada por actualizar ni enviar a Cloudant se notifica, a través de correo electrónico, a una lista de distribución que ha finalizado el proceso en el mismo correo se especifica que no hubo actividad por realizar.

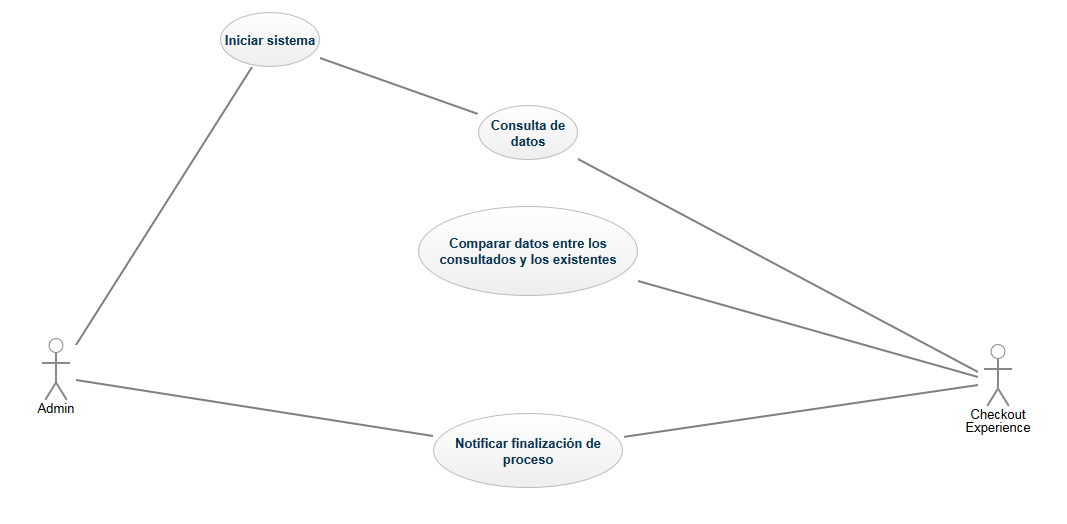


Figura 11: Diagrama de caso de uso

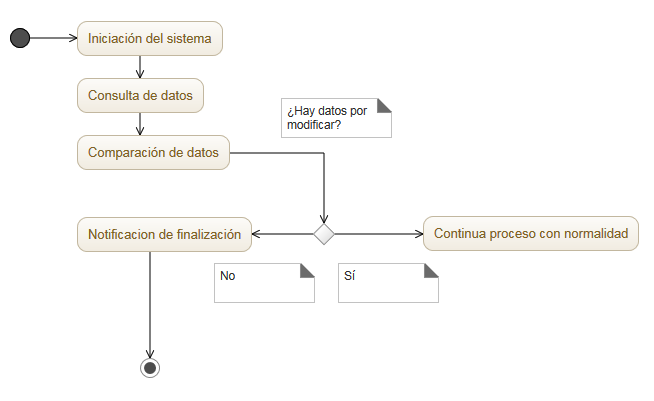


Figura 12: Diagrama de actividades

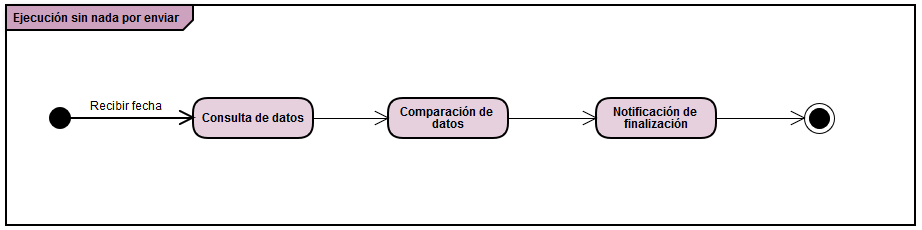


Figura 13: Diagrama de estados

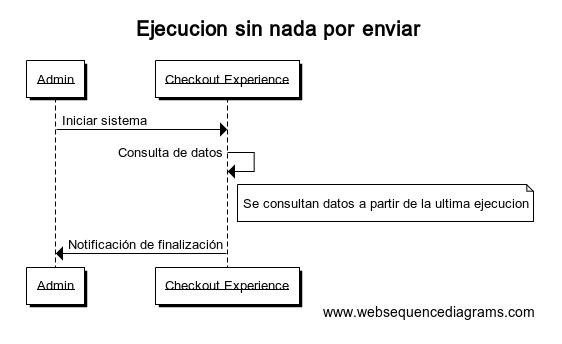


Figura 14: Diagrama de secuencia

### Tercer caso

Al hacer uso de una conexión con Cloudant mediante una API puede suceder situaciones externas al sistema como que la API no esté en línea o que no haya conexión a Internet por lo que en esos casos el sistema debe ser capaz de intentar la conexión, al menos, en cinco ocasiones antes de finalizar el proceso debido a la imposibilidad de conectar con la API.

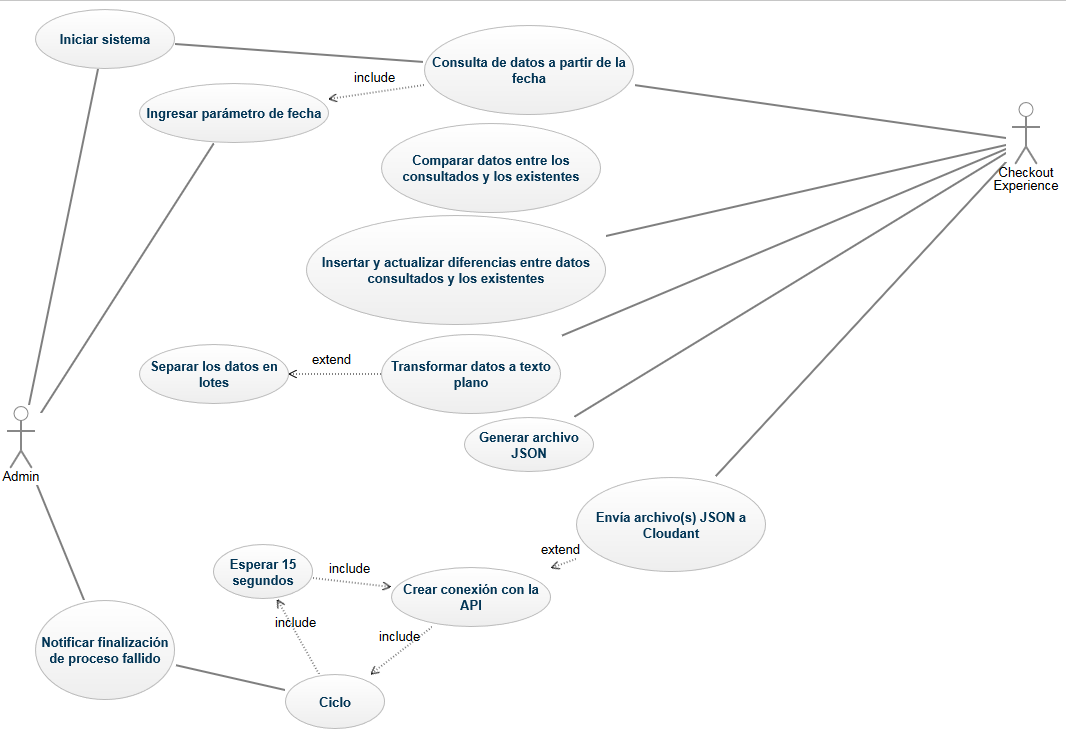


Figura 15: Diagrama de caso de uso

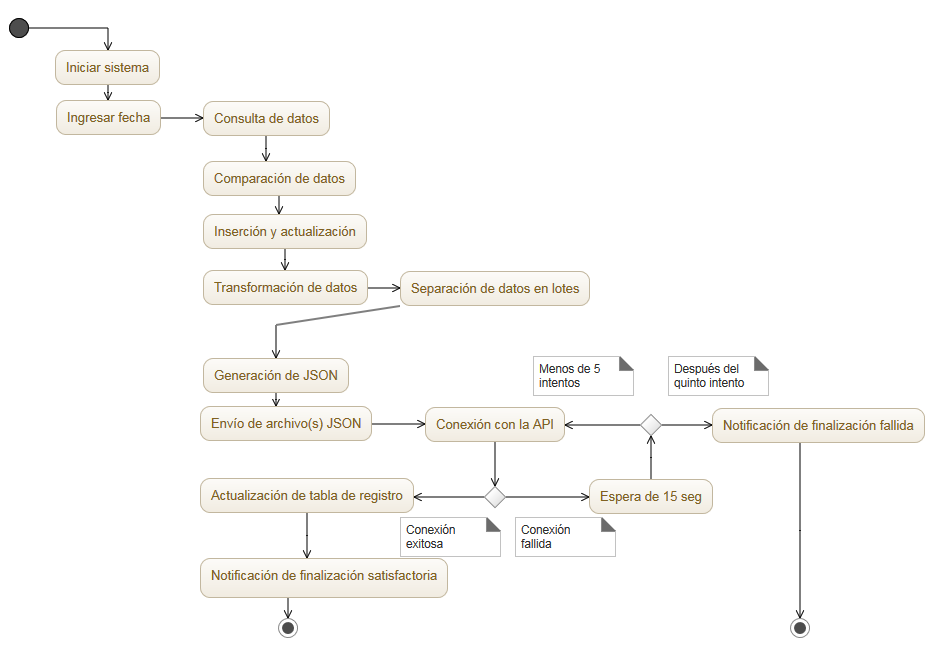


Figura 16: Diagrama de actividades

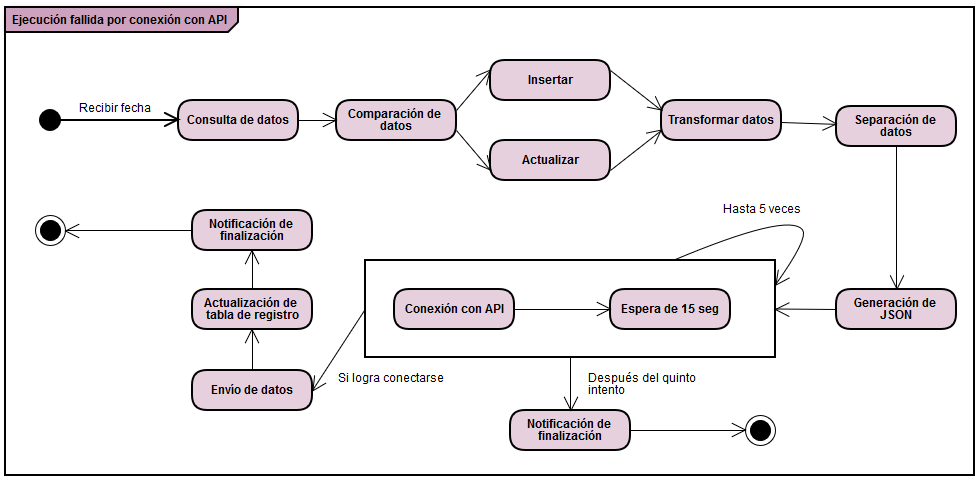


Figura 17: Diagrama de estados

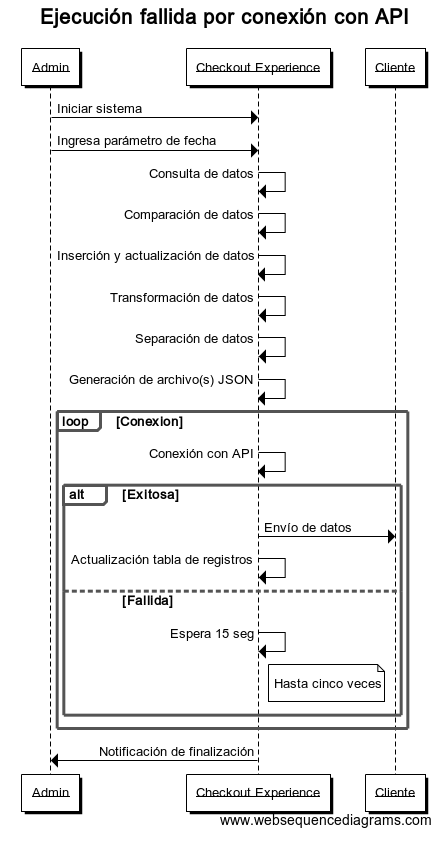


Figura 18: Diagrama de secuencia

## Dependencias con otros elementos de software o hardware de terceros

Al ser un requerimiento el trabajar con tecnologías de la propia IBM se listan a continuación los elementos de software utilizados para el desarrollo de la solución:

* IBM InfoSphere DataStage 11.3

Se utilizan los siguientes lenguajes de programación:

* **Basic**: mismo que fue utilizado en rutinas de servidor en la herramienta de DataStage
* **SQL**: utilizado para consultas complejas y estructuradas
* **JavaScript:** para la composición de los archivos JSON

Las bases de datos utilizadas son:

* IBM Db2 Database
* IBM Cloudant

DataStage como herramienta principal corre en una instancia donde está configurada una de las bases de datos de Db2 que utiliza como sistema operativo:

* AIX Server

Y para el desarrollo de la solución en DataStage se utiliza un diseñador con interfaz gráfica que utiliza como sistema operativo:

* Windows Server 2012R2

Finalmente, para la conexión para el envío de información entre Db2 y Cloudant se utiliza una API.

A continuación, será presentado el diagrama de actividades llevado a cabo para el desarrollo del proyecto Languis.

# Cronograma de actividades

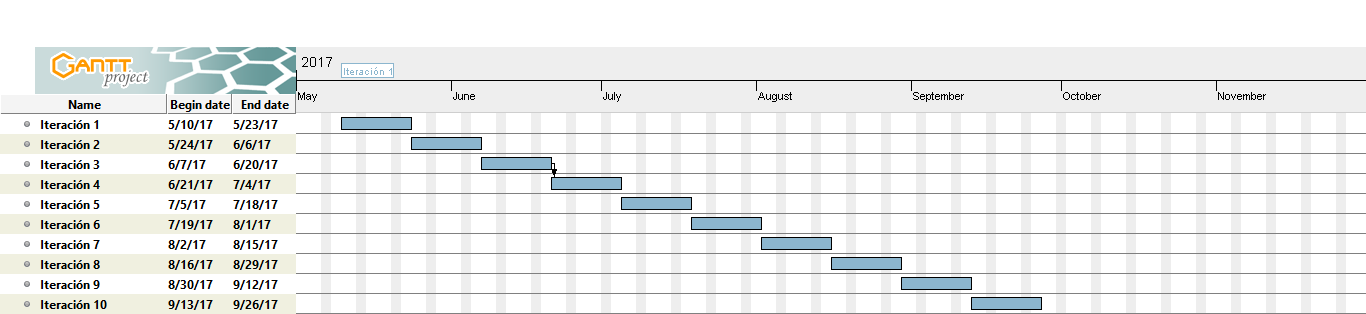


Figura 19: Diagrama de Gantt

## Iteración I

De 10 de mayo de 2017 al 23 de mayo de 2017

Como parte del proyecto Languis se necesita una solución técnica para llevar cotizaciones de los representantes de ventas de la compañía IBM.

Se requiere una prueba de concepto para el diseño de la solución con los siguientes requerimientos:

- Rápida replicación

- Alta disponibilidad de la tienda y del servicio

- Tiempo veloces para las consultas

Esta historia es para descubrimiento, se deben tener las siguientes consideraciones:

- Definir cuáles serán los criterios de selección para las cotizaciones

- Se necesita añadir campos adicionales

Preguntas para responder de la investigación:

- ¿Cómo será el estatus secundario de las cotizaciones para que el sistema SAP las administre?

- ¿Cuál será el canal de distribución a ser utilizado para obtener las cotizaciones de los representantes de ventas?

## Iteración II

De 24 de mayo de 2017 al 6 de junio de 2017

* Configurar el ambiente de desarrollo de DataStage, haciendo uso de su versión 11.3
* Crear una prueba de concepto para enviar información desde la base de datos de DB2 hacia la base de datos de Cloudant

## Iteración III

De 7 de junio de 2017 al 20 de junio de 2017

* Prueba de concepto para demostrar que se es posible conectar directamente la base de datos de Cloudant usando DataStage 11.3

## Iteración IV

De 21 de junio de 2017 al 4 de julio de 2017

* Seleccionar la arquitectura correspondiente para el flujo de datos
* Identificar los servidores a conectar para la base de datos de Cloudant
* Encontrar los puntos de control de acceso entre las bases de datos Db2 y Cloudant (Llaves públicas y privadas)

## Iteración V

De 5 de julio de 2017 al 18 de julio de 2017

* Revisar el diseño del proyecto para estandarizar la arquitectura del diseño de las tareas
* Añadir etapas de revisión de errores para solucionar problemas con las tareas, por ejemplo: campos vacíos, tipo de dato, conexión con las bases de datos.
* Añadir una columna para cada tipo de datos existente en DataStage con la finalidad de asegurar que se pueden manejar todos los tipos de datos.
* Configurar parámetros y variables de ambiente para facilitar la instalación y ejecución en diferentes ambientes

## Iteración VI

De 19 de julio de 2017 al 1 de agosto de 2017

Hasta ahora el proyecto en DataStage es capaz de construir un conjunto de archivos JSON con el siguiente formato:

"{

"serviceQuotes”: [ {

"NUMERO\_COTIZACION “: "000000",[[1]](#footnote-1)

"CODIGO\_COTIZACION”: "E0003",

"NUMERO\_PARTE”: "12345",

"MONEDA”: "EUR"

}, {

"NUMERO\_COTIZACION”: "000000",

"CODIGO\_COTIZACION”: "E0003",

"NUMERO\_PARTE”: "12345",

"MONEDA”: "CHF"

}, {

"NUMERO\_COTIZACION”: "000000",

"CODIGO\_COTIZACION”: "E0003",

"NUMERO\_PARTE”: "12345",

"MONEDA”: "USD"

} ]

}”

La siguiente etapa consiste en enviar ese archivo a través de la conexión a la API especificada para la conexión con la base de datos de Cloudant.

## Iteración VII

De 2 de agosto de 2017 al 15 de agosto de 2017

Se integrarán la siguiente lógica para que el proyecto de Checkout Experience sea capaz de:

* Añadir un parámetro que permita ingresar una fecha específica (timestamp), de tal forma que el proyecto pueda transferir información no solamente desde la última vez que se ejecutó, si no que desde cualquier fecha.
* Añadir lógica para manejar fallas de conexiones con la API de la siguiente forma:
* Durante la llamada a la API y si esta no responde exitosamente se intentará de nueva forma hacer la conexión esperando 15 segundos antes, de tal forma se reintenta conectar durante un tiempo estimado de 1:45 minutos:  
    
  a las 10:00:00 hrs – 1er intento – fallado  
  -- espera 15 segundos --  
  a las 10:00:15 hrs – 2do intento – fallado  
  -- espera 30 segundos --  
  a las 10:00:45 hrs – 3er intento – fallado  
  -- espera 60 segundos --  
  a las 10:01:00 hrs – 4to intento – fallado   
    
  tiempo total de espera: 1:45 minutos
* Añadir lógica para no actualizar cuando no exista ningún registro para ser enviado.
* Añadir capacidad de enviar por lotes los registros, de tal forma que sea una transferencia de información más ligera y eficaz: Si durante la ejecución del proyecto han sido seleccionadas 750 cotizaciones (registros), se enviarán ocho lotes, serán siete lotes de 100 cotizaciones cada uno y un único lote con el resto de las cotizaciones.

## Iteración VIII

De 16 de agosto de 2017 al 29 de agosto de 2017

* Para esta iteración ya estamos cerca de finalizar el desarrollo por lo que es necesario comenzar a enviar todo al ambiente de pruebas para validar que funcione correctamente aunado a lo siguiente:
* Hacer uso de una nueva tabla en donde se mantendrán los registros a ser enviados hacia Cloudant
* Todas las columnas que proceden de ODS deben ser enviadas hacia el IBM Marketplace
* Todos los registros en el archivo JSON deben estar convertidos en su tipo de dato concreto, por ejemplo:  
  {  
   “Nombre”: “tipo de dato string”,  
   “NumeroEntero”: 0,  
   “NumeroDecimal”: 0.0,  
   “Booleanos”: true  
  }
* Todas las columnas que funcionan como bandera deben mostrarse en el JSON con true o false, según sea el caso, por ejemplo:  
  Cero (0) -> “false”, Uno (1) -> “true”
* El proyecto al final de la iteración debe estar corriendo en los ambientes de pruebas (FVT y UAT).

## Iteración IX

De 30 de agosto de 2017 al 12 de septiembre de 2017

El producto final está casi listo, lo que se debe hacer para el resto de la iteración es:

1. Crear un mapa de etiquetas en donde se pueda ubicar los nombres de los valores enviados a través del archivo JSON y las columnas de donde provienen esos valores
2. Continuar con las pruebas en el último ambiente (UAT) de tal forma que se puedan minimizar problemas de producción una vez que esté corriendo en este ambiente y siendo utilizados por el usuario final

## Iteración X

De 13 de septiembre de 2017 al 26 de septiembre de 2017

Para que el cliente final sea capaz de utilizar este proyecto es necesario realizar el movimiento al ambiente de producción para que pueda ser utilizado.

# Resultados obtenidos

## Análisis de resultados

Se ha implementado una solución robusta, capaz de cumplir cada uno de los objetivos especificados de tal forma que cumple a la medida y resuelve un problema.

El sistema, Checkout Experience, es capaz de conectarse a los almacenes de datos de la IBM que viven en la interfaz de ODS y transmitir a su vez, de forma segura y cumpliendo con todos los requerimientos de conformidad establecidos por la compañía, a una base de datos jerárquica como lo es IBM Cloudant.

Gracias a esto es posible mantener casi en tiempo real lo datos siempre disponibles para ser consumidos por los representantes de ventas y clientes de la IBM, ya que una de las funcionalidades del sistema es que sea capaz de ejecutarse en periodos determinados de tiempo, automatizando la actividad de ETL de Checkout Experience.

Otra de las características implementadas en esta solución es la capacidad de ser auto-recuperable, esto debido a que durante su ejecución pueden ocurrir diversas situaciones, en donde podría no ser utilizable el sistema, tales como fallas con la API de Cloudant o que no existan datos para ser enviados, por lo que se prevén estas situaciones para que la intervención de un administrador de sistemas sea mínima, básicamente, limitándose a ejecutar manualmente el sistema por necesidades específicas, tales como, ejecutar a partir de cierta fecha.

Los algoritmos de minería de datos implementados para la solución son un factor determinante en el desarrollo satisfactorio de la misma, ya que es debido a ellos que Checkout Experience puede recuperar hasta millones de registros y manipularlos a su vez dentro de un intervalo de tiempo significante por lo corto que resulta.

La capacidad de notificar a una serie de individuos sobre el resultado de la ejecución es algo que permite su adecuada monitorización sin necesidad de estar interviniendo directamente con el sistema, ya que estas notificaciones suceden mediante correo electrónico, en ese mismo correo enviado tras su ejecución, sin importar si es exitosa, fallida o sin actividad por realizar, incluye un registro de todo lo que hizo Checkout Experience desde su inicialización hasta su finalización

## Capturas de Pantalla

### Gestión de la tecnología de la información

En este módulo se muestran como los datos están y han sido integrados, respectivamente en DB2 y Cloudant.

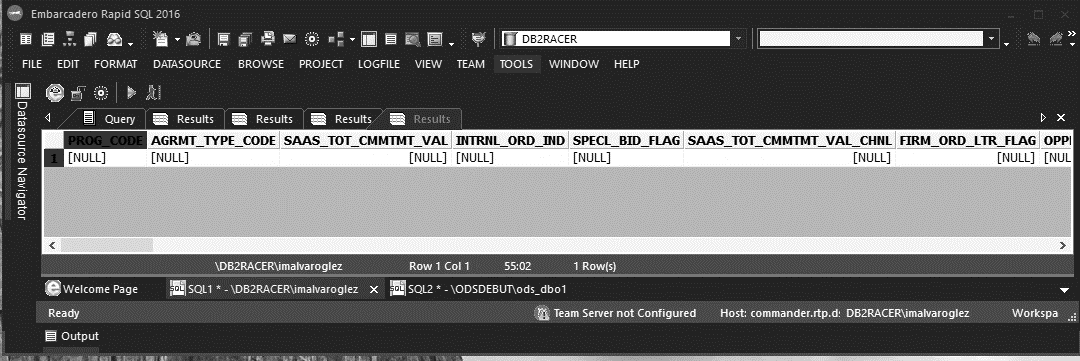


Figura 20: Datos en DB2

En la figura 20 se muestra un registro obtenido de una consulta compleja ejecutada directamente en un sistema gestor de bases de datos, para finalidad de demostración y por cuestiones de conformidad y privacidad de los datos solamente se muestra el formato en que los datos son consumidos por parte de Checkout Experience.

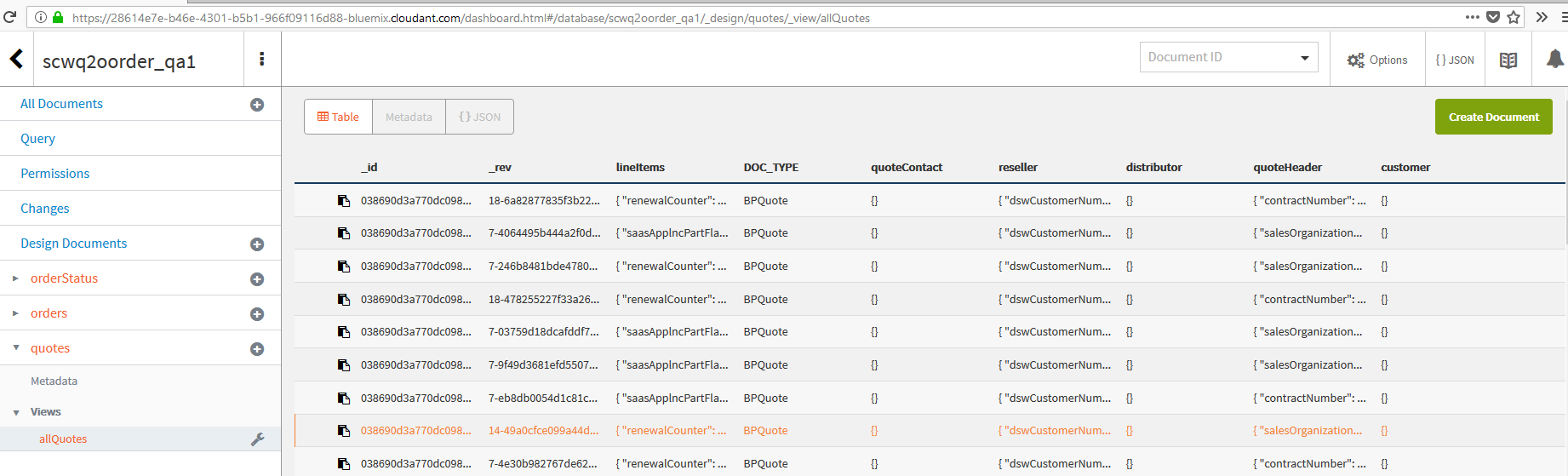


Figura 21: Registros en IBM Cloudant

En la figura 21 se muestra cómo se ven los datos en la base de datos No-SQL de IBM Cloudant.

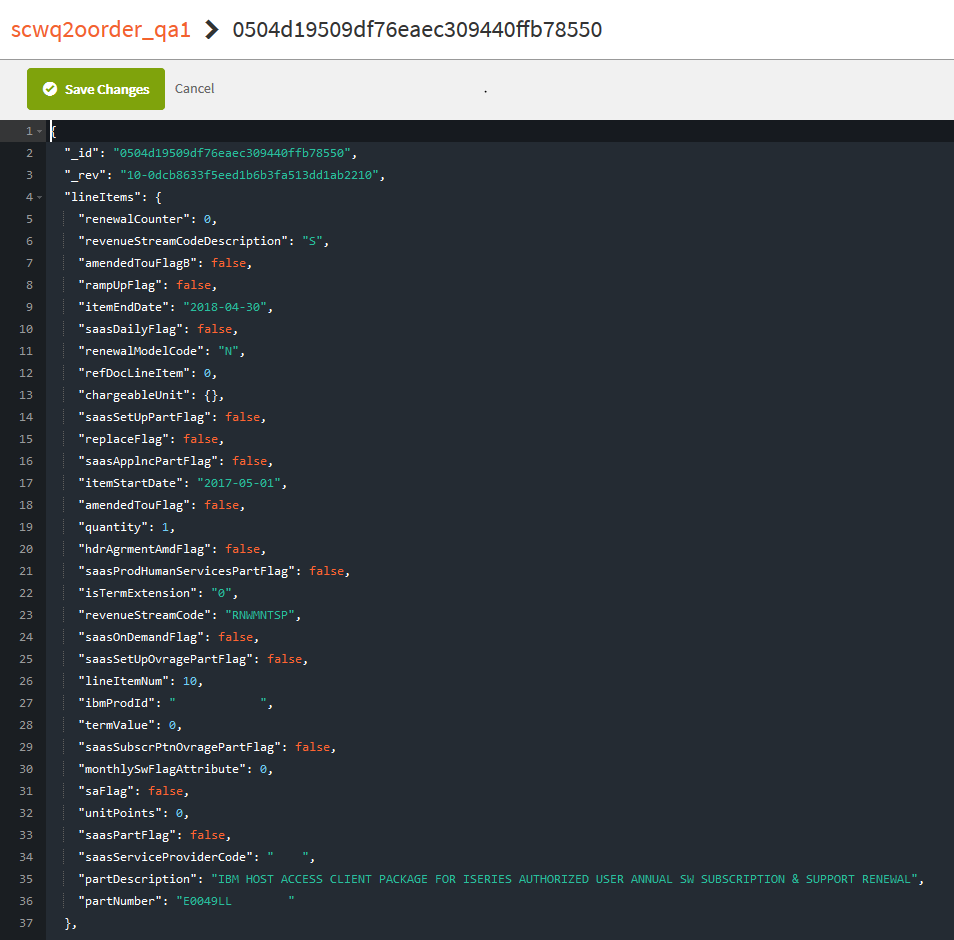


Figura 22: Registro en formato JSON en IBM Cloudant

Es en la figura 22 donde se confirma uno de los objetivos, que era mantener el tipo de dato en el archivo JSON. Los datos de tipo cadena o fecha se muestran en el documento entre comillas dobles, los datos de tipo entero o flotante se escriben por si mismos y los datos que equivalen n a banderas desde el origen ahora se muestran con un verdadero o falso.

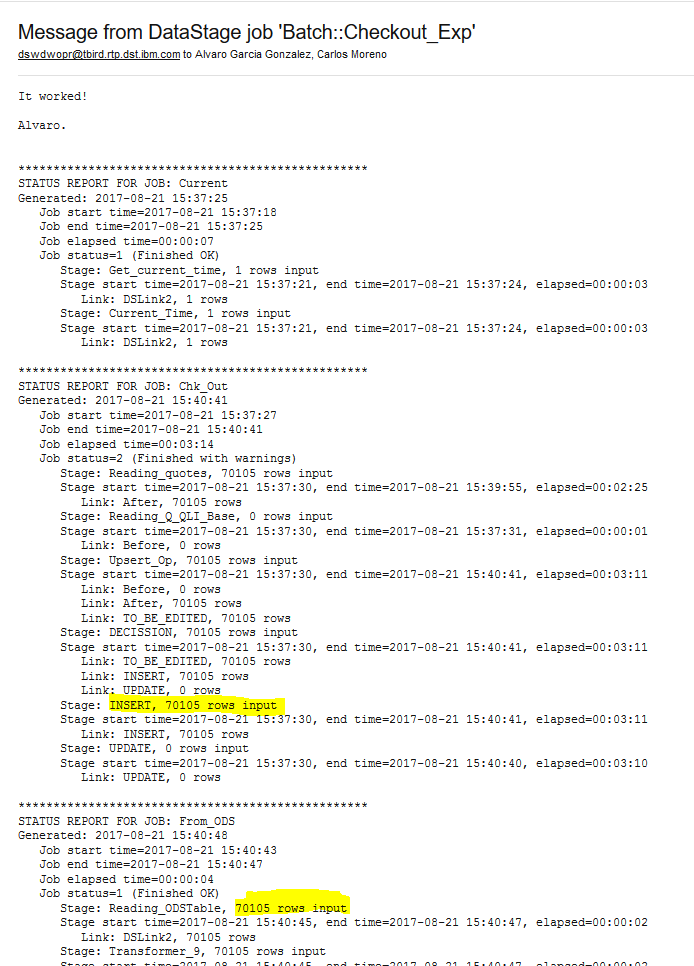


Figura 23: Correo electrónico enviado por Checkout Experience

En la figura 23 se muestra uno de los correos electrónicos enviados por el sistema, Checkout Experience, en donde se incluye el estatus final del mismo y un registro de todo lo que se ejecutó.

### Sistemas robustos, paralelos y distribuidos

En este módulo se presenta la solución implementada para lograr la conexión con Cloudant a través de la API y donde se utiliza la misma como medio de transporte para los archivos JSON que contienen los datos de las cotizaciones a ser utilizadas en el IBM Marketplace.

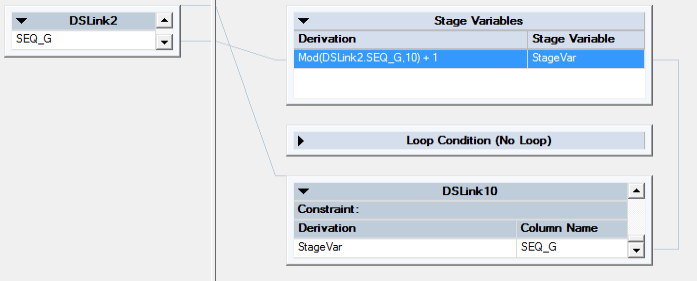


Figura 24: Etapa de conversión y previa separación de datos

Esta etapa es importante previo a la generación de archivo(s) JSON porque realiza la conversión de los datos a enviar hacia Cloudant a formato de texto simple y a su vez a marcarlos para que en la siguiente etapa puedan ser divididos en lotes.

Una de las columnas de los registros, “SEQ\_G”, es utilizada para almacenar el número de registro en la tabla base, este número se refiere, por ejemplo, a la función *número* en diversos sistemas gestores de bases de datos.

Previo a ser almacenado el registro en el archivo de texto plano se aplica un algoritmo creado para generar el módulo resultado de una división entre el valor de “SEQ\_G” para ese registro sobre (en este caso, el número determina que serán lotes por enviarse).

A ese resultado se le suma para que el resultado final a asignarse en la columna “SEQ\_G” sea entre y N (Donde N es la cantidad de lotes a generar, en este caso 10) con la finalidad de evitar que el índice sea 0 y facilitar la comprensión del usuario final al momento de consumir los datos.

Este resultado es almacenado en el archivo de texto plano.

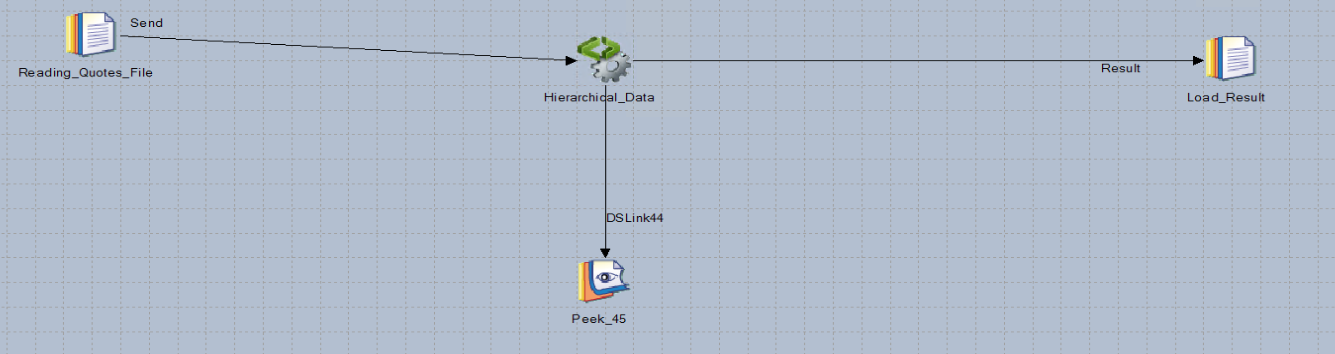


Figura 25: Etapa jerárquica de Checkout Experience

En la figura 25 se muestra la etapa en donde se obtienen los datos en texto plano a través del objeto *Reading\_Quotes\_File* que a su vez envía a través del flujo hacía el objeto *Hierarchichal\_Data*.

Este objeto es, sin duda alguna, uno de los pilares principales de la aplicación ya que realiza el siguiente proceso:

* Separación de los datos del texto plano en base a la columna “SEQ\_G”.
* Conversión de ese segmento de datos en un archivo de tipo JSON en base a un formato establecido, el cual va conforme a la necesidad de almacenar los datos en Cloudant por parte del cliente.
* Creación de la conexión a través de un puerto SSL con la API para generar una transmisión segura de los datos.
* Transmisión de cada archivo JSON generado.

Al finalizar este proceso el resultado del código HTTP de respuesta de la conexión, el archivo JSON enviado y la respuesta por parte de la API, se almacena en una variable de tipo *nvarchar* que es enviada a siguiente flujo para que el objeto *Load\_Result* genere un archivo de salida.

Un ejemplo de una prueba preliminar en el ambiente de desarrollo (Dev) con cuatro columnas por registro es mostrado en la figura 26.

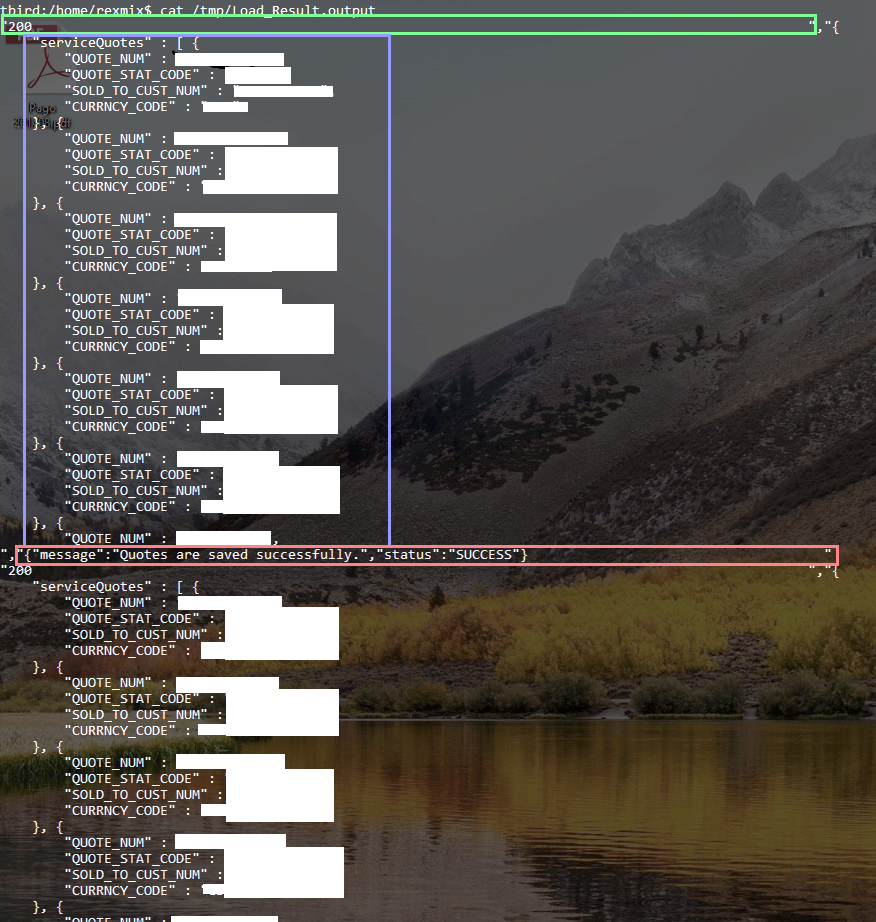


Figura 26: Archivo de salida

En la figura 26 se muestra un archivo de salida que es el resultado final de la transferencia de datos hacia Cloudant. Este archivo se divide en tres partes. La primera parte, con el color verde, es el código de respuesta HTTP de la API, este código nos sirve para detectar que clase de error ocurre en caso de fallos o para señalar la entrega satisfactoria de la información.

La segunda parte, señalada con el color azul, es un ejemplo de los datos enviados desde DB2, ya en formato JSON, hacia Cloudant. Aunado a esto se puede apreciar que se envían datos en lotes, con lo que se logra el objetivo de dividir los datos para una transferencia de datos más eficiente.[[2]](#footnote-2)

La tercera y última parte, señala de color rojo, es un mensaje conformado con la respuesta que da la API en formato JSON, en este ejemplo, la misma muestra el mensaje de que las cotizaciones han sido guardadas satisfactoriamente en Cloudant y el estatus de satisfactorio con lo cual se comprueba en correcto funcionamiento de la transferencia de datos.

### Cómputo flexible

Para este módulo fue diseñada la solución principal de todo el sistema la cual se encarga de realizar dos consultas.

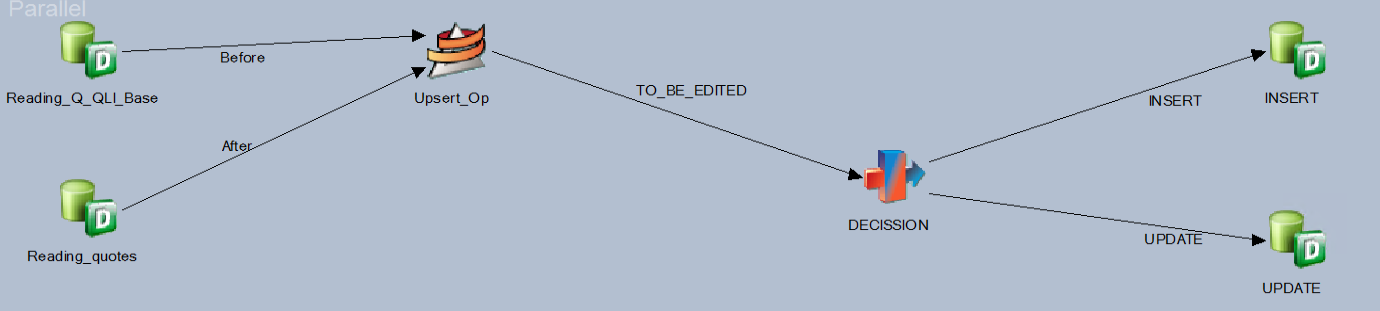


Figura 27: Etapa principal de Checkout Experience

En la figura 27 se muestran (de izquierda a derecha) los primeros dos íconos simbolizan las consultas a la base de datos de Db2, en el primer ícono: *Reading\_Q\_QLI\_Base* es donde se hace una consulta a la tabla base, en el segundo ícono, *Reading\_Quotes*, se realiza la consulta compleja a las distintas tablas de donde se obtendrán todos los datos a comparar contra la tabla base.

Los datos obtenidos en la consulta del objeto *Reading\_Quotes* son los datos actualizados por lo que todos los cambios se realizarán conforme a esta consulta.

El flujo de ambas consultas muestra como el resultado es enviado a través del flujo hacia el objeto *Upsert\_Op*, éste objeto junto al siguiente llamado *Decission* han sido modificados de forma que puedan cumplir con el cometido de esta etapa del proyecto.

En el objeto *Upsert\_Op* inicia una comparación a gran escala donde en base a una llave primaria definida se buscan modificaciones en el resto de los campos de cada registro. Además de todos los campos obtenidos por las consultas complejas del paso anterior se agrega una columna temporal en donde se definirá un número entero que puede ser entre 1 y 3 para identificar los registros que serán insertados y los que serán actualizados, respectivamente.

Por ejemplo, si en el registro con clave primaria 123 se detecta una diferencia en el campo “Nombre de producto” se agrega el valor de “3” en la columna temporal correspondiente a ese registro para que sea actualizado en la tabla base.

Otro ejemplo, es cuando un registro no se encuentra en la tabla base, significa que a este registro le será asignado el valor “1” en la columna temporal correspondiente a ese registro para que sea insertado en la tabla base.

Por último, un caso que no debe suceder, debido a que los registros sobre ventas jamás deben eliminarse de ninguna base de datos, si se da el caso donde un registro de la tabla base no coincida con el de la consulta compleja se completa el proceso de comparación de datos para que una vez finalizada esta etapa sea finalizada el proceso e inmediatamente notificar al grupo de administradores del sistema sobre esta(s) incongruencia(s) en la información obtenida.

Una vez finalizada la etapa de la comparación de los datos, el conjunto de datos generado a partir de esto es enviado hacia el objeto *Decission,* en donde este objeto modificado tomará la determinación de enviar los datos por flujos diferentes dependiendo del valor en la columna temporal, todos aquellos que tengan un valor nulo quedan descartados debido a que no presentan diferencias entre sí.

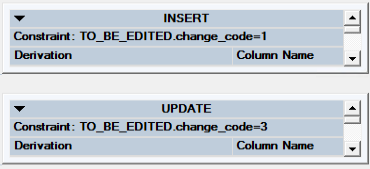


Figura 28: Configuración del objeto Decission

En la figura 28 se muestra como está configurado éste objeto para determinar cuál flujo seguir, el nombre en mayúscula de cada recuadro indica el nombre del flujo a seguir y la restricción (Constraint) indica que valor debe tener el registro en la columna temporal (change\_code) para que sea envíado por el flujo.

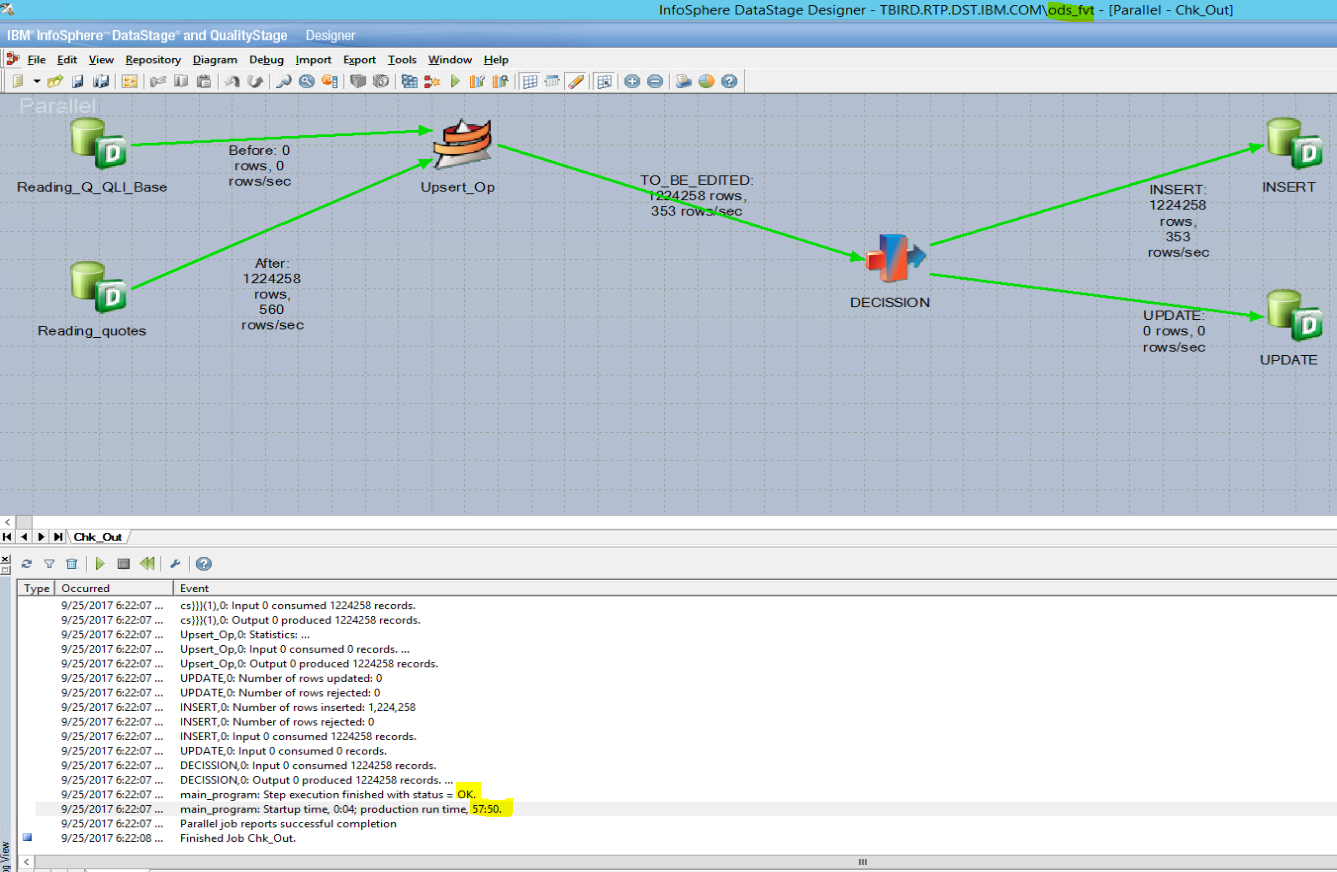


Figura 29: Ejecución de la etapa principal de Checkout Experience

La figura 29 muestra el resultado de una ejecución en el ambiente de pruebas (FVT) en donde la tabla base está vacía, por lo que todos los resultados obtenidos de *Reading\_quotes* (consulta compleja) son insertados en la tabla base.

# Conclusiones

La problemática, cómo ha sido mencionada anteriormente, ha sido la cantidad de datos a ser enviados en conjunto a la necesidad de mantener una constante actualización de los mismos. Por lo que para un problema nuevo fue necesario implementar una solución nueva, capaz de solventar estos puntos de una forma robusta y estable.

Además de esto, uno de los objetivos indirectos y no mencionados previamente, es automatizar un proceso que pudiese implicar al menos la intervención directa de un administrador del sistema. Si el sistema es capaz de encargarse de un proceso de esta manera un recurso para la empresa está disponible para otras actividades que sí necesiten su presencia.

A continuación, son descritas las conclusiones sobre la realización de este proyecto, comenzando por el módulo esencial para el mismo:

## Gestión de la tecnología de la información

Durante la planeación del desarrollo de la arquitectura se consultaron distintas herramientas ETL, en busca en encontrar una que pudiera cumplir con los requisitos solicitados. Esta idea nace a partir de distintos artículos hicieran referencia a las herramientas ETL como una forma de lograr que datos heterogéneos se integren en un entorno homogéneo. [18] [19]

Es por ello que la elección de utilizar una herramienta ETL de este tipo fue minimizar las etapas o interfaces involucradas en un proceso como este, el cual requiere de la mayor disposición de las herramientas disponibles para hacer un uso más efectivo y poder mantener actualizada la información lo más cercano a tiempo real.

La determinación de utilizar DataStage sobre otras herramientas ETL fue debido a las ventajas que ésta ofrece con respecto a otras, ventajas que son:

* Curva de aprendizaje, la cual permite un rápido desarrollo y un reducido mantenimiento haciendo uso de su interfaz gráfica.
* Gran rendimiento en servidores con sistemas operativos con Windows o Unix

De esa forma los representantes de ventas son capaces de ser más eficientes al momento de generar cotizaciones que a su vez pueden ser generadas las órdenes de compra y tener un impacto real sobre los ingresos de la compañía.

El primer módulo del proyecto fue destinado para lograr la gestión de datos entre dos tipos de tecnologías de bases de datos como lo son las bases tradicionales y las jerárquicas.

Y es esa su principal característica, realizar una conexión prácticamente directa entre ambas bases de datos. Fue un objetivo que Checkout Experience es capaz de lograr gracias al sistema de flujos creado y a las etapas implementadas para crear esta conexión.

La API proporcionada por el cliente permite que el sistema únicamente se preocupe por hacer el envío de los datos en formato JSON para que la misma API sea la responsable de procesar, previos a ser insertados, los datos de cualquier forma que lo necesite el cliente.

Para la lograr esta etapa del proyecto fue fundamental la integración del siguiente módulo.

## Sistemas robustos, paralelos y distribuidos

El sistema hace uso de una etapa jerárquica creada para ese tipo de cargas de datos hacía la nube, etapa que fue de las más complejas debido a la falta de documentación existente sobre el tema y sobre todo a la cantidad de campos enviados por cada registro, los cuales son cerca de 160, por lo que realizar el mapeo entre el dato proveniente de la base de datos tradicional hacia el valor del JSON fue un arduo trabajo que implicó conocer el tipo de dato, longitud y formato.

Este trabajo correspondiente al mapeo de los datos fue complicado, pero una de las virtudes de la herramienta de desarrollo es generar archivos en donde esos mapeos se guardan para poder ser reutilizados en otras etapas o proyectos, inclusive.

Una forma ingeniosa para ser capaz de dividir los datos fue incluir una columna con el número de registro en aquellos datos ya cargados al sistema para después, al valor presente en esa columna, aplicar una fórmula que obtuviese el resultado de una operación módulo. De esa forma se mantienen controlados los archivos JSON a generar y la transmisión final de esos archivos hacia la API fluye de una forma más eficaz ya que son manejados (archivos) de menor tamaño. Este no fue un objetivo establecido por el cliente, pero realmente me pareció algo bastante acertado.

Al final de esta etapa se generan archivos de registro, que, si bien usualmente estos archivos son utilizados en casos de fallas o comportamientos inadecuados, se trató de generar un archivo que pudiese ser fácilmente leído y procesado por el administrador de sistemas, es por ello por lo que se incluyen las partes más importantes como lo es el código de respuesta HTTP, los datos enviados en el archivo JSON y la respuesta de la API.

Este archivo log fue el único realizado de esta forma ya que es el principal objetivo del sistema, sin embargo, el resto de las etapas también generar un archivo de registro y a su vez cada etapa es capaz de responder a cualquier clase de finalización, ya sea: exitosa o fallida. Al finalizar el proceso se genera un correo electrónico en donde se incluyen todos los registros generados hasta el momento de finalización que es enviado a una lista de distribución previamente generada, de esta forma el o los administrador(es) de sistemas está(n) informado(s) en todo momento sobre la actividad del sistema.

Reitero lo compleja que fue esta etapa por lo que implicó desarrollar, pero esto fue debido a que la extracción de los datos fuese capaz de que el desarrollo no necesitara mayores esfuerzos en la etapa de cómputo flexible.

## Cómputo flexible

Y es que esta, creo que es una de las mayores virtudes del sistema. La obtención de los datos de un almacén de datos con miles de millones de registros no fue ni es una tarea sencilla, ya que implica el amplio conocimiento de los datos para ser capaz de poder generar la discriminación adecuada de los mismos para extraer solamente los necesarios que al final de todo son los únicos que le importan al cliente.

Y en este caso, primero la consulta compleja es capaz de obtener todos estos datos haciendo uso de múltiples enlaces entre todas las tablas utilizadas y distintas condiciones necesarias para la discriminación mencionada.

Para este proceso se aplicaron diversos algoritmos evolutivos para seguir las técnicas de reducción de datos aplicadas en KDD (Conocimiento en bases de datos, por sus siglas en inglés).

La primera etapa de estos algoritmos aplicados es conocida como preprocesamiento, tarea que implicó mayor dedicación por el tiempo que esta consume.

El algoritmo PAELLA fue utilizado para realizar limpieza de los datos a ser utilizados y de esa forma aumentar la calidad de los datos al nivel requerido mediante las técnicas de análisis selectivo que ofrece este, este proceso consistió en la eliminación de los datos erróneos o inconsistentes. [20]

La integración de los datos mediante la combinación de múltiples tablas para generar nuevos registros fue llevada a cabo siguiendo técnicas similares a las que aplica el recién aceptado algoritmo MMeMeR [21] tales como la combinación de dos o más tablas que presentan diferente información sobre los mismos objetos. La combinación de datos incluyó la agregación, la cual consistió en operaciones donde se obtuvieron nuevos valores mediante la unión de información de varios registros o tablas.

Esta tarea también comprendió así mismo operaciones relativas a construcción de datos tales como la producción de atributos derivados, nuevas muestras completas y transformaciones de los valores de atributos ya existentes.

Como parte de esas transformaciones de datos, la modificación de algunos registros como por ejemplo, los campos referentes a banderas en donde en lugar de tomar el valor de o fueron aplicados casos para cuando esos valores aparecieran, se obtuviera un falso o verdadero.

De esta forma se ejemplifica que las modificaciones sintácticas llevadas a cabo sobre los datos no suponen un cambio en el significado de los mismos y a su vez se permite ser más claro al momento de integrar los datos al archivo JSON ya que este muestra los datos de una forma más comprensible.

Una vez obtenidos los datos, la etapa capaz de discernir entre los datos obsoletos y los actualizados entre el almacén de datos y la tabla base del sistema fue desarrollada de tal forma que únicamente tuviese que tomar como referencia una clave primaria y a partir de esa información, validar cada campo de cada registro para determinar si ese registro sería insertado o actualizado.

En este procedimiento, específicamente, es donde radica la virtud del sistema ya que es capaz de hacer comparaciones de millones de registros en tiempos mínimos y óptimos, por lo que nos permite mantener la uniformidad en los datos procesados que a su vez evita que pudiesen ocurrir duplicados en los datos de Cloudant.

Una de las claves para que el sistema sea capaz de interactuar con cantidades masivas de información es debido al algoritmo para deduplicación o comprensión inteligente de datos [22] implementado.

En el 2013 Nikita A. Pande y Namrata D. Ghuse encontraron una aplicación del algoritmo luciérnaga, un algoritmo meta heurístico, basado en el comportamiento de alumbramiento de las luciérnagas.

“La razón más importante del brillo de una luciérnaga es el acto es como sistema indicador para ser el objetivo de otras luciérnagas” [23] de tal forma que el algoritmo funciona para encontrar réplicas de los registros. Las ventajas de la implementación de este algoritmo, como han sido mencionadas a lo largo de este documento, es que comparado con otros algoritmos el uso del algoritmo luciérnaga arroja un mejor rendimiento.

Esta parte de evitar los duplicados se evita haciendo uso de dos columnas de tipo *timestamp,* columnas donde en la primera de ellas se almacena la fecha y hora exacta en que el registro fue agregado a la tabla base y en la segunda se almacena la fecha y hora exacta en que el registro fue modificado, de esta forma en la etapa subsecuente, mediante una consulta a la tabla base se obtienen solamente los registros que hayan sido insertados o modificados a partir de la fecha especificada o la última ejecución, según corresponda.

En términos generales no fue un proyecto sencillo de desarrollar por todos los temas relacionados en el desarrollo del mismo. Mismo que implicó el conocimiento de los datos radicados en almacén de datos de ODS y los datos que SAP envía a esta interfaz, el conocimiento de un tipo de base de datos relativamente nueva como lo es una de tipo jerárquica que únicamente consume datos de formato JSON.

Creo fervientemente que este sistema es capaz de cumplir con su objetivo principal, que finalmente es mantener actualizada la información de una base de datos que radica en la nube y que será utilizada por representantes de ventas a fin de mejorar el proceso de venta.

Esto permitirá a la IBM recibir un impacto positivo en los ingresos, e inclusive, a los mismos representantes de ventas podrán tener un mejor desempeño y eficiencia, por lo que la automatización de este innovador proceso terminará siendo un total acierto.

# Referencias bibliográficas

|  |  |
| --- | --- |
| [1] | I. Fernández de Marcos, «razonypalabra,» Revista Razon y Palabra, 17 Febrero 2006. [En línea]. Available: http://www.razonypalabra.org.mx/anteriores/n49/bienal/Mesa%205/IsabelDavara.pdf. [Último acceso: 20 Noviembre 2017]. |
| [2] | IFAI, «IFAI,» Instituto Nacional de Transparencia, Acceso a la Información y Protección de datos Personales, 27 Septiembre 2013. [En línea]. Available: http://abcavisosprivacidad.ifai.org.mx/. [Último acceso: 20 Noviembre 2017]. |
| [3] | M. Pena, «Safecreative,» Safecreative Blog, 8 Junio 2010. [En línea]. Available: http://es.safecreative.net/2010/06/08/que-hacer-en-caso-de-plagio-u-otro-uso-ilicito-de-nuestro-contenido/. [Último acceso: 20 Noviembre 2017]. |
| [4] | A. Cornejo, «Normas y estándares en proyectos de T.I.,» 29 Enero 2015. [En línea]. Available: https://normasyestandaresproyectosti.wordpress.com/2015/01/29/iso-12207/. [Último acceso: 20 Noviembre 2017]. |
| [5] | Lloyd's Register LRQA, «Lloyd's Register LRQA,» 12 Diciembre 2016. [En línea]. Available: http://www.lrqa.es/certificaciones/iso-9001-norma-calidad/. [Último acceso: 20 Noviembre 2017]. |
| [6] | ISOTools, «Blog Calidad y Excelencia,» 11 Septiembre 2015. [En línea]. Available: https://www.isotools.org/2015/09/11/en-que-consiste-la-norma-iso-27001/. [Último acceso: 20 Noviembre 2017]. |
| [7] | J. Izquierdo, «IEBS School,» IEBS School, 25 Septiembre 2014. [En línea]. Available: http://www.iebschool.com/blog/por-que-utilizar-metodologia-agiles-agile-scrum/. [Último acceso: 20 Noviembre 2017]. |
| [8] | S. Lamayzi, «Universidad de la Castilla,» 1998. [En línea]. Available: http://alarcos.esi.uclm.es/per/fruiz/cur/mso/comple/iso14764.pdf. [Último acceso: 20 Noviembre 2017]. |
| [9] | I. Castañeda, «Universidad Nacional de Colombia,» [En línea]. Available: http://dis.unal.edu.co/~icasta/GGP/xDBD/2013\_02\_18\_ieee830/GGP\_IEEE\_830\_vL4.pdf. [Último acceso: 20 Noviembre 2017]. |
| [10] | J. Aguilar, «Informmatico-juan,» 08 Diciembre 2012. [En línea]. Available: http://informmatico-juan.blogspot.mx/. [Último acceso: 20 Noviembre 2017]. |
| [11] | Instituto Nacional del Emprendor, «Instituto Nacional del Emprendor,» Instituto Nacional del Emprendor, [En línea]. Available: https://www.inadem.gob.mx/empresas-integradoras/. [Último acceso: 20 Noviembre 2017]. |
| [12] | Revista Cloud Computing, «Revista Cloud Computing,» Revista Cloud Computing, 4 Mayo 2014. [En línea]. Available: https://www.revistacloudcomputing.com/2014/05/cloud-marketplace-el-mercado-de-soluciones-empresariales-en-la-nube-de-ibm/. [Último acceso: 20 Noviembre 2017]. |
| [13] | IBM, «IBM Knowledge Center,» IBM, 31 Mayo 2017. [En línea]. Available: https://www.ibm.com/support/knowledgecenter/en/SSGMCP\_4.1.0/com.ibm.cics.ts.messages.doc/cics\_mc/ASxx\_abend\_codes/ASCA.html. [Último acceso: 20 Noviembre 2017]. |
| [14] | IBM, «IBM Knowledge Center,» IBM, 02 12 2012. [En línea]. Available: https://www.ibm.com/support/knowledgecenter/es/SSYLSL\_9.3.0/com.ibm.help.wcc.concepts.doc/c\_wcc\_overview.html. [Último acceso: 16 02 2018]. |
| [15] | J. Herrera, «juanherrera.files.wordpress.com,» 11 2008. [En línea]. Available: https://juanherrera.files.wordpress.com/2008/11/investigacion-cuantitativa.pdf. [Último acceso: 20 Noviembre 2017]. |
| [16] | B. González, D. Hernández, M. Jiménez, M. I. Marrero y A. Sanabria, «Universidad de La Laguna,» 14 01 2013. [En línea]. Available: https://campusvirtual.ull.es/ocw/pluginfile.php/6115/mod\_resource/content/1/tema9/ME9-muestreo.pdf. [Último acceso: 14 02 2018]. |
| [17] | Unión Europea, «Official Journal of the European Union,» 6 Mayo 2016. [En línea]. Available: http://ec.europa.eu/justice/data-protection/reform/files/regulation\_oj\_en.pdf. [Último acceso: 20 Noviembre 2017]. |
| [18] | R. Espinosa, «Herramientas ETL. ¿Qué son, para que valen? Productos más conocidos,» Dataprix, 25 Febrero 2010. [En línea]. Available: http://www.dataprix.com/blogs/respinosamilla/herramientas-etl-que-son-para-que-valen-productos-mas-conocidos-etl-s-open-sour. [Último acceso: 11 Septiembre 2017]. |
| [19] | P. Rochina, «¿Qué es y cuáles son las aplicaciones del Text Mining?,» INESEM, 25 04 2017. [En línea]. Available: https://revistadigital.inesem.es/informatica-y-tics/text-mining/. [Último acceso: 11 Septiembre 2017]. |
| [20] | Herrera, Francisco; Cano, J.-R., «Soft Computing and Intelligent Information Systems,» 3 Enero 2008. [En línea]. Available: http://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/0678\_Herrera-Cano-ssi06.pdf. [Último acceso: 11 Septiembre 2017]. |
| [21] | A. Goyal, R. Chowdhury y P. Snupam, «Semantic Scholar,» 08 Agosto 2017. [En línea]. Available: https://pdfs.semanticscholar.org/f9ed/f6c8521289bd96d4288324ed88342a11d016.pdf. [Último acceso: 11 Septiembre 2017]. |
| [22] | J. Zavala, «Seguridad informática,» Seguridad informática, 12 Octubre 2008. [En línea]. Available: http..... [Último acceso: 20 Noviembre 2017]. |

# Referencia de figuras

[Figura 1: Logo de la compañía 10](#_Toc507437461)

[Figura 2: Organigrama de la compañía 11](#_Toc507437462)

[Figura 3: Diseño de la solución 27](#_Toc507437463)

[Figura 4: Diagrama entidad relación DB2 29](#_Toc507437464)

[Figura 5: Documento No-SQL Cloudant 30](#_Toc507437465)

[Figura 6: Diagrama de flujo Checkout Experience 32](#_Toc507437466)

[Figura 7: Diagrama de caso de uso 32](#_Toc507437467)

[Figura 8: Diagrama de actividades 33](#_Toc507437468)

[Figura 9: Diagrama de estados 34](#_Toc507437469)

[Figura 10: Diagrama de secuencia 35](#_Toc507437470)

[Figura 11: Diagrama de caso de uso 36](#_Toc507437471)

[Figura 12: Diagrama de actividades 36](#_Toc507437472)

[Figura 13: Diagrama de estados 37](#_Toc507437473)

[Figura 14: Diagrama de secuencia 37](#_Toc507437474)

[Figura 15: Diagrama de caso de uso 38](#_Toc507437475)

[Figura 16: Diagrama de actividades 39](#_Toc507437476)

[Figura 17: Diagrama de estados 39](#_Toc507437477)

[Figura 18: Diagrama de secuencia 40](#_Toc507437478)

[Figura 19: Diagrama de Gantt 41](#_Toc507437479)

[Figura 20: Datos en DB2 46](#_Toc507437480)

[Figura 21: Registros en IBM Cloudant 46](#_Toc507437481)

[Figura 22: Registro en formato JSON en IBM Cloudant 47](#_Toc507437482)

[Figura 23: Correo electrónico enviado por Checkout Experience 48](#_Toc507437483)

[Figura 24: Etapa de conversión y previa separación de datos 49](#_Toc507437484)

[Figura 25: Etapa jerárquica de Checkout Experience 49](#_Toc507437485)

[Figura 26: Archivo de salida 51](#_Toc507437486)

[Figura 27: Etapa principal de Checkout Experience 52](#_Toc507437487)

[Figura 28: Configuración del objeto Decission 53](#_Toc507437488)

[Figura 29: Ejecución de la etapa principal de Checkout Experience 53](#_Toc507437489)

1. La información presentada ha sido reemplazada por datos sin valor debido a que por acuerdo de confidencialidad y seguridad no es posible mostrar un dato real. [↑](#footnote-ref-1)
2. Los valores de los campos mostrados han sido retirados de la figura 26 por acuerdo de conformidad [↑](#footnote-ref-2)