



UNIVERSIDAD
**SAN IGNACIO
DE LOYOLA**

FACULTAD DE INGENIERÍA

Carrera de Ingeniería Empresarial y de Sistemas

IMPLEMENTACIÓN DE LA PLATAFORMA IBM INFOSPHERE INFORMATION SERVER PARA MEJORAR LA INTEGRACIÓN DE DATOS EN SEGUROS GENERALES

**Trabajo de Suficiencia Profesional para optar el Título Profesional
de Ingeniero Empresarial y Sistemas**

JORGE LUIS CHIPANA CHAVEZ

Asesor:

Dr. Marco Antonio Salcedo Huarcaya

**Lima – Perú
2019**

Índice General

Introducción	5
Generalidades de la empresa	6
Datos generales	6
Nombre o razón social de la empresa	6
Ubicación de la empresa	6
Giro de la empresa	7
Tamaño de la empresa	7
Breve reseña histórica de la empresa	7
Organigrama de la empresa	8
Misión, visión y política	8
Productos y clientes	9
Premios y certificaciones	10
Relación de la empresa con la sociedad	10
Planteamiento del problema que fue abordado	11
Caracterización del área en que se participó	11
Antecedentes y definición del problema	12
Problema general.	15
Problemas específicos.	15
Objetivos	16
Objetivo general.	16
Objetivos específicos.	16
Justificación	17
Relación de los objetivos con logros.	18
Alances y limitaciones	20
Alcances.	20
Limitaciones.	20
MARCO TEÓRICO	21
IBM InfoSphere Information Server	21
Fases de integración de datos	22
Descubrimiento y análisis.	22
Diseño.	23
Desarrollo.	23
Despliegue.	23
Componentes de la plataforma	24

InfoSphere DataStage.	24
InfoSphere Data Click.	25
InfoSphere FastTrack.	25
InfoSphere Information Analyzer.	25
InfoSphere Information Governance Catalog.	25
InfoSphere Information Services Director.	26
InfoSphere QualityStage.	26
DESARROLLO DEL PROYECTO	27
Objetivo 1: Analizar los componentes internos y externos con relación a la arquitectura de integración de datos en la empresa Seguros Generales, Lima, 2017	31
Planificación del proyecto.	31
Definición de requerimientos del negocio.	31
Diseño de Arquitectura de Hardware.	32
Diseño de Arquitectura de Componentes IBM InfoSphere Information Server.	35
Requerimientos de sistema IBM InfoSphere Information Server.	37
Preparación de los servidores.	38
Revisión de los servidores y requerimientos de hardware y software.	38
Instalación de plataforma IBM InfoSphere Information Server.	40
Configuración de plataforma IBM InfoSphere Information Server.	46
Objetivo 2: Diseñar modelos de integración de datos basados en casos de uso de la empresa Seguros Generales, Lima, 2017.	53
Modelos y lineamientos de integración de datos.	53
Objetivo 3: Implementar una arquitectura orientada a servicios, eliminar duplicidad de funciones, uso eficiente de recursos de hardware, reducir el esfuerzo en tareas administrativas y despliegue procesos de integración de datos de la empresa Seguros Generales, Lima, 2017.	60
Taller de administración de plataforma IBM InfoSphere Information Server.	60
Taller de desarrollo de integración de datos	61
ANÁLISIS Y RESULTADOS	62
Análisis	62
Resultados	64
Conclusiones	68
Recomendaciones	69
LISTA DE FIGURAS	70

LISTA DE TABLAS	72
REFERENCIAS	73
ANEXOS	74
Anexo 1: Project Chapter	74

Introducción

Hoy en día se genera una inmensa cantidad de datos segundo a segundo, no solo tenemos datos provenientes de sistemas tradicionales sino también de redes sociales, video, voz, lo que hoy se conoce cómo big data. Las empresas se enfrentan al desafío de cómo capitalizar y lograr valor con los datos, para encontrar soluciones y ventajas competitivas a partir de la recopilación, almacenamiento, análisis y transformación de los datos.

IBM InfoSphere Information Server ofrece a las empresas una plataforma única para la integración, calidad y gobierno de datos, lo que permite crear una base unificada capaz de soportar las necesidades de una arquitectura de información unificada. La plataforma proporciona una amplia gama de funciones de transformación, calidad y gobierno datos a partir del perfilamiento, estandarización, coincidencia y enriquecimiento de los datos.

El proyecto de implementación de la plataforma de integración de datos de IBM en la empresa Seguros Generales consistió en el dimensionamiento de recursos de hardware y software, diseño de arquitectura de integración de datos, definición de lineamientos y plantillas de componentes de integración de datos y transferencia de conocimientos en administración de la plataforma para el soporte y operatividad. Con ello, se logró entregar al cliente una plataforma robusta y escalable, que permite perfilar, transformar y gobernar datos estructurados y no estructurados. Además, aprovechar la funcionalidad de procesamiento paralelo capaz de procesar grandes volúmenes de datos.

Generalidades de la empresa

Datos generales

IBM del Perú S.A.C. es una empresa transnacional con más 85 años de actividad en el país. Desde entonces, IBM se ha transformado con el fin de atender las necesidades de sus clientes y trabajar junto con ellos para reinventar sus negocios. En los últimos años nuevas tecnologías como: Inteligencia Artificial, Analítica Avanzada, Blockchain, Computación Cognitiva, Ciencia de Datos, Internet de las Cosas y Cloud están marcando la diferencia en la manera de hacer negocios, innovación, e IBM presenta un amplio portafolio de productos y servicios afines a los objetivos empresariales.

Nombre o razón social de la empresa

IBM del Perú S.A.C.

Ubicación de la empresa

Dirección: Av. Javier Prado Este 6230, La Molina, Lima, Perú.

Teléfono: +511 625 6000

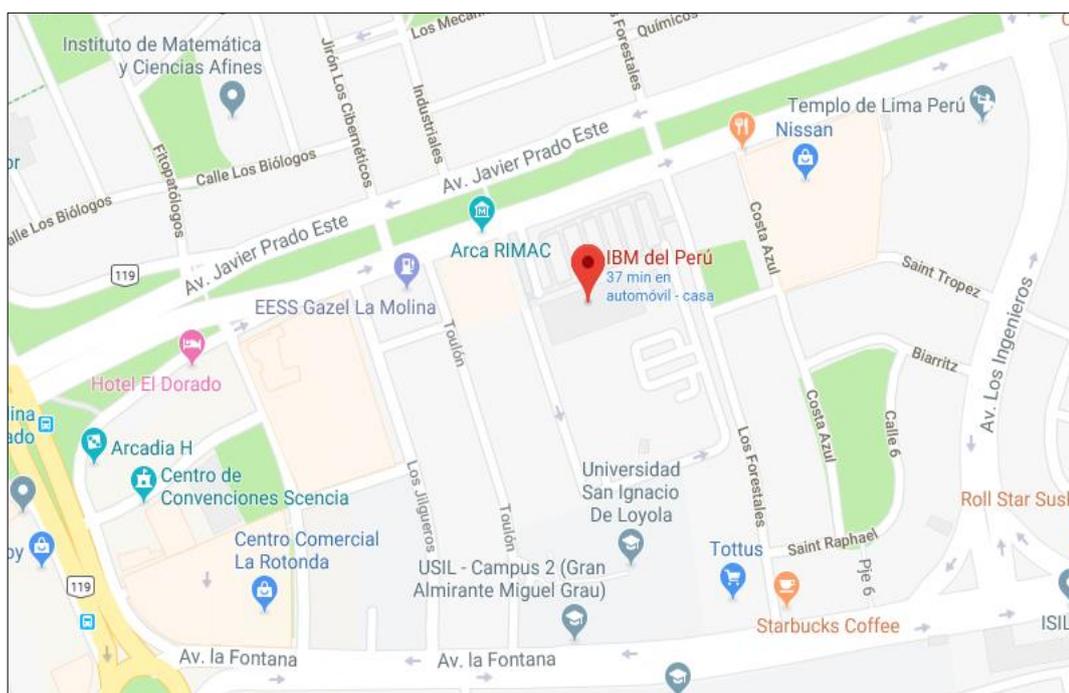


Figura 1. Ubicación IBM del Perú en el distrito de la Molina, Lima, Perú. Tomado de Google Maps.

Giro de la empresa

Dedicada a investigación, desarrollo, fabricación y comercialización de tecnologías de información, hardware y software. Además, servicios de TI, outsourcing, integración de sistemas, financiamiento y servicios de consultoría de negocio.

Tamaño de la empresa

IBM es una empresa transnacional grande, siendo IBM del Perú S.A.C. la filial en el país, está presente en más de 170 países y cuenta con más de 386 mil colaboradores en el mundo. En el 2018 ha registrado ingresos 93, 800 millones de dólares, distribuidas en prestaciones servicios 52%, comercialización de hardware 22%, 20% en venta de software y 6% en financiamiento y otros conceptos.

Breve reseña histórica de la empresa

Computing Tabulating Recording Corporation fue fundada el 15 de junio de 1911 en Binghamton, Estados Unidos, pero el 14 de febrero de 1924 cambió su nombre a International Business Machines (IBM) y llega al Perú en 1932. Desde entonces se ha mantenido a la vanguardia tecnológica, investigando, desarrollando no solo productos como hardware y software sino también servicios de TI, outsourcing, financiamiento y servicios de consultoría de negocio, lo que permite ofrecer soluciones tecnológicas a cualquier tipo de cliente desde usuarios particulares y grandes empresas de cualquier sector de actividad.

Organigrama de la empresa

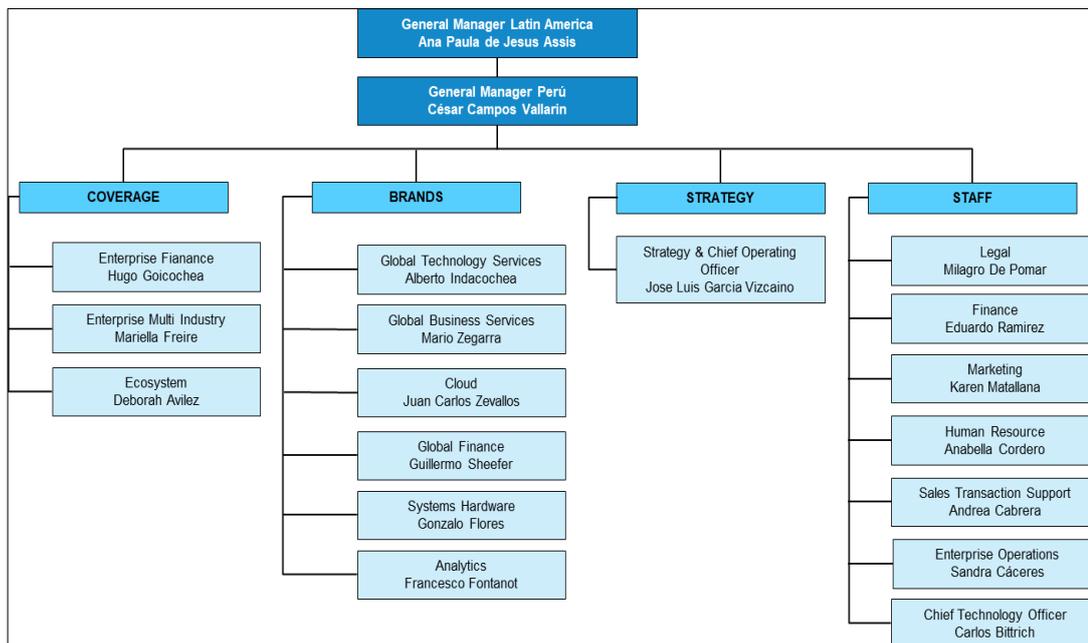


Figura 2. Organigrama IBM del Perú. Elaboración propia.

Misión, visión y política

IBM tiene por misión desarrollar ventajas competitivas en la industria de tecnologías de información. La red de profesionales de IBM traduce las tecnologías de información en valor comercial para sus clientes.

IBM tiene por visión ser líder en la industria de tecnologías de información.

IBM tiene por política manejar de manera responsable la recopilación, almacenamiento y administración de los datos.

Productos y clientes

IBM tiene un amplio portafolio de productos y servicios adaptables a las necesidades de las empresas, cada unidad de negocio de IBM ofrece a los clientes soluciones para desarrollar su negocio de la manera más eficiente y sumar valor. A continuación, una descripción de las unidades de negocio de IBM:

IBM Global Financing, financiamiento de tecnologías de información TI. Con soluciones de financiamiento y arrendamiento que ayudan a las empresas a conseguir un importante ahorro en adquisiciones de TI.

IBM Software, provee la gama más amplia de middleware (software de infraestructura) con software IBM y las soluciones de industria ayudan a innovar para crear soluciones específicas para sectores industriales que transforman su empresa y proporcionan un valor empresarial real.

IBM Global Business Services, cuenta con profesionales dedicados a transformar las organizaciones con nuevos modelos de negocio, desarrollando estrategias e implementando soluciones que aportan al crecimiento de las empresas.

IBM Global Technology Services, servicios de outsourcing reduciendo costos de TI y flexibilidad en la transformación de infraestructura y aplicaciones.

IBM System & Technology Group, los principios centrales de la colaboración, apertura, virtualización y optimización como elementos esenciales para una infraestructura integrada.

Premios y certificaciones

IBM ha logrado certificaciones corporativas para las normas ISO 9001, ISO 14001 & ISO 50001, ISO 20000, OHSAS 18001, etc. (según el sitio <https://www.ibm.com/support/pages/ibm-iso-management-system-certifications>)

Relación de la empresa con la sociedad

En octubre del 2018 el Ministerio de Desarrollo e Inclusión Social (MIDIS) e IBM del Perú organizaron "IBM Coding Day: HACKATHON por un Perú #SinAnemia". La Hackathon consistió en una "maratón" donde diversos equipos desarrollaron soluciones innovadoras y tecnológicas a un problema específico. El objetivo fue generar un ambiente para diseñar soluciones innovadoras que contribuyan a disminuir la anemia en niños y niñas entre 6 y 36 meses.

Planteamiento del problema que fue abordado

Caracterización del área en que se participó

IBM opera en más de 170 países, en el Perú desde 1932, cuenta con más de 24 laboratorios de desarrollo y 8 centros de investigación repartidos por todo el mundo.

IBM ha diversificado su oferta de producto con la producción de hardware, software, financiamiento de tecnologías de información, servicios de seguridad, servicios de consultoría, servicios de infraestructura de tecnologías de información y servicios de continuidad del negocio.

El área que participó en el presente trabajo es el área de IBM Software, que está conformada por los equipos de venta de licencias de software, preventa técnica, servicios de consultoría de software y servicio de soporte.

El proceso inicia con el levantamiento de información y el análisis de la oportunidad a razón de un requerimiento de cliente. A continuación, se plantean y elaboran varias propuestas de licencias de software y servicios de consultoría que reúnan las condiciones las de licenciamiento, soporte y alcance de actividades del proyecto de implementación del software.

Antecedentes y definición del problema

El área de arquitectura de TI de Seguros Generales es el responsable de la planificación e implementación de la integración de datos. El principal objetivo es la facilitar información confiable, centralizada y oportuna a las áreas de negocio como marketing, ventas y cobranzas en la organización.

Las áreas de negocio de Seguros Generales han estado implementando integración de datos para poblar sus modelos de datos de manera aislada, sin un marco de referencia o lineamientos de diseño y gobierno de datos, lo que no solo generó duplicidad en infraestructura de TI (servidores, arreglos de disco, cpu, memoria y licenciamiento de software), sino también componentes manuales y redundantes. En consecuencia, los datos son poco confiables, aislados y poco oportunos afectándose la toma de decisiones y estrategias de negocio como campañas de ventas, desarrollo de productos, seguimiento de siniestros, entre otros.

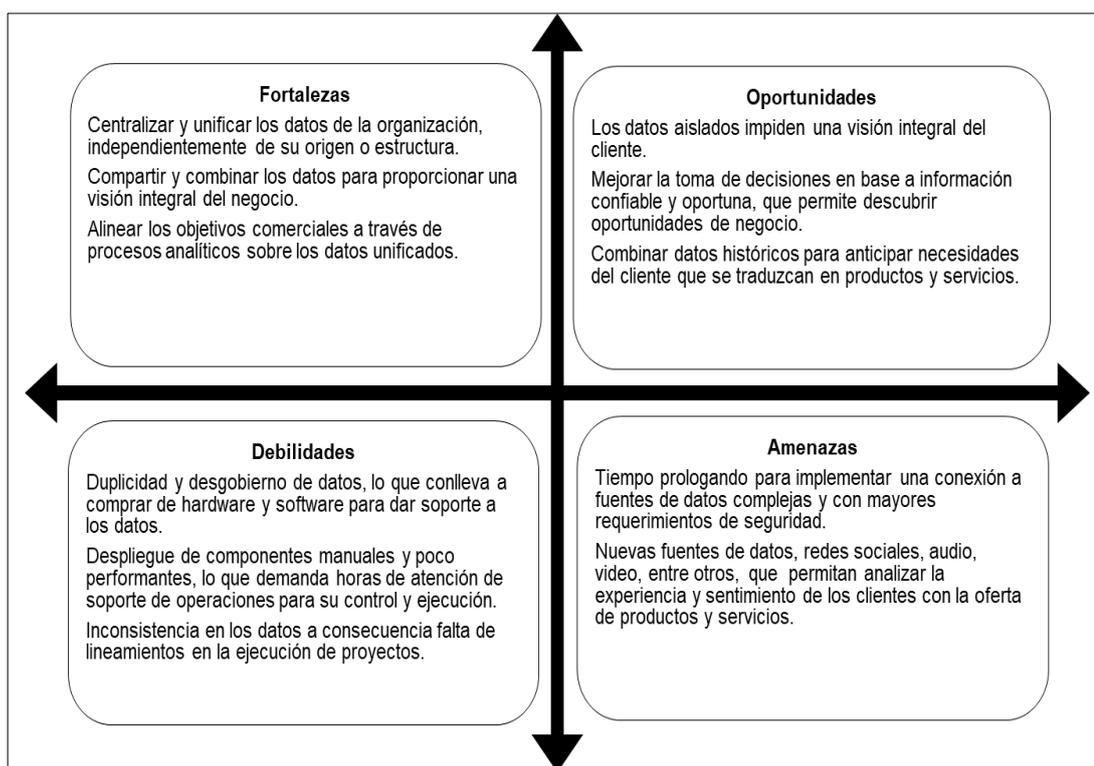


Figura 3. Matriz FODA de integración de datos en Seguros Generales.

En base al análisis se evidencia el impacto de la integración datos en la organización y como esta afecta a logro de los objetivos del negocio. Elaboración propia.

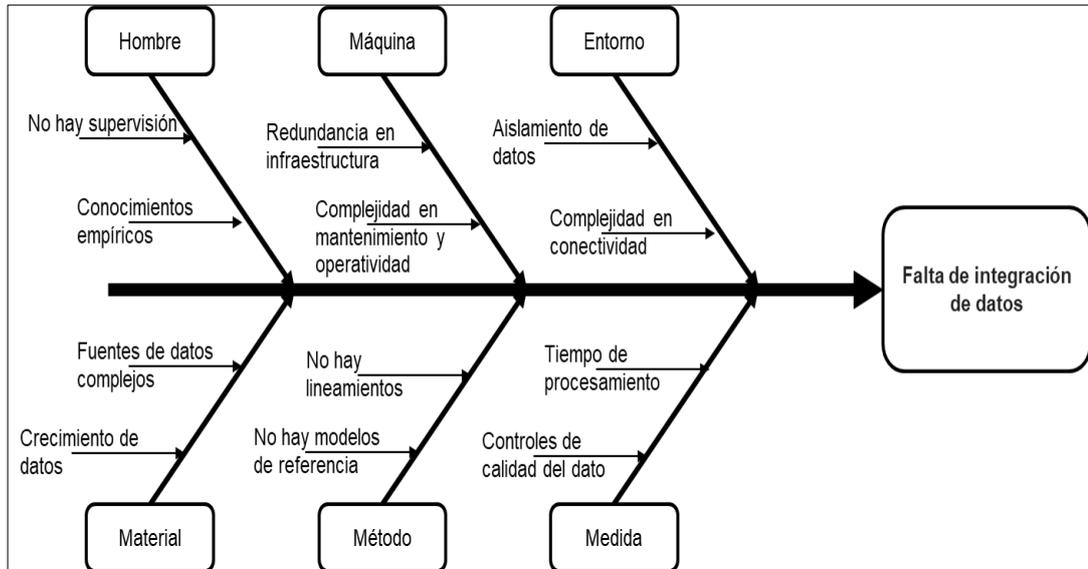


Figura 4. Análisis Ishikawa del problema en Seguros Generales. Se exponen las causas principales de los problemas de la integración de datos. Elaboración propia.

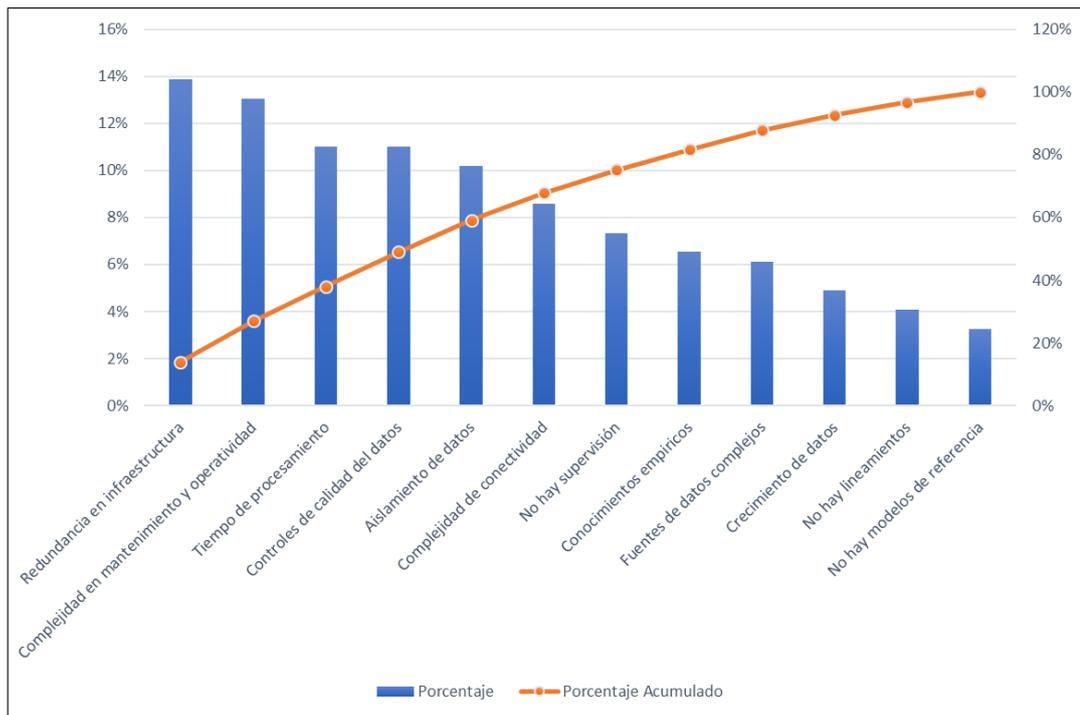


Figura 5. Diagrama de Pareto determina las principales causas del problema en Seguros Generales. El proyecto se enfocó en resolver las principales causas del problema falta de integración de datos. Redundancia en infraestructura, complejidad en mantenimiento y operatividad, tiempo de procesamiento, controles de calidad del dato, aislamiento de datos, complejidad de conectividad y no hay supervisión son las que concentran el 80% de causas. Elaboración propia.

A inicios de 2017 el área de Arquitectura de TI de Seguros Generales, luego de una reestructuración del área, realizar un levantamiento de información para identificar la infraestructura hardware y licenciamiento de software que ha sido adquirido e implementado desde 2010 para soportar integración de datos en la empresa. En junta de comité donde se discutió el plan estratégico anual las áreas de negocio o usuarias evidenciaron que sus metas y objetivos se ven afectados en gran parte debido a decisiones tomadas a partir de los datos que reciben. Lo que ha ocasionado que los costos de mantenimiento de hardware y licenciamiento de software que adquirieron se hayan incrementado cada vez más y no se ha obtenido el beneficio esperado.

Tabla 1

Resumen de inventario de adquisiciones de servidores del 2010 al 2016.

Servidores Base Datos	Servidores Herramientas ETL	Servidores Herramientas Visualización
85	45	15

Nota ETL= Extracción, Transformación y Carga.

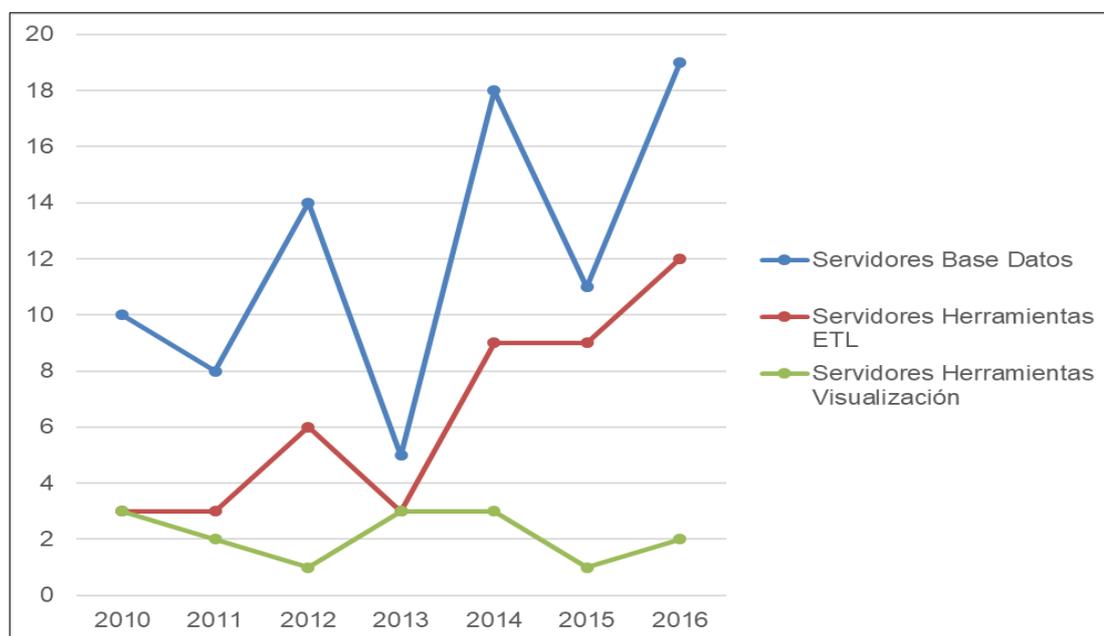


Figura 6. Inventario anual de adquisiciones de infraestructura de hardware y software. Las cifras representan la adquisición progresiva de servidores de bases de datos, herramientas ETL y de visualización requeridas para la integración de datos por las áreas usuarias. Elaboración propia.

Tabla 2

Resumen de costos por adquisiciones de servidores desde 2010 a 2016

Año	Servidores Base Datos		Servidores Herramientas ETL		Servidores Herramientas Visualización	
	Anual \$	Acumulado \$	Anual \$	Acumulado \$	Anual \$	Acumulado \$
2010	24,000.00	24,000.00	36,000.00	36,000.00	66,000.00	66,000.00
2011	19,200.00	43,200.00	36,000.00	72,000.00	44,000.00	132,000.00
2012	33,600.00	76,800.00	72,000.00	144,000.00	22,000.00	176,000.00
2013	12,000.00	88,800.00	36,000.00	180,000.00	66,000.00	198,000.00
2014	43,200.00	132,000.00	108,000.00	288,000.00	66,000.00	264,000.00
2015	26,400.00	158,400.00	108,000.00	396,000.00	22,000.00	330,000.00
2016	45,600.00	204,000.00	144,000.00	540,000.00	44,000.00	352,000.00

Nota ETL= Extracción, Transformación y Carga.

Problema general.

¿Cuánto mejora la implementación de la plataforma IBM InfoSphere Information Server la integración de datos en la empresa Seguros Generales?

Problemas específicos.

Redundancia en infraestructura de hardware y licenciamiento de software para el soporte de la operatividad de integración de datos.

Desconocimiento en modelos de referencia para la implementación de integración de datos.

Incumplimiento en la entrega oportuna de información a los usuarios del negocio debido a lentitud en el procesamiento de datos. La información está disponible entre 24 a 40 horas desde su ingesta al repositorio data warehouse.

Objetivos

Objetivo general.

Implementación de la plataforma IBM Infosphere Information Server para unificar la arquitectura de integración de datos de la empresa Seguros Generales, Lima, 2017

Objetivos específicos.

Analizar los componentes internos y externos con relación a la arquitectura de integración de datos en la empresa Seguros Generales, Lima, 2017.

Diseñar modelos de integración de datos basados en casos de uso de la empresa Seguros Generales, Lima, 2017.

Implementar una arquitectura orientada a servicios, eliminar duplicidad de funciones, uso eficiente de recursos de hardware, reducir el esfuerzo en tareas administrativas y despliegue procesos de integración de datos de la empresa Seguros Generales, Lima, 2017.

Justificación

Hoy en día muchas empresas buscan lograr ventajas competitivas a través del correcto tratamiento y análisis de datos, estructurados o no estructurados, tomando conciencia de que los datos es la principal materia prima en la toma de decisiones.

Integración de datos se implementa usualmente en data warehouse mediante el uso de hardware y software especializado que permita alojar grandes repositorios de datos internos o externos. Ello implica, gran inversión económica, especialización técnica y funcional. Sin embargo, no contar con una plataforma de integración de datos capaz de conectar, procesar y entregar en tiempo adecuado los datos, afecta severamente los objetivos de integración de datos.

IBM InfoSphere Information Server es una plataforma de integración de datos que permite a las empresas obtener más valor de los datos complejos y heterogéneos inmersos en sus sistemas de información. Proporciona una única plataforma integrada, con una arquitectura escalable vertical y horizontal, que permite a las empresas entender, limpiar, transformar y entregar información confiable.

Relación de los objetivos con logros.

Analizar los componentes internos y externos con relación a la arquitectura de integración de datos en la empresa Seguros Generales, Lima, 2017.

Implementación de integración de datos sobre una plataforma que permite unificar los componentes de la arquitectura de información de la organización.

Escalabilidad para cumplir los requerimientos de procesamiento de volúmenes de datos, independientemente de su origen, manteniendo la integridad de estos.

Integración al ecosistema de TI de Seguros Generales, permitiendo reducir y optimizar la infraestructura de TI, hardware y software, en términos de costos de licenciamiento y mantenimiento.

Diseñar modelos de integración de datos basados en casos de uso de la empresa Seguros Generales, Lima, 2017.

Diseño de modelos de implementación de integración de datos que permiten establecer un marco de trabajo, impulsando la colaboración entre los usuarios del negocio y de TI para entender el significado, estructura y contenido de la información entre sus fuentes de datos.

Implementar una estrategia completa de integración de datos para incrementar la eficiencia operativa, compartiendo activos de información para que los usuarios interactúen según sus roles y necesidades de información.

Implementar una arquitectura orientada a servicios, eliminar duplicidad de funciones, uso eficiente de recursos de hardware, reducir el esfuerzo en tareas administrativas y despliegue procesos de integración de datos de la empresa Seguros Generales, Lima, 2017.

Estandarizar y consolidar la arquitectura de integración de datos en una única plataforma. Las iniciativas de inteligencia de negocios desarrollaron una vista más cohesiva de los datos al obtener una fuente común de información.

Realizar talleres para usuarios administradores y desarrolladores de la plataforma IBM InfoSphere Information Server permitió afianzar conocimientos necesarios para una óptima operatividad de la integración de datos.

Alances y limitaciones

Alcances.

El presente trabajo buscó mejorar la integración de datos a través de la implementación de la plataforma de IBM InfoSphere Information Server. Lo cual, se tradujo en consolidación de la arquitectura de integración datos, en términos de infraestructura, lograr reducción de costos de mantenimiento y licenciamiento adquiridos. Establecer un marco de trabajo colaborativo que permita el desarrollo de iniciativas de inteligencia de negocios.

Las tareas de migración de datos no formaron parte del alcance del proyecto. Sin embargo, el objetivo de reducción de tiempos de procesamiento de datos fue alcanzado mediante aplicación de los modelos de integración de datos. El cliente ejecutó un proyecto de migración de datos, con otro proveedor de servicios, inmediatamente después de finalizado este proyecto.

El proyecto tuvo una estimación de 6 semanas entre marzo y abril de 2017, en las oficinas de Seguros Generales del Perú. El cliente fue responsable de otorgar y facilitar los accesos a las instalaciones y a los ambientes tecnológicos.

Limitaciones.

No es posible especificar el nombre de la empresa cliente donde se desarrolló el proyecto, nombres de integrante del equipo de IBM Software Services y de cliente, debido a políticas de confidencialidad IBM.

MARCO TEÓRICO

IBM InfoSphere Information Server

En 2005 IBM adquiere el software DataStage, una herramienta ETL, que permite la integración de datos de múltiples sistemas y el procesamiento de grandes volúmenes de datos. DataStage permite a través de una interfaz gráfica el diseño de trabajos para la extracción, transformación, validación y carga de datos procedentes de aplicaciones empresariales como Sibel, Oracle y Microsoft SQL Server.

IBM realizó adaptaciones al software DataStage para integrarla a la plataforma IBM InfoSphere Information Server. Con ello, se posicionó como una herramienta de integración de datos que permite a los usuarios mover, transformar, validar y cargar datos operacionales, transaccionales y analíticos.

IBM InfoSphere Information Server ofrece una plataforma única para la integración, calidad y gobierno de datos. De este modo, facilitar al negocio y al equipo de TI a colaborar para comprender el significado, la estructura y el contenido de diversas fuentes de información.

Entre las funcionalidades más importante de la plataforma está el repositorio común de metadatos que almacena metadatos importados, configuraciones de proyectos, informes y resultados para todos los componentes de InfoSphere Information Server. Al compartirse los datos importados en el repositorio de metadatos, otros usuarios pueden hacer uso de los activos importados en otros componentes de InfoSphere Information Server.

IBM expone que “InfoSphere Information Server supports all phases of an effective data integration project. These phases constantly evolve as the lifecycle of the project grows and changes. By providing key data integration capabilities, InfoSphere Information Server addresses each phase to ensure that your project is successful” (2017). Recuperado el 6 de diciembre de 2019, de https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_11.7.0/com.ibm.swg.im.iis.productization.iisinfsv.overview.doc/topics/cisoarchoverview.html

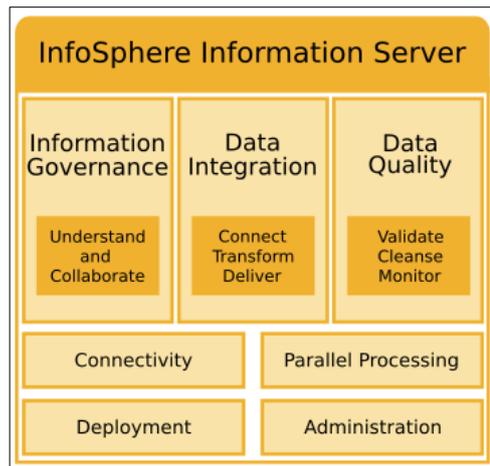


Figura 7. Funciones de integración de la plataforma IBM InfoSphere Information Server. Capacidades clave para implementar una estrategia de integración de datos. Tomado de IBM Knowledge Center. *Introduction to InfoSphere Information Server*. Recuperado de https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_11.7.0/com.ibm.swg.im.iis.productization.ii.sinfosv.overview.doc/topics/cisoinroduction.html

Fases de integración de datos

InfoSphere Information Server soporta las fases de un proyecto de integración de datos bajo un enfoque metadatos común y componentes que optimizan la colaboración entre los roles de usuario.

Descubrimiento y análisis.

InfoSphere Information Server permite descubrir automáticamente la estructura de los datos, analizar el significado, las relaciones de integridad referencial y el linaje de información. Utilizar un repositorio de metadatos común facilita información sobre el origen, el uso y los cambios de cada dato.

Los analistas de datos pueden generar especificaciones de integración y reglas de datos que permitan monitorear el comportamiento de la información, logrando los siguientes objetivos:

Comprender las fuentes de datos y las relaciones de integridad referencial.

Reducir el riesgo de usar datos incorrectos.

Usar y compartir los activos de información.

Diseño.

InfoSphere Information Server permite diseñar modelos de datos físicos y modelos de datos lógicos, asegurando que la arquitectura de información pueda soportar cambios a medida que ocurran.

Los analistas pueden diseñar reglas de calidad que analicen muestras de datos para determinar su calidad, de modo que pueda tomar acción sobre algún evento que afecten al proyecto. Este enfoque garantiza la confiabilidad e integridad de sus datos al monitorear constantemente los cambios.

Desarrollo.

InfoSphere Information Server permite la estandarización, validación, coincidencia y fusión de datos. Además, transforma y enriquece los datos para garantizar el contexto solicitado. Para ello, se encuentran disponibles cientos de funciones de transformación preconstruidas que combinan y reestructuran.

Despliegue.

InfoSphere Information Server permite mover los datos a través de etapas en el proyecto. Al término de diseñar y desarrollar los procesos de extracción, transformación y carga, puede monitorear las actividades de ejecución, el consumo de los recursos del sistema y gestionar la carga de trabajo.

Mediante la importación, exportación y administración activos de metadatos comunes los activos están disponibles para los usuarios de la plataforma.

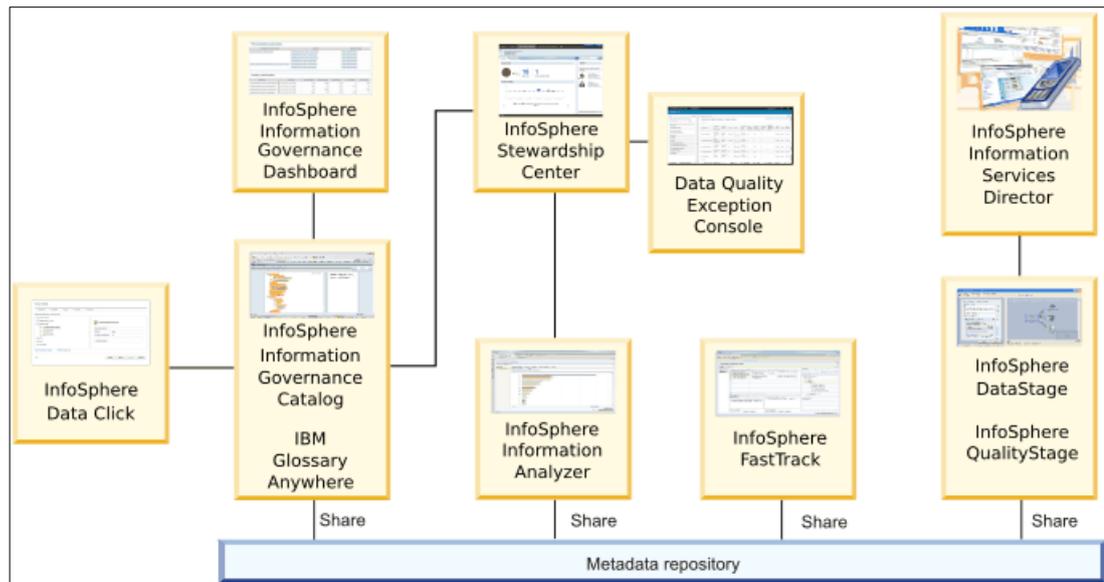


Figura 8. Componentes de la plataforma IBM InfoSphere Information Server.

Los componentes de la plataforma, a través de metadatos comunes, trabajan juntos para crear una solución de integración de datos unificada. Tomado de IBM Knowledge Center. *Data integration phases*. Recuperado de https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_11.7.0/com.ibm.swg.im.iis.productization.ii.sinfsv.overview.doc/topics/cisocapabilities.html

Componentes de la plataforma

InfoSphere Information Server está compuesto por componentes, y cada uno provee funciones distintas para la integración de datos. En conjunto, apalancan la entrega de información confiable a la organización.

InfoSphere DataStage.

Este componente permite la integración de datos, provee de conectividad directa a los orígenes de datos, dispone de múltiples funciones para la transformación y enriquecimiento de los datos. Se integra y adopta las capacidades de procesamiento en paralelo del hardware multiprocesador permitiendo a las organizaciones trabajar con grandes volúmenes de datos.

InfoSphere Data Click.

Este componente permite mover datos entre bases de datos, no aplica ningún cambio sobre los datos.

InfoSphere FastTrack.

Este componente permite automatizar el flujo de trabajo del proyecto de integración de datos. Los usuarios pueden diseñar plantillas de procesos de integración de datos consumiendo metadatos comunes para desplegar Jobs DataStage, lo que permite reducir el tiempo de desarrollo.

InfoSphere Information Analyzer.

Este componente permite perfilar y analizar datos, con ello tener un diagnóstico de estado de los datos en términos de dominio de valores, formatos y clasificación de datos.

InfoSphere Information Governance Catalog.

Este componente permite comprender y gobernar su información. Establece un lenguaje común a través de diccionario de negocio y vincularlo a los metadatos comunes.

InfoSphere Information Services Director.

Este componente facilita la implementación procesos de integración de datos, Jobs DataStage, como servicios.

InfoSphere QualityStage.

Este componente provee de funciones para la limpieza, estandarización, coincidencia y fusión de datos. Lo componen etapas que están inmersos en IBM InfoSphere DataStage.

DESARROLLO DEL PROYECTO

Para el desarrollo del proyecto se tuvo en cuenta la metodología de ciclo de vida de Kimball. El cual es tomada como referencia por las organizaciones para el diseño y desarrollo de sistemas de inteligencia de negocio y data warehouse.

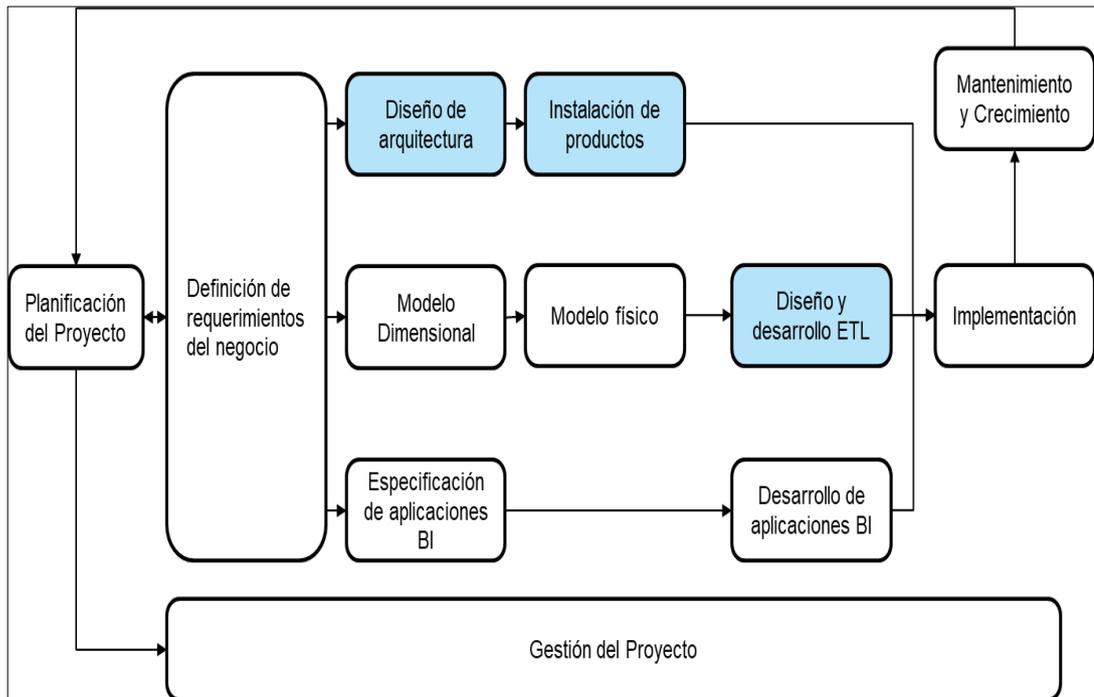


Figura 9. Metodología del ciclo de vida de Kimball.

El proyecto abordó las actividades de diseño de arquitectura, instalación de productos y diseño y desarrollo de ETL. Elaboración propia.

Seguros Generales tenía en marcha el proyecto de desarrollo de data warehouse bajo el enfoque de la metodología de ciclo de vida de Kimball. Sin embargo, en relación las actividades concernientes a integración de datos no tenían definida la hoja de ruta a seguir. Lo que incrementaba la incertidumbre de que el proyecto de data warehouse no cumpliera con los objetivos del negocio por la incapacidad de integrar los datos.

El proyecto de implementación de la plataforma IBM InfoSphere Information Server se enfocó en implementar las actividades de diseño de arquitectura, instalación de productos y diseño y desarrollo de ETL, que son clave en sistemas de inteligencia de negocios. Fue necesario considerar lo establecido por Seguros Generales en las actividades de planificación de proyecto y definición de requerimientos de negocio, para alinear las necesidades de la organización e implementar una arquitectura de integración de datos que cumpla las expectativas.

IBM Software para la ejecución del proyecto designó un gerente de proyecto y un consultor, especialista en la plataforma IBM InfoSphere Information Server. El gerente de proyecto se encargó de todo lo relacionado a la gestión del proyecto, desde coordinaciones con el cliente, agenda de reuniones, elaboración de actas, elaboración y seguimiento del cronograma.

Se participó como consultor, por lo que la responsabilidad de los aspectos técnicos, desde el diseño de la arquitectura de integración de datos, instalación y configuración del software, transferencia de conocimientos, diseño de plantillas ETL, lineamientos y buenas prácticas de desarrollo de ETL para la integración de datos en la plataforma IBM InfoSphere Information Server fueron las actividades principales.

Seguros Generales asignó un gerente de proyecto y un arquitecto de TI, quienes participaban también del proyecto de data warehouse, con quienes nos reunimos para revisar el plan de actividades del proyecto, prerequisites, premisas y entregables. Con la aprobación del plan de actividades el proyecto fue ejecutado en 6 semanas con lo que se logró cumplir los objetivos y expectativas del cliente.

Tabla 3

Cronograma de actividades del proyecto de implementación de la plataforma IBM InfoSphere Information Server

Tarea	Responsable	Semana 1		Semana 2		Semana 3		Semana 4		Semana 5		Semana 6	
		6-Mar	10-Mar	13-Mar	17-Mar	20-Mar	24-Mar	27-Mar	31-Mar	3-Apr	7-Apr	10-Apr	14-Apr
Diseño de Arquitectura													
Arquitectura y pre requisitos de hardware y software	IBM	■											
Preparación de servidores y software base	SG		■										
Revisión de servidores y pre requisitos de HW y SW	IBM		■										
Instalación													
Instalación IBM InfoSphere Information Server	IBM			■									
Configuración de seguridad	IBM				■								
Configuración de conexión a bases de datos	IBM					■							
Configuración de proyectos de integración de datos	IBM						■						
Pruebas de configuración de plataforma	IBM							■					
Documentación de instalación y configuración	IBM									■			

Nota HW y SW= Hardware y Software.

Tabla 4

Cronograma de actividades del proyecto de implementación de la plataforma IBM InfoSphere Information Server

Tarea	Responsable	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6				
		6- Mar	10- Mar	13- Mar	17- Mar	20- Mar	24- Mar	27- Mar	31- Mar	3- Abr	7- Abr
Plantillas y lineamiento de procesos de integración de datos											
Documento de plantilla de procesos de integración de datos	IBM										
Documento de lineamientos de procesos de integración de datos	IBM										
Transferencia de conocimiento											
Workshop Administración de plataforma IBM InfoSphere Information Server	IBM										
Workshop Desarrollo de procesos de integración de datos	IBM										

Nota. Workshop=Taller.

Objetivo 1: Analizar los componentes internos y externos con relación a la arquitectura de integración de datos en la empresa Seguros Generales, Lima, 2017

Planificación del proyecto.

Seguros generales planteó que la capacidad de integrar datos para analizarlos responde a un sentido de urgencia motivados por factores externos, la competencia en el mercado de seguros. Las gerencias tenían participación en las reuniones de definición de los requerimientos de negocio.

Arquitectura de TI de Seguros Generales orquestó la implementación de la plataforma IBM InfoSphere Information Server para soportar la arquitectura de integración de datos. Las capacidades de la plataforma como procesamiento de grandes volúmenes de datos, conectividad a fuentes de datos, esquema de seguridad, gestión de metadatos, extracción, transformación y carga de datos cumplen con las necesidades técnicas y de negocio.

El equipo multi disciplinario de arquitectos técnicos, administradores de base de datos, arquitectos de datos, desarrolladores fueron involucrados como usuarios directos de la plataforma IBM InfoSphere Information Server. Con ello, reducir la curva de aprendizaje y que la operatividad sea eficiente.

Definición de requerimientos del negocio.

Seguros generales mapeó los requerimientos del negocio y las correlacionó a la realidad de los datos. Las fuentes de datos estaban en modelos relacionales, un 70% correspondía a clientes, vehículos, pólizas y siniestros, en base datos de datos Oracle. Un 20% de las fuentes de información de colaboradores en Microsoft S SQL Server y 10 % restante de contabilidad en Teradata.

Entre los principales requerimientos del negocio estaba la capacidad de combinar los datos de diferentes orígenes de bases de datos para explorar y analizar productos y campañas de ventas, procesar grandes volúmenes de datos en caso de siniestros y fraudes. El factor tiempo de procesamiento fue un factor importante, ya que se requería que la información fuera procesada y disponible para el día siguiente.

Diseño de Arquitectura de Hardware.

Se planteó un diseño de la arquitectura de hardware alineado a las recomendaciones del producto y que cumplir los requerimientos de negocio en relación con integración de datos.

Tabla 5

Requerimientos de hardware y software base

Server	Hardware	Sistema Operativo		RAM	CPU	Disco	Función	Ambiente
1	System x86-64	Red Hat Enterprise Linux (RHEL) Server 7.4		32GB	4	500GB	Metadatos Servicios	Desarrollo
2	System x86-64	Red Hat Enterprise Linux (RHEL) Server 7.4		64GB	8	2 TB	Engine	Desarrollo
3	System x86-64	Red Hat Enterprise Linux (RHEL) Server 7.4		32GB	4	500GB	Metadatos Servicios	Productivo
4	System x86-64	Red Hat Enterprise Linux (RHEL) Server 7.4		64GB	8	2 TB	Engine	Productivo

Nota. RHEL=Red Hat Enterprise Linux.

En base a los requerimientos de hardware y de software base (sistema operativo) se diseñó la distribución de discos, volumn gruoup y file systems, de cada servidor con el fin de que las operaciones de disco sean distribuidas equitativamente.

Tabla 6

Definición de requerimientos de hardware y software del servidor Metadatos/Servicios

Server	<<IP>>/<<Host Name>>			
Función	Metadata/Service s			
Sistema Operativo	RHEL 7.4 - x86-64			
Topología	SMP			
Cores	4			
Memoria RAM	32			
Volumn Group	Filesystem	Size GB	Mounted on	Descripción
vg_root	fs_vg_root	20	/	Root
vg_root	fs_vg_home	20	/home	Home directory
vg_opt	fs_vg_opt_lib	30	/opt	Library
vg_iismeta	fs_vg_issmeta	100	/Data/xmeta	Metadata Repository
vg_iismeta	fs_vg_issiadb	100	/Data/iadb	Analyzer Repository
vg_iismeta	fs_vg_issdsodb	100	/Data/dsodb	Operation Console Repository
vg_iis	fs_vg_iis_db2	20	/opt/IBM/db2	Installation db2
vg_iis	fs_vg_iis_was	20	/opt/IBM/WebSphere	Installation WebSphere
vg_root	fs_vg_tmp	20	/tmp	Temporary OS

Nota. RHEL=Red Hat Enterprise Linux.

Tabla 7

Definición de requerimientos de hardware y software del servidor Engine

Server	<<IP>>/<<Host Name>>			
Función	Engine			
Sistema Operativo	RHEL 7.4 - x86-64	RAM	64	
Topología	SMP	Core	8	
Volume Group	Filesystem	Size GB	Mounted on	Descripción
vg_root	fs_vg_root	20	/	Root
vg_root	fs_vg_home	20	/home	Home directory
vg_opt	fs_vg_opt_lib	50	/opt	Library
vg_iis	fs_vg_iis_eng	20	/opt/IBM/InformationServer	Installation Engine
vg_opt	fs_vg_opt_pryds	100	/ProyectosDs	DataStage Projects
vg_nodoa	fs_vg_ds_nod01	100	/datasets/nodo1	Dataset area
vg_nodob	fs_vg_ds_nod02	100	/datasets/nodo2	Dataset area
vg_nodoc	fs_vg_ds_nod03	100	/datasets/nodo3	Dataset area
vg_nodod	fs_vg_ds_nod04	100	/datasets/nodo4	Dataset area
vg_nodoa	fs_vg_sc_nod01	100	/scratches/nodo1	Scratch area
vg_nodob	fs_vg_sc_nod02	100	/scratches/nodo2	Scratch area
vg_nodoc	fs_vg_sc_nod03	100	/scratches/nodo3	Scratch area
vg_nodod	fs_vg_sc_nod04	100	/scratches/nodo4	Scratch area
vg_stg	fs_vg_stg_inbound	100	/Inbound	Inbound Area
vg_stg	fs_vg_stg_staging	100	/Staging	Staging Area
vg_stg	fs_vg_stg_outbound	100	/Outbound	Outbound Area
vg_root	fs_vg_tmp	50	/tmp	Temporary OS
vg_stg	fs_vg_tmpds	50	/tmpds	Temporary DataStage

Nota. RHEL=Red Hat Enterprise Linux.

Diseño de Arquitectura de Componentes IBM InfoSphere Information Server.

Se presentó al cliente la arquitectura de componentes de IBM InfoSphere Information Server sobre el diseño de arquitectura de hardware establecido. El objetivo fue mostrar como los componentes de la plataforma se integran.

Basado en las recomendaciones de producto se recomendó que los componentes de las capas de metadatos y servicios sean instalados en un servidor dedicado para sus funciones. Mientras que, los componentes de la capa de engine sean instalados en otro servidor dedicados para maximizar el procesamiento durante la ejecución de los procesos de integración de datos.

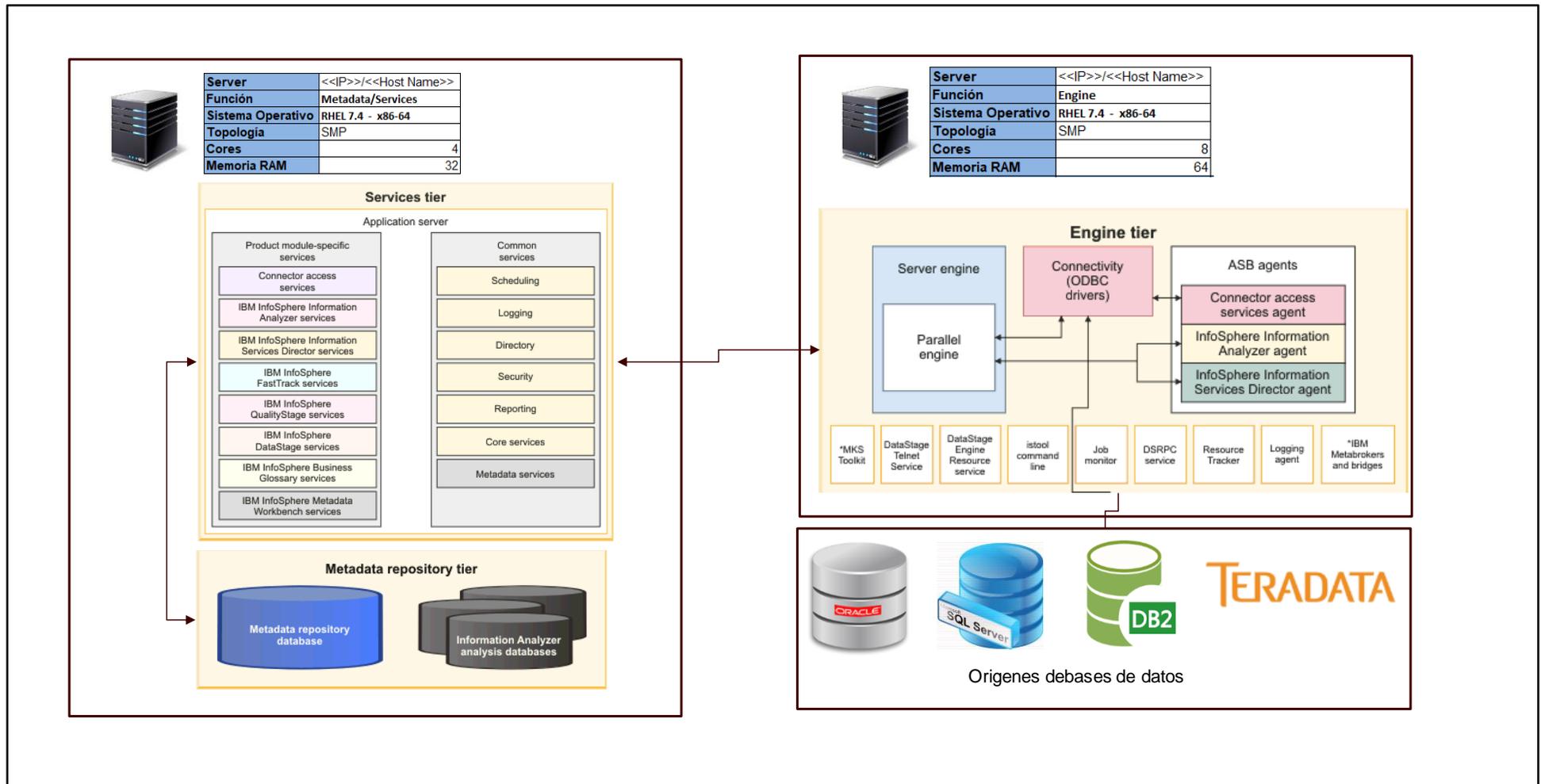


Figura 10. Arquitectura de componentes IBM InfoSphere Information Server.
Arquitectura propuesta para la implementación de la plataforma IBM InfoSphere Information Server. Elaboración propia.

Requerimientos de sistema IBM InfoSphere Information Server.

Para la instalación y configuración del software de la plataforma IBM InfoSphere Information Server se solicitó los siguientes requerimientos:

Seguridad.

Credenciales de usuario administrador (root)

Credenciales de usuario de dominio

Credenciales de usuario de orígenes de datos de datos, Oracle, db2, Microsoft Sql server y Teradata, con permisos de lectura, escritura, creación, eliminación de tablas.

Configuración de punto de conexión a la red interna

Software.

Instalación de librerías Linux

glibc-2.12-1.107.el6.x86_64

libXp-1.0.0-15.1.el6.x86_64

libXau-1.0.5-1.el6.x86_64

nss-softokn-freebl-3.14.3-9.el6.x86_64

pam-1.1.1-4.el6_0.1.x86_64

libaio-0.3.107-10.el6.x86_64

libstdc++-4.4.7-3.el6.x86_64

compat-libstdc++-33-3.2.3-69.el6.x86_64

libgcc-4.4.7-3.el6.x86_64

Instalación y configuración de clientes de orígenes de base de datos, Oracle, db2, Microsoft Sql server y Teradata.

Licencia.

Descarga y copia de instaladores del software IBM InfoSphere Information Server con la cuenta del cliente.

Preparación de los servidores.

Aprobados los diseños de arquitectura de hardware, componentes y los requerimientos de sistema el cliente se encargó de preparar los servidores con las características definidas. El arquitecto de TI del cliente elaboró la solicitud de hardware y software que luego se presentó al comité de TI del cliente.

Se sustentó ante el comité TI y respondí las interrogantes de los equipos seguridad, arquitectura, redes, operaciones de TI. Finalmente, el comité aprobó la preparación de los servidores y su entrega en los plazos establecidos en cronograma de actividades.

Revisión de los servidores y requerimientos de hardware y software.

Esta actividad consistió en la validar todo lo solicitado al cliente con respecto a la arquitectura de hardware, componentes y requerimientos de sistema.

En este tipo de proyectos IBM dispone de herramientas para la revisión y diagnóstico de las características de los servidores donde se instalarán productos IBM. En este caso se ejecutó la herramienta ISALite y en las imágenes a continuación muestro el resultado de la validación de los servidores.

Multi-Tier [New InfoSphere Information Server 11.3 or 11.5 - Fresh install]		
Result	Status Code	Description
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library libpthread.so.0 is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library libdl.so.2 is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library libcrypt.so.1 is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library libstdc++.so.6 is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library libm.so.6 is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library libXp.so.6 64bit is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library libgcc_s.so.1 is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library libc.so.6 is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2022I	Ensure that the required library ld-linux.so.2 is installed.
<input checked="" type="checkbox"/> PASSED	CDIPR2003I	Ensure that the InfoSphere Information Server installation directory /opt/IBM/InformationServer is empty or does not exist.
<input checked="" type="checkbox"/> PASSED	CDIPR2033I	Ensure that the InfoSphere Information Server installation directory path /opt/IBM/InformationServer does not contain spaces.
<input checked="" type="checkbox"/> PASSED	CDIPR2034I	Ensure that the directory path /opt/IBM/InformationServer length is not more than 83 characters.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /opt/IBM/InformationServer is 3350 MB. This drive or partition has 28351 MB available.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /tmp is 260 MB. This drive or partition has 6710 MB available.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /dev/mapper/vg_issuer-lv_issuer_db2 is 695 MB. This drive or partition has 28530 MB available.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /dev/mapper/vg_data-lv_data_opt is 3350 MB. This drive or partition has 28351 MB available.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /dev/mapper/vg_root-lv_root_05 is 520 MB. This drive or partition has 6710 MB available.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /dev/mapper/vg_root-lv_root_02 is 1000 MB. This drive or partition has 4692 MB available.

Figura 11. Validación de librerías y espacio en disco en el servidor metadatos/servicios.

Reporte de la herramienta de análisis de pre-requisitos de instalación de la plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Metadata Repository Tier [New InfoSphere Information Server 11.3 or 11.5 - Fresh install]		
Result	Status Code	Description
<input checked="" type="checkbox"/> PASSED	CDIPR2119I	Ensure that the operating system is at the correct fix pack level and has the required software packages for the DB2 installation.
<input checked="" type="checkbox"/> PASSED	CDIPR2021I	Ensure that required kernel parameter kernel.shmall is optimal.
<input checked="" type="checkbox"/> PASSED	CDIPR2021I	Ensure that required kernel parameter kernel.msgmni is optimal.
<input checked="" type="checkbox"/> PASSED	CDIPR2021I	Ensure that required kernel parameter kernel.msgmax is optimal.
<input checked="" type="checkbox"/> PASSED	CDIPR2021I	Ensure that required kernel parameter kernel.msgmnb is optimal.
<input checked="" type="checkbox"/> PASSED	CDIPR2021I	Ensure that required kernel parameter kernel.sem is optimal.
<input checked="" type="checkbox"/> PASSED	CDIPR2021I	Ensure that required kernel parameter kernel.shmmax is optimal.
<input checked="" type="checkbox"/> PASSED	CDIPR2015I	Ensure that the DB2 instance ID db2inst1 can write to the temporary directory /tmp.
<input checked="" type="checkbox"/> PASSED	CDIPR2121I	Ensure that the directory /opt/IBM/db2/V10 is writable by the root user.
<input checked="" type="checkbox"/> PASSED	CDIPR2004I	Ensure that the directory /opt/IBM/db2/V10 either exists and is writable or does not exist but can be created.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /opt/IBM/db2/V10 is 695 MB. This drive or partition has 28530 MB available.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /tmp is 260 MB. This drive or partition has 6710 MB available.
<input checked="" type="checkbox"/> PASSED	CDIPR2016I	The disk space requirement for /var is 1000 MB. This drive or partition has 4692 MB available.
<input checked="" type="checkbox"/> PASSED	CDIPR2017I	Ensure that the operating system tools required by DB2 are available.
<input checked="" type="checkbox"/> PASSED	CDIPR2121I	Ensure that the directory /usr/local/bin is writable by the root user.
<input checked="" type="checkbox"/> PASSED	CDIPR2004I	Ensure that the directory /usr/local/bin either exists and is writable or does not exist but can be created.
<input checked="" type="checkbox"/> PASSED	CDIPR2012I	Ensure that the DB2 instance port 50000 is available.
<input checked="" type="checkbox"/> PASSED	CDIPR2077I	Ensure that the DB2 installation directory /opt/IBM/db2/V10 is empty.

Figura 12. Validación de parámetros kernel y espacio en disco en el servidor metadatos/servicios. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Lográndose validar todos los requerimientos de hardware, software y de sistema fue posible cerrar el primer hito del proyecto “Diseño de Arquitectura” en los tiempos del cronograma de actividades.

Instalación de plataforma IBM InfoSphere Information Server.

En esta actividad se ejecutó los pasos de instalación del software IBM InfoSphere Information Server personalizando las opciones de instalación según las características de cada servidor entregado por el cliente.

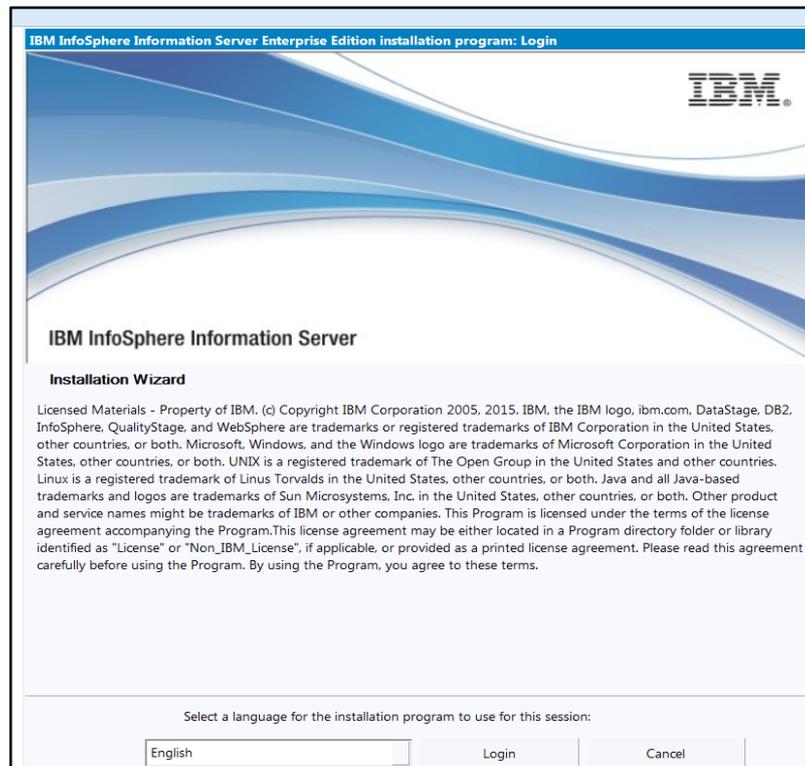


Figura 13. Inicio de instalación del software.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

El primer servidor estará dedicado a la instalación de los componentes de capas de metadatos y servicios

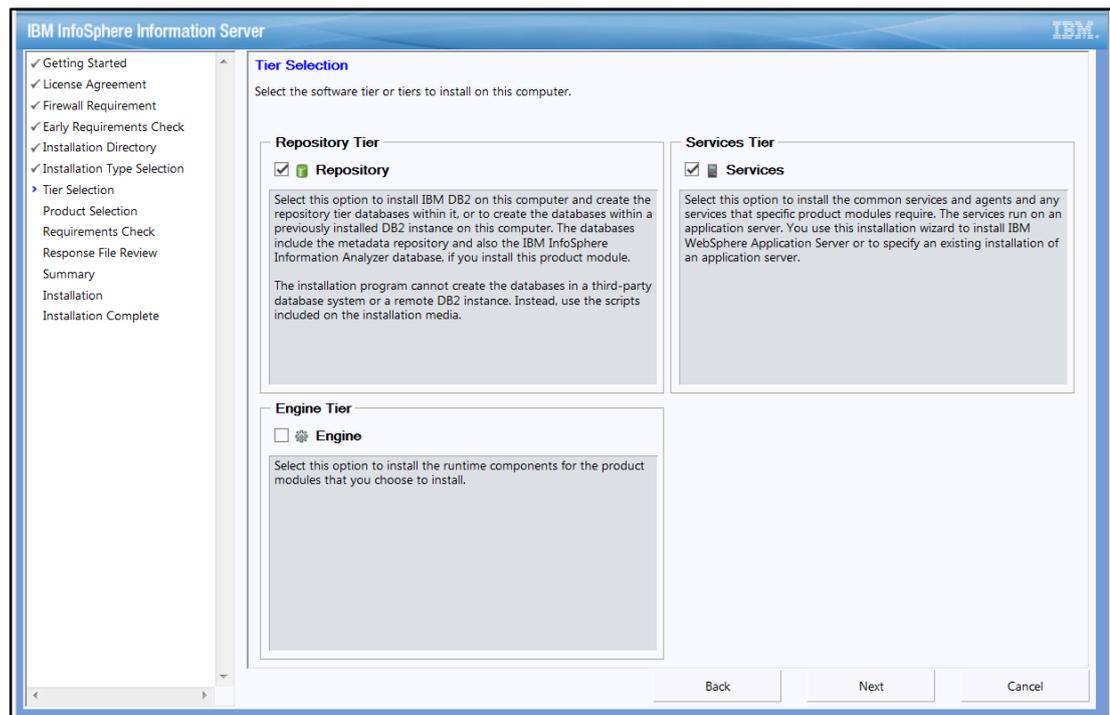


Figura 14. Selección de las capas de repositorio metadatos y servicios.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

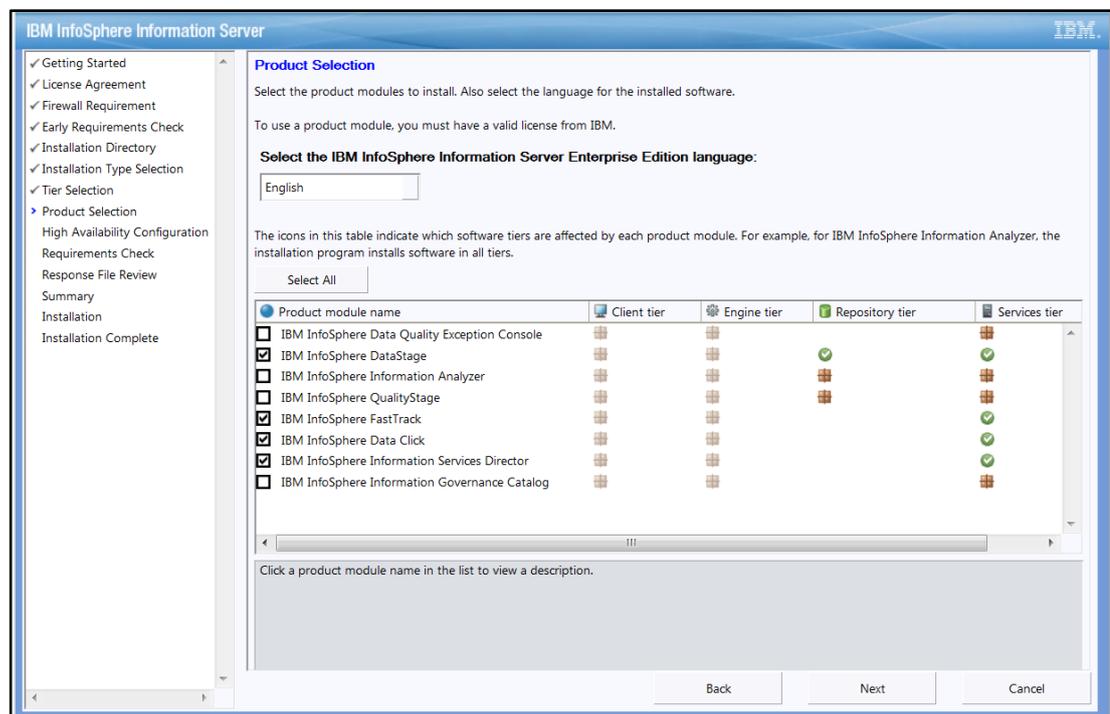


Figura 15. Selección de los componentes de integración de datos.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

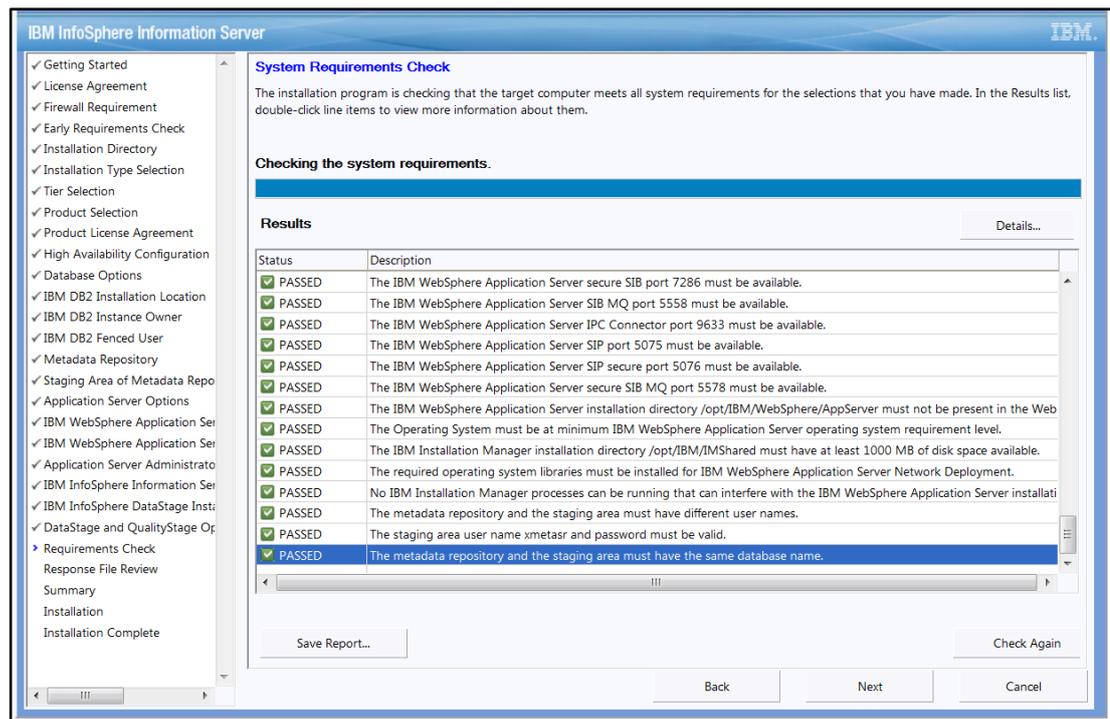


Figura 16. Validación de requerimientos de sistema.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

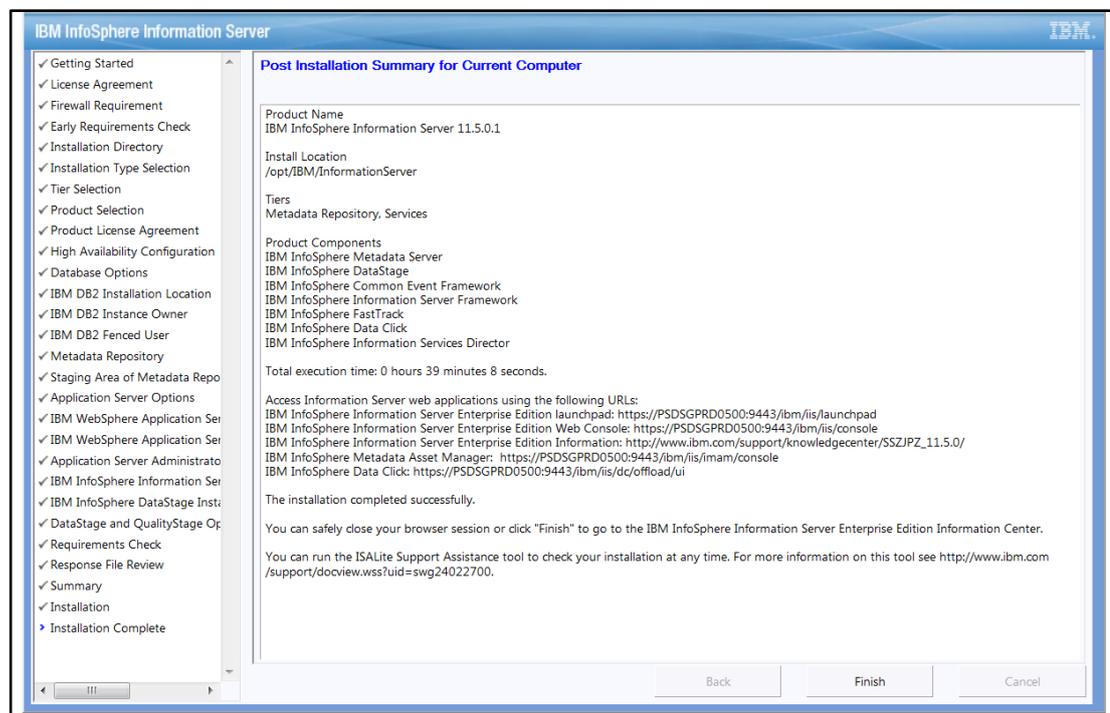


Figura 17. Finalización de instalación de las capas de metadatos y servicios.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Inicio de instalación del software para la capa de engine

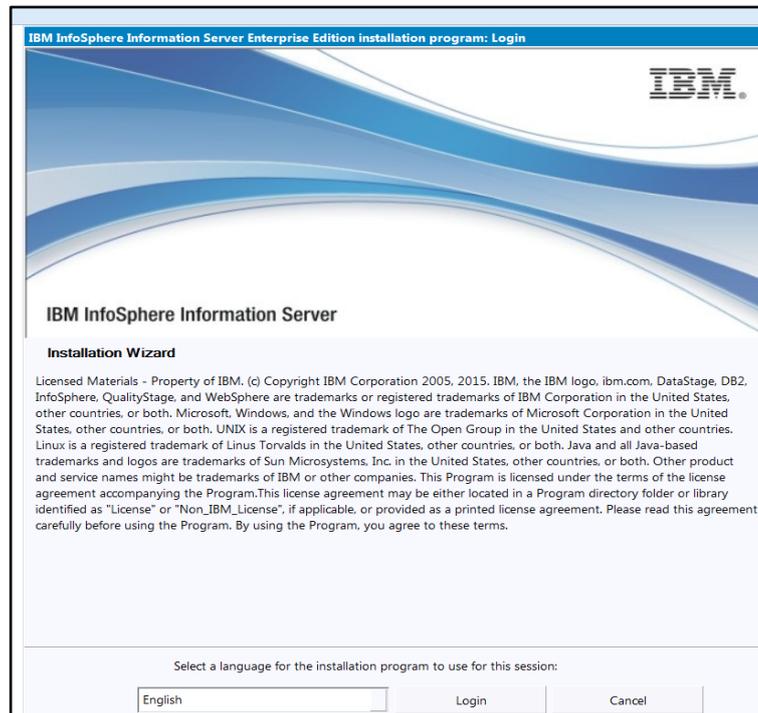


Figura 18. Inicio de instalación del software.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

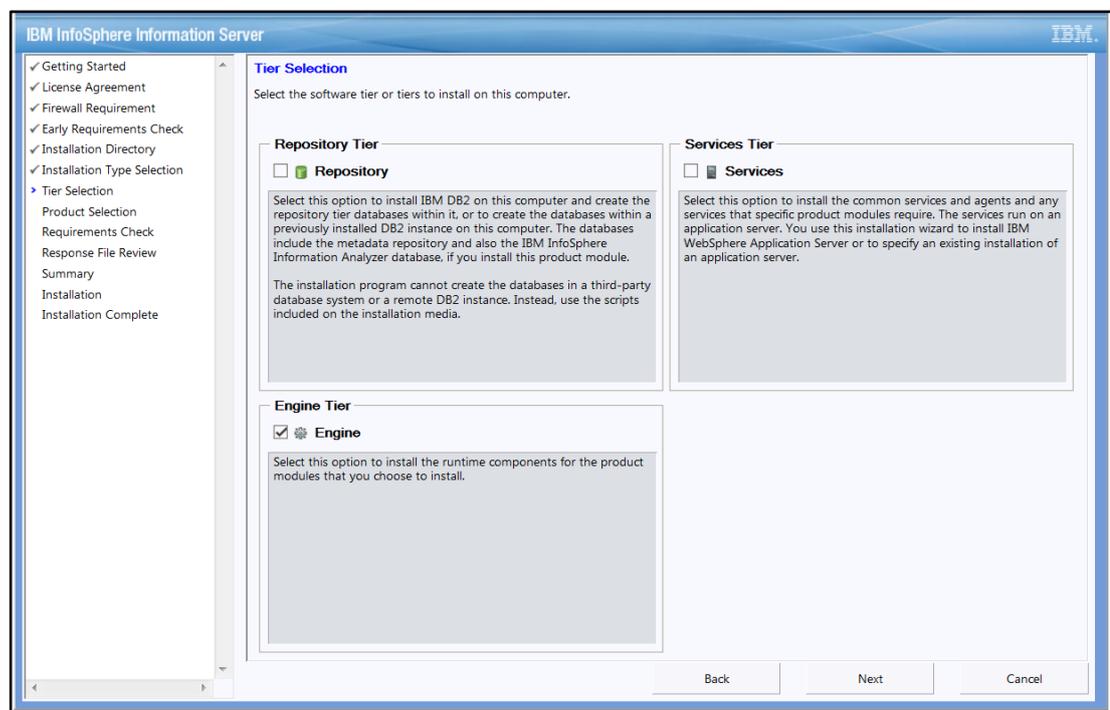


Figura 19. Selección de la capa de engine. Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

IBM InfoSphere Information Server Enterprise Edition Location

The IBM InfoSphere Information Server Enterprise Edition product modules must be registered and licensed. Enter the host name or IP address and the port number of the computer where you want to register this installation. For WebSphere cluster installations, enter the hostname and port number for the HTTP dispatcher that this cluster uses.

Host
PSDSGPRD0500

Port
9443

The IBM InfoSphere Information Server Enterprise Edition Administrator
isadmin

The IBM InfoSphere Information Server Enterprise Edition Administrator Password
••••••••

Back Next Cancel

Figura 20. Registro de capa engine a la capa de servicios.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

System Requirements Check

The installation program is checking that the target computer meets all system requirements for the selections that you have made. In the Results list, double-click line items to view more information about them.

Checking the system requirements.

Results Details...

Status	Description
✓ PASSED	The operating system must have sufficient memory.
✓ PASSED	Ensure that the current user has umask set to 022.
✓ PASSED	Ensure that the number of open file descriptors is greater or equal to 10240.
✓ PASSED	Ensure that the unsupported third party authentication systems are disabled.
✓ PASSED	No IBM InfoSphere Information Server engine processes can be running on the target computer.
✓ PASSED	Ensure kernel parameters are valid for DSEngine.
✓ PASSED	The tools that are required by the IBM InfoSphere Information Server engine must be present. Required tools are ed, ipcs, sed.
✓ PASSED	Ensure nosuid restriction has not been specified on the file system that contains executables on the path /opt/IBM/InformationServer.
✓ PASSED	Ensure IBM InfoSphere DataStage Administrator user dsadm exists.
✓ PASSED	Ensure the IBM InfoSphere DataStage administrative user dsadm owns its home directory /home/dsadm.
✓ PASSED	User dsadm must have read permission for Information Server image folder /Instaladores/is-suite.
✓ PASSED	Ensure that user dsadm is able to execute a script cleanly.
✓ PASSED	Ensure DataStage administrative user dsadm home directory exists.
✓ PASSED	DataStage engine requires glibc installed

Save Report... Check Again

Back Next Cancel

Figura 21. Validación de requerimientos de sistema.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

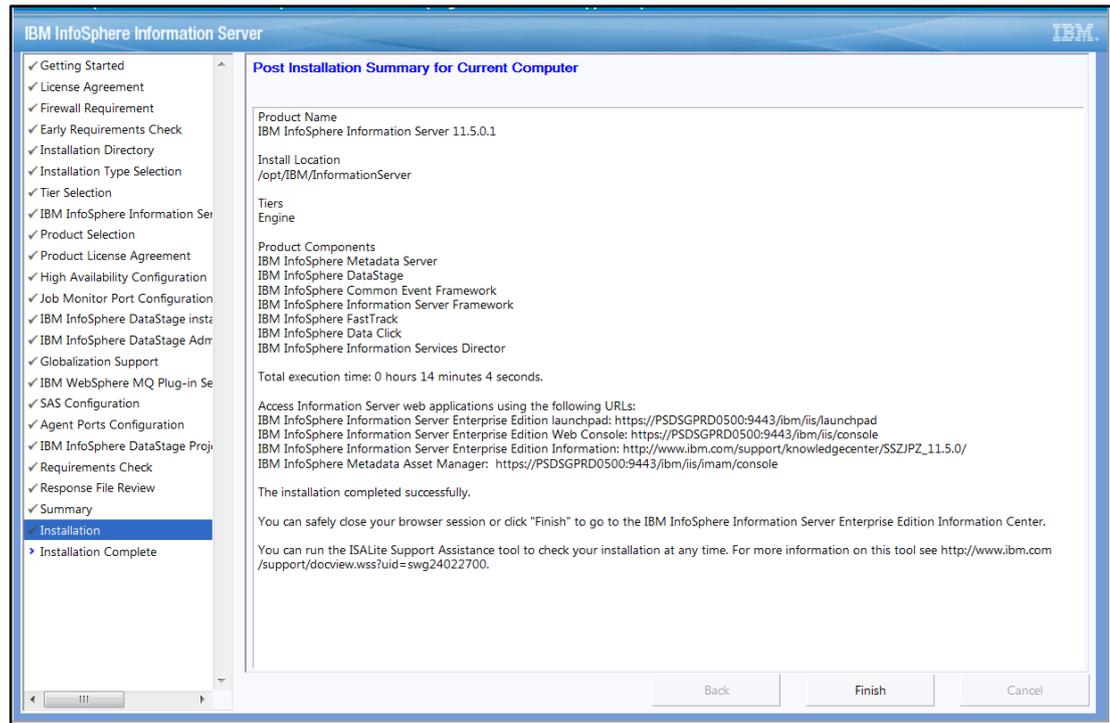


Figura 22. Finalización de instalación de la capa de engine.

Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Configuración de plataforma IBM InfoSphere Information Server.

Estas actividades consistieron en la adaptación de parámetros de la plataforma para que se integre adecuadamente a las características físicas del servidor (discos, memoria y CPU) y a los orígenes de bases de datos y directorio activo) del cliente

Seguridad.

Esta actividad consistió en integrar el directorio activo sincronizando las cuentas de usuario de dominio a la plataforma IBM InfoSphere Information Server mediante la configuración de un repositorio federado en el componente IBM WebSphere Application Server.

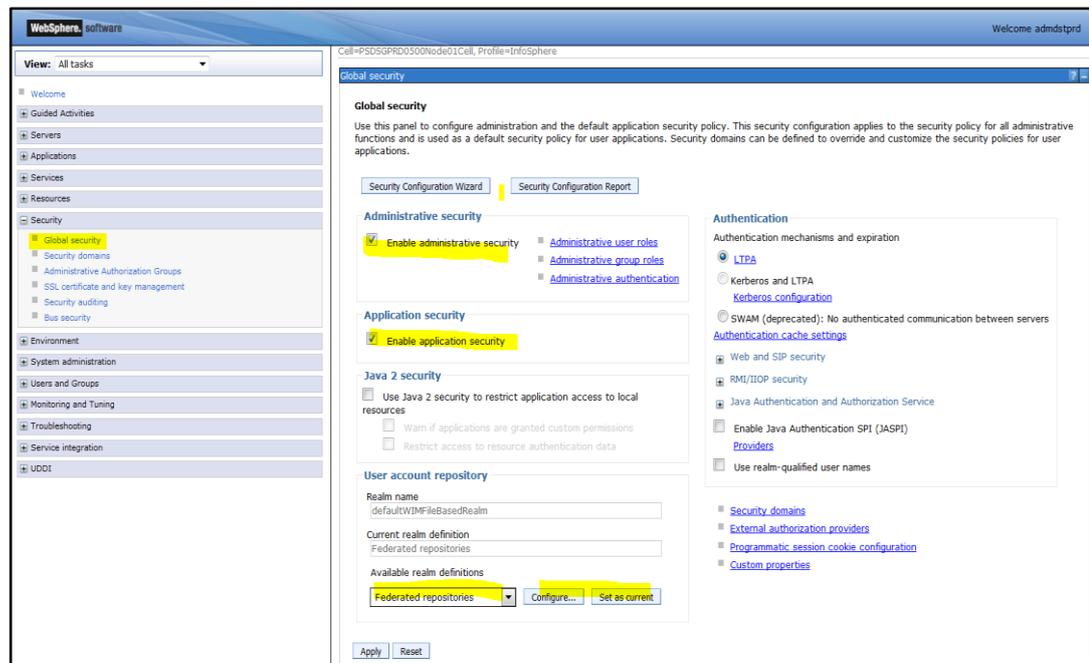


Figura 23. Configuración repositorio federado.

Configuración de seguridad de la plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Conexiones a bases de datos.

Esta actividad consistió en integrar la plataforma IBM InfoSphere Information Server a los clientes de bases de datos instalados y configurados en el Servidor Engine.

```
[root@PSDSGPRD0600 DSEngine]# cat dsenv |grep -A 6 ORACLE_BASE
export ORACLE_BASE=/opt/oracle
export ORACLE_HOME=$ORACLE_BASE/product/11.2.0/client_1
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$ORACLE_HOME/lib
export PATH=$PATH:$ORACLE_HOME/bin
export NLS_LANG=AMERICAN_AMERICA.AL32UTF8
export TNS_ADMIN=$ORACLE_HOME/network/admin
export SQLPATH=$ORACLE_HOME/bin

[root@PSDSGPRD0600 admin]# cat tnsnames.ora |grep -A 8 IGWPRD
IGWPRD =
(DESCRIPTION =
  (ADDRESS = (PROTOCOL = TCP) (HOST = psscnpd0100.pps.com.pe) (PORT = 1521))
  (CONNECT_DATA =
    (SERVER = DEDICATED)
    (SERVICE_NAME = igwprd)
  )
)
```

Figura 24. Configuración parámetros bases de datos Oracle.

Configuración de conectividad de la plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Proyectos de integración de datos.

Esta actividad consistió la creación y configuración de un proyecto de integración de datos con el componente IBM InfoSphere DataStage. El proyecto permitió que el equipo de integración de datos del cliente lo tome como referencia para futuros proyecto en la plataforma.

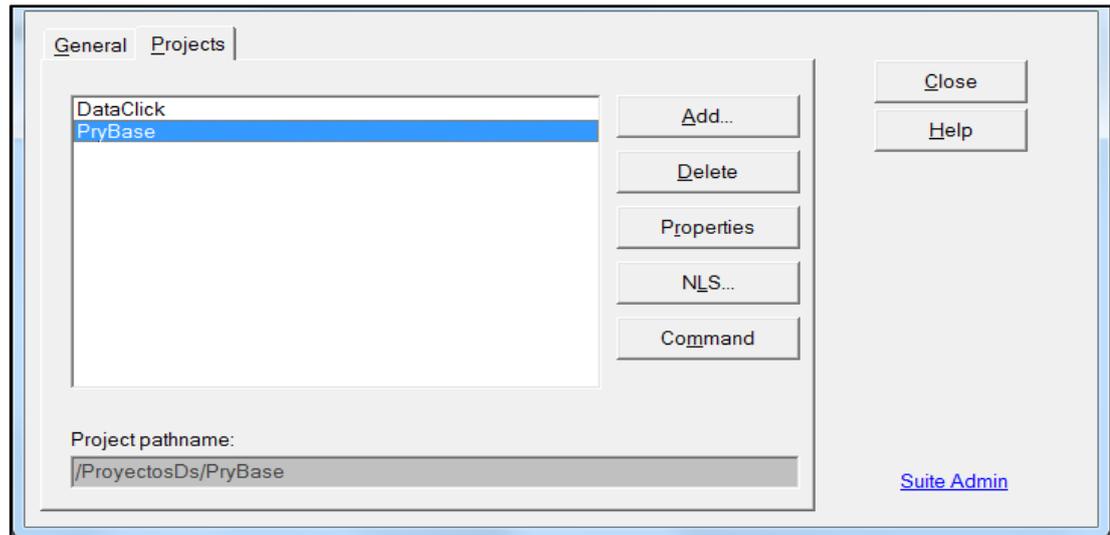


Figura 25. Configuración proyecto base de integración de datos.

Configuración de proyecto DataStage para la integración de datos. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

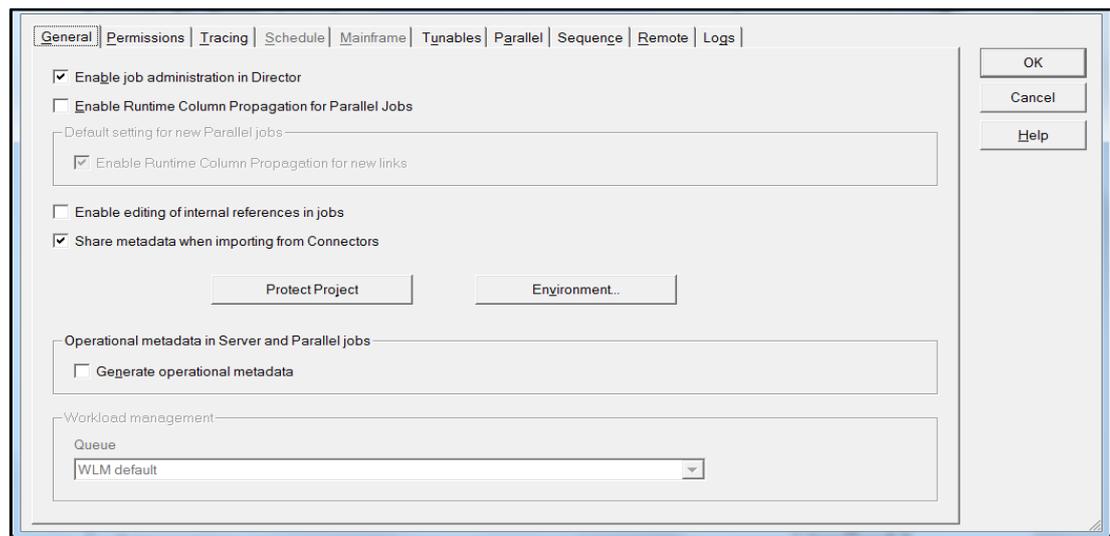


Figura 26. Configuración parámetros proyecto base de integración de datos.

Configuración de proyecto DataStage para la integración de datos. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

La configuración para integrar las capacidades de disco y los CPU del servidor Engine se realizó mediante la configuración de nodos de procesamiento. Esto permitió el aprovechamiento de las capacidades de procesamiento paralelo de la plataforma durante el desarrollo y ejecución de procesos de integración de datos con el componente IBM InfoSphere DataStage

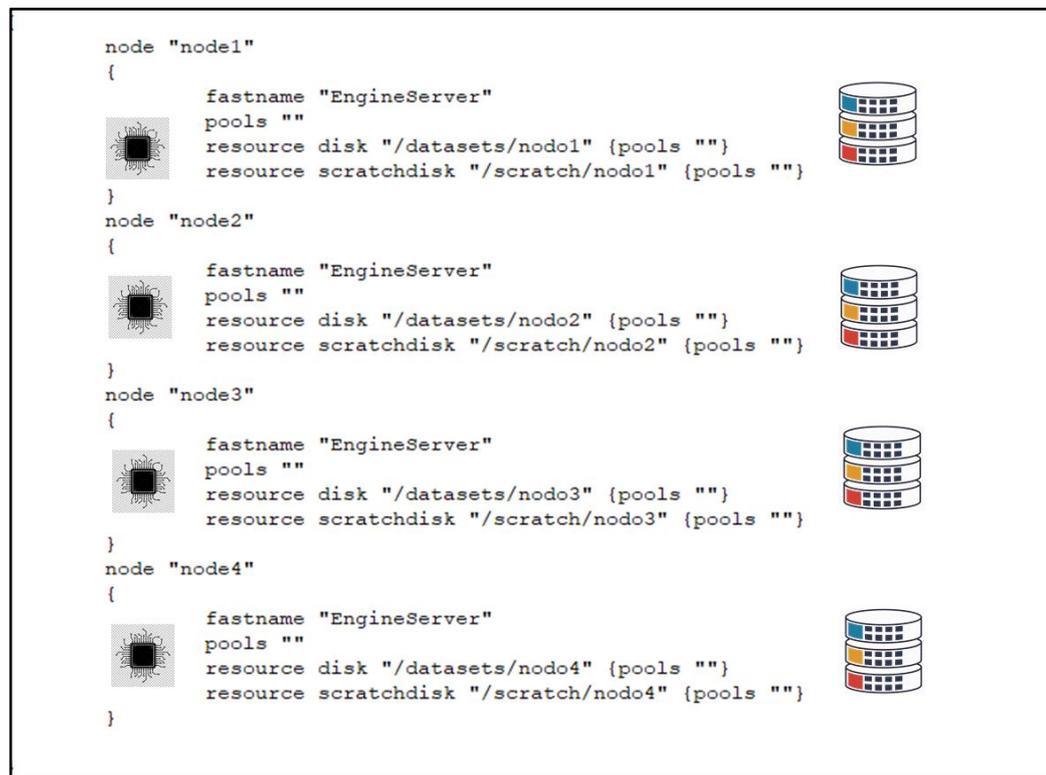


Figura 27. Configuración nodos de procesamiento.
Configuración de proyecto DataStage para la integración de datos. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Pruebas de configuración.

Esta actividad consistió en la verificación del correcto funcionamiento de los componentes de la plataforma, desde las consolas web de administración hasta el desarrollo de procesos de integración de datos para medir las capacidades de procesamiento en términos de volúmenes de datos y tiempo de procesamiento.

El componente de monitoreo de recursos y servicios de la plataforma, la consola de Operaciones está basada en un entorno web.

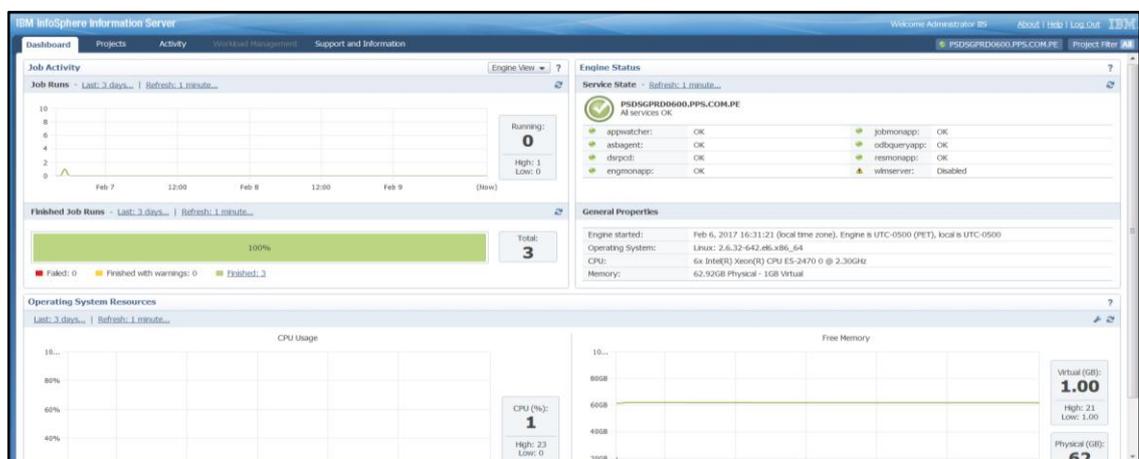


Figura 28. Consola web de monitoreo “Consola de Operaciones”.

Configuración monitoreo de la plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

La consola web administrativa, esta permite la configuración de roles de usuario dentro de la plataforma.



Figura 29. Consola web administrativa.

Configuración administrativa de la plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

El desarrollo de un trabajo de integración de datos con el componente IBM InfoSphere DataStage, se procesan alrededor de 10 millones de registros y se utilizan conexiones de bases de datos Oracle de forma nativa y odbc. El tiempo de procesamiento 45 segundos, lo cual es un buen indicador de performance frente a los tiempos de procesamiento actuales con el uso de scripts y programas que superan las 3 y 4 horas para el mismo volumen de datos

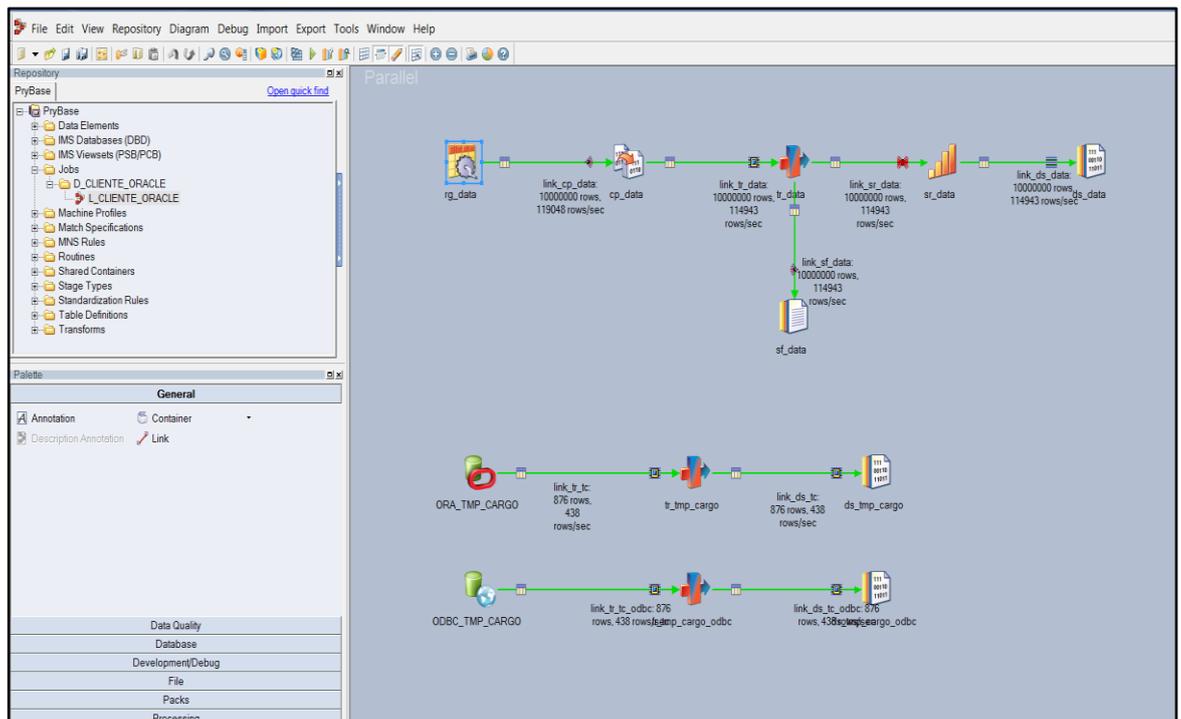


Figura 30. Lienzo de un trabajo de integración de datos con IBM InfoSphere DataStage. Pruebas de la configuración de la plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Nota ODBC= Open Database Connectivity.

Documentación de instalación y configuración.

Esta actividad consistió en la elaboración de documento de instalación y configuración de la plataforma, en formato IBM, por medio de captura de pantallas acompañadas de descripciones de cada actividad.

El documento reúne las configuraciones de línea base de la plataforma, con este documento el equipo de administradores de TI del cliente puede replicar la instalación en otros servidores, diseñar el plan de respaldo y mantenimiento de componentes.

Objetivo 2: Diseñar modelos de integración de datos basados en casos de uso de la empresa Seguros Generales, Lima, 2017.

Modelos y lineamientos de integración de datos.

Documento de modelos de integración de datos.

Esta actividad consistió en la elaboración de documentación que detalle los modelos de integración de datos recomendados basados en parámetros como volumen de datos, complejidad de lógica de transformaciones, cantidad de orígenes y destinos de bases de datos.

Los procesos de integración de datos con el componente IBM InfoSphere DataStage utilizan lienzos y objetos “stages”, los cuales permiten de manera gráfica organizarlos e implementar lógica para extraer, transformar y cargar datos aprovechando las capacidades del procesamiento paralelo de la plataforma.

En la imagen a continuación se muestra el modelo lineal de integración de datos. El modelo consiste en un solo lienzo que parte de un archivo fuente, lógica de transformación simple y una conexión de base datos como destino o carga. El diseño es recomendado para procesar volúmenes de datos menores e iguales a 1 millón de registros, en términos de tamaño de archivos están entre 1GB a 2GB y en tiempo de procesamiento puede estar entre 40 a 120 segundos.

- Volumen de Datos: <=1 Mill reg
- Nro de Transformaciones : <= 5 y simples transformaciones
- De punto a punto : 1 Origen a 1 Destino
- Nro de Stages en el Job: <=5

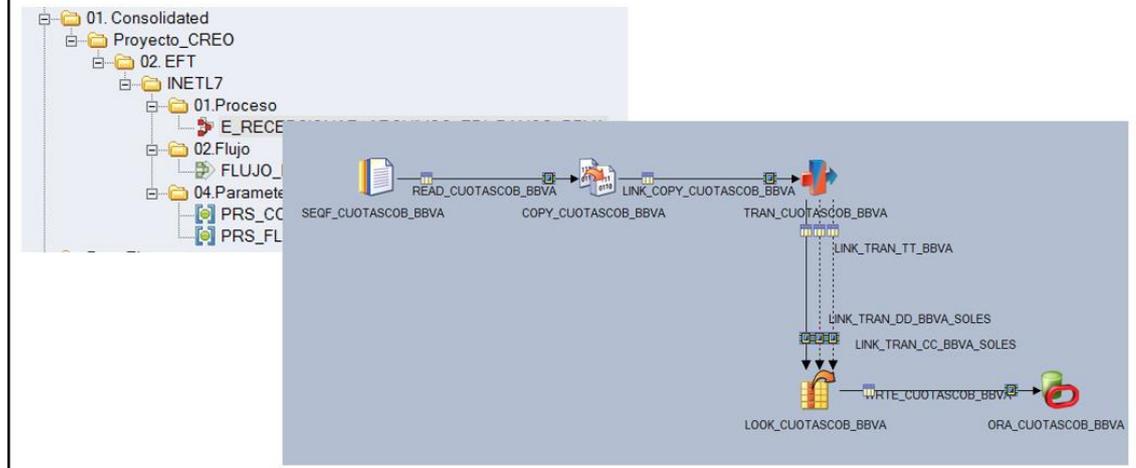


Figura 31. Modelo de integración de datos lineal.

Tomado de Chipana, J. (2017). *Modelos de integración de datos con IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

En la imagen a continuación se muestra el modelo ETL (Extracción, Transformación y Carga) de integración de datos. El modelo consiste en un uno o más lienzos que parte de uno o más fuentes de datos (archivos, tablas de bases de datos), lógica de transformación simple a compleja y una o más conexión de base datos como destino o carga. El diseño es recomendado para procesar volúmenes de datos mayores a 1 millón de registros, en términos de tamaño de archivos mayores a 2GB y en tiempo de procesamiento puede estar entre 120 a 360 segundos.

- **Volumen de Datos:** > 1 Mill reg
- **Nro de Transformaciones :** > 5 y simples y complejas transformaciones
- **De muchos a mucho :** Origen a Destino
- **Nro de Stages en el Job:** > 5

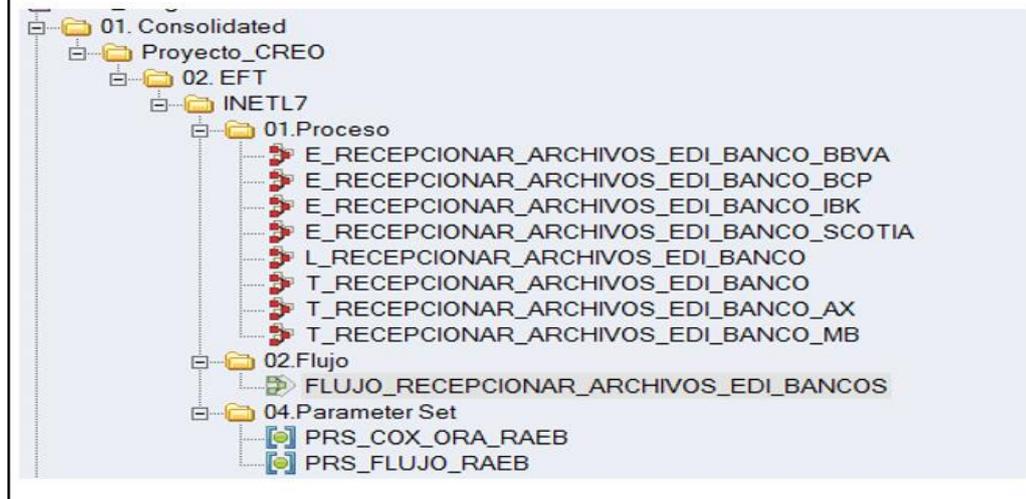


Figura 32. Modelo de integración de datos ETL.

Tomado de Chipana, J. (2017). *Modelos de integración de datos con IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

En la imagen se aprecia el lienzo donde se orquesta la secuencia de ejecución de uno o más procesos de integración de datos. Siendo posible paralelizar múltiples procesos de extracción, transformación y cargar de datos.

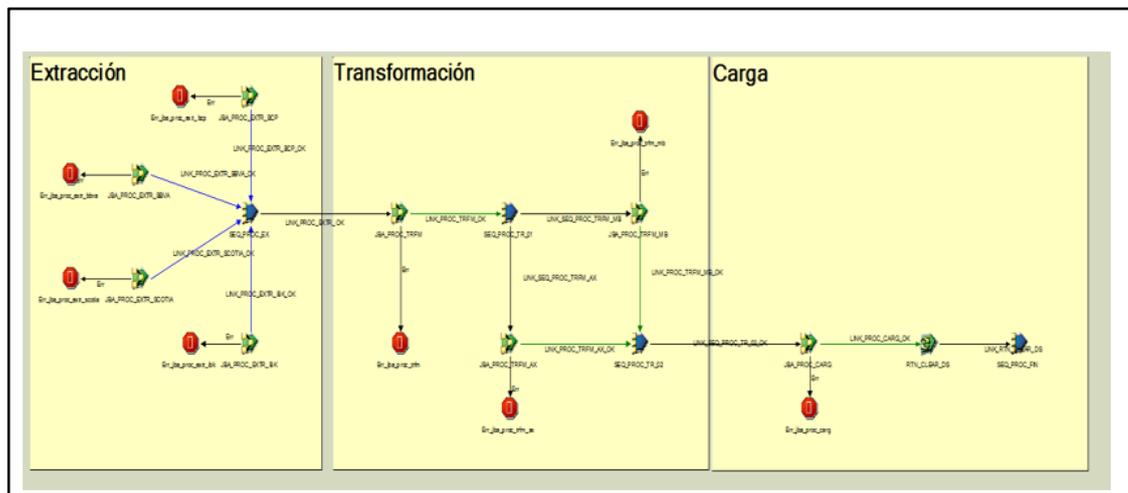


Figura 33. Modelo de secuencia de integración de datos.

Tomado de Chipana, J. (2017). *Modelos de integración de datos con IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Documento de lineamientos de integración de datos.

Esta actividad consistió en la elaboración de documentación que detalla los lineamientos para la implementación proyectos y trabajos de integración de datos con el componente IBM InfoSphere DataStage.

En las imágenes se aprecian la manera de como estructurar los proyectos de integración de datos, desde la definición de nombre del proyecto hasta la estructura de carpetas en el servidor.

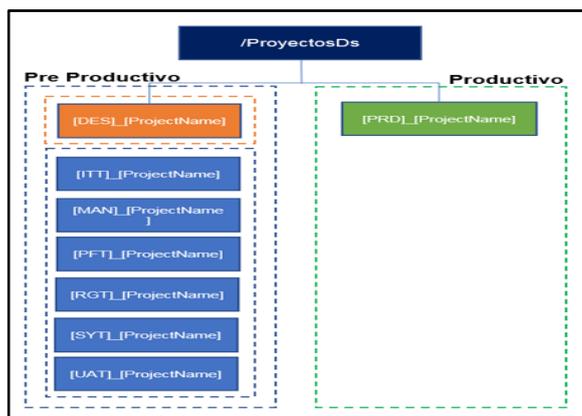


Figura 34. Modelo de estructura de proyectos de integración de datos.

Tomado de Chipana, J. (2017). *Lineamientos de integración de datos con IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

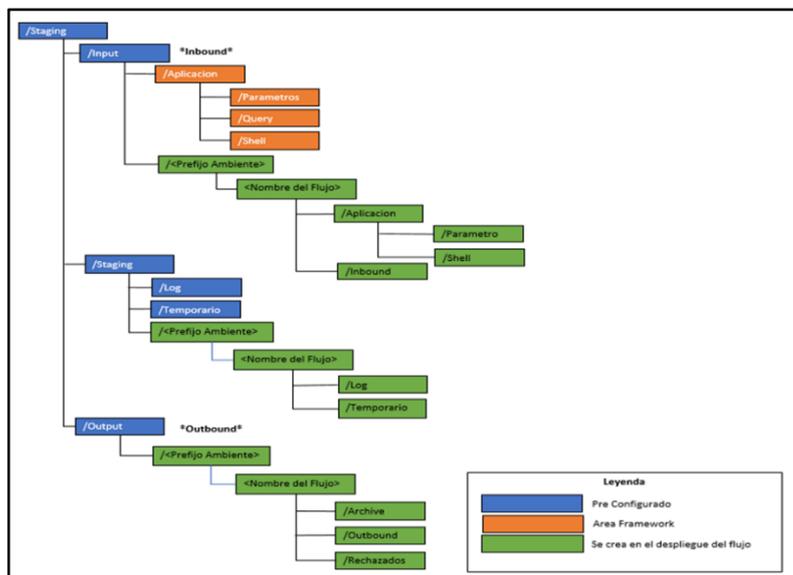


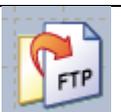
Figura 35. Modelo de estructura carpetas a nivel de servidor para proyectos de integración de datos. Tomado de Chipana, J. (2017). *Lineamientos de integración de datos con IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

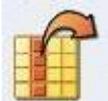
Se estableció un diccionario común para la definición de objetos como parámetros, procesos, proyectos, archivos, entidades, etc. La correcta definición de nombres a los objetos que forman parte de los proyectos y procesos de integración de datos es un aspecto crítico.

Definir un lenguaje común facilita el entendimiento de la lógica implementada en el proceso de integración de datos. Permite que las modificaciones a nivel de codificación sean más rápidas y sencillas de implementar.

Tabla 8

Diccionario de dominio de definición de nombres objetos en procesos de integración de datos.

stage	nomenclatura estándar	descripción	ejemplo
	ora_<nombre_tabla>	conexión nativa a oracle	ora_cliente
	<fuente>_<nombre_tabla>	conexión base de datos	ora_operaciones
	ftpe_<nombre_servidor>	conexión ftp entre servidores	ftpe_psbaodes01
	seqf_<nombre_archivo>	manipula archivos secuenciales	seqf_arch_contratos
	dats_<nombre_archivo>	manipula datasets	dats_arch_cliente
	copy_<nombre_stage_fuente>	realiza la copia de un input	copy_backup_cliente
	tran_<descripción>	realizar mapeos y transformaciones de datos	tran_valida_estado_cliente
	rowg_<descripción>	genera registros en tiempo de ejecución	rowg_teste_operación

stage	nomenclatura estándar	descripción	ejemplo
	shrd_<nombreproceso>	contenedor reutilizable compuesto de lógica	shrd_asigna_clave
	look_<descripción>	realiza de búsqueda de integridad referencial	look_garantia_operación
	join_<descripción>	realiza de búsqueda de integridad referencial	join_cliente_contrato
	merg_<descripción>	realiza de búsqueda de integridad referencial	merg_cliente_dirección
	funn_< descripción>	realiza la consolidación de "n" entradas	funn_cliente_activo_inactivo
	filt_<descripción>	realizar filtros de registros a partir de una condición dada	filt_cliente_ativo
	rmvd_<descripción>	realiza eliminación de registros duplicados a partir de una clave determinada	rmvd_operación_cliente
	sort_<descripción>	ordena una entrada de datos	sqrt_dt_operación
	agrg_<descripción>	realizar operaciones de agregación	agrg_contrato_cliente

Nota. Stage= Etapa u objeto con función específica.

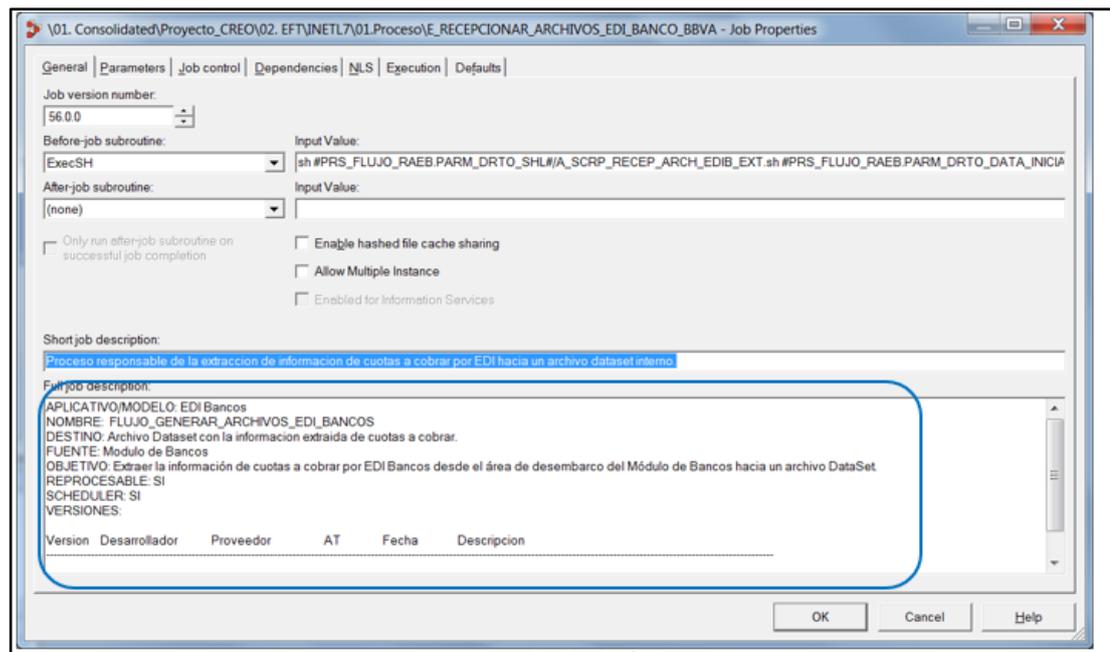


Figura 36. Documentación de procesos de integración de datos.
 Tomado de Chipana, J. (2017). *Lineamientos de integración de datos con IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Objetivo 3: Implementar una arquitectura orientada a servicios, eliminar duplicidad de funciones, uso eficiente de recursos de hardware, reducir el esfuerzo en tareas administrativas y despliegue procesos de integración de datos de la empresa Seguros Generales, Lima, 2017.

Esta actividad consistió en capacitación a miembros del equipo técnico del cliente con perfiles de administración y desarrollo.

Taller de administración de plataforma IBM InfoSphere Information Server.

En esta actividad se coordinó 4 sesiones diarias de 4 horas para participantes con el perfil de administración de la plataforma. Los participantes debían tener conocimientos de administración de servidores systems x86, Linux RHEL, y administración de bases de datos. Con estos conocimientos, los participantes les resultaron más sencillo interiorizar los conceptos de la plataforma IBM InfoSphere Information Server y con ello dar soporte a los usuarios y desarrolladores.

El taller inicia con una introducción funcional de los componentes de IBM InfoSphere Information Server, para luego enfocarse en las tareas de administración de la plataforma, en los servidores instalados, incluyéndose administración de usuarios, administración de sesiones y administración de reportes. Las tareas administrativas pueden ser realizadas con el uso de clientes de escritorio como también desde líneas de comando.

A continuación de los tópicos desarrollados en el taller de administración:

- Introducción IBM InfoSphere Information Server

- Tareas administrativas con Information Server Clients

- Autenticación y seguridad de usuarios

- Gestión de sesiones

- Gestión de reportes

- Herramientas administrativas

- Gestión de repositorio de metadatos

Taller de desarrollo de integración de datos

En esta actividad se coordinó 4 sesiones diarias de 4 horas para participantes con el perfil de analistas de datos. Los participantes debían tener conocimientos en lenguajes de programación como java, sql y bash script y modelamiento de bases de datos. Con estos conocimientos, los participantes les resultaron más sencillo interiorizar los conceptos de la plataforma IBM InfoSphere Information Server y con ello desarrollar los procesos de integración de datos siguiendo los lineamientos y plantillas definidas para lograr el óptimo desempeño de los procesos en la plataforma.

El taller inicia con una introducción funcional de los componentes de IBM InfoSphere Information Server, para luego enfocarse en técnicas de desarrollo de procesos de integración de datos con el componente IBM InfoSphere DataStage.

A continuación de los tópicos desarrollados en el taller de desarrollo de procesos de integración de datos:

- Introducción IBM InfoSphere Information Server
- Introducción a la arquitectura parallel framework
- Diseño, compilación y ejecución de Jobs
- Técnicas de particionamiento y combinación de datos
- Técnicas de ordenamiento de datos
- Tipos de datos en parallel framework
- Componentes reusables
- Optimización de procesos

ANÁLISIS Y RESULTADOS

Análisis

Seguros Generales realizó un balance de gastos en diciembre 2016 relacionado a integración de datos, adquisición de hardware y software realizadas por las áreas de negocio, costo de mantenimiento y soporte de la infraestructura adquirida y costos por inconsistencia de datos. Este último fue reportado por el área de marketing y ventas como resultado de la campaña 2016 de venta de seguros, inconsistencia y duplicidad los datos de cliente, el cliente ya cuenta con el producto, fueron algunos de los factores que afectaron a la campaña.

Tabla 9

Balance de gastos relacionados a integración de datos 2016

Concepto	Monto
Bases de datos	\$ 45,600.00
Herramientas ETL	\$ 144,000.00
Herramientas visualización	\$ 44,000.00
Costo de mantenimiento y soporte de la infraestructura	\$ 125,000.00
Costo por inconsistencia de datos	\$ 822,480.00
Total	\$1,181,080.00

Nota. ETL= Extracción, Transformación y Carga.

Los gastos relacionados a integración de datos representan 33% de los ingresos anuales por venta de seguros. Un porcentaje que desde 2010 se ha incrementado, en función a los datos en la organización, y se han convertido en un riesgo crítico para alcanzar las metas.

Reducir inconsistencia en los datos de clientes permitió entregar al equipo de fuerza de ventas datos más confiables y con ello reducir los índices de operaciones rechazadas. En 2017 se realizaron un aproximado de 358 mil operaciones de venta de seguros, de los cuales solo las 98 mil operaciones resultaron exitosas. Factores como no se encontró al cliente y el cliente ya tenía el producto resultaban ser por inconsistencia de en los datos. Al final de 2018, con la plataforma IBM InfoSphere Information Server implementada, de 209 mil operaciones de venta de seguros realizadas 155 mil fueron exitosas. Mientras que, 4 mil operaciones derivan de inconsistencia en los datos.

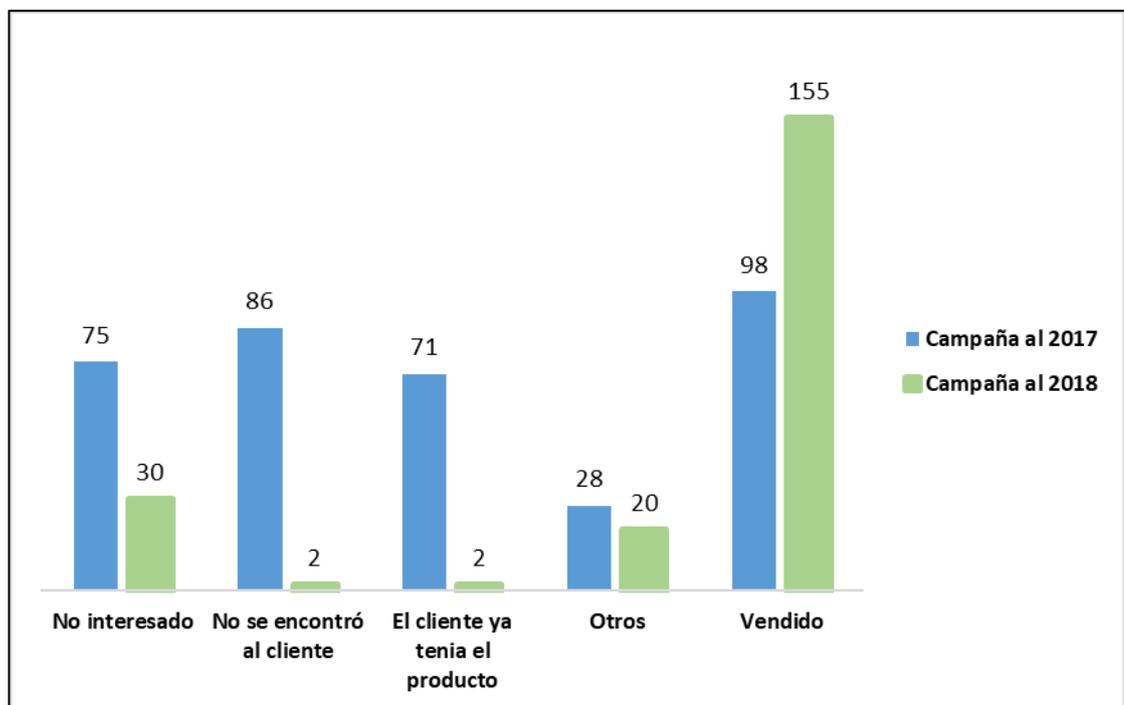


Figura 37 Comparativo de operaciones en campaña de venta de seguros 2017 y 2018. Las cifras representan miles de operaciones. Tomado de Chipana, J. (2018). Caso de éxito IBM InfoSphere Information Server de Seguros Generales. Material inédito. IBM.

Resultados

Arquitectura de TI de Seguros Generales, como parte del proyecto de desarrollo de data warehouse, buscó que la integración de datos sea capaz cumplir con los objetivos del negocio y reducir los gastos relacionados a integración datos con la implementación la plataforma IBM InfoSphere Information Server.

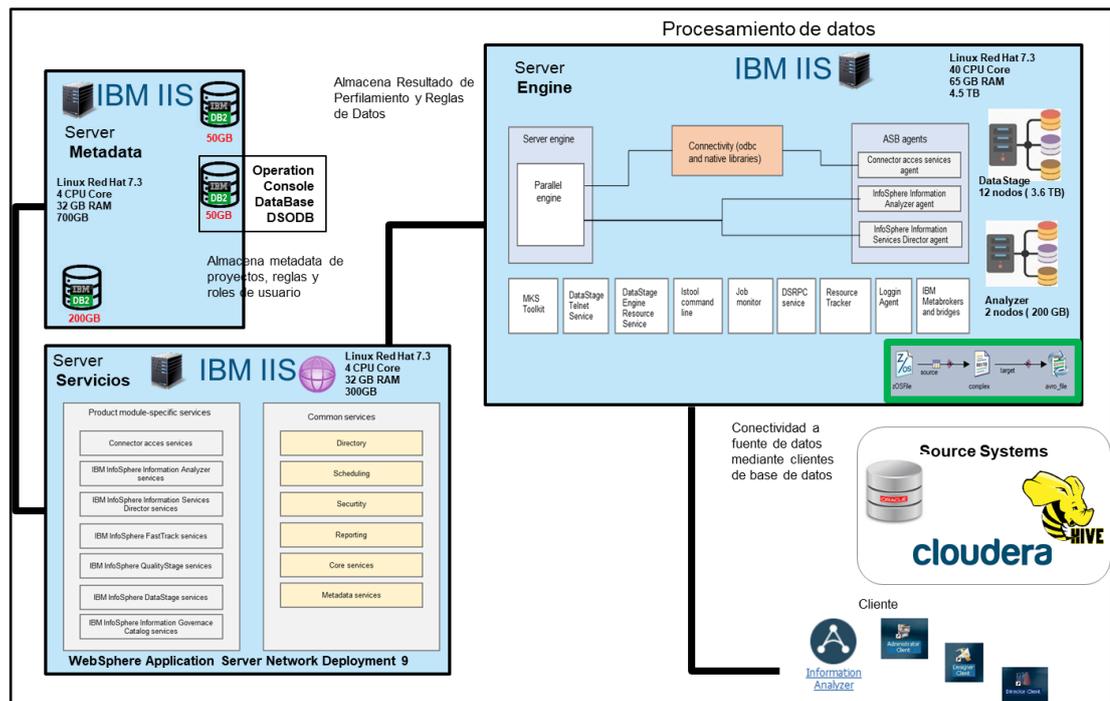


Figura 38 Arquitectura final de integración de datos en Seguros Generales.
Tomado de Chipana, J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Material inédito. IBM.

Alineada a la metodología de Kimball se consideró lo definido por Seguros Generales en las actividades de planificación de proyecto y definición de requerimientos para lograr con éxito las actividades de diseño de arquitectura, instalación de productos y diseño y desarrollo de ETL en el proyecto de implementación de la plataforma IBM InfoSphere Information Server.

El proyecto de implementación de la plataforma IBM InfoSphere Information Server fue desarrollado en 2 meses, con un costo de inversión de \$114,021.47 aproximadamente. El objetivo de recupero fue lograr el 12% de \$98,340.00 de gastos mensual de integración de datos en 12 meses.

Tabla 10

Resumen general inversión del proyecto de implementación de la plataforma IBM InfoSphere Information Server

Resumen general inversión	Total
Equipo de trabajo	\$ 37,840.00
Inversión tecnológica	\$ 65,580.84
Reserva de contingencia (5%)	\$ 5,171.04
Total, línea base	\$ 108,591.88
Reserva de gestión (5%)	\$ 5,429.59
Presupuesto del proyecto (CIP)	\$ 114,021.47

Nota. CIP= Costo de inversión de proyecto.

Tabla 11

Resumen de beneficios esperado mensual del proyecto de implementación de la plataforma IBM InfoSphere Information Server

Beneficios	
Promedio de ventas mensuales en seguros (vida, vehicular, salud, soat y primas)	\$ 298,000.00
Gastos por Integración de datos (33%)	\$ 98,340.00
Recupero por Integración de datos (12%)	\$ 11,880.00

Nota. Objetivo de recupero del 12% en 12 meses, incluyen los 2 meses de proyecto.

Tabla 12

Flujo de trabajo de proyecto de implementación de la plataforma IBM InfoSphere Information Server

	Mes 0	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
Beneficios		\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80
Flujo Neto	\$ (114,021.47)	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80	\$ 11,800.80
Acumulado	\$ (114,021.47)	\$ (102,220.67)	\$ (90,419.87)	\$ (78,619.07)	\$ (66,818.27)	\$ (55,017.47)	\$ (43,216.67)	\$ (31,415.87)	\$ (19,615.07)	\$ (7,814.27)	\$ 3,986.53	\$ 15,787.33	\$ 27,588.13

Nota. La proyección de recupero del proyecto es de 12 meses.

Tabla 13

Resumen de VAN, TIR y ROI del proyecto de implementación de la plataforma IBM InfoSphere Information Server

Concepto	Valor	Observación
Tasa de descuento mensual (COK) %	0.9%	El COK anual es de 12%
Valor actual neto (VAN)	\$19,226.43	Viable
Tasa interna de retorno (TIR)	3.5%	Aceptado, siendo mayor al COK
Retorno de la inversión (ROI)	124.20%	

Nota. VAN=Valor Actual Neto, TIR= Tasa Interna de Retorno y ROI=Retorno sobre la Inversión.

Conclusiones

La inversión del proyecto de implementación de la plataforma IBM InfoSphere Information Server fue de \$ 114,021.47, el VAN de la inversión en 2017 fue de \$ 19,226.43. Al resultar positivo indicó que el proyecto generó rentabilidad. De igual manera lo mostrado por el valor TIR de 3.5%, comparada con la tasa de descuento mensual COK 0.9% es mayor, lo que significa que está por encima de la mínima exigible. Con un ROI de 124.20 % el proyecto retornará la inversión a partir del mes 10. Se concluye que el proyecto fue rentable.

Con relación al objetivo, analizar los componentes internos y externos con relación a la arquitectura de integración de datos en la empresa Seguros Generales, Lima, 2017. Con el análisis y la implementación de la plataforma IBM InfoSphere Information Server se logró unificar los componentes de la arquitectura de información de la organización.

Con relación al objetivo, diseñar modelos de integración de datos basados en casos de uso de la empresa Seguros Generales, Lima, 2017. Con la implementación de modelos y lineamientos de integración de datos se logró establecer un marco de referencia y una plataforma enfocada en la colaboración de activos de información.

Con relación al objetivo, implementar una arquitectura orientada a servicios, eliminar duplicidad de funciones, uso eficiente de recursos de hardware, reducir el esfuerzo en tareas administrativas y despliegue procesos de integración de datos de la empresa Seguros Generales, Lima, 2017. El desarrollo del taller de administración de la plataforma IBM InfoSphere Information Server logró que el equipo de soporte y operaciones de TI, responsable del soporte, adoptase los conocimientos necesarios para cumplir con las tareas de administración y con ello asegurar la funcionalidad de la plataforma.

Se logró establecer un vínculo de confianza con el cliente e IBM para futuros proyectos.

Recomendaciones

Tras la implementación la plataforma IBM InfoSphere Information Server en Seguros Generales, se recomendó a Seguros Generales un plan de mantenimiento de la plataforma. Lo que consistió en tareas automatizadas para la depuración de proyectos, trabajos, archivos binarios de datos y archivos temporales a nivel de sistema operativo. Con el objetivo de garantizar la operatividad de la plataforma y su interacción con los usuarios.

Cuando un cliente adquiere una licencia de software IBM, tiene el derecho perpetuo a utilizarla la versión de software adquirida. Sin embargo, el soporte ante incidentes está supeditada al ciclo de vida de la versión del software. En este sentido, IBM ofrece una amplia adquirir los servicios de soporte y renovación de licencia que se adaptan a las necesidades y proyectos del cliente. Por lo tanto, es recomendable adquirir un plan de soporte de servicios y renovación de licencia para mitigar los riesgos de quedarse sin soporte del software.

LISTA DE FIGURAS

<i>Figura 1. Ubicación IBM del Perú en el distrito de la Molina, Lima, Perú. Tomado de Google Maps.</i>	6
<i>Figura 2. Organigrama IBM del Perú. Elaboración propia.</i>	8
<i>Figura 3. Matriz FODA de integración de datos en Seguros Generales.</i>	12
<i>Figura 4. Análisis Ishikawa del problema en Seguros Generales.</i>	13
<i>Figura 5. Diagrama de Pareto determina las principales causas del problema en Seguros Generales.</i>	13
<i>Figura 6. Inventario anual de adquisiciones de infraestructura de hardware y software.</i>	14
<i>Figura 7. Funciones de integración de la plataforma IBM InfoSphere Information Server.</i>	22
<i>Figura 8. Componentes de la plataforma IBM InfoSphere Information Server.</i>	24
<i>Figura 9. Metodología del ciclo de vida de Kimball.</i>	27
<i>Figura 10. Arquitectura de componentes IBM InfoSphere Information Server.</i>	36
<i>Figura 11. Validación de librerías y espacio en disco en el servidor metadatos/servicios.</i>	38
<i>Figura 12. Validación de parámetros kernel y espacio en disco en el servidor metadatos/servicios. Tomado de Chipana, J. (2017). Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales. Material inédito. IBM.</i>	39
<i>Figura 13. Inicio de instalación del software.</i>	40
<i>Figura 14. Selección de las capas de repositorio metadatos y servicios.</i>	41
<i>Figura 15. Selección de los componentes de integración de datos.</i>	41
<i>Figura 16. Validación de requerimientos de sistema.</i>	42
<i>Figura 17. Finalización de instalación de las capas de metadatos y servicios.</i>	42
<i>Figura 18. Inicio de instalación del software.</i>	43
<i>Figura 19. Selección de la capa de engine. Proceso de instalación plataforma IBM InfoSphere Information Server. Tomado de Chipana, J. (2017). Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales. Material inédito. IBM.</i>	43
<i>Figura 20. Registro de capa engine a la capa de servicios.</i>	44
<i>Figura 21. Validación de requerimientos de sistema.</i>	44
<i>Figura 22. Finalización de instalación de la capa de engine.</i>	45
<i>Figura 23. Configuración repositorio federado.</i>	46
<i>Figura 24. Configuración parámetros bases de datos Oracle.</i>	47
<i>Figura 25. Configuración proyecto base de integración de datos.</i>	48
<i>Figura 26. Configuración parámetros proyecto base de integración de datos.</i>	48
<i>Figura 27. Configuración nodos de procesamiento.</i>	49
<i>Figura 28. Consola web de monitoreo “Consola de Operaciones “.</i>	50
<i>Figura 29. Consola web administrativa.</i>	50
<i>Figura 30. Lienzo de un trabajo de integración de datos con IBM InfoSphere DataStage.</i>	51
<i>Figura 31. Modelo de integración de datos lineal.</i>	54
<i>Figura 32. Modelo de integración de datos ETL.</i>	55
<i>Figura 33. Modelo de secuencia de integración de datos.</i>	55
<i>Figura 34. Modelo de estructura de proyectos de integración de datos.</i>	56
<i>Figura 35. Modelo de estructura carpetas a nivel de servidor para proyectos de integración de datos. Tomado de Chipana, J. (2017). Lineamientos de integración</i>	

<i>de datos con IBM InfoSphere Information Server de Seguros Generales. Material inédito. IBM.</i>	56
<i>Figura 36. Documentación de procesos de integración de datos.</i>	59
<i>Figura 37 Comparativo de operaciones en campaña de venta de seguros 2017 y 2018. Las cifras representan miles de operaciones. Tomado de Chipana, J. (2018). Caso de éxito IBM InfoSphere Information Server de Seguros Generales. Material inédito. IBM.</i>	63
<i>Figura 38 Arquitectura final de integración de datos en Seguros Generales.</i>	64

LISTA DE TABLAS

Tabla 1 Resumen de inventario de adquisiciones de servidores del 2010 al 2016	14
Tabla 2 Resumen de costos por adquisiciones de servidores desde 2010 a 2016	15
Tabla 3 Cronograma de actividades del proyecto de implementación de la plataforma IBM InfoSphere Information Server	29
Tabla 4 Cronograma de actividades del proyecto de implementación de la plataforma IBM InfoSphere Information Server	30
Tabla 5 Requerimientos de hardware y software base	32
Tabla 6 Definición de requerimientos de hardware y software del servidor Metadatos/Servicios	33
Tabla 7 Definición de requerimientos de hardware y software del servidor Engine	34
Tabla 8 Diccionario de dominio de definición de nombres objetos en procesos de integración de datos	57
Tabla 9 Balance de gastos relacionados a integración de datos 2016	62
Tabla 10 Resumen general inversión del proyecto de implementación de la plataforma IBM InfoSphere Information Server	65
Tabla 11 Resumen de beneficios esperado mensual del proyecto de implementación de la plataforma IBM InfoSphere Information Server	66
Tabla 12 Flujo de trabajo de proyecto de implementación de la plataforma IBM InfoSphere Information Server	66
Tabla 13 Resumen de VAN, TIR y ROI del proyecto de implementación de la plataforma IBM InfoSphere Information Server	67

REFERENCIAS

- Chipana J. (2017). *Instalación y configuración plataforma IBM InfoSphere Information Server de Seguros Generales*. Lima, Perú: [s.n]
- Chipana J. (2017). *Modelos de integración de datos con IBM InfoSphere Information Server de Seguros Generales*. Lima, Perú: [s.n]
- Chipana J. (2017). *Tareas Administrativas. IBM InfoSphere Information Server Administration Workshop*. Lima, Perú: [s.n]
- Chipana J. (2017). *IBM Parallel Framework. IBM InfoSphere Information Server Developers Workshop*. Lima, Perú: [s.n]
- Churk B., Tuvia A., Naveen D., Stephen J., Mark L., Sachiko T. (2012). *IBM InfoSphere Information Server Deployment Architectures*. En InfoSphere Information Server Overview (cap. 1). Recuperado el 1 agosto 2019, de <http://www.redbooks.ibm.com/redbooks/pdfs/sg248028.pdf>
- Lerm J., Christesen P. (2010). *InfoSphere DataStage Parallel Framework Standard Practice*. En Data Integration with InfoSphere Information Server DataStage (cap. 1). Recuperado el 1 agosto 2019, de <http://www.redbooks.ibm.com/redbooks/pdfs/sg247830.pdf>
- Kimball R., Ross M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition*. En Kimball DW/BI Lifecycle Overview (cap. 17). Recuperado el 1 agosto 2019, de <http://www.essai.rnu.tn/Ebook/Informatique/The%20Data%20Warehouse%20Toolkit,%203rd%20Edition.pdf>
- IBM. (2007). *Conozca IBM, Productos y Soluciones*. En IBM Be blue, think green. Recuperado el 1 agosto 2019, de https://www.ibm.com/expressadvantage/mx/pdf/Folleto_Conozca_IBM.pdf

ANEXOS

Anexo 1: Project Charter

IBM Gerencia de Proyectos					
CONTROL DE VERSIONES					
Versión	Fecha por	Revisada por	Aprobada por	Fecha	Motivo
2.0	BU		NV	06.12.2019	Actualización
PROJECT CHARTER					
Nombre del Proyecto			Símbolos del Proyecto		
IMPLEMENTACIÓN DE LA PLATAFORMA IBM INFOSPHERE INFORMATION			IIS		
Descripción del Proyecto: (qué, quién, cómo, cuándo y dónde)					
<p>El proyecto consistió en mejorar la integración de datos a través de la implementación de la plataforma de IBM InfoSphere Information Server.</p> <p>1. Diseño de Arquitectura (Arquitectura y pre requisitos de hardware, y pre requisitos de HW y SW)</p> <p>2. Instalación y configuración de la plataforma IBM InfoSphere Information Server</p> <p>3. Plantillas y lineamiento de procesos de integración de datos</p> <p>Tiempo ejecución: 6 semanas entre marzo y abril de 2017, en las oficinas de Seguros Generales del Perú</p> <p>Modalidad : Tiempo y materiales</p> <p>Definición del Producto del Proyecto: Descripción del producto, servicio o capacidad a generar.</p> <p>Diseño de Arquitectura:</p> <ul style="list-style-type: none"> - Arquitectura y pre requisitos de hardware y software - Revisión de servidores y pre requisitos de hardware y software <p>Instalación de plataforma IBM InfoSphere Information Server</p> <ul style="list-style-type: none"> - Instalación IBM InfoSphere Information Server - Configuración de seguridad - Configuración de conexión a bases de datos - Configuración de proyectos de integración de datos - Pruebas de configuración de plataforma <p>Documentación de instalación y configuración</p> <p>Plantillas y lineamiento de procesos de integración de datos</p> <ul style="list-style-type: none"> - Documento de plantilla de procesos de integración de datos - Documento de lineamiento de procesos de integración de datos <p>Transferencia de conocimiento</p> <ul style="list-style-type: none"> - Workshop Administración de plataforma IBM InfoSphere Information Server - Workshop Desarrollo de procesos de integración de datos <p>Definición de Requisitos del Proyecto: Descripción de requerimientos funcionales, no funcionales, de calidad, etc., del proyecto/producto</p> <p>El cliente tiene los siguientes requisitos:</p> <ul style="list-style-type: none"> - El responsable del cliente habilite el hardware y software base para instalación del servidor IBM InfoSphere Information Server 11.X conforme a los requerimientos de sistema - No se considera desarrollar ni modificar ni configurar alguna herramienta de scheduler u otra ajena a Datastage utilizada por el Cliente - No se contempla configuración de alta disponibilidad - Este servicio contempla solamente instalación del servidor de InfoSphere Information Server 11.X. No se contempla esfuerzo para la instalación de otros servicios adicionales. - La instalación de IBM InfoSphere Information Server 11.X no incluye la instalación y/o configuración de ningún Parche al Sistema Operativo. - Las bases de datos de IBM IIS (smeta, dsodb, iadbi), usuarios y permisos se crean durante la instalación de servidor InfoSphere Information Server 11.X - Configuración se entienda como la adaptación de parámetros conforme a las recomendaciones de los manuales de instalación del software. - Esta configuración no tiene ninguna relación con ningún tipo de medidas de performance o tests en ningún ambiente y de ningún software o proceso. - Es necesario que el cliente provea una estación de trabajo cliente, para labores de Desarrollo/Administración, la cual deberá ser Windows 7 SP 2 - CPU Intel > 2.0 Ghz - 2 GB en RAM o Superior. - 40GB en disco. - Este servicio no incluye revisión de la arquitectura o análisis de capacidad de la infraestructura para la solución. En caso de requerirse, IBM ofrece un servicio de Capacity Planning, el cual no hace parte de la presente propuesta - Se considera la configuración de los drivers ODBC a conexión a base de datos, previamente el cliente debe haber instalado el cliente de la base de datos ajena a IBM que desea conectar en la capa de engine. - La configuración de la seguridad no considera una configuración de Single Sign On - El cliente será responsable de la instalación de los parches de sistema operativo requeridos para la instalación de los aplicativos. parámetros de núcleo (Kernel) del sistema operativo. - El cliente será responsable de la creación de filesystem para la creación de los espacios según estándar de cliente en cada ambiente - Es responsabilidad del cliente asegurar la conectividad nativa a los servidores de base datos - Documentación de la instalación de IBM InfoSphere Information Server es en formato IBM 					
Objetivos del Proyecto: (qué, hacia los cuales se debe dirigir, el trabajo del proyecto en términos de la triple restricción)					
Concepto	Objetivos	Criterio de Éxito			
1. Alcance	Diseño de Arquitectura Instalación de plataforma IBM InfoSphere Information Server Plantillas y lineamiento de procesos de integración de datos Transferencia de conocimiento	Seguimiento continuo de los avances de las tareas y los entregables por parte del cliente			
2. Tiempo	Concluir el proyecto en el plazo establecido	Concluir el proyecto en 6 semanas entre marzo y abril de 2017			
3. Costo	Cumplir con el costo estimado del proyecto de \$ 114,021.47 Dolares Americanos	No exceder del presupuesto del proyecto			
Finalidad del Proyecto: (en último, propósito general, u objetivo de nivel superior por el cual se ejecuta el proyecto, enlace con programas, portafolios, o estrategias de la organización)					
Generar ingresos a los clientes					
Justificación del Proyecto: (Motivos, razones, o argumentos que justifican la ejecución del proyecto)					
Justificación Cualitativa		Justificación Cuantitativa			
Unificar los componentes de la organización.		Integración de Datos			
Escalabilidad para cumplir los requerimientos de procesamiento de volúmenes de datos		Integración de Datos			
Diseño de modelos de implementación de integración de datos que permitan establecer un marco de trabajo, impulsando la colaboración entre los usuarios del negocio y de TI para entender el significado, estructura y contenido de la información entre sus fuentes de datos.		Integración de Datos			
Reducir costos de HW y SW		VAN			
Designación del Project Manager del Proyecto.					
Nombre	Designación	Niveles de autoridad			
Reporta a	NV				
Supervisa a	Contratista a definir	Exigir el cumplimiento de los entregables			
Cronograma de Hitos del Proyecto.					
Hito o Evento Significativo		Fecha Programada			
Diseño de Arquitectura		10-Marzo-2017			
Instalación de plataforma IBM InfoSphere Information Server		31-Marzo-2017			
Plantillas y lineamiento de procesos de integración de datos		07-Abril-2017			
Transferencia de conocimiento		14-Abril-2017			
Organizaciones o Grupos Organizacionales que Intervienen en el Proyecto.					
Organización o Grupo Organizacional		Rol que desempeña			
Seguros Generales del Perú		Cliente			
IBM Recurso 1		Gerente Proyecto IBM			
IBM Recurso 2		Consultor IBM			
Principales Amenazas del Proyecto (Riesgos Negativos)					
Habilitación del hardware y software base para instalación del servidor IBM InfoSphere Information Server 11.X conforme a los requerimientos del sizing y requerimientos de sistema					
Desarrollar y/o modificar configuración alguna de herramienta de scheduler u otra ajena a Datastage utilizada por el Cliente					
Configuración de alta disponibilidad					
El cliente debe instalar los parches de sistema operativo requeridos para la instalación de los aplicativos					
El cliente debe configurar filesystem para la creación de los espacios según estándar de cliente en cada ambiente					
Documentación de la instalación de IBM InfoSphere Information Server es en formato IBM					
Principales Oportunidades del Proyecto (Riesgos Positivos)					
Eliminar duplicidad de funciones, uso eficiente de recursos de hardware, reducir el esfuerzo en tareas administrativas y despliegue					
Procesos de integración de datos de la empresa Seguros Generales					
Diseñar modelos de integración de datos basados en casos de uso de la empresa Seguros Generales					
Diseñar e implementar la arquitectura de integración de datos en la empresa Seguros Generales					
Presupuesto Preliminar del Proyecto. (Monto \$ / (DOLARES)					
Concepto		Monto \$ / (DOLARES)			
Equipo de Trabajo		\$ 2,840.00			
Inversión Tecnológica		\$ 5,280.84			
Reserva de Contingencia (5%)		\$ 1,171.04			
Total \$		\$ 9,291.88			
Reserva de Gestión (5%)		\$ 4,645.94			
Presupuesto del proyecto		\$ 114,021.47			
sponsor que autoriza el Proyecto.					
Nombre	Empresa	Cargo	Fecha		
Arquitectura - TI	Seguros Generales del Perú	Arquitectura - TI	01 Marzo 2017		