

**UNIVERSIDAD NACIONAL
JOSÉ FAUSTINO SÁNCHEZ CARRIÓN**



**FACULTAD DE INGENIERÍA INDUSTRIAL, SISTEMAS E
INFORMÁTICA**

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

**INTELIGENCIA DE NEGOCIOS Y EL MONITOREO DE
SERVIDORES EN EL CENTRO DE DATOS DE UNA
EMPRESA DE CLOUD, LIMA - 2019**

**TESIS PARA OPTAR EL TÍTULO PROFESIONAL EN
INGENIERÍA DE SISTEMAS**

AUTOR

MIGUEL ANTONIO CARREÑO VALERIO

ASESOR

**ING. JAVIER ALBERTO MANRIQUE QUIÑONEZ
C. I. P. 48354**

HUACHO-PERÚ

2019

Dedicatoria

A mis queridos padres y hermanos que
estuvieron siempre apoyándome para lograr
mis metas profesionales y a mi esposa e hijo
por motivarme a conseguir el éxito
profesional.

Agradecimiento

A mi asesor Ing. Javier Alberto Manrique
Quiñonez, por todo su apoyo y orientación
brindado en el desarrollo de la presente
investigación.

Índice General

Introducción	1
CAPÍTULO 1: PLANTEAMIENTO METODOLÓGICO	2
1.1. Descripción de la realidad problemática	2
1.2. Formulación del problema	9
1.2.1. Problema general	9
1.2.2. Problemas específicos	9
1.3. Objetivos	9
1.3.1. Objetivo general	9
1.3.2. Objetivos específicos	10
1.4. Justificación de la investigación	10
1.4.1. Económico	10
1.4.2. Tecnológico	10
1.4.3. Operativo	10
1.4.4. Académico	10
1.4.5. Personal	11
1.5. Delimitación de la investigación	11
1.5.1. Delimitación espacial	11
1.5.2. Delimitación temporal	11
1.6. Viabilidad de la investigación	11
1.6.1. Viabilidad técnica	11
1.6.2. Viabilidad operativa	11
1.6.3. Viabilidad económica	11
1.7. Alcance	11
CAPÍTULO 2: MARCO TEÓRICO	12
2.1. Antecedentes de la investigación	12
2.1.1. Investigaciones internacionales	12
2.1.2. Investigaciones nacionales	14
2.2. Bases teóricas	18
2.2.1. Monitoreo	18
2.2.2. Monitoreo de servidores	18
2.2.3. Métricas de servidores	19
2.2.4. Manejo de datos de monitoreo	20

2.2.5. Sistemas de información	21
2.2.6. Inteligencia de Negocios	21
2.2.7. Componentes de Inteligencia de Negocios	22
2.2.8. Tecnología OLAP	25
2.2.9. Diferencias entre las tecnologías OLTP y OLAP	26
2.2.10. Arquitectura de Inteligencia de Negocios	27
2.2.11. Metodología de Ralph Kimball	28
2.2.12. SQL Server Analysis Services (SSAS)	31
2.2.13. SQL Server Integration Services (SSIS)	31
2.3. Procedimiento para realizar prueba de hipótesis de diferencia de medias	31
2.4. Definiciones conceptuales	32
2.5. Formulación de la hipótesis	34
2.5.1. Hipótesis general	34
2.5.2. Hipótesis específicas	34
2.6. Operacionalización de variables e indicadores	35
CAPÍTULO 3: METODOLOGÍA	36
3.1. Diseño metodológico	36
3.1.1. Tipo de investigación	36
3.1.2. Nivel de investigación	36
3.1.3. Diseño de investigación	36
3.1.4. Enfoque	36
3.2. Población y muestra	37
3.2.1. Población	37
3.2.2. Muestra	37
3.3. Técnicas e instrumentos de recolección de datos	37
3.3.1. Técnica empleada	37
3.3.2. Descripción de los instrumentos	38
3.4. Técnicas para el procesamiento de la información	38
3.5. Matriz de consistencia	38
CAPÍTULO 4: RESULTADOS	39
4.1. Descripción de los datos obtenidos	39
4.2. Análisis de resultados	39
4.3. Prueba de hipótesis	43

4.3.1. Comprobación de la hipótesis específica 1	43
4.3.2. Comprobación de la hipótesis específica 2	46
4.3.3. Comprobación de la hipótesis específica 3	49
4.3.4. Comprobación de la hipótesis general	52
CAPÍTULO 5: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES	54
5.1. Discusión	54
5.2. Conclusiones	57
5.3. Recomendaciones	58
REFERENCIAS BIBLIOGRÁFICAS	59
ANEXOS	62

Índice de tablas

Tabla 1. Diferencias entre las tecnologías OLTP y OLAP	27
Tabla 2. Operacionalización de variables e indicadores	35
Tabla 3. Cálculo de la mejora	39
Tabla 4. Datos depurados de los tiempos que toma la generación de reportes	40
Tabla 5. Datos depurados de los tiempos que toma la ejecución de reportes	41
Tabla 6. Datos depurados de los tiempos que toma el análisis de reportes	42
Tabla 7. Comprobación de prueba de hipótesis general	53
Tabla 8. Cronograma de tareas para el desarrollo del data mart	65
Tabla 9. Matriz de dimensiones vs hechos	67
Tabla 10. Lista de dimensiones	69

Índice de figuras

Figura 1. Arquitectura del servicio de cloud de la empresa	3
Figura 2. Proceso de ventas del servicio de cloud	5
Figura 3. Subproceso de monitoreo de servidores	6
Figura 4. Subproceso de solicitud de reportes	7
Figura 5. Componentes de Inteligencia de Negocios	22
Figura 6. Análisis multidimensional	26
Figura 7: Arquitectura de Inteligencia de Negocios	27
Figura 8: Ciclo de vida del data warehouse (Enfoque Kimball)	28
Figura 9: Resultados de la prueba de normalidad de los datos de la actividad de generación de reportes	44
Figura 10: Resultados de la prueba T Student entre el antes y el después de la implementación de BI para la actividad de generación de reportes	45
Figura 11: Resultados de la prueba de normalidad de los datos de la actividad de ejecución de reportes	47
Figura 12: Resultados de la prueba T Student entre el antes y el después de la implementación de BI para la actividad de ejecución de reportes	48
Figura 13: Resultados de la prueba de normalidad de los datos de la actividad de análisis de reportes	50
Figura 14: Resultados de la prueba T Student entre el antes y el después de la implementación de BI para la actividad de análisis de reportes	51
Figura 15: Gráfico del análisis dimensional	68
Figura 16: Diseño físico de la base de datos del data mart	75
Figura 17: Arquitectura de la solución de Inteligencia de Negocios	78
Figura 18: Arquitectura del proceso de ETL	79
Figura 19: Diagrama de la base de datos stage	80
Figura 20: Paquetes del proceso de ETL	81
Figura 21: Paquete ETL DimDate	83
Figura 22: Paquete ETL DimTime	84
Figura 23: Paquete ETL DimCustomer	86
Figura 24: Paquete ETL DimMetric	88
Figura 25: Paquete ETL DimMonitor	89
Figura 26: Paquete ETL DimServer	92

Figura 27: Paquete ETL FactMonitoring	95
Figura 28: Paquete ETL procesar cubo	96
Figura 29: Paquete ETL registrar proceso	96
Figura 30: Proyecto del cubo OLAP	97
Figura 31: Fuente de datos	98
Figura 32: Vista de la fuente de datos	98
Figura 33: Cubo OLAP	99
Figura 34: Estructura Dim Customer	100
Figura 35: Estructura Dim Date	100
Figura 36: Estructura Dim Metric	101
Figura 37: Estructura Dim Monitor	102
Figura 38: Estructura Dim Server	102
Figura 39: Estructura Dim Time	103
Figura 40: Reporte de estado de servidores	104
Figura 41: Reporte de estado de servidores detallado últimas 5 horas	105
Figura 42: Reporte de estado de servidores detallado últimos 3 días	106
Figura 43: Reporte Top N del porcentaje de uso de disco	107
Figura 44: Reporte de umbral excedido del porcentaje de uso de memoria	107

Anexos

Anexo 1. Matriz de consistencia	63
Anexo 2. Planificación del proyecto	64
Anexo 3. Definición de requerimientos	66
Anexo 4. Modelo dimensional	69
Anexo 5. Diseño físico y arquitectura	75
Anexo 6. Diseño del sistema de extracción, transformación y carga	79
Anexo 7. Implementación	97
Anexo 8. Modelo de cuestionario	108

Inteligencia de Negocios y el Monitoreo de Servidores en el Centro de Datos de una Empresa de Cloud, Lima – 2019

Business Intelligence and Sever Monitoring in the Data Center of a Cloud Company, Lima - 2019

Miguel Antonio Carreño Valerio ¹.

Resumen

Objetivo: Evaluar si la implementación de una solución de Inteligencia de Negocios mejora el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud en Lima, año 2019. **Metodología:** Investigación aplicada, de nivel pre-experimental y de enfoque cuantitativo. Debido a que la población es pequeña, se consideró como la muestra toda la población, 10 reportes por mes, para medir el tiempo de las actividades de generación de reportes, de ejecución y de análisis de reportes que permitieron obtener datos pre y post implementación de la solución de Inteligencia de Negocios. El diseño metodológico fue O1 X O2, para cada grupo de observación. Los datos extraídos siguen una distribución normal por lo que fueron procesados a través de la prueba paramétrica T Student para comparación de medias de tiempo de generación de reportes, ejecución de reportes y análisis de reportes. Se usó el software Minitab 18.1 para la obtención de los p-valor con las que se tomó la decisión de aprobar o rechazar la hipótesis nula. **Resultados:** Se demostró que la implementación de una solución de Inteligencia de Negocios permite mejorar el subproceso de monitoreo de servidores, reduciendo considerablemente el tiempo de las actividades de generación, ejecución y análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud en Lima, año 2019 logrando un 79.27%, 92.02% y 18.81% de mejora respectivamente. **Conclusiones:** Se corroboró que la implementación de una solución de Inteligencia de Negocios mejora el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud en Lima, año 2019.

Palabras clave: Inteligencia de Negocios, Data mart, Monitoreo, Tiempo, Ejecución, Análisis, Prueba T Student.

Abstract

Objective: Evaluate if the implementation of a Business Intelligence solution improves the server monitoring sub-process in the data center of a cloud company in Lima, 2019. **Methodology:** Applied research, pre-experimental level and quantitative approach. I have taken the whole population as a sample, due to this is small. There is in total 10 reports per month, so they were used to measuring the time activities of reports, reports execution and reports analysis; by using these reports I was able to obtain data before and after implementation of the Business Intelligence solution. The methodological design was O1 X O2, for each observation group. The data obtained follows a normal distribution, so they were processed through the Student's T parametric test for comparison of time averages of reports generation, reports execution and reports analysis. Minitab 18.1 software was used to obtain the values what allowed to make the decision for approving or rejecting the null hypothesis. **Results:** It was demonstrated that the implementation of a Business Intelligence solution allows the improvement of the server monitoring sub-process, considerably reducing the time of activities of generation, execution and reports analysis of the server monitoring sub-process in the data center of a cloud company in Lima, 2019, achieving 79.27%, 92.02% and 18.81% improvement respectively. **Conclusions:** It was corroborated that the implementation of a Business Intelligence solution improves the server monitoring sub-process in the data center of a cloud company in Lima, 2019.

Key words: Business Intelligence, Data mart, Monitoring, Time, Execution, Analysis, Student's t-test.

¹ Escuela Profesional de Ingeniería de Sistemas. Facultad de Ingeniería Industrial, Sistemas e Informática. Universidad Nacional José Faustino Sánchez Carrión. Huacho, Perú.

Introducción

Las organizaciones siempre están buscando ser más competitivas para lograr el éxito empresarial, por ello es fundamental obtener ventaja convirtiendo los datos en información para apoyar la toma de decisiones y la definición de estrategias de negocio acertadas.

La evolución tecnológica ha permitido el desarrollo de sistemas complejos que apoyan las actividades diarias de las organizaciones. Esto ha generado una sobreabundancia de datos y una necesidad de información que no son procesados debido a que no fueron creados con ese objetivo.

A partir de estas necesidades, nace el concepto de Inteligencia de Negocios, que básicamente es un conjunto de estrategias y herramientas orientadas a mejorar la administración y creación de conocimiento mediante el análisis de los datos presentes en la organización. Permite gestionar una empresa a partir de su propia información generada. Con esto se busca atender las necesidades de información de ejecutivos y analistas para ampliar el entendimiento del negocio y así poder definir las estrategias de negocio.

El presente trabajo de investigación tiene por objetivo implementar una solución de Inteligencia de Negocios, desarrollando un data mart para el área de monitoreo de servidores de la línea de negocio de cloud de la empresa IBM, orientada a proporcionar una potente herramienta para que los usuarios puedan acceder a la información de manera oportuna y precisa, y así poder tomar las más acertadas decisiones para mantener el normal funcionamiento de los servidores.

CAPÍTULO 1: PLANTEAMIENTO METODOLÓGICO

1.1. Descripción de la realidad problemática

En la actualidad, existen muchos centros de datos (data centers) al rededor del mundo, en donde se concentra el equipamiento informático y electrónico para el procesamiento de los datos de una organización, siendo los servidores uno de los recursos más importantes.

La empresa IBM (International Business Machines) es una importante multinacional estadounidense de tecnología y consultoría. IBM fabrica y comercializa tanto hardware como software, ofreciendo servicios de infraestructura, alojamiento en cloud y consultoría en una amplia gama de áreas relacionadas con la informática.

Uno de los principales servicios que brinda IBM es el alojamiento o hosting en cloud. Este servicio brinda acceso a diferentes aplicaciones y sitios web usando recursos de cloud. A diferencia del alojamiento tradicional, las soluciones no se despliegan en un único servidor, sino que está compuesto por una amplia red de servidores físicos y virtuales conectados que albergan aplicaciones o sitios web, garantizando una mayor escalabilidad y flexibilidad.

IBM constantemente está buscando expandir su servicio de cloud, es por ello por lo que recientemente ha adquirido parte de otra empresa que también brinda servicios de cloud. Con esta adquisición ha heredado la administración de un centro de datos, el cual posee una gran cantidad de servidores que son usados por muchos clientes, siendo IBM responsable de asegurar el óptimo funcionamiento de estos recursos. En la figura 1 se visualiza la arquitectura del servicio de cloud adquirido.

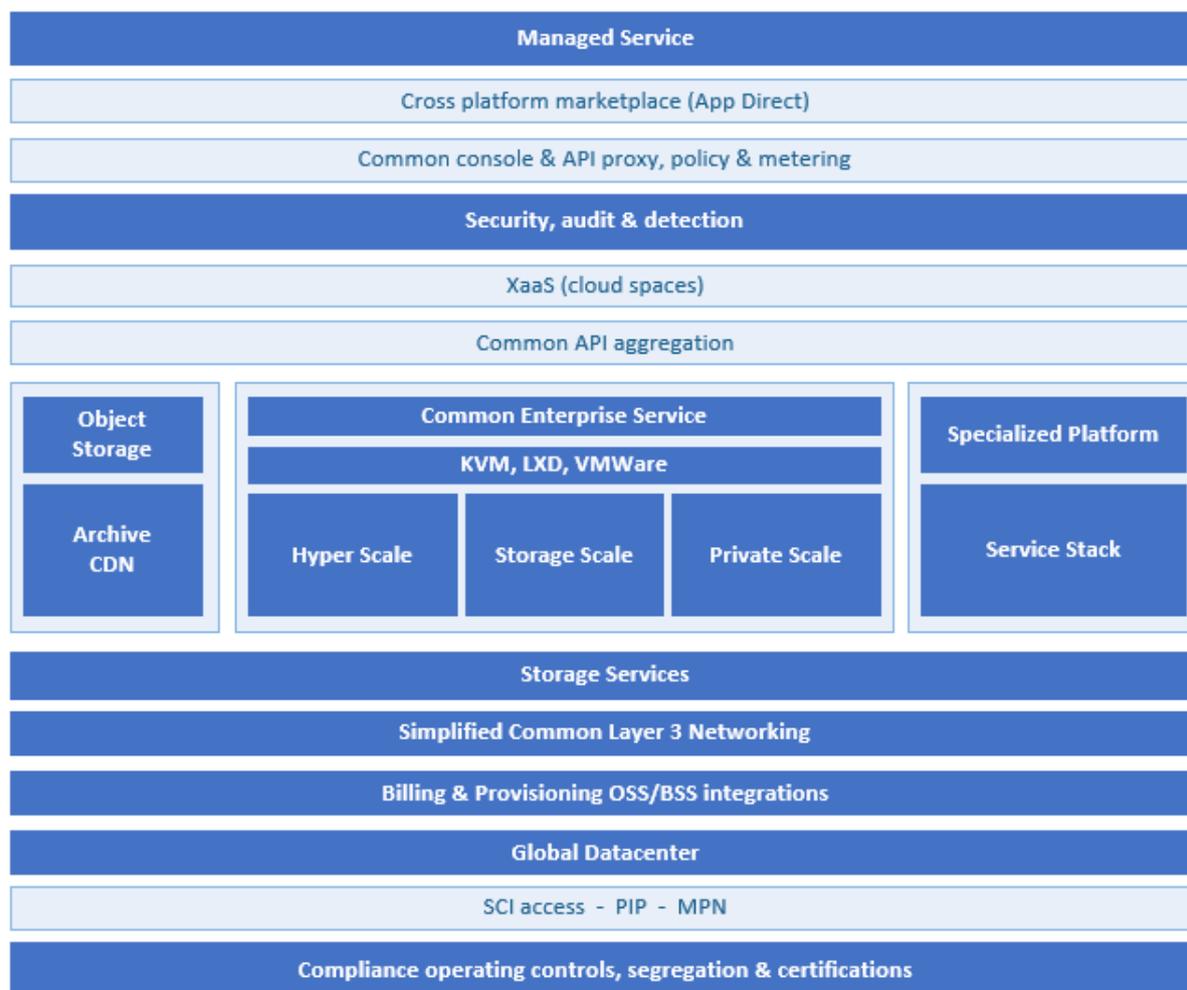


Figura 1. Arquitectura del servicio de cloud de la empresa

El proceso de monitoreo de servidores del centro de datos está a cargo del área de monitoreo de servidores, dicha área tiene la función de prevenir situaciones que afecten el correcto funcionamiento de los mismos y con ello garantizar que todos los recursos de los servidores estén siendo usados dentro de los parámetros considerados normales. Cuando se presenta alguna anomalía, se toman las acciones correctivas e inclusive se notifica al cliente si este se está excediendo en el uso de los recursos contratados.

Para realizar un monitoreo eficiente, los usuarios necesitan información precisa y oportuna sobre el uso de recursos de los servidores contratados por los clientes, este trabajo es muy importante porque evitará causar inconvenientes en la disponibilidad de los servidores que están siendo usados. Si el servicio presentara fallas, el cliente estaría insatisfecho y

optaría por cancelar o no renovar el contrato, siendo algo muy negativo para la reputación y rentabilidad de la empresa.

El centro de datos tiene configurado siete sistemas de monitoreo y cada uno de ellos almacena una enorme cantidad de datos con una frecuencia de cada quince minutos. Los datos capturados por estos sistemas son consolidados en un solo repositorio central de datos. Todos los reportes de monitoreo se acceden al mencionado repositorio para mostrar la información que será usada por los analistas de monitoreo. Estos reportes son compartidos mediante un portal web colaborativo.

La empresa tiene definido una serie de actividades para el proceso de ventas del servicio de cloud, el cual inicia cuando el cliente solicita una cotización del servicio. El asesor de ventas revisa la solicitud, luego elabora la cotización de acuerdo a las necesidades del cliente y se lo envía. El cliente evalúa la cotización, si hay alguna observación, lo notifica al asesor de ventas para que hagan las modificaciones correspondientes. Si todo está en orden envía una confirmación al asesor de ventas para que elabore el contrato. El asesor de ventas envía el contrato al cliente. El cliente verifica que el contrato esté correcto. Si está correcto procede a la firma del contrato; de lo contrario se notifica al asesor de ventas para que haga las correcciones necesarias. Después de la firma del contrato, el asesor de ventas procede a generar la orden de implementación. Luego el analista de infraestructura implementa el servicio. Finalmente, el analista de monitoreo inicia las actividades propias de subproceso de monitoreo de servidores. En la figura 2 se visualiza el proceso de ventas del servicio de cloud.

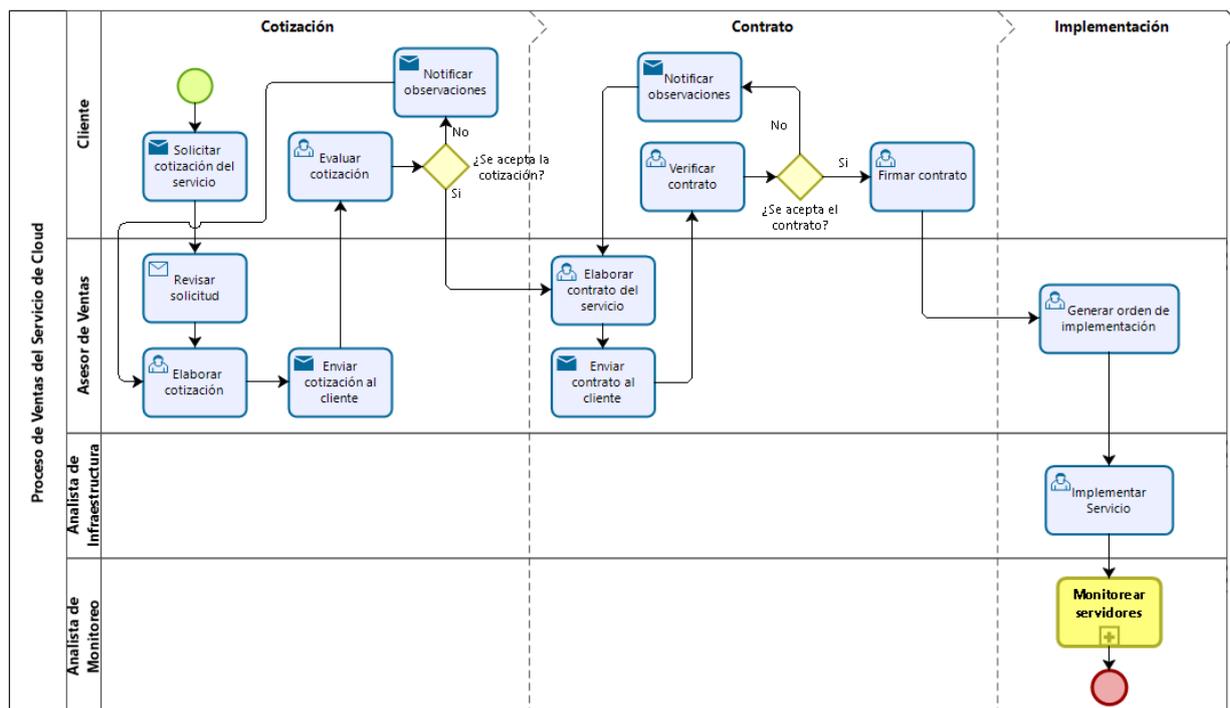


Figura 2. Proceso de ventas del servicio de cloud

La descripción narrativa del subproceso de monitoreo de servidores es el siguiente: el analista de monitoreo monitorea los servidores. Si no ocurre ningún incidente, no hay ninguna acción a tomar. Pero si se presenta algún incidente, el analista de monitoreo necesitará acceder a la información sobre la actividad de los servidores ingresando al portal web de reportes. Si el analista de monitoreo encuentra un reporte que satisfaga sus necesidades de información, este solicita un reporte al área de TI (Tecnología de la Información) de acuerdo a sus necesidades. Pero si existe el reporte, procederá a ejecutarlo, y así obtendrá los resultados del reporte. El analista de monitoreo analiza el reporte para determinar el origen del problema o encontrar patrones para determinar las acciones a tomar. Cuando el incidente no es responsabilidad del cliente el analista de monitoreo determina la solución y la ejecuta. Si el problema persiste, nuevamente se realiza la revisión de la información. Si el incidente es responsabilidad del cliente, el analista de monitoreo identifica al cliente y luego envía un informe del incidente al asesor de servicio.

El asesor de servicio revisa el informe del incidente y luego prepara una propuesta de modificación del servicio contratado y se lo envía al cliente. El cliente evalúa la propuesta, y si no está de acuerdo puede rechazarlo y lo hace saber al asesor de servicio. En el caso de que el cliente acepte la propuesta, pero exista alguna observación, se notifica la observación al asesor de servicio para que este realice las correcciones de caso. Una vez aceptada la propuesta por parte del cliente, este envía la aceptación de la propuesta al asesor de servicio. A su vez, el asesor de servicio envía la aceptación de la propuesta al analista de monitoreo para generar la orden de implementación. Finalmente, el analista de infraestructura implementa la modificación del servicio. En la figura 3 se visualiza el subproceso de monitoreo de servidores.

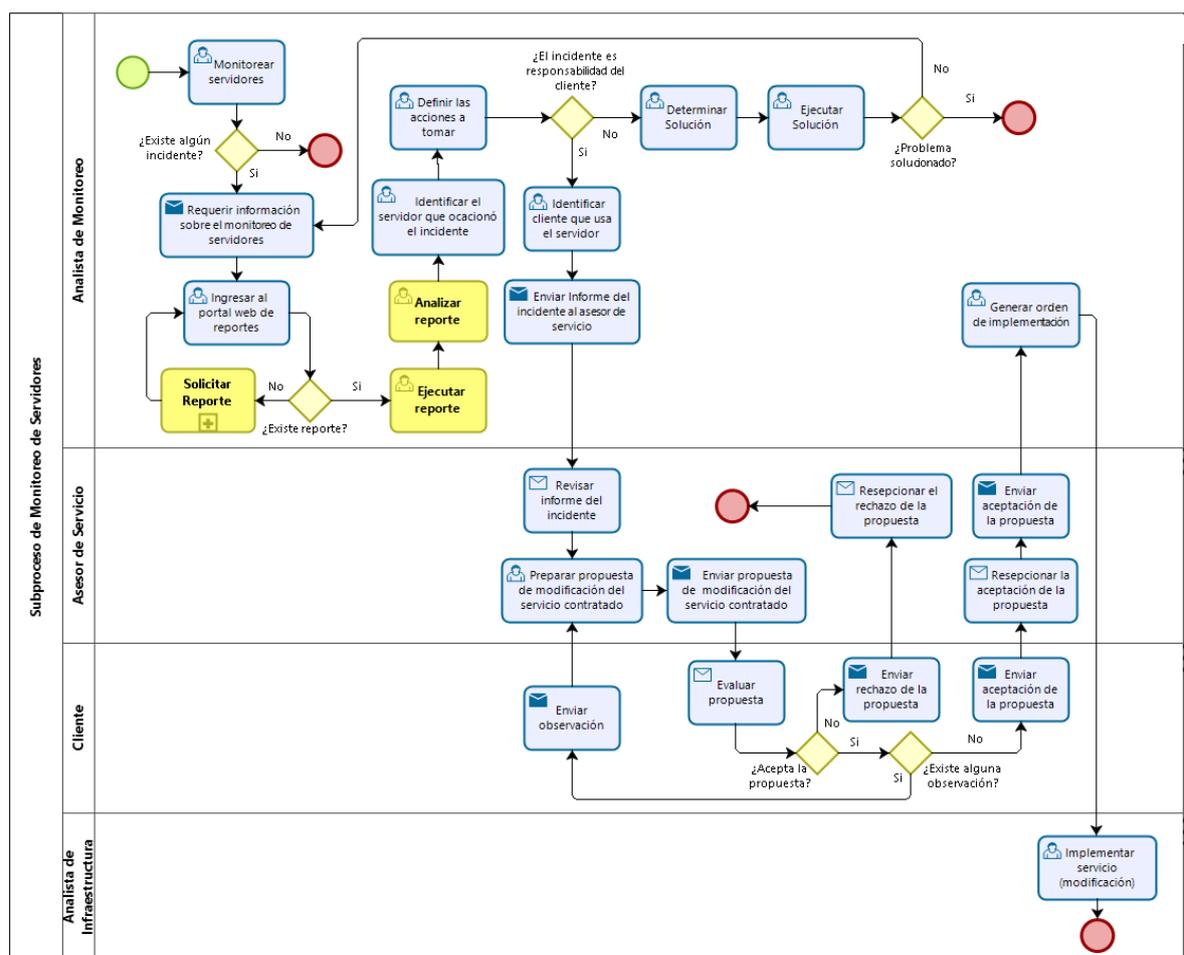


Figura 3. Subproceso de monitoreo de servidores

La necesidad de nuevos reportes es frecuente, es por ello que existe un subproceso para solicitarlos. El subproceso inicia cuando el analista de monitoreo prepara un documento con las especificaciones del reporte y crea un ticket en el sistema de gestión de tickets solicitando la generación de un nuevo reporte en donde se incluyen las especificaciones técnicas detalladas de este. El supervisor de TI revisa el ticket y si se requiere información adicional, le envía una notificación al analista de monitoreo. De lo contrario realiza la priorización del ticket para asignarlo a un analista de datos. El analista de datos revisa las especificaciones del reporte y luego procede a generar el reporte. Si en el transcurso de la generación del reporte se requiere información adicional, lo solicita al analista de monitoreo. Finalizado la generación del reporte, el analista de datos procede a solicitar la validación del reporte al analista de monitoreo y si hubiera alguna observación se hacen las correcciones necesarias. Finalmente, el analista de datos procede con el despliegue del reporte en producción y cierra el ticket relacionado a la solicitud. En la figura 4 se visualiza el subproceso de solicitud de reportes.

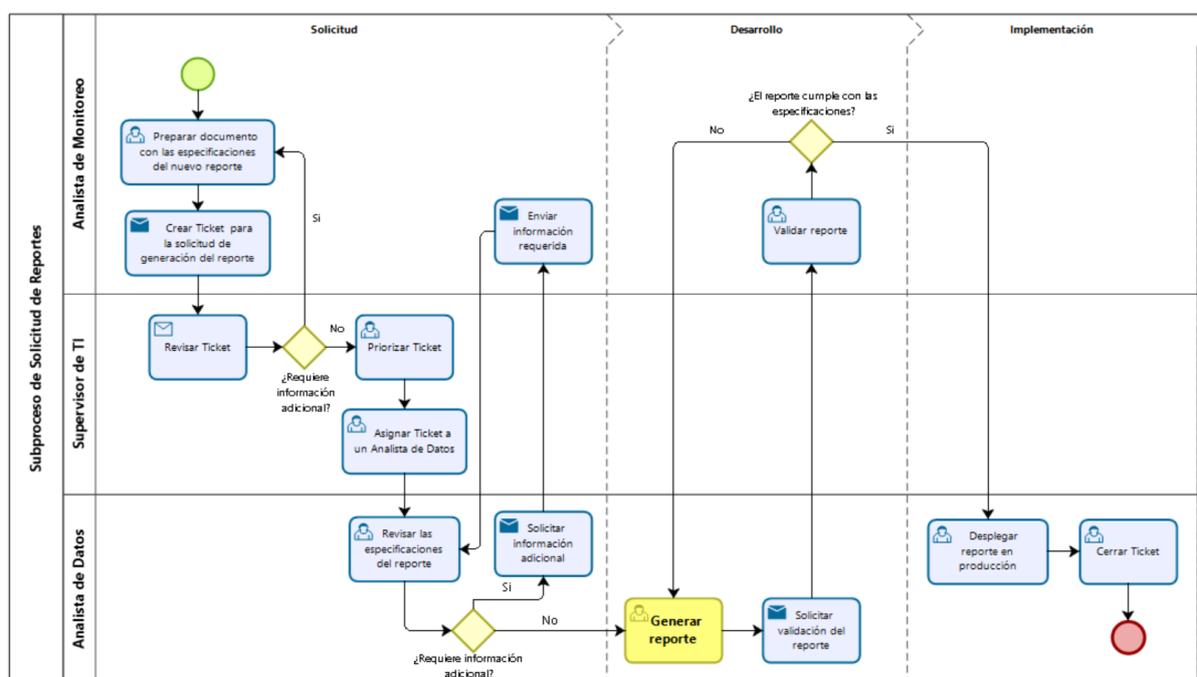


Figura 4. Subproceso de solicitud de reportes

Algunos analistas del equipo encargado del desarrollo de reportes han expresado que a menudo que les resulta muy engorroso realizar su trabajo, debido a la complejidad de las especificaciones técnicas de los reportes solicitados, por ejemplo, mostrar información relacionando múltiples perspectivas de análisis como fecha, tipo de recurso, cliente, servidor, etc. También cabe precisar que la arquitectura del repositorio de datos no es la más adecuada para realizar este tipo de consultas, teniendo como resultado un retraso en la creación de reportes y perjudicando la oportuna toma de decisiones relacionadas al monitoreo de servidores.

Otro problema frecuente es la demora en el tiempo que toma la ejecución de un reporte, debido al gran volumen de datos consultados y la inadecuada arquitectura del repositorio de datos mencionado previamente. Para mitigar esto, hay tareas programadas de eliminación definitiva de datos de una antigüedad mayor de tres meses. Para la mayoría de los informes solo es necesario datos recientes, sin embargo, existe la necesidad de analizar datos antiguos para tener una mayor perspectiva histórica del uso de los recursos por parte de los analistas de monitoreo, cosa que no es posible actualmente.

En muchas ocasiones, los analistas de monitoreo han mencionado que realizar el análisis de los reportes es un poco complicado por muchas razones, entre las que destacan las siguientes: no existe un adecuado nivel de detalle de la información mostrada en el reporte o tienen que recurrir a la unión de más de un reporte en hojas de Excel para realizar cálculos que no están incluidos en el reporte. Se tiene conocimiento que estas deficiencias en los reportes se deben a que resulta complicado cumplir con todos los requerimientos de los analistas de monitoreo con la arquitectura relacional de datos con que se cuenta.

Si no se toman acciones para solucionar los problemas anteriormente expuestos, la empresa podría estar ante la posibilidad de perder valiosos clientes y con ello afectar negativamente su rentabilidad.

Para hacer frente a estos problemas, se propone la implementación de una solución de Inteligencia de Negocios, desarrollando un data mart que permita a los usuarios acceder a la información de manera oportuna y precisa. Se profundizará más en el tema durante el transcurso de la investigación.

1.2. Formulación del problema

1.2.1. Problema general

¿La implementación de una solución de Inteligencia de Negocios mejorará el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019?

1.2.2. Problemas específicos

¿La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019?

¿La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019?

¿La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019?

1.3. Objetivos

1.3.1. Objetivo general

Implementar una solución de Inteligencia de Negocios para mejorar el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.

1.3.2. Objetivos específicos

Implementar una solución de Inteligencia de Negocios para mejorar el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima 2019.

Implementar una solución de Inteligencia de Negocios para mejorar el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima 2019.

Implementar una solución de Inteligencia de Negocios para mejorar el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima 2019.

1.4. Justificación de la investigación

1.4.1. Económico

Reducir gastos ocasionados por la actividad de búsqueda de información para la toma de decisiones relacionada al monitoreo de servidores, debido a que se implementará un data mart para la obtención acertada y oportuna de la información.

1.4.2. Tecnológico

Permite a la empresa IBM trabajar con herramientas de Inteligencia de Negocios para brindar soporte al requerimiento de información.

1.4.3. Operativo

Es de gran beneficio para la alta dirección y el área de monitoreo de servidores, permitiéndoles superar las deficiencias en la obtención de información para la toma de decisiones.

1.4.4. Académico

Aplicar todos los conocimientos adquiridos en el transcurso de mi formación universitaria para implementar una solución frente al problema encontrado.

1.4.5. Personal

Adquirir nuevos conocimientos en el transcurso de mi investigación y con ello tener una sólida base metodológica para poder aplicar mis conocimientos en un futuro.

1.5. Delimitación de la investigación

1.5.1. Delimitación espacial

El presente trabajo de investigación se realizará en la empresa IBM, en la sucursal ubicada en la ciudad de Lima – Perú, así como en las instalaciones de la UNJFSC – Huacho, ya que se cuenta con asesorías de docentes.

1.5.2. Delimitación temporal

El presente trabajo de investigación se realizará entre los meses de diciembre del 2018 y enero del 2019.

1.6. Viabilidad de la investigación

1.6.1. Viabilidad técnica

Se cuenta con la tecnología necesaria para realizar la investigación, tanto como los equipos y la infraestructura adecuada.

1.6.2. Viabilidad operativa

Se cuenta con el acceso a la información necesaria para el desarrollo de la investigación y el apoyo de los usuarios involucrados.

1.6.3. Viabilidad económica

Se cuenta con el apoyo de la empresa para realizar la investigación.

1.7. Alcance

La implementación de la solución de Inteligencia de Negocios tendrá como alcance el subproceso de monitoreo de servidores del centro de datos de IBM, como fue mencionado anteriormente, este centro de datos fue heredado producto de una adquisición de la empresa.

CAPÍTULO 2: MARCO TEÓRICO

2.1. Antecedentes de la investigación

2.1.1. Investigaciones internacionales

(Buitrago M. & González S., 2016) En su trabajo de investigación, titulado: *Implementación de un sistema de inteligencia de negocios para el apoyo en la toma de decisiones en una empresa integradora de soluciones de telecomunicación*, Caracas – Venezuela 2016, Universidad Central de Venezuela.

Tuvo como finalidad describir y mostrar los resultados del proceso de implementación de una solución de Inteligencia de Negocios y el impacto logrado de la obtención de indicadores de gestión en una empresa integradora de soluciones de telecomunicación, que presentaba problemas a la hora de generar reportes.

A través de esta solución, se generan reportes relacionados a los estados financieros, así como también indicadores que ayudan a la gerencia media y alta con información fundamental para la toma de decisiones. La metodología utilizada es la del ciclo de vida Ralph Kimball. La plataforma tecnológica utilizada como respaldo para el desarrollo de la solución, es la aplicación de Oracle Business Inteligencie Standard Edition One. El uso del enfoque metodológico y la aplicación indicada, ofrece como resultado una herramienta que integra la información vital para la toma de decisiones en la organización.

Concluyeron lo siguiente:

Se logra reducir muy considerablemente los tiempos de los reportes y consultas analíticas del negocio, gracias a las múltiples bondades del sistema de Inteligencia de Negocios.

Con base en las distintas pruebas realizadas y en la buena recepción por parte de los usuarios técnicos y funcionales, se puede afirmar que el sistema de Inteligencia de Negocios disminuye notablemente no solo los tiempos de desarrollo de reportes con fines analíticos, sino los reportes y consultas en sí mismas, excediendo las expectativas en algunos casos.

(González Pinos, 2013) En su trabajo de investigación, titulado: *Análisis costo / beneficio de la implementación de Business Intelligence en PYMEs de la ciudad de Quito, del sector de redes y telecomunicaciones*, Quito – Ecuador 2013, Universidad Internacional del Ecuador.

Se analizaron distintas herramientas de Business Intelligence que podrían implementarse en las PYMES del sector de redes y telecomunicaciones de la ciudad de Quito. Sin embargo, a través del caso de estudio de la PYME HARDCOM S.A. y de acuerdo a la investigación realizada en cuanto al manejo de datos e información en ésta empresa y de sus procesos de toma de decisiones, se propuso el uso de la herramienta Power Pivot de Excel 2010, con la cual, mediante una simulación, se obtuvieron reportes que respondieron a las preguntas estratégicas planteadas por el gerente de la compañía, demostrando así la utilidad de su implementación en la administración de la gerencia.

El análisis costo/beneficio de la implementación de esta herramienta se realizó mediante un modelo de “Análisis Comparativo” usando dos escenarios de estudios: en el primero se obtuvo el costo del proceso de toma de decisiones de la gerencia sin el uso de la herramienta Power Pivot; y en el segundo se determinó el costo del proceso de toma de decisiones con el uso de ésta. Con los valores obtenidos se detalló el beneficio final.

Concluyendo lo siguiente:

El beneficio principal que se determinó con el uso de esta herramienta de BI en la administración gerencial es la eficiencia en el tiempo y en la obtención de información y con ello una mejor y rápida toma de decisiones del gerente.

Con el uso de la herramienta, el gerente general estimó una eficiencia de hasta el 60% en el tiempo de manejo de la información, lo que le permite administrar su empresa con eficiencia y eficacia.

2.1.2. Investigaciones nacionales

(Espinoza Vargas & Palomino Ruiz, 2016) En su trabajo la investigación, titulado: *Desarrollo de un Datamart para optimizar la generación de información estratégica de apoyo a la toma de decisiones en la vicepresidencia de banca comercial de Interbank Perú*, Lima – Perú 2016, Universidad de San Martín de Porres.

Tuvo por objetivo disminuir el alto consumo de recursos mediante el desarrollo de un Datamart que optimice la generación de información estratégica para el apoyo a la toma de decisiones.

La metodología usada para el desarrollo de la investigación fue la metodología de Ralph Kimball, la cual está basada en el ciclo de vida dimensional del negocio.

Concluyendo lo siguiente:

El Datamart ha permitido optimizar la generación de información estratégica para el apoyo a la toma de decisiones de tal forma que esto se verá reflejado en incrementos de utilidades.

Mediante el Datamart se logró disminuir el tiempo de generación de información estratégica a una hora y 37 minutos, lo cual equivale a un 90.6%.

Mediante el Datamart se logró reducir en un 100% las labores operacionales de los 4 analistas encargados del proceso de generación de información estratégica, ya que cuenta con una ejecución automática.

(Zamora Saldarriaga, 2017) En su trabajo de investigación, titulado:

Implementación de un Datamart para la mejora en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado

Nacional, Lima – Perú 2017, Universidad Inca Garcilaso de la Vega.

La finalidad esta investigación fue la implementación de un Datamart para la mejora en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional; para ello se empleó la metodología de Ralph Kimball y las herramientas tecnológicas Power BI y Power Pivot para entregar al usuario el acceso a la información del Datamart.

Según los resultados obtenidos mediante la encuesta elaborada en el presente trabajo de investigación se demostró que el Datamart cumple con los indicadores de accesibilidad, fiabilidad, eficiencia y disponibilidad.

El Datamart influyó satisfactoriamente para la mejora en la toma de decisiones del control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional, asimismo la accesibilidad, fiabilidad, usabilidad y eficiencia del Datamart propuesto, influyeron positivamente para el cumplimiento de los procesos.

Concluyendo lo siguiente:

Se determinó que la implementación del datamart influyó positivamente en cuanto a la satisfacción de los usuarios con el funcionamiento de los reportes

en un estado óptimo bajo cualquier circunstancia con data fiable para el análisis de la demanda eléctrica y una posterior toma de decisiones.

Por medio de la implementación del Datamart, la eficiencia en consultar grandes volúmenes de datos y la reducción de horas/hombre en la elaboración de reportes gráficos permitieron mejorar la toma de decisiones para el control de la demanda eléctrica.

(Carhuallanqui Bastidas, 2017) En su trabajo de investigación, titulado: *Diseño de una solución de inteligencia de negocios como herramienta de apoyo a la toma de decisiones en el área de ventas de la empresa farmacéutica Dispefarma*, Lima – Perú 2017, Universidad Nacional Mayor de San Marcos.

Se propone el diseño de una solución de Inteligencia de Negocios en el área de ventas de la empresa Dispefarma como una herramienta de apoyo para la toma de decisiones, donde el principal problema radica en la desintegración de la información y la dificultad de su análisis desde diferentes perspectivas según sea la necesidad de los usuarios.

Por ello se plantea diseñar una solución de inteligencia de negocios que permita integrar la información y poder analizarla de tal forma que brinde reportes de acuerdo a las necesidades de los usuarios.

En el marco metodológico se opta por utilizar la metodología de Ralph Kimball.

Como muestra se tomó a las personas involucradas en el área de ventas de la empresa Dispefarma, a quienes se les aplicó un pre y post test, la recolección de los datos se realizó a través de encuestas, entrevistas, reportes y la observación. El procesamiento de los datos se realizó con la herramienta Microsoft Office Excel.

Se realizaron Dashboards e indicadores de gestión para el proceso de ventas de la organización bajo diversas perspectivas que permite un mejor análisis de la información.

Concluyendo lo siguiente:

Se redujo el tiempo promedio en elaboración de informes solicitados por la gerencia en un 90.13%, es decir de 7.6 a 0.75 horas.

La solución de Inteligencia de Negocios mejoró significativamente la calidad de la información haciendo más confiables los reportes generados a partir de ella brindando un soporte para la toma de decisiones en la empresa.

(Inca Damián & Zavala Sanchez, 2016) En su trabajo de investigación, titulado: *Desarrollo de una solución de inteligencia de negocios para la mejora del proceso de toma de decisiones en el área de administración tributaria de la Municipalidad Distrital de San Bartolo*, Lima – Perú 2016, Universidad Autónoma del Perú.

Se planteó el desarrollo de una solución de Inteligencia de Negocios, que permita reducir los tiempos en el proceso de obtención de los reportes y a su vez disminuir el esfuerzo desplegado en dicho proceso.

Dicha solución permitirá mejorar el proceso de toma de decisiones en el área de Administración Tributaria de la Municipalidad Distrital de San Bartolo, haciendo uso de la metodología de Ralph Kimball.

El área de Administración Tributaria de la Municipalidad Distrital de San Bartolo diariamente maneja grandes cantidades de información, pero debido a que su sistema actual no soporta el adecuado manejo de grandes volúmenes de información, se establece que en dicha área tiene el inconveniente de usar toda esa información que sirva de apoyo a la toma de decisiones de la gerencia.

Concluyendo lo siguiente:

Se observa que se logró reducir el tiempo empleado para generar los reportes en el área de Administración Tributaria en un 96 % aproximadamente.

Se observa que se logró reducir en un 67 % aproximadamente el tiempo que el gerente de Administración Tributaria empleaba en el análisis de la información obtenida en los reportes.

Se observa que se logró incrementar en un 75 % aproximadamente el número de reportes generados al día.

Se observa que se logró incrementar en un 63 % aproximadamente el nivel de satisfacción del gerente de Administración Tributaria respecto al proceso de obtención de los reportes en los reportes.

2.2. Bases teóricas**2.2.1. Monitoreo**

Según Wikipedia, monitoreo significa ser consciente del estado de un sistema, para observar una situación de cambios que se pueda producir con el tiempo, para lo que se precisa un monitor o dispositivo de medición de algún tipo.

(Vera Castro & Arreaga Pérez, 2015) indican que el propósito del monitoreo consiste en detectar de manera oportuna las fortalezas y deficiencias de los procesos, con el objetivo de hacer ajustes para una óptima gestión de las iniciativas.

2.2.2. Monitoreo de servidores

Es la recopilación y el análisis regulares de los datos para garantizar que los servidores estén trabajando de manera óptima y brinden su función esperada. Los datos utilizados para el monitoreo de servidores básicamente son métricas. Por ejemplo, el monitoreo de un servidor examinaría las métricas del sistema operativo del servidor (CPU, memoria, red y rendimiento de disco). Los datos de cada una de estas métricas se analizan

para minimizar, o idealmente prevenir, las interrupciones o consumo excesivo de recursos del servidor (Heroix, s.f.).

Según (Kerkelä, 2018), el monitoreo de servidores es un proceso de recopilación continua de datos sobre los servidores con respecto a la disponibilidad, las operaciones, la seguridad y el rendimiento del servidor. El monitoreo de servidores es ampliamente utilizado en los campos de negocios, que dependen de TI. El monitoreo es esencial, ya que la disponibilidad de los servicios tiene un gran impacto en las organizaciones, lo que ocasiona pérdidas de costos y reputación durante el tiempo de inactividad o la interrupción del servicio.

Según (Kerkelä, 2018), el objetivo principal de monitoreo de servidores, en general, es poder recopilar información sobre el rendimiento del servidor y los eventos de error, diagnosticar el problema y proporcionar información sobre las causas de los problemas a las partes interesadas, como los clientes. Al establecer el monitoreo en el servidor, los problemas se pueden prevenir antes de que se manifiestan realizando acciones preventivas basadas en el seguimiento. La disponibilidad es obviamente una característica importante del servidor, ya que, si no hay disponibilidad, todo proceso relacionado con el servidor se ve afectado directamente. Los problemas de disponibilidad llegan a causar costos para la empresa durante el tiempo que el servidor está inactivo.

2.2.3. Métricas de servidores

Las métricas se pueden dividir en categorías como métricas de trabajo y métricas de recursos (Kerkelä, 2018).

Métricas de trabajo

Estas métricas miden cómo funciona el sistema e incluye indicadores como la cantidad de operaciones y porcentaje de éxito de las operaciones, los resultados de error de las operaciones, el desempeño, lo que califica la eficiencia con la que funcionan. Las métricas de

trabajo miden cómo funciona el servidor, al centrarse en el rendimiento del sistema completo, incluidos todos sus componentes sin centrarse en los datos de los componentes individuales.

Métricas de recursos

Se centran en la supervisión del estado de bajo nivel del servidor. Las métricas de recursos incluyen la velocidad del disco, la carga de la CPU, la carga de memoria, la disponibilidad de recursos, el conteo de hilos y la velocidad de la red. Con estas métricas es posible el diagnóstico en la ocurrencia de problemas. Las métricas de recursos se centran en la medición del rendimiento de componentes individuales del sistema.

2.2.4. Manejo de datos de monitoreo

Cuando se trata del monitoreo de servidores, hay un número finito de métricas de rendimiento y disponibilidad y todos pueden clasificarse como "datos". Teniendo en cuenta las mejores prácticas para el monitoreo de servidores, es mejor dividir el manejo de los datos en cuatro fases (Heroix, s.f.):

Capturar

Hay todo tipo de datos disponibles para el monitoreo de servidores. El desafío no es determinar qué datos recopilar, sino las mejores prácticas sobre cómo capturar óptimamente los datos. Una implementación de monitoreo de servidores no solo debe ser rápida, ligera y confiable, sino que también debe ser resistente a las interrupciones de la red.

Almacenar

El almacenamiento a largo plazo de los datos capturados en una base de datos es esencial para proporcionar un contexto histórico de los problemas relacionados con servidores. La mejor práctica es retener todos los datos relevantes que pueden venir en una variedad de formas.

Evaluar, notificar y corregir

Esencial para cualquier estrategia de monitoreo de servidores es evaluar con precisión los datos capturados, alertar proactivamente al personal sobre los problemas y, cuando sea posible, iniciar acciones correctivas.

Visualizar

Una interfaz web para controlar y visualizar los datos es muy importante. Una buena práctica para cualquier estrategia de monitoreo de servidores es la capacidad de mostrar qué problemas se están desarrollando en tiempo real, así como informar sobre lo que históricamente ha ocurrido.

2.2.5. Sistemas de Información

Según los autores (Andreu, Ricart, & Valor, 1991), un sistema de información es el conjunto formal de procesos que operando sobre una colección de datos estructurada de acuerdo con las necesidades de una empresa, recopila, elabora y distribuye la información necesaria para la operación de dicha empresa y para las actividades de dirección y control correspondientes, apoyando, al menos en parte, los procesos de toma de decisiones necesarios para desempeñar las funciones de negocio de la empresa de acuerdo con su estrategia de negocio.

2.2.6. Inteligencia de Negocios

Es un concepto que proviene del término inglés “Business Intelligence (BI)”. (Ramos, 2011) indica que Inteligencia de Negocios es el conjunto de estrategias y tecnologías que nos van a ayudar a convertir los datos en información de calidad y dicha información en conocimiento que nos permita una toma de decisiones más acertadas y nos ayude así a mejorar nuestra competitividad.

Por otra parte, (Curto Díaz & Conesa Caralt, 2010) mencionan que se entiende por Inteligencia de Negocios al conjunto de metodologías, aplicaciones, prácticas y

capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de la organización.

(Medina La Plata, 2012) señala que la capacidad para tomar decisiones de negocio precisas y de forma rápida se ha convertido en una de las claves para que una empresa llegue al éxito. Sin embargo, los sistemas de información tradicionales suelen presentar una estructura muy inflexible para este fin. Aunque su diseño se adapta en mayor o menor medida para manejar los datos de la empresa, no permite obtener la información de estos. La Inteligencia de Negocios se apoya en un conjunto de herramientas que facilita la extracción, la depuración, el análisis y el almacenamiento de los datos generados en una organización, con la velocidad adecuada para generar conocimiento y apoyar la toma de decisiones de los directivos y los usuarios en general en una organización.

2.2.7. Componentes de Inteligencia de Negocios

En la figura 5, se puede apreciar los distintos componentes que forman parte de Inteligencia de Negocios.

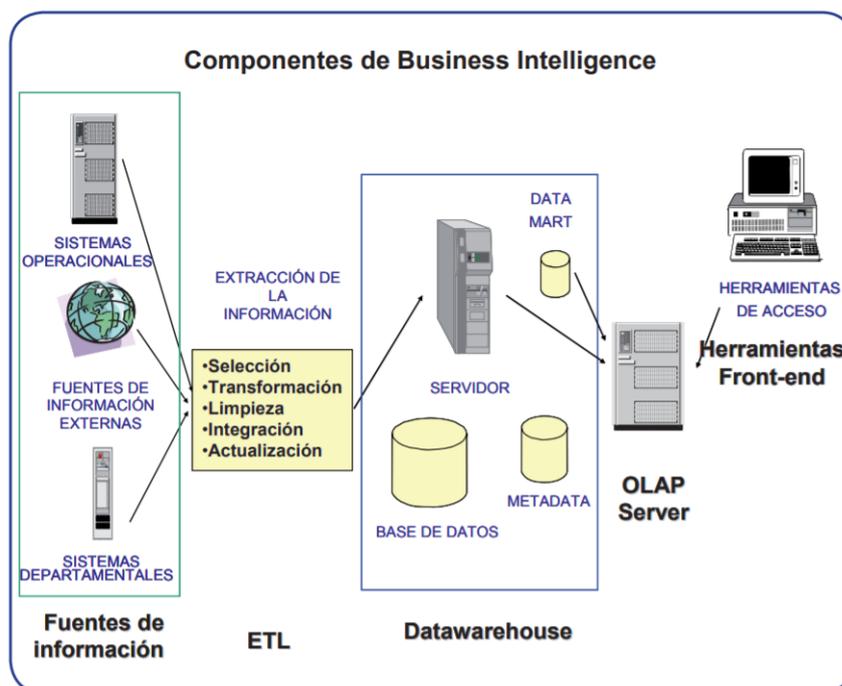


Figura 5. Componentes de Inteligencia de Negocios. (Cano, 2007)

Fuentes de Información

Son los diferentes lugares de donde se extrae la información para alimentar el data warehouse. Básicamente, estas fuentes podrían ser sistemas operacionales o transaccionales, como sistemas desarrollados a medida, archivos de Excel, archivos de texto, etc.

Data warehouse

Según (Curto Díaz & Conesa Caralt, 2010), el data warehouse es un repositorio de datos que proporciona una visión global, común e integrada de los datos de la organización, independientemente de cómo se vayan a utilizar posteriormente por los consumidores o usuarios, con las propiedades siguientes: estable, coherente, fiable y con información histórica. Al abarcar un ámbito global de la organización y con un amplio alcance histórico, el volumen de datos puede ser muy grande (centenas de terabytes). Las bases de datos relacionales son el soporte técnico más comúnmente usado para almacenar las estructuras de estos datos y sus grandes volúmenes.

(Curto Díaz & Conesa Caralt, 2010) indican que el data warehouse está compuesto por los siguientes elementos:

a) Tabla de hechos (Fact Table): Es la representación en el data warehouse de los procesos de negocio de la organización. Por ejemplo, una venta puede identificarse como un proceso de negocio, y se puede considerar la tabla de hecho ventas.

b) Dimensión: Es la representación en el data warehouse de una vista para un cierto proceso de negocio. Si regresamos al ejemplo de una venta, para la misma tenemos el cliente que ha comprado, la fecha en la que se ha realizado, etc.

c) Métricas: Son los indicadores de negocio de un proceso de negocio. Aquellos conceptos cuantificables que permiten medir nuestro proceso de negocio. Por ejemplo, en una venta tenemos el importe de la misma.

(Curto Díaz & Conesa Caralt, 2010) indican que existen principalmente dos tipos de esquemas para estructurar los datos en un almacén de datos:

a) *Esquema en estrella (Star):* A nivel de diseño, consiste en una tabla de hechos en el centro y una o varias tablas de dimensión por cada punto de vista de análisis.

b) *Esquema en copo de nieve (Snowflake):* Es un esquema de representación derivado del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas, y aparecen nuevas uniones.

Data mart

Es un subconjunto de los datos del data warehouse cuyo objetivo es responder a un determinado análisis, función o necesidad, con una población de usuarios específica. Al igual que en un data warehouse, los datos están estructurados en modelos de estrella o copo de nieve, y un data mart puede ser dependiente o independiente de un data warehouse. El data mart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. (Curto Diaz et al. , 2010)

Proceso de extracción, transformación y carga (ETL)

Según (Cano, 2007), ETL es proceso que trata de recuperar los datos de las fuentes de información y alimentar el data warehouse. Este parte del proceso de construcción del data warehouse es costosa y consume una parte significativa de todo el proceso, por ello requiere recursos, estrategia, habilidades especializadas y tecnologías. El proceso ETL se divide en cinco subprocesos:

a) *Extracción:* Este proceso recupera los datos físicamente de las distintas fuentes de información. En este momento disponemos de los datos en bruto.

b) *Limpieza:* Este proceso recupera los datos en bruto y comprueba su calidad, elimina los duplicados y, cuando es posible, corrige los valores erróneos y completa

los valores vacíos.

c) **Transformación:** Este proceso recupera los datos limpios y de alta calidad y los estructura y resume en los distintos modelos de análisis. El resultado de este proceso es la obtención de datos limpios, consistentes, resumidos y útiles.

d) **Integración:** Este proceso valida que los datos que cargamos en el data warehouse son consistentes con las definiciones y formatos del data warehouse; los integra en los distintos modelos de las distintas áreas de negocio que hemos definido en el mismo.

e) **Actualización:** Este proceso es el que nos permite añadir los nuevos datos en el data warehouse.

2.2.8. Tecnología OLAP

Según (Curto Díaz & Conesa Caralt, 2010) se entiende por OLAP, o proceso analítico en línea, al método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o jerarquía de objetos como en un sistema u organización multidimensional, y cuyo objetivo es recuperar y manipular datos y combinaciones de los mismos a través de consultas o incluso informes.

(Curto Díaz & Conesa Caralt, 2010) señala que existen diferentes tipos de OLAP, que principalmente difieren en cómo se guardan los datos:

a) **MOLAP (Multidimensional OLAP):** Es la forma clásica de OLAP y frecuentemente es referida con dicho acrónimo. MOLAP utiliza estructuras de bases de datos generalmente optimizadas para la recuperación de los mismos. Es lo que se conoce como bases de datos multidimensionales (o, más coloquialmente, cubos).

b) **ROLAP (Relational OLAP):** Trabaja directamente con las bases de datos relacionales, que almacenan los datos base y las tablas dimensionales como tablas relacionales mientras se crean nuevas tablas para guardar la información agregada.

c) **HOLAP (Hybrid OLAP):** Es una base de datos en la que los datos se

dividen en almacenaje relacional y multidimensional.

(Curto Díaz & Conesa Caralt, 2010) menciona que OLAP permite el análisis multidimensional. Ello significa que la información está estructurada en ejes (puntos de vista de análisis) y celdas (valores que se están analizando). Como elementos comunes dentro de las diferentes tipologías OLAP tenemos:

a) **Esquema:** Es una colección de cubos, dimensiones, tablas de hecho y roles.

b) **Cubo:** Es una colección de dimensiones asociadas a una tabla de hecho.

Un cubo permite cruzar la información entre tablas de hecho a partir de sus dimensiones comunes. En la figura 6, se puede entender mejor lo mencionado anteriormente.

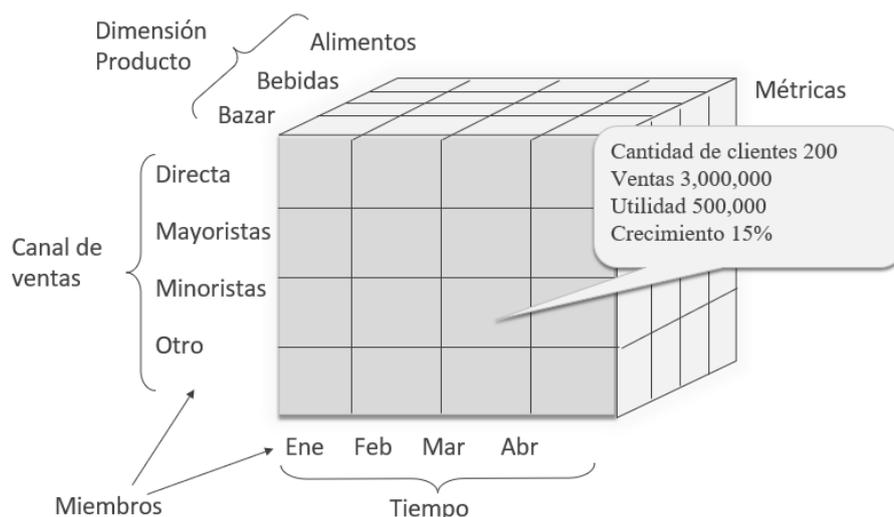


Figura 6. Análisis multidimensional.

2.2.9. Diferencia entre las tecnologías OLTP y OLAP

Mientras que la tecnología OLTP se caracteriza por estar actualizada constantemente por varios usuarios a través de transacciones operacionales sobre datos individuales, la tecnología OLAP es utilizada por personal de niveles ejecutivos que requieren datos con alto grado de agregación y desde distintas perspectivas (dimensiones), como ser: totales de venta por región, por producto, por período de tiempo, etc. (Mendez, Mártire,

Britos, & García-Martínez, 2003). En la Tabla 1, se visualizan las diferencias entre las tecnologías OLTP y OLAP.

Tabla 1.

Diferencias entre las tecnologías OLTP y OLAP (Mendez, et al., 2003).

OLTP	OLAP
Atomizado	Sumarizado
Datos históricos	Datos actuales
Un registro a la vez	Muchos registros a la vez
Orientado a la información operativa	Orientado a la información estratégica
Datos relacionales	Datos multidimensionales
Consultas simples predefinidas	Consultas ad-hoc
Volumen de datos acotados	Grandes volúmenes de datos

2.2.10. Arquitectura de Inteligencia de Negocios

Existen muchas variantes de la arquitectura de Inteligencia de Negocios, pero básicamente se podría representar tal como se puede visualizar en la figura 7.

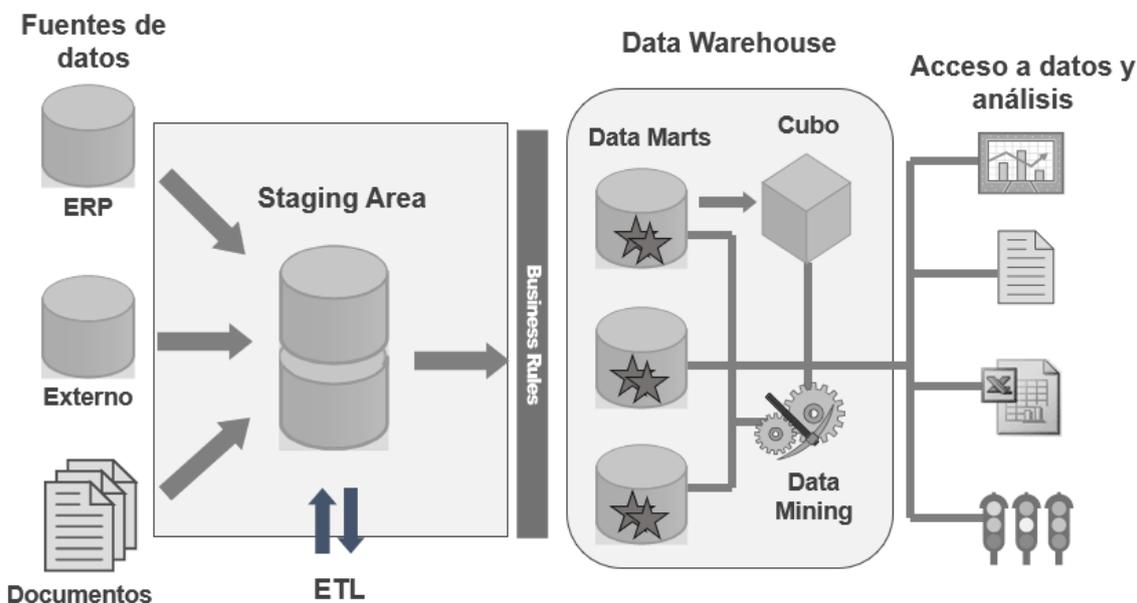


Figura 7. Arquitectura de Inteligencia de Negocios.

2.2.11. Metodología de Ralph Kimball

También conocida como Bottom - Up, especifica que el data warehouse corporativo no es más que la unión de los diferentes data marts. Esta característica permite que sea más flexible y sencillo de implementar, pues podemos construir un data mart como primer elemento del sistema de análisis, y luego ir añadiendo otros que comparten las dimensiones ya definidas o incluyen otras nuevas, en la figura 8 se puede visualizar el ciclo de vida del data warehouse. (Espinosa, 2010).

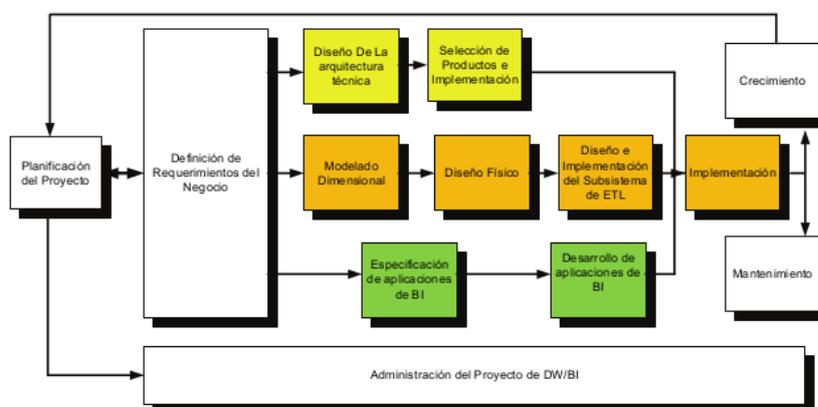


Figura 8. Ciclo de vida del data warehouse, (Kimball & Ross, 2013)

Fase 1: Planificación del proyecto

En esta fase se determina el propósito del proyecto del data warehouse, los objetivos específicos y el alcance de este, los principales riesgos y una aproximación inicial a las necesidades de información. (Rivadera, 2010)

Fase 2: Definición de requerimientos de negocio

La definición de los requerimientos es en gran medida un proceso de entrevistar al personal de negocio y técnico, pero siempre conviene tener un poco de preparación previa. Se debe aprender tanto como se pueda sobre el negocio, los competidores, la industria y los clientes de este. Hay que leer todos los informes posibles de la organización; rastrear los documentos de estrategia interna; entrevistar a los empleados, analizar lo que se

dice en la prensa acerca de la organización, la competencia y la industria. Se deben conocer los términos y la terminología del negocio. (Rivadera, 2010)

Fase 3: Modelo dimensional

La creación del modelo dimensional es un proceso dinámico y altamente iterativo. El diseño comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados de la matriz de requerimientos. Esta fase consta principalmente de los siguientes pasos: elegir el proceso de negocio, establecer el nivel de granularidad, elegir las dimensiones e identificar medidas y las tablas de hechos. (Rivadera, 2010).

Fase 4: Diseño físico

Se enfoca en la selección de la estructura necesaria para dar soporte al diseño lógico. Algunas de las principales actividades de esta fase son la definición de los estándares de nombres y configuraciones específicas del entorno de la base de datos (Rivadera, 2010).

Fase 5: Diseño e implementación del subsistema de ETL

Es la base sobre la cual se alimenta el data warehouse. Si el sistema ETL se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de estos, consolidar la información proveniente de distintos sistemas, y finalmente cargar la información en el data warehouse. (Rivadera, 2010)

Fase 6: Diseño de la arquitectura técnica

Los ambientes del data warehouse requieren la integración de numerosas tecnologías. Se debe tener en cuenta tres factores: los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras planificadas para establecer el diseño de la arquitectura técnica del ambiente del data warehouse. (Rojas Saldívar, 2014).

Fase 7: Selección del producto e implementación

Utilizando como marco el diseño de arquitectura técnica es necesario evaluar y seleccionar componentes específicos de la arquitectura como la plataforma de hardware, el motor de base de datos, la herramienta de ETL o el desarrollo pertinente, herramientas de acceso, etc. Una vez evaluados y seleccionados los componentes determinados se procede con la instalación y prueba de estos en un ambiente integrado del data warehouse. (Rojas Saldívar, 2014)

Fase 8: Especificación y desarrollo de aplicaciones de BI

No todos los usuarios del data warehouse necesitan el mismo nivel de análisis. Es por ello, que en esta etapa se identifican los diferentes roles o perfiles de usuarios para determinar los diferentes tipos de aplicaciones necesarias sobre la base al alcance de los diferentes perfiles (gerencial, analista del negocio, vendedor, etc.).

Una vez que se ha cumplido con todas las actividades de la especificación y se tiene la posibilidad de trabajar con algunos datos de prueba, comienza el desarrollo de la aplicación. El desarrollo de las aplicaciones de los usuarios finales involucra configuraciones de la meta data y construcción de reportes específicos (Rojas Saldívar, 2014).

Fase 9: Implementación

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio. Existen factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos, se encuentran la capacitación, el soporte técnico, la comunicación, las estrategias de retroalimentación. (Rojas Saldívar, 2014)

Fase 10: Administración del proyecto

El gerenciamiento del proyecto asegura que las actividades del ciclo de vida dimensional del negocio se lleven en sincronizadas y en la mejor forma. Entre las actividades

principales se encuentra el monitoreo del estado del proyecto, la comunicación entre los requerimientos del negocio y las restricciones de información para poder manejar correctamente las expectativas en ambos sentidos (Rojas Saldívar, 2014).

2.2.12. SQL Server Analysis Services (SSAS)

SSAS es un motor de datos que ofrece funciones de procesamiento analítico en línea (OLAP) para aplicaciones de Inteligencia de Negocios. SSAS admite OLAP y permite diseñar, crear y administrar estructuras multidimensionales que contienen datos agregados desde otros orígenes de datos, como bases de datos relacionales. (Ramos, 2011)

2.2.13. SQL Server Integration Services (SSIS)

SSIS es una plataforma para la creación de soluciones empresariales de transformación e integración de datos. SSIS sirve para resolver complejos problemas empresariales mediante la copia y descarga de archivos, el envío de mensajes de correo electrónico como respuesta de eventos, la actualización del almacenamiento de datos y la limpieza. (Ramos, 2011)

2.3. Procedimiento utilizado para realizar la prueba de hipótesis

Las actividades del subproceso de monitoreo de servidores fueron: de generación de reportes, de ejecución de reportes y la de análisis de reportes.

Las tareas del procedimiento utilizado para realizar la prueba de hipótesis para cada actividad de investigación fueron:

i) Obtención y limpieza de los datos de prueba de la actividad de investigación (se consideró solo reportes que fueron creados en un mes durante el periodo de investigación).

ii) Realización de la prueba de normalidad de los datos usando la prueba Anderson-Darling. Se plantea la siguiente hipótesis:

H_0 : Los datos tienen una distribución normal

H_1 : Los datos no tienen una distribución normal

Se establece la regla de decisión:

Valor de $p \leq \alpha$ rechazar H_0 a nivel de α

Valor de $p > \alpha$ No rechazar H_0 a nivel de α

Para nuestro estudio el valor de significancia es de $\alpha = 0.05$.

Con el software Minitab 18.1 y la prueba Anderson-Darling, se realiza la prueba de normalidad. Toma de la decisión de rechazar la hipótesis nula si el valor de p es menor o igual que el valor de significancia o no rechazar la hipótesis nula en caso de que el valor de p es mayor que el valor de significancia.

iii) Formulación de las hipótesis estadísticas (nula y alternativa) basado en el resultado de la prueba de normalidad. En el presente estudio de investigación, los datos cumplieron los requisitos para pruebas paramétricas por lo que se utilizó la Prueba T Student basado en prueba de medias.

iv) Especificación el nivel de significancia α .

v) Selección del software para realizar la prueba estadística. En nuestro caso se utilizó Minitab versión 18.1

vi) Establecimiento de la regla de decisión basado en el valor de p .

vii) Cálculo del valor de p utilizando el software Minitab versión 18.1

viii) Toma de la decisión de rechazar la hipótesis nula si el valor de p es menor o igual que el valor de significancia o no rechazar la hipótesis nula en caso de que el valor de p es mayor que el valor de significancia.

ix) Realización de la conclusión respectiva.

2.4. Definiciones conceptuales

Centro de datos (Data Center): Es un espacio que alberga los equipos como servidores, dispositivos de red, etc., necesarios para el procesamiento de datos de una

organización. Cuenta con la energía eléctrica necesaria, ventilación adecuada y sistemas de seguridad.

Servidor: Es un ordenador físico central integrado a una red informática que ejecuta aplicaciones que proporcionan servicios a otros programas, denominados clientes.

Hosting: Es donde se alquila o compra un espacio en un Data Center para un servidor propio y las instalaciones provee energía y conexión de red.

Cloud: Es un término inglés que significa “nube”, este permite brindar servicios de computación a través de una red, que generalmente es internet. Cloud es una red enorme de servidores remotos conectados para funcionar como un único sistema.

Monitoreo de servidores: Es uno de los aspectos más importantes que debe tener en cuenta cualquier empresa de hosting, debido a que ayuda a prevenir problemas en el funcionamiento de los servidores. Esto se logra eficientemente mediante el uso de sofisticados sistemas que se conectan a los servidores para reportar las actividades de estos.

Monitor: En un centro de datos un monitor es un dispositivo que captura datos relacionados al funcionamiento del software o hardware y los almacena en una base de datos local.

Hardware: Se refiere a la parte física de un sistema informático, poseen componentes electrónicos y mecánicos, un ejemplo claro es el servidor.

Software: Es el soporte lógico de un sistema informático, que está compuesto por un conjunto de componentes lógicos necesarios que realizan tareas específicas.

Ticket: Se usa para realiza el seguimiento de requerimientos de usuarios. Generalmente existe un sistema para la gestión de tickets, en donde el usuario puede crear uno y hacer seguimiento hasta su atención y cierre.

Jerarquía: Es un conjunto de miembros organizados en niveles, de mayor a menor. Las jerarquías se establecen de acuerdo a las relaciones entre los miembros.

Granularidad: Se refiere al nivel de detalle de la información almacenada en el data warehouse. Por ejemplo, en una jerarquía la granularidad se podría definir desde el más alto nivel hasta el menor nivel.

Generar reporte: Se refiere a una actividad que consiste en diseñar un nuevo reporte en base a requerimientos específicos de usuarios que necesitan visualizar información.

Ejecutar reporte: Es una actividad que consiste en obtener los resultados de un reporte, el tiempo que demora en mostrar estos resultados depende de muchos factores, como la cantidad de datos, arquitectura del sistema de reportes, etc.

Analizar reporte: Es una actividad que consiste en examinar detenidamente los datos que muestra un reporte en busca de información relevante.

Desplegar reportes en producción: Se refiere a la actividad de poner en funcionamiento los reportes para que los usuarios puedan acceder a ellos. El entorno de hardware en donde se alojan estos reportes se llama producción.

Paquete: En la herramienta SQL Server Integration Services, paquete es un conjunto de tareas relacionadas al proceso de ETL.

2.5. Formulación de la hipótesis

2.5.1. Hipótesis general

La implementación de una solución de Inteligencia de Negocios mejora el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.

2.5.2. Hipótesis específicas

La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima – 2019.

La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima – 2019.

La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.

2.6. Operacionalización de variables e indicadores

La operacionalización de las variables e indicadores se visualiza en la tabla 2.

Tabla 2.

Operacionalización de variables e indicadores.

Variable	Dimensión	Indicador	Unidad de medida	Técnica	Instrumento
Inteligencia de Negocios	Estado	Implementado	Implementado	Observación	Hoja de control
			No Implementado	Observación	Hoja de control
Monitoreo de servidores	Tiempo de generación de reportes	Tiempo de generación de cada reporte de monitoreo de servidores	Minutos que toma la generación de cada reporte	Encuesta	Cuestionario
	Tiempo de ejecución de reportes	Tiempo de ejecución de cada reporte de monitoreo de servidores	Segundos que toma la ejecución de cada reporte	Extracción de datos de la base de datos	Sentencia SQL
	Tiempo de análisis de reportes	Tiempo de análisis de cada reporte de monitoreo de servidores	Minutos que toma el análisis de cada reporte	Encuesta	Cuestionario

CAPÍTULO 3: METODOLOGÍA

3.1. Diseño metodológico

3.1.1. Tipo de investigación

De acuerdo con la naturaleza de la investigación, el presente estudio reúne las condiciones metodológicas de una investigación aplicada porque busca convertir el conocimiento teórico en conocimiento práctico.

3.1.2. Nivel de investigación

El presente estudio es de nivel experimental, específicamente pre-experimental, debido a que se realizó la medición a 10 reportes creados en un mes antes y después de la implementación de la solución de Inteligencia de Negocios.

O₁ X O₂

Para cada grupo de observación (sin y con la solución de Inteligencia de Negocios) se tomaron 10 casos de generación de reportes.

O₁: Se hace referencia al grupo de observación que contiene la muestra de 10 casos de generación de reportes que fueron obtenidos sin el uso de una solución de BI.

O₂: Se hace referencia al grupo de observación que contiene la muestra de 10 casos de generación de reportes que fueron obtenidos con el uso de la solución de BI.

3.1.3. Diseño de investigación

La presente investigación es de diseño experimental, debido a que analizó cómo la variable independiente de la investigación afecta el tiempo de generación de reportes, el tiempo de ejecución de reportes y el tiempo del análisis de reportes.

3.1.4. Enfoque

La investigación presenta un enfoque cuantitativo, debido a que se hará uso de la estadística para el análisis de los datos y la prueba de hipótesis.

3.2. Población y muestra

3.2.1. Población

La población para la presente investigación está compuesta por los reportes más importantes elaborados mensualmente, que son un total de 10 reportes.

3.2.2. Muestra

Para el cálculo de la muestra se decidió trabajar con el 100% de población ya que la población es pequeña, un total de 10 reportes mensuales.

3.3. Técnicas e instrumentos de recolección de datos

Para el desarrollo de la investigación fue necesario emplear las siguientes técnicas e instrumentos para la recolección de los datos, cada una de ellas sirvió para asegurar una adecuada investigación.

3.3.1. Técnica empleada

Se usó la siguiente técnica:

Observación

Se observó con el fin de determinar si está implementado o no la solución de Inteligencia de Negocios.

Encuesta

Se realizaron encuestas con el fin de recopilar información del desarrollo de las actividades de generación de reportes y análisis de reportes y los datos se plasmaron en un cuestionario.

Extracción de datos de la base de datos

Se realizó la extracción de los tiempos de ejecución de reportes desde una base de datos de Microsoft SQL Server Reporting Services donde están desplegados los reportes generados antes y después de implementar la solución de Inteligencia de Negocios.

3.3.2. Descripción de los instrumentos

La información necesaria se obtuvo a través de los siguientes instrumentos:

Cuestionario

Se usaron cuestionarios de preguntas para obtener los tiempos de generación y análisis de cada reporte antes y después de la implementación de la solución de Inteligencia de Negocios.

Sentencia SQL

Se usó una sentencia SQL para obtener los tiempos de ejecución de cada reporte.

Análisis documental

Se realizó el análisis de documentos sobre trabajos de investigación, libros, información en Internet, revistas, etc. Así como documentación interna de la empresa sobre el proceso de ventas del servicio de cloud, el subproceso de monitoreo de servidores y el subproceso de solicitud de reportes.

3.4. Técnicas para el procesamiento de la información

Para el procesamiento y análisis de los datos de la investigación, se utilizó la técnica de diferencias de medias dado que los datos tienen una distribución normal. Se aplicó la prueba estadística de T Student para los indicadores de tiempo de generación de reportes, ejecución de reportes y análisis de reportes, dado que se hizo la comparación de datos tomados antes y después de la implementación de la solución de Inteligencia de Negocios. Adicionalmente se contó con el apoyo de un programa computarizado de análisis estadístico llamado Minitab 18.1 y la hoja de cálculo Excel.

3.5. Matriz de consistencia

La matriz de consistencia se presenta en el anexo 1

CAPÍTULO 4: RESULTADOS

4.1. Descripción de los datos obtenidos

Se seleccionaron 10 reportes generados mensualmente del subproceso de monitoreo de servidores para obtener el tiempo de generación, ejecución y análisis de estos.

Los datos sobre los tiempos de generación de reportes (antes y después de la implementación de BI) que se presentan en la tabla 4, fueron obtenidos a través de un cuestionario que se presenta en el Anexo N°08, que se hicieron a los analistas de datos, quienes son los encargados de generar los reportes.

Los datos sobre los tiempos de ejecución de reportes (antes y después de la implementación de BI) que se presentan en la tabla 5, fueron obtenidos a través de una sentencia SQL en la base de datos de Microsoft SQL Server Reporting Services, en donde se registra el tiempo de ejecución de cada reporte, se tomaron las 3 últimas ejecuciones de cada uno de los 10 reportes y se calculó el promedio para cada reporte.

Los datos sobre los tiempos de análisis de reportes (antes y después de la implementación de BI) que se presentan en la tabla 6, fueron obtenidos a través de un cuestionario que se presenta en el Anexo N°08, que se hicieron a los analistas de monitoreo, quienes son los que usan los reportes para extraer información relevante para el subproceso de monitoreo de servidores.

4.2. Análisis de resultados

El cálculo de la mejora se visualiza en la tabla 3.

Tabla 3

Cálculo de la mejora

ACTIVIDAD	BI NO IMPLEMENTADO	BI IMPLEMENTADO	% MEJORA
Generación de reportes (<i>promedio en minutos</i>)	120	25	79.27%
Ejecución de reportes (<i>promedio en segundos</i>)	23.8	1.9	92.02%
Análisis de reportes (<i>promedio en minutos</i>)	21.8	17.7	18.81%

Tabla 4

Datos depurados de los tiempos que toma la generación de reportes

NRO	REPORTE	TIEMPOS EN MINUTOS		
		BI NO IMPLEMENTADO	BI IMPLEMENTADO	DIFERENCIA
1	Exceeded Threshold Disk Usage	120	25	95
2	Exceeded Threshold Logical Disk Usage	115	20	95
3	Exceeded Threshold Memory Usage	135	32	103
4	Exceeded Threshold Processor Usage	110	21	89
5	Exceeded Threshold Volume Usage	115	22	93
6	TopN Disk Usage	125	30	95
7	TopN Logical Disk Usage	120	27	93
8	TopN Memory Usage	135	35	100
9	TopN Processor Usage	105	18	87
10	TopN Process Volume Usage	120	20	100

Tabla 5

Datos depurados de los tiempos que toma la ejecución de reportes

NRO	REPORTE	TIEMPOS EN SEGUNDOS		
		BI NO IMPLEMENTADO	BI IMPLEMENTADO	DIFERENCIA
1	Exceeded Threshold Disk Usage	35	2	33
2	Exceeded Threshold Logical Disk Usage	27	3	24
3	Exceeded Threshold Memory Usage	15	2	13
4	Exceeded Threshold Processor Usage	42	3	39
5	Exceeded Threshold Volume Usage	34	2	32
6	TopN Disk Usage	23	1	22
7	TopN Logical Disk Usage	15	1	14
8	TopN Memory Usage	17	2	15
9	TopN Processor Usage	12	2	10
10	TopN Process Volume Usage	18	1	17

Tabla 6

Datos depurados de los tiempos que toma el análisis de reportes

NRO	REPORTE	TIEMPOS EN MINUTOS		
		BI NO IMPLEMENTADO	BI IMPLEMENTADO	DIFERENCIA
1	Exceeded Threshold Disk Usage	30	20	10
2	Exceeded Threshold Logical Disk Usage	25	20	5
3	Exceeded Threshold Memory Usage	20	15	5
4	Exceeded Threshold Processor Usage	30	25	5
5	Exceeded Threshold Volume Usage	25	20	5
6	TopN Disk Usage	15	14	1
7	TopN Logical Disk Usage	20	16	4
8	TopN Memory Usage	18	15	3
9	TopN Processor Usage	20	18	2
10	TopN Process Volume Usage	15	14	1

4.3. Prueba de hipótesis

Las pruebas de hipótesis seguirán el procedimiento presentado en el ítem 2.3
Procedimiento utilizado para realizar la prueba de hipótesis.

4.3.1. Comprobación de la hipótesis específica 1

La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores.

i) Obtención y limpieza de los datos de prueba del proceso de investigación respectivo (se consideró solo reportes que fueron creados en el periodo de estudio).

Los datos obtenidos y limpiados de la actividad de generación de reportes se presentan en la tabla 4.

ii) Realización de la prueba de normalidad de los datos usando la prueba Anderson-Darling de la actividad de generación de reportes.

Se planteó la siguiente hipótesis:

H_0 : Los datos de la actividad de generación de reportes se ajustan a una distribución normal

H_1 : Los datos de la actividad de generación de reportes no se ajustan a una distribución normal

Se estableció la siguiente regla de decisión:

Valor de $p \leq \alpha$ Rechazar H_0 a nivel α

Valor de $p > \alpha$ No rechazar H_0 a nivel α

Para nuestro estudio el valor de significancia es de 0.05.

Con el software Minitab 18.1 se obtuvo los siguientes resultados que se visualizan en la figura 9 y se observa que el valor de p es de 0.517.

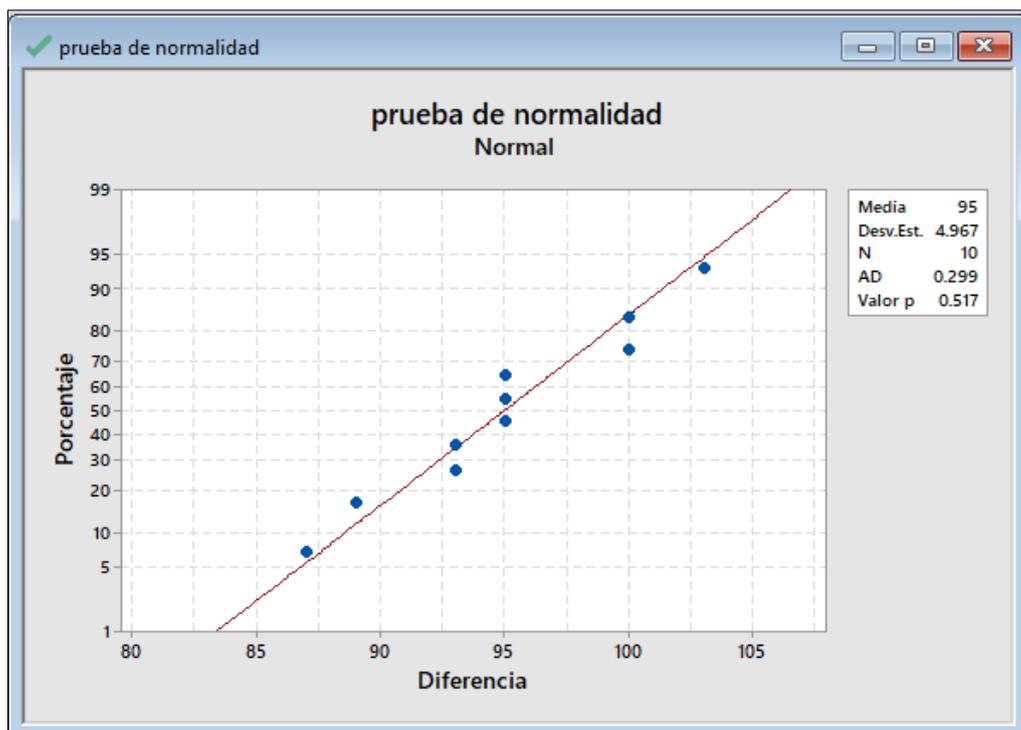


Figura 9. Resultados de la prueba de normalidad de los datos de la actividad de generación de reportes.

Dado que el valor de $p = 0.517$ y el valor de $p \geq 0.05$, no se rechaza H_0 a nivel de α y se rechaza la hipótesis alternativa, por lo que se concluye que los datos son de naturaleza normal por lo que la hipótesis de la investigación tiene que ser probada mediante una prueba paramétrica alternativa.

iii) Formulación de las hipótesis estadísticas (nula y alternativa) basado en el resultado de la prueba de normalidad.

μ_1 : Media del tiempo de generación de reportes antes de la implementación de BI

μ_2 : Media del tiempo de generación de reportes después de la implementación de BI.

Hipótesis nula e hipótesis alternativa:

Hipótesis nula $H_0: \mu_1 \leq \mu_2$

Hipótesis alternativa $H_1: \mu_1 > \mu_2$

iv) Especificación del nivel de significancia.

Se ha convenido que el nivel de significancia será del 0.05

v) Selección del software para realizar la prueba estadística. En nuestro caso se utilizó Minitab versión 18.1

vi) Establecimiento de la regla de decisión basado en el valor de p.

Si valor $p \leq \alpha$ Rechazar H_0 al nivel α

Si valor $p > \alpha$ No rechazar H_0 al nivel α

vii) Cálculo del valor de p con la prueba T Student entre el antes y el después de la implementación de BI utilizando el software Minitab versión 18.1.

En la figura 10 se presentan los resultados de la prueba T Student entre el antes y el después de la implementación de BI del tiempo que toma la generación de reportes.

Prueba		
Hipótesis nula	$H_0: \mu_1 - \mu_2 = 0$	
Hipótesis alterna	$H_1: \mu_1 - \mu_2 > 0$	
Valor T	GL	Valor p
26.55	14	0.00000000000001120

Figura 10. Resultados de la prueba T Student entre el antes y el después de la implementación de BI para la actividad de generación de reportes.

viii) Toma de la decisión de rechazar la hipótesis nula si el valor de p es menor o igual que el valor de significancia o no rechazar la hipótesis nula en caso de que el valor de p es mayor que el valor de significancia.

Dado que el valor de $p = 0.00000000000001120$ y el valor de $p \leq 0.05$ se rechaza H_0 a nivel de α y se acepta la hipótesis alternativa.

ix) Realizar la conclusión respectiva. Se concluye que la implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores.

4.3.2. Comprobación de la hipótesis específica 2

La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores.

i) Obtención y limpieza de los datos de prueba del proceso de investigación respectivo (se consideró solo reportes que fueron creados en el periodo de estudio).

Los datos obtenidos y limpiados de la actividad de ejecución de reportes se presentan en la tabla 5.

ii) Realización de la prueba de normalidad de los datos usando la prueba Anderson-Darling de la actividad de ejecución de reportes.

Se planteó la siguiente hipótesis:

H_0 : Los datos de la actividad de ejecución de reportes se ajustan a una distribución normal

H_1 : Los datos de la actividad de ejecución de reportes no se ajustan a una distribución normal

Se estableció la siguiente regla de decisión:

Valor de $p \leq \alpha$ Rechazar H_0 a nivel α

Valor de $p > \alpha$ No rechazar H_0 a nivel α

Para nuestro estudio el valor de significancia es de 0.05.

Con el software Minitab 18.1 se obtuvo los siguientes resultados que se visualizan en la figura 11 y se observa que el valor de p es de 0.327.

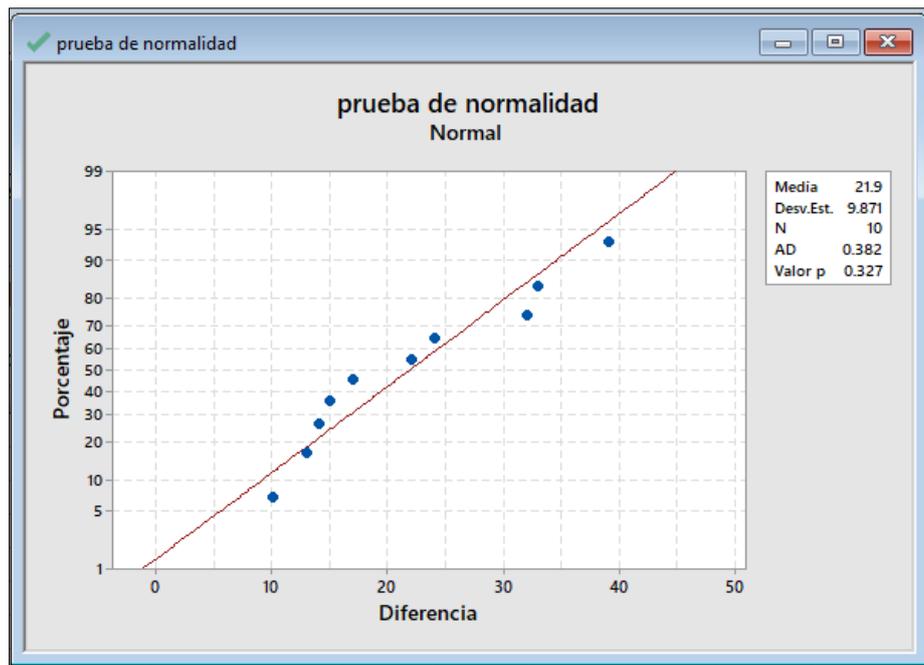


Figura 11. Resultados de la prueba de normalidad de los datos de la actividad de ejecución de reportes.

Dado que el valor de $p = 0.327$ y el valor de $p \geq 0.05$, no se rechaza H_0 a nivel de α y se rechaza la hipótesis alternativa, por lo que se concluye que los datos son de naturaleza normal por lo que la hipótesis de la investigación tiene que ser probada mediante una prueba paramétrica alternativa.

iii) Formulación de las hipótesis estadísticas (nula y alternativa) basado en el resultado de la prueba de normalidad.

μ_1 : Media del tiempo de ejecución de reportes antes de la implementación de BI

μ_2 : Media del tiempo de ejecución de reportes después de la implementación de BI.

Hipótesis nula e hipótesis alternativa:

Hipótesis nula $H_0: \mu_1 \leq \mu_2$

Hipótesis alternativa $H_1: \mu_1 > \mu_2$

iv) Especificación del nivel de significancia.

Se ha convenido que el nivel de significancia será del 0.05

v) Selección del software para realizar la prueba estadística.

En nuestro caso se utilizó Minitab versión 18.1

vi) Establecimiento de la regla de decisión basado en el valor de p.

Si valor $p \leq \alpha$ Rechazar H_0 al nivel α

Si valor $p > \alpha$ No rechazar H_0 al nivel α

vii) Cálculo del valor de p con la Prueba T Student entre el antes y el después de la implementación de BI utilizando el software Minitab versión 18.1.

En la figura 12 se presenta los resultados de la prueba T Student entre el antes y el después de la implementación de BI del tiempo que toma la ejecución de reportes.

Prueba		
Hipótesis nula	$H_0: \mu_1 - \mu_2 = 0$	
Hipótesis alterna	$H_1: \mu_1 - \mu_2 > 0$	
Valor T	GL	Valor p
6.74	9	0.0000423537898599

Figura 12. Resultados de la prueba T Student entre el antes y el después de la implementación de BI para la actividad de ejecución de reportes.

viii) Toma de la decisión de rechazar la hipótesis nula si el valor de p es menor o igual que el valor de significancia o no rechazar la hipótesis nula en caso de que el valor de p es mayor que el valor de significancia.

Dado que el valor de $p = 0.0000423537898599$ y el valor de $p \leq 0.05$ se rechaza H_0 a nivel de α y se acepta la hipótesis alternativa.

ix) Realizar la conclusión respectiva. Se concluye que la implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores.

4.3.3. Comprobación de la hipótesis específica 3

La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores.

i) Obtención y limpieza de los datos de prueba del proceso de investigación respectivo (se consideró solo reportes que fueron creados en un mes durante el periodo de estudio).

Los datos obtenidos y limpiados de la actividad de análisis de reportes se presentan en la tabla 6.

ii) Realización de la prueba de normalidad de los datos usando la prueba Anderson-Darling de la actividad de análisis de reportes.

Se planteó la siguiente hipótesis:

H_0 : Los datos de la actividad de análisis de reportes se ajustan a una distribución normal

H_1 : Los datos de la actividad de análisis de reportes no se ajustan a una distribución normal

Se estableció la siguiente regla de decisión:

Valor de $p \leq \alpha$ Rechazar H_0 a nivel α

Valor de $p > \alpha$ No rechazar H_0 a nivel α

Para nuestro estudio el valor de significancia es de 0.05.

Con el software Minitab 18.1 se obtuvo los siguientes resultados que se visualizan en la figura 13 y se observa que el valor de p es de 0.206.

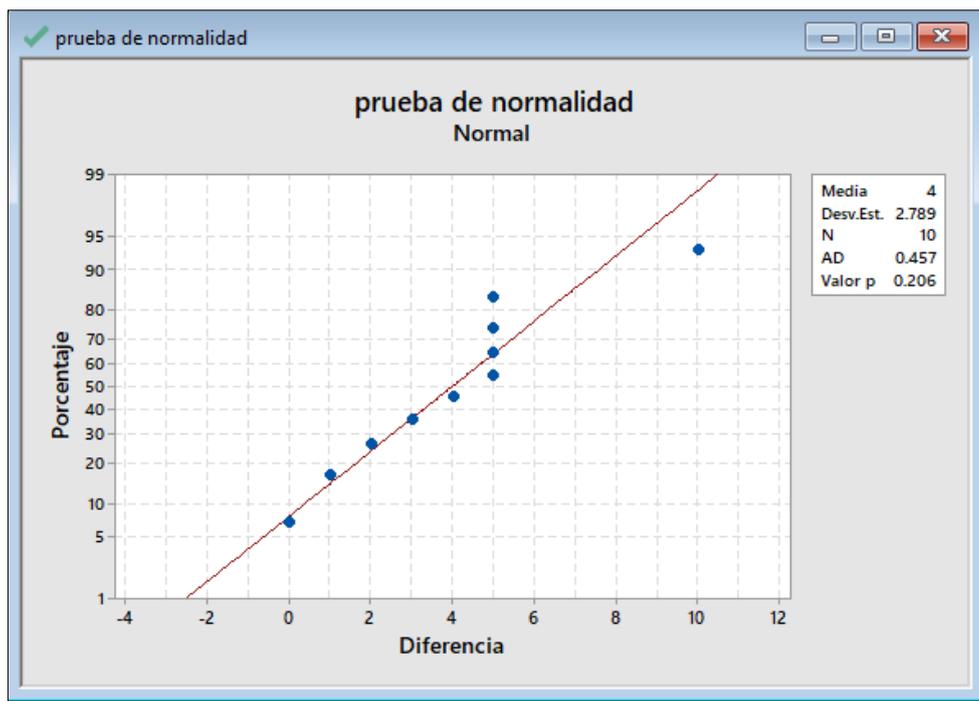


Figura 13. Resultados de la prueba de normalidad de los datos de la actividad de análisis de reportes.

Dado que el valor de $p = 0.206$ y el valor de $p \geq 0.05$, no se rechaza H_0 a nivel de α y se rechaza la hipótesis alternativa, por lo que se concluye que los datos son de naturaleza normal por lo que la hipótesis de la investigación tiene que ser probada mediante una prueba paramétrica alternativa.

iii) Formulación de las hipótesis estadísticas (nula y alternativa) basado en el resultado de la prueba de normalidad.

μ_1 : Media del tiempo de análisis de reportes antes de la implementación de BI

μ_2 : Media del tiempo de análisis de reportes después de la implementación de

BI.

Hipótesis nula e hipótesis alternativa:

Hipótesis nula $H_0: \mu_1 \leq \mu_2$

Hipótesis alternativa $H_1: \mu_1 > \mu_2$

iv) Especificación del nivel de significancia.

Se ha convenido que el nivel de significancia será del 0.05

v) Selección del software para realizar la prueba estadística.

En nuestro caso se utilizó Minitab versión 18.1

vi) Establecimiento de la regla de decisión basado en el valor de p.

Si valor $p \leq \alpha$ Rechazar H_0 al nivel α

Si valor $p > \alpha$ No rechazar H_0 al nivel α

vii) Cálculo del valor de p con la Prueba T Student entre el antes y el después de la implementación de BI utilizando el software Minitab versión 18.1.

En la figura 14 se presentan los resultados de la Prueba T Student entre el antes y el después de la implementación de BI del tiempo que toma el análisis de reportes.

Prueba		
Hipótesis nula	$H_0: \mu_1 - \mu_2 = 0$	
Hipótesis alterna	$H_1: \mu_1 - \mu_2 > 0$	
Valor T	GL	Valor p
1.98	15	0.033

Figura 14. Resultados de la prueba T Student entre el antes y el después de la implementación de BI para la actividad de análisis de reporte.

viii) Toma de la decisión de rechazar la hipótesis nula si el valor de p es menor o igual que el valor de significancia o no rechazar la hipótesis nula en caso de que el valor de p es mayor que el valor de significancia.

Dado que el valor de $p = 0.0331446667471780$ y el valor de $p \leq 0.05$ se rechaza H_0 a nivel de α y se acepta la hipótesis alternativa.

ix) Realizar la conclusión respectiva. Se concluye que la implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores.

4.3.4. Comprobación de la hipótesis general

La comprobación de la hipótesis general se realizó teniendo como base las conclusiones de las hipótesis específicas que se presentan en la tabla 7.

Tabla 7
Comprobación de prueba de hipótesis general

Hipótesis	Media	Hipótesis Nula e Hipótesis alternativa	Valor de p	Sign.	Regla de decisión	Resultado	Conclusión
Hipótesis específica 1							
La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.	μ_1 : Media del tiempo de generación de reportes antes de la implementación de BI. μ_2 : Media del tiempo de generación de reportes después de la implementación de BI.	Hipótesis nula $H_0: \mu_1 \leq \mu_2$ Hipótesis alternativa $H_1: \mu_1 > \mu_2$	$p = 0.0000000000001120$	0.05	Si valor $p \leq \alpha$ Rechazar H_0 al nivel α Si valor $p > \alpha$ No rechazar H_0 al nivel α	Dado que el valor de $p = 0.0000000000001120$ y el valor de $p \leq 0.05$ se rechaza H_0 a nivel de α y se acepta la hipótesis alternativa.	La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.
Hipótesis específica 2							
La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.	μ_1 : Media del tiempo de ejecución de reportes antes de la implementación de BI. μ_2 : Media del tiempo de ejecución de reportes después de la implementación de BI.	Hipótesis nula $H_0: \mu_1 \leq \mu_2$ Hipótesis alternativa $H_1: \mu_1 > \mu_2$	$p = 0.0000423537898599$	0.05	Si valor $p \leq \alpha$ Rechazar H_0 al nivel α Si valor $p > \alpha$ No rechazar H_0 al nivel α	Dado que el valor de $p = 0.0000423537898599$ y el valor de $p \leq 0.05$ se rechaza H_0 a nivel de α y se acepta la hipótesis alternativa.	La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.
Hipótesis específica 3							
La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.	μ_1 : Media del tiempo de análisis de reportes antes de la implementación de BI. μ_2 : Media del tiempo de análisis de reportes después de la implementación de BI.	Hipótesis nula $H_0: \mu_1 \leq \mu_2$ Hipótesis alternativa $H_1: \mu_1 > \mu_2$	$p = 0.0331446667471780$	0.05	Si valor $p \leq \alpha$ Rechazar H_0 al nivel α Si valor $p > \alpha$ No rechazar H_0 al nivel α	Dado que el valor de $p = 0.0331446667471780$ y el valor de $p \leq 0.05$ se rechaza H_0 a nivel de α y se acepta la hipótesis alternativa.	La implementación de una solución de Inteligencia de Negocios mejora el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.
Dado que:							
<ul style="list-style-type: none"> ❖ La hipótesis específica 1 comprueba que la implementación de una solución de BI mejora el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima -2019; ❖ La hipótesis específica 2 comprueba que la implementación de una solución de BI mejora el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima -2019; ❖ La hipótesis específica 3 comprueba que la implementación de una solución de BI mejora el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima -2019; 							
Entonces							
<ul style="list-style-type: none"> ❖ Se comprueba que la implementación de una solución de BI mejora el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019. 							
Hipótesis general							
La implementación de una solución de Inteligencia de Negocios mejora el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.		La implementación de una solución de Inteligencia de Negocios mejora el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.					

CAPÍTULO 5: DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES

5.1. Discusión

Con respecto a (Buitrago M. & González S., 2016), en la investigación *Implementación de un sistema de inteligencia de negocios para el apoyo en la toma de decisiones en una empresa integradora de soluciones de telecomunicación*, tenemos las siguientes coincidencias:

Que con el uso de Inteligencia de Negocios se logra reducir muy considerablemente los tiempos de los reportes y consultas analíticas del negocio. También coincidimos en que se disminuye notablemente los tiempos de análisis de los datos. Cabe resaltar que la metodología usada también fue la de Ralph Kimball.

Entre las divergencias tenemos que (Buitrago M. & González S., 2016) utilizaron como plataforma tecnológica la aplicación de Oracle Business Intelligence mientras que en mi investigación utilicé Microsoft SQL Server Análisis Services 2012.

Con respecto a (González Pinos, 2013), en la investigación *Análisis costo / beneficio de la implementación de Business Intelligence en pymes de la ciudad de Quito, del sector de redes y telecomunicaciones*, tenemos las siguientes coincidencias:

Propone el uso de una herramienta de BI que resulte eficiente en el tiempo y en la obtención de información y con ello una mejor y rápida toma de decisiones. Con el uso de la herramienta, el gerente general estimó una eficiencia de hasta el 60% en el tiempo de manejo de la información.

Entre las divergencias tenemos que (González Pinos, 2013) mejoró los aspectos de obtención y análisis de información, mientras que en mi investigación se basa en las mejoras en la generación, ejecución y análisis de reportes.

Con respecto a (Espinoza Vargas & Palomino Ruiz, 2016), en la investigación *Desarrollo de un Datamart para optimizar la generación de información estratégica de apoyo a la toma de decisiones en la vicepresidencia de banca comercial de Interbank Perú*, tenemos las siguientes coincidencias:

Resalta que mediante el Datamart se logró disminuir el tiempo de generación de información estratégica a 1 hora y 37 minutos, lo cual equivale a un 90.6%. También coincidimos en el uso de la metodología de Ralph Kimball para el desarrollo de la investigación.

Entre las divergencias tenemos que (Espinoza Vargas & Palomino Ruiz, 2016) utilizaron como plataforma tecnológica la aplicación de IBM Cognos mientras que en mi investigación utilicé Microsoft SQL Server Análisis Services 2012.

Con respecto a (Zamora Saldarriaga, 2017), en la investigación *Implementación de un Datamart para la mejora en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional*, tenemos las siguientes coincidencias:

Que mediante la implementación de un Datamart el tiempo de atención de los requerimientos de información se redujo considerablemente. También mejoró el tiempo de ejecución de las consultas de grandes volúmenes de datos. Se usó la metodología de Ralph Kimball para el desarrollo de la solución.

Entre las divergencias tenemos que (Zamora Saldarriaga, 2017) utilizó como plataforma tecnológica la aplicación de Power BI y Power Pivot mientras que en mi investigación utilicé Microsoft SQL Server Análisis Services 2012, la cual es una herramienta más completa y potente, y que efectos de mi investigación fue necesaria. Adicionalmente el autor realizó el análisis de satisfacción del cliente, el cual en mi investigación no fue considerado.

Con respecto a (Carhuallanqui Bastidas, 2017), en la investigación *Diseño de una solución de inteligencia de negocios como herramienta de apoyo a la toma de decisiones en el área de ventas de la empresa farmacéutica Dispefarma*, tenemos las siguientes coincidencias:

Uno de los principales problemas que tienen las empresas es la dificultad del análisis de los datos desde diferentes perspectivas según sea la necesidad de los usuarios. Por ello se tiene que diseñar una solución de inteligencia de negocios que permita integrar la información y poder analizarla de tal forma que brinde reportes de acuerdo a las necesidades de los usuarios. Se redujo el tiempo promedio en elaboración de informes solicitados por la gerencia en un 90.13%, es decir de 7.6 a 0.75 horas. Al igual que mi investigación, el autor usó la metodología de Ralph Kimball.

Entre las divergencias tenemos que (Carhuallanqui Bastidas, 2017) realizó la investigación sobre el nivel de satisfacción de los usuarios y la calidad de la información, mientras que en mi investigación no se tomaron en cuenta esos aspectos.

Con respecto a (Inca Damián & Zavala Sanchez, 2016), en la investigación *Desarrollo de una solución de inteligencia de negocios para la mejora del proceso de toma de decisiones en el área de administración tributaria de la Municipalidad Distrital de San Bartolo*, tenemos las siguientes coincidencias:

Se planteó el desarrollo de una solución de Inteligencia de Negocios, que permita reducir los tiempos en el proceso de obtención de los reportes y a su vez disminuir el esfuerzo desplegado en dicho proceso. Se observa que se logró reducir el tiempo empleado para generar los reportes en el área de Administración Tributaria en un 96% aproximadamente. También se observa que se logró reducir en un 67 % aproximadamente el tiempo que el gerente de Administración Tributaria empleaba en el análisis de la información

obtenida en los reportes. Adicionalmente se puede ver que la metodología de desarrollo usada fue la de Ralph Kimball.

Entre las divergencias tenemos que (Inca Damián & Zavala Sanchez, 2016) realizan la investigación sobre el nivel de satisfacción de los usuarios, mientras que en mi investigación no se tomó en cuenta ese aspecto.

5.2. Conclusiones

Después de concluida la investigación tenemos las siguientes conclusiones:

Se evaluó y demostró que la implementación de una solución de Inteligencia de Negocios mejora la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, logrando una mejora de 79.27%, beneficiando al área de TI, ya que podrán tener más tiempo para dedicarse a otras actividades y también al área de monitoreo de servidores porque podrán tener los reportes oportunamente.

Se evaluó y demostró que la implementación de una solución de Inteligencia de Negocios mejora la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, logrando una mejora de 92.02%, lo cual beneficia a los analistas de monitoreo ya que podrán ejecutar los reportes tantas veces lo necesiten.

Se evaluó y demostró que la implementación de una solución de Inteligencia de Negocios mejora la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, logrando una mejora de 18.81%, brindando una mejor experiencia de análisis.

Se evaluó y demostró que la implementación de una solución de Inteligencia de Negocios mejora el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud.

5.3. Recomendaciones

Se recomienda lo siguiente:

Ampliar la implementación de la solución Inteligencia de Negocios a otras áreas del centro de datos como por ejemplo el área de soporte técnico y aprovisionamiento de energía eléctrica, estas áreas también manejan volúmenes considerables de datos.

Implementar minería de datos, resultaría muy beneficioso para encontrar patrones que ayuden a prever incidentes con los servidores. Ya se cuenta con un repositorio centralizado de los datos de monitoreo de servidores que serviría como fuente de datos para los modelos de minería de datos.

Capacitar adecuadamente a los usuarios sobre el uso del cubo de información creado, por ejemplo, para acceder a los datos desde Excel los usuarios necesitan cierto conocimiento del cubo y cómo usarlo adecuadamente.

El tiempo ahorrado a causa de la implementación de la solución de Inteligencia de Negocios, se debería usar en actividades de prevención.

REFERENCIAS BIBLIOGRÁFICAS

- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la investigación*. México: McGraw-Hill.
- Andreu, R., Ricart, J., & Valor, J. (1991). *Estrategia y sistema de información*. España: McGraw-Hill.
- Ramos, S. (2011). *Microsoft Business Intelligence: vea el cubo medio lleno*. España: SolidQPress.
- Medina La Plata, E. (2012). *Business Intelligence. Una guía práctica*. Perú: Universidad Peruana de Ciencias Aplicadas.
- Curto Diaz, J. & Conesa Caralt, J. (2010). *Introducción al Business Intelligence*. España: Universitat Oberta de Catalunya.
- Kimball R. & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Estados Unidos: John Wiley & Sons.
- Espinosa R. (2010). *Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional*. Recuperado de <https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimencional/>
- Cano, J. L. (2007). *Business Intelligence: competir con información*. España: Fundación Cultural.
- Mendez, A., Mártire, A., Britos, P. & Garcia-Martínez, R. (2003). *Fundamentos de Data Warehouse*. Recuperado de <http://artemisa.unicauca.edu.co/~ecaldon/docs/bd/fundamentosdedatawarehouse.pdf>
- Rivadera, G. R. (2010). *La Metodología de Kimball para el Diseño de almacenes*. Argentina: Universidad Católica de Salta. Recuperado de <http://www.ucasal.edu.ar/hm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>

- Rojas Saldívar, A. (2014). *Implementación de un Data Mart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el Departamento de Finanzas de la Contraloría General de la República*. (Tesis de pregrado). Universidad de San Martín de Porres, Chiclayo, Perú.
- Carhuallanqui Bastidas, J. L. (2017). *Diseño de una solución de inteligencia de negocios como herramienta de apoyo a la toma de decisiones en el área de ventas de la empresa farmacéutica Dispefarma*. (Tesis de pregrado). Universidad Nacional Mayor de San Marcos, Lima, Perú.
- Espinoza Vargas, J. M. & Palomino Ruiz, C. A. (2016). *Desarrollo de un Datamart para optimizar la generación de información estratégica de apoyo a la toma de decisiones en la vicepresidencia de banca comercial de Interbank Perú*. (Tesis de pregrado). Universidad San Martín de Porres, Lima, Perú.
- Inca Damián, R. N. & Zavala Sánchez, L. F. (2016). *Desarrollo de una solución de inteligencia de negocios para la mejora del proceso de toma de decisiones en el área de administración tributaria de la Municipalidad Distrital de San Bartolo*. (Tesis de pregrado). Universidad Autónoma del Perú, Lima, Perú.
- Zamora Saldarriaga, J. E. (2017). *Implementación de un Datamart para la mejora en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional*. (Tesis de pregrado). Inca Garcilaso de la Vega, Lima, Perú.
- Buitrago M., Carlos M. & González S., Juan A. (2016). *Implementación de un sistema de inteligencia de negocios para el apoyo en la toma de decisiones en una empresa integradora de soluciones de telecomunicación*. (Tesis de pregrado). Universidad Central de Venezuela, Caracas, Venezuela.

González Pinos, S. A. (2013). *Análisis costo / beneficio de la implementación de Business Intelligence en pymes de la ciudad de Quito, del sector de redes y telecomunicaciones.*

(Tesis de pregrado). Universidad Internacional del Ecuador, Quito, Ecuador.

Heroix. (s.f.). *What is Server Monitoring?*. Recuperado el 06 de 11 de 2018, de Heroix:

<https://www.heroix.com/download/WhatisServerMonitoring.pdf>

Kerkeleä, J. (2018). *Continuous Integration Server Performance.* (Tesis de maestría).

University of Oulu, Finlandia.

ANEXOS

Anexo 1 – Matriz de consistencia

Inteligencia de Negocios y el monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019					
Problema	Objetivos	Hipótesis	Variables	Indicadores	Metodología
<p>Problema general ¿La implementación de una solución de Inteligencia de Negocios mejorará el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019?</p> <p>Problemas específicos ¿La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019?</p> <p>¿La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019?</p> <p>¿La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019?</p>	<p>Objetivo general Implementar una solución de Inteligencia de Negocios para mejorar el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.</p> <p>Objetivos específicos Implementar una solución de Inteligencia de Negocios para mejorar el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.</p> <p>Implementar una solución de Inteligencia de Negocios para mejorar el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.</p> <p>Implementar una solución de Inteligencia de Negocios para mejorar el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.</p>	<p>Hipótesis general La implementación de una solución de Inteligencia de Negocios mejorará el subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.</p> <p>Hipótesis específicas La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de generación de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.</p> <p>La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de ejecución de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.</p> <p>La implementación de una solución de Inteligencia de Negocios mejorará el tiempo de la actividad de análisis de reportes del subproceso de monitoreo de servidores en el centro de datos de una empresa de cloud, Lima - 2019.</p>	<p>Variable X: Inteligencia de Negocios</p> <p>Dimensión: Estado</p> <p>Variable Y: Monitoreo de servidores</p> <p>Dimensiones: Tiempo de generación de reportes</p> <p>Tiempo de ejecución de reportes</p> <p>Tiempo de análisis de reportes</p>	<p>i. BI Implementado ii. BI No Implementado</p> <p>Tiempo de generación de cada reporte de monitoreo de servidores, en minutos.</p> <p>Tiempo de ejecución de cada reporte de monitoreo de servidores, en segundos.</p> <p>Tiempo de análisis de cada reporte de monitoreo de servidores, en minutos.</p>	<p>Población: 10 reportes generados por mes.</p> <p>Muestra: Se decidió trabajar con toda la población, un total de 10 reportes por mes.</p> <p>Tipo de Investigación: Investigación aplicada.</p> <p>Nivel de Investigación: Pre-experimental.</p> <p>O₁ X O₂</p> <p>Para cada grupo de observación (sin y con la solución de BI) se tomarán 10 casos de generación de reportes.</p> <p>Enfoque: Cuantitativo.</p> <p>Estadístico de prueba: Prueba de hipótesis T Student para cada indicador (tiempo).</p> <p>Instrumentos:</p> <ul style="list-style-type: none"> ✓ Análisis documental ✓ Observación ✓ Cuestionario

Anexo 2 – Planificación del proyecto

Alcance

El alcance del desarrollo del data mart abarca toda la información de monitoreo de servidores, el cual proviene de siete fuentes de datos (monitores).

Metodología de desarrollo

La metodología elegida para el desarrollo del data mart fue la de Ralph Kimball, el cual se ha definido anteriormente.

Tareas identificadas

Las tareas más importantes que se definieron se describen a continuación:

- Identificación de las fuentes de datos.
- Evaluación de los datos.
- Definición de la arquitectura de solución de BI.
- Creación de las bases de datos en donde se almacenará la información cargada. Se crearon dos bases de datos: SDPBI_Stage y SDPBI_DW, la primera se usa como un repositorio temporal que sirve como apoyo al proceso de extracción, transformación y carga, y la segunda es el data mart en donde se almacenan los datos que servirán como fuente de datos para el cubo multidimensional.
- Creación de los procesos de extracción, transformación y carga de los datos, con el apoyo de la herramienta Microsoft SQL server Integration Services.
- Creación del cubo multidimensional con la herramienta Microsoft SQL server Analysis Services.

Programación de las tareas identificadas

En la tabla 7 se muestra la planificación del proyecto:

Tabla 8.

Cronograma de tareas para el desarrollo del data mart

N°	Tareas	Semana							
		1	2	3	4	5	6	7	8
1	Identificación de las fuentes de datos.								
2	Evaluación de los datos								
3	Definición de la arquitectura de solución de BI.								
4	Creación de las bases de datos								
5	Creación de los procesos de extracción, transformación y carga de los datos								
6	Creación del cubo multidimensional								

Anexo 3 – Definición de requerimientos

Para la definición de los requerimientos se tuvieron muchas reuniones con todos los usuarios involucrados en el proyecto, dando como resultado lo siguiente:

Requerimientos funcionales

- El sistema debe permitir hacer proyecciones de las métricas de consumo de recursos de servidores.
- El sistema debe permitir crear reportes con un máximo nivel de detalle.
- El sistema debe permitir migrar todos los reportes existentes relacionados al monitoreo de servidores que hayan sido usados al menos una vez en los últimos 6 meses.
- El sistema debe permitir obtener información de métricas de servidores usados por cada cliente (Customer).
- El sistema debe mostrar información por año, trimestre, mes, día, hora y minuto.
- El sistema debe soportar reprocesos de los datos de forma manual y automática.
- El sistema debe permitir obtener información de cada monitor.

Requerimientos no funcionales

- El acceso a la información debe ser en el mejor tiempo posible.
- La frecuencia de carga debe ser de cinco minutos como máximo.
- El data mart debe funcionar sobre la plataforma de Microsoft Windows Server 2012, utilizando el manejador de bases de datos Microsoft SQL Server 2012, para la carga de datos se debe usar Microsoft SQL Server 2012 Integration

Services y para la construcción del cubo multidimensional se debe usar Microsoft SQL Server 2012 Analysis Services.

- Adecuada administración de la seguridad de los datos.
- Para la creación de reportes se usará Microsoft SQL Server 2012 Reporting Services.
- El sistema debe tener alta disponibilidad.
- El data mart tendrá siete fuentes de información, y queda abierta la posibilidad de adicionar más fuentes.

Matriz de dimensiones vs hechos

La tabla 9 muestra la relación que existe entre las dimensiones y las métricas.

Tabla 9.

Matriz de dimensiones vs hechos

Hechos	Métricas	Dimensiones					
		Date	Time	Customer	Server	Monitor	Metric
Monitoring	Value Avg	X	X	X	X	X	X
	Value Min	X	X	X	X	X	X
	Value Max	X	X	X	X	X	X
	Threshold				X		X

Gráfico del análisis dimensional

La figura 15 muestra el análisis dimensional.

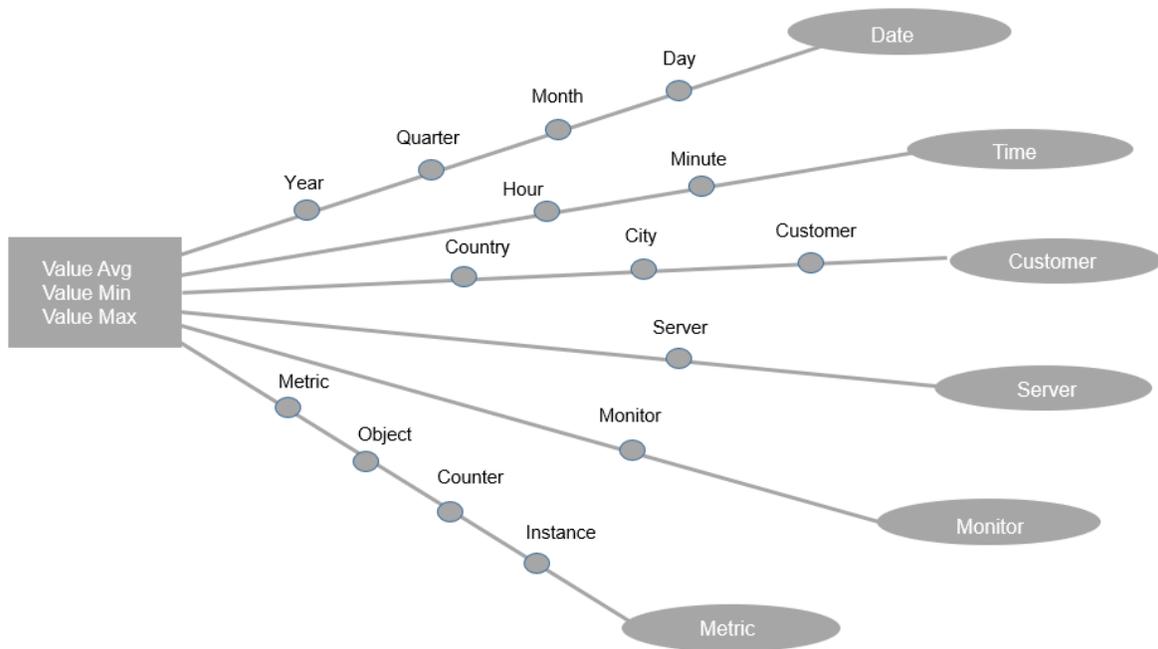


Figura 15. Gráfico del análisis dimensional.

Anexo 4 – Modelo dimensional

El modelo dimensional se ha definido usando el esquema estrella, el cual fue descrito anteriormente. Las tablas dimensionales tendrán una clave primaria simple, y la tabla de hechos, tendrá también una clave simple.

Dimensiones

Como resultado del análisis de requerimientos se obtuvieron las siguientes dimensiones, los cuales se muestran en la tabla 10.

Tabla 10.

Lista de dimensiones

N°	Dimensión
1	Date
2	Time
3	Customer
4	Server
5	Monitor
6	Metric

1. Dimensión Date

Esta dimensión contiene las fechas por las que se analizarán las métricas de monitoreo de servidores, esta dimensión no considera la hora, llegando solo hasta el nivel de día como el menor nivel de granularidad.

La dimensión tiene dos jerarquías:

Jerarquía 1	Jerarquía 2
Dates	Date Quarter
.Year	.Year
..Month	..Quarter
...Day	...Month
Day

La dimensión tiene los siguientes atributos:

Nombre	Descripción	Formato	Ejemplo de datos
DateID	Código de la fecha, es la concatenación de año, mes y día	Numérico	20180101, 20180102, ...
Date	Fecha	Fecha	2018-01-01, 2018-01-02, ...
MonthID	Código del mes, es la concatenación de año y mes.	Numérico	201801, 201802, ...
Month	Mes del año, es concatenación de la abreviatura del mes y el año	Cadena	Jan 2018, Feb 2018, ...
QuarterID	Código del trimestre, es la concatenación de año y el número de trimestre	Numérico	20181, 20182, ...
Quarter	Trimestre del año, es la concatenación del trimestre (Q1, Q2, Q3 o Q4) y el año	Cadena	Q1 – 2018, Q2 – 2018, ...
Year	Año	Numérico	2018, 2019, ...

2. Dimensión Time

Esta dimensión contiene las horas y minutos por las que se analizarán las métricas de monitoreo de servidores.

La dimensión tiene una sola jerarquía:

Jerarquía 1
Date
.Hour
..Minute

La dimensión tiene los siguientes atributos:

Nombre	Descripción	Formato	Ejemplo de datos
TimeID	Código de la hora	Numérico	10001, 10419, 12358, ...
Minute	Minuto, es la concatenación de la hora y el minuto	cadena	00:01, 04:19, 23:58, ...
Hour	Hora, representado por dos dígitos	cadena	00, 04, 23, ...

3. Dimensión Customer

Esta dimensión contiene a todos los clientes que tienen servicios contratados con la empresa.

La dimensión tiene los siguientes atributos:

Nombre	Descripción	Formato	Ejemplo de datos
CustomerID	Código del cliente, es un correlativo	Numérico	1, 2, 3, ...
CustomerCode	Código del cliente en el sistema transaccional	Numérico	12587104, 12582384, ...
CustomerName	Nombre del cliente	Cadena	NEPON DICKINSON COMPANY, INC.
CustomerShortName	Nombre corto del cliente	Cadena	NEPON COMPANY, ...
CustomerStatus	Estado del cliente	Cadena	ACT, INA, ...
ServiceManager	Nombre del administrador de la cuenta	Cadena	Jeffrey Merle, Tim McLaren, ...
City	Ciudad de donde proviene el cliente	Cadena	Miami, New York, ...
Country	País de donde proviene el cliente	Cadena	USA, ...
BusinessUnit	Nombre de la unidad de negocio del cliente	Cadena	UMCloud, ...
LineOfBusiness	Nombre de la línea de negocio del cliente	Cadena	Colo, Cloud, ...
SiteID	Código del cliente de migración	Cadena	1-28KZLQ, 8758, ...

4. Dimensión Server

Esta dimensión contiene todos los servidores del data center.

La dimensión tiene los siguientes atributos:

Nombre	Descripción	Formato	Ejemplo de datos
ServerID	Código del servidor, es un correlativo	Numérico	1, 2, 3, ...
ServerCode	Código del servidor en el sistema transaccional	Numérico	7394731, 7394734, ...
ServerName	Nombre del servidor	Cadena	DOWB384354, DOBD385535, ...
OperatingSystem	Sistema operativo del servidor	Cadena	Unix, Windows, ...
NumberCPU	Número de CPUs del servidor	Cadena	16, 4, 8, ...
PublicIP	Dirección IP pública del servidor	Cadena	165.112.40.57, 164.119.17.26, ...
BackendIP	Dirección IP backend del servidor	Cadena	10.181.197.97, 10.129.119.8, ...

5. Dimensión Monitor

Esta dimensión contiene todos los monitores, los cuales son las fuentes que almacenan datos de la actividad de los servidores.

La dimensión tiene los siguientes atributos:

Nombre	Descripción	Formato	Ejemplo de datos
MonitorID	Código del monitor, es un correlativo	Numérico	1, 2, 3, ...
MonitorCode	Código del monitor en el sistema transaccional	Numérico	1, 2, 3, ...
MonitorName	Nombre del monitor	Cadena	NetIQ-DLT2100B\INSTANCE2\QDB11, NetIQ-DLT2100B\INSTANCE2\QDB12, ...

6. Dimensión Metric

Esta dimensión contiene las descripciones de las métricas sobre la actividad de los servidores.

La dimensión posee una sola jerarquía:

Jerarquía 1
Metrics
.Metric
..Object
...Counter
....Instance

La dimensión tiene los siguientes atributos:

Nombre	Descripción	Formato	Ejemplo de datos
MetricID	Código de la métrica, es un correlativo	Numérico	1, 2, 3, ...
Metric	Nombre de la métrica	cadena	DiskSpace, Services, ...
Object	Es una subdivisión de la métrica	cadena	FileSystem, LogicalDisk, ...
Counter	Es la unidad de medida de la métrica	cadena	Percent in use, Free Megabytes, ...
Instance	Instancia en donde se aplica la métrica, por ejemplo una unidad de disco, nombre de archivo, etc.	cadena	C:, D:, /var/crash, ...

Tabla de hechos o Fact Table

Como resultado del análisis de requerimientos, se obtuvo una sola tabla de hechos, llamada Fact Monitoring.

1. Fact Monitoring

La Fact Monitoring consolidará todos los datos de monitoreo, el cual será cargado al cubo multidimensional.

La tabla de hechos tiene los siguientes atributos:

Nombre	Descripción	Formato	Dimensión
RowID	Código de la lectura, es un correlativo	Numérico	
DateID	Llave para relacionar la Fact Monitoring con la Dimensión Date	Numérico	DimDate
TimeID	Llave para relacionar la Fact Monitoring con la Dimensión Time	Numérico	DimTime
MonitorID	Llave para relacionar la Fact Monitoring con la Dimensión Monitor	Numérico	DimMonitor
CustomerID	Llave para relacionar la Fact Monitoring con la Dimensión Customer	Numérico	DimCustomer
ServerID	Llave para relacionar la Fact Monitoring con la Dimensión Server	Numérico	DimServer
MetricID	Llave para relacionar la Fact Monitoring con la Dimensión Metric	Numérico	DimMetric
ValueAvg	Es el valor promedio de la métrica	Numérico	
ValueMin	Es el valor mínimo de la métrica	Numérico	
ValueMax	Es el valor máximo de la métrica	Numérico	
Threshold	Es el valor umbral máximo que puede alcanzar una métrica	Numérico	

Anexo 5 – Diseño físico y arquitectura

Diseño físico del data mart

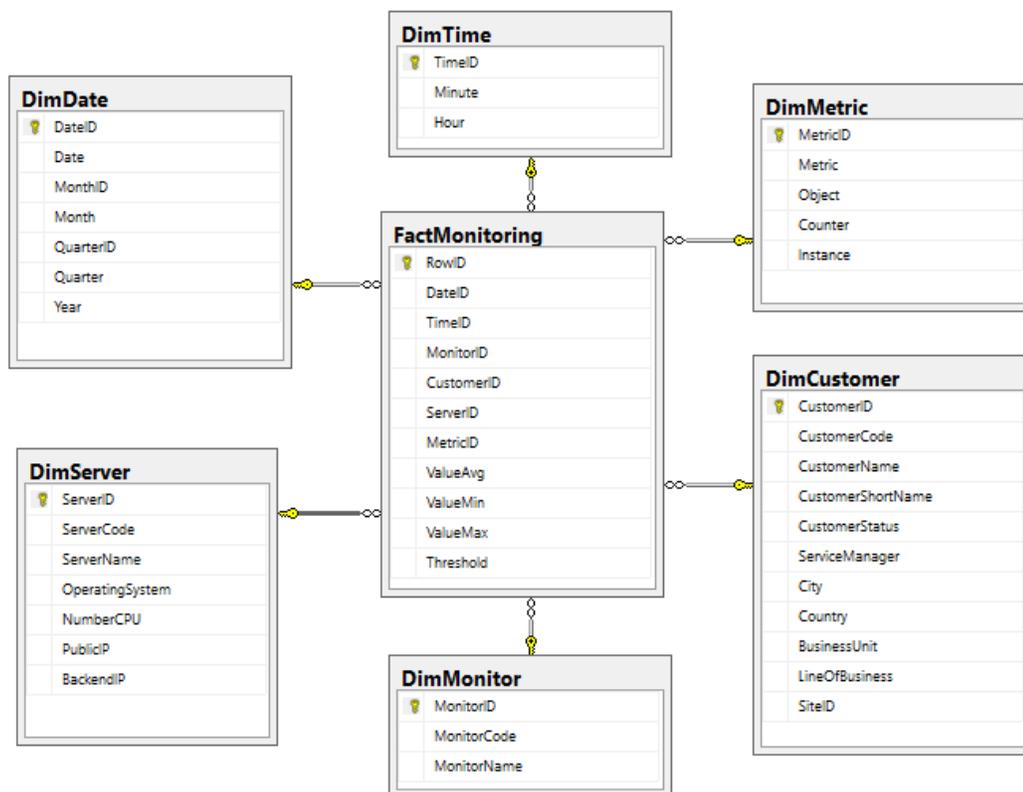


Figura 16. Diseño físico del data mart.

1. Dimensión Date: DimDate

La dimensión Date está conformada por:

Nombre de columna	Tipo de dato	Nulo	Descripción	PK	FK
DateID	int	No	Código de la fecha, es la concatenación de año, mes y día	Si	No
Date	date	No	Fecha	No	No
MonthID	int	No	Código del mes, es la concatenación de año y mes	No	No
Month	varchar(50)	No	Mes del año, es concatenación de la abreviatura del mes y el año.	No	No
QuarterID	int	No	Código del trimestre, es la concatenación de año y el número de trimestre	No	No
Quarter	Char(10)	No	Trimestre del año, es la concatenación del trimestre y el año	No	No
Year	smallint	No	Año	No	No

2. Dimensión Time: DimTime

La dimensión Time está conformada por:

Nombre de columna	Tipo de dato	Nulo	Descripción	PK	FK
TimeID	int	No	Código de la fecha, es la concatenación de año, mes y día	Si	No
Minute	Char(5)	No	Fecha	No	No
Hour	Char(5)	No	Código del mes, es la concatenación de año y mes	No	No

3. Dimensión Server: DimServer

La dimensión Server está conformada por:

Nombre de columna	Tipo de dato	Nulo	Descripción	PK	FK
ServerID	int	No	Código del servidor, es un correlativo	Si	No
ServerCode	int	No	Código del servidor en el sistema transaccional	No	No
ServerName	varchar(500)	No	Nombre del servidor	No	No
OperatingSystem	varchar(100)	No	Sistema operativo del servidor.	No	No
NumberCPU	varchar(50)	Si	Número de CPUs del servidor	No	No
PublicIP	varchar(100)	Si	Dirección IP pública del servidor	No	No
BackendIP	varchar(100)	Si	Dirección IP backend del servidor	No	No

4. Dimensión Customer: DimCustomer

La dimensión Customer está conformada por:

Nombre de columna	Tipo de dato	Nulo	Descripción	PK	FK
CustomerID	int	No	Código del cliente, es un correlativo	Si	No
CustomerCode	int	No	Código del cliente en el sistema transaccional	No	No
CustomerName	varchar(500)	No	Nombre del cliente	No	No
CustomerShortName	varchar(200)	Si	Nombre corto del cliente	No	No
CustomerStatus	varchar(50)	No	Estado del cliente	No	No
ServiceManager	varchar(500)	Si	Nombre del administrador de la cuenta	No	No
City	varchar(100)	Si	Ciudad de donde proviene el cliente	No	No
Country	varchar(100)	Si	País de donde proviene el cliente	No	No
BusinessUnit	varchar(50)	Si	Nombre de la unidad de negocio del cliente	No	No

LineOfBusiness	varchar(50)	Si	Nombre de la línea de negocio del cliente	No	No
SiteID	varchar(20)	No	Código del cliente de migración	No	No

5. Dimensión Monitor: DimMonitor

La dimensión Monitor está conformada por:

Nombre de columna	Tipo de dato	Nulo	Descripción	PK	FK
MonitorID	int	No	Código del monitor, es un correlativo	Si	No
MonitorCode	int	No	Código del monitor en el sistema transaccional	No	No
MonitorName	varchar(200)	No	Nombre del monitor	No	No

6. Dimensión Metric: DimMetric

La dimensión Metric está conformada por:

Nombre de columna	Tipo de dato	Nulo	Descripción	PK	FK
MetricID	int	No	Código de la métrica, es un correlativo	Si	No
Metric	varchar(100)	No	Nombre de la métrica	No	No
Object	varchar(500)	No	Es una subdivisión de la métrica	No	No
Counter	varchar(500)	No	Es la unidad de medida de la métrica	No	No
Instance	varchar(500)	No	Instancia en donde se aplica la métrica, por ejemplo una unidad de disco, nombre de archivo, etc.	No	No

7. Fact Monitoring: FactMonitoring

La Fact Monitoring está conformada por:

Nombre de columna	Tipo de dato	Nulo	Descripción	PK	FK
RowID	bigint	No	Código de la lectura, es un correlativo	Si	No
DateID	int	No	Llave para relacionar la Fact Monitoring con la Dimensión Date	No	Si
TimeID	int	No	Llave para relacionar la Fact Monitoring con la Dimensión Time	No	Si
MonitorID	int	No	Llave para relacionar la Fact Monitoring con la Dimensión Monitor	No	Si
CustomerID	int	No	Llave para relacionar la Fact Monitoring con la Dimensión Customer	No	Si

ServerID	int	No	Llave para relacionar la Fact Monitoring con la Dimensión Server	No	Si
MetricID	int	No	Llave para relacionar la Fact Monitoring con la Dimensión Metric	No	Si
ValueAvg	float	Si	Es el valor promedio de la métrica	No	No
ValueMin	float	Si	Es el valor mínimo de la métrica	No	No
ValueMax	float	Si	Es el valor máximo de la métrica	No	No
Threshold	float	Si	Es el valor umbral máximo que puede alcanzar una métrica	No	No

Arquitectura de la solución de Inteligencia de Negocios

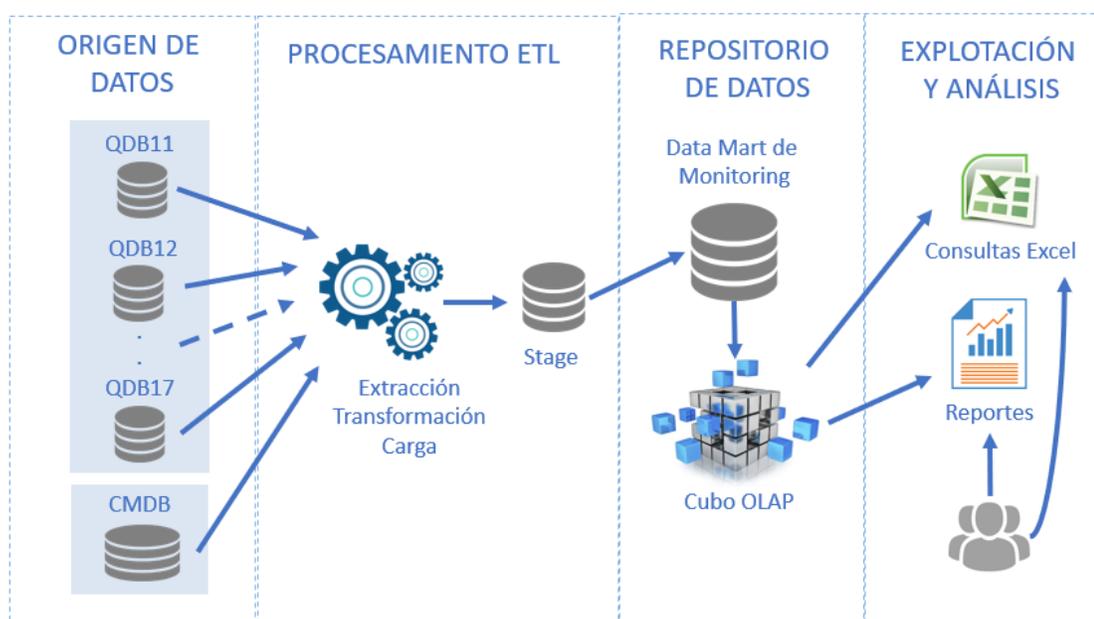


Figura 17. Arquitectura de la solución de Inteligencia de Negocios.

Anexo 6 – Diseño del sistema de extracción, transformación y carga

Para el desarrollo de los procesos de ETL se usó la herramienta de Microsoft SQL Server Integration Services. La arquitectura del proceso de ETL se muestra en la figura 18.

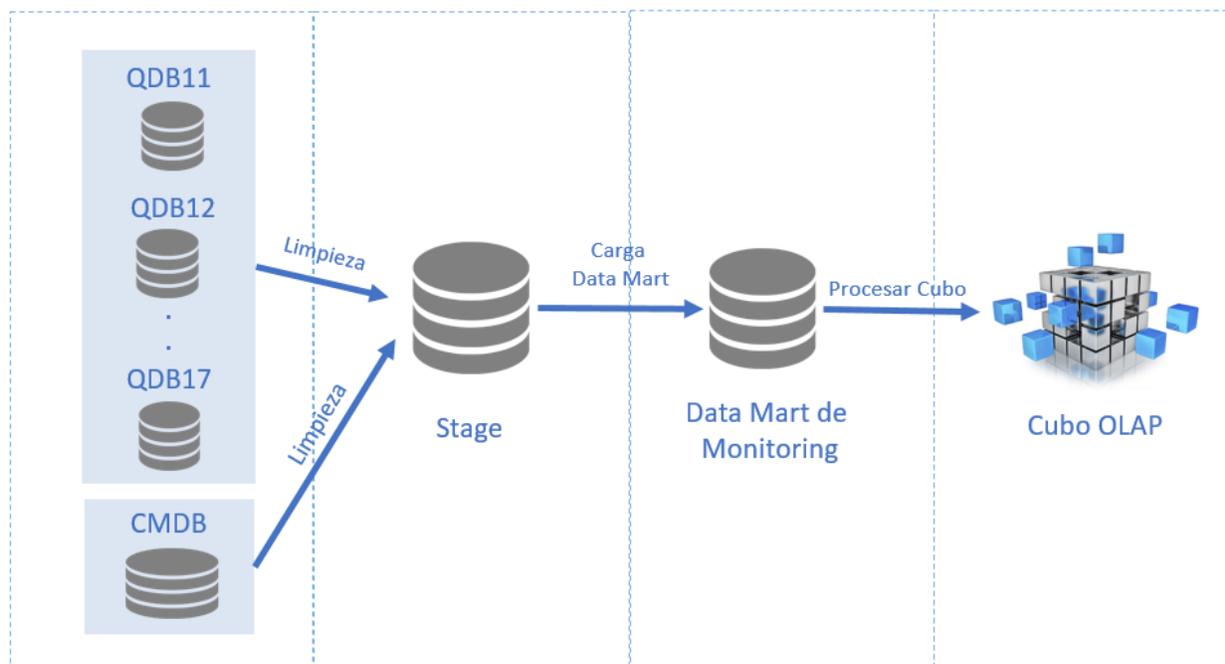


Figura 18. Arquitectura del proceso de ETL.

La arquitectura del proceso ETL contempla una base de datos Stage, y se usa como un repositorio intermedio para el proceso de extracción, transformación y carga de los datos. El diagrama de la base de datos stage se muestra en la figura 19.

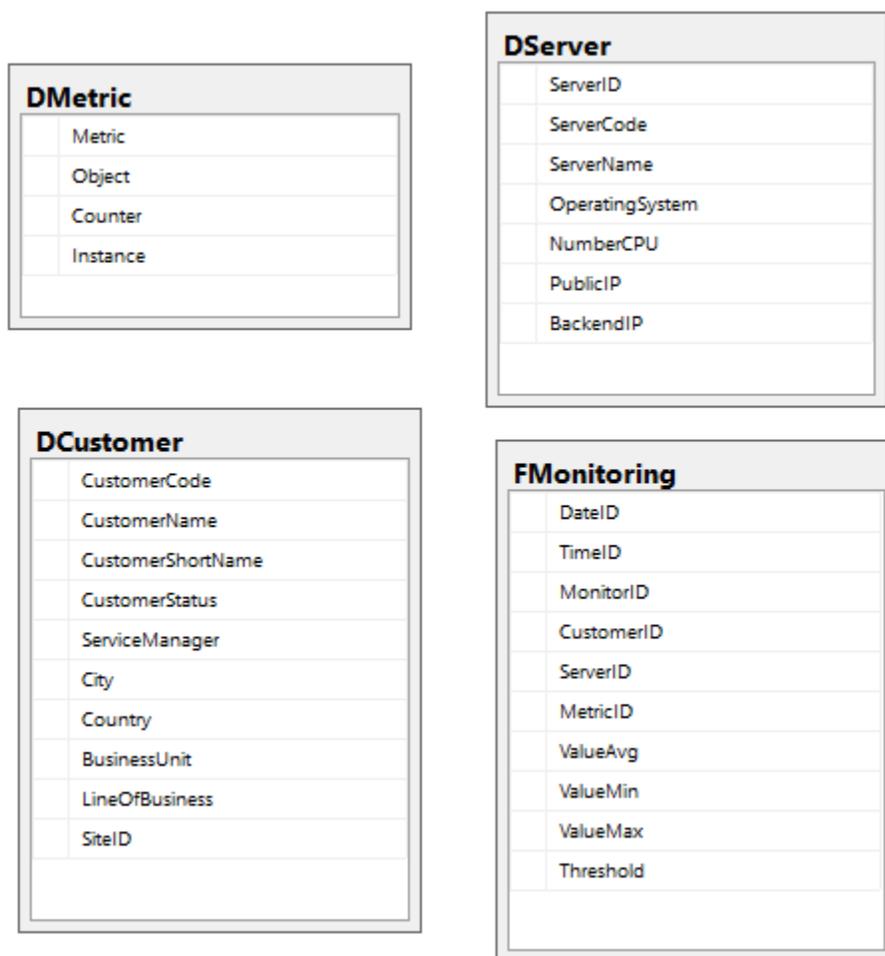


Figura 19. Diagrama de la base de datos Stage.

El proceso de ETL contempla un paquete por cada dimensión o tabla de hechos definido en el diseño del data mart, adicionalmente se han creado paquetes para la administración del proceso y para la carga de los datos desde el data mart al Cubo multidimensional. Los paquetes se muestran en la figura 20.

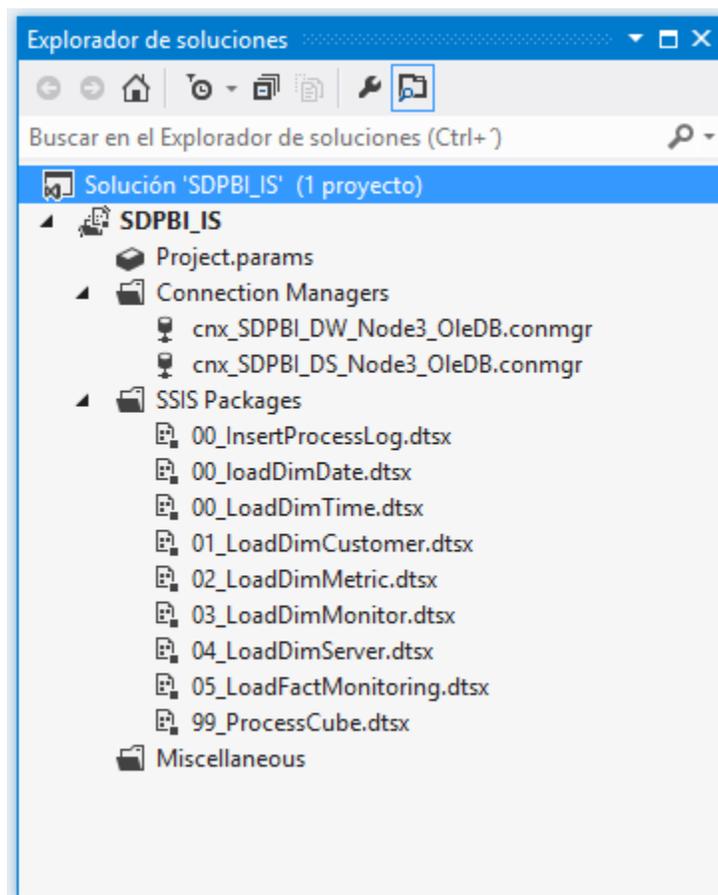


Figura 20. Paquetes del proceso de ETL.

➤ Dimensión Date

a. Descripción

El proceso consiste en cargar la dimensión Date mediante un procedimiento almacenado en la base de datos SQL Server. La frecuencia de ejecución de este proceso es solo en el momento del despliegue del sistema de BI en producción.

b. Fuentes de datos

Tipo: Procedimiento almacenado			
Base de datos: SDPBI_Stage			
Procedimiento Almacenado: usp_GetDimDate			
Descripción: Al ejecutar el procedimiento almacenado se obtendrán las fechas necesarias para el data mart			
Campo	Llave	Tipo de dato	Formato
DateID	PK	Int	Numérico
Date		Date	Fecha
MonthID		Int	Numérico
Month		Varchar(50)	Cadena
QuarterID		Int	Numérico
Quarter		Char(10)	Cadena
Year		Smallint	Numérico

c. Tabla destino

Base de datos: SDPBI_DW		
Tabla: dbo.DimDate		
Campo	Tipo de dato	Mapeo
DateID	Int	usp_GetDimDate.DateID
Date	Date	usp_GetDimDate.Date
MonthID	Int	usp_GetDimDate.MonthID
Month	Varchar(50)	usp_GetDimDate.Month
QuarterID	Int	usp_GetDimDate.QuarterID
Quarter	Char(10)	usp_GetDimDate.Quarter
Year	Smallint	usp_GetDimDate.Year

d. Descripción del paquete

El nombre del paquete es “00_loadDimDate.dtsx” y consta básicamente de dos pasos, el primero ejecuta el procedimiento almacenado y el segundo inserta el resultado en la dimensión Date, en la figura 21 se muestra el flujo de este paquete.

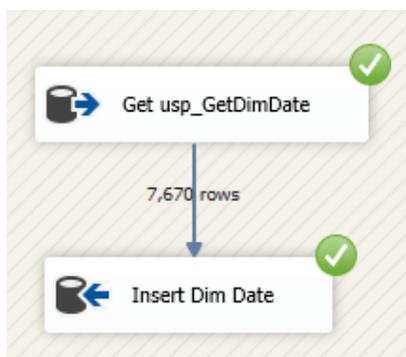


Figura 21. Paquete ETL DimDate.

➤ Dimensión Time

a. Descripción

El proceso consiste en cargar la dimensión Time mediante un procedimiento almacenado en la base de datos SQL Server. La frecuencia de ejecución de este proceso es solo en el momento del despliegue del sistema de BI en producción.

b. Fuentes de datos

Tipo: Procedimiento almacenado			
Base de datos: SDPBI_Stage			
Procedimiento Almacenado: usp_GetDimTime			
Descripción: Al ejecutar el procedimiento almacenado se obtendrán las horas y minutos necesarias para el data mart			
Campo	Llave	Tipo de dato	Formato
TimeID	PK	Int	Numérico
Minute		Char(5)	Cadena
Hour		Char(2)	Cadena

c. Tabla destino

Base de datos: SDPBI_DW		
Tabla: dbo.DimTime		
Campo	Tipo de dato	Mapeo
TimeID	Int	usp_GetDimTime.DateID
Minute	Char(5)	usp_GetDimTime.Date
Hour	Char(2)	usp_GetDimDate. MonthID

d. Descripción del paquete

El nombre del paquete es “00_LoadDimTime.dtsx” y consta básicamente de dos pasos, el primero ejecuta el procedimiento almacenado y el segundo inserta el resultado en la dimensión Time, en la figura 22 se muestra el flujo de este paquete.

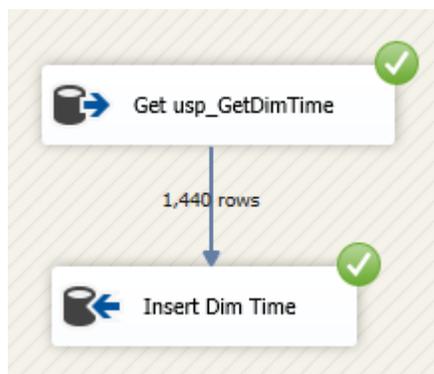


Figura 22. Paquete ETL DimTime

➤ Dimensión Customer

a. Descripción

El proceso consiste en cargar la dimensión Customer desde la tabla CI_Instance. La frecuencia de ejecución de este proceso es cada 5 minutos.

b. Fuentes de datos

Tipo: Tabla			
Base de datos: CMDB			
Tabla: dbo.CI_Instance			
Descripción: Contiene la información de muchas entidades, siendo uno de ellas la entidad cliente (Customer)			
Campo	Llave	Tipo de dato	Formato
InstanceID	PK	Bigint	Numérico
Name		Nvarchar(250)	Cadena
Status		Nvarchar(100)	Cadena

Tipo: Tabla			
Base de datos: CMDB			
Tabla: dbo.CI_InstanceAttributeValue			
Descripción: Contiene la información de atributos de muchas entidades, siendo uno de ellas la entidad cliente (Customer)			
Campo	Llave	Tipo de dato	Formato
InstanceAttributeValueID	PK	Bigint	Numérico
InstanceID		Bigint	Numérico
ValueString		Nvarchar(max)	Cadena

c. Tabla destino

Base de datos: SDPBI_DW		
Tabla: dbo.DimCustomer		
Campo	Tipo de dato	Mapeo
CustomerID	Int	Autogenerado
CustomerCode	Int	CI_Instance.InstanceID
CustomerName	Varchar(500)	CI_Instance.Name
CustomerShortName	Varchar(200)	CI_InstanceAttributeValue.ValueString
CustomerStatus	Varchar(50)	CI_Instance.Status
ServiceManager	Varchar(500)	CI_InstanceAttributeValue.ValueString
City	Varchar(100)	CI_InstanceAttributeValue.ValueString
Country	Varchar(100)	CI_InstanceAttributeValue.ValueString
BusinessUnit	Varchar(50)	CI_InstanceAttributeValue.ValueString
LineOfBusiness	Varchar(50)	CI_InstanceAttributeValue.ValueString
SiteID	Varchar(20)	CI_InstanceAttributeValue.ValueString

d. Descripción del paquete

El nombre del paquete es “01_LoadDimCustomer.dtsx” y básicamente consiste en cargar el código y el nombre del Customer desde la tabla CI_Instance. Adicionalmente se extraen los demás campos de una tabla maestra de atributos llamada CI_InstanceAttributeValue, en donde cada atributo tiene un código y la relación con un elemento de la tabla CI_Instance. Después de realizar la limpieza de los datos, estos se insertan en una tabla Stage y posteriormente

se insertan en la dimensión (DimCustomer). La carga es incremental, si el cliente existe ya no se carga, en la figura 23 se muestra el flujo de este paquete.

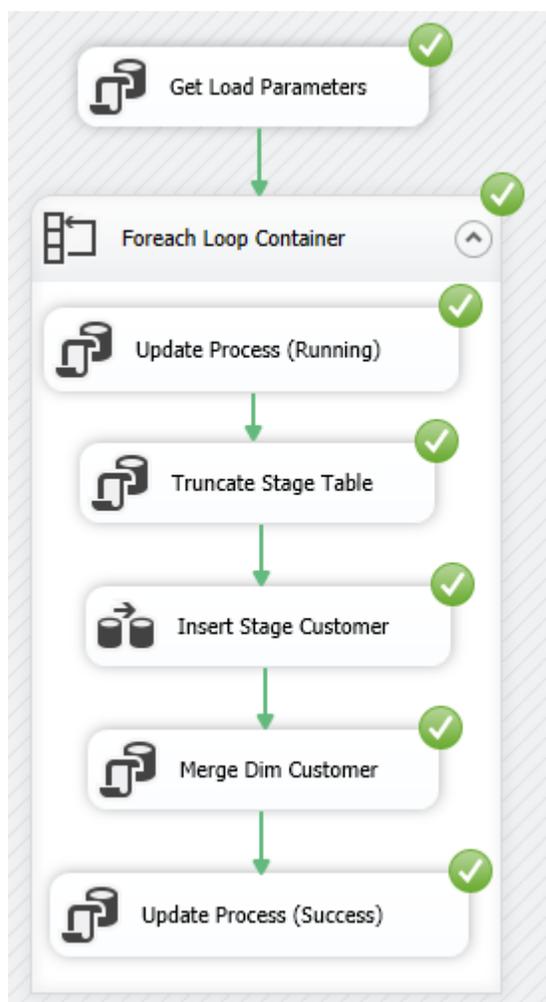


Figura 23. Paquete ETL DimCustomer.

➤ Dimensión Metric

a. Descripción

El proceso consiste en cargar la dimensión Metric desde la tabla ArchiveDataHeader, existe un campo llamado Legend de donde se puede extraer los campos necesarios para la dimensión. La frecuencia de ejecución de este proceso es cada 5 minutos.

b. Fuentes de datos

Tipo: Tabla			
Base de datos: QDB11, QDB12, QDB13, QDB14, QDB15, QDB16, QDB17			
Tabla: dbo.ArchiveDataHeader			
Descripción: Contiene la información de todas las métricas reportadas por los monitores			
Campo	Llave	Tipo de dato	Formato
Legend		Nvarchar(128)	Cadena

c. Tabla destino

Base de datos: SDPBI_DW		
Tabla: dbo.DimMetric		
Campo	Tipo de dato	Mapeo
MetricID	Int	Autogenerado
Metric	Varchar(100)	ArchiveDataHeader.Legend
Object	Varchar(500)	ArchiveDataHeader.Legend
Counter	Varchar(500)	ArchiveDataHeader.Legend
Instance	Varchar(500)	ArchiveDataHeader.Legend

d. Descripción del paquete

El nombre del paquete es “02_LoadDimMetric.dtsx” y básicamente consiste en cargar los datos desde la tabla dbo.ArchiveDataHeader. La información se extrae de cada una de las 7 instancias de monitoreo. Después de realizar la limpieza de los datos, estos se insertan en una tabla Stage y posteriormente se insertan en la dimensión (DimMetric). La carga es incremental, si la métrica existe ya no se carga, en la figura 24 se muestra el flujo de este paquete.

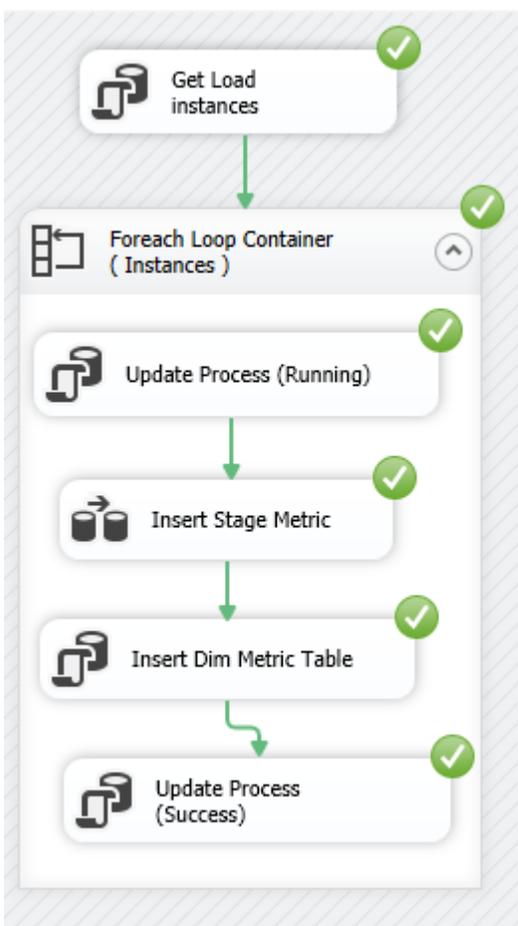


Figura 24. Paquete ETL DimMetric.

➤ Dimensión Monitor

a. Descripción

El proceso consiste en cargar la dimensión Monitor desde la tabla JobQDB_t. La frecuencia de ejecución de este proceso es cada 5 minutos.

b. Fuentes de datos

Tipo: Tabla			
Base de datos: PortalMonitorDB			
Tabla: dbo.JobQDB_t			
Descripción: Contiene la información de todos los monitores (servidor y base de datos)			
Campo	Llave	Tipo de dato	Formato
QDBID	PK	Bigint	Numérico
Name		Nvarchar(250)	Cadena

c. Tabla destino

Base de datos: SDPBI_DW		
Tabla: dbo.DimMonitor		
Campo	Tipo de dato	Mapeo
MonitorID	Int	Autogenerado
MonitorCode	Int	JobQDB_t.MonitorID
MonitorName	Varchar(500)	JobQDB_t.MonitorName

d. Descripción del paquete

El nombre del paquete es “03_LoadDimMonitor.dtsx” y básicamente consiste en cargar el código y el nombre del Monitor desde la tabla JobQDB_t, el campo nombre está compuesto por el nombre del servidor, el nombre de la instancia y el nombre de la base de datos. Después de realizar la limpieza de los datos, estos se insertan en una tabla Stage y posteriormente se insertan en la dimensión (DimMonitor). La carga es incremental, si el monitor existe ya no se carga, en la figura 25 se muestra el flujo de este paquete.

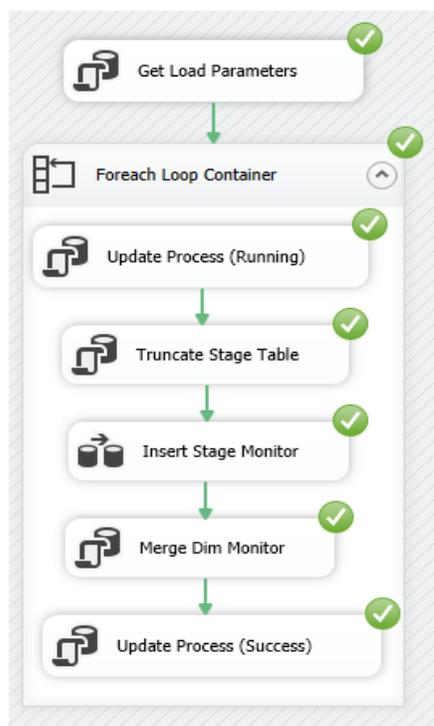


Figura 25. Paquete ETL DimMonitor.

➤ Dimensión Server

a. Descripción

El proceso consiste en cargar la dimensión Server desde las tablas ServerMaster y CI_Instance. La frecuencia de ejecución de este proceso es cada 5 minutos.

b. Fuentes de datos

Tipo: Tabla			
Base de datos: QDB11, QDB12, QDB13, QDB14, QDB15, QDB16, QDB17			
Tabla: dbo.ServerMaster			
Descripción: Contiene la información de todas los servidores configurados en el data center.			
Campo	Llave	Tipo de dato	Formato
ServerName		Nvarchar(500)	Cadena

Tipo: Tabla			
Base de datos: CMDB			
Tabla: dbo.CI_Instance			
Descripción: Contiene la información de atributos de muchas entidades, siendo uno de ellas la entidad servidor (server)			
Campo	Llave	Tipo de dato	Formato
InstanceID	PK	Bigint	Numérico
Name		Nvarchar(250)	Cadena

Tipo: Tabla			
Base de datos: CMDB			
Tabla: dbo.CI_InstanceAttributeValue			
Descripción: Contiene la información de atributos de muchas entidades, siendo uno de ellas la entidad cliente (Customer)			
Campo	Llave	Tipo de dato	Formato
InstanceAttributeValueID	PK	Bigint	Numérico
InstanceID		Bigint	Numérico
ValueString		Nvarchar(max)	Cadena

c. Tabla destino

Base de datos: SDPBI_DW		
Tabla: dbo.DimServer		
Campo	Tipo de dato	Mapeo
ServerID	Int	Autogenerado
ServerCode	Int	CI_Instance.InstanceID
ServerName	Varchar(500)	CI_Instance.Name
OperatingSystem	Varchar(100)	CI_InstanceAttributeValue.ValueString
NumberCPU	Varchar(50)	CI_InstanceAttributeValue.ValueString
PublicIP	Varchar(100)	CI_InstanceAttributeValue.ValueString
BackendIP	Varchar(100)	CI_InstanceAttributeValue.ValueString

d. Descripción del paquete

El nombre del paquete es “04_LoadDimServer.dtsx” y básicamente consiste en cargar el código y el nombre del servidor desde la tabla CI_Instance. Adicionalmente se cargan los demás campos desde una tabla maestra de atributos llamada CI_InstanceAttributeValue, en donde cada atributo tiene un código y la relación con un elemento de la tabla CI_Instance. Después de realizar la limpieza de los datos, estos se insertan en una tabla Stage y posteriormente se insertan en la dimensión (DimServer). La carga es incremental, si el servidor existe ya no se carga, en la figura 26 se muestra el flujo de este paquete.

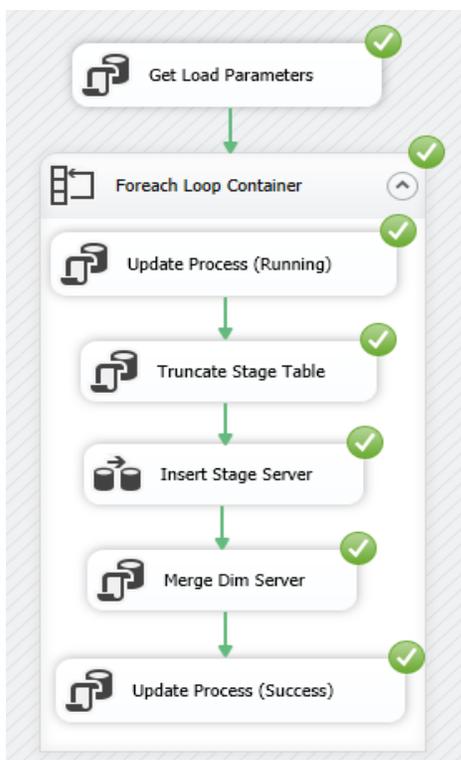


Figura 26. Paquete ETL DimServer.

➤ Fact Monitoring

a. Descripción

El proceso consiste en cargar la tabla de hechos Fact Monitoring desde las tablas ArchiveData and ArchiveDataHeader. La frecuencia de ejecución de este proceso es cada 5 minutos.

b. Fuentes de datos

Tipo: Tabla			
Base de datos: QDB11, QDB12, QDB13, QDB14, QDB15, QDB16, QDB17			
Tabla: dbo.ArchiveData			
Descripción: Contiene los datos de las lecturas de las métricas de monitoreo de servidores.			
Campo	Llave	Tipo de dato	Formato
DataID	PK	Bigint	Numérico
PerformanceDate	PK	Datetime	Fecha
Value		Float	Cadena

Tipo: Tabla			
Base de datos: QDB11, QDB12, QDB13, QDB14, QDB15, QDB16, QDB17			
Tabla: dbo.ArchiveDataHeader			
Descripción: Contiene el detalle de los datos de las lecturas de las métricas de monitoreo de servidores, es la cabecera de los datos.			
Campo	Llave	Tipo de dato	Formato
DataID	PK	Bigint	Numérico
MachineName		Varchar(128)	Cadena
KPName		Varchar(256)	Cadena
Legend		Varchar(128)	Cadena

Tipo: Tabla			
Base de datos: PortalMonitorDB			
Tabla: dbo.Job_t			
Descripción: Contiene la información de la configuración de servidores.			
Campo	Llave	Tipo de dato	Formato
JobID	PK	Int	Numérico
JobName		Varchar(128)	Cadena

Tipo: Tabla			
Base de datos: PortalMonitorDB			
Tabla: dbo.JobParameterType_t			
Descripción: Contiene la información de los tipos de parámetros de configuración de servidores.			
Campo	Llave	Tipo de dato	Formato
JobParameterTypeID	PK	Int	Numérico
JobParameterType		Varchar(128)	Cadena

Tipo: Tabla			
Base de datos: PortalMonitorDB			
Tabla: dbo.JobParameter_t			
Descripción: Contiene la información de los parámetros de configuración de servidores.			
Campo	Llave	Tipo de dato	Formato
JobParameterID	PK	Int	Numérico
JobID	FK	Int	Numérico
JobParameterTypeID	FK	Int	Numérico
JobParameterValue		Varchar(1000)	Cadena

Tipo: Tabla			
Base de datos: CMDB			
Tabla: dbo.CI_Instance			
Descripción: Contiene la información de clientes (Customer) y servidores (Server)			
Campo	Llave	Tipo de dato	Formato
InstanceID	PK	Int	Numérico

c. Tabla destino

Base de datos: SDPBI_DW		
Tabla: dbo.FactMonitoring		
Campo	Tipo de dato	Mapeo
RowID	Bigint	Autogenerado
DateID	Int	DimDate.DateID
TimeID	Int	DimDate.TimeID
MonitorID	Int	DimDate.MonitorID
CustomerID	Int	DimDate.CustomerID
ServerID	Int	DimDate.ServerID
MetricID	Int	DimDate.MetricID
ValueAvg	Float	ArchiveData.Value
ValueMin	Float	ArchiveData.Value
ValueMax	Float	ArchiveData.Value
Threshold	Float	JobParameter_t.JobParameterValue

d. Descripción del paquete

El nombre del paquete es “05_LoadFactMonitoring.dtsx” y básicamente consiste en cargar los datos de lecturas de métricas de los servidores desde las tablas ArchiveData y ArchiveDataHeader. La información se extrae de cada una de las 7 instancias de monitoreo. Después de realizar la limpieza de los datos, estos se insertan en una tabla stage y posteriormente se insertan en la tabla de hechos (FactMonitoring). La carga es incremental de acuerdo a un rango de fechas, si el proceso falla, en la siguiente ejecución se cargarán los datos desde la última fecha de procesamiento, en la figura 27 se muestra el flujo de este paquete.



Figura 27. Paquete ETL FactMonitoring.

➤ **Procesar Cubo**

La frecuencia de ejecución de este proceso es cada 5 minutos. El nombre del paquete es “99_ProcessCube.dtsx” y básicamente consiste en procesar el cubo multidimensional, en la figura 28 se muestra el flujo de este paquete.

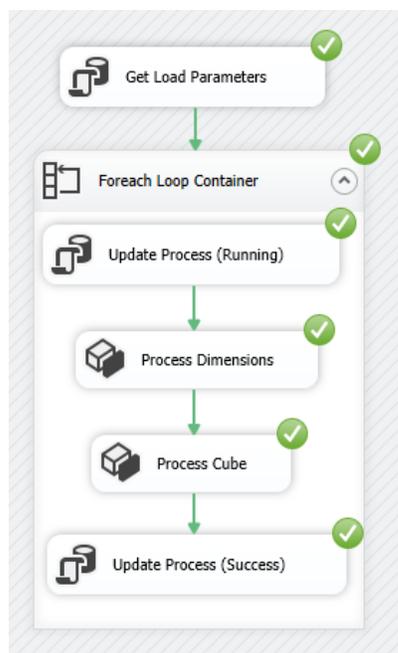


Figura 28. Paquete ETL procesar cubo.

➤ Registrar proceso

Este paquete se encarga de inicializar la ejecución de todos los demás paquetes, realiza el cálculo de los rangos de fecha de carga por cada paquete e inserta registros para el control del proceso. El nombre del paquete es “00_InsertProcessLog.dtsx”, en la figura 29 se muestra el flujo de este paquete.

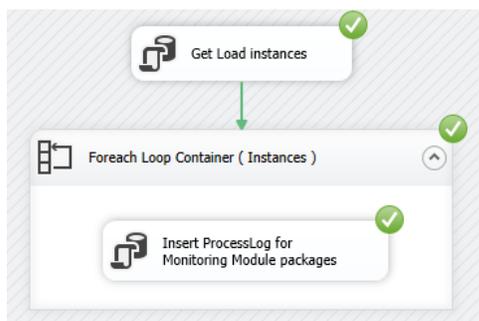


Figura 29. Paquete ETL registrar proceso.

Anexo 6 – Implementación

Construcción del cubo multidimensional

Para la construcción de cubo multidimensional se usó la herramienta de Microsoft SQL Server Analysis Services. En la figura 30 se muestra el proyecto del cubo OLAP.

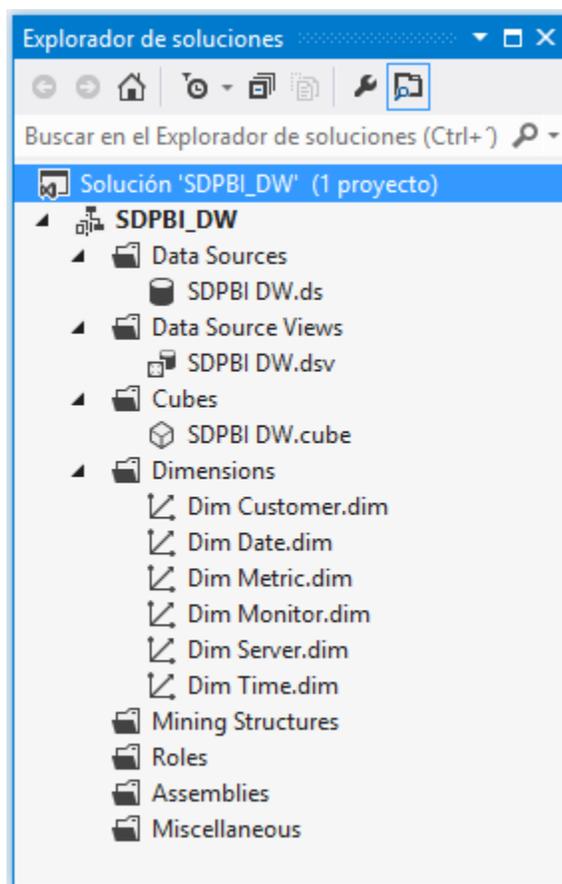


Figura 30. Proyecto del cubo OLAP.

➤ Fuente de datos

La fuente de datos es la base de datos “SDPBI_DW”, en donde se encuentra el modelo dimensional que consta de 6 dimensiones y una tabla de hechos, en la figura 31 se muestra la configuración de la fuente de datos en la herramienta.

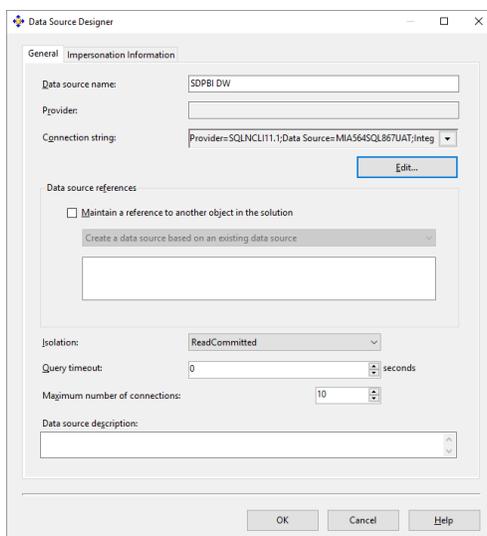


Figura 31. Fuente de datos.

➤ Vista de la fuente de datos

En la vista de la fuente de datos se puede apreciar todas las tablas del modelo dimensional que servirán como base para la construcción del cubo OLAP, en la figura 32 se muestra la vista de la fuente de datos en la herramienta.

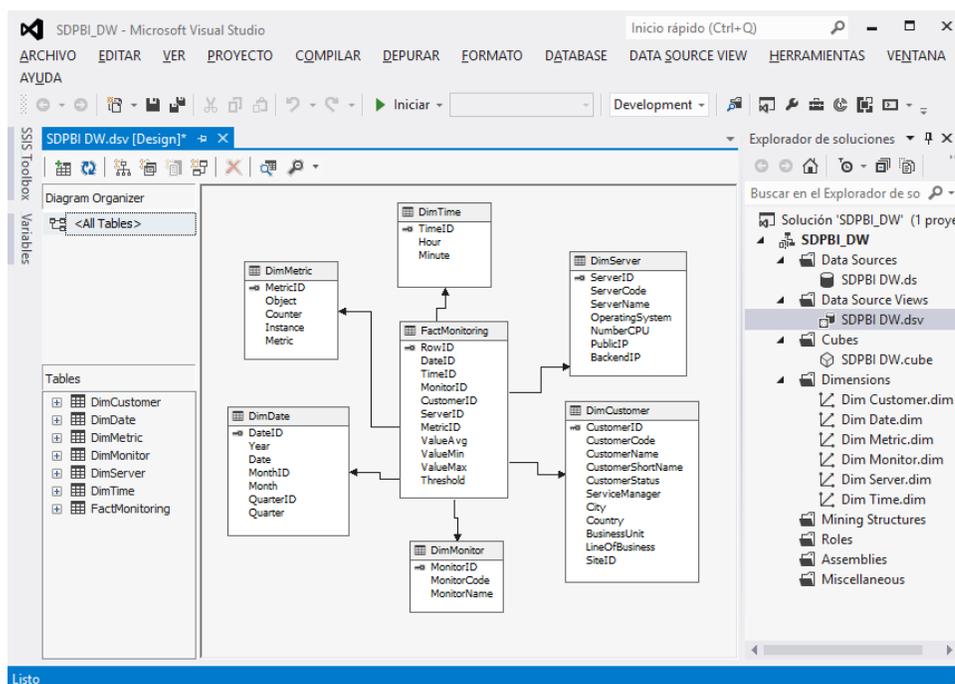


Figura 32. Vista de la fuente de datos.

➤ Cubo OLAP

El cubo OLAP fue construido a partir de la tabla FactMonitoring y las 6 dimensiones que se encuentran en el data mart. El cubo consta de 4 métricas agrupadas en un grupo de métricas denominado “Monitoring”, en la figura 33 se muestra la estructura del cubo en la herramienta.

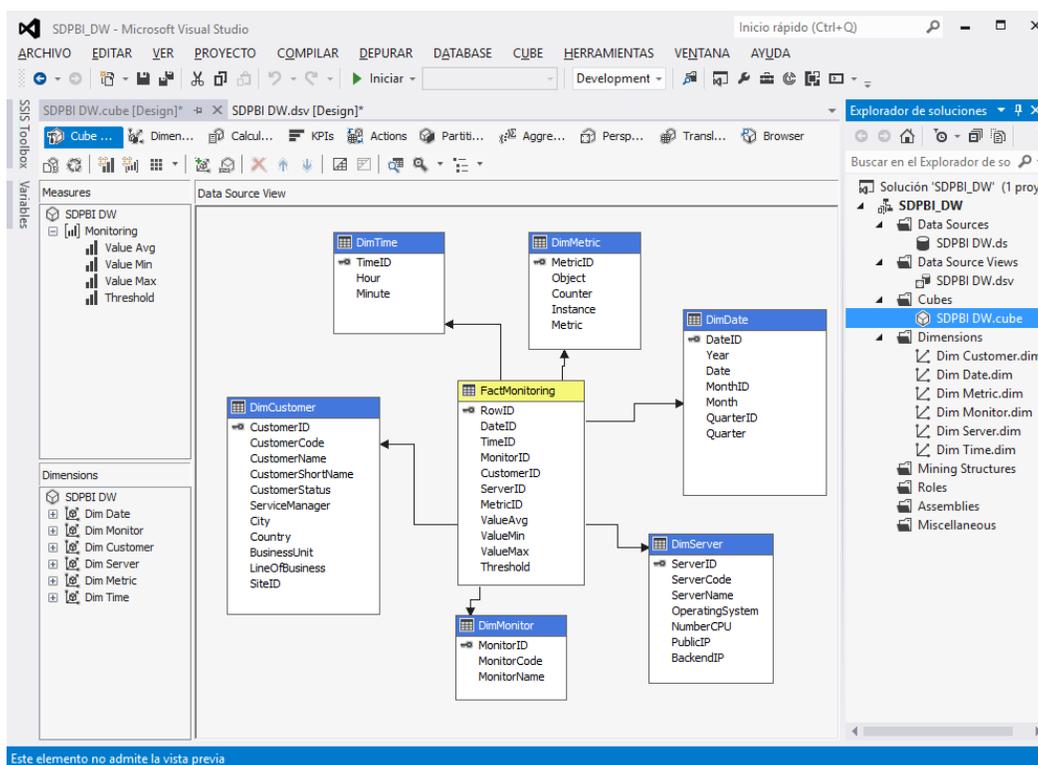


Figura 33. Cubo OLAP.

➤ Dimensiones

Las dimensiones fueron creadas de acuerdo a las especificaciones de los requerimientos de negocio, siendo un total de 6 dimensiones.

a. Dimensión Customer

Esta dimensión fue construida a partir de la tabla DimCustomer que se encuentra en el data mart, en la figura 34 se muestra la estructura de la dimensión Customer.

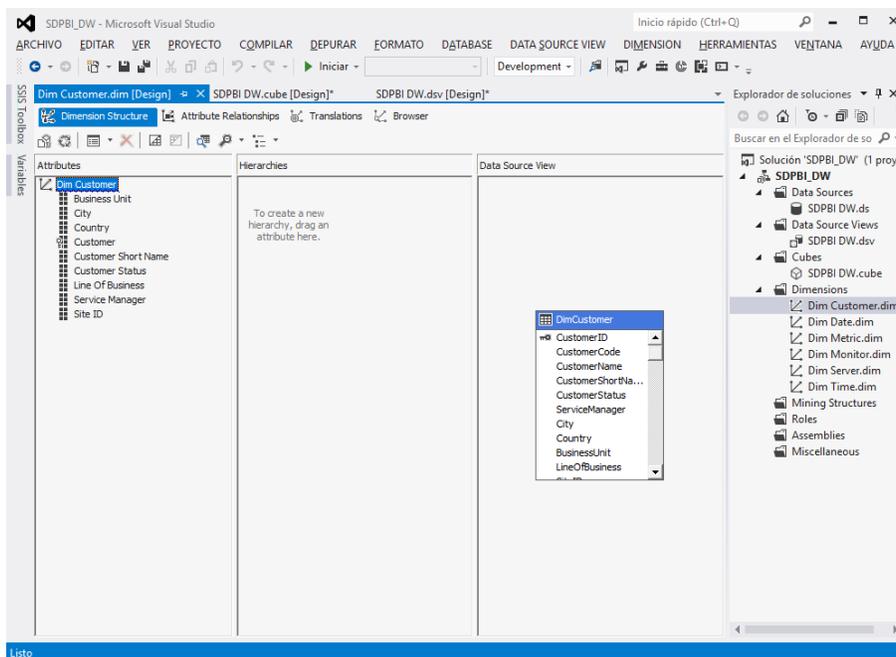


Figura 34. Estructura Dim Customer.

b. Dimensión Date

Esta dimensión fue construida a partir de la tabla DimDate que se encuentra en el data mart, en la figura 35 se muestra la estructura de la dimensión Date.

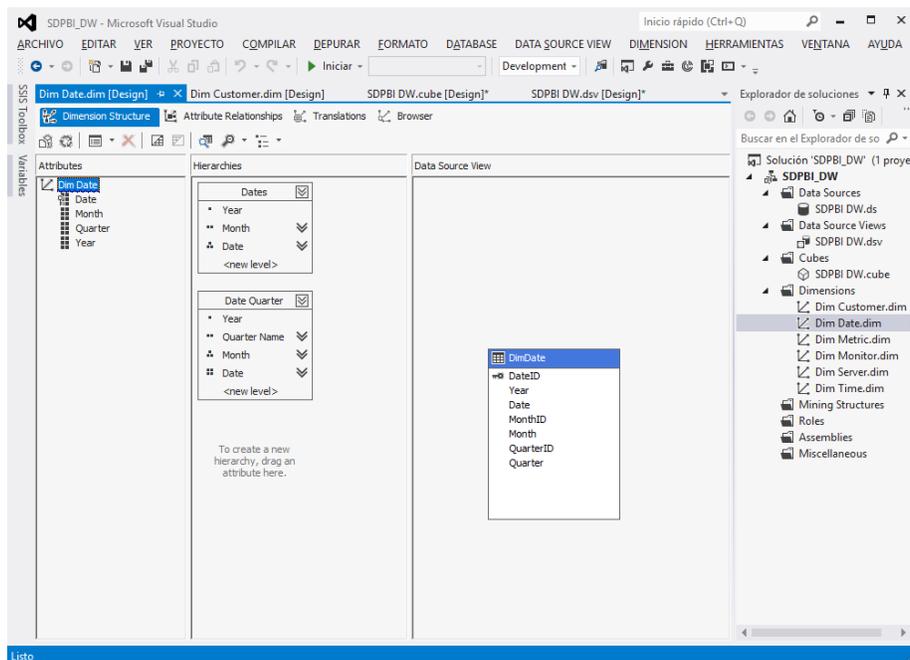


Figura 35. Estructura Dim Date.

c. Dimensión Metric

Esta dimensión fue construida a partir de la tabla DimMetric que se encuentra en el data mart, en la figura 36 se muestra la estructura de la dimensión Metric.

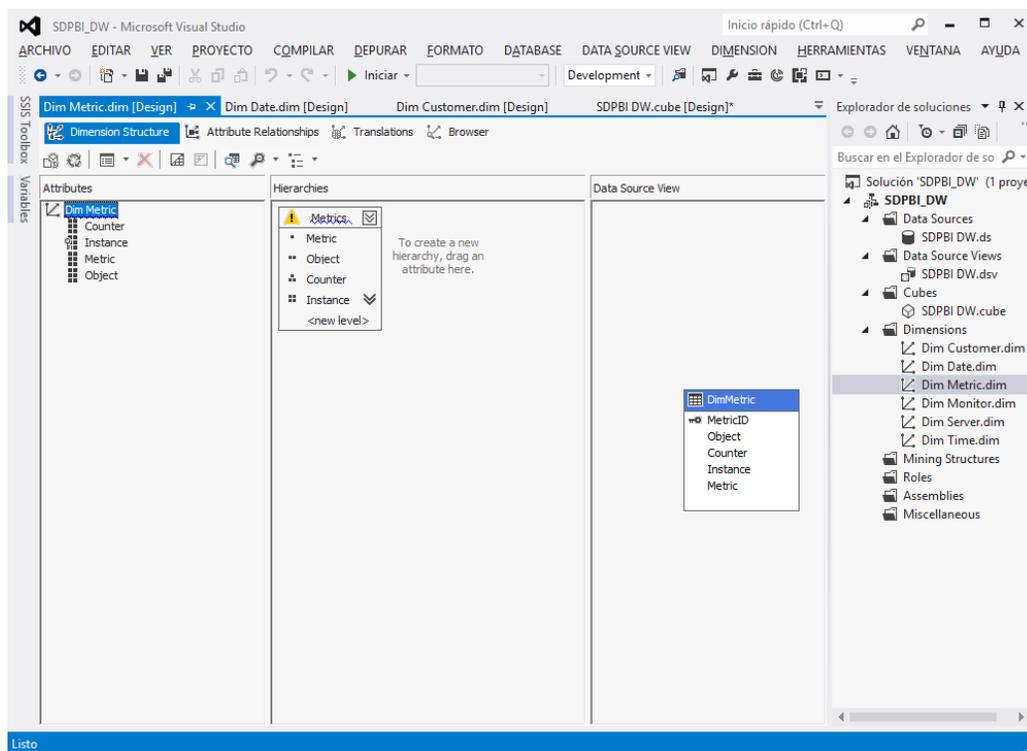


Figura 36. Estructura Dim Metric.

d. Dimensión Monitor

Esta dimensión fue construida a partir de la tabla DimMonitor que se encuentra en el data mart, en la figura 37 se muestra la estructura de la dimensión Monitor.

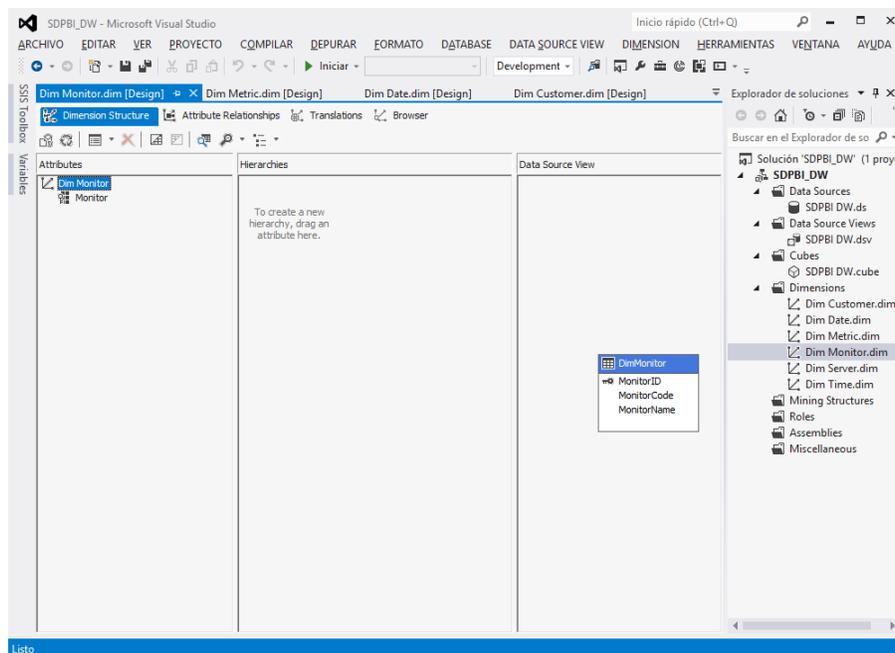


Figura 37. Estructura Dim Monitor.

e. Dimensión Server

Esta dimensión fue construida a partir de la tabla DimServer que se encuentra en el data mart, en la figura 38 se muestra la estructura de la dimensión Server.

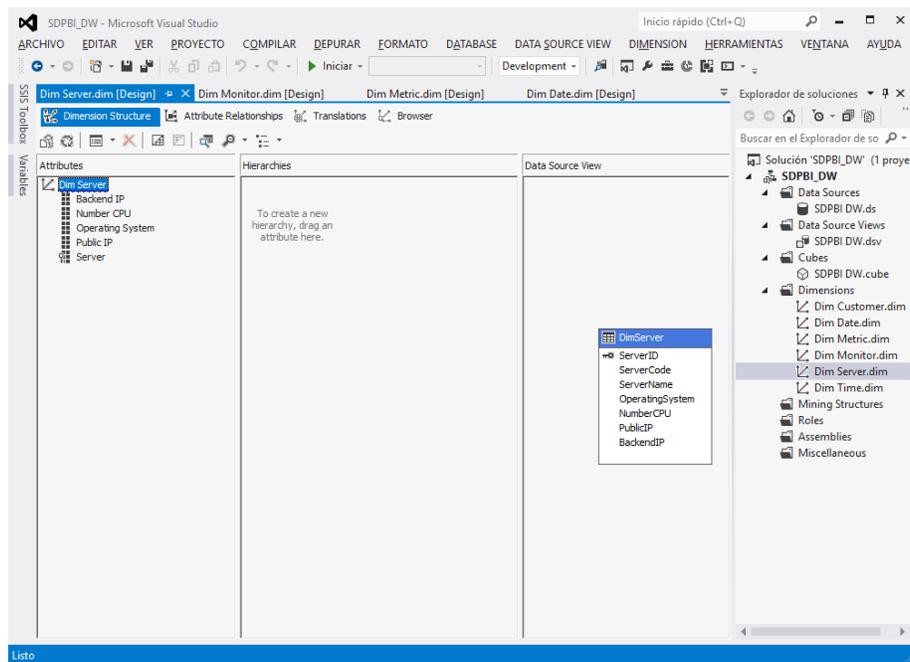


Figura 38. Estructura Dim Server.

f. Dimensión Time

Esta dimensión fue construida a partir de la tabla DimTime que se encuentra en el data mart.

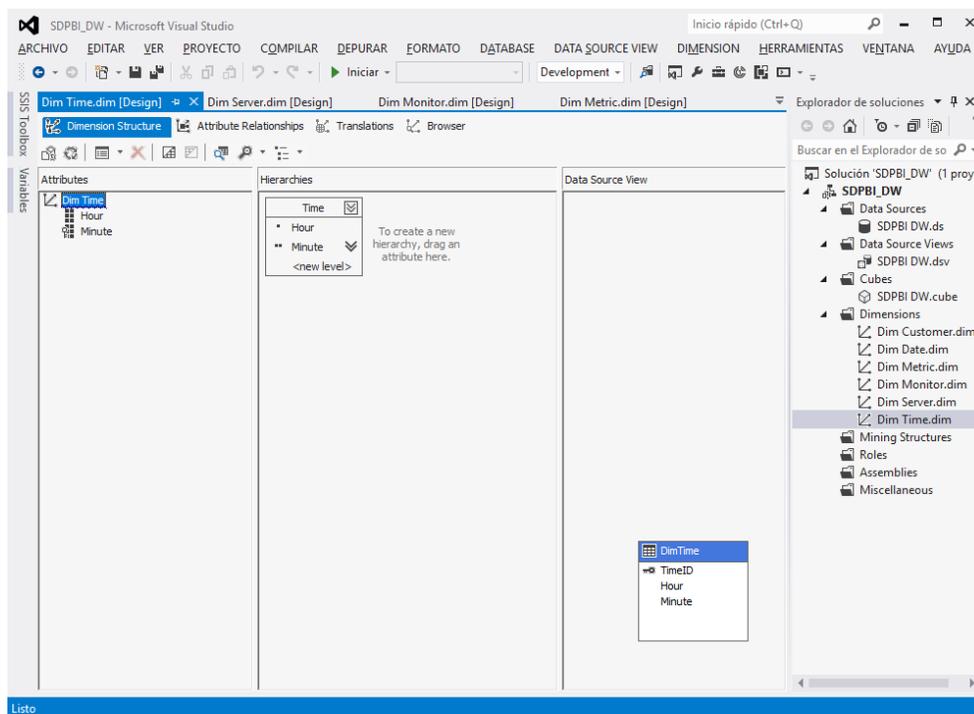


Figura 39. Estructura Dim Time.

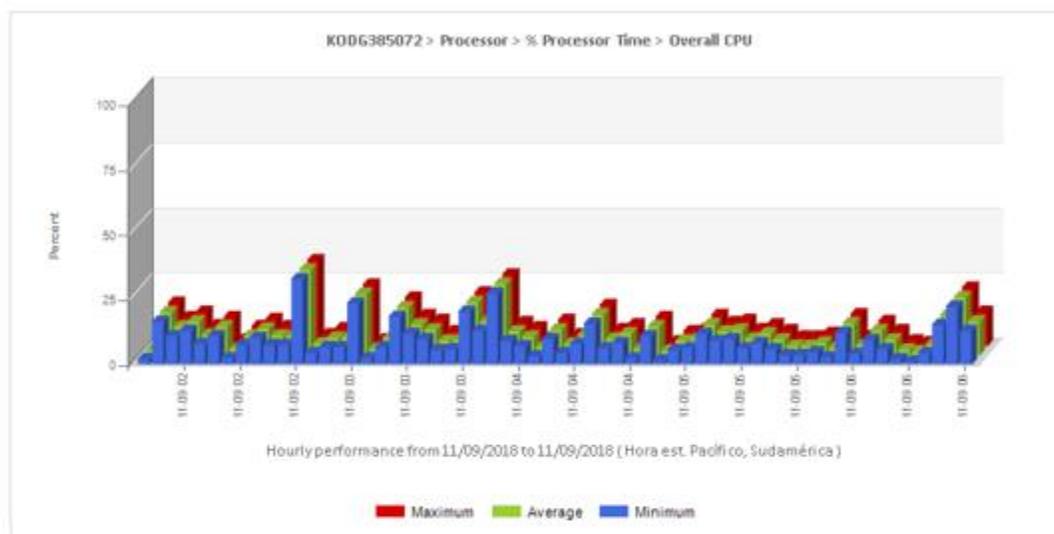
Construcción de reportes

A continuación, se muestran algunos reportes creados a partir del cubo OLAP desarrollado.



Figura 40. Reporte de estado de servidores.

Metric: KODG385072 > Processor > % Processor Time > Overall CPU
 Last Data Point Collection Time: 11/09/2018 06:53 Hora est. Pacifico, Sudamérica
 Data Point Value: [Avg = 12,9] [Min = 12,9] [Max = 12,9] %
 Maximum CPU usage (%) threshold: 99 %



Hourly data points for the past 3 days:

Entity	Date / Time	Minimum	Average	Maximum
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	12,894	12,894	12,894
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	22,269	22,269	22,269
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	15,238	15,238	15,238
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	4,301	4,301	4,301
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	1,176	1,176	1,176
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	1,957	1,957	1,957
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	5,863	5,863	5,863
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	9,378	9,378	9,378
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	3,910	3,910	3,910
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	12,113	12,113	12,113
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-06	2,738	2,738	2,738
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-05	4,691	4,691	4,691
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-05	3,519	3,519	3,519
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-05	3,519	3,519	3,519
KODG385072 > Processor > % Processor Time > Overall CPU	11-09-05	5,863	5,863	5,863

Figura 41. Reporte de estado de servidores detallado últimas 5 horas.

Metric: DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU
 Last Data Point Collection Time: 10/03/2018 11:00 Hora est. Pacifico, Sudamérica
 Data Point Value: [Avg = 37,8] [Min = 32,7] [Max = 44,2] %
 Maximum CPU usage (%) threshold: 99 %



Hourly data points for the past 3 days:

Entity	Date / Time	Minimum	Average	Maximum
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-11	32,692	37,756	44,231
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-10	32,692	37,147	41,923
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-09	32,692	37,115	45,000
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-08	32,691	37,275	46,923
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-07	31,923	38,365	44,615
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-06	31,535	36,730	42,308
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-05	31,154	36,250	43,462
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-04	31,154	37,403	51,923
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-03	29,988	49,101	70,000
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-02	31,525	32,431	34,231
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-01	31,538	32,212	33,846
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-03-12	30,769	32,532	37,308
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-02-11	30,769	31,731	34,615
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-02-10	30,769	32,308	35,000
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-02-09	31,154	32,404	36,538
DCA2-NQRM-41 > Processor > % Processor Time > Overall CPU	10-02-08	30,769	34,904	53,846

Figura 42. Reporte de estado de servidores detallado últimos 3 días.

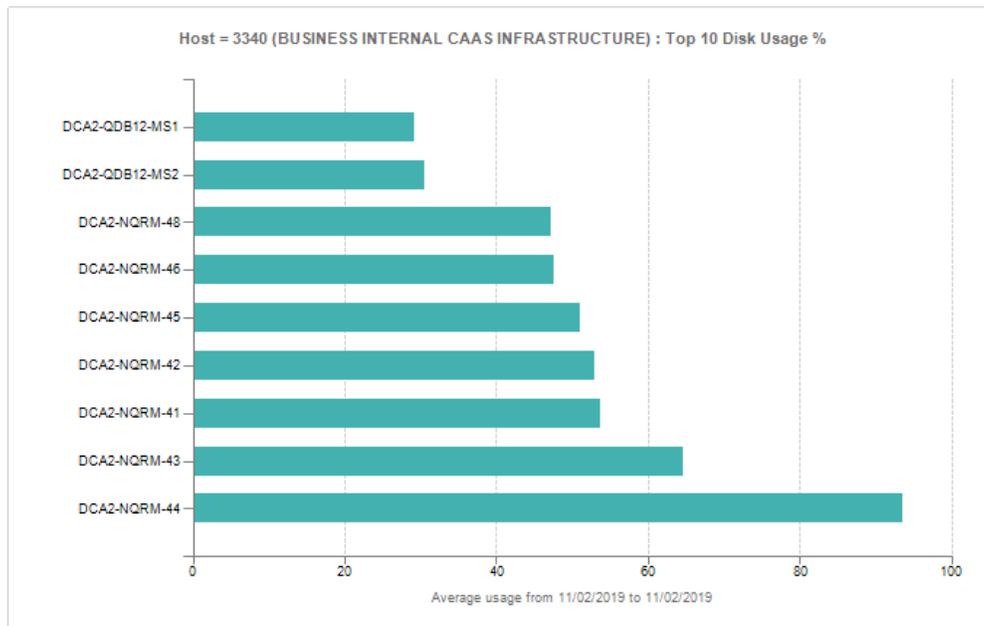


Figura 43. Reporte Top N del porcentaje de uso de Disco.

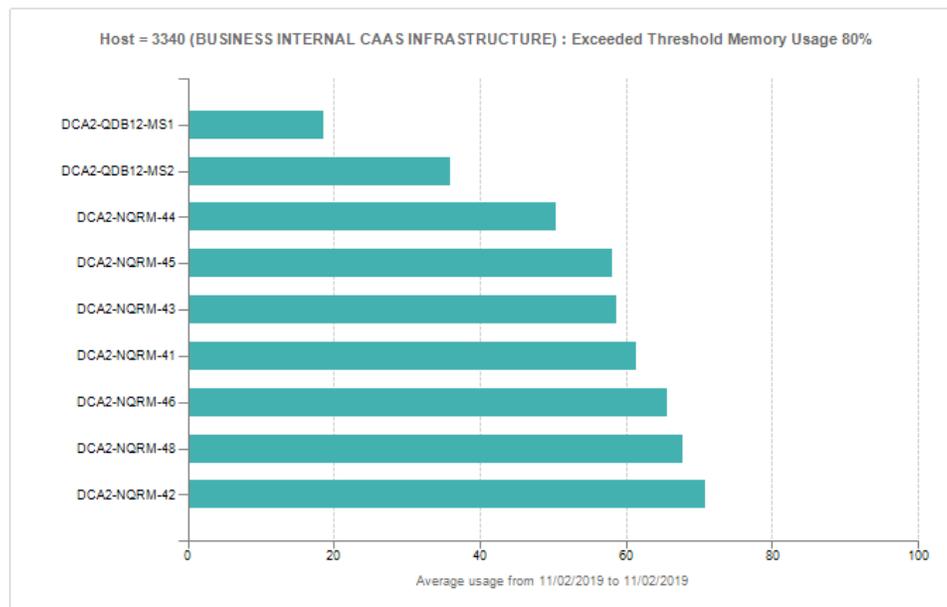


Figura 44. Reporte de umbral excedido del porcentaje de uso de Memoria.

Anexo 8 – Modelo de cuestionario

1. Cuestionario para obtener el tiempo de generación de reportes dirigido a los analistas de datos.

Indique cuanto tiempo tomó la generación de cada reporte si es que aplica		En minutos
N°	Pregunta	
P.1	¿Cuánto tiempo le tomó la generación del reporte “Exceeded Threshold Disk Usage”	
P.2	¿Cuánto tiempo le tomó la generación del reporte “Exceeded Threshold Logical Disk Usage”	
P.3	¿Cuánto tiempo le tomó la generación del reporte “Exceeded Threshold Memory Usage”	
P.4	¿Cuánto tiempo le tomó la generación del reporte “Exceeded Threshold Processor Usage”	
P.5	¿Cuánto tiempo le tomó la generación del reporte “Exceeded Threshold Volume Usage”	
P.6	¿Cuánto tiempo le tomó la generación del reporte “TopN Disk Usage”	
P.7	¿Cuánto tiempo le tomó la generación del reporte “TopN Logical Disk Usage”	
P.8	¿Cuánto tiempo le tomó la generación del reporte “TopN Memory Usage”	
P.9	¿Cuánto tiempo le tomó la generación del reporte “TopN Processor Usage”	
P.10	¿Cuánto tiempo le tomó la generación del reporte “TopN Process Volume Usage”	

2. Cuestionario para obtener el tiempo de análisis de reportes dirigido a los analistas de monitoreo.

Indique cuanto tiempo tomó el análisis de cada reporte si es que aplica		En minutos
N°	Pregunta	
P.1	¿Cuánto tiempo le tomó el análisis del reporte “Exceeded Threshold Disk Usage”	
P.2	¿Cuánto tiempo le tomó el análisis del reporte “Exceeded Threshold Logical Disk Usage”	
P.3	¿Cuánto tiempo le tomó el análisis del reporte “Exceeded Threshold Memory Usage”	
P.4	¿Cuánto tiempo le tomó el análisis del reporte “Exceeded Threshold Processor Usage”	
P.5	¿Cuánto tiempo le tomó el análisis del reporte “Exceeded Threshold Volume Usage”	
P.6	¿Cuánto tiempo le tomó el análisis del reporte “TopN Disk Usage”	
P.7	¿Cuánto tiempo le tomó el análisis del reporte “TopN Logical Disk Usage”	
P.8	¿Cuánto tiempo le tomó el análisis del reporte “TopN Memory Usage”	
P.9	¿Cuánto tiempo le tomó el análisis del reporte “TopN Processor Usage”	
P.10	¿Cuánto tiempo le tomó el análisis del reporte “TopN Process Volume Usage”	

JURADO EVALUADOS DE LA TESIS

INTELIGENCIA DE NEGOCIOS Y EL MONITOREO DE SERVIDORES EN EL CENTRO
DE DATOS DE UNA EMPRESA DE CLOUD, LIMA - 2019

ING. VICTOR FREDY ESPEZUA SERRANO
PRESIDENTE

ING. ANA DORIS M. BARRERA
SECRETARIO

ING. EDDY IVAN QUISPE SOTO
VOCAL

ING. JAVIER ALBERTO MANRIQUE QUIÑONEZ
ASESOR