



PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INGENIERÍA

INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

**DISERTACIÓN PREVIA A LA OBTENCIÓN DEL TÍTULO DE INGENIERA
EN SISTEMAS Y COMPUTACIÓN**

**“ANÁLISIS COMPARATIVO DE LAS HERRAMIENTAS DE BIG DATA EN
LA FACULTAD DE INGENIERÍA DE LA PONTIFICIA UNIVERSIDAD
CATÓLICA DEL ECUADOR”**

AUTORAS:

PAOLA ALEXANDRA FLORES AVENDAÑO

ANDREA ESTEFANÍA VILLACÍS VERA

DIRECTOR:

Mtr. JAVIER CÓNDOR

QUITO, Septiembre 2017

RESUMEN

Uno de los desafíos de la información es el de administrar, buscar, almacenar y analizar datos para apoyar a la toma de decisiones; para lo cual, los métodos tradicionales de procesamiento han tenido que evolucionar, con el fin de proporcionar respuestas en tiempo real y considerando el menor costo posible. Este fenómeno es conocido como **Big Data**, en donde todos los datos pueden transformarse en información útil y se pueden utilizar para mejorar la toma de decisiones en cualquier organización, mejorar la eficiencia, reducir costos y aumentar ingresos.

La tecnología denominada “Big Data”, juntos con sus herramientas y su aplicabilidad en el ambiente universitario permitiría encontrar indicadores para tomar decisiones. Las herramientas que nos proporciona Big Data son una manera de procesar datos y extraer el valor de los mismos, permitiendo el manejo de grandes volúmenes de información, de una manera más atractiva que utilizando algún algoritmo y adaptándose a las necesidades del usuario.

Este trabajo, se aplicará a la Facultad de Ingeniería de la PUCE¹, cuya organización recibe diariamente información de estudiantes, docentes y administradores. Con los datos se realiza un análisis comparativo de las herramientas tanto de software libre, licenciadas y en la nube, se pueden ajustar de mejor manera a las necesidades de la facultad y que junto a la gestión de Big Data sirva para el manejo y explotación adecuada de dicha información.

¹ PUCE: Pontificia Universidad Católica del Ecuador

DEDICATORIA

A Dios en primer lugar, que es el pilar en mi familia y en mi vida, el que me dio las fuerzas para seguir adelante con todos obstáculos que tuve durante mi carrera. A mis padres, quienes me han formado con valores y que, con su apoyo incondicional, estuvieron alentándome siempre a superarme cada día, y nunca darme por vencida, con su frase “lo que se empieza, se termina”, gracias a ellos estoy aquí, en el último peldaño de esta etapa. A mi hermano, al cual considero como mi mejor amigo y una gran inspiración en mi vida, por él puedo ser una mejor persona día a día, dándole ejemplo de lucha constante, y demostrándole que con perseverancia se consiguen las cosas que te propongas. En general, a todos los que me conocen y han aportado positivamente a lo largo de mi vida.

Andrea.

Dedico esta tesis a Dios por darme fortaleza en los momentos más difíciles y por hacer realidad un sueño tan anhelado. A mis padres, Marcelo Flores y Mariana Avendaño, por haber confiado en mí y ser un constante apoyo durante toda mi carrera universitaria. Y a toda mi familia por ser un pilar fundamental para nunca rendirme y llegar hasta donde me encuentro ahora.

Paola.

AGRADECIMIENTOS

Agradezco a Dios por la fortaleza necesaria en toda mi carrera, en mis momentos de debilidad y ser la luz en mis noches de deberes y trabajos. Le doy gracias a mis padres Luis y Nelly que fueron siempre los que creyeron y siguen creyendo en mí, los que a pesar de toda circunstancia me han dado su apoyo incondicional y la fuerza para seguir adelante sin dejar que nada me tumbe; les agradezco por haberme brindado una excelente educación, rica en valores, enseñanzas y conocimientos durante toda mi vida, ellos son mi ejemplo de vida. Sin dejar de lado a mi querido hermano Luis Alfredo, le agradezco por ser una parte importante en mi vida, él es mi motivo por el cuál lucho cada día por darle el mejor ejemplo, estoy agradecida por llenar mi vida de risas y amor cuando lo he necesitado, sin dejar de lado a mi director de tesis Mtr.Javier Córdor, que por su esfuerzo y constancia nos ha sabido guiar durante todo este proceso, al igual que nuestros revisores Suyana Arcos y Alfredo Calderón que estuvieron en el seguimiento del mismo.

Andrea.

El presente trabajo me gustaría agradecer en primer lugar a Dios por bendecirme para llegar hasta donde he llegado, por haberme guiado a lo largo de mi carrera y por brindarme una vida llena de felicidad, aprendizaje y experiencia.

Le doy gracias a mis padres Marcelo y Mariana por apoyarme en todo momento, por desvelarse conmigo en las largas noches en las que su compañía era mis ganas para seguir con mi sueño, por los valores que me han inculcado y por la oportunidad que me han brindado para poder tener una excelente educación en todo el transcurso de mi vida.

Agradezco a todos los profesores que me han instruido con sus conocimientos y experiencia durante todos estos años y principalmente a Mtr. Javier Córdor y a todos los ingenieros que nos brindaron apoyo y la comprensión que nos ha brindado durante el desarrollo de la respectiva tesis.

Y finalmente doy gracias a todas las personas que me apoyaron durante todo el transcurso de mi vida educativa y a los grandes amigos que pude conseguir en estos años.

Paola.

TABLA DE CONTENIDO

RESUMEN	i
DEDICATORIA	ii
AGRADECIMIENTOS	iii
ÍNDICE DE FIGURAS	x
ÍNDICE DE ILUSTRACIONES.....	xii
ÍNDICE DE TABLAS	xiv
CAPÍTULO 1. MARCO TEÓRICO.....	15
1.1. Datos.....	15
1.1.1. Tipos de datos.....	15
1.1.2. Base de datos.....	16
1.2. Base de Datos SQL.....	16
1.3. Base de Datos NoSQL.....	16
1.4. Data Warehouse y Data Mining	18
1.4.1. Data Warehouse (DW).....	18
1.4.2. Data Mining (DM)	19
1.5. Cloud Computing	19
1.6. Business Intelligence (BI)	21
1.7. Big Data.....	22
1.7.1. Características de Big Data	22
1.7.2. Aplicaciones típicas de Big Data.....	23
1.7.3. El trabajo con Big Data	23
1.7.4. Beneficios de Big Data.....	24

1.8.	Bases de datos multidimensionales	26
1.9.	Cubos Multidimensionales	28
1.10.	OLAP	28
1.11.	Cubos OLAP.....	29
1.12.	ETL (Extract, Transform and Load).....	29
1.13.	Herramientas de extracción de información	29
1.13.1.	QlikView	30
1.13.2.	Pentaho	31
1.13.3.	Jaspersoft.....	32
1.13.4.	Tableau	32
1.13.5.	Microsoft Power BI.....	33
1.14.	Cuadrante de Gartner para BI y Analíticas (2017)	34
1.15.	Software Libre	35
1.16.	Software Propietario	35
CAPÍTULO 2. HERRAMIENTAS DE BIG DATA.....		36
2.1.	Herramientas de software libre y Open Source	36
2.1.1.	Hadoop	36
2.1.2.	Apache Spark	39
2.1.3.	MongoDB.....	40
2.2.	Herramientas licenciadas.....	42
2.2.1.	Google Cloud Platform	42
2.2.2.	Oracle Big Data Appliance.....	44
2.2.3.	IBM Watson Analytics.....	45

2.3. Herramientas en la nube (Cloud Computing).....	46
2.3.1. Hadoop on-cloud	46
2.3.2. Microsoft Azure	47
2.3.3. Amazon Web Services (AWS).....	48
2.4. Cuadros comparativos de las herramientas de Big Data	49
2.4.1. Cuadro comparativo entre Apache Spark y Hadoop	49
2.4.2. Cuadro comparativo entre MongoDB y Spark.....	50
2.4.3. Cuadro comparativo entre Hadoop y MongoDB	50
2.4.4. Cuadro comparativo entre AWS, Microsoft Azure y Google....	51
2.5. Base de datos Relacionales vs. Base de datos No Relacionales.....	53
2.5.1. Comparación en velocidad de los motores de base de datos.....	53
CAPÍTULO 3. INGENIERÍA – SITUACIÓN ACTUAL.....	56
3.1. PUCE.....	56
3.1.1. Misión.....	56
3.1.2. Visión	57
3.1.3. Valores institucionales	58
3.1.4. Sedes PUCE	58
3.1.5. Organigrama Estructural – PUCE.....	58
3.1.6. Autoridades PUCE Quito	59
3.1.7. Oferta Pregrado	60
3.1.8. Estudiantes Pregrado Nuevos y Antiguos (2016-02)	61
3.2. Facultad de Ingeniería	62
3.2.1. Misión PUCE	62

3.2.2.	Visión PUCE	62
3.2.3.	Compromiso PUCE.....	62
3.2.4.	Características distintivas de la Facultad de Ingeniería.....	62
3.2.5.	Carreras de Ingeniería	63
3.2.6.	Autoridades de la Facultad de Ingeniería	64
3.2.7.	Estudiantes de la Facultad de Ingeniería Semestre 2016-02.....	64
3.2.8.	Número de Alumnos Facultad de Ingeniería (5 últimos años)...	64
3.3.	Ingeniería de Sistemas y Computación	65
3.3.1.	Misión.....	65
3.3.2.	Visión	65
3.3.3.	Objetivos	65
3.3.4.	Perfil profesional	65
3.3.5.	Campo laboral	66
3.4.	Ingeniería Civil.....	66
3.4.1.	Misión.....	66
3.4.2.	Visión	66
3.4.3.	Campo Laboral	67
3.4.4.	Características del Ingeniero Civil	67
CAPÍTULO 4. BIG DATA PARA INGENIERÍA.....		69
4.1.	Requerimientos.....	69
4.2.	Diseño.....	70
4.3.	Implementación	71
4.4.1.	Generación de base de datos con información de prueba	73

4.4.2. Aplicación MongoDB de los datos oficiales de la Facultad	80
4.5. Herramienta QlikView	81
4.6. Herramienta Microsoft Power BI	83
CAPÍTULO 5. CONCLUSIONES Y RECOMENDACIONES	88
5.1. Conclusiones.....	88
5.2. Recomendaciones	90
GLOSARIO	94
BIBLIOGRAFÍA	98
ANEXOS	108

ÍNDICE DE FIGURAS

Figura 1-1. Interacción Clave – Valor. (Acens, 2017).....	17
Figura 1-2. Base de Datos Relacional vs. NoSQL. (Martin, 2017).	17
Figura 1-3. Typical Architecture for a Data Warehouse. (Oracle, 2017).	18
Figura 1-4. Knowledge Discovery Process. (Zentut, 2017).....	19
Figura 1-5. Cloud Computing. (Taylor & Lily, 2016).	19
Figura 1-6. La Nube. (Galarza, Tuamá, Zaccardi, Encinas, & Morales, 2013).	21
Figura 1-7. Beneficios de la Inteligencia de Negocios. (Dean, 2015)	21
Figura 1-8. Proceso de Big Data. (Flores A. & Villacís V., 2017).	24
Figura 1-9. NoSQL vs. SQL. (Czajkowski, 2017).....	25
Figura 1-10. Esquema de Estrella. (Banchero, 2015).	26
Figura 1-11. Esquema de Copo de Nieve. (Banchero, 2015).....	27
Figura 1-12. Esquema Constelación. (Banchero, 2015).	27
Figura 1-13. Cubo Multidimensional. (Guarneros Sánchez, 2012).	28
Figura 1-14. Proceso ETL. (lighpath, 2017).	29
Figura 1-15. Logotipo QlikView. (Qlik, 2017).....	30
Figura 1-16. Logotipo Pentaho Business Intelligence. (renacens, 2017).....	31
Figura 1-17. Logotipo Jaspersoft. (Jaspersoft, 2017).....	32
Figura 1-18. Logotipo Tableau. (Tableau, 2017).....	32
Figura 1-19. Logotipo Microsoft Power BI. (Microsoft, 2017).....	33
Figura 1-20. Cuadrante Mágico Gartner. (Tableau, 2017).....	34
Figura 2-1. Logotipo Hadoop. (Apache Hadoop, 2017).....	36

Figura 2-2. Bloques de datos escritos en HDFS. (Pérez Marqués, 2015).....	37
Figura 2-3. Flujo de datos en un proceso de MapReduce. (Pérez Marqués, 2015). ...	38
Figura 2-4. Logotipo Apache Spark. (Apache Spark, 2017).....	39
Figura 2-5. Logotipo MongoDB. (MongoDB, 2017).	40
Figura 2-6. Logotipo Google Cloud Platform. (Google Cloud Platform, 2017).....	42
Figura 2-7. Arquitectura en Google Cloud. (Google Cloud Platform, 2017).	43
Figura 2-8. Logotipo Oracle Big Data Appliance. (Oracle, 2017).....	44
Figura 2-9. Logotipo IBM Watson Analytics. (IBM, 2017)	45
Figura 2-10. Logotipo Hadoop on-cloud. (Apache Hadoop, 2017).....	46
Figura 2-11. Logotipo Microsoft Azure. (Microsoft, 2017).	47
Figura 2-12. Logotipo Amazon Web Services. (Amazon Web Services, 2017).	48
Figura 2-13. PostgreSQL CRUD. (Correa, 2013).....	53
Figura 2-14. MySQL CRUD. (Correa, 2013).	53
Figura 2-15. MongoDB CRUD. (Correa, 2013).	54
Figura 2-16. Ecosistema Big Data 2016. (Turck, 2016).....	55
Figura 3-1. Organigrama Estructural – Sede Quito. (PUCE, 2017).....	58
Figura 3-2. Carreras Ingeniería PUCE. (Flores A. & Villacís V., 2017).	63
Figura 3-3. Autoridades Ingeniería. (Flores A. & Villacís V., 2017).	64
Figura 4-1. Ejemplo de Cubos OLAP. (Flores A. & Villacís V., 2017).....	70
Figura 4-2. Cubo Estudiante. (Flores A. & Villacís V., 2017).	70
Figura 4-3. Cubo Docente. (Flores A. & Villacís V., 2017).....	70
Figura 4-4. Cubo Cátedra. (Flores A. & Villacís V., 2017).....	71

ÍNDICE DE ILUSTRACIONES

Ilustración 4-1. Inicio terminal. (Flores A. & Villacís V., 2017).....	71
Ilustración 4-2. Proceso principal MongoDB. (Flores A. & Villacís V., 2017).	72
Ilustración 4-3. Proceso secundario MongoDB. (Flores A. & Villacís V., 2017).	72
Ilustración 4-4. Datos Prueba Estudiante. (Flores A. & Villacís V., 2017).	73
Ilustración 4-5. Datos Prueba Docente. (Flores A. & Villacís V., 2017).	73
Ilustración 4-6. Datos Prueba Cátedra. (Flores A. & Villacís V., 2017).	74
Ilustración 4-7. Importe de datos. (Flores A. & Villacís V., 2017).	74
Ilustración 4-8. Datos importados en Robo 3T. (Flores A. & Villacís V., 2017).	75
Ilustración 4-9. Datos Estudiante Robo 3T. (Flores A. & Villacís V., 2017).	75
Ilustración 4-10. Comando muestra de datos. (Flores A. & Villacís V., 2017).	76
Ilustración 4-11. Datos Estudiante Terminal. (Flores A. & Villacís V., 2017).	76
Ilustración 4-12. Datos Docente Robo 3T. (Flores A. & Villacís V., 2017).	77
Ilustración 4-13. Comando muestra de datos. (Flores A. & Villacís V., 2017).	77
Ilustración 4-14. Datos Docente Terminal. (Flores A. & Villacís V., 2017).	78
Ilustración 4-15. Datos Cátedra Robo 3T. (Flores A. & Villacís V., 2017).	78
Ilustración 4-16. Comando muestra de datos. (Flores A. & Villacís V., 2017).	79
Ilustración 4-17. Datos Cátedra Terminal. (Flores A. & Villacís V., 2017).	79
Ilustración 4-18. Importe de datos reales. (Flores A. & Villacís V., 2017).	80
Ilustración 4-19. Datos Docente Robo 3T. (Flores A. & Villacís V., 2017).	80
Ilustración 4-20. Datos Reales Terminal. (Flores A. & Villacís V., 2017).	81
Ilustración 4-21. Filtros QlikView. (Flores A. & Villacís V., 2017).	81

Ilustración 4-22. Filtros QlikView. (Flores A. & Villacís V., 2017).	82
Ilustración 4-23. Tabla Docentes QlikView. (Flores A. & Villacís V., 2017).....	82
Ilustración 4-24. Resultado Gráficos I QlikView. (Flores A. & Villacís V., 2017)..	82
Ilustración 4-25. Resultado Gráficos II QlikView. (Flores A. & Villacís V., 2017).	83
Ilustración 4-26. Interface Power BI. (Flores A. & Villacís V., 2017).....	84
Ilustración 4-27. Power BI con datos y filtros. (Flores A. & Villacís V., 2017).	84
Ilustración 4-28. Importe de datos en Power BI. (Flores A. & Villacís V., 2017). ...	85
Ilustración 4-29. Filtros Power BI. (Flores A. & Villacís V., 2017).....	85
Ilustración 4-30. Filtros II Power BI. (Flores A. & Villacís V., 2017).....	86
Ilustración 4-31. Gráficos I Power BI. (Flores A. & Villacís V., 2017).....	86
Ilustración 4-32. Gráficos II Power BI. (Flores A. & Villacís V., 2017).....	87

ÍNDICE DE TABLAS

Tabla 2-1. Hadoop. (Flores A. & Villacís V., 2017).....	36
Tabla 2-2. Apache Spark. (Flores A. & Villacís V., 2017).....	39
Tabla 2-3. MongoDB. (Flores A. & Villacís V., 2017).	41
Tabla 2-4. Google Cloud Platform. (Flores A. & Villacís V., 2017).....	42
Tabla 2-5. Oracle Big Data Appliance. (Flores A. & Villacís V., 2017).....	44
Tabla 2-6. IBM Watson Analytics. (Flores A. & Villacís V., 2017).	45
Tabla 2-7. Hadoop on-cloud. (Flores A. & Villacís V., 2017).	46
Tabla 2-8. Microsoft Azure. (Flores A. & Villacís V., 2017).....	47
Tabla 2-9. Amazon Web Services. (Flores A. & Villacís V., 2017).....	48
Tabla 2-10. Spark vs. Hadoop. (Flores A. & Villacís V., 2017).....	49
Tabla 2-11. MongoDB vs. Spark. (Solid IT, 2017).	50
Tabla 2-12a. Hadoop vs. MongoDB. (Open Software Integrators, LLC, 2017).....	50
Tabla 2-13a. AWS vs. Azure vs. Google Platform. (apser. IT plug & play, 2015). ..	51
Tabla 2-14. Resultado de las pruebas de BDD. (Correa, 2013).....	54
Tabla 3-1. Autoridades PUCE Quito. (PUCE, 2017).	59
Tabla 3-2a. PUCE, Oferta académica actual. (PUCE, 2017).....	60
Tabla 3-3. Estudiantes Semestre 2016-02. (Dirección Informática PUCE, 2017).....	61
Tabla 3-4. Estudiantes Facultad Ingeniería. (Dirección Informática PUCE, 2017)...	64
Tabla 3-5. Alumnos últimos 5 años Facultad de Ingeniería. (PUCE, 2017).....	64

CAPÍTULO 1. MARCO TEÓRICO

En este capítulo se describirá los conceptos que se encuentran relacionados con Big Data, al igual que el estado del arte dará fundamento teórico al presente trabajo.

1.1. Datos

“Son un el conjunto básico de hechos referentes a una persona, cosa o transacción. Incluyen cosas como: tamaño, cantidad, descripción, volumen, tasa, nombre o lugar. Según O'Brien, la data usualmente no es útil hasta que está sujeta a un proceso de valor añadido: 1) su forma es agregada, manipulada y organizada, 2) su contenido es analizado y evaluado, 3) Es puesta en un contexto para el usuario humano” (UCLA, 2006).

1.1.1. Tipos de datos

Los tipos de datos se clasifican en: estructurados, no estructurados y semi-estructurados; los cuales se mencionarán a continuación.

Datos estructurados: Son aquellos en donde los datos tienen definido su longitud, su formato y que conforman las bases de datos relacionales, las hojas de cálculo y los archivos. Suelen ser fechas, números de cédula, números de cuenta en un banco, etc.

Datos no estructurados: Son datos sin tipos predefinidos, es decir carecen de un formato específico. Se almacenan como documentos u objetos sin estructura uniforme y se carece de control sobre ellos. Fotografía, video, audio son ejemplos de este tipo de datos.

Datos semi-estructurados: Es una combinación entre los datos estructurados y no estructurados, es decir, tienen un tipo de estructura implícita. Un ejemplo son las páginas web.

1.1.2. Base de datos

“Una base de datos es una herramienta para recopilar y organizar información. Las bases de datos pueden almacenar información sobre personas, productos, pedidos u otras cosas” (Microsoft, 2017).

“Una base de datos consiste en alguna colección de datos persistentes e independientes usados por una organización determinada.” (Johnson P. & Rubio L., 2017).

1.2. Base de Datos SQL

Son una base de datos relacional, es decir, un conjunto de tablas, cada una de las cuales tiene un nombre único asignado. Cada tabla consiste de una serie de atributos (columnas o campos) y usualmente almacenan una gran cantidad de listas ordenadas de elementos o tuplas (registros, filas o coordenadas). Cada una de estas lista ordenadas en una tabla relacionada, representa un producto que se identifica con una única clave y que se describe por una serie de atributos. Se puede acceder a los datos relacionados a través de consultas (*queries*) desarrolladas en un lenguaje de programación relacionado llamado SQL (García Reyes, 2012).

1.3. Base de Datos NoSQL

“Hablar de bases de datos NoSQL es hablar de estructuras que nos permiten almacenar información en aquellas situaciones en las que las bases de datos relacionales generan ciertos problemas debido principalmente a problemas de escalabilidad y rendimiento de las bases de datos relacionales donde se dan cita miles de usuarios concurrentes y con millones de consultas diarias. Además de lo comentado anteriormente, las bases de datos NoSQL son sistemas de almacenamiento de información que no cumplen con el esquema entidad–relación. Tampoco utilizan una estructura de datos en forma de tabla donde se van almacenando los datos, sino que para el almacenamiento hacen uso de otros formatos como clave–valor” (Acens, 2017).

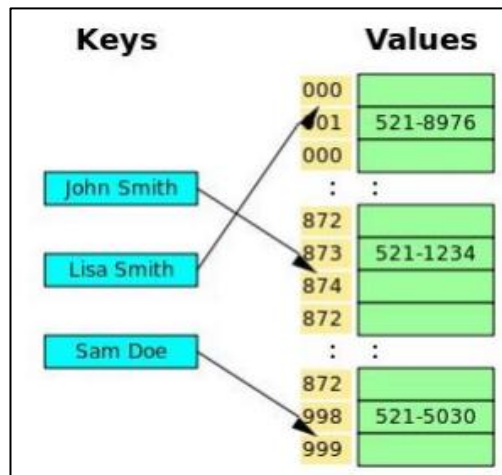


Figura 1-1. Interacción Clave – Valor. (Acens, 2017).

“Una diferencia clave entre las bases de datos de NoSQL y las bases de datos relacionales tradicionales, es el hecho de que NoSQL es una forma de almacenamiento no estructurado” (Martin, 2017).

“Esto significa que NoSQL no tiene una estructura de tabla fija como las que se encuentran en las bases de datos relacionales” (Martin, 2017). Véase en **Figura 1-2.**

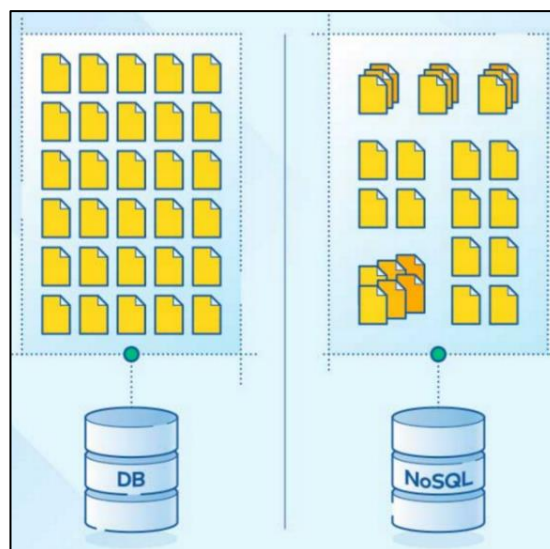


Figura 1-2. Base de Datos Relacional vs. NoSQL. (Martin, 2017).

1.4. Data Warehouse y Data Mining

Data Warehouse es el lugar donde todos los datos de la organización son almacenados, y que Data Mining se encarga de explorar y analizar dicha información en busca de patrones o tendencias que sirvan para la toma de decisiones de la organización.

1.4.1. Data Warehouse (DW)²

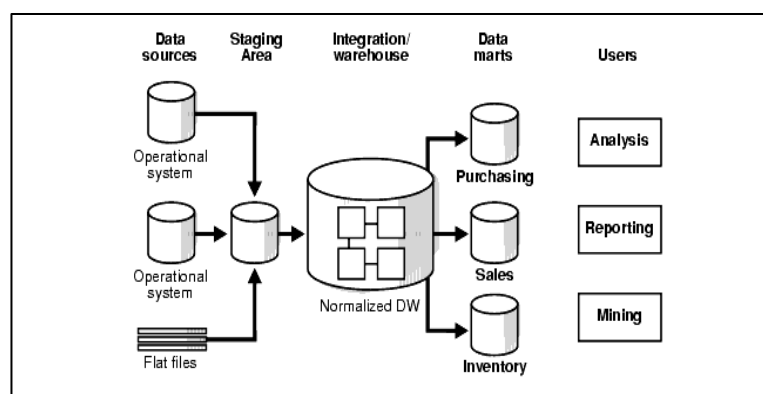


Figura 1-3. Typical Architecture for a Data Warehouse. (Oracle, 2017).

Castilla Alcalá, Albarrán Lozano, & de Pablos Heredero (1998) afirmaron que Data Warehouse es una herramienta tecnológica que facilita la gestión de la información que se encuentra almacenada en las grandes bases de datos empresariales, intentando resolver los problemas clásicos con los que se enfrentan los sistemas de información tradicionales como costes, redundancia de datos, dispersión de la información, etc.

Hernández Lamillar (2005) opina que el Data Warehouse se construye y actúa sobre los datos operacionales. Al ser los datos operacionales sólo informativos y no tener funciones de actualización sobre los datos operativos, permite sacar las máximas ventajas de rendimiento, tanto con el acceso a los datos, como con el SGBD³ base para el Data Warehouse, como se ve en la *Figura 1-3*.

² DW (Data Warehouse) Es una base de datos corporativa que se caracteriza...

³ SGBD (Sistema de gestión de base de datos) o en inglés Database management ...

1.4.2. Data Mining (DM)⁴

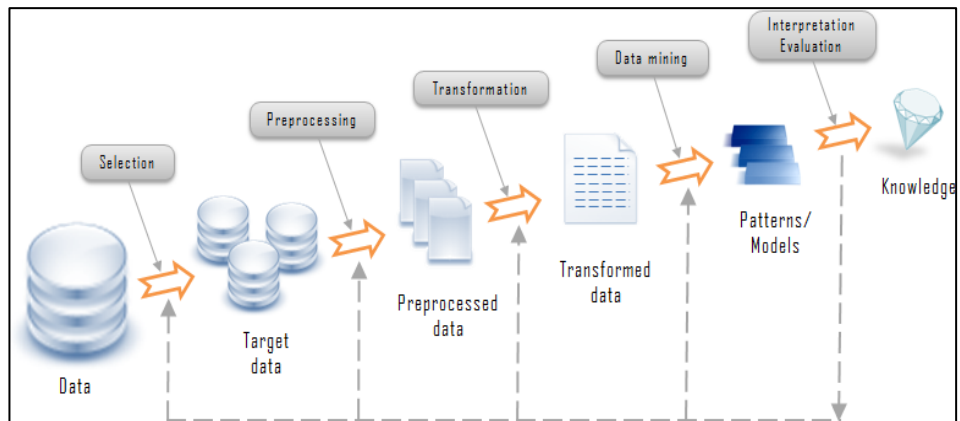


Figura 1-4. Knowledge Discovery Process. (Zentut, 2017).

Data Mining o Minería de datos, es la extracción de conocimiento a partir de grandes cantidades de datos almacenados en repositorios, sobre los que se aplican diferentes métodos con el fin de extraer patrones. (García Reyes, 2012). Véase la *Figura 1-4*.

1.5. Cloud Computing



Figura 1-5. Cloud Computing. (Taylor & Lily, 2016).

⁴ DM (Data Mining): Es el conjunto de técnicas y tecnologías que permiten ...

Modelo de prestación de servicios de computación, información y aplicaciones a través de internet donde la mayoría del software se ejecuta en la propia red y también allí se almacenan los datos. Véase la *Figura 1-5*.

Cloud Computing posee características como la formas de pago, que puede ser por suscripción o por uso; la escalabilidad que es adaptar los recursos en función de las necesidades; la virtualización que tiene la capacidad de separar el hardware del software y la ubicuidad, un acceso desde diferentes dispositivos a través de Internet (Rey, 2012).

Existen tres niveles de prestación de servicios en Cloud Computing: IaaS, PaaS, SaaS; los servicios SaaS son soportados por la plataforma PaaS y estos consumen indirectamente infraestructura como servicio que es el IaaS, esto se debe a que cada uno está sustentado sobre los inferiores.

IaaS (Infrastructure as a Service), es la capa núcleo del servicio. “Se contrata capacidad de proceso (CPU) y almacenamiento. En este caso sólo se paga por lo que se usa” (Galarza, Tuamá, Zaccardi, Encinas, & Morales, 2013). Un ejemplo de este servicio es Azure de Microsoft.

PaaS (Platform as a Service), es la capa intermedia, permite que las utilidades necesarias para el ciclo de vida completo de construir y distribuir aplicaciones web estén disponibles en Internet, si necesidad de descargar o instalar software (Curto Díaz, 2010). Un ejemplo es Google App Engine.

SaaS (Software as a Service), “es la capa externa y es un modelo de despliegue de software en el que una aplicación es alojada como un servicio ofrecido a los clientes” (Curto Díaz, 2010).

Es una aplicación que es brindada para el usuario final, en donde debe pagar un alquiler para su respectivo uso. Uno de los ejemplos más conocidos es Dropbox.

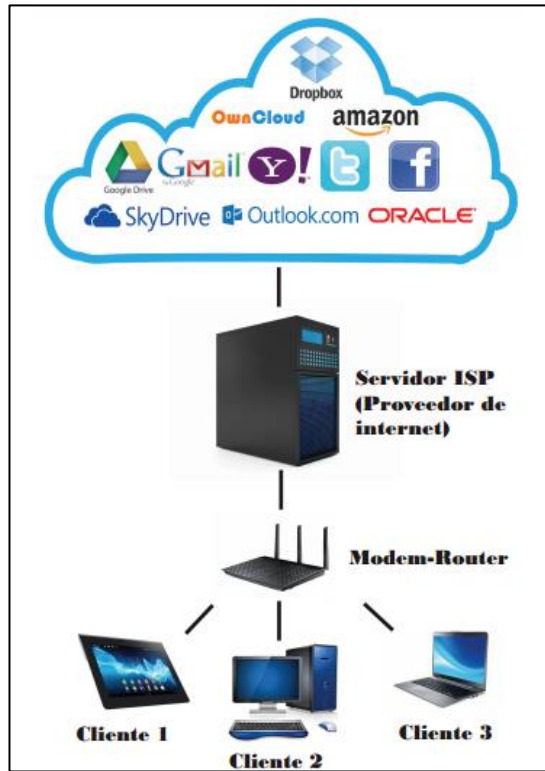


Figura 1-6. La Nube. (Galarza, Tuamá, Zaccardi, Encinas, & Morales, 2013).

1.6. Business Intelligence (BI)⁵

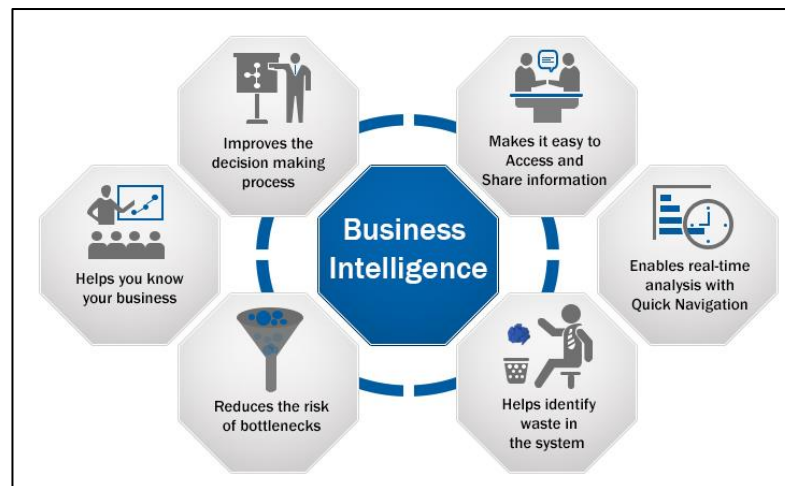


Figura 1-7. Beneficios de la Inteligencia de Negocios. (Dean, 2015)

⁵ BI (Business Intelligence) es la habilidad para transformar los datos en ...

La inteligencia de negocios es la destreza para tomar decisiones empresariales, se logra mediante el uso de metodologías, aplicaciones y tecnologías las que permiten reunir, depurar, transformar datos y aplicar en ellos técnicas analíticas de extracción de conocimiento, generando el conocimiento sobre los problemas y oportunidades del negocio para que pueden ser corregidos y aprovechados respectivamente (Rosado Gómez & Rico Bautista, 2010).

1.7. Big Data

“Se conoce como Big Data el tratamiento y análisis de grandes cantidades de datos, cuyo tamaño hace imposible manejarlos con las herramientas de bases de datos y analíticas convencionales” (Pérez, 2015).

“El término “Big Data” suele aplicarse a conjuntos de datos que superan la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable y por los medios habituales de procesamiento de la información” (Pérez Marqués, 2015).

“Big Data es un conjunto de soluciones que incluyen software, hardware y cloud computing (computación en la ‘nube’ de Internet) y que permiten administrar grandes bases de datos” (Noticias Financieras, 2015).

1.7.1. Características de Big Data

Se acostumbra a referirse a datos masivos que cumplen tres características, las conocidas 3V's: gran *volumen* de datos, *velocidad* de estos datos y *variedad* del origen de los mismos. Pero hay una cuarta V, cada vez más importante, que es la *visualización*. (Tascón & Coullaut, 2016), (Caballero & Martín, 2015).

Volumen: Se debe almacenar y trabajar con una gran cantidad de datos, obligando a que se deban utilizar nuevas infraestructuras, las cuales deben ser escalables y distribuidas.

Velocidad: Está asociado con la proliferación de nuevas fuentes de datos y la necesidad de poder utilizarlos más rápidamente; además de poder transformarlos en información útil en el menor tiempo posible, es así que los tiempos necesarios en ser creados, almacenados, analizados y visualizados se deben reducir para poder acercarse al “tiempo real”.

Variedad: Es la diversidad de los tipos de datos y sus diferentes fuentes de obtención, es decir los datos vienen en diferentes tipos de datos estructurados, semiestructurados, no estructurados y requiere de diferentes tipos de análisis y herramientas para su interpretación, generando información nueva.

Visualización: Poder visualizar la información de los datos de una forma comprensible es otra de las funciones necesarias del Big Data. La visualización ha ido adquiriendo cada vez más importancia. (Tascón & Coullaut, 2016).

Valor. El valor económico de los diferentes datos varía significativamente y el desafío esencial es identificar la información valiosa, transformarla y extraer los datos para su análisis (Pérez Marqués, 2015).

1.7.2. Aplicaciones típicas de Big Data

Existen aplicaciones utilizadas en Big Data, las cuales se encuentran en áreas donde existen grandes volúmenes de datos, que se pueden analizar, procesar y presentar la información que está oculta en el almacenamiento de los datos.

Las áreas más comunes que utilizan Big Data por su gran cantidad de información son: patrones de detección de fraude, medicina, educación, área financiera, turismo, entretenimiento, entre otros.

1.7.3. El trabajo con Big Data

Para trabajar con grandes cantidades de datos en Big Data se sigue el siguiente diagrama (*Figura 1-8*):

- a. *Recolección de datos*: Es emitir, procesar y recoger información dentro de un cierto contexto.
- b. *Almacenamiento*: El repositorio donde se almacena gran volumen de datos, por lo general escalable. “La llegada del modelo de almacenamiento y trabajo en ‘nube’ está también ayudando a la consolidación del nuevo mundo del Big Data” (Tascón & Coullaut, 2016).
- c. *Análisis*: Con el análisis se puede revelar y procesar la información que parecía oculta en el almacenamiento de los datos (Tascón & Coullaut, 2016).
- d. *Visualización*: En Big Data se utiliza diferentes herramientas para visualizar los datos que fueron analizados; tales como gráficas, tablas, de manera que sea comprensible para el usuario.

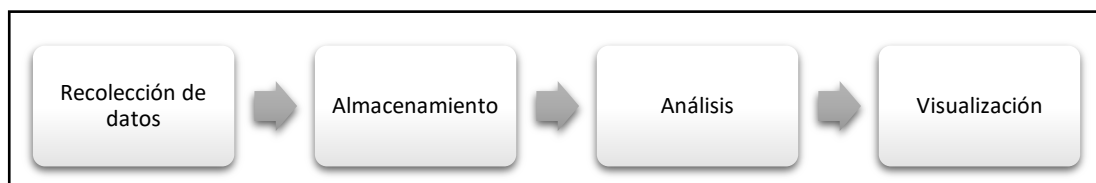


Figura 1-8. Proceso de Big Data. (Flores A. & Villacís V., 2017).

1.7.4. Beneficios de Big Data

Entre los beneficios que proporciona Big Data se encuentra: mejorar la capacidad para manejar y organizar datos, permitiendo implementar un análisis predictivo para ser más eficaz y proactivo; además de tomar mejores decisiones, con reacción rápida, mayor innovación y ventaja competitiva, generando estrategias de marketing con un análisis previo de la información recogida y reduciendo “la latencia en los procesos críticos de la organización a fin de tener en tiempo real el comportamiento de las variables requeridas para la toma de decisiones” (Cock Gomez, 2014)

Lo anterior da como consecuencia una mayor seguridad, disminuyendo el riesgo y logrando la detección de fraudes.

Existen varios factores para determinar si se elige el modelo de base de datos relacional o no relacional, entre los cuales se destaca el origen de los datos, es decir, de donde proviene cada tipo de dato; el modelo relacional maneja datos estructurados, mientras que el no relacional maneja varios tipos de datos y no necesariamente todos iguales, por ejemplo, pueden ser formatos XML, JSON, redes sociales, entre otros.

Por otro lado, las empresas pequeñas o medianas no manejan gran volumen de información, por lo que una base relacional bien estructurada es suficiente para el correcto manejo de la información. “Incluso los comercios con ventas por internet a menudo pueden limitarse al modelo relacional si se trata de una página sencilla, con catálogo de productos más o menos estable (al menos en formato) y número relativamente controlado de accesos” (Caballero & Martín, 2015).

Otros factores son las **consultas y relaciones manejadas en la base de datos, y la estructura de la base de datos**, tomando en cuenta que Big Data es la solución cuando la empresa u organización posee cambios constantes en la información y en su estructura, al igual cuando la información proviene de diferentes orígenes de datos.

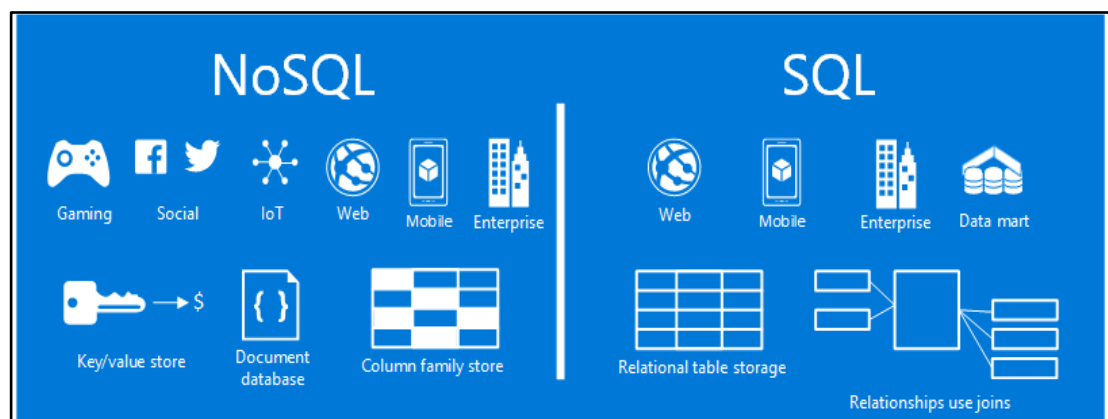


Figura 1-9. NoSQL vs. SQL. (Czajkowski, 2017).

1.8. Bases de datos multidimensionales

Es una base de datos diseñada para procesamiento analítico on-line (OLAP). Estructurada como hipercubo con un eje por dimensión, “proveen una estructura que permite tener acceso flexible a los datos, para explorar y analizar sus relaciones y resultados, las cuales pueden ser visualizadas en un cubo multidimensional” (Guarneros Sánchez, 2012).

En las bases de datos multidimensionales implican tres variantes posibles de modelamiento:

- *Esquema de estrella:*

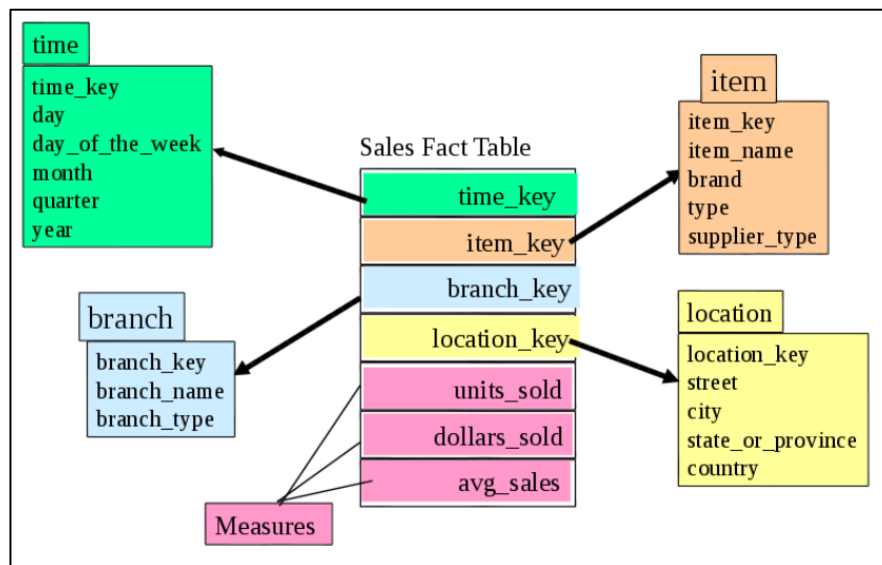


Figura 1-10. Esquema de Estrella. (Banchero, 2015).

“Es el esquema más utilizado, en donde DW contiene: una gran tabla central (Fact Table) que contiene el volumen de datos sin redundancia y un conjunto de tablas relacionadas (Dimension Tables) una por cada dimensión” (Banchero, 2015).

- *Esquema de copo de nieve*

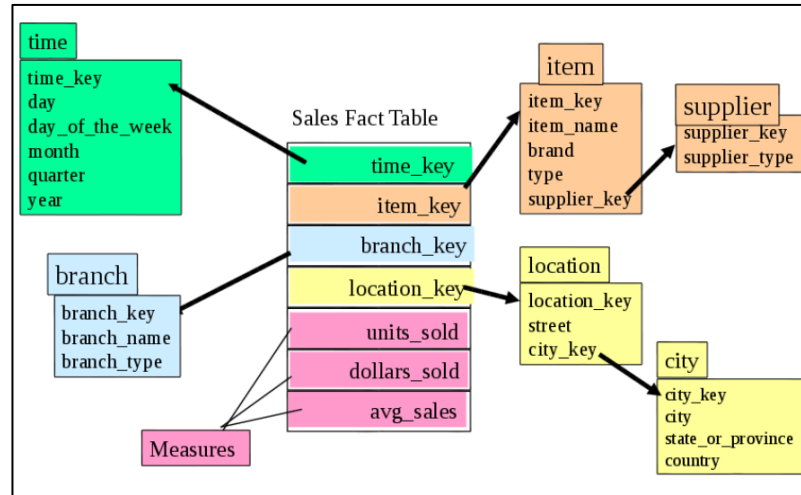


Figura 1-11. Esquema de Copo de Nieve. (Banchero, 2015).

En este tipo de esquema, algunas tablas de dimensiones son normalizadas, adicionando tablas, el gráfico forma una figura de copo de nieve; permite que las tablas tengan un mantenimiento adecuado y ahorra espacio de almacenamiento (Banchero, 2015).

- *Esquema constelación o copo de estrellas*

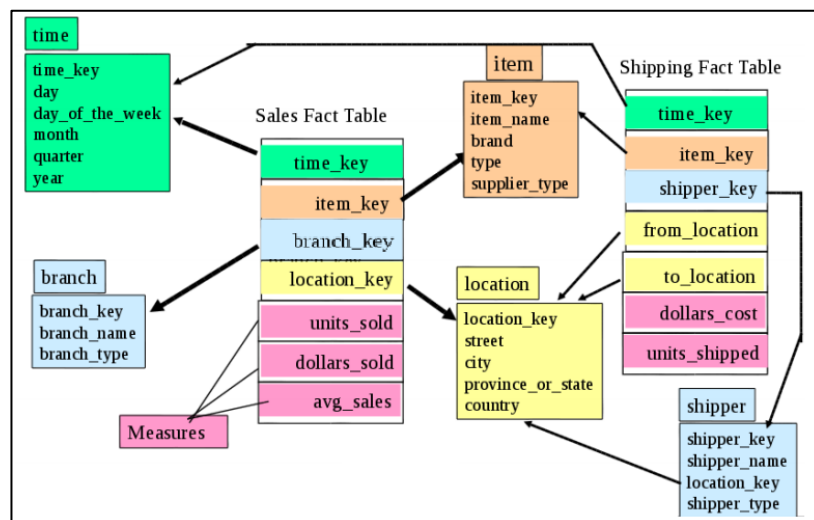


Figura 1-12. Esquema Constelación. (Banchero, 2015).

“Son múltiples tablas de hechos que comparten tablas de dimensiones visto como una colección de esquemas de estrella” (Banchemo, 2015).

1.9. Cubos Multidimensionales

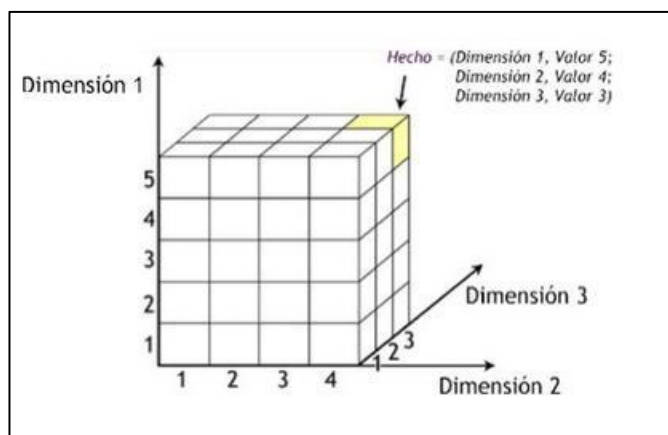


Figura 1-13. Cubo Multidimensional. (Guarneros Sánchez, 2012).

“Los cubos multidimensionales tienen la función de representar los datos planos que se encuentran almacenados en filas y columnas, en una matriz de N dimensiones. Mediante la creación y consulta de los mismos se puede tener acceso a los datos almacenados en el Data Mart para explorarlos, analizar las relaciones existentes entre ellos y obtener resultados” (Manzano de Armas, Molina Pérez, & Castro Soler, 2017).

1.10. OLAP

Es un paradigma o solución utilizada en el campo de la llamada inteligencia empresarial (Business Intelligence), cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Se trata de un método para dar respuestas rápidas a las peticiones de bases de datos complejas. Y generalmente está diseñados para: Apoyar las necesidades de análisis complejos de toma de decisiones, analizar los datos de un número de diferentes perspectivas (dimensiones de negocios) y soporte de análisis complejos contra la entrada de grandes (a nivel atómico) conjuntos de datos (Gallardo, 2012).

1.11. Cubos OLAP

“Es una estructura de datos que supera las limitaciones de la base de datos relacionales y proporciona un análisis rápido de datos. Los cubos pueden mostrar y sumar grandes cantidades de datos, a la vez que proporcionan a los usuarios acceso mediante búsqueda a los puntos de datos. De este modo, los datos se pueden resumir o reorganizar según sea necesario” (Microsoft, 2016).

1.12. ETL (Extract, Transform and Load)

“Extract, Transform and Load (Extraer, transformar y cargar, frecuentemente abreviado a ETL) Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data marts o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio” (Morales Mateo, 2012).

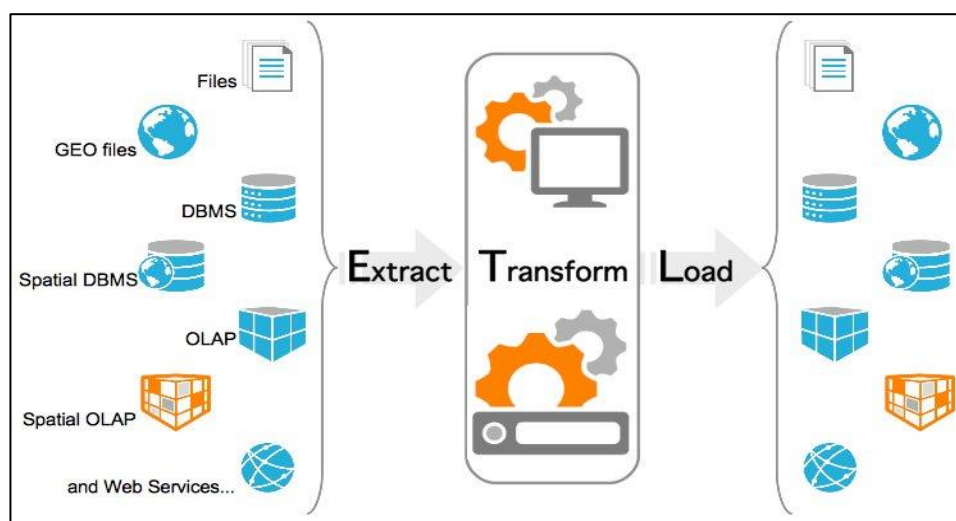


Figura 1-14. Proceso ETL. (lighpath, 2017).

1.13. Herramientas de extracción de información

Las herramientas de extracción de la información es el paso primordial para el análisis de datos, es decir, “la ciencia que examina datos en bruto con el propósito de sacar conclusiones sobre la información” (Rouse, 2012).

“El análisis de datos es usado en varias industrias para permitir que las compañías y las organizaciones tomen mejores decisiones empresariales y también es usado en las ciencias para verificar o reprobador modelos o teorías existentes” (Rouse, 2012).

“El análisis de datos se distingue de la extracción de datos por su alcance, su propósito y su enfoque sobre el análisis. Los extractores de datos clasifican inmensos conjuntos de datos usando software sofisticado para identificar patrones no descubiertos y establecer relaciones escondidas” (Rouse, 2012).

“El análisis de datos se centra en la inferencia, el proceso de derivar una conclusión basándose solamente en lo que conoce el investigador” (Rouse, 2012).

1.13.1. QlikView



Figura 1-15. Logotipo QlikView. (Qlik, 2017).

Entre las herramientas para la extracción de información Open Source y el análisis de datos tenemos a **QlikView**, anteriormente conocida como QlikTech es una herramienta de escritorio que ayuda en el campo de Business Intelligence. Surge en Suecia en el año de 1993 siendo fundadores de esta herramienta Accel Partners y Jerusalem Venture Partners (JVP).

QlikView interpreta gráficos tecnológicamente avanzados, accede a datos de dispositivos móviles, explora las relaciones que existen entre los datos, generando información a tiempo real.

Qlik se basa en 4 características básicas que debe tener una herramienta para el análisis de datos que son: calidad, comprensión, interacción y conocimiento; es lo que da significado de Qlik.

1.13.2. Pentaho



Figura 1-16. Logotipo Pentaho Business Intelligence. (renacens, 2017).

Pentaho también cubre necesidades de Análisis de los Datos e Informes empresariales, las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Entre una de las características más relevantes tenemos la administración y distribución de reportes, análisis y dashboards. Para el análisis de datos, tiene un despliegue gráfico interactivo, con tablas de datos donde los resultados son explorables y navegables incluso en los gráficos (Arias Moreno, Guzmán Tapia, Solis Alarcón, Yacolca Neyra, & Zumaeta Mejía, 2012).

Todos los componentes del módulo Pentaho Reporting y Pentaho Análisis pueden formar parte de un Dashboard realiza varias funciones como la identificación de indicadores (KPI, Key Performance Indicators), monitoreo de datos, seguimiento de excepción, alertas basadas en reglas del negocio, entre otras (Arias Moreno, Guzmán Tapia, Solis Alarcón, Yacolca Neyra, & Zumaeta Mejía, 2012).

Pentaho también cuenta con Pentaho Reporting que permite la distribución de los resultados del análisis en múltiples formatos, todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto; permiten también la programación de tareas y ejecución automática de informes con una determinada periodicidad (Arias Moreno, Guzmán Tapia, Solis Alarcón, Yacolca Neyra, & Zumaeta Mejía, 2012).

1.13.3. Jaspersoft



Figura 1-17. Logotipo Jaspersoft. (Jaspersoft, 2017).

Jaspersoft también nos proporciona herramientas para el manejo de datos, “es la Plataforma de Inteligencia de Negocios Multi-Cliente de TIBCO⁶. Un software de autoservicio BI, de clase empresarial, que permite generar tableros de análisis embebidos para grandes organizaciones. Escalable y asequible, los usuarios de negocio pueden obtener respuestas a sus preguntas rápidamente a través de soluciones de informes, cuadros de mando y análisis de autoservicio” (Jaspersoft, 2017).

“De código abierto basada en estándares se integra fácilmente en su infraestructura de TI existente, independientemente del servidor de aplicaciones, el servidor Web o el sistema operativo que utilice” (Jaspersoft, 2017).

1.13.4. Tableau



Figura 1-18. Logotipo Tableau. (Tableau, 2017).

Tableau es una herramienta que proporciona un análisis profundo y de utilidad, sus dos activos más importante son su gente y sus datos (tanto big data como datos de menor volumen). Admite análisis a medida por parte de cualquier usuario, de este modo otorga a todo el personal la capacidad para ver y comprender mejor sus datos (Tableau, 2017).

⁶ TIBCO: Empresa dedicada a crear productos tanto hardware como software para ...

“Tableau es una herramienta de visualización de datos interactiva, es decir, el usuario tiene la posibilidad de interactuar con los datos: comparar, filtrar, conectar unas variables con otras... Además, la plataforma y los paneles que se pueden crear con la herramienta son muy visuales (facilita la comprensión rápida de los datos)” (Cía, 2015).

“También tiene algunas ventajas interesantes cuando manejas bases de datos: acepta formatos con Excel, Access y texto; puedes acceder a muchas bases de datos comunes como Microsoft SQL Server, MySQL, Oracle o Greenplum; y también tienes la posibilidad de usar la API de Tableau para la extracción sistemática de datos” (Cía, 2015).

1.13.5. Microsoft Power BI



Figura 1-19. Logotipo Microsoft Power BI. (Microsoft, 2017).

“Power BI es un conjunto de aplicaciones de análisis de negocios que permite analizar datos y compartir información. Los paneles de Power BI ofrecen a los usuarios una vista de 360 grados con sus métricas más importantes en un mismo lugar. La información se actualiza en tiempo real y está disponible en todos sus dispositivos. Asimismo, puede acceder a sus datos e informes desde cualquier lugar con las aplicaciones móviles de Power BI Mobile, que se actualizan automáticamente con los cambios que se realizan en los datos” (Microsoft, 2017).

“Si es un analista de datos que proporciona informes y análisis a su organización, Power BI le permite liberar todo su potencial creativo y alcanzar la máxima productividad en sus tareas” (Microsoft, 2017).

1.14. Cuadrante de Gartner para BI y Analíticas (2017)

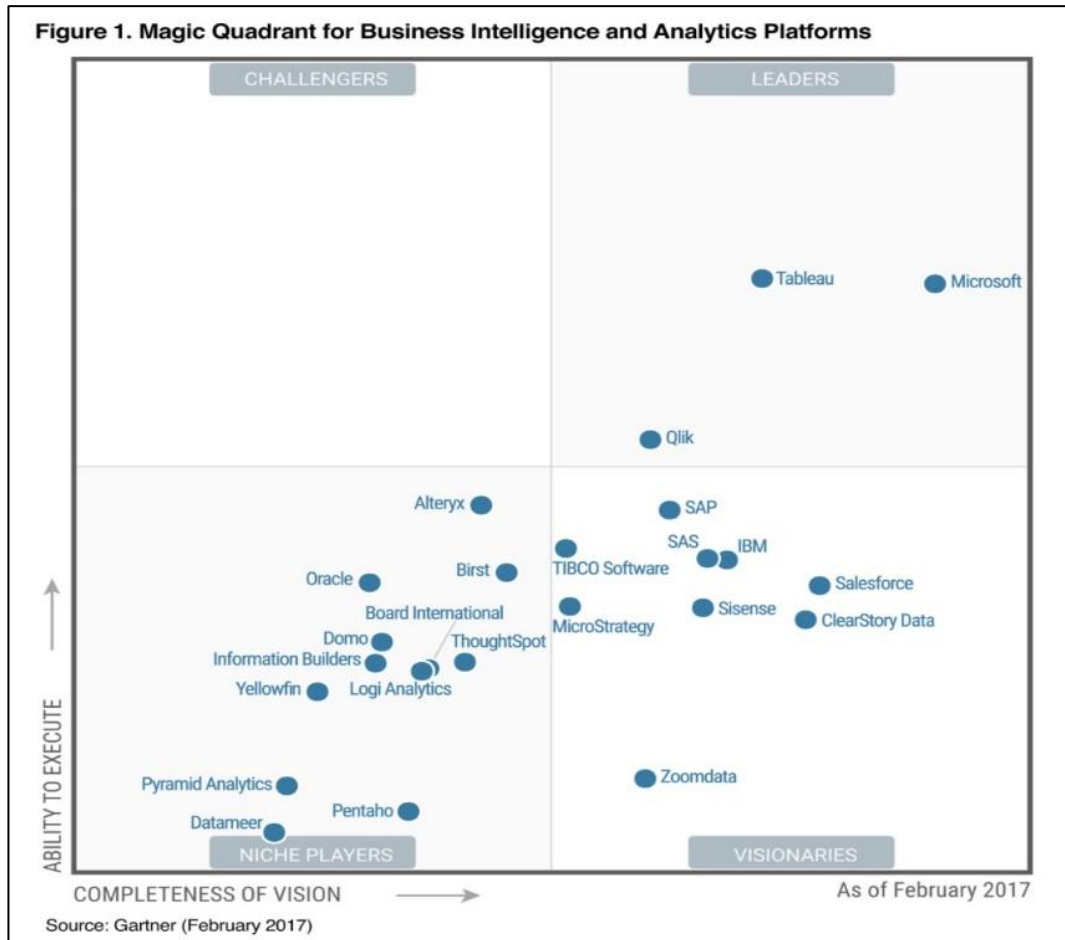


Figura 1-20. Cuadrante Mágico Gartner. (Tableau, 2017).

Como se puede ver en la **Figura 1-20**, Qlik, Tableau y Microsoft Power BI lideran como herramientas para el análisis de datos, contando con características relevantes con respecto a las demás herramientas, poseen una ventaja competitiva grande por su escalabilidad y manejo de datos para BI.

Por otra parte herramientas Open Source como TIBCO-Jaspersoft que en comparación con el año anterior mantiene su misma posición en el cuadrante de Visionario. Pentaho el presente año se posiciona en el cuadrante de Jugadores de Nicho con herramientas básicas para la extracción de información y análisis de datos.

1.15. Software Libre

Software libre hace referencia a la libertad de los usuarios para ejecutar, copiar, distribuir, estudiar, modificar el software y distribuirlo modificado.

En general, se debe cumplir con cuatro tipos de libertades para que un programa pueda ser considerado en este campo. La interfaz gráfica suele ser muy sencilla y por tanto poco amigable. Su costo es muy bajo y mayormente gratuito.

Existen aplicaciones para todas las plataformas y el usuario no depende del creador del software. La figura más representativa es Richard Stallman (Vega, 2012).

1.16. Software Propietario

Software propietario es todo programa informático en el que el usuario tiene limitaciones para usarlo, modificarlo o redistribuirlo (esto último con o sin modificaciones).

Generalmente la interfaz gráfica es más completa por lo que es considerada avanzada y entendible por un usuario común.

Las aplicaciones de este tipo pueden tener un costo mayor y el usuario depende del creador del software. Destaca Bill Gates como pionero en este software (Vega, 2012).

CAPÍTULO 2. HERRAMIENTAS DE BIG DATA

En este capítulo se describirá las herramientas de software libre, propietarias y en la nube, para posteriormente realizar el análisis comparativo entre éstas.

2.1. Herramientas de software libre y Open Source

Las herramientas de software libre garantizan a los usuarios, mediante una licencia, una serie de libertades que no otorga el propietario del software privativo. A continuación se describirán las características más relevantes que tiene Big Data en cuanto a herramientas de software libre.

2.1.1. Hadoop



Figura 2-1. Logotipo Hadoop. (Apache Hadoop, 2017).

Desarrollador(es)	Apache Software Foundation
Sistema Operativo	Multiplataforma (Windows, Linux, OS X)
Licencia	Apache License 2.0
Lenguaje de Programación	Java
Última Versión	Versión 2.0
Sitio Oficial	http://hadoop.apache.org/

Tabla 2-1. Hadoop. (Flores A. & Villacís V., 2017).

“Es una infraestructura digital de desarrollo creada en código abierto bajo licencia Apache” (Pérez Marqués, 2015). Hadoop gestiona una gran cantidad de datos en petabytes, lo que hace posible que el funcionamiento de sus aplicaciones sea el adecuado. Detecta los fallos y vuelve a ejecutar la instrucción, probando con otros caminos de nodos, para que los resultados que se obtienen no sean inconsistentes.

Permite desarrollar tareas con datos de cantidad masiva, dividiéndolas en partes y distribuyéndolas a un conjunto de máquinas. El análisis realizado es en petabytes de datos, en entornos distribuidos formados por varias máquinas sencillas. En la actualidad, la tecnología avanza rápidamente, por esta razón trabajar de esta forma es una manera razonable, debido a la gran cantidad de información que se maneja y son utilizadas en grandes empresas como son Yahoo, Google, Twitter, Facebook, etc. El núcleo de Hadoop está formado por dos componentes el sistema de ficheros distribuido HDFS⁷ y el modelo de procesamiento MapReduce.

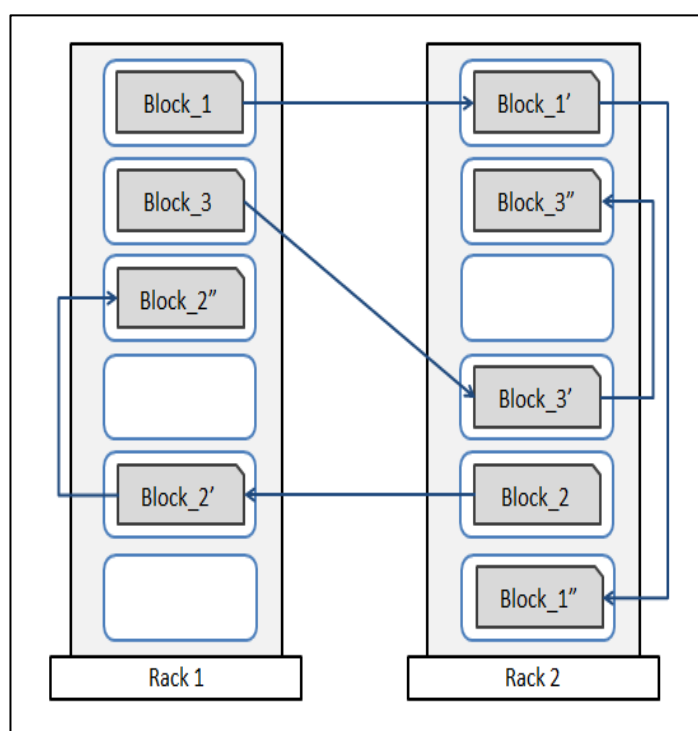


Figura 2-2. Bloques de datos escritos en HDFS. (Pérez Marqués, 2015).

HDFS permite crear sistemas de ficheros utilizando servidores “commodity”, los cuales permiten el paradigma de llevar los datos a la computación y ofrecen redundancia, capacidad y rendimiento.

⁷ Hadoop Distributed File System, es de código abierto, la cual permite a las ...

“Los datos en el clúster de Hadoop son divididos en pequeñas piezas llamadas *bloques* y distribuidas a través del clúster. De esta manera, las funciones *map* y *reduce* pueden ser ejecutadas en pequeños subconjuntos y esto provee de la escalabilidad necesaria para el procesamiento de grandes volúmenes” (Pérez Marqués, 2015).

En la **Figura 2-2**, se explica de manera más detallada como los bloques de datos son escritos hacia el HDFS. Observamos que cada bloque es almacenado 3 veces y al menos un bloque se almacena en un diferente Rack para lograr redundancia (Pérez Marqués, 2015).

En cuanto a **MapReduce**, es un modelo de procesamiento distribuido, está muy ligado a HDFS; ya que toma ventaja de que los ficheros se encuentran divididos en partes lo cual permite la optimización de cómputo.

MapReduce se refiere principalmente a dos procesos independientes, el primero es el *map* que toma los datos y los convierte en datos agrupados, cada uno de ellos son separados en tuplas; el segundo proceso es *reduce* que con la salida del map procesa los datos y combina las tuplas en conjuntos más pequeños.

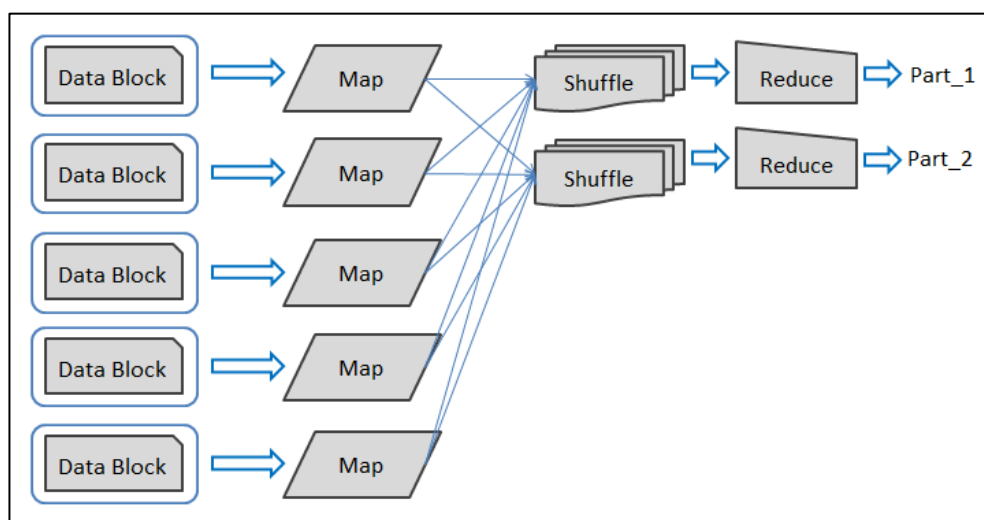


Figura 2-3.Flujo de datos en un proceso de MapReduce. (Pérez Marqués, 2015).

Como podemos observar en la Figura 2-3, la fase entre *map* y *reduce* denominada *shuffle* determina a que nodo se dirigen las tuplas del proceso *map* las cuales van a ser asignadas a las tareas del proceso *reduce*.

Tanto MapReduce como HDFS colaboran en conjunto, permiten el almacenamiento y procesamiento de grandes cantidades de datos de una manera eficiente.

2.1.2. Apache Spark



Figura 2-4. Logotipo Apache Spark. (Apache Spark, 2017).

Desarrollador(es)	Foundation, UC Berkeley AMPLab, Databricks
Sistema Operativo	Multiplataforma (Microsoft Windows, OS X, Linux)
Licencia	Apache License 2.0
Lenguaje de Programación	Scala, Java, Python, R
Última Versión	Versión 2.1.0
Sitio Oficial	spark.apache.org

Tabla 2-2. Apache Spark. (Flores A. & Villacís V., 2017).

Spark es una plataforma que está diseñada para mejorar la velocidad y el rendimiento de las aplicaciones Big Data, brindando rapidez a las consultas y el tratamiento de grandes paquetes de datos, lo cual permite la optimización de la gestión de Big Data y esto lo hace mediante la repartición de los procesos a través de la memoria de distintas máquinas.

Empezó como un proyecto de investigación en la Universidad de California en Berkeley, en la cual se estaba desarrollando Hadoop y en donde observaron que MapReduce no era lo suficientemente eficiente para procesos de algoritmos iterativos o consultas interactivas; es así que le redireccionaron a Spark para que éste pueda dar soporte para persistencia en memoria y un eficiente sistema de tolerancia a fallos.

Spark está escrito en Scala, el cual es más conciso, permitiendo más expresividad en menos líneas de código. No necesita Hadoop para ser ejecutado, sin embargo, es compatible tanto con Hadoop como con su estructura de datos, permitiendo la capacidad de procesar datos de cualquier fuente. Existen varios componentes que definen el ecosistema de Apache Spark, entre los cuales se encuentran:

Spark Core Engine, es el motor responsable para la planificación, distribución y monitorización de aplicaciones ejecutadas en un clúster; *Spark SQL*, el cual da soporte para las consultas interactivas.

Spark Streaming, es un componente que permite procesar y analizar datos caso en tiempo real; *MLlib*, es la librería de aprendizaje automático de Spark y proporciona una gama amplia de algoritmos de alta calidad.

GraphX, es un motor para el análisis de grafos, permitiendo al usuario crear, transformar y obtener conclusiones sobre datos estructurados; *SparkR*, es un paquete que integra Spark con el lenguaje estadístico R (Macías & Gómez, 2015).

2.1.3. MongoDB



Figura 2-5. Logotipo MongoDB. (MongoDB, 2017).

Desarrollador(es)	MongoDB Inc.
Sistema Operativo	Multiplataforma
Licencia	GNU AGPL v3.0 (drivers: licencia Apache)
Lenguaje de Programación	C++
Última Versión	Versión 3.2.11
Sitio Oficial	https://www.mongodb.org/

Tabla 2-3. MongoDB. (Flores A. & Villacís V., 2017).

MongoDB es una base de datos NoSQL orientada a documentos, se caracteriza porque almacena los datos en documentos de tipo JSON⁸ con un esquema dinámico denominado BSON⁹ (Sarasa, 2016).

“Organizaciones de todos los tamaños están usando MongoDB para crear nuevos tipos de aplicaciones, mejorar la experiencia del cliente, acelerar el tiempo de comercialización y reducir costes” (MongoDB, 2017).

“Es una base de datos ágil que permite a los esquemas cambiar rápidamente cuando las aplicaciones evolucionan, proporcionando siempre la funcionalidad que los desarrolladores esperan de las bases de datos tradicionales, tales como índices secundarios, un lenguaje completo de búsquedas” (MongoDB, 2017).

Los principales tipos de datos soportados por esta herramienta son: nulo, booleanos, números reales y enteros, cadenas, fechas, expresiones regulares para las consultas, documentos embebidos, identificadores de objetos, código JavaScript (MongoDB, 2017).

Las colecciones se agrupan en bases de datos, por lo tanto una única instancia puede gestionar distintas bases de datos y agrupándoles en cero o más colecciones (MongoDB, 2017).

⁸ JSON, JavaScript Object Notation–Notación de Objetos de JavaScript, es un ...

⁹ BSON, Binary Jason es una serialización codificada en binario de documentos ...

Finalmente, el objetivo de MongoDB es brindar un servicio que sea escalable, con gran capacidad de rendimiento y gran disponibilidad, permitiendo que las organizaciones de pequeño, mediano y gran tamaño puedan alcanzar sus objetivos empresariales más amplios, entre los principales y que la mayoría de empresas desea es disminuir precios, permitiendo aumentar el tiempo de comercialización y reducir riesgos (MongoDB, 2017).

2.2.Herramientas licenciadas

Las herramientas de software propietario el usuario debe considerar el precio y las licencias pues a menudo éstas son pagadas.

A continuación, se describirán las características más relevantes que tiene Big Data en cuanto a herramientas de software propietario.

2.2.1. Google Cloud Platform



Figura 2-6. Logotipo Google Cloud Platform. (Google Cloud Platform, 2017).

Desarrollador(es)	Google Inc.
Licencia	Propietaria
Lenguaje de Programación	Java, Python, Go, Ruby
Versión Inicial	Octubre 2011
Sitio Oficial	cloud.google.com

Tabla 2-4. Google Cloud Platform. (Flores A. & Villacís V., 2017).

Google Cloud Platform utiliza la infraestructura base, el análisis de datos y el sistema de aprendizaje automático de Google, para encontrar respuestas con mayor rapidez y crear mejores productos. Conformada por una infraestructura diseñada para el futuro, segura, global, de alto rendimiento, rentable y que se encuentra en constante evolución. (Google Cloud Platform, 2017)

Permite el análisis de datos a escala de toda la web, con la familiaridad de SQL y en una arquitectura completamente gestionada. Algunas características significativas son: almacenamiento de datos totalmente administrado, procesamiento por lotes y de transmisiones, exploración de datos, Hadoop/Spark y mensajería fiable, controla el acceso y la visibilidad de los recursos en una plataforma protegida por el modelo de seguridad de Google (Google Cloud Platform, 2017).

Permite desarrollar e implementar tus aplicaciones con nuestra interfaz de línea de comandos y otras herramientas para desarrolladores. Almacenamientos de objetos y bases de escalables, resistentes y de alto rendimiento para tus aplicaciones (Google Cloud Platform, 2017).

Como se puede observar en la **Figura 2-7**. La arquitectura de Big Data está representada de la siguiente manera: **Cloud Pub/Sub** es un servidor a gran escala, que permite enviar y recibir mensajes entre aplicaciones independientes, puede aprovechar la flexibilidad de la nube para desacoplar los sistemas y componentes alojados en la plataforma en la nube o en otros lugares en Internet.

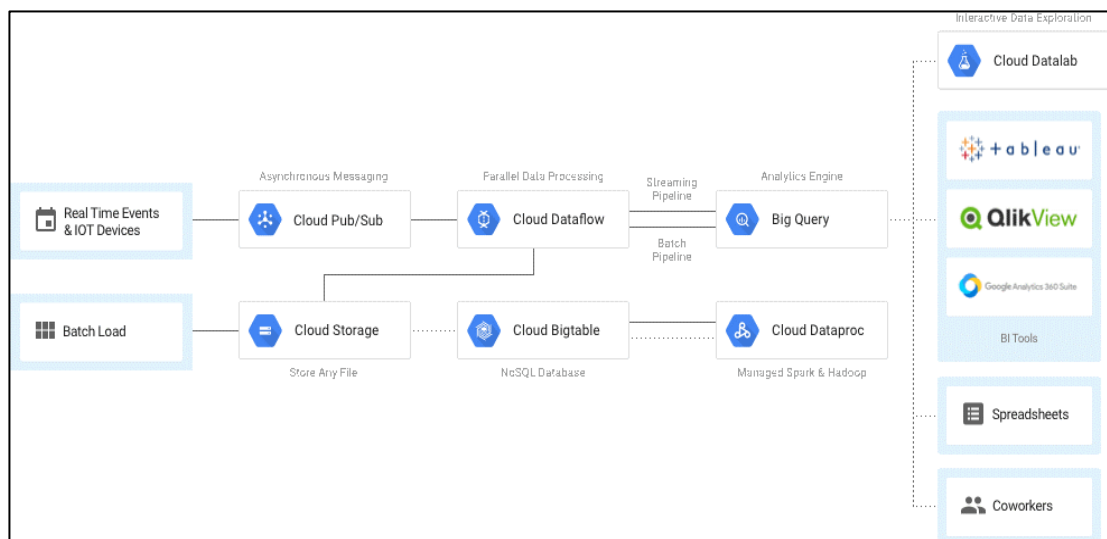


Figura 2-7. Arquitectura en Google Cloud. (Google Cloud Platform, 2017).

Cloud Dataflow, ofrece un modelo de programación unificado y un servicio gestionado para la ejecución de una amplia gama de patrones de procesamiento de datos. *Big Query* es un sistema que permite analizar los datos en tiempo real, no requiere de un administrador de base de datos, usa SQL familiar con una potente plataforma de análisis de Big Data.

Cloud Storage es el lugar donde se almacenan los archivos dentro de la nube de Google. *Cloud Dataproc* un servicio manejado por Hadoop y Spark, para procesar fácilmente grandes cantidades de datos, utilizando las herramientas potentes y abiertas en el ecosistema de Apache (Google Cloud Platform, 2017).

2.2.2. Oracle Big Data Appliance



Figura 2-8. Logotipo Oracle Big Data Appliance. (Oracle, 2017)

Desarrollador(es)	Oracle Corporation
Licencia	Propietaria
Fecha de lanzamiento	Octubre 3, 2011
Versión	12.1.2.7.0
Sitio Oficial	https://www.oracle.com/

Tabla 2-5. Oracle Big Data Appliance. (Flores A. & Villacís V., 2017).

Oracle Big Data Appliance es una solución de infraestructura que servirá tanto a empresas que poseen grandes proyectos de datos como a organizaciones que aumentan sus necesidades a lo largo del tiempo. Se centra en la idea de que puede ampliar su actual arquitectura de información empresarial para incorporar datos grandes (Oracle, 2015).

Es un sistema que combina el hardware con una pila de datos en el software para ofrecer una solución completa y fácil de implementar, para adquirir y organizar gran cantidad de datos. Posee una configuración de rack completo con 18 servidores, con una capacidad total de almacenamiento de 648TB (Oracle, 2017).

2.2.3. IBM Watson Analytics

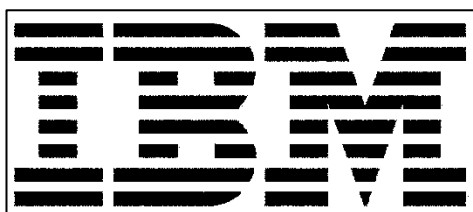


Figura 2-9. Logotipo IBM Watson Analytics. (IBM, 2017)

Desarrollador(es)	IBM
Licencia	Propietaria
Lenguaje de Programación	R, Java, Python
Versión Inicial	10.2.2 (2014)
Sitio Oficial	https://www.ibm.com/analytics/us/en/

Tabla 2-6. IBM Watson Analytics. (Flores A. & Villacís V., 2017).

Ofrece una variedad de soluciones Big Data y analítica para impulsar un negocio, permitiendo a las organizaciones trabajar con datos para responder a las preguntas de negocio más complicadas, descubrir patrones y poner en práctica ideas innovadoras. Posee un ecosistema integral de datos, funcionalidades cognitivas y analíticas, disponible en el cloud, on premise o en despliegues híbridos.

IBM Watson ayuda a las organizaciones descubrir información de valor permitiendo mejorar los procesos de negocio y las ideas que impulsan resultados innovadores. Las características principales son: Gestión de datos y analítica, servicio de datos en cloud, soluciones de analítica utilizando datos meteorológicos, analítica predictiva y visualización de datos creados a su medida.

2.3.Herramientas en la nube (Cloud Computing)

Las herramientas en la nube permiten manejar grandes cantidades de datos mediante lo que hoy conocemos por Internet, sin preocuparnos por su capacidad de almacenamiento. A continuación, se describirán las características más relevantes que tiene Big Data en cuanto a este tipo de herramientas.

2.3.1. Hadoop on-cloud

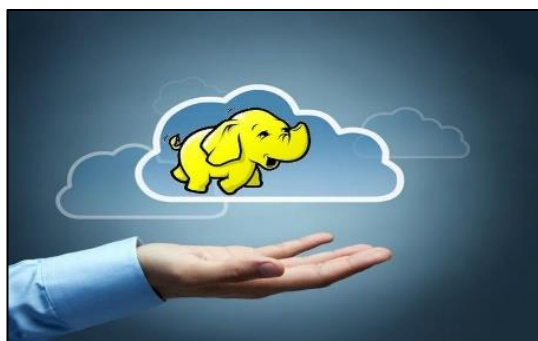


Figura 2-10. Logotipo Hadoop on-cloud. (Apache Hadoop, 2017).

Desarrollador(es)	Apache Software Foundation
Sistema Operativo	Multiplataforma (Windows, Linux, OS X)
Licencia	Apache License 2.0
Lenguaje de Programación	Java
Última Versión	Versión 2.0
Sitio Oficial	https://www.cloudera.com/products/cloud.html

Tabla 2-7. Hadoop on-cloud. (Flores A. & Villacís V., 2017).

Hasta el momento se ha podido determinar que Hadoop es un conjunto que está formado por varios componentes, que permiten aprovechar las características de Big Data en clúster y que para su correcto funcionamiento y disponibilidad del mismo se deben requerir de equipos físicos, pero el avance tecnológico que va en aumento cada día, ha permitido el cómputo en la nube, lo cual permite que ya no se requiera de dichos equipos físicos, los cuales pueden ser muy costosos.

Actualmente, no tiene importancia si se los encuentra en una habitación o son equipos virtuales localizados en cualquier parte del mundo, debido a que los componentes están diseñados para funcionar en equipos de costos reducidos. Esto da varios beneficios a aquellas pequeñas o medianas empresas y uno de ellos es que no requieren un costo de inversión inicial elevada, además utiliza máquinas virtuales alquiladas las cuales podrán ser pagadas por medio del tiempo en que estas sean utilizadas.

2.3.2. Microsoft Azure



Figura 2-11. Logotipo Microsoft Azure. (Microsoft, 2017).

Desarrollador(es)	Microsoft
Licencia	Microsoft CLUF (EULA)
Lenguaje de Programación	PHP 5.5, .NET
Género	SaaS (Software as a Service)
Sitio Oficial	https://azure.microsoft.com/

Tabla 2-8. Microsoft Azure. (Flores A. & Villacís V., 2017).

Microsoft Azure, es una plataforma que se encuentra en la nube de Microsoft, actúa como SaaS teniendo la ventaja de nube híbrida, es decir, se puede tener todos los centros de datos y la nube pública al mismo tiempo, de esta manera, actúa de forma inmediata. Azure nos da la accesibilidad de crear servicios y aplicaciones con cualquier dispositivo.

Otra característica a rescatar de esta herramienta, es que “obtiene la libertad de crear e implementar donde quiera, utilizando las herramientas, las aplicaciones y los marcos que prefiera” (Microsoft, 2017), así como también adaptarse a cualquier demanda, es decir, solo se paga por lo que se usa.

2.3.3. Amazon Web Services (AWS)¹⁰



Figura 2-12. Logotipo Amazon Web Services. (Amazon Web Services, 2017).

Desarrollador(es)	Amazon
Comercial	Si
Lanzamiento	2006
Estado Actual	Online
Género	Aplicación informática y servicio de alojamiento de Internet
Sitio Oficial	http://aws.amazon.com/

Tabla 2-9. Amazon Web Services. (Flores A. & Villacís V., 2017).

En 2016, comenzó a proporcionar servicios de infraestructura de TI para empresas en formas de servicios web, conocido en la actualidad como “Cloud Computing”.

Existen ciertos beneficios al utilizar esta herramienta, uno de ellos es la oportunidad de reemplazar importantes gastos anticipados en infraestructura con costos variables reducidos que se escalan en su negocio (Amazon Web Services, 2017).

Otro de los beneficios es que ya no tienen que planificar ni adquirir servidores y otra infraestructura de TI con semanas o meses de antelación, de esta manera, permitiendo disponer de dichos servidores en cuestión de minutos y con esto se obtienen resultados rápidamente (Amazon Web Services, 2017).

¹⁰ AWS (Amazon Web Services) Colección de servicios de computación en la ...

Es una plataforma de servicios en la nube el cual ofrece potencias de cómputo, almacenamiento de base de datos, entrega de contenido permitiendo a las empresas a escalar y crecer. Una de las ventajas de esta herramienta es que independiente del lenguaje y del sistema operativo, la empresa puede determinar la plataforma de desarrollo o el modelo de programación que tenga más sentido para su negocio.

Es tecnológicamente segura y duradera, con agilidad y elasticidad instantánea, proporcionando una infraestructura global y masiva en la cloud permitiendo la innovación, experimentación e iteración con una sorprendente rapidez. Es de bajo costo y se lo paga según su uso, es decir independientemente de lo que se use o si se utiliza pocas horas o todo un día (Amazon Web Services, 2017).

2.4. Cuadros comparativos de las herramientas de Big Data

A continuación se compara las diferentes herramientas de Big Data para posteriormente escoger, según nuestro criterio y conocimiento, la más adecuada para la estructura de la facultad.

2.4.1. Cuadro comparativo entre Apache Spark y Hadoop

SPARK	HADOOP
Trabaja in memory y, por lo tanto, todos los procesos se aceleran.	Los datos están en disco y eso hace que todo resulte más lento.
Necesita más memoria para el almacenamiento.	Las necesidades de almacenamiento son inferiores.
Su rendimiento puede verse mermado debido a la necesidad de utilizar aplicaciones pesadas.	Al ocuparse de eliminar los datos cuando no son ya necesarios, no produce pérdidas de rendimiento significativas para aplicaciones pesadas.
Spark necesita ejecutarse en HDFS para acceder a permisos de nivel de archivo.	Proporciona a sus usuarios todos los beneficios de los avances obtenidos en los proyectos de seguridad de Hadoop (Knox Gateway o Sentry son algunos ejemplos).

Tabla 2-10. Spark vs. Hadoop. (Flores A. & Villacís V., 2017).

2.4.2. Cuadro comparativo entre MongoDB y Spark

Nombre	MongoDB	Spark
DB-Engines Ranking (DB-Motores de Clasificación)	Puntuación: 335.00 Rango #5 En general #1 Almacén de Documentos	Puntuación: 10.50 Rango #37 En general
Sitio web	https://www.mongodb.org/	spark.apache.org
Desarrollador	MongoDB, Inc	Apache Software Foundation
Versión Inicial	2009	2014
Lanzamiento Actual	3.4.4, abril de 2017	v2.1.1, mayo de 2017
Licencia	Fuente abierta (AGPL versión 3)	Fuente abierta (Apache 2.0)
Lenguaje de implementación	C++	Scala, Java, Python, R
Sistemas Operativos	Multiplataforma	Multiplataforma
Triggers	No	No
Foreign Keys	No	No
Capacidades de memoria	Si (En memoria del motor de almacenamiento introducido con MongoDB versión 3.2)	No
Control de acceso	Derechos de acceso para usuarios y roles	No

Tabla 2-11. MongoDB vs. Spark. (Solid IT, 2017).

2.4.3. Cuadro comparativo entre Hadoop y MongoDB

Nombre	Hadoop	MongoDB
Tecnología	Java	C++ código nativo
Largest System (Sistema más grande)	4000 nodos, 100 TB	100 nodos, 5 TB
Uso típico	Análisis de informes, “map-reduce” en grandes conjuntos de datos.	Sistemas de alto volumen, conjunto de datos de tamaño moderado.

Tabla 2-12a. Hadoop vs. MongoDB. (Open Software Integrators, LLC, 2017).

Nombre	Hadoop	MongoDB
Estructura Empresarial	Sin fines de lucro (Apache), respaldado por múltiples proveedores.	Con fines de lucro (10gen)
Mejor Característica	Flexibilidad, implementación de “map-reduce”	Peso ligero, mejor rendimiento.
Mayor Limitación	Único punto de fallo.	Cada nodo tiene un único subproceso.
Madurez	Ampliamente implementado, pero menos desarrollado.	Más reciente, menos ampliamente implementado, pero más desarrollado.
Complejidad	Alta	Moderado.

Tabla 2-12b. Hadoop vs. MongoDB. (Open Software Integrators, LLC, 2017).

2.4.4. Cuadro comparativo entre AWS, Microsoft Azure y Google.

	Amazon Web Services	Microsoft Azure	Google Cloud Platform
Nacimiento Servicio	2006	2010	2011
Coste	Pago por hora o fracción. Descuentos en contrataciones de 1 0 3 años. Precio servidor pequeño: 39 € Precio servidor mediano: 134€	Pago por minutos. No tiene opciones de ahorro. Precio servidor pequeño: 43€ Precio servidor mediano: 180€	Pago por minutos. Descuentos proporcionales a las horas de consumo. Precio servidor pequeño: 35€ Precio servidor mediano: 143€

Tabla 2-13a. AWS vs. Azure vs. Google Platform. (apser. IT plug & play, 2015).

	Amazon Web Services	Microsoft Azure	Google Cloud Platform
Backups	Realiza 3 copias en misma zona geográfica. Publicidad de replicar copias a otras zonas.	Realiza 3 copias en misma zona geográfica. Publicidad de replicar copias a otras zonas.	Por defecto realiza las copias en todas las plataformas alrededor del mundo.
Disponibilidad Mundial	11 centros de datos. 37 puntos de distribución contenido.	20 centros de datos. 32 puntos de distribución contenido.	4 centros de datos. 160 puntos de distribución contenido.
Soporte	Soporte gratuito. Contacto por email en horas de oficina 49€/mes. Contacto en 24x7 con 1 hora de tiempo de respuesta. 10% de la facturación (min. 100€/mes).	Soporte gratuito. Contacto web en 24x7. 8 horas de tiempo de respuesta. 24.66€/mes. Contacto telefónico 24x7. 2horas de tiempo de respuesta. 256€/mes.	Soporte gratuito. 4 horas laborables de tiempo de respuesta. 150€/mes. 1 hora de tiempo de respuesta. 9% de la facturación (min. 40€/mes).
Tipos Servidores	53	25	18
Otros Servicios en la Nube	Almacenamiento. Bases de datos. DNS. VDI.	Almacenamiento. Bases de datos. Suite Ofimática. Correo Electrónico.	Almacenamiento. Bases de datos. Suite Ofimática. Correo Electrónico. Registro dominios y DNS.
Seguridad	20 certificaciones.	25 certificaciones.	6 certificaciones.

Tabla 2-13b. AWS vs. Azure vs. Google Platform. (apser. IT plug & play, 2015).

2.5. Base de datos Relacionales vs. Base de datos No Relacionales

2.5.1. Comparación en velocidad de los motores de base de datos

La comparación se realizó entre los motores de base de datos: MySQL, PostgreSQL y MongoDB; A continuación se efectúa pruebas para “crear 10000 (diez mil) tuits, encontrarlos (por su id único) y actualizar su texto (guardándolo en DB nuevamente), encontrar nuevamente los 10000 tuits y actualizar tres veces el contenido de cada uno, guardando en DB cada vez que se actualice y encontrar cada uno de los 10000 tuits (por su id) y eliminarlo” (Correa, 2013).

```
cuxibamba:postgresql macool$ ruby crud.rb
Rehearsal -----
creating 10000 tweets          35.680000    0.980000    36.660000 ( 42.242559)
find and update them          21.870000    1.220000    23.090000 ( 32.856776)
updating each one three times 63.530000    3.490000    67.020000 ( 93.805948)
removing them                  10.970000    0.960000    11.930000 ( 17.294459)
----- total: 138.700000sec

              user      system      total      real
creating 10000 tweets          32.210000    0.930000    33.140000 ( 38.820269)
find and update them          23.690000    1.350000    25.040000 ( 35.445241)
updating each one three times 70.610000    4.060000    74.670000 (106.387315)
removing them                  14.950000    1.150000    16.100000 ( 23.539710)
```

Figura 2-13. PostgreSQL CRUD. (Correa, 2013).

```
cuxibamba:mysql macool$ ruby crud.rb
Rehearsal -----
creating 10000 tweets          39.190000    0.800000    39.990000 ( 44.653293)
find and update them          21.950000    0.960000    22.910000 ( 28.654262)
updating each one three times 64.350000    2.740000    67.090000 ( 81.312197)
removing them                  13.400000    0.810000    14.210000 ( 18.313242)
----- total: 144.200000sec

              user      system      total      real
creating 10000 tweets          35.330000    0.730000    36.060000 ( 40.181477)
find and update them          23.100000    0.960000    24.060000 ( 29.710288)
updating each one three times 69.420000    2.920000    72.340000 ( 87.323573)
removing them                  17.130000    0.930000    18.060000 ( 22.551120)
```

Figura 2-14. MySQL CRUD. (Correa, 2013).

```

cuxibamba:mongodb macool$ ruby crud.rb
Rehearsal -----
creating 10000 tweets      5.050000  0.170000  5.220000 ( 5.212017)
find and update them     11.020000  0.520000  11.540000 ( 12.733068)
updating each one three times 34.080000  1.580000  35.660000 ( 39.534616)
removing them            8.170000  0.500000  8.670000 ( 9.995912)
----- total: 61.090000sec

                user      system      total      real
creating 10000 tweets      4.650000  0.160000  4.810000 ( 4.806377)
find and update them     15.530000  0.800000  16.330000 ( 18.129270)
updating each one three times 46.710000  2.410000  49.120000 ( 54.862238)
removing them            12.240000  0.760000  13.000000 ( 14.901368)
    
```

Figura 2-15. MongoDB CRUD. (Correa, 2013).

Los resultados obtenidos de las pruebas de velocidad de cada motor de base de datos realizadas anteriormente, están especificados en la siguiente tabla (Tabla 2-14):

Descripción (Tiempo=Segundos)	PostgreSQL	MySQL	MongoDB
Almacenar diez mil tuits.	38.82	40.18	4.81
Encontrar y actualizar diez mil tuits.	35.45	29.71	18.13
Encontrar los tuits y actualizar tres veces cada uno.	106.39	87.32	54.86
Eliminar los tuits.	23.54	22.55	14.9
TOTAL	204.2	179.76	92.7

Tabla 2-14. Resultado de las pruebas de BDD. (Correa, 2013).

CAPÍTULO 3. INGENIERÍA – SITUACIÓN ACTUAL

En este capítulo se describirá la situación actual de la Facultad de Ingeniería de la PUCE, teniendo en cuenta que los datos a analizar son de estudiantes, docentes y cátedras de la misma facultad como caso de estudio.

3.1. PUCE

“La PUCE, fundada en 1946, es la primera universidad privada del país y, por consiguiente, la más antigua. Está dirigida y administrada por la Compañía de Jesús (Jesuitas). Se la considera como una de las cinco principales universidades del país por sus programas de docencia, investigación y vinculación con la comunidad. Goza de prestigio entre los empleadores y los medios de comunicación” (Jesuitas Ecuador, 2017).

3.1.1. Misión

“LA PUCE:

1. es una comunidad académica que, de modo riguroso y crítico, contribuye a la tutela y desarrollo de la dignidad humana y de la herencia cultural mediante la investigación, la docencia y los diversos servicios ofrecidos a las comunidades locales, nacionales e internacionales

2. presta particular atención a las dimensiones éticas de todos los campos del saber y del actuar humano, tanto a nivel individual como social. En este marco, propugna el respeto a la dignidad y derechos de la persona humana y sus valores trascendentes, apoya y promueve la implantación de la justicia en todos los órdenes de la existencia, promueve la preservación del medio ambiente y el respeto a la vida

3. goza de la autonomía inherente a su condición de universidad, que le es necesaria para cumplir sus funciones eficazmente. Ejerce dicha autonomía con responsabilidad, y consiguientemente cumple con la rendición social de cuentas, tal y como lo determina la Ley

4. garantiza a sus miembros la libertad académica, salvaguardando los derechos de la persona y de la comunidad dentro de las exigencias de la verdad y del bien común

5. dirige su actividad hacia la formación integral del ser humano. Por ello trata de formar a sus miembros intelectual y éticamente para el servicio a la sociedad en el ejercicio profesional y en el compromiso con el desarrollo sustentable del país

6. pretende la integración del saber mediante el examen de la realidad con los métodos propios de cada disciplina académica y propiciando, al mismo tiempo, el diálogo entre estas para que se enriquezcan mutuamente

7. promueve el compromiso de todos los miembros de la comunidad universitaria para la consecución de los fines institucionales a través del diálogo y la participación, de conformidad con el presente Estatuto h) como universidad particular ofrece una alternativa específica en el ámbito académico conforme a su propio Estatuto y reglamentos, e

8. como universidad católica, se inspira en los principios cristianos; propugna la responsabilidad del ser humano ante Dios, el respeto a la dignidad y derechos de la persona humana y a sus valores trascendentales; apoya y promueve la implantación de la justicia en todos los órdenes de la existencia; propicia el diálogo de las diversas disciplinas con la fe, la reflexión sobre los grandes desafíos morales y religiosos, y la praxis cristiana” (PUCE, 2017).

3.1.2. Visión

“La PUCE, es el referente nacional en formación integral e inclusiva con impacto social. La innovación, agilidad y compromiso identifican su cultura organizacional.

Es reconocida internacionalmente por su producción científica y la calidad de sus estudiantes y docentes” (PUCE, 2017).

3.1.3. Valores institucionales

- “1. La responsabilidad del ser humano ante Dios
2. La excelencia académica
3. El respeto
4. La confianza mutua
5. La equidad y la justicia
6. La responsabilidad compartida
7. El reconocimiento del otro
8. La valoración académica
9. La transparencia
10. El respeto a la vida” (PUCE, 2017).

3.1.4. Sedes PUCE

La PUCE cuenta con cinco sedes las cuales son Sede Ambato, Sede Esmeraldas, Sede Santo Domingo, Sede Ibarra y Sede Manabí (PUCE, 2017).

3.1.5. Organigrama Estructural – PUCE

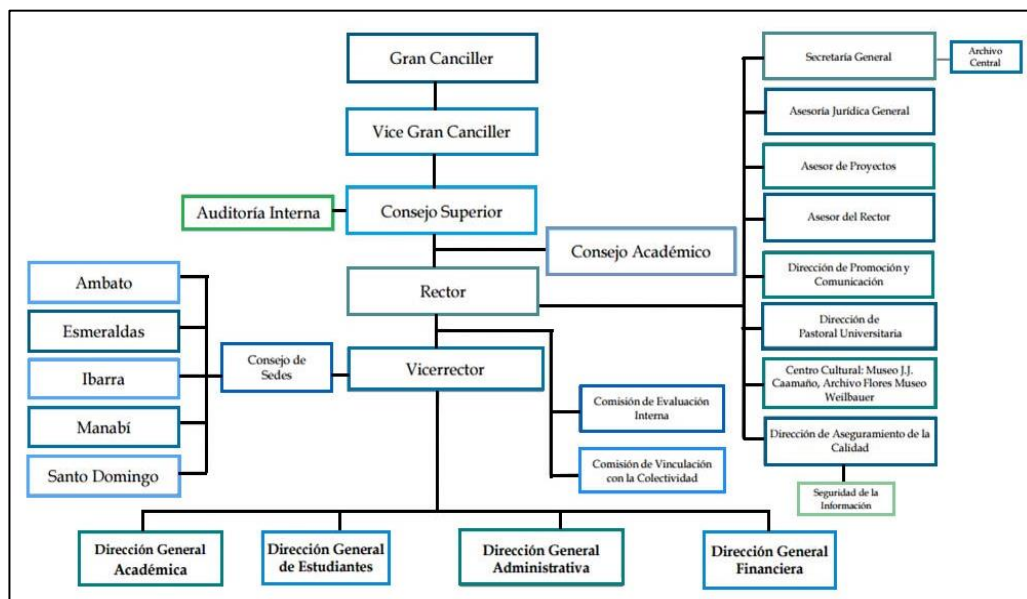


Figura 3-1. Organigrama Estructural – Sede Quito. (PUCE, 2017).

3.1.6. Autoridades PUCE Quito

CARGO	RESPONSABLE
Gran Canciller	Excelentísimo Monseñor Fausto Trávez Trávez, OFM
Vice Gran Canciller	Máster Gustavo Andrés Calderón Schmidt SJ.
Rector	Doctor Fernando Ponce León, SJ.
Vicerrector	Doctor Fernando Barredo Heinert, SJ.
Directora General Académica	Doctora Graciela Monesterolo Lencioni
Director General Administrativo	Máster Javier España Mera (E)
Director General de Estudiantes	Ingeniero Juan Carlos Latorre
Director General Financiero	Máster Javier España Mera
Secretario General	Doctor Santiago Jaramillo Herdoiza

Tabla 3-1. Autoridades PUCE Quito. (PUCE, 2017).

3.1.7. Oferta Pregrado

FACULTAD	CARRERAS EN CURSO	CARRERAS EN PROCESO DE APROBACIÓN
Arquitectura, Diseño y Artes	Arquitectura	Arquitectura
	Artes Visuales	Artes Visuales
	Diseño	Diseño Gráfico
		Diseño de Productos
Ciencias Administrativas y Contables	Administración de Empresas	Administración de Empresas
	Contabilidad y Auditoría	Contabilidad y Auditoría
		Negocios internacionales
		Finanzas
		Mercadotecnia
Ciencias de la Educación	Educación (mención Educación Inicial)	Educación Inicial
Ciencias Exactas y Naturales	Ciencias biológicas	Biología
	Ciencias químicas	Química
	Microbiología	Microbiología
Ciencias Filosófico-Teológicas	Ciencias Filosóficas	Filosofía
	Ciencias Teológicas	Teología
Ciencias Humanas	Antropología Sociocultural	Arqueología
	Arqueología	Antropología
	Ciencias Geográficas y Medio Ambiente	Turismo
	Ciencias Geográficas y Planificación Territorial	Hospitalidad y Hotelería
	Ciencia Históricas	Historia
	Ecoturismo	Sociología
	Gestión Hotelera	Ciencia Políticas
	Historia	Relaciones Internacionales

Tabla 3-2a. PUCE, Oferta académica actual. (PUCE, 2017).

FACULTAD	CARRERAS EN CURSO	CARRERAS EN PROCESO DE APROBACIÓN
Ciencias Humanas	Historia del Arte	
	Sociología mención en Desarrollo	
	Enseñanza de lenguas	
	Traducción	
	Negocios y Relaciones internacionales	
Economía	Economía	Economía
Enfermería	Enfermería	Enfermería
	Nutrición humana	Nutrición y Dietética
	Terapia física	Fisioterapia
Ingeniería	Ingeniería civil	Ingeniería civil
	Ingeniería en sistemas	Tecnologías de información
		Sistemas de información
Jurisprudencia	Derecho	Derecho
Medicina	Medicina	Medicina
	Bioquímica Clínica	Laboratorio Clínico
Psicología	Psicología educativa	Psicología
	Psicología clínica	Psicología clínica
	Psicología organizacional	
Trabajo Social	Gestión Social	Gestión Social

Tabla 3-2b. PUCE, Oferta académica actual. (PUCE, 2017).

3.1.8. Estudiantes Pregrado Nuevos y Antiguos (2016-02)

Tipo de Estudio	Ecuatoriano		Extranjero		Total
	ANTIGUO	NUEVO	ANTIGUO	NUEVO	
CURSOS ABIERTOS	72	99	55	108	334
POSTGRADO	1.040	36	7		1.083
PREGRADO	8.480	1.177	37	36	9.730
	9.592	1.312	99	144	11.147

Tabla 3-3. Estudiantes Semestre 2016-02. (Dirección Informática PUCE, 2017).

3.2. Facultad de Ingeniería

La Facultad de Ingeniería de la PUCE, forma profesionales de alto nivel, capacitados para satisfacer las necesidades de la sociedad y de las empresas, está conformada por las escuelas de Ingeniería Civil e Ingeniería en Sistemas; fundada en 1961, siendo como precursora la escuela de Civil.

3.2.1. Misión PUCE

“Investigación, transmisión del saber y saber hacer, servicio en beneficio de la comunidad a la que pertenece y se debe” (PUCE, 2017).

3.2.2. Visión PUCE

“Fundamentado en el pensamiento y en las directrices pedagógicas ignacianas se consolidará como un sistema nacional integrado (SINAPUCE) competitivo y autosostenible con infraestructura tecnológica de vanguardia” (PUCE, 2017).

3.2.3. Compromiso PUCE

“Ser una comunidad de maestros, estudiantes y trabajadores, atraídos por un mismo imán: La búsqueda del saber y la verdad” (PUCE, 2017).

3.2.4. Características distintivas de la Facultad de Ingeniería

Desde el punto de vista académico, la Escuela de Ingeniería Civil está dividida en áreas académicas, en las que se agrupan las materias con contenido afín desde el punto de vista de contenidos. Las áreas académicas existentes son: Matemáticas, Física, Construcciones, Hidráulica Sanitaria, Geotecnia, Transportes y vías, Estructuras y Formación general.

La Escuela de Ingeniería de Sistemas, en cambio, está compuesta por las siguientes áreas académicas: Matemáticas, Hardware, Programación, Software de Base, Desarrollo de Sistemas, Teoría de Sistemas, Aplicaciones Matemáticas y Formación General.

Cada una de las áreas está compuesta por un profesor coordinador, y con tareas, objetivos y fines específicos.

La Facultad de Ingeniería está apoyada por laboratorios con equipos y tecnología de punta, que sirven como complemento a la formación académica que se recibe en aulas. Los laboratorios que se encuentran a disposición de los estudiantes son:

Escuela de Ingeniería Civil:

La escuela de civil consta con los siguientes laboratorios: Laboratorio de Resistencia de Materiales, Laboratorio de Suelos y Geotecnia, Laboratorio de Hormigones y Gabinete de Topografía (PUCE, 2017).

Escuela de Ingeniería de Sistemas

La escuela de sistemas consta con los siguientes laboratorios: Laboratorio de Computación, que cuenta con equipos y programas para bases de datos, sistemas operativos, multiplataforma para desarrollo de sistemas y Laboratorio de Redes, con equipos LAN, WAN para interconectividad y comunicaciones (PUCE, 2017).

3.2.5. Carreras de Ingeniería

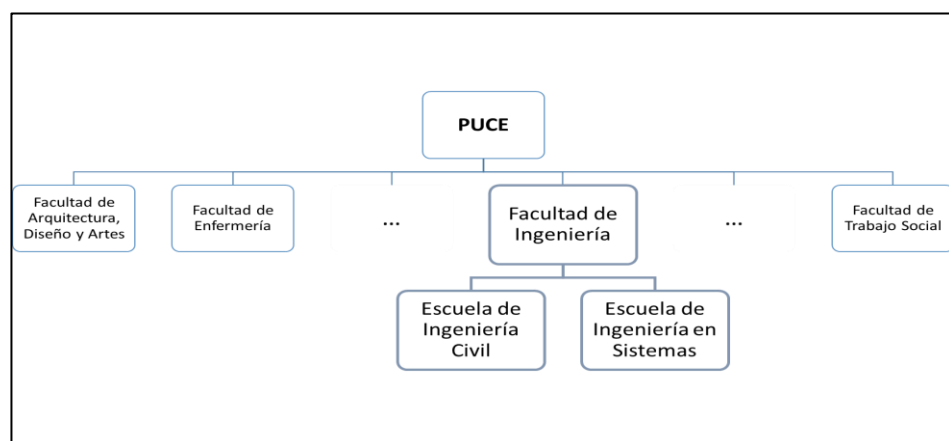


Figura 3-2. Carreras Ingeniería PUCE. (Flores A. & Villacís V., 2017).

3.2.6. Autoridades de la Facultad de Ingeniería

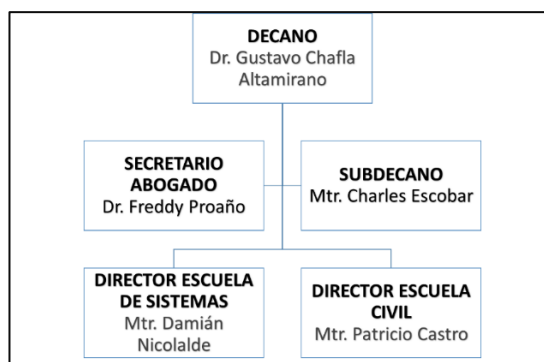


Figura 3-3. Autoridades Ingeniería. (Flores A. & Villacís V., 2017).

3.2.7. Estudiantes de la Facultad de Ingeniería Semestre 2016-02.

FACULTAD	CARRERA	ANTIGUO	NUEVO	TOTAL
Facultad de Ingeniería	Ingeniería Civil	605	50	655
	Ingeniería de Sistemas y Computación	247	32	279
SUBTOTAL		852	82	934

Tabla 3-4. Estudiantes Facultad Ingeniería. (Dirección Informática PUCE, 2017).

3.2.8. Número de Alumnos Facultad de Ingeniería (5 últimos años)

Semestres	Facultad de Ingeniería				Total
	Ingeniería Civil		Ingeniería de Sistemas y Computación		
	Antiguos	Nuevos	Antiguos	Nuevos	
2012-01	424	81	245	65	815
2012-02	473	63	268	45	849
2013-01	495	52	273	38	858
2013-02	525	65	276	43	909
2014-01	548	74	284	44	950
2014-02	561	53	270	30	914
2015-01	548	63	272	33	916
2015-02	560	59	275	28	922
2016-01	583	110	257	37	987
2016-02	605	50	247	32	934
TOTAL	5992		3062		9054

Tabla 3-5. Alumnos últimos 5 años Facultad de Ingeniería. (PUCE, 2017).

3.3. Ingeniería de Sistemas y Computación

La Escuela de Ingeniería en Sistemas y Computación, fundada en 1961, comenzó siendo lo que ahora conocemos como la Facultad de Ingeniería de la PUCE, teniendo en el presente año 56 años de fundación.

3.3.1. Misión

“La carrera de Ingeniería de Sistemas y Computación de la PUCE es un conjunto de estudios integrados sistemáticamente, que permite formar a los estudiantes en un marco de principios cristianos, con pedagogía ignaciana e investigación formativa, habilitando a sus egresados para el ejercicio de la ingeniería en las áreas de las tecnologías de la información y la comunicación; de una manera profesional, responsable, buscando la innovación y el servicio a la sociedad” (PUCE, 2017).

3.3.2. Visión

“En el año 2017 la carrera de Ingeniería de Sistemas y Computación de la PUCE mantiene su acreditación académica nacional e internacional” (PUCE, 2017).

3.3.3. Objetivos

- “Formar Ingenieros en Sistemas con sólidos conocimientos técnicos en diferentes áreas de: telecomunicaciones, sistemas de información, permitiendo cubrir la demanda de profesionales en el ámbito privado y público.
- Formar profesionales con capacidad objetiva, analítica y crítica, comprometidos en mejorar el nivel de vida en el País” (PUCE, 2017).

3.3.4. Perfil profesional

“El Ingeniero en Sistemas es un profesional capaz y comprometido con el desarrollo de la sociedad por su ingenio en el desarrollo de sistemas de información. El cambio continuo de la tecnología obliga a que el profesional gestione y aproveche los recursos tecnológicos disponibles” (PUCE, 2017).

3.3.5. Campo laboral

“El Ingeniero en Sistemas puede desempeñarse en cualquier tipo de organización o empresa en el departamento de sistemas, planificando, diseñando, desarrollando, auditando e implantando sistemas de información en intranets, extranets, redes LAN y WAN. En general, el ingeniero en sistemas se desempeña como el arquitecto y el administrador de las tecnologías de la información. Además, su formación le permite ejercer exitosamente la función de gerente” (PUCE, 2017).

3.4. Ingeniería Civil

La Escuela de Ingeniería Civil, fundada en 1961, comenzó siendo lo que ahora conocemos como la Facultad de Ingeniería de la PUCE, teniendo en el presente año 56 años de fundación.

3.4.1. Misión

“La carrera de Ingeniería Civil es una comunidad académica de vanguardia en el Ecuador que, inspirada en el espíritu de la Pontificia Universidad Católica del Ecuador, forma con rigor científico, proyección técnica y espíritu emprendedor, profesionales capacitados para cumplir labores de planificación, diseño y construcción de obras de infraestructura para el desarrollo del país” (PUCE, 2017).

3.4.2. Visión

“En el 2018, la Carrera de Ingeniería Civil de la PUCE, fundamentada en el pensamiento y en las directrices pedagógicas ignacianas, será reconocida nacional e internacionalmente por su gestión ética en servicio de la comunidad; por una estructura académica moderna para la formación de profesionales con responsabilidad social; por los resultados de su investigación científica y por realizar su gestión con el apoyo de un sistema técnico, innovador y efectivo, con procesos eficientes y recursos humanos capacitados y comprometidos con la misión institucional” (PUCE, 2017).

3.4.3. Campo Laboral

“Nuestro Ingeniero Civil se constituye en promotor del desarrollo nacional, a través de habilidades y competencias adquiridas durante su formación profesional, que le permiten incursionar con solvencia en los siguientes campos:

Ejecución de diseños estructurales, hidráulicos, sanitarios, estudios de suelos, geotecnia, topografía, tráfico y transportes, pavimentos, trazado de vías, cimentaciones, puentes, fiscalización de obras civiles, ejecución y administración de proyectos de construcción, consultoría en los diversos componentes de obras de Ingeniería Civil, programación de obra, elaboración de presupuestos de construcción, evaluación de proyectos de Ingeniería, entre otros” (PUCE, 2017).

3.4.4. Características del Ingeniero Civil

- “El graduado de Ingeniería Civil fundamenta sus acciones en una sólida formación científica, en la que las ciencias básicas constituyen el pilar fundamental y el cimiento sobre el que descansa la tecnología de la Ingeniería.
- Debe poseer una sólida cultura informática, basada en la excelencia en el conocimiento y manejo de los recursos computacionales disponibles.
- Tiene la capacidad de discernir la interrelación entre el entorno social y la tecnología, de tal manera que está facultado para definir los problemas e existentes, plantear las posibles soluciones, analizar y predecir su comportamiento e implicaciones, y seleccionar la mejor solución, todo dentro de un marco de profunda conciencia social.
- Es capaz de actuar con responsabilidad social, velando por el adecuado equilibrio entre los actores del desarrollo: las personas, recursos materiales, y medio ambiente. Tiene capacidad para orientar su acción hacia el mejoramiento del nivel de vida de la población, sin afectar negativamente su entorno.

- Tiene conocimiento de elementos de gestión administrativa que le permiten ejercer un liderazgo en la conducción de grupos humanos y el manejo óptimo de los recursos, orientados a lograr calidad, productividad y competitividad.
- Tiene una sólida formación moral y ética, que le califique para el manejo de los recursos materiales y humanos que estarán a su disposición. Es consciente de su compromiso moral con la sociedad, sus colegas de la profesión y sus empleados” (PUCE, 2017).

CAPÍTULO 4. BIG DATA PARA INGENIERÍA

En este capítulo, se maneja toda la información relacionada con estudiantes y docentes de la Facultad de Ingeniería, por lo tanto, a continuación se definirá los requerimientos para determinar las herramientas, y posteriormente generar y diseñar los cubos de información que nos permitirá hacer el análisis que se necesita.

4.1. Requerimientos

Los requerimientos obtenidos fueron consultados anteriormente con algunas de las autoridades principales de la Facultad de Ingeniería, como el anterior y el actual Director de Carrera, los cuales se especifican a continuación:

- Determinar en qué cátedra se desenvuelven mejor cada docente de la facultad.
- Conocer en qué periodo los docentes de la facultad tuvieron mejor desempeño.
- Conocer el desempeño de los estudiantes por periodo.
- Determinar el cambio de dedicación a la facultad por los diferentes años.
- Establecer en qué cátedra se desempeña mejor un curso de estudiantes de la facultad.
- Establecer en qué cátedra se desempeña mejor cada estudiante de la facultad.
- Conocer el porcentaje de graduados por escuela.
- Conocer el porcentaje o promedio de estudiantes por periodo y por género.
- Determinar la diferencia de estudiantes entre la escuela de civil y de sistemas.
- Determinar el nivel académico de los docentes por periodo.
- Conocer las variantes del número de estudiantes por periodo y por escuela.
- Determinar la cantidad de estudiantes por periodo y por cátedra.

4.2. Diseño

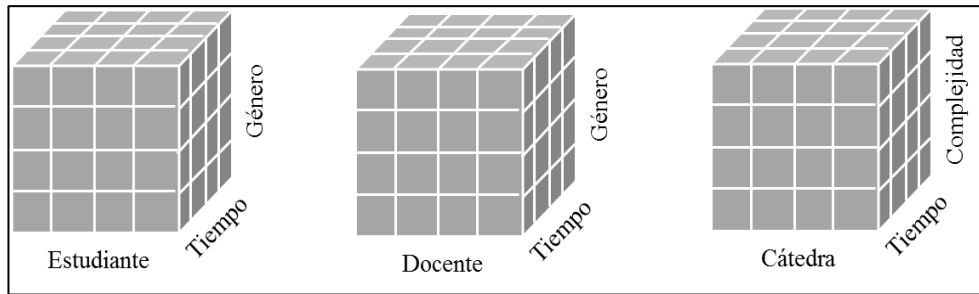


Figura 4-1. Ejemplo de Cubos OLAP. (Flores A. & Villacís V., 2017).

En esta etapa se ejemplificó de la mejor las posibles consultas generales realizadas con los datos que se tienen, como podemos observar en la **Figura 4-1**.

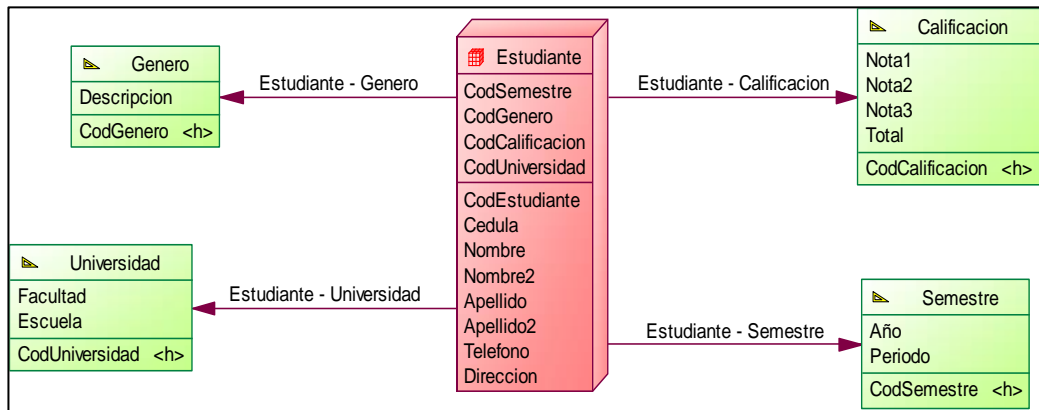


Figura 4-2. Cubo Estudiante. (Flores A. & Villacís V., 2017).

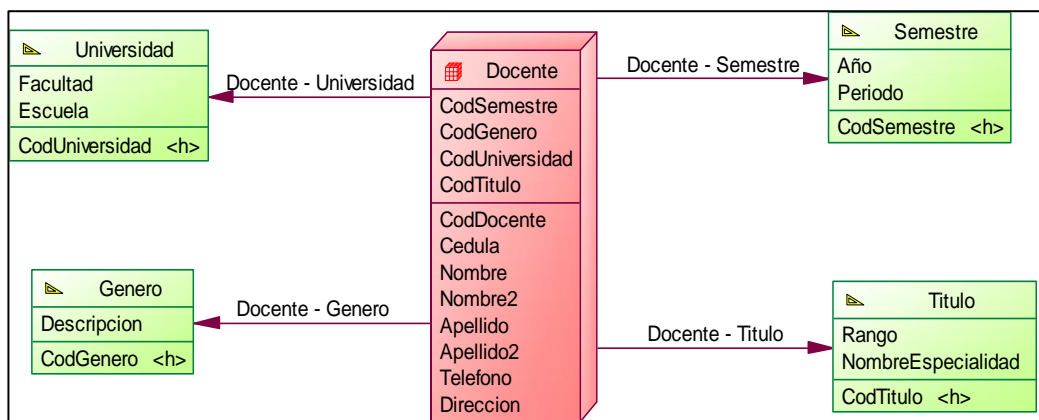


Figura 4-3. Cubo Docente. (Flores A. & Villacís V., 2017).

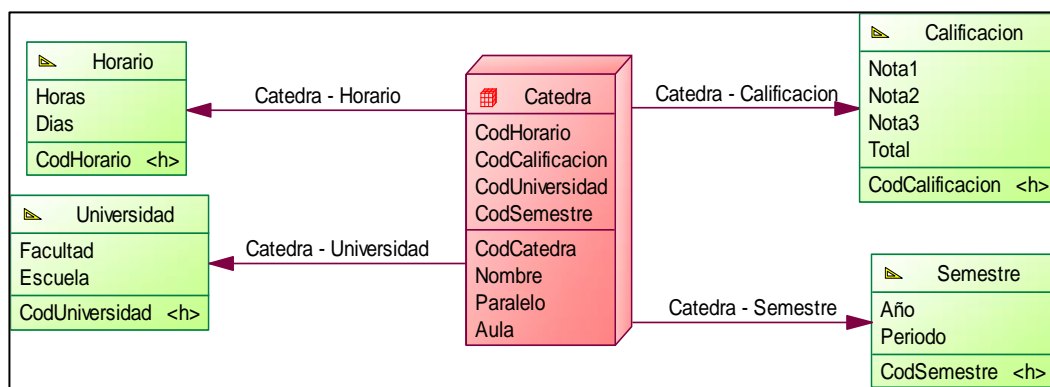


Figura 4-4. Cubo Cátedra. (Flores A. & Villacís V., 2017).

Como se observa en la *Figura 4-2*, *Figura 4-3* y *Figura 4-4*, los cubos multidimensionales son una representación visual de los datos que se almacenan, cada entidad o cubo posee sus respectivas dimensiones que serán los datos a consultar de esa entidad.

4.3. Implementación

- Como una opción antes de la recepción de datos se implementará con datos de prueba realizados en Microsoft Excel 2013.
- Se implementará usando la herramienta de MongoDB con los datos proporcionados.
- Para el diseño estructural de la información almacenada, se usará Power Designer versión 16.5.

4.4. Herramienta MongoDB

- Se abre la terminal

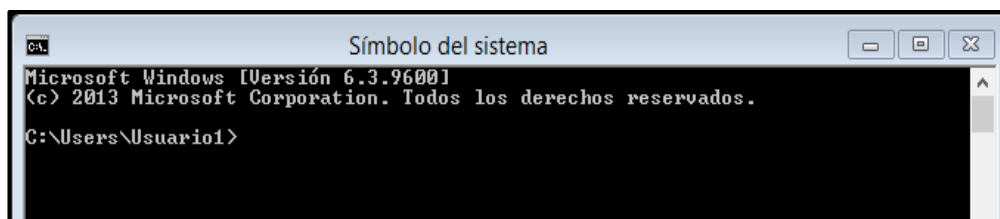
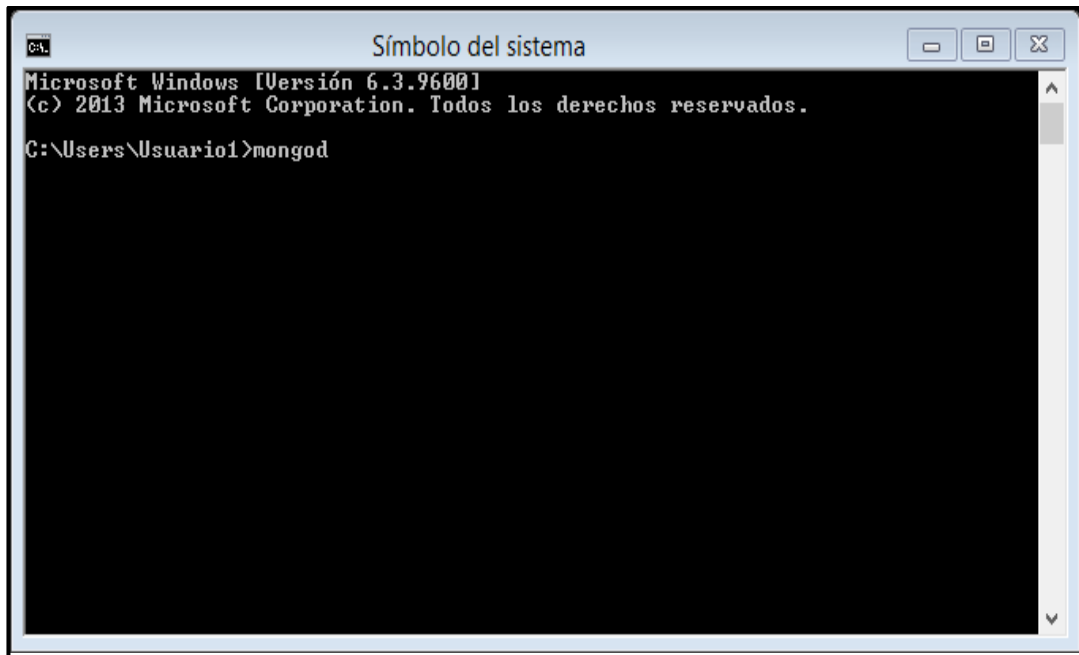


Ilustración 4-1. Inicio terminal. (Flores A. & Villacís V., 2017).

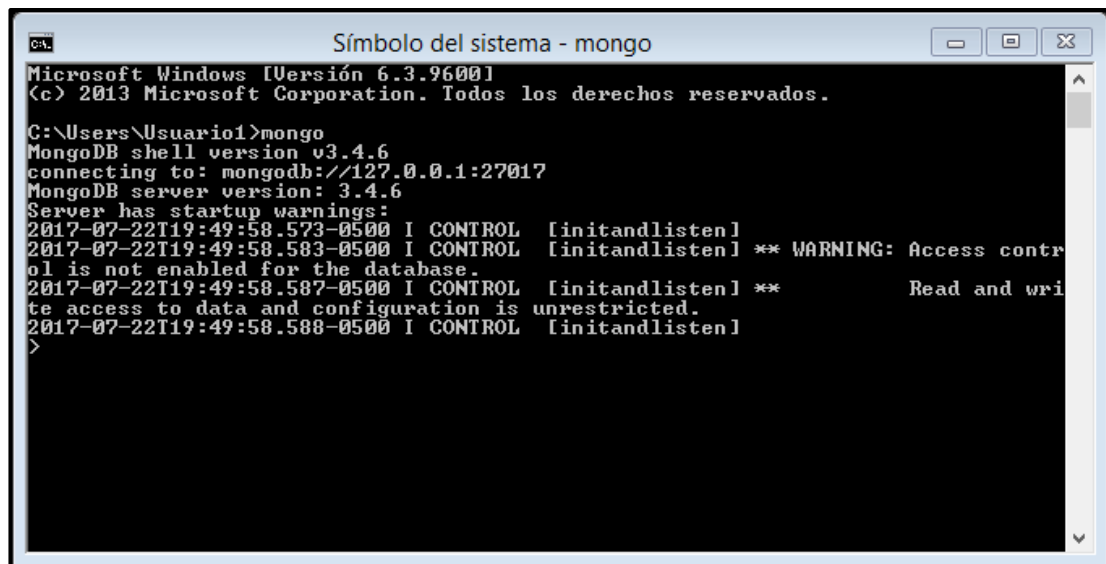
- Y se escribe “mongod”, esperando que se abra la conexión



```
C:\> mongod
```

Ilustración 4-2. Proceso principal MongoDB. (Flores A. & Villacís V., 2017).

- En otro terminal escribimos “mongo”



```
C:\> mongo
MongoDB shell version v3.4.6
connecting to: mongodb://127.0.0.1:27017
MongoDB server version: 3.4.6
Server has startup warnings:
2017-07-22T19:49:58.573-0500 I CONTROL [initandlisten]
2017-07-22T19:49:58.583-0500 I CONTROL [initandlisten] ** WARNING: Access control
ol is not enabled for the database.
2017-07-22T19:49:58.587-0500 I CONTROL [initandlisten] **      Read and wri
te access to data and configuration is unrestricted.
2017-07-22T19:49:58.588-0500 I CONTROL [initandlisten]
>
```

Ilustración 4-3. Proceso secundario MongoDB. (Flores A. & Villacís V., 2017).

4.4.1. Generación de base de datos con información de prueba

Cada una de las bases de datos generadas serán en Excel y guardadas con formato csv.

- *Estudiante*

CodEstudiante	Cedula	Nombre	Nombre2	Apellido	Apellido2	Fecha de Nacimiento	Genero	Direccion	Telefono	Año	Periodo	Nota 1	Nota 2	Nota 3	Total	Facultad	Escuela
1	1709281902	Juan	Andrés	Dávalos	Cisneros	02/01/1992	Masculino	Ponceano Al	2809891	2013	1	13	13	15	41	Ingeniería	Civil
2	1720210020	Daniela	Rafaela	Hernández	Jimenez	10/04/1990	Femenino	Mariana de	2301910	2012	2	12	12	14	38	Ingeniería	Sistemas
3	1756278190	Camilo	Stefano	Andrade	Suarez	01/02/1993	Masculino	Valle de los	2790191	2014	1	10	10	10	36	Ingeniería	Civil
4	1709101912	Carlos	Roberto	Centro	Ramirez	02/09/1992	Masculino	Amazonas y	2901920	2011	1	9	9	19	37	Ingeniería	Civil
5	1729219913	Carolina	Jacinta	Alarcón	López	10/06/1990	Femenino	La Floresta	2993810	2010	2	15	15	20	50	Ingeniería	Civil
6	1782910013	Javier	Leonardo	Flores	Dualla	01/02/1989	Masculino	6 de Diciembre	2981911	2015	1	10	10	14	34	Ingeniería	Civil
7	1782819121	Luis	Alberto	Balderas	Flores	08/11/1990	Masculino	Amazonas y	2618181	2013	1	11	11	12	34	Ingeniería	Civil
8	1763212212	Laura	Elizabeth	Baltazar	Cedeño	11/02/1989	Femenino	Av. América	2891091	2016	2	12	12	16	40	Ingeniería	Civil
9	1704619335	Humberto	Alejandro	Blanco	Velasco	17/07/1990	Masculino	Colón y Rein	2900191	2014	2	13	13	17	43	Ingeniería	Civil
10	1722417134	Juan	José	Cantú	Treviño	18/01/1995	Masculino	Av. Gran Col	2781910	2014	1	15	15	17	47	Ingeniería	Civil
11	173431223	Darío	Carlos	Contreras	Pavía	15/04/1990	Masculino	Sector la Ma	2918991	2012	1	10	10	18	38	Ingeniería	Sistemas
12	172617128	Natalia	Esperanza	Cortés	Trujillo	17/08/1990	Femenino	Chimborazo	2471819	2013	1	9	9	19	37	Ingeniería	Sistemas
13	1788919103	Héctor	Alejandro	Cruz	González	13/11/1990	Masculino	Sector Univ	2891911	2011	2	14	14	18	46	Ingeniería	Sistemas
14	1728181919	Claudia	Valeria	Dávila	Montero	18/01/1991	Femenino	La Gasca	2879909	2011	1	14	14	17	45	Ingeniería	Sistemas
15	1788838800	José	Luis	Cuevas	Nolasco	14/03/1995	Masculino	Estadio de l	2888651	2012	2	10	10	16	36	Ingeniería	Sistemas

Ilustración 4-4. Datos Prueba Estudiante. (Flores A. & Villacís V., 2017).

- *Docente*

CodDocente	Cedula	Nombre	Nombre2	Apellido	Apellido2	Fecha de Nacimiento	Genero	Direccion	Telefono	Dedicacion
1	1709266902	Luis	Abel	Cruz		01/02/1980	Masculino	Sector Ilaquito	2877181	PHD
2	1782929102	Sergio	Ignacia	Corrales	Carmona	02/07/1976	Masculino	Av. 10 de Agosto y Corea	2891929	Master
3	1786281919	Carlos	Rafael	Durán	Suárez	11/05/1979	Masculino	América y Mañosa	2458201	Master
4	1781728189	César	Luis	Erazo	Bernal	17/04/1970	Masculino	La Mañosa	2677282	Master
5	1728288821	Francisco	Enrique	Escobar	Beltrán	11/01/1972	Masculino	América y Naciones Unidas	2549299	PHD
6	1798298911	Maria	Enriqueta	Fernández	Martínez	10/09/1981	Femenino	La Shyris Y Portugal	2734581	Master
7	1776543719	Jorge	Alberto	Figuerola	Vallé	18/07/1964	Masculino	Eloy Alfaro y Gaspar de Villaroel	2987182	Master
8	1763222212	Alma	Estela	Flores	Martínez	19/08/1967	Femenino	Las Palmeras y Eloy Alfaro	2372940	Master
9	1756526711	Mario	Fernando	Gallardo	Vara	18/05/1973	Masculino	Simon Bolívar y 17 de Septiembre	2340029	Master
10	1722538392	Juan	Pablo	García	Ledesma	15/02/1966	Masculino	Abdón Calderón y 24 de Mayo	2439201	Especializado
11	1773431223	José	Alfonso	García	López	12/12/1970	Masculino	Las Camellas y esq.	2859200	Especializado
12	1786292912	Norma	Nelia	Villacís	Cruz	11/03/1973	Femenino	De los Pinos	2343212	Especializado
13	1786281819	Sarah	Cristina	Escobedo	Torres	13/01/1978	Femenino	Diego de Vásquez y Rosa Campuzano	2443908	Especializado
14	1726382921	Victoria	Guadalupe	Martínez	Endara	18/07/1965	Femenino	La Prensa y Gustavo Lemos	2674100	Especializado
15	1762562819	Clemente	Federico	Delgado	Sánchez	14/01/1963	Masculino	Machala y Cuicocha	2389703	Especializado

Ilustración 4-5. Datos Prueba Docente. (Flores A. & Villacís V., 2017).

- *Cátedra*

Nombre	Cod_Docente	Horario	Dias	Paralelo	Anio	Periodo	Cod_Estudiante
Física	LCruz289	7:00-9:00	Lunes/Martes	1	2013	1	JDávalos898
Matemática	SCorral282	9:00-11:00	Martes/Miércoles	2	2014	2	DHernandez777
Ética	CDurán246	16:00-18:00	Viernes	1	2012	1	CAndrade267
Ética	CEraza261	18:00-20:00	Lunes/Martes	3	2011	1	CCastro243
Química	FEscobar258	9:00-11:00	Martes/Miércoles	2	2010	2	CAlarcón111
Matemáticas discretas	MFernández278	9:00-11:00	Lunes/Martes	1	2015	1	JFlores234
Química	JFigueroa296	12:00-14:00	Miércoles/Jueves	2	2013	1	LBalderas342
Jesucristo	AFlores232	7:00-9:00	Miércoles/Jueves	2	2016	2	LBaltazar287
Matemáticas discretas	MGallardo236	7:00-9:00	Lunes/Martes	3	2014	2	HBlanco438
Jesucristo	JGarcía242	7:00-9:00	Jueves/Viernes	1	2014	1	JCantu432
Jesucristo	JGarcía284	16:00-18:00	Lunes/Martes	2	2012	1	Dcontreras190
Hidrología	NVillacís236	16:00-18:00	Lunes/Martes	1	2013	1	NCortés490
Matemáticas discretas	SEscobedo246	16:00-18:00	Lunes	3	2011	2	HCruz103
Hidrología	VMartínez266	12:00-14:00	Lunes/Martes	1	2011	1	CDávila290
Hidrología	CDelgado232	12:00-14:00	Miércoles	1	2012	2	JCuevas132

Ilustración 4-6. Datos Prueba Cátedra. (Flores A. & Villacís V., 2017).

- Importamos las respectivas bases de datos realizadas en Excel con formato CSV a MongoDB, en una base llamada “FacultadIngeniería”, se debe copiar la ubicación en donde se encuentren nuestras bases de datos (Estudiante, Docente, Cátedra); en el terminal ponemos el siguiente código:

```

C:\> 2017-07-22T20:19:21.602-0500 I NETWORK [initandlisten] shutdown: going to close
listening sockets...
2017-07-22T20:19:21.603-0500 I NETWORK [initandlisten] shutdown: going to flush
diaglog...
2017-07-22T20:19:21.603-0500 I CONTROL [initandlisten] now exiting
2017-07-22T20:19:21.603-0500 I CONTROL [initandlisten] shutting down with code:
100

C:\Users\Usuario1>mongoimport -d FacultadIngenieria -c Estudiante --type csv --file
C:\Users\Usuario1\Documents\PUCE\Tesis\BaseDeDatos\Estudiante.csv --headerline
2017-07-22T20:20:21.093-0500 connected to: localhost
2017-07-22T20:20:22.500-0500 imported 15 documents

C:\Users\Usuario1>mongoimport -d FacultadIngenieria -c Docente --type csv --file
C:\Users\Usuario1\Documents\PUCE\Tesis\BaseDeDatos\Docente.csv --headerline
2017-07-22T20:21:06.386-0500 connected to: localhost
2017-07-22T20:21:06.905-0500 imported 15 documents

C:\Users\Usuario1>mongoimport -d FacultadIngenieria -c Catedra --type csv --file
C:\Users\Usuario1\Documents\PUCE\Tesis\BaseDeDatos\Catedra.csv --headerline
2017-07-22T20:21:24.456-0500 connected to: localhost
2017-07-22T20:21:26.182-0500 imported 15 documents

C:\Users\Usuario1>
    
```

Ilustración 4-7. Importe de datos. (Flores A. & Villacís V., 2017).

- Se puede observar que en Robomongo la base de datos ha sido generada, con sus respectivas colecciones.

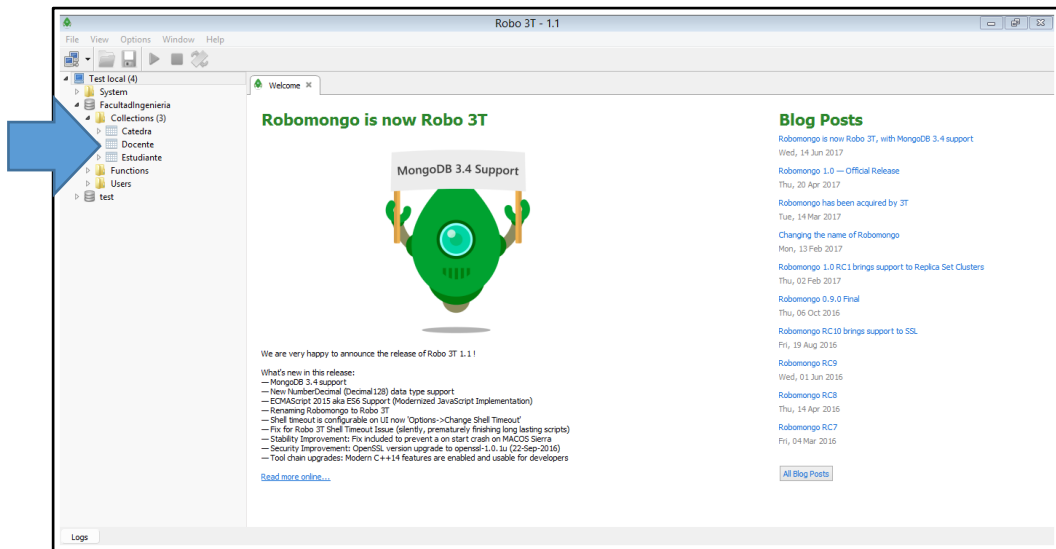


Ilustración 4-8. Datos importados en Robo 3T. (Flores A. & Villacís V., 2017).

- Los datos se pueden observar en Robomongo o por terminal, representados en las siguientes figuras:
 - *Estudiante*

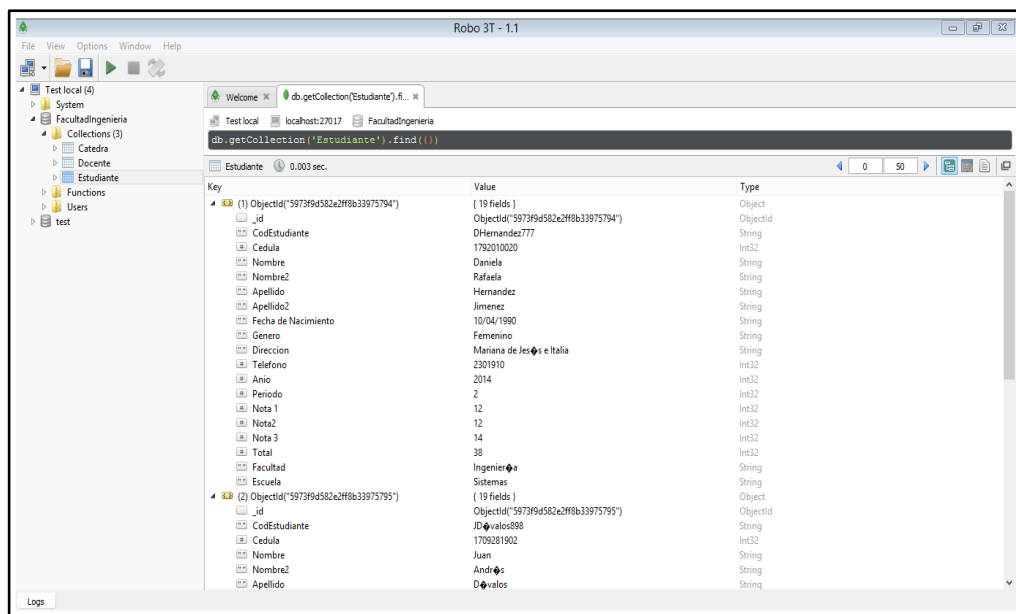


Ilustración 4-9. Datos Estudiante Robo 3T. (Flores A. & Villacís V., 2017).

Para poder mostrar los datos en MongoDB, se lo puede hacer por medio de consola, ingresando el comando que está representado en el cuadro subrayado en la **Ilustración 4-10**.

```

C:\Users\Usuario1>mongo
MongoDB shell version v3.4.6
connecting to: mongod://127.0.0.1:27017
MongoDB server version: 3.4.6
Server has startup warnings:
2017-07-22T20:32:17.960-0500 I CONTROL [initandlisten]
2017-07-22T20:32:17.960-0500 I CONTROL [initandlisten] ** WARNING: Access control
ol is not enabled for the database.
2017-07-22T20:32:17.961-0500 I CONTROL [initandlisten] **          Read and wri
te access to data and configuration is
unrestricted.
2017-07-22T20:32:17.962-0500 I CONTROL [initandlisten]

> use FacultadIngenieria
switched to db.FacultadIngenieria
> db.Estudiante.find().pretty()
    
```

Ilustración 4-10. Comando muestra de datos. (Flores A. & Villacís V., 2017).

```

{
  "_id" : ObjectId("5973f9d582e2ff8b33975795"),
  "CodEstudiante" : "JD?valos898",
  "Cedula" : 1709201902,
  "Nombre" : "Juan",
  "Nombre2" : "Andr?s",
  "Apellido" : "D?valos",
  "Apellido2" : "Cisneros",
  "Fecha de Nacimiento" : "02/01/1992",
  "Genero" : "Masculino",
  "Direccion" : "Ponceano Alto",
  "Telefono" : 2809891,
  "Año" : 2013,
  "Periodo" : 1,
  "Nota 1" : 13,
  "Nota2" : 13,
  "Nota 3" : 15,
  "Total" : 41,
  "Facultad" : "Ingenier?a",
  "Escuela" : "Civil"
}

{
  "_id" : ObjectId("5973f9d582e2ff8b33975796"),
  "CodEstudiante" : "CAndrade267",
  "Cedula" : 1756270190,
  "Nombre" : "Camilo",
  "Nombre2" : "Stefano",
  "Apellido" : "Andrade",
  "Apellido2" : "Suarez",
  "Fecha de Nacimiento" : "01/02/1993",
  "Genero" : "Masculino",
  "Direccion" : "Ualle de los Chillos",
  "Telefono" : 2790191,
  "Año" : 2012,
  "Periodo" : 1,
  "Nota 1" : 10,
  "Nota2" : 10,
  "Nota 3" : 16,
  "Total" : 36,
  "Facultad" : "Ingenier?a",
  "Escuela" : "Civil"
}

{
  "_id" : ObjectId("5973f9d582e2ff8b33975797"),
  "CodEstudiante" : "LBalderas342",
  "Cedula" : 1770201921,
  "Nombre" : "Luis",
  "Nombre2" : "Alberto",
  "Apellido" : "Balderas",
  "Apellido2" : "Flores",
  "Fecha de Nacimiento" : "00/11/1990",
  "Genero" : "Masculino",
  "Direccion" : "Amazonas y Gaspar de Villaruel",
  "Telefono" : 2618181,
  "Año" : 2013,
  "Periodo" : 1,
}
    
```

Ilustración 4-11. Datos Estudiante Terminal. (Flores A. & Villacís V., 2017).

○ Docente

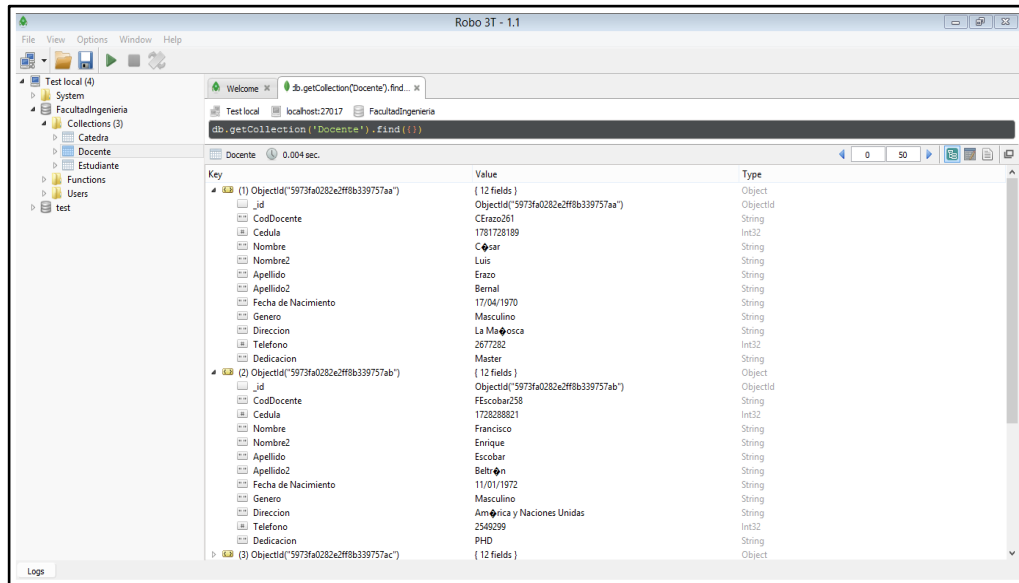


Ilustración 4-12. Datos Docente Robo 3T. (Flores A. & Villacís V., 2017).

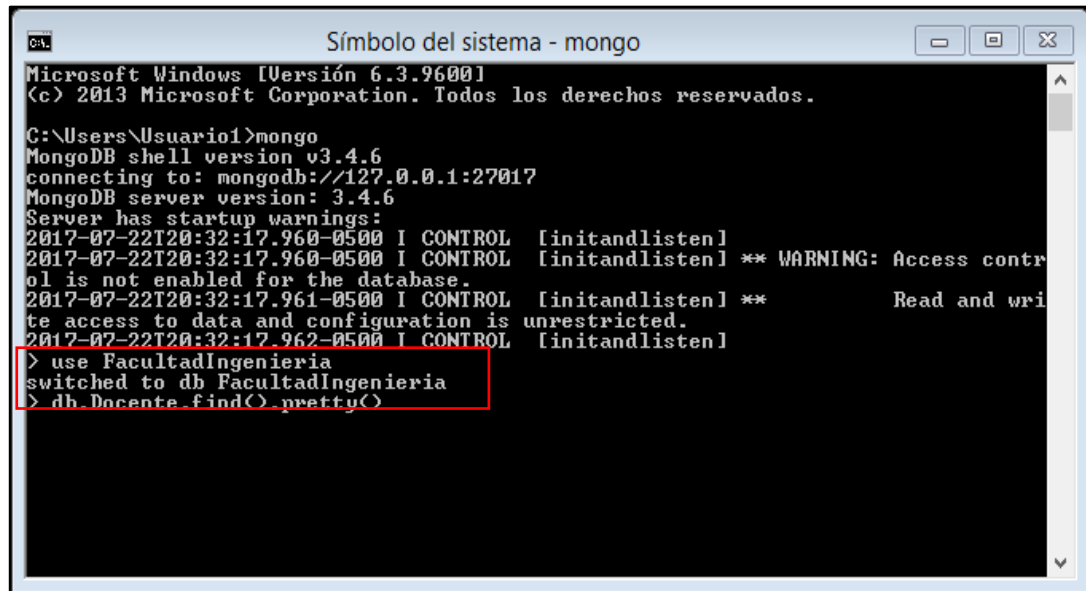


Ilustración 4-13. Comando muestra de datos. (Flores A. & Villacís V., 2017).

```

"Simbolo del sistema - mongo"
"Dedicacion" : "Master"
{
  "_id" : ObjectId("5973fa0282e2ff8b339757af"),
  "CodDocente" : "MGallardo236",
  "Cedula" : 1756526711,
  "Nombre" : "Mario",
  "Nombre2" : "Fernando",
  "Apellido" : "Gallardo",
  "Apellido2" : "Uara",
  "Fecha de Nacimiento" : "18/05/1973",
  "Genero" : "Masculino",
  "Direccion" : "Simon Bolivar y 17 de Septiembre",
  "Telefono" : 2340029,
  "Dedicacion" : "Master"
}
{
  "_id" : ObjectId("5973fa0282e2ff8b339757b0"),
  "CodDocente" : "JGarc?a242",
  "Cedula" : 1722538392,
  "Nombre" : "Juan",
  "Nombre2" : "Pablo",
  "Apellido" : "Garc?a",
  "Apellido2" : "Ledesma",
  "Fecha de Nacimiento" : "15/02/1966",
  "Genero" : "Masculino",
  "Direccion" : "Abd?n Calder?n y 24 de Mayo",
  "Telefono" : 2439201,
  "Dedicacion" : "Especializado"
}
{
  "_id" : ObjectId("5973fa0282e2ff8b339757b1"),
  "CodDocente" : "JGarc?a284",
  "Cedula" : 1773431223,
  "Nombre" : "Jos?",
  "Nombre2" : "Alfonso",
  "Apellido" : "Garc?a",
  "Apellido2" : "L?pez",
  "Fecha de Nacimiento" : "12/12/1970",
  "Genero" : "Masculino",
  "Direccion" : "Las Camelias y esq.",
  "Telefono" : 285200,
  "Dedicacion" : "Especializado"
}
{
  "_id" : ObjectId("5973fa0282e2ff8b339757b2"),
  "CodDocente" : "NUllac?s236",
  "Cedula" : 1786292912,
  "Nombre" : "Norma",
  "Nombre2" : "Norma",
  "Apellido" : "Ullac?s",
  "Apellido2" : "Cruz",
  "Fecha de Nacimiento" : "11/03/1973",
  "Genero" : "Femenino",
  "Direccion" : "De los Pinos",
  "Telefono" : 2343212,
  "Dedicacion" : "Especializado"
}

```

Ilustración 4-14. Datos Docente Terminal. (Flores A. & Villacís V., 2017).

○ Cátedra

Key	Value	Type
(1) ObjectId("5973fa1402e2ff8b339757be")	{ 9 fields }	Object
_id	ObjectId("5973fa1402e2ff8b339757be")	ObjectId
Nombre	F?sica	String
Cod_Docente	LCruz289	String
Horario	7:00-9:00	String
Dias	Lunes/Martes	String
Paralelo	1	Int32
Anio	2013	Int32
Periodo	1	Int32
Cod_Estudiante	09valad098	String
(2) ObjectId("5973fa1402e2ff8b339757bf")	{ 9 fields }	Object
_id	ObjectId("5973fa1402e2ff8b339757bf")	ObjectId
Nombre	Etica	String
Cod_Docente	CDu?as246	String
Horario	16:00-18:00	String
Dias	Viernes	String
Paralelo	1	Int32
Anio	2012	Int32
Periodo	1	Int32
Cod_Estudiante	CAndrade267	String
(3) ObjectId("5973fa1402e2ff8b339757c0")	{ 9 fields }	Object
(4) ObjectId("5973fa1402e2ff8b339757c1")	{ 9 fields }	Object
(5) ObjectId("5973fa1402e2ff8b339757c2")	{ 9 fields }	Object
(6) ObjectId("5973fa1402e2ff8b339757c3")	{ 9 fields }	Object
(7) ObjectId("5973fa1402e2ff8b339757c4")	{ 9 fields }	Object
(8) ObjectId("5973fa1402e2ff8b339757c5")	{ 9 fields }	Object
(9) ObjectId("5973fa1402e2ff8b339757c6")	{ 9 fields }	Object

Ilustración 4-15. Datos Cátedra Robo 3T. (Flores A. & Villacís V., 2017).

```

C:\Users\Usuario1>mongo
MongoDB shell version v3.4.6
connecting to: mongod://127.0.0.1:27017
MongoDB server version: 3.4.6
Server has startup warnings:
2017-07-22T20:32:17.960-0500 I CONTROL [initandlisten]
2017-07-22T20:32:17.960-0500 I CONTROL [initandlisten] ** WARNING: Access control
is not enabled for the database.
2017-07-22T20:32:17.961-0500 I CONTROL [initandlisten] **          Read and write
access to data and configuration is unrestricted.
2017-07-22T20:32:17.962-0500 I CONTROL [initandlisten]
> use FacultadIngenieria
switched to db FacultadIngenieria
> use FacultadIngenieria
switched to db FacultadIngenieria
> db.Catedra.find(<>.pretty(<>)
```

Ilustración 4-16. Comando muestra de datos. (Flores A. & Villacís V., 2017).

```

{"Nombre": "Matemáticas discretas",
"Cod_Docente": "MGallardo236",
"Horario": "7:00-9:00",
"Dias": "Lunes/Martes",
"Paralelo": 3,
"Año": 2014,
"Periodo": 2,
"Cod_Estudiante": "HBlanco438"}
<
{"_id": ObjectId("5973fa1482e2ff8b339757c7"),
"Nombre": "Jesucristo",
"Cod_Docente": "JGarc?a242",
"Horario": "7:00-9:00",
"Dias": "Jueves/Viernes",
"Paralelo": 1,
"Año": 2014,
"Periodo": 1,
"Cod_Estudiante": "JCant?432"}
<
{"_id": ObjectId("5973fa1482e2ff8b339757c8"),
"Nombre": "Jesucristo",
"Cod_Docente": "JGarc?a284",
"Horario": "16:00-18:00",
"Dias": "Lunes/Martes",
"Paralelo": 2,
"Año": 2012,
"Periodo": 1,
"Cod_Estudiante": "Dcontreras190"}
<
{"_id": ObjectId("5973fa1482e2ff8b339757c9"),
"Nombre": "Matemáticas discretas",
"Cod_Docente": "SEscobedo246",
"Horario": "16:00-18:00",
"Dias": "Lunes",
"Paralelo": 3,
"Año": 2011,
"Periodo": 2,
"Cod_Estudiante": "HCruz103"}
<
{"_id": ObjectId("5973fa1482e2ff8b339757ca"),
"Nombre": "Hidrología",
"Cod_Docente": "NUillac?s236",
"Horario": "16:00-18:00",
"Dias": "Lunes/Martes",
"Paralelo": 1,
"Año": 2013,
"Periodo": 1,
"Cod_Estudiante": "NCort?s490"}
<
{"_id": ObjectId("5973fa1482e2ff8b339757cb"),
"Nombre": "Hidrología",
"Cod_Docente": "UMart?nez266",
```

Ilustración 4-17. Datos Cátedra Terminal. (Flores A. & Villacís V., 2017).

4.4.2. Aplicación MongoDB de los datos oficiales de la Facultad

A continuación, se aplicará los datos oficiales de la facultad a MongoDB, para observar los resultados.

Para importar los datos en MongoDB, se lo puede hacer por medio de consola, ingresando el comando que está representado en el cuadro subrayado en la *Ilustración 4-18*.

```

2017-08-09T12:18:08.403-0500 I CONTROL [initandlisten] now exiting
2017-08-09T12:18:08.404-0500 I CONTROL [initandlisten] shutting down with code:
48
C:\Users\Usuario1>mongoimport -d FacultadIngenieriaPUCE -d Docente --type csv --
file C:\Users\Usuario1\Documents\PUCE\Tesis\DocentesIngenieria.csv --headerline
2017-08-09T12:18:37.030-0500 E QUERY no collection specified
2017-08-09T12:18:37.093-0500 I QUERY using filename 'DocentesIngenieria' as collectio
n
2017-08-09T12:18:37.103-0500 I QUERY connected to: localhost
2017-08-09T12:18:37.793-0500 I QUERY imported 245 documents
C:\Users\Usuario1>
    
```

Ilustración 4-18. Importe de datos reales. (Flores A. & Villacís V., 2017).

En la *Ilustración 4-19*, se puede observar los datos de docentes por medio de Robomongo.

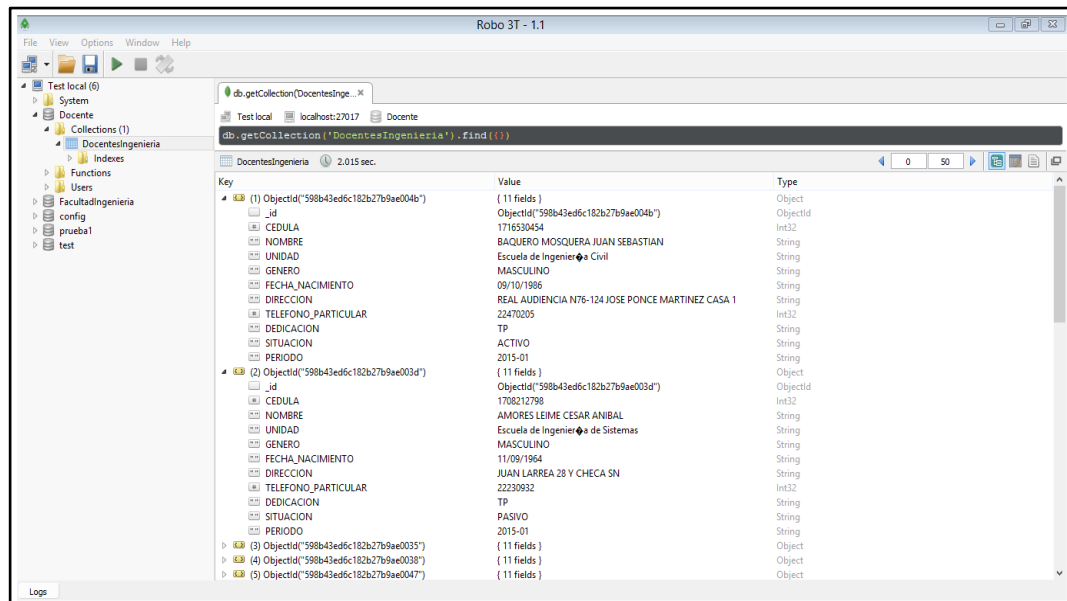


Ilustración 4-19. Datos Docente Robo 3T. (Flores A. & Villacís V., 2017).

En la **Ilustración 4-20**. Se puede observar los datos de los docentes por medio del terminal.

```

C:\Users\Usuario1>use Docente
'use' no se reconoce como un comando interno o externo,
programa o archivo por lotes ejecutable.

C:\Users\Usuario1>mongo
MongoDB shell version v3.4.6
connecting to: mongod://127.0.0.1:27017
MongoDB server version: 3.4.6
Server has startup warnings:
2017-08-08T10:41:02.577-0500 I CONTROL [main]
2017-08-08T10:41:02.577-0500 I CONTROL [main] ** WARNING: Access control is not
enabled for the database.
2017-08-08T10:41:02.578-0500 I CONTROL [main] ** Read and write access
to data and configuration is unrestricted.
2017-08-08T10:41:02.578-0500 I CONTROL [main]
mongos> use Docente
switched to db Docente
mongos> db.DocentesIngenieria.find().pretty()
{
  "_id" : ObjectId("598b43ed6c182b27b9ae0033"),
  "CEDULA" : 1708774004,
  "NOMBRE" : "ACOSTA TRUJILLO ENRIQUE",
  "UNIDAD" : "Escuela de Ingenier?a Civil",
  "GENERO" : "MASCULINO",
  "FECHA NACIMIENTO" : "05/11/1938",
}
    
```

Ilustración 4-20. Datos Reales Terminal. (Flores A. & Villacís V., 2017).

4.5. Herramienta QlikView

QlikView es una herramienta de extracción, que permite una excelente visualización de datos, con mayor flexibilidad y análisis de información, con una pantalla amigable y de fácil uso. A continuación, se presentará esta herramienta con los datos oficiales de los Docentes de la facultad.

En la **Ilustración 4-21** e **Ilustración 4-22**. Se encuentran los filtros y los campos de los datos importados que se requieren para su respectivo análisis: incluyendo los más relevantes para obtener resultados que sean útil para la toma de decisiones en la facultad.

PERIODO	Count	ESCUELA	Count	EDAD	Count
2012-01	166	Escuela de Ingeniería Civil	27	3	
2012-02	181	Escuela de Ingeniería de Sistemas	29	1	
2013-01	187		30	3	
2013-02	194		31	2	
2014-01	203		32	2	
2014-02	243		33	2	
2015-01	245		34	4	
2015-02	241		35	6	
2016-01	236		36	6	
2016-02	260		37	3	
			38	7	
			39	6	

Ilustración 4-21. Filtros QlikView. (Flores A. & Villacís V., 2017).

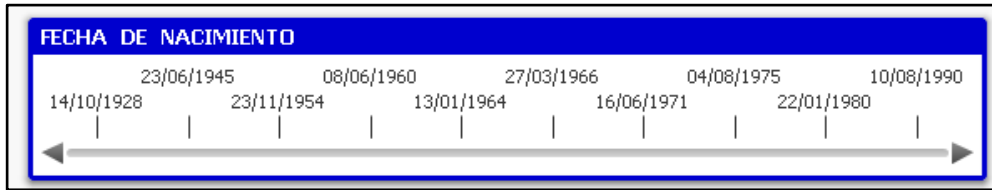


Ilustración 4-22. Filtros QlikView. (Flores A. & Villacís V., 2017).

En la *Ilustración 4-23*. Se observa la tabla de docentes, con los datos más importantes de los mismos.

DOCENTES FACULTAD INGENIERÍA			
CEDULA	NOMBRE	DIRECCION	TELEFONO_PARTICULAR
C4X01LWPR	BIERWIRTH HANNES	PASAJE ROSALES TUMBACO ALTO	NULL
0102007176	AGUILAR JARAMILLO JORGE ARTURO	TOMÁS BERMUR GUANGUILTAGUA N39-107	022252292
0102716099	CASTILLO ASTUDILLO JAIME HERNAN	GENERAL DUMA LAS VIOLETAS N47-211	023340360
0200838217	MENDIZABAL FREIRE MAURO JAVIER	RAFAEL BUSTAMANTE PIO JARAMILLO E6-239	022401838
0201547544	VERDEZOTO VASCONEZ PAOLA KATHERINE	AV. ELOY ALFARO CALLE CAPRI 27	023460540
0300317435	OCHOA NUÑEZ FRANCISCO GUILLERMO	VALLADOLID LUGO E14-49	022558309
0300878394	TOBAR SOLANO CARLOS BORIS	CARCELÉN, CALLE VELASCO IBARRA CARVAJAL 63B	022802581
0400081998	RAZA CALDERON NELSON ANIBAL	RUMIÑAHUI, JOSÉ ARCE AV. DEL MAESTRO NO 61-22	022531490
0400631271	MORA VARELA JAIME HUMBERTO	PASAJE I PASAJE H 146	023807418
0400747333	CUAICAL ENRIQUEZ LUIS ALFREDO	AVENIDA UNIVERSITARIA EUSTORGIO SALGADO 19-110	022978176
0400777993	MICHELENA ROSERO CESAR GUSTAVO	SABANILLA REAL AUDIENCIA OE2-155	023465426

Ilustración 4-23. Tabla Docentes QlikView. (Flores A. & Villacís V., 2017).

En la *Ilustración 4-24* e *Ilustración 4-25*, se encuentran los gráficos de los resultados, por medio de los filtros realizados según se requiera.

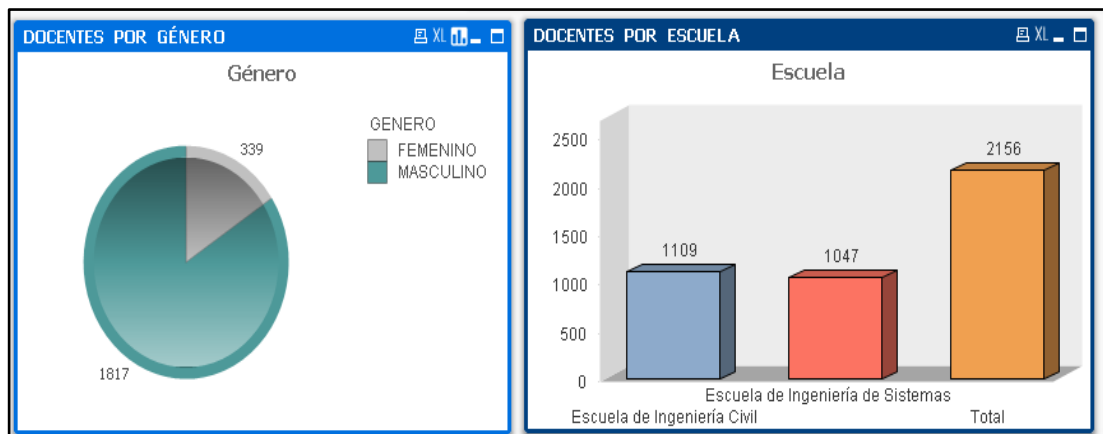


Ilustración 4-24. Resultado Gráficos I QlikView. (Flores A. & Villacís V., 2017).

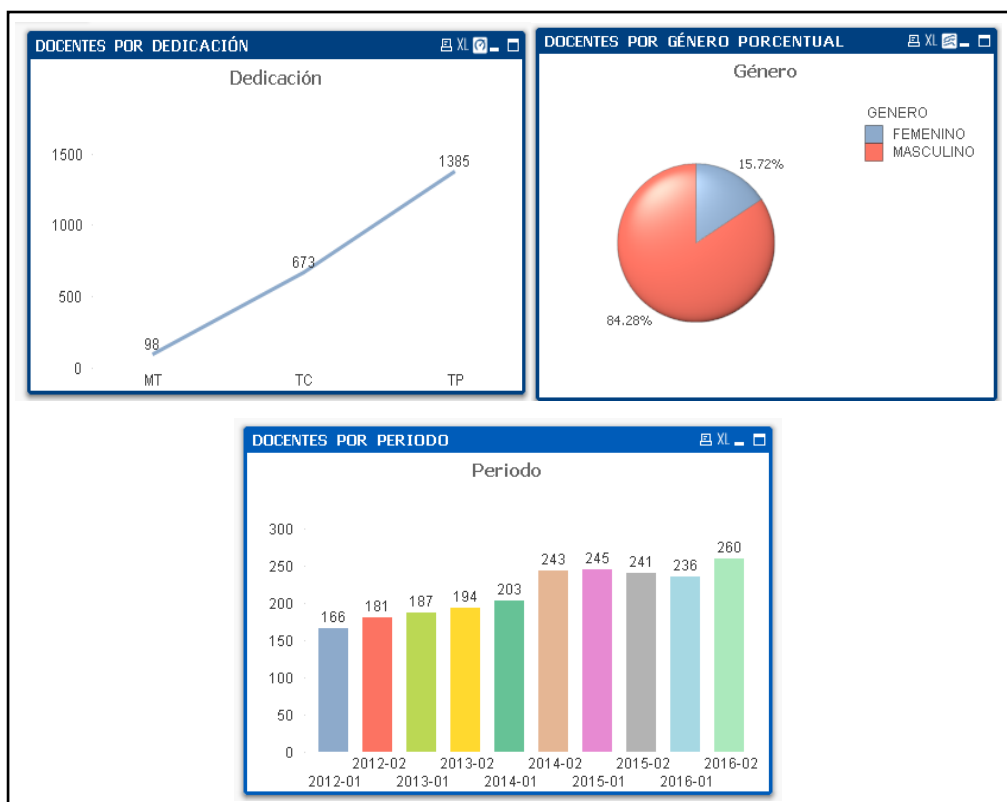


Ilustración 4-25. Resultado Gráficos II QlikView. (Flores A. & Villacís V., 2017).

4.6. Herramienta Microsoft Power BI

Microsoft Power BI es una herramienta de extracción de información que nos ayuda a visualizar los datos de forma amigable, por las pruebas realizadas anteriormente, esta herramienta es la más adaptable para la Facultad de Ingeniería de la PUCE y aparte de estar liderando el Cuadrante de Gartner. A continuación, se presentará esta herramienta con los datos oficiales de los Docentes de la facultad.

Para empezar con esta herramienta en la *Ilustración 4-26* se presenta la interfaz general, a lado derecho se encuentran los filtros y los campos de los datos importados. Dependiendo de cómo se desea los filtros, se va arrastrando al lienzo en blanco que se encuentra en el centro como se visualiza en la *Ilustración 4-27*.

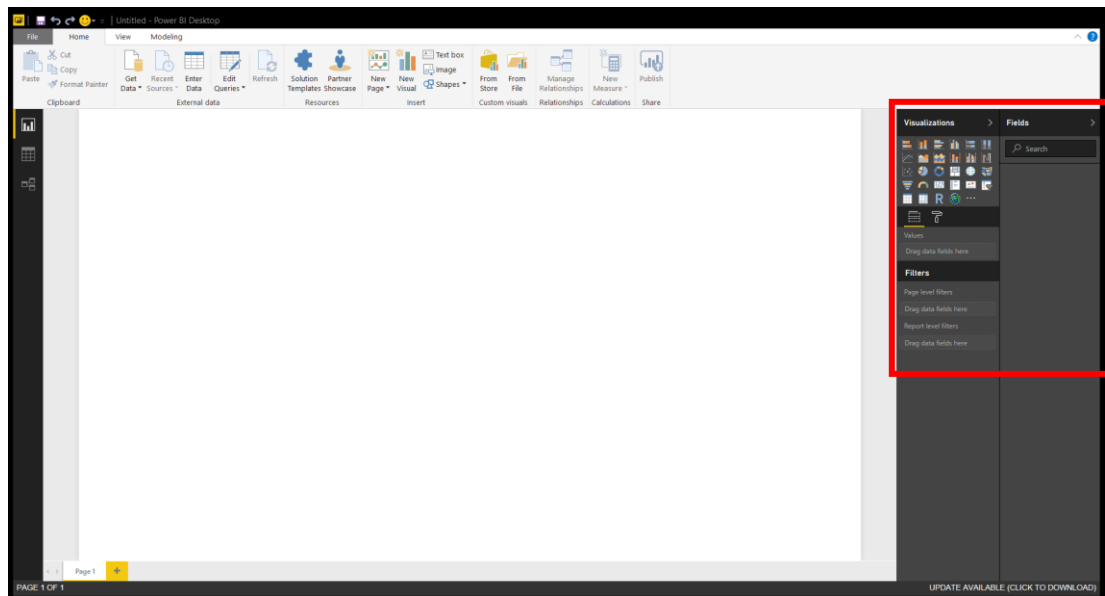


Ilustración 4-26. Interface Power BI. (Flores A. & Villacís V., 2017).

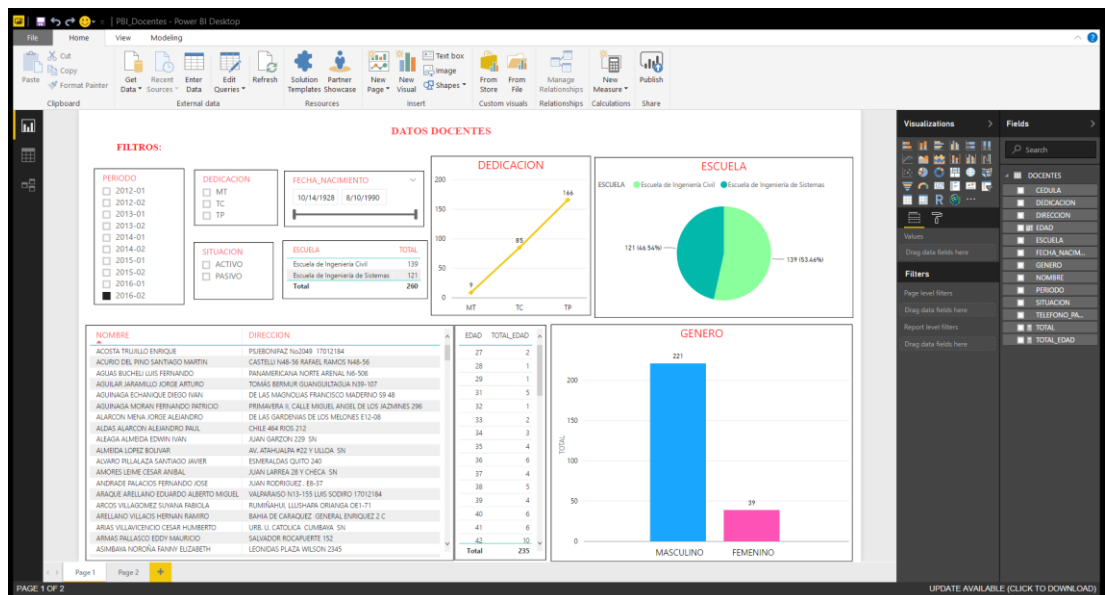


Ilustración 4-27. Power BI con datos y filtros. (Flores A. & Villacís V., 2017).

Al momento de importar los datos con los que vamos a trabajar se puede visualizar la información como se observa en la *Ilustración 4-28*.

CARRERA	NOMBRE	ESCUELA	GENERO	FECHA_NACIMIENTO	DEDICACION	TELEFONO_PARTICULAR	SITUACION	PERIODO	EDAD
570890240	ENRIQUEZ RIVERA ROBERT ARTURO	Escuela de Ingeniería de Sistemas	MASCULINO	8-16-1966	12 DE OCTUBRE 1976 Y ROCA 5N	022991700	TP	PASIVO	2012-01
170923243	JARAMILLO ORELLANA MARK EDUARDO	Escuela de Ingeniería de Sistemas	MASCULINO	9-29-1968	12 DE OCTUBRE 1976 Y ROCA 5N	022991700	TP	PASIVO	2012-01
170383937	MEJIA MANTILLA GALO FERNANDO	Escuela de Ingeniería de Sistemas	MASCULINO	11-23-1954	12 DE OCTUBRE 1976 Y ROCA 5N	022991700	TP	ACTIVO	2012-01
170871787	ROLAN CREPO GIOVINNI RAFAEL	Escuela de Ingeniería de Sistemas	MASCULINO	11-14-1965	12 DE OCTUBRE 1976 Y ROCA 5N	022991700	TP	PASIVO	2012-01
170424877	TORRES PROAÑO EDGAR PORFIRIO	Escuela de Ingeniería de Sistemas	MASCULINO	5-29-1955	12 DE OCTUBRE 1976 Y ROCA 5N	022991700	TP	PASIVO	2012-01
1707958746	VALDIVESIO TORRES XAVIER RENATO	Escuela de Ingeniería Civil	MASCULINO	12-21-1968	12 DE OCTUBRE 1976 Y ROCA 5N	022991700	TP	PASIVO	2012-01
1701620567	VERGARA JARAMILLO LUIS AUGUSTO	Escuela de Ingeniería Civil	MASCULINO	7-19-1945	12 DE OCTUBRE 1976 Y ROCA 5N	022991700	TP	PASIVO	2012-01
170102570	GONZALEZ ZAMBRANO MARIA CRISTINA	Escuela de Ingeniería de Sistemas	FEMENINO	12-1-1973	12 DE OCTUBRE 1976 Y ROCA 5N	022991700	TP	PASIVO	2012-01
170723090	ARAUQUE ARELLANO EDUARDO ALBERTO MIGUEL	Escuela de Ingeniería Civil	MASCULINO	7-11-1961	VALPARAISO N13-155 LUIS SCORIO 17012184	022991700	TP	PASIVO	2012-01
170723090	ARAUQUE ARELLANO EDUARDO ALBERTO MIGUEL	Escuela de Ingeniería de Sistemas	MASCULINO	7-11-1961	VALPARAISO N13-155 LUIS SCORIO 17012184	022991700	TP	PASIVO	2012-01
170825267	ARELLANO VILLACIS HERNAN RAMIRO	Escuela de Ingeniería Civil	MASCULINO	7-10-1964	BAHIA DE CARAQUEZ GENERAL ENRIQUEZ 2 C	022991700	TP	PASIVO	2012-01
171274059	BERNI GALLANO JORGE JAVIER	Escuela de Ingeniería de Sistemas	MASCULINO	1-2-1975	12 DE OCTUBRE 1976 Y ROCA 17012184	022991700	TP	PASIVO	2012-01
170538930	HERRMANN FERNANDEZ PATRICIA LUCIA	Escuela de Ingeniería de Sistemas	FEMENINO	7-8-1959	NO EXISTE 5N	022991700	TP	PASIVO	2012-01
171232049	JARRIN TORRES LE OGLE ELIZABETH	Escuela de Ingeniería de Sistemas	FEMENINO	11-29-1975	12 DE OCTUBRE 1976 Y ROCA 17012184	022991700	TP	PASIVO	2012-01
180189869	GUICHAMORENO QUINDO EDUARDO	Escuela de Ingeniería de Sistemas	MASCULINO	7-8-1965	VIA BARRIO LOS ANGELES 289	022991700	TP	ACTIVO	2012-01
170475491	ORTEGA ORTEGA RICARDO	Escuela de Ingeniería Civil	MASCULINO	11-15-1956	LEZARZAS 5N	022991700	TP	PASIVO	2012-01
8199913487	RESTREPO MONTOYA GUSTAVO HERNANDO	Escuela de Ingeniería de Sistemas	MASCULINO	9-1-1963	12 DE OCTUBRE 1976 Y ROCA 17012184	022991700	TP	PASIVO	2012-01
170034946	SANTOS NARANJAZ BENIGNO IGNACIO	Escuela de Ingeniería Civil	MASCULINO	12-18-1947	12 DE OCTUBRE 1976 Y ROCA 17012184	022991700	TP	PASIVO	2012-01
170807600	FERNANDEZ CARDENAS FANNY	Escuela de Ingeniería de Sistemas	FEMENINO	7-6-1964	RAMON ROCA LEONIDAS PLAZA 21-210	022564161	TP	ACTIVO	2012-01
170584591	GARCÉS ABAD EUGENIA PATRICIA	Escuela de Ingeniería de Sistemas	FEMENINO	9-23-1959	VALLADOLID CORDERO N24-219	022238474	TP	ACTIVO	2012-01
171232024	GARCÍA PARDEDES YOLANDA CECILIA	Escuela de Ingeniería de Sistemas	FEMENINO	12-4-1973	PSAJE MAURICIO PLAZA PEDRO PONCE S/N	022551526	TP	ACTIVO	2012-01
150017113	GUTIERREZ GARCIA NATI LOTHY	Escuela de Ingeniería de Sistemas	FEMENINO	12-7-1959	NAURETIN 7B Y PARRA REAL ALDONCIA DEL 17012184	022290923	TP	PASIVO	2012-01
171615695	HERSERA SRAZO MARÍA FERNANDA	Escuela de Ingeniería Civil	FEMENINO	8-3-1980	URB. RIECEL CALLE A N 40523 Y MOTOLONES MOTOLONES 40-5280	022447882	TP	PASIVO	2012-01
170521827	HOLTE FIONA	Escuela de Ingeniería Civil	FEMENINO	6-11-1975	CALLE F EDUARDO CARIBALLO N-6332	022100002	TP	PASIVO	2012-01
091225590	INFANTES MANTILLA CATHERINE DE LOUISE	Escuela de Ingeniería de Sistemas	FEMENINO	8-31-1973	PORTETE 6 DE DICIEMBRE, TORRE A DPTO. 202	023330621	TP	PASIVO	2012-01
171404562	MASARANTA CARRION SUSANA MARGARITA	Escuela de Ingeniería de Sistemas	FEMENINO	11-9-1976	ESQUEJUA ZARUMBA OBR-05	022647614	TP	ACTIVO	2012-01
170705320	MORALES MALES MARIA DEL PILAR	Escuela de Ingeniería de Sistemas	FEMENINO	9-8-1961	VENACRUZ 9 17012184	022437886	TP	PASIVO	2012-01
170584755	MURDOZ ARDOLZA TANIA FRANCISCA	Escuela de Ingeniería Civil	FEMENINO	7-19-1964	JUAN DIAZ RODRIGO MUREL N37-224	023310174	TP	ACTIVO	2012-01
170584755	MURDOZ ARDOLZA TANIA FRANCISCA	Escuela de Ingeniería de Sistemas	FEMENINO	7-19-1964	JUAN DIAZ RODRIGO MUREL N37-224	023310174	TP	ACTIVO	2012-01
170401606	PRO ZAMBRANO KATIA PATRICIA	Escuela de Ingeniería de Sistemas	FEMENINO	10-9-1954	EL RICA, GUARUMOS 6 DE DICIEMBRE 448	022432298	TP	PASIVO	2012-01
170914817	SARMIENTO SARMIENTO NARCISA MARISOL	Escuela de Ingeniería de Sistemas	FEMENINO	3-27-1966	AV OCCIDENTAL LEGARDA NO 62183	021596001	TP	ACTIVO	2012-01
170914817	SARMIENTO SARMIENTO NARCISA MARISOL	Escuela de Ingeniería Civil	FEMENINO	3-27-1966	AV OCCIDENTAL LEGARDA NO 62183	022396001	TP	ACTIVO	2012-01
170417844	TURRIAGA EGUIGUIREN CAROLINA	Escuela de Ingeniería de Sistemas	FEMENINO	5-31-1964	SAN IGNACIO 418 Y CORUÑA 5N	022502942	TP	ACTIVO	2012-01
170423458	VILLACIS VALLAFUERTE CECILIA B	Escuela de Ingeniería de Sistemas	FEMENINO	2-6-1956	MAÑOSCA Y FCO CRUZ MIRANDA EL BALCON 30 5N	022259549	TP	PASIVO	2012-01
171362071	ACURIO DEL PINO SANTIAGO MARTIN	Escuela de Ingeniería de Sistemas	MASCULINO	11-8-1975	CASTELLANOS 56 RAFAEL RAMOS N48-56	022400878	TP	ACTIVO	2012-01
110200276	AGUIAR JARAMILLO JORGE ARTURO	Escuela de Ingeniería de Sistemas	MASCULINO	7-12-1963	TOMÁS BERNAR GUANUQUIGUA N69-037	022232292	TP	ACTIVO	2012-01

Ilustración 4-28. Importe de datos en Power BI. (Flores A. & Villacís V., 2017).

Los campos para filtrar la información se obtuvieron de los requerimientos, los cuales son Periodo, que es el periodo académico de cada semestre de la facultad; Dedicación, que es el tiempo de permanencia en la facultad de cada docente y se dividen en MT (Medio Tiempo), TC (Tiempo Completo) y TP (Tiempo Parcial); la Situación que refleja la situación actual de cada docente que son Activo o Pasivo; Y el filtro de la Escuela que son las carreras de la Facultad de Ingeniería.

PERIODO

- 2012-01
- 2012-02
- 2013-01
- 2013-02
- 2014-01
- 2014-02
- 2015-01
- 2015-02
- 2016-01
- 2016-02

DEDICACION

- MT
- TC
- TP

FECHA_NACIMIENTO

10/14/1928 8/10/1990

SITUACION

- ACTIVO
- PASIVO

ESCUELA **TOTAL**

Escuela de Ingeniería Civil	77
Escuela de Ingeniería de Sistemas	89
Total	166

Ilustración 4-29. Filtros Power BI. (Flores A. & Villacís V., 2017).

Con la fecha de nacimiento podemos sacar la edad y poder filtrar la información por edades, observando también el nombre de los docentes, como se puede ver en la **Ilustración 4-30**.

NOMBRE	DIRECCION	EDAD	TOTAL_EDAD
CASTRO CARRERA FERNANDO FRANCISCO	ATACAZO CALLE GENERAL ENRIQUEZ 142	50	1
CASTRO MERINO EDWIN PATRICIO	LAS PALMERAS LOS CEDROS 45	51	5
CAZCO CAZCO MARCO ANTONIO	DE LOS NOGALES DE LAS PALMERAS 1492	52	7
CHACON TORAL OCTAVIO FRANCISCO JOSE	LA CUMBRE 400 Y CARLOS MONTUFAR . 17012184	53	10
CHAFLA ALTAMIRANO GUSTAVO XAVIER	CALLE CHIMBORAZO AV INTEROCEANICA 234	54	5
CHANCAY BERMELLO BOGAR ADOLFO	LUIS CORDERO ULPIANO PÁEZ S/N	55	7
CODA TROSSARELLO SESTILLO MATEO	LLANO CHICO, BARRIO LA DELICIA MANUEL BENITEZ S/N	56	4
CONDOR CRUZ JAVIER WILFRIDO	ELOY ALFARO N44-403 HIGUERAS DPTO.3-1	57	2
DE LA CRUZ DOMINGUEZ FABIAN IGNACIO	COTOCOLLAO, GENERAL GUERRERO RUMIHURCO N63-224	58	6
ECHVERRIA FEJOO CARLOS AUGUSTO	PASAJE D LOS NOGALES S4 - 157	60	2
ECHVERRIA RODRIGUEZ CARLOS	JULIO TEODORO SALEM 766 17012184	61	5
EGAS ACOSTA CARLOS ROBERTO	AV. CORDOVA GALARZA 13 DE JUNIO CASA 114	62	2
EGAS VAREA DIEGO	VICENTE PAJUJEO MANUEL VALDIVIEZO 311	63	5
ENRIQUEZ CAJAS PAUL	PASAJE GUACAMAYOS ED. SEBASTIÁN FRESNOS S/N	64	1
ENRIQUEZ REYES ROBERT ARTURO	12 DE OCTUBRE 1076 Y ROCA SN	65	2
ESPINOSA VITERI LUIS OSWALDO	CONOCOTO, URBANIZACIÓN PUERTA DEL VALLE PSJ. B 106		
ESTRELLA CADENA CARLOS ALBERTO	INGLATERRAS46VANCOUVERT 17012184		
FERNANDEZ CARDENAS FANNY	RAMON ROCA LEONIDAS PLAZA 21-210		
GAMBOA TROYA PEDRO FABIAN	EL VERGEL DE PUSUQUI 200 SN		
		Total	145

Ilustración 4-30. Filtros II Power BI. (Flores A. & Villacís V., 2017).

La Ilustración 4-31. Gráficos I Power BI. **Ilustración 4-31** y la **Ilustración 4-32** se encuentran los gráficos de los resultados por medio de los filtros realizados según se requiera.

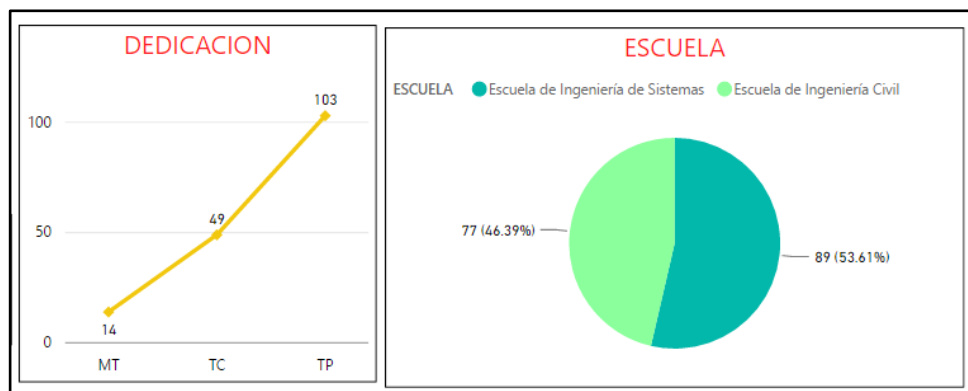


Ilustración 4-31. Gráficos I Power BI. (Flores A. & Villacís V., 2017).

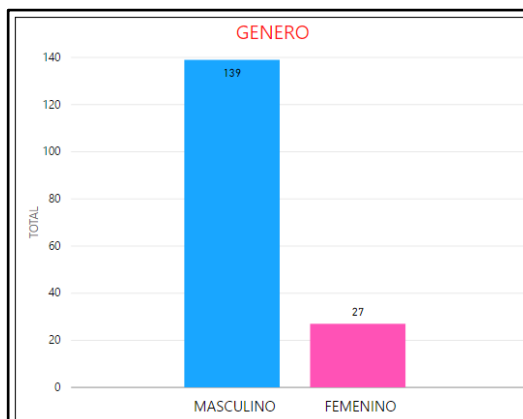


Ilustración 4-32. Gráficos II Power BI. (Flores A. & Villacís V., 2017).

En este capítulo podemos concluir que la herramienta de MongoDB no es recomendable para un análisis con pocos datos, como con nuestro caso de estudio; no se llega completamente al concepto de Big Data, es por eso que se usan para un mejor resultado herramientas de extracción de información, que para este caso son más eficientes.

Después de pruebas con herramientas de extracción como QlikView y Microsoft Power BI, escogimos la segunda opción, por las ventajas que ofrece y ser más adaptable a los requerimientos propuestos por autoridades de la facultad.

Si queremos hablar de Big Data, se puede fácilmente hacer un estudio más amplio que requiera no sólo un caso de estudio como es el de la Facultad de Ingeniería de la PUCE, sino otras facultades y otras universidades donde los datos puedan ser homologables y se pueda comparar, además tener información variada como redes sociales, papers, etc., que puedan cumplir con lo que llamamos Big Data.

CAPÍTULO 5. CONCLUSIONES Y RECOMENDACIONES

En este capítulo se detallará las conclusiones y recomendaciones que se han propuesto después del presente trabajo.

5.1. Conclusiones

- En la Pontificia Universidad Católica del Ecuador existe falta de información respecto a estudiantes como docentes, debido a que los datos solicitados, no eran de conocimiento público, tanto en la página oficial de la PUCE como en Redes Sociales.
- Las herramientas de extracción de información son de utilidad para la analítica de datos, reportes, cuadros; los cuáles intervienen con el área de Business Intelligence que es parte del ecosistema de Big Data.
- Las bases de datos NoSQL son estructuras de almacenamiento que no cumplen con el esquema de entidad – relación, es decir, los datos son almacenados de forma independiente de cualquier relación que se pueda tener con algún otro grupo de datos.
- Los datos en una base de datos NoSQL son una colección de objetos, los cuales para su tratamiento en la base de datos, no se tiene el concepto de entidad; son objetos que tienen otros objetos de igual o diferente características.
- La ventaja que aportan en este estudio las bases de datos NoSQL es su gran adaptabilidad, es decir, se puede guardar en una misma columna archivos de diferente tipo de dato como son gráficos, formato XML, objetos, entre otros.
- Big Data posee una gran selección de herramientas como software libre y open source, propietarias y en la nube, las cuales benefician a organizaciones con gran cantidad de datos que requieren de consultas más rápidas y ágiles.
- La función de la persona que gestiona los datos en Big Data tiene la responsabilidad de que estos no sean datos muertos, es decir, que los datos se conviertan en información relevante para la empresa u organización.
- En Big Data hay que limitar el alcance de los datos a manejar, para que no se conviertan en información basura.

- Hadoop es una plataforma de código abierto que ostenta el liderazgo en la actualidad como herramienta para analizar grandes cantidades de datos.
- Apache Spark está diseñado para ser accesible a través de sencillas interfaces de programación de aplicaciones (API, siglas en inglés de Application Programming Interface) para lenguajes de programación como Python, Java, Scala, SQL o R.
- MongoDB es una base de datos NoSQL orientada a documentos, por lo cual podemos decir que toda la información almacenada en esta herramienta se guarda como documentos BSON.
- En la actualidad, las organizaciones están viendo el valor de los datos que tiene la misma, lo que es una ventaja competitiva saber manejar estos datos.
- Los datos históricos que se manejan en la Facultad de Ingeniería de la PUCE, tienen un cierto grado de confidencialidad, lo que no se puede compartir públicamente cierta información.
- Cuando se comparan datos entre varias instituciones, estos deben ser homogéneos y homologables.
- El manejo de datos tiene particularidades para lo que es relacional, transaccional, cubo de información, documental y Big Data en sí; por lo tanto existen herramientas especializadas para cada una de ellas.
- La información solicitada, ha sido comparada con el sitio oficial web de la PUCE, con la nueva plataforma IGNATIUS, para la consistencia de los datos.
- Los datos que se manejan se vuelven de manera inmediata disponibles al público o a la audiencia que está destinada.
- La información en la plataforma IGNATIUS refleja información diferente de la que nosotros tenemos como datos detallados por parte de la Facultad como de la Dirección General.
- Los únicos datos oficiales que se posee en este trabajo son los de Docentes, debido a que por privacidad de los mismos la Facultad de Ingeniería de la PUCE junto con la Dirección de Aseguramiento de la Calidad no los concede de forma inmediata.

- Por medio de un análisis comparativo de velocidad de las herramientas de Big Data se pudo concluir que la más rápida en ejecución es MongoDB.
- MongoDB al ser una herramienta para el manejo de una gran cantidad de información y datos, no es recomendable para datos de la Facultad de Ingeniería de la PUCE ya que estos se pueden trabajar sin ningún problema con herramientas de extracción de información.

5.2. Recomendaciones

- Se sugiere que en la página oficial de la Pontificia Universidad Católica del Ecuador, se pueda visualizar el número de estudiantes por año e información que pueda ser relevante para este tipo de análisis.
- Se sugiere que la información de las carreras de la universidad en la página web oficial, debe ser estándar, es decir, que se refleje los mismos datos en todas.
- Se sugiere utilizar una base de datos NoSQL para manejar gran cantidad de información en estructuras distribuidas.
- Las bases de datos NoSQL puede tener mucha información de varios tipos, es por eso que se sugiere que la información tomada sea útil para la organización y no sólo una agrupación de datos en una consulta.
- Se recomienda tener en claro que las herramientas de extracción de información, son parte del proceso de Big Data, más no el ecosistema completo del mismo.
- Para un manejo adecuado de Big Data se recomienda utilizar las herramientas de extracción de información para la creación de reportes dinámicos, y análisis de datos estadísticos.
- Se sugiere revisar la información de las licencias adquiridas, debido a que no todas son gratuitas, independientemente si es software libre o propietario.
- El uso de Big Data se recomienda en empresas donde la cantidad de datos ya no se pueda manejar en forma ágil con bases de datos relacionales o el tiempo de espera de las consultas es extenso.
- Se recomienda para la implementación de Big Data investigar acerca de su infraestructura para que con esos conocimientos se pueda acelerar las consultas y

optimizar la analítica de los datos obtenidos, como resultado teniendo reportes de mayor utilidad para la organización.

- Para dar valor a los datos, se recomienda tomar en cuenta al realizar una estructura de base de datos desde la fase de diseño en adelante.
- Se sugiere que las organizaciones se tomen el tiempo adecuado junto a usuarios expertos para validar la calidad de los datos.
- Se sugiere que los datos históricos sean analizados por el departamento que corresponda de la PUCE, en este caso, el departamento asignado denominado “Dirección de Aseguramiento de la Calidad”, para la validez de la información.
- Se recomienda que los datos a comparar se deben comprar, vender o intercambiar con el criterio de homogeneidad.
- Se sugiere mantener actualizada la información en el sitio web oficial de la PUCE, debido a que algunos de los datos tomados no concuerdan con la situación actual.
- Se recomienda que los datos en cada uno de las facultades sean homologables y tengan el mismo criterio de obtención de los mismos.
- Se sugiere a la facultad que al otorgar los datos lo hagan de manera anónima, es decir, modificando los nombres para conservar la privacidad de la persona.
- Se sugiere que para llegar al concepto de Big Data y utilizar herramientas para la misma se debe comparar mucho más que datos de una sola facultad, es decir, de facultades por universidades, redes sociales, papers, etc.
- Se sugiere que en la página oficial de la Pontificia Universidad Católica del Ecuador, se pueda visualizar el número de estudiantes por año e información que pueda ser relevante para este tipo de análisis.
- Se sugiere que la información de las carreras de la universidad en la página web oficial debe ser estándar, es decir, que se refleje los mismos datos en todas.
- Se sugiere utilizar una base de datos NoSQL para manejar gran cantidad de información en estructuras distribuidas.

- Las bases de datos NoSQL puede tener mucha información de varios tipos, es por eso que se sugiere que la información tomada sea útil para la organización y no sólo una agrupación de datos en una consulta.
- Se recomienda tener en claro que las herramientas de extracción de información son parte del proceso de Big Data, más no el ecosistema completo del mismo.
- Para un manejo adecuado de Big Data se recomienda utilizar las herramientas de extracción de información para la creación de reportes dinámicos, y análisis de datos estadísticos.
- Se sugiere revisar la información de las licencias adquiridas, debido a que no todas son gratuitas, independientemente si es software libre o propietario.
- El uso de Big Data se recomienda en empresas donde la cantidad de datos ya no se pueda manejar en forma ágil con bases de datos relacionales o el tiempo de espera de las consultas es extenso.
- Se recomienda para la implementación de Big Data investigar acerca de su infraestructura para que con esos conocimientos se pueda acelerar las consultas y optimizar la analítica de los datos obtenidos, como resultado teniendo reportes de mayor utilidad para la organización.
- Para dar valor a los datos, se recomienda tomar en cuenta al realizar una estructura de base de datos desde la fase de diseño en adelante.
- Se sugiere que las organizaciones se tomen el tiempo adecuado junto a usuarios expertos para validar la calidad de los datos.
- Se sugiere que los datos históricos sean analizados por el departamento que corresponda de la PUCE, en este caso, el departamento asignado denominado “Dirección de Aseguramiento de la Calidad”, para la validez de la información.
- Se recomienda que los datos a comparar se deben comprar, vender o intercambiar con el criterio de homogeneidad.
- Se sugiere mantener actualizada la información en el sitio web oficial de la PUCE, debido a que algunos de los datos tomados no concuerdan con la situación actual.

- Se recomienda que los datos en cada una de las facultades sean homologables y tengan el mismo criterio de obtención de los mismos.
- Se sugiere a la facultad que al otorgar los datos lo hagan de manera anónima, es decir, modificando los nombres para conservar la privacidad de la persona.
- Se sugiere que para llegar al concepto de Big Data y utilizar herramientas para la misma se debe comparar mucho más que datos de una sola facultad, es decir, de facultades por universidades, redes sociales, papers, etc.

GLOSARIO

A

Algoritmos iterativos

Son aquellos algoritmos que llegan a un resultado a través de una iteración mediante un ciclo definido o indefinido. · 40

Amazon Web Services

Es una colección de servicios de computación en la nube, ofrecidas a través de Internet por Amazon.com · 48

B

Bases de datos relacionales

Es una base de datos que cumple con el modelo relacional, permite establecer interconexiones (relaciones) entre los datos (que están guardados en tablas), y a través de dichas conexiones relacionar los datos de ambas tablas. · 15

BSON

Binary JSON, es un formato binario que es un superconjunto de JSON, el cual optimiza espacio, rendimiento y aporta funciones extra sobre JSON. Es un formato de intercambio de datos usado principalmente para su almacenamiento y transferencia en la base de datos MongoDB. · 41

Business Intelligence

Business Intelligence es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. · 21

C

Clúster

Se aplica a conjuntos o conglomerados de computadoras construidos mediante la utilización de componentes de hardware comunes y que se comportan como si fuesen una única computadora. · 38

D

Data Mart

Son almacenes de datos (DW) pequeños que están centrados en un tema y responden a los intereses de determinadas áreas de negocio dentro de una organización. Se caracteriza por disponer de la estructura óptima de datos para analizar la información al detalle. · 28

Data Mining

Es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos que expliquen el comportamiento de los datos en un determinado contexto. · 19

Data Warehouse

Es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. La creación de un datawarehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence. · 18

F

Ficheros

Es un conjunto de información que se almacena en forma virtual para ser leído y/o accedido por medio de una computadora. · 37

J

JSON

JavaScript Object Notation (JSON) es un estándar abierto, legible por máquina que facilita el intercambio de datos, y junto con XML es el formato principal para el intercambio de datos utilizado en la web moderna. JSON admite todos los tipos de datos básicos que se esperan . 41

O

On premise

Es el software y tecnología que se encuentra dentro de los límites físicos de una empresa. . 45

Es el software y tecnología que se encuentra dentro de los límites físicos de una empresa. Se puede traducir como aplicaciones locales. . 45

R

R

Es un lenguaje y entorno de programación para análisis estadístico y gráfico. . 39

Rack

Es un bastidor destinado a alojar equipamiento electrónico, informático y de comunicaciones. . 38

S

SGBD

(Sistema de gestión de base de datos) o en inglés Database management system (DBMS), es una agrupación de programas que sirven para definir, construir y manipular una base de datos. . 18

Sistemas de Información

Un sistema de información se puede definir técnicamente como un conjunto de componentes relacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar la toma de decisiones y el control en una organización. · 18

Software libre

Se refiere a la libertad de los usuarios para ejecutar, copiar, distribuir, estudiar, cambiar y mejorar el software. · 36

T

TIBCO

Empresa dedicada a crear productos tanto hardware como software para implementar soluciones empresariales basadas en arquitectura SOA · 32

BIBLIOGRAFÍA

- Acens. (2017). *acens.com*. Recuperado el 2 de Mayo de 2017, de <https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>
- Aguilar, L. J. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. México DF, México: Alfaomega Grupo Editor. Recuperado el 11 de Diciembre de 2016, de https://books.google.es/books?hl=es&lr=lang_es&id=1GywDAAAQBAJ&oi=fnd&pg=PT6&dq=big%20data%20&ots=_VV5Q6-kZS&sig=oYxS-ZEa0IjH7TQxGVc6XVyx6FU#v=onepage&q&f=false
- Alegsa, L. (5 de Diciembre de 2010). *Alegsa.com*. Obtenido de Dirección de Informática y Tecnología: <http://www.alegsa.com.ar/Dic/sghbd.php>
- Amazon Web Services. (2017). *AWS / Cloud Computing - Servicios de informática en la nube*. Recuperado el 16 de Mayo de 2017, de Amazon Web Services, Inc.: <https://aws.amazon.com/es/>
- Apache Hadoop. (2017). *What Is Apache Hadoop?* Recuperado el 05 de Mayo de 2017, de <http://hadoop.apache.org>
- Apache Spark. (2017). *Apache Spark*. Recuperado el 5 de Mayo de 2017, de Lightning-fast cluster computing: <http://spark.apache.org/>
- apser. IT plug & play. (25 de Noviembre de 2015). *APSER CLOUD SERVICES*. Recuperado el 14 de Junio de 2017, de <http://www.apser.es/blog/2015/11/25/comparativa-amazon-web-services-vs-microsoft-azure-vs-google-cloud-platform/>
- Arias Moreno, J., Guzmán Tapia, P., Solis Alarcón, J., Yacolca Neyra, R. J., & Zumaeta Mejía, D. (13 de Octubre de 2012). *SCRIBD*. Recuperado el 8 de Junio de 2017, de Pentaho Bi: <https://es.scribd.com/doc/109789629/Pentaho-Bi>

- Banchero, S. (Septiembre de 2015). *Bases de Datos Masivas*. (U. N. Luján, Ed.) Recuperado el 15 de Julio de 2017, de http://www.labredes.unlu.edu.ar/sites/www.labredes.unlu.edu.ar/files/site/data/bdm/clase_3_db_multidimensionales.pdf
- Caballero, R., & Martín, E. (2015). *Las bases de Big Data*. Madrid: Catarata.
- Castilla Alcala, G., Albarrán Lozano, I., & de Pablos Heredero, C. (1998). EL PROCESO DE IMPLANTACIÓN DEL DATA WAREHOUSE EN LA ORGANIZACIÓN: ANÁLISIS DE UN CASO. *Investigaciones Europeas de Dirección y Economía de la Empresa*, 4(3), 73-92. Recuperado el Marzo de 2017
- Choudhary, A. (5 de Mayo de 2015). *About Apache Spark, Lightning-fast cluster computing (Big Data)*. Recuperado el 8 de Junio de 2017, de LinkedIn: <https://www.linkedin.com/pulse/apache-spark-lightning-fast-cluster-computing-big-data-choudhary>
- Cía, J. F. (1 de Abril de 2015). *BBVA API_MARKET*. Recuperado el 9 de Junio de 2017, de bbvaopen4u.com: <https://bbvaopen4u.com/es/actualidad/tableau-contra-el-mundo-la-batalla-del-business-intelligence-se-calienta>
- Cock Gomez, G. (Noviembre de 2014). *UMNG trabajo de Investigacion GIOVANNI COCK VI.pdf*. Recuperado el 11 de Diciembre de 2016, de repository.unimilitar.edu.co: <http://repository.unimilitar.edu.co/bitstream/10654/13166/1/UMNG%20trabajo%20de%20Investigacion%20GIOVANNI%20COCK%20V1.pdf>
- Correa, M. A. (4 de Febrero de 2013). *macool.me*. Recuperado el 14 de Junio de 2017, de bitácora de @macool: <http://macool.me/mysql-vs-postgresql-vs-mongodb-velocidad/04>

- Curto Díaz, J. (2010). *Introducción al Business Intelligence*. Barcelona, España: Editorial UOC. Recuperado el 22 de Abril de 2017, de <http://site.ebrary.com/lib/pucesp/detail.action?docID=10647203>
- Czajkowski, L. (07 de Mayo de 2017). *DZone / Database Zone*. Recuperado el 22 de Septiembre de 2017, de <https://dzone.com/articles/a-comparison-of-sql-and-nosql-to-simplify-your-dat>
- Dean, T. (25 de Julio de 2015). *Bussiness 2 Community*. Recuperado el 22 de Abril de 2017, de Gain the Competitive Edge with Business Intelligence Software & Analytics: <http://www.business2community.com/business-intelligence/gain-competitive-edge-business-intelligence-software-analytics-01280045#ROFqivRqHmPdghWZ.99>
- Dip, P. (1 de Agosto de 2009). *Tecnología e Informática*. Recuperado el 30 de Abril de 2017, de latecnologiavirtual.blogspot.com/2009/08/datos.html
- Dirección Informática PUCE. (2017). *Matriculados Nuevos y Antiguos por Facultad*. Quito: PUCE.
- Flores A., P. A., & Villacís V., A. E. (2017). *Análisis Comparativo de Herramientas de Big Data para la Facultad de Ingeniería de la Pontificia Universidad Católica del Ecuador*. Quito: PUCE.
- Fundéu BBVA. (s.f.). *Fundéu*. Recuperado el 6 de Abril de 2017, de Macrodatos e inteligencia de datos, alternativas a big data: <http://www.fundeu.es/recomendacion/macrodatosalternativa-abig-data-1582/>
- Galarza, B., Tuamá, C., Zaccardi, G., Encinas, D., & Morales, M. (Septiembre de 2013). *Implementaciones de Cloud Computing y aplicaciones en el ámbito universitario*. Recuperado el 22 de Abril de 2017, de Instituto de ingeniería y agronomía:

<http://conaiisi.frc.utn.edu.ar/PDFsParaPublicar/PostersAlumnos/PosterFinalA4.pdf>

Gallardo, D. (2012). *PROYECTO DE TESIS: OLAP*. Recuperado el 15 de Julio de 2017, de <https://repositorio.espe.edu.ec/bitstream/21000/5123/2/DT-ESPEL-0900.pdf>

García Reyes, R. (Noviembre de 2012). *Minería de Datos para la Toma de Decisiones e Inteligencia de Negocios: Aplicaciones en la Mercadotecnia*. (U. N. México, Ed.) Recuperado el 1 de Mayo de 2017, de <http://132.248.9.195/ptd2012/noviembre/098505131/098505131.pdf>

García, D. L. (2012). *TFM - David López GarcíaS.pdf*. Recuperado el 11 de Diciembre de 2016, de [repositorio.unican.es: https://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADaS.pdf?sequence=1](https://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADaS.pdf?sequence=1)

Google Cloud Platform. (2017). *Google Cloud Computing, Hosting Services & APIs / Google Cloud Platform*. Recuperado el 23 de Mayo de 2017, de <https://cloud.google.com/>

Guarneros Sánchez. (10 de Abril de 2012). *McGus Iix*. Recuperado el 15 de Julio de 2017, de Bases de datos multidimensionales (BDM): <https://gusmc.wikispaces.com/1.2+Bases+de+datos+multidimensionales+%28BDM%29>

IBM. (2017). *IBM Analytics - Home - United States*. Recuperado el 31 de Mayo de 2017, de [Ibm.com: https://www.ibm.com/analytics/us/en/](https://www.ibm.com/analytics/us/en/)

Instituto Tecnológico de Sonora. (s.f.). *Introducción a los sistemas de información*. Recuperado el 6 de Abril de 2017, de http://biblioteca.itson.mx/oa/dip_ago/introduccion_sistemas/p3.htm

Jaspersoft. (2017). *TIBC Jaspersoft*. Recuperado el 9 de Junio de 2017, de <http://www.jasperlatam.com/>

- Jesuitas Ecuador. (2017). *Jesuitas Ecuador*. Recuperado el 22 de 07 de 2017, de <http://www.jesuitas.ec/pontificia-universidad-catolica-del-ecuador-puce/>
- Johnson P., F., & Rubio L., J. M. (2017). *Base de Datos*. Valparaíso: Universidad Católica de Valparaíso. Recuperado el 1 de Mayo de 2017, de <http://zeus.inf.ucv.cl/~jrubio/docs/2009-01/INF%20340/Capitulo%20I.pdf>
- lighpath. (18 de Mayo de 2017). *lighpath*. Recuperado el 22 de Julio de 2017, de APRENDIENDO SOBRE PROCESOS ETL: <http://www.lightpath.io/aprendiendo-sobre-procesos-etl/>
- López Rodríguez, J. (2012). *Licenciatura en Documentación: Bases de datos documentais*. Recuperado el 6 de Abril de 2017, de <http://docencia.lbd.udc.es/bdd/teoria/tema2/2.3.1.-ElModeloRelacional.pdf>
- Lublinsky, B., Smith, K., & Yakubovich, A. (2014). *Hadoop Soluciones Big Data*. Madrid, España: Ediciones Anaya Multimedia.
- Luna Ramirez, E., & García M., F. (2006). *El repositorio de metadatos en un data warehouse*. Arica, Chile: Red Revista de Facultad de Ingeniería.
- Macías, M., & Gómez, M. (2015). *Introducción a Apache Spark: para empezar a programar el big data*. Barcelona, España: Editorial UOC.
- Malpica Velasco, J. A. (s.f.). *Introducción a la Inteligencia Artificial*. Recuperado el 01 de 02 de 2017, de Inteligencia Artificial y conciencia: <http://www.cs.upc.edu/~bejar/ia/transpas/teoria/1-IA-introduccion.pdf>
- Manzano de Armas, J., Molina Pérez, D., & Castro Soler, A. (17 de Abril de 2017). *Implementación de un Data Mart para un servicio de Dosimetría Externa*. Recuperado el 15 de Julio de 2017, de https://www.researchgate.net/profile/Jose_Manzano_De_Armas3/publication/276920105_IMPLEMENTACION_DE_UN_DATA_MART_PARA_UN_SERVICIO_DE_DOSIMETRIA_EXTERNA/links/555b9c9608ae6aea0816c871

/IMPLEMENTACION-DE-UN-DATA-MART-PARA-UN-SERVICIO-DE-DOSIMETRIA-EXTERNA.pd

Martín, A., Chávez, S. B., Rodríguez, N. R., Valenzuela, A., & Murazzo, M. A. (7 de Junio de 2013). *Bases de datos NoSQL en cloud computing*. Recuperado el 22 de Abril de 2017, de SEDICI: <http://sedici.unlp.edu.ar/handle/10915/27121>

Martin, S. (20 de Abril de 2017). *Bases de datos NoSQL : Guía definitiva*. Recuperado el 07 de Julio de 2017, de Pandorafms: <https://blog.pandorafms.org/es/bases-de-datos-nosql/>

Microsoft. (Julio de 2016). *Microsoft | Developer Network*. Recuperado el 15 de Julio de 2017, de [https://msdn.microsoft.com/es-es/library/hh916543\(v=sc.12\).aspx](https://msdn.microsoft.com/es-es/library/hh916543(v=sc.12).aspx)

Microsoft. (2017). *Microsoft*. Recuperado el 1 de Mayo de 2017, de <https://support.office.com/es-es/article/Conceptos-b%C3%A1sicos-sobre-bases-de-datos-a849ac16-07c7-4a31-9948-3c8c94a7c204>

Microsoft. (2017). *Microsoft | Power BI*. Recuperado el 9 de Junio de 2017, de <https://powerbi.microsoft.com/es-es/what-is-power-bi/>

Microsoft. (2017). *Microsoft Azure*. Recuperado el 25 de Abril de 2017, de <https://azure.microsoft.com/es-es/overview/what-is-azure/>

MongoDB. (2017). *MongoDB Atlas Database as a Service*. Recuperado el 05 de Mayo de 2017, de <https://www.mongodb.com/>

Morales Mateo, Y. S. (22 de Septiembre de 2012). *basededatoscms.files.wordpress.com*. Recuperado el 10 de Agosto de 2017, de <https://basededatoscms.files.wordpress.com/2012/09/interacion-de-datos-y-almacenes-de-datos.pdf>

Noticias Financieras. (26 de Octubre de 2015). *Big data, herramienta inteligente para los negocios*. Recuperado el 13 de Marzo de 2017, de ProQuest:

<http://puceec.summon.serialssolutions.com/#!/search?ho=t&l=es-ES&q=big%20data%20herramientas>

Open Software Integrators, LLC. (2017). *Scribd*. Recuperado el 14 de Junio de 2017, de MongoDB vs Hadoop. Big Solutions for Big Problems: <https://es.scribd.com/document/261862150/Hadoop-vs-MongoDB>

Oracle. (2015). *The Surprising Economics of Engineered*. (Oracle, Ed.) Recuperado el 1 de Junio de 2017, de <http://www.oracle.com/us/technologies/big-data/eng-systems-for-big-data-esg-wp-2852701.pdf>

Oracle. (2017). *Docs.oracle.com*. Recuperado el 4 de Abril de 2017, de Data Warehousing Concepts: https://docs.oracle.com/cd/A87860_01/doc/server.817/a76994/concept2.gif

Oracle. (2017). *Oracle: Big Data for the Enterprise*. Recuperado el 1 de Junio de 2017, de <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>

Palma, C., Palma, W., & Pérez, R. (2009). *Data mining: el arte de anticipar* (Primera ed.). Santiago de Chile, Chile: RIL editores.

Paniagua, S. (10 de Agosto de 2014). *Bidooop, data science con un simple drag and drop*. Recuperado el 16 de Mayo de 2017, de <http://www.sorayapaniagua.com/2014/08/10/bidooop-data-science-con-un-simple-drag-and-drop/>

Pérez Marqués, M. (2015). *Big Data. Técnicas, herramientas y aplicaciones*. Madrid: RC Libros.

Pérez, C. (2015). *Big Data Analytics con herramientas de SAS, IBM, ORACLE y MICROSOFT*. Madrid: Ibergarceta Publicaciones.

Prabhu, S., & Venkatesan, N. (2006). *Data Mining and Warehousing (1)*. New Age International. Recuperado el 11 de Diciembre de 2016

- Pragsis Bidoop. (2017). *Big Data & Analytics Experts / Pragsis Bidoop*. Recuperado el 16 de Mayo de 2017, de Bidoop.es: <http://www.bidoop.es/?language=en>
- PUCE. (2017). *Pontificia Universidad Católica del Ecuador*. Recuperado el 14 de Mayo de 2017, de <http://www.puce.edu.ec/autoridades.php>
- Qlik. (2017). *Qlik*. Recuperado el 16 de Mayo de 2017, de https://www.google.com.ec/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwj_odPl6_TTAhXCWCYKHc1dCFQQjRwIBw&url=http%3A%2F%2Fwww.aliosoluciones.com%2Fqlikview&psig=AFQjCNFKI-eo-qHdwhVimGzzhewh04HE4Q&ust=1495038518826195
- renacens. (2017). *INTELIGENCIA DE NEGOCIO - BI*. Recuperado el 8 de Junio de 2017, de APROVECHE EL GRAN VALOR DE LA INFORMACIÓN: <http://www.renacens.es/web/business-intelligence---bi>
- Rey, A. (1 de Marzo de 2012). *Cloud computing: El servicio de almacenamiento en la nube*. Recuperado el 8 de Junio de 2017, de http://www.huesca.es/_archivos/ficheros/bibliotecas_2651.pdf
- Rosado Gómez, A., & Rico Bautista, D. W. (2010). INTELIGENCIA DE NEGOCIOS: ESTADO DEL ARTE. *Universidad Tecnológica de Pereira*, 1(44), 321-326.
- Rouse, M. (Noviembre de 2012). *SearchDataCenter*. Recuperado el 6 de Junio de 2017, de TechTarget: <http://searchdatacenter.techtarget.com/es/definicion/Analisis-de-Datos>
- Sage. (2015). *DEL USO DEL BIG DATA, Beneficios.Big Data para pymes*. Recuperado el 11 de Diciembre de 2016, de Sage Spain: http://www.sage.es/~media/markets/es/recursos%20de%20negocio/infografias/infografia%20biga%20data%20pymes/infografia_sage-big-data-para-pymes.pdf

- Sarasa, A. (2016). *Introducción a las bases de datos NoSQL usando MongoDB*. Barcelona, España: Editorial UOC.
- Sinnexus. (s.f.). *Sinnexus, Business Intelligence, Informática Estratégica*. Recuperado el 6 de Abril de 2017, de Datawarehouse: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx
- Solid IT. (2017). *DB-Engines. Knowledge Base of Relational and NoSQL Database Management Systems*. Recuperado el 14 de Junio de 2017, de System Properties Comparison MongoDB vs. Spark SQL: <https://db-engines.com/en/system/MongoDB%3BSpark+SQL>
- Tableau. (2017). *Tableau*. Recuperado el 9 de Junio de 2017, de <https://www.tableau.com/es-es>
- Tascón, M., & Coullaut, A. (2016). *Big Data y el Internet de las cosas. Qué hay detrás y cómo nos va a cambiar*. Madrid: Catarata.
- Taylor, & Lily. (2 de Diciembre de 2016). *jombangkit*. Recuperado el 4 de Abril de 2017, de <http://www.jombangkit.org/>
- Turck, M. (2016). *Matt Turck*. Recuperado el 18 de Abril de 2017, de <http://mattturck.com/big-data-landscape/>
- UCLA. (13 de Diciembre de 2006). *www.ucla.edu.ve*. Recuperado el 10 de Agosto de 2017, de <http://www.ucla.edu.ve/dcee/Departamentos/coordinaciones/informaticai/documentos/resumen%20tema1.pdf>
- Van Rijmenam, M. (2014). *Think Bigger : Developing a Successful Big Data Strategy for Your Business (1)*. AMACOM. Recuperado el 11 de Diciembre de 2016
- Vega, D. (14 de Marzo de 2012). *Software libre y Software propietario: cuadro comparativo*. Recuperado el 30 de Mayo de 2017, de

<http://sis19upt.blogspot.com/2012/03/software-libre-y-software-propietario.html>

Vera, A. (Enero de 2017). *Steemit Beta*. Recuperado el 7 de Abril de 2017, de Big Data: <https://steemit.com/spanish/@javf1016/big-data>

Wikipedia. (30 de Mayo de 2017). *Cloud computing comparison*. Recuperado el 1 de Junio de 2017, de https://en.wikipedia.org/wiki/Cloud_computing_comparison

WORLD OF DATA. (11 de Julio de 2014). *WORLDofDATA.es*. Recuperado el 10 de Diciembre de 2016, de <http://worldofdata.es/6-grandes-frases-sobre-big-data/>

Zentut. (2017). *What is Data Mining*. Recuperado el 4 de Abril de 2017, de Zentut: <http://www.zentut.com/data-mining/what-is-data-mining/>

ANEXOS

ANEXO A

Tipo: Documento

Nombre del archivo: Aranceles2016.pdf

Descripción: Aranceles 2016-2017 por carrera de la PUCE Matriz.

ANEXO B

Tipo: Imagen

Nombre del archivo: AutoridadesPuceEsmeraldas.png

Descripción: Autoridades de la PUCE en la Sede de Esmeraldas.

ANEXO C

Tipo: Imagen

Nombre del archivo: Big-Data-Landscape-2016-v18-FINAL.png

Descripción: Imagen representativa del ecosistema de Big Data clasificada por tipos.

ANEXO D

Tipo: Documento

Nombre del archivo: PUCE-ING-IngenierIa-Civil.pdf

Descripción: Malla Curricular de la carrera de Ingeniería Civil de la PUCE.

ANEXO E

Tipo: Documento

Nombre del archivo: PUCE-ING-IngenierIa-Sistemas.pdf

Descripción: Malla Curricular de la carrera de Ingeniería en Sistemas y Computación de la PUCE.

ANEXO F

Tipo: Documento

Nombre del archivo: 04 ISC 328 Manual MongoDB.pdf

Descripción: Manual de instalación y uso de MongoDB.

ANEXO G

Tipo: Documento

Nombre del archivo: hadoop_tutorial.pdf

Descripción: Manual de uso Hadoop.

ANEXO H

Tipo: Documento

Nombre del archivo: cloudera-spark.pdf

Descripción: Manual de uso Spark.

ANEXO I

Tipo: Documento

Nombre del archivo: MongoDBInstalacionLinux.pdf

Descripción: Manual de instalación de MongoDB para Linux.

ANEXO J

Tipo: Documento

Nombre del archivo: ORGANIGRAMA-GENERAL-2014_ibarra.pdf

Descripción: Organigrama Directivo General de la PUCE, Sede Ibarra

ANEXO K

Tipo: Documento

Nombre del archivo: PUCE-ORGANIGRAMA-032017.pdf

Descripción: Organigrama Directivo PUCE, Matriz Quito

ANEXO L

Tipo: Documento

Nombre del archivo: organigrama-estructural-puceesmeraldas2016.pdf

Descripción: Organigrama Directivo PUCE, Sede Esmeraldas

ANEXO M

Tipo: Documento

Nombre del archivo: ORGANIGRAMA-PUCE-22-NOV-2016Manabi.pdf

Descripción: Organigrama Directivo PUCE, Sede Manabí

ANEXO N

Tipo: Documento

Nombre del archivo: ORGANIGRAMA-PUCE-22-NOV-2016Manabi.pdf

Descripción: Organigrama Directivo PUCE, Sede Manabí

ANEXO O

Tipo: Documento

Nombre del archivo: organigrama2017puceSantoDomingo.pdf

Descripción: Organigrama Directivo PUCE, Sede Santo Domingo

ANEXO P

Tipo: Documento

Nombre del archivo: eng-systems-for-big-data-esg-wp-2852701.pdf

Descripción: Paper acerca de Oracle Big Data, especificando la importancia, características y relevancia del mismo.

ANEXO Q

Tipo: Documento

Nombre del archivo: Apser-tabla-comparativa-grandes-cloud.jpg

Descripción: Tabla Comparativa de Herramientas en la Nube, describiendo las características principales entre ellas.

ANEXO R

Tipo: Documento

Nombre del archivo:

Microsoft_Press_ebook_Introducing_Power_BI_PDF_mobile.pdf

Descripción: Introducción a Microsoft Power BI, con una guía práctica para su respectivo uso, obteniendo una visión general del mismo.

ANEXO S

Tipo: Documento

Nombre del archivo: Analyzing Sales Data Tutorial.pdf

Descripción: Análisis de datos de ventas en Excel, importación de datos desde Excel a Microsoft Power BI, para determinar cómo es el uso de la herramienta para extracción de datos.

ANEXO T

Tipo: Documento

Nombre del archivo: QV Tutorial.pdf

Descripción: Tutorial sobre QlikView, explicación y uso de la herramienta, con ejercicios prácticos.

ANEXO U

Tipo: Documento

Nombre del archivo: -Pentaho Data Integration Cookbook, Second Edition.pdf

Descripción: Tutorial sobre Pentaho, con ejercicios prácticos y su respectiva explicación.

ANEXO V

Tipo: Documento

Nombre del archivo: Jaspersoft-Studio-User-Guide.pdf

Descripción: Tutorial sobre Jaspersoft, introducción y uso de la herramienta.

ANEXO W

Tipo: Documento

Nombre del archivo: tableau-tutorial-for-beginners.pdf

Descripción: Tutorial para principiantes sobre Tableau, introducción y uso de la herramienta.

ANEXO X

Tipo: Documento

Nombre del archivo: GCP.pdf

Descripción: Introducción a Google Cloud Platform.

ANEXO Y

Tipo: Documento

Nombre del archivo: Watson-Analytics-tutorial-2017-07-19.pdf

Descripción: Introducción a Watson Analytics, definiciones y ejercicios prácticos.

ANEXO Z

Tipo: Documento

Nombre del archivo: microsoft_azure_tutorial.pdf

Descripción: Tutorial de Microsoft Azure.

ANEXO AA

Tipo: Documento

Nombre del archivo: amazon_web_services_tutorial.pdf

Descripción: Manual de uso Amazon Web Services.