

COORPORACIÓN UNIVERSITARIA MINUTO DE DIOS - UNIMINUTO

ESPECIALIZACIÓN TECNOLÓGICA EN HERRAMIENTAS PARA LA

GESTIÓN DE DATOS BIG DATA

**ESTADO DEL ARTE DEL BIGDATA**

**Imagen que contiene Logotipo

Descripción generada automáticamente**

**Integrantes**

GUSTAVO ANDRES DIAZ PRIETO (000792780)

JHON MAURICIO BOHORQUEZ (000376969)

DAVID ALBERTO DIAZ SIERRA (000870422)

GIOVANNY PRIETO ANGULO (000860570)

OSCAR MAURICIO GARCIA RODRIGUEZ (000855059)

ANGELO RIOS RAMIREZ (000544544)

**Bogotá D.C.**

**04.03.2023**

Bogotá. 2023

Contenido

[1. INTRODUCCIÓN 3](#_Toc128660066)

[2. BIG DATA 4](#_Toc128660067)

[2.1 Historia del Big Data 4](#_Toc128660068)

[2.2 Impacto del Big Data en el mundo 5](#_Toc128660069)

[2.3 Tecnologías de análisis de datos 5](#_Toc128660070)

[2.4 Herramientas para la gestión de datos 6](#_Toc128660071)

[2.5 Beneficios del Big Data 13](#_Toc128660072)

[3. VISUALIZACIÓN DE DATOS 14](#_Toc128660073)

[3.1 PowerBi 14](#_Toc128660074)

[3.2 DataStudio 17](#_Toc128660075)

[4. CONCLUSIONES Y RETOS PARA EL BIG DATA 19](#_Toc128660076)

1. INTRODUCCIÓN

La capacidad actual para producir información (**datos**) ha crecido exponencialmente respecto a

años anteriores. La enorme cantidad de datos a disposición hace necesario el desarrollo de herramientas que permitan el procesado y análisis de estos, para identificar y extraer la información relevante; hoy en día las compañías no saben qué hacer con el gran volumen de datos e información almacenada en diferentes medios o bases de datos, los cuales pueden ser de gran importancia, principalmente en la toma de decisiones.

La eficiente captura y análisis de los datos redundará en la competitividad de las empresas, desde la fabricación, cadena de suministro, hasta los sistemas de gestión, procesos que pueden ser mejorados a través del uso de Big Data y entrar en una era de interoperabilidad entre entidades que puedan suministrar información de calidad que aporten para el crecimiento en todos los sectores.

Para dar sentido al término Big Data y que la industria disponga de mecanismos para la toma de mejores decisiones, conocer mejor su negocio, generar posibles oportunidades de negocio y verificar o refutar teorías y modelos existentes, es necesario incorporar dos términos más, necesarios, para el desarrollo de Big Data:

1. Big Data Analytics: Ciencia de examinar datos en bruto con el propósito de sacar conclusiones sobre esa información. Implica aplicar un proceso algorítmico o mecánico para obtener conocimiento. Por ejemplo, aplicar un proceso para buscar correlaciones significativas entre varias series de datos.
2. Cloud Computing: Modelo para permitir el acceso conveniente y bajo demanda a un conjunto compartido de recursos computacionales configurables (por ejemplo, redes, servidores, almacenamiento, aplicaciones y servicios), que se pueden aprovisionar y liberar rápidamente con un esfuerzo mínimo de gestión o una interacción entre el proveedor de servicios [[1]](#footnote-1)
3. BIG DATA

Cuando hablamos de Big Data nos referimos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles. Por esta razón se describen a continuación una serie de aspectos que se deben tener en cuenta para conocer lo que es Big Data.

* 1. Historia del Big Data

El término "Big Data" se refiere a conjuntos de datos extremadamente grandes y complejos que no pueden ser procesados y analizados mediante herramientas y técnicas tradicionales. La historia de Big Data se remonta a la década de 1990, cuando las empresas comenzaron a recopilar grandes cantidades de datos en sus sistemas empresariales. Sin embargo, no fue hasta la década de 2000 cuando el término "Big Data" comenzó a utilizarse con frecuencia.

En 2001, Doug Laney, analista de Gartner, definió el concepto de "Big Data" como un conjunto de tres características: Volumen, Velocidad y Variedad. Estos tres V's se convirtieron en la base para la comprensión de Big Data.

El surgimiento de la web y la explosión de los dispositivos móviles y las redes sociales han generado una enorme cantidad de datos, lo que ha llevado a una necesidad aún mayor de herramientas y técnicas para procesar y analizar grandes conjuntos de datos. Las tecnologías de Big Data, como Apache Hadoop, Apache Spark, y otras plataformas de Big Data han sido desarrolladas para abordar estos desafíos.

Desde entonces, la tecnología de Big Data ha avanzado rápidamente y ha habido muchos desarrollos importantes en la historia de Big Data. En 2010, Google lanzó su tecnología de procesamiento de datos a gran escala llamada "Google File System". En 2012, Facebook creó y lanzó "Hadoop Distributed File System" (HDFS) que se convirtió en uno de los marcos de Big Data más populares. En el mismo año, Apache Spark fue lanzado como una alternativa a Hadoop que proporciona una velocidad de procesamiento más rápida.

En 2014, la adopción de Big Data se expandió más allá del mundo empresarial y se convirtió en una parte importante de la investigación científica y de la salud. Las aplicaciones de Big Data en la investigación de la salud y la medicina han permitido a los científicos identificar patrones y tendencias en grandes conjuntos de datos para ayudar en el diagnóstico y tratamiento de enfermedades.

En los últimos años, la Inteligencia Artificial (IA) y el Aprendizaje Automático (AA) se han integrado en las tecnologías de Big Data para hacer que los procesos de análisis sean aún más precisos y eficientes. Las herramientas de IA y AA permiten que los algoritmos de análisis aprendan y se ajusten automáticamente a medida que procesan más datos.

En conclusión, la historia de Big Data es relativamente corta, pero ha experimentado un crecimiento exponencial en los últimos años. La capacidad de procesar y analizar grandes conjuntos de datos ha permitido a las empresas tomar decisiones más informadas y ha permitido avances en la investigación y la ciencia. Se espera que el crecimiento y la evolución de Big Data continúen en el futuro con la aparición de nuevas tecnologías y aplicaciones.

* 1. Impacto del Big Data en el mundo
  2. Tecnologías de análisis de datos

Dada la magnitud y complejidad de los desarrollos, métodos, técnicas, etc. que actualmente se incluyen bajo la denominación Big Data, dividiremos las tendencias en Tecnologías de análisis de datos y las Herramientas para la gestión de datos a Gran Escala:

* **Microsoft Power BI**: Líder de 2020 como plataforma de analítica de datos e inteligencia de negocio. Gracias a sus dos versiones Power BI Desktop (gratuita) y, Power BI Pro (precio por usuario), permiten al usuario crear potentes cuadros de mando (Dashboard) de forma muy sencilla e intuitiva.
* **Tableau**: Se trata de una plataforma de análisis extremo a extremo (end to end), ya que permite preparar los datos antes de su posterior procesado con la herramienta Prep. También incorpora un sistema de arrarstrar y soltar en la hoja de trabajo (Tableau Desktop) para generar potentes gráficas dinámicas, sobre las que podemos resaltar datos en concreto, hacer zoom, etc
* **Qlik:** Qlik View y Qlik Sense, siendo esta última la más moderna y que, de hecho, supone la transición de View a Sense. Qlik Sense, se utiliza como un gestor de transformación, una plataforma que utiliza un novedoso motor asociativo de análisis (incorporando Inteligencia Artificial) para sugerir conocimientos y, automatizar procesos.
* **ThoughtSpot:** Permite realizar operaciones de computación paralela, es posible procesar 100 billones de filas en 100 segundos. Aprende sobre las analíticas que realiza el usuario para ofrecer insights, ocultos para mejorar la toma de decisiones a través de cuadros de mando. Los precios por uso de la aplicación están sujetos por la capacidad de procesamiento de datos.
* **MicroStrategy:** Como plataforma de análisis de datos, proporciona todas las funciones que necesita cualquier compañía para desarrollar aplicaciones que permitan a través del dato, conocer la información de la empresa y acelerar los negocios. Además de la capacidad de generar informes personalizados o, cuadros de mando, también incluye aplicaciones de aprendizaje automático (Machine Learning), Big Data (conexión nativa con Hadoop), Cloud computing, etc
* **Looker:** Enfocada para extraer el máximo valor de los datos, Looker ofrece dentro de sus soluciones herramientas BI para realizar análisis modernos, realizando cuadros de mando en tiempo real, fácil acceso a los datos, flujos de trabajo (workflow) basados en los datos, es capaz de conectarse con múltiples fuentes de datos como Redshift, Snowflake, BigQuery y, otras 50 herramientas SQL
* **Tibco Spotfire:** Esta herramienta, incluye IA basada en el manejo de los datos, para incluir funciones de analítica avanzada, como, por ejemplo, sistemas de recomendación en búsqueda. Incluye módulos para realizar: Location Analytics, análisis en tiempo real a través de Data Streams, Data Wranling o, analítica productiva utilizando funciones de Python, R, SAS y Matlab para realizar ciencia de datos. Su precio por licencia es de 1.300$ por año y usuario[[2]](#footnote-2)
  1. Herramientas para la gestión de datos

Herramientas para la gestión de datos

Herramientas de gestión de datos son un conjunto de herramientas de software diseñadas para ayudar a los usuarios a almacenar, administrar, manipular y analizar grandes cantidades de datos. Estas herramientas facilitan la recopilación y el análisis de información para la toma de decisiones informadas. Algunas herramientas populares para la gestión de datos incluyen bases de datos relacionales, bases de datos no relacionales, herramientas de análisis de datos, herramientas de visualización de datos, herramientas de limpieza de datos, herramientas de migración de datos, etc. Estas herramientas pueden ayudar a los usuarios a organizar y analizar datos de una manera eficiente y proporcionar información valiosa para ayudar a los usuarios a tomar mejores decisiones.

Algunas de las herramientas para la gestión de datos de mayor impacto en el mundo son:

1. Apache Hadoop

2. Apache Spark

3. Microsoft Azure

4. IBM Watson

5. SAS

6. Oracle Database

7. MongoDB

8. Microsoft SQL Server

9. Tableau

10. AWS Redshift

Algunas de las Herramientas para la gestión de datos más utilizadas en Colombia.

1. MongoDB

2. Oracle Database

3. Tableau

4. Microsoft SQL Server

5. Microsoft Access

6. Apache Hadoop

7. IBM DB2

8. Apache Spark

9. Talend Data Integration

10. QlikView

Las herramientas para la gestión de datos se pueden clasificar por grupos. Estos grupos son:

1. Herramientas de administración de bases de datos: estas herramientas ayudan a los usuarios a crear, administrar y almacenar los datos en una base de datos. Estas herramientas se utilizan para crear tablas, consultas, formularios, informes y otras características de una base de datos.

2. Herramientas de análisis de datos: estas herramientas permiten a los usuarios analizar y manipular los datos almacenados en una base de datos. Estas herramientas permiten hacer preguntas sobre los datos y obtener respuestas para ayudar a tomar decisiones.

3. Herramientas de visualización de datos: estas herramientas ayudan a los usuarios a ver los datos

Python

Python se puede usar para gestionar datos. Es un lenguaje de programación que se puede usar para escribir scripts para procesar, manipular y analizar datos. Además, muchas de las bibliotecas de Python, como Pandas, se pueden usar para manipular y analizar datos. Python también se puede usar para automatizar tareas repetitivas en el procesamiento de datos, como la limpieza de datos y la preparación de datos.

Características:

1. Lenguaje de programación de alto nivel: Python es un lenguaje de programación de alto nivel, lo que significa que es fácil de leer y entender.

2. Interpretado: Python es un lenguaje de programación interpretado, lo que significa que no necesita compilar el código antes de su ejecución.

3. Multiparadigma: Python es un lenguaje de programación multiparadigma, lo que significa que admite diferentes paradigmas, como programación orientada a objetos, programación funcional, programación imperativa, etc.

4. Multiplataforma: Python es un lenguaje de programación multiplataforma, lo que significa que se puede ejecutar en diferentes sistemas operativos.

5. Amplia biblioteca de herramientas: Python viene con una amplia biblioteca de herramientas que se pueden usar para desarrollar, depurar y probar código.

StudioR

StudioR es una plataforma de análisis de datos basado en la nube que se puede usar para realizar análisis avanzados de datos y obtener resultados de forma rápida. Esta plataforma se puede utilizar para transformar, limpiar, modelar y visualizar datos en tiempo real. También se puede utilizar para crear análisis predictivos y obtener conclusiones de los datos. Esta plataforma se puede integrar con otros sistemas de análisis de datos, como Apache Spark, Apache Hadoop y otros. Esto le permite extraer datos de diferentes fuentes y aprovechar la potencia de estas herramientas para obtener resultados con mayor rapidez.

Características:

1. Visualización en tiempo real: StudioR ofrece una amplia gama de gráficos interactivos para ayudar al usuario a comprender mejor los datos.

2. Preparación de datos: StudioR ofrece herramientas para ayudar al usuario a limpiar, transformar y normalizar los datos antes de su análisis.

3. Análisis predictivo: StudioR ofrece la capacidad de realizar análisis predictivos para ayudar al usuario a tomar decisiones informadas.

4. Detección de errores: StudioR ofrece herramientas para ayudar al usuario a detectar y corregir errores en los datos.

5. Compatibilidad con lenguajes: La herramienta es compatible con los principales lenguajes de programación, como Python, R, Julia y Scala.

MongoDB

MongoDB es una base de datos no relacional de código abierto que permite a los usuarios almacenar y administrar grandes volúmenes de datos. Es una herramienta flexible y escalable que se puede usar para muchas aplicaciones y casos de uso, desde aplicaciones web y móviles hasta sistemas de análisis de datos. Su interfaz de línea de comandos orientada a documentos facilita el almacenamiento y recuperación de datos, mientras que sus herramientas de administración permiten a los usuarios administrar fácilmente la base de datos. Además, MongoDB proporciona funcionalidad de replicación y alta disponibilidad, lo que garantiza que los datos estén siempre seguros y disponibles.

Características:

1. Almacenamiento de datos en formato de documento: MongoDB almacena y administra datos en formato de documento en lugar de la estructura relacional de tablas de bases de datos relacionales, lo que permite una mayor flexibilidad para el almacenamiento de datos.

2. Escalabilidad: MongoDB es fácilmente escalable y puede manejar grandes volúmenes de datos sin tener que cambiar la estructura de la base de datos.

3. Rendimiento: MongoDB se ha diseñado para ofrecer un rendimiento óptimo en entornos de producción.

4. Seguridad: MongoDB ofrece funcionalidad de seguridad para proteger los datos, como autenticación, encriptación, control de acceso, auditoría y más.

5. Herramientas de administración: MongoDB proporciona herramientas de administración fáciles de usar para ayudar a los usuarios a administrar y mantener la base de datos.

6. Replicación y alta disponibilidad: MongoDB proporciona funcionalidad de replicación y alta disponibilidad para garantizar que los datos estén siempre seguros y disponibles.

Oracle Database

Oracle Database es una base de datos relacional de código abierto que permite a los usuarios almacenar, administrar y manipular grandes volúmenes de datos. Ofrece robustas herramientas de administración, una amplia variedad de lenguajes de consulta y herramientas de análisis para ayudar a los usuarios a obtener información útil de sus datos. Oracle Database también ofrece funcionalidad de replicación y alta disponibilidad para garantizar que los datos estén seguros y disponibles en todo momento. Esta herramienta es la base de datos relacional de código abierto más popular en el mundo, y se utiliza ampliamente en empresas de todos los tamaños.

Características:

1. Seguridad: Oracle Database ofrece funcionalidad de seguridad para proteger los datos, como autenticación, encriptación, control de acceso, auditoría y más.

2. Herramientas de administración: Oracle Database proporciona herramientas de administración fáciles de usar para ayudar a los usuarios a administrar y mantener la base de datos.

3. Lenguajes de consulta: Oracle Database ofrece una amplia variedad de lenguajes de consulta para ayudar a los usuarios a obtener información útil de sus datos.

4. Rendimiento: Oracle Database se ha diseñado para ofrecer un rendimiento óptimo en entornos de producción.

5. Replicación y alta disponibilidad: Oracle Database proporciona funcionalidad de replicación y alta disponibilidad para garantizar que los datos estén siempre seguros y disponibles.

6. Integración de aplicaciones: Oracle Database ofrece un conjunto de herramientas de integración de aplicaciones para facilitar la conexión de la base de datos a otras aplicaciones.

Tableau

Tableau es una herramienta de visualización de datos de código abierto que permite a los usuarios crear gráficos y visualizaciones a partir de sus datos. Esta herramienta proporciona una interfaz intuitiva para la creación de visualizaciones, lo que permite a los usuarios crear gráficos complejos en poco tiempo. También ofrece herramientas para el análisis de datos avanzados, que permiten a los usuarios profundizar en sus datos para obtener información valiosa. Además, Tableau se conecta a una variedad de bases de datos relacionales y no relacionales, lo que facilita la conexión de la herramienta a sus datos.

Características:

1. Visualización de datos: Tableau permite a los usuarios crear gráficos y visualizaciones a partir de sus datos con una interfaz intuitiva.

2. Herramientas de análisis avanzadas: Tableau ofrece herramientas para el análisis avanzado de datos, que permiten a los usuarios profundizar en sus datos para obtener información valiosa.

3. Capacidad de conexión: Tableau se conecta a una variedad de bases de datos relacionales y no relacionales, lo que facilita la conexión de la herramienta a sus datos.

4. Compatibilidad con dispositivos móviles: Tableau ofrece soporte para dispositivos móviles, lo que permite a los usuarios acceder a sus visualizaciones desde cualquier lugar.

5. Herramientas de compartir: Tableau proporciona herramientas para compartir visualizaciones con otros usuarios para facilitar la colaboración.

Microsoft SQL Server

Microsoft SQL Server es una base de datos relacional de código abierto que permite a los usuarios almacenar, administrar y manipular grandes volúmenes de datos. Esta herramienta proporciona una amplia variedad de lenguajes de consulta para ayudar a los usuarios a obtener información útil de sus datos, así como herramientas de administración para ayudar a los usuarios a administrar y mantener la base de datos. Además, Microsoft SQL Server ofrece funcionalidad de replicación y alta disponibilidad para garantizar que los datos estén siempre seguros y disponibles. Esta herramienta es una de las bases de datos relacionales de código abierto más populares en el mundo.

Características:

1. Lenguajes de consulta: Microsoft SQL Server ofrece una amplia variedad de lenguajes de consulta para ayudar a los usuarios a obtener información útil de sus datos.

2. Herramientas de administración: Microsoft SQL Server proporciona herramientas de administración fáciles de usar para ayudar a los usuarios a administrar y mantener la base de datos.

3. Rendimiento: Microsoft SQL Server se ha diseñado para ofrecer un rendimiento óptimo en entornos de producción.

4. Seguridad: Microsoft SQL Server ofrece funcionalidad de seguridad para proteger los datos, como autenticación, encriptación, control de acceso, auditoría y más.

5. Replicación y alta disponibilidad: Microsoft SQL Server proporciona funcionalidad de replicación y alta disponibilidad para garantizar que los datos estén siempre seguros y disponibles.

6. Integración de aplicaciones: Microsoft SQL Server ofrece un conjunto de herramientas de integración de aplicaciones para facilitar la conexión de la base de datos a otras aplicaciones.

Microsoft Access

Microsoft Access es una herramienta de base de datos relacional desarrollada por Microsoft. Está diseñada para ayudar a los usuarios a crear y administrar bases de datos de forma sencilla. Las bases de datos creadas con Access pueden contener tablas, consultas, formularios e informes, entre otros elementos. Está disponible para uso en ordenadores personales con Windows. Access también se utiliza para conectar bases de datos a aplicaciones de Internet, como páginas web.

Características:

1. Tablas: las tablas contienen los detalles sobre los datos almacenados en la base de datos.

2. Consultas: las consultas le permiten buscar y filtrar datos específicos de la base de datos.

3. Formularios: los formularios proporcionan una interfaz de usuario para ingresar, modificar y visualizar los datos.

4. Informes: los informes le permiten presentar los datos de una manera fácilmente legible.

5. Macros: las macros le permiten asignar tareas a la base de datos con un solo clic.

6. Páginas web: Access le permite conectar una base de datos a una página web. Esto le permite crear aplicaciones web que utilicen los datos almacenados en la base de datos.

Apache Hadoop

Apache Hadoop es una plataforma de código abierto para almacenamiento y procesamiento distribuido de datos. Está diseñado para ejecutar aplicaciones en clústeres de hardware de bajo costo. Hadoop es una de las plataformas de procesamiento de datos más populares para manejar grandes volúmenes de datos en una variedad de aplicaciones. Está diseñado para escalar fácilmente desde una sola computadora hasta miles de máquinas, proporcionando una excelente velocidad y rendimiento. La plataforma Hadoop proporciona diversas herramientas y frameworks, como HBase, Hive, Pig, y Spark, para facilitar el almacenamiento y procesamiento de datos.

Características:

1. Almacenamiento escalable y distribuido: Hadoop está diseñado para escalar fácilmente desde una sola computadora hasta miles de máquinas.

2. Procesamiento de datos paralelo: Hadoop proporciona un mecanismo para procesar grandes volúmenes de datos de manera paralela. Esto permite a Hadoop procesar grandes cantidades de datos mucho más rápido que una sola computadora.

3. Herramientas y frameworks para el almacenamiento y el procesamiento de datos: Hadoop incluye herramientas y frameworks tales como HBase, Hive, Pig y Spark para facilitar el almacenamiento y procesamiento de datos.

4. Capacidad de procesamiento de datos no estructurados: Hadoop es capaz de procesar grandes cantidades de datos no estructurados como imágenes, videos y audio.

5. Escalabilidad: Hadoop es una plataforma de computación escalable, lo que significa que se pueden agregar más nodos para procesar más datos.

6. Seguridad: Hadoop proporciona mecanismos de seguridad para proteger los datos almacenados en los clústeres. Estos mecanismos incluyen autenticación, autorización y auditoría.

* 1. Beneficios del Big Data

El Big Data, o la gestión y análisis de grandes cantidades de datos, puede ofrecer una amplia gama de beneficios en diferentes áreas. Aquí hay 16 beneficios potenciales del Big Data:

1. Mejora la toma de decisiones: el análisis de grandes cantidades de datos puede ayudar a las empresas a tomar decisiones más informadas y basadas en datos.
2. Reducción de costos: el análisis de Big Data puede ayudar a las empresas a identificar áreas de ineficiencia y reducir costos.
3. Identificación de oportunidades de mercado: el Big Data puede ayudar a las empresas a identificar nuevas oportunidades de mercado al analizar las tendencias y los patrones de compra de los consumidores.
4. Optimización de la cadena de suministro: el Big Data puede ayudar a las empresas a optimizar la cadena de suministro al analizar el rendimiento de los proveedores, la gestión de inventarios y la demanda del mercado.
5. Personalización de la experiencia del cliente: el análisis de Big Data puede ayudar a las empresas a personalizar la experiencia del cliente al analizar los patrones de compra y los comportamientos de los consumidores.
6. Mejora la seguridad: el análisis de Big Data puede ayudar a mejorar la seguridad al identificar patrones de comportamiento sospechosos y detectar posibles amenazas.
7. Identificación de fraudes: el análisis de Big Data puede ayudar a detectar fraudes en transacciones financieras y prevenir posibles pérdidas.
8. Mejora la eficacia del marketing: el Big Data puede ayudar a las empresas a mejorar la eficacia del marketing al identificar los canales de marketing más efectivos y al analizar el comportamiento de los consumidores.
9. Identificación de tendencias de mercado: el Big Data puede ayudar a las empresas a identificar tendencias emergentes en el mercado al analizar grandes cantidades de datos.
10. Mejora la calidad del producto: el Big Data puede ayudar a las empresas a mejorar la calidad del producto al analizar los comentarios de los clientes y la retroalimentación.
11. Aumento de la productividad: el análisis de Big Data puede ayudar a las empresas a aumentar la productividad al identificar áreas de ineficiencia y mejorar los procesos de negocio.
12. Optimización de la publicidad: el Big Data puede ayudar a las empresas a optimizar la publicidad al analizar el comportamiento de los consumidores y la efectividad de las campañas publicitarias.
13. Análisis de riesgos: el Big Data puede ayudar a las empresas a analizar los riesgos asociados con diferentes proyectos y decisiones empresariales.
14. Identificación de problemas de calidad: el Big Data puede ayudar a las empresas a identificar problemas de calidad en los productos al analizar los datos de producción y los comentarios de los clientes.
15. Mejora la eficacia de la investigación y el desarrollo: el Big Data puede ayudar a las empresas a mejorar la eficacia de la investigación y el desarrollo al analizar grandes cantidades de datos y descubrir nuevas oportunidades de innovación.
16. Reducción de errores: el análisis de Big Data puede ayudar a reducir los errores en los procesos de negocio al identificar las áreas de ineficiencia.
17. VISUALIZACIÓN DE DATOS
    1. PowerBi

**¿Qué es?**

Power BI es una herramienta que se utiliza principalmente para crear cuadros de mando que faciliten la toma de decisiones. La información se puede actualizar de manera automatizada o manual y permite la compartición de los informes mediante la propia herramienta.

Esto permite pasar de un sistema con varias herramientas de gestión (un ERP, un CRM, varias tablas de control en Excel) y un controller que cruza información entre ellas habitualmente en Excel para entregar a Dirección informes en Excel o en papel a una situación en la que el controller prepara los informes Power BI que necesita el gerente recopilando la información más importante en un cuadro de mandos que se actualiza automáticamente.

**Principales funciones**

* permitir importar datos y moldearlos a voluntad,
* facilita cruzar datos, hacer cálculos y gráficas sofisticadas de manera relativamente sencilla
* permite que los cálculos se reutilicen tantas veces como sea necesario con la información actual recogida en las bases de datos de origen, por lo que se puede contar con información actualizada y libre de errores humanos.
* Esto se traduce en tablas y gráficas que permiten que el usuario interactúe con la información seleccionando períodos de estudio, tipologías de clientes o de productos o aquellas otras dimensiones que necesite visualizar la persona que consume la información, como puede verse en el video.
* Permite importar datos de diferentes herramientas y fuentes (Google Analytics, Excel, SEM, MailChimp, instituciones oficiales, empresas…) para presentarlos como nos interese y tener una visión global de nuestra organización, detectando posibles puntos a mejorar, oportunidades de negocio, fortalezas, etc.
* Unifica: puede hacerse en la nube o localmente lo que permite almacenar y organizar la información (tanto externa como interna) para lograr una mayor eficiencia. Con ello, a su vez, se pueden cruzar datos, elaborar informes o gráficos útiles para la gestión diaria, presentación de proyectos o reuniones con clientes.
* Segmenta: los datos se pueden segmentar para visualizar lo que nos interese en cada momento.

**Ventajas de Power BI**

Las empresas generan cada vez más datos y tienen la necesidad de ordenarlos y procesarlos con este tipo de soluciones. Power BI ayuda a las organizaciones a interpretar lo que sucedió en el pasado y a predecir lo que podría suceder en el futuro. La visualización de la información de tu empresa, a través de cuadros de mando, te ayudará a mejorar la toma de decisiones. A continuación, procedemos a resumir algunas de las principales ventajas de este software:

* La información se actualiza en tiempo real.
* Permite visualizar datos de forma sencilla sin necesitar conocimientos específicos de programación por lo que la experiencia del usuario es muy buena.
* Es gratuito, con versión en la nube y descargable en escritorio, aunque también hay opción de pago (el Power BI Pro, con mayor capacidad de almacenamiento, más medidas de seguridad y refrescamiento de datos con mayor frecuencia).
* Colaboración: permite el trabajo colaborativo entre diferentes departamentos de una empresa.
* Acceso desde diferentes dispositivos: tanto Android como iOS.
* Buena integración con multitud de plataformas y herramientas de otros fabricantes.
* Polivalente: sus amplias posibilidades la convierten en una herramienta útil para diferentes departamentos de una empresa.

**Desventajas de Power BI**

Power BI es una herramienta poderosa y útil para visualizar y analizar datos, sin embargo, algunas de las desventajas que se pueden mencionar son:

* Costo: Power BI es una herramienta comercial, lo que significa que hay costos asociados a su uso. Aunque hay una versión gratuita, las funciones y características más avanzadas están disponibles en las versiones de pago, lo que puede ser costoso para las pequeñas empresas o los usuarios individuales.
* Curva de aprendizaje: Power BI puede ser difícil de aprender para los usuarios sin experiencia en el análisis de datos y la visualización. Aprender a utilizar la herramienta puede requerir una curva de aprendizaje pronunciada.
* Integración limitada con otras herramientas: Aunque Power BI se integra bien con otras herramientas de Microsoft, la integración con otras herramientas de terceros puede ser limitada, lo que puede ser un inconveniente para las empresas que utilizan múltiples herramientas.
* Limitaciones en la personalización: Aunque Power BI permite cierto grado de personalización, hay algunas limitaciones en cuanto a la personalización de las visualizaciones y los informes. Esto puede ser un problema para los usuarios que desean crear informes y visualizaciones altamente personalizados.
* Requerimientos de hardware: Para trabajar con grandes cantidades de datos, Power BI puede requerir hardware potente, lo que puede ser costoso para los usuarios y las empresas. Además, los usuarios pueden experimentar tiempos de carga prolongados y un rendimiento deficiente si no cuentan con el hardware adecuado.

**¿Cómo funciona?**

Power BI es una herramienta gratuita con actualizaciones mensuales y gran cantidad de información técnica disponible desde su web. Su sencillo uso tampoco exige un aprendizaje previo para comenzar a compartir y administrar consultas. La posibilidad de compartir los datos relevantes con todo el equipo y de consultar los paneles de BI en tiempo real permite solucionar con mayor solvencia cualquier contratiempo que le surja a la empresa.

Power BI es una herramienta con una interfaz al estilo de Microsoft Office, suficientemente intuitiva para usuarios familiarizados con programas de bases de datos y de la Suite de Microsoft, y varios menús.

Permite crear y compartir dashboards e informes y compartir.

En el menú superior podemos seleccionar opciones tales como

* Obtención de datos
* Editar consultas
* Crear objeto visual
* Añadir botones, imágenes, formas
* Administrar relaciones entre tablas
* Crear medidas
* Publicar dashboard

En el menú de la parte lateral se pueden seleccionar:

* Visualizaciones
* Valores
* Filtros
* Campos

Una vez modelados los datos, se pueden generar de forma sencilla las visualizaciones, para las cuales solo habrá que seleccionar las métricas y dimensiones a visualizar, y arrastrar el tipo de datos que se deseen mostrar en el informe.

* 1. DataStudio

Descripción:

Looker es una plataforma de análisis de datos que fue fundada en Santa Cruz, California en 2012 por Lloyd Tabb, un ex ingeniero de Netscape y LiveOps. La idea detrás de Looker era crear una herramienta de análisis de datos que fuera fácil de usar para los usuarios empresariales sin experiencia en programación, se basa en una arquitectura de datos moderna que permite a los usuarios acceder y analizar grandes cantidades de datos en tiempo real. En lugar de utilizar un almacén de datos tradicional, Looker utiliza una técnica conocida como "exploración de datos" que permite a los usuarios navegar y explorar los datos de manera más eficiente.

Modelado de Datos:

Looker Studio permite a los usuarios modelar datos de diferentes fuentes para crear una vista única y coherente de los datos para su análisis. Los modelos de datos se crean utilizando LookML, un lenguaje de modelado específico de Looker que se utiliza para definir los datos que se utilizarán para el análisis.

Creación de Informes:

Una vez que se ha creado el modelo de datos, los usuarios pueden crear informes y paneles personalizados utilizando la interfaz de usuario intuitiva de Looker Studio. Los informes pueden incluir tablas, gráficos, mapas y otros elementos visuales que ayudan a los usuarios a entender los datos y tomar decisiones informadas.

Exploración de Datos:

Los usuarios pueden explorar los datos en tiempo real utilizando la función de exploración de datos de Looker Studio. Esta función permite a los usuarios navegar por los datos utilizando diferentes filtros y dimensiones, y ver los resultados en tiempo real.

Seguridad y Colaboración:

Looker Studio ofrece opciones de seguridad y colaboración para garantizar que los datos sean seguros y estén protegidos. Los usuarios pueden establecer permisos de acceso y restricciones de nivel de usuario, lo que les permite compartir información de manera segura con otras personas en su organización.

Integraciones:

Looker Studio se integra con una variedad de herramientas de terceros, como Google Analytics, Salesforce y Zendesk, lo que permite a los usuarios agregar datos de diferentes fuentes y realizar un análisis más completo.

Looker Studio es una plataforma de análisis de datos que ofrece una amplia variedad de gráficos y visualizaciones para ayudar a los usuarios a comprender mejor sus datos. A continuación, se describen algunos de los tipos de gráficos y visualizaciones que se pueden utilizar en Looker Studio:

Gráfico de líneas:

Este tipo de gráfico se utiliza para mostrar cómo cambian los valores de una o más métricas a lo largo del tiempo. Las líneas se trazan en un eje X (tiempo) y un eje Y (valor) para mostrar las tendencias y patrones a lo largo del tiempo.

Gráfico de barras:

Este tipo de gráfico se utiliza para comparar valores entre diferentes categorías. Las barras se trazan en un eje X (categoría) y un eje Y (valor) para mostrar las diferencias entre las categorías.

Gráfico de dispersión:

Este tipo de gráfico se utiliza para mostrar la relación entre dos variables numéricas. Los valores se trazan en un eje X (variable 1) y un eje Y (variable 2) para mostrar si hay una relación entre las variables.

Gráfico circular:

Este tipo de gráfico se utiliza para mostrar cómo se dividen los valores entre diferentes categorías. Los valores se trazan en un círculo y se dividen en secciones para mostrar la proporción de cada categoría.

Mapa:

Esta visualización se utiliza para mostrar datos geográficos. Los datos se trazan en un mapa y se pueden utilizar diferentes colores y tamaños de puntos para mostrar diferentes valores.

Tabla:

Esta visualización se utiliza para mostrar datos en una tabla. Los datos se presentan en filas y columnas, y se pueden utilizar diferentes opciones de formato y cálculo para mostrar diferentes resúmenes de los datos.

Gráfico de área:

Este tipo de gráfico se utiliza para mostrar cómo cambia una o más métricas a lo largo del tiempo, y también para mostrar cómo se acumulan los valores. Las áreas se trazan en un eje X (tiempo) y un eje Y (valor) para mostrar las tendencias y patrones a lo largo del tiempo.

Locker Studio puede conectarse a una amplia variedad de bases de datos, incluyendo:

Bases de datos relacionales: como MySQL, PostgreSQL, SQL Server, Oracle, SQLite, etc.

Bases de datos NoSQL: como MongoDB, Cassandra, Redis, Couchbase, etc.

Bases de datos en la nube: como Amazon RDS, Google Cloud SQL, Azure SQL Database, etc.

Bases de datos de Big Data: como Hadoop, Hive, Impala, etc.

Locker Studio también puede conectarse a bases de datos locales o remotas, dependiendo de la configuración del entorno en el que se esté utilizando.

1. CONCLUSIONES Y RETOS PARA EL BIG DATA

* La transformación digital está impulsando la tecnología en todas las direcciones, a medida que profundizamos en el internet de las cosas, el aprendizaje automático y el Big Data. En 2020 existían 9700 millones de dispositivos conectados a internet; en 2030 se espera que esa cifra alcance los 29.400 millones. La inmensa cantidad de datos generados por ellos hará que el Big Data tenga un papel absolutamente esencial. Dentro de los mas relevante para el uso de Big Data tenemos:
  + Business Intelligence
  + Tecnología de nube
  + Los datos como servicio (DaaS)
  + Tecnología y bienestar (sector salud)
  + Vehículos sin conductor
  + Big data para ayudar en la investigación del cambio climático
  + El análisis en tiempo real gana tracción tras la pandemia
  + Más procesamiento del lenguaje natural (PLN)
  + Automatización del análisis del big data[[3]](#footnote-3)

En Colombia, el Big Data se está convirtiendo en una tendencia creciente entre las empresas de todos los sectores. Esto se debe a la creciente demanda de datos y al aumento de los costos de almacenamiento. El Big Data se está utilizando para mejorar la toma de decisiones estratégicas, para identificar patrones y tendencias y para mejorar la eficiencia operativa.

Retos del Big Data en Colombia.

Estandarización de datos: Uno de los principales desafíos en el Big Data en Colombia es la estandarización de los datos. Esto se debe a que los datos provienen de diferentes fuentes y hay una gran variedad de formatos de datos.

1. https://www.atiga.es/web/wp-content/uploads/2017/03/Estado-del-Arte-Big-Data-CC-DA.pdf [↑](#footnote-ref-1)
2. https://blogs.imf-formacion.com/blog/tecnologia/las-10-mejores-herramientas-de-analisis-de-datos-202006/ [↑](#footnote-ref-2)
3. https://www.cyberclick.es/numerical-blog/tendencias-en-data-y-analytics-para-2023 [↑](#footnote-ref-3)