CECO

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE

INSTITUCIÓN DE EDUCACIÓN TÉCNICA

www.ceco.edu.co

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

WORKSHOP EN CIENCIA DE DATOS LOW CODE (KNIME)

FACILITADOR: ING. MARCIAL REYES
MGTR. BIGDATA Y ANALYTICS

SOLEDAD, 18 DE NOVIEMBRE DE 2023

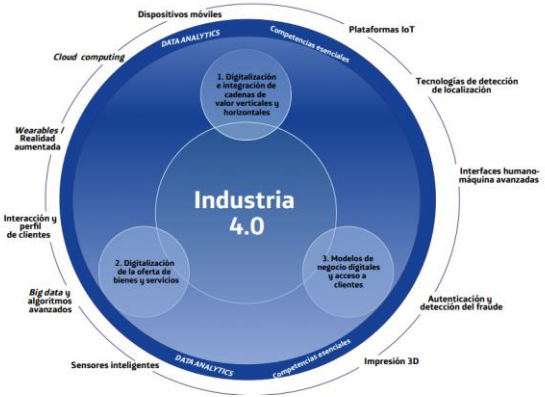
SERVICIOS INTEGRALES PARA EL DESARROLLO ONG
SOLUCIONES DE INGENIERÍA Lagit SAS
SISE Digital

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

LA CUARTA REVOLUCIÓN INDUSTRIAL

La Cuarta Revolución Industrial es el resultado del dinamismo de las **tecnologías** y de la combinación de **sistemas digitales** y físicos para mejorar la calidad de vida del ser humano. En tal sentido, la filosofía en el ámbito laboral y en los procesos del **mundo 4.0** debe transformarse; la enseñanza debe superar lo convencional y avanzar hacia el **uso de herramientas digitales**, la **inteligencia artificial** y el **análisis de datos** para formar personas altamente competitivas que respondan a las exigencias de este mundo 4.0.

Tecnologías disruptivas de la cuarta revolución industrial: Inteligencia artificial, Realidad aumentada, Blockchain,Big data y analítica de datos, Machine learning, Robótica.



!LOS DATOS COMO EL NUEVO PETROLEO¡

En 2006 Clive Humby, uno de los primeros científicos de datos de la historia dijo, “los datos son el nuevo petróleo” que luego Michael Palmer amplió diciendo, “los datos son valiosos, pero si no están refinados, en realidad no se pueden usar. El Petróleo debe transformarse en gas, plástico, productos químicos, etc. para crear una entidad valiosa que impulse una actividad rentable; por lo tanto, los datos deben desglosarse y analizarse para que tengan valor”.

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

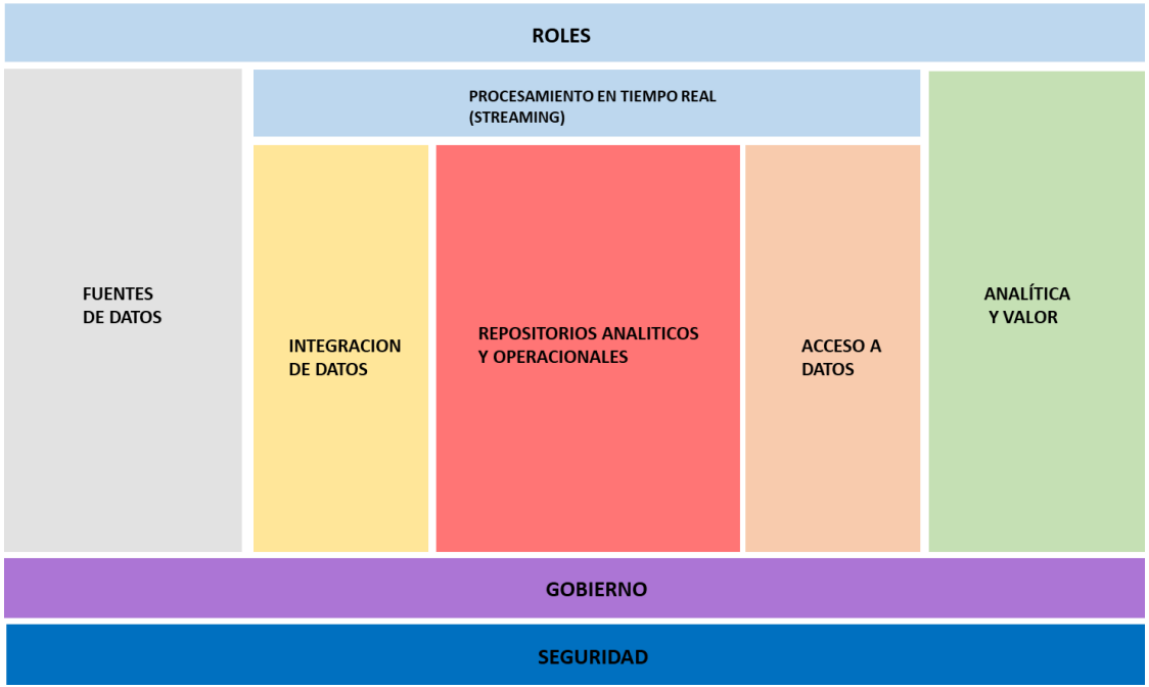
¿QUE ES BIGDATA?

La tasa de crecimiento de los datos generados y almacenados ha aumentado exponencialmente. En un artículo de 1965, [10] Gordon Moore estimó que la densidad de los transistores en una placa de circuito integrado se duplicaba cada dos años.
Conocida como Ley de Moore, esta tasa de crecimiento se ha aplicado a todos los aspectos de la informática, desde las velocidades del reloj hasta memoria. Se estima que las tasas de crecimiento de los volúmenes de datos son más rápidas que la Ley de Moore, con volúmenes de datos más más del doble cada dieciocho meses. Esta explosión de datos está creando oportunidades para nuevas formas de combinar y el uso de datos para encontrar valor, además de presentar desafíos importantes debido al tamaño de los datos que se gestionan.
Un cambio significativo se produce en la cantidad de datos no estructurados. Históricamente, los datos estructurados suelen ha sido el foco de la mayoría de los análisis empresariales y se ha manejado mediante el uso del modelo de datos relacionales.
Recientemente, la cantidad de datos no estructurados (como microtextos, páginas web, datos de relaciones, imágenes y vídeos) se ha disparado, lo que ha provocado el deseo de incorporar estos datos no estructurados para generar valor adicional. El beneficio central del análisis de Big Data es la capacidad de procesar grandes cantidades y diversos tipos de información.

Big Data consiste en conjuntos de datos extensos principalmente con las características de: volumen, velocidad, variedad y/o variabilidad, que requieren una arquitectura escalable para un almacenamiento, manipulación y análisis eficientes.

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

ARQUITECTURA DE BIGDATA



¿QUE ES LA CIENCIA DE DATOS?

La ciencia de datos combina matemáticas y estadística, programación especializada, análisis avanzados, inteligencia artificial (IA) y machine learning con experiencia en distintas materias para descubrir información práctica oculta en los datos de una organización. Esta información se puede utilizar como guía para la toma de decisiones y la planificación estratégica.

El aumento del volumen de orígenes de datos y, por lo tanto, de datos, ha convertido a la ciencia de datos en uno de los campos de más rápido crecimiento de todos los sectores. Como resultado, no es de extrañar que el rol de científico de datos haya sido calificado como la “profesión más sexi del siglo XXI” por Harvard Business Review (enlace externo a IBM).

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

NECESIDAD DE PROFESIONALES EN DATOS

Un reciente estudio en los Estados Unidos del McKinsey Global Institute, plantea que un reto urgente e importante para el país lo constituye la escasez del talento analítico y de gestión necesario para sacar el máximo provecho al Big Data. En dicho reporte se expone que, tan solo en EEUU, 5 millones de puestos de trabajo requerirán de habilidades en el análisis de datos en el 2018. Además se necesitarán 1,5 millones de directivos y analistas con profundas habilidades analíticas y técnicas que puedan realizar las preguntas correctas a los datos y utilizar los resultados que brinde dicho análisis efectivamente (Loukides.M, 2011).

Fuente. https://www.proquest.com/openview/c35c739c6bd258013b8575fab99c9a5c/1?pq-origsite=gscholar&cbl=2043236

Según el área de trabajo desempeñada, una buena muestra de directivos de grandes empresas españolas apunta a que las profesiones digitales que más aumentarán en el futuro serán las asociadas al big data y la ciberseguridad.
De hecho, otros estudios comienzan a constatar cierta carencia de profesionales cualificados que satisfagan las necesidades de las empresas, lo que está provocando un aumento del número de vacantes de empleo, especialmente en el caso de los desarrolladores de aplicaciones móviles, de soluciones de big data o los especialistas en ciberseguridad. Una carencia de profesionales que sorprende, al coincidir con la abundancia de personas desempleadas de nuestro país (España tiene la segunda tasa de paro más alta de la UE-28), y que exige el reforzamiento del sistema educativo y de las políticas activas de empleo, sobre la base de las habilidades requeridas por estos nuevos puestos de trabajo

Fuente. https://repositorio.upn.edu.pe/bitstream/handle/11537/31835/Libro-Economia-de-los-Datos-Ontiveros.pdf?sequence=1&isAllowed=y

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

ROLES GENERADOS EN BIG DATA

Chief Data Officer (CDO): Miembro de la dirección ejecutiva encargado de liderar la gestión de datos y analítica asociada con el negocio, es responsable de los diferentes equipos especializados en datos en la empresa.

Data Scientist (científico de los datos): Interpreta y transforma los grandes volúmenes de datos en información útil para la empresa. Se caracteriza por tener habilidades en matemáticas, estadística, programación, es creativo y cuenta con habilidades comunicativas para explicar los resultados de su trabajo.

Data Steward (administrador de datos): Se encarga de mantener la calidad, disponibilidad y seguridad de los datos. Posee conocimientos en los procesos del negocio e identifica cómo son usados dentro de la empresa.

Business Data Analyst (analista de datos): Participa en el análisis de los datos con el fin de recolectar las necesidades del cliente para sustentarlas de manera clara al científico de datos.

Data Artist: Es un experto en Business Analytics, encargado de mostrar de manera sencilla (gráficos, infografías y herramientas visuales), los resultados del análisis de los datos para comprender grandes volúmenes de información.

Estadístico: Sus funciones se centran en obtener, analizar e interpretar datos cualitativos y cuantitativos usando los métodos estadísticos.

ROLES GENERADOS EN BIG DATA

Data Engineer: Es el responsable de proporcionar los datos requeridos por el científico de datos. Tiene gran conocimiento en bases de datos, arquitecturas de clúster, lenguajes de programación y sistemas de procesamiento de datos.

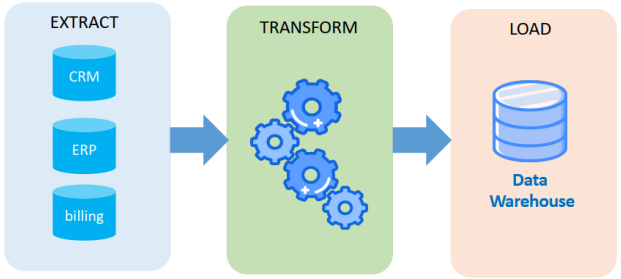
Administrador de bases de datos: Tiene conocimientos fuertes en el manejo de bases de datos, típicamente relacionales (CAOBA et. al. , 2017)

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

PROCESOS ETL: EXTRACT, TRANSFORM AND LOAD

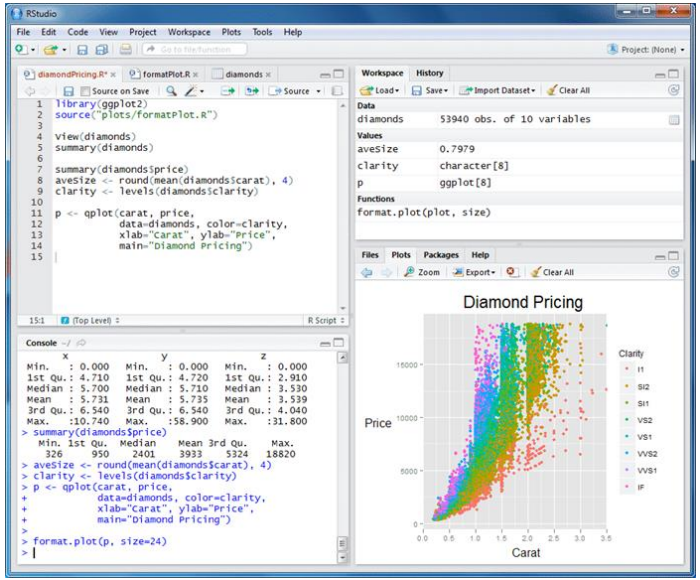
El proceso de ETL es uno de los conceptos más importante en el BIGDATA, ya que se obtiene la información de varias fuentes, se procesa y se carga en un repositorio de datos (Base de datos, archivo, etc)

- El proceso ETL consta de tres fases: extracción, transformación y carga. De cada uno de estos procesos es necesario conocer el funcionamiento y sus claves.
- El proceso ETL es una parte importante de la integración de datos, y da cohesión a aplicaciones y sistemas que consumen datos.
- Es un proceso continuo, con un flujo de trabajo bien definido. Durante este proceso, los datos se extraen inicialmente de una o más fuentes. Luego, los datos se limpian, se enriquecen, se transforman y finalmente, se almacenan en Data Warehouse.



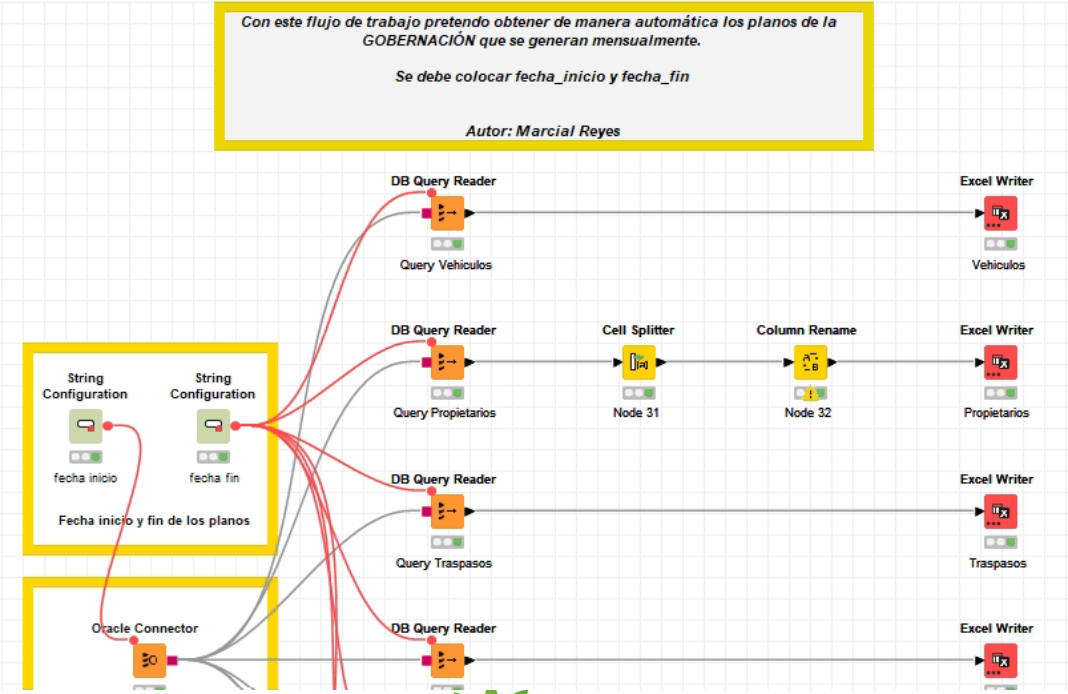
CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

ALGUNOS UTILIZADOS PARA CIENCIA DE DATOS



CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

KNIME ANALITYCS PLATFORM LOW CODE



CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

¿QUE ES KNIME?

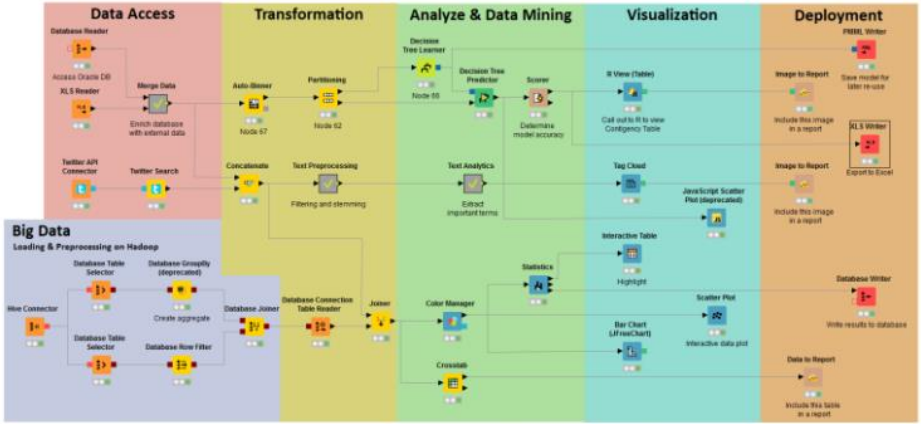
KNIME (Konstanz information mINer) constituye una de las técnicas de minería de datos con un entorno de desarrollo totalmente gratuito, su origen se remonta a un proyecto desarrollado originalmente en el departamento de bioinformática y minería de datos de la Universidad de Constanza, Alemania, bajo la supervisión del profesor Michael Berthold. Actualmente, sigue creciendo en desarrollo, así como prestando servicios de formación y consultoría como parte de la Empresa GmbH, radicada en Zúrich, Suiza.

Dentro de las principales características de esta herramienta se cuenta que está desarrollado sobre la plataforma Eclipse y programado esencialmente en java, por lo que puede ser ejecutado en diferentes Sistemas Operativos. Fue creada como herramienta gráfica pudiéndose diseñar un workflow, para representar un proceso de minería de datos, incluyendo nodos, que encapsulan distintos algoritmos.

Otro de las características a destacar en KNIME, de acuerdo a lo expresado por Cubero y Berzal (2011) es que al utilizarse se sigue el diseño de un workflow que modela las distintas etapas de un proyecto de minería de datos. En este sentido, se proporciona distintos nodos agrupados en fichas, como, por ejemplo: a) Entrada de datos [IO > Read]. b) Salida de datos [IO > Write]. c) Preprocesamiento [Data Manipulation], para filtrar, discretizar, normalizar, filtrar, seleccionar variables… d) Minería de datos [Mining], para construir modelos (reglas de asociación, clustering, clasificación, MDS, PCA…). e) Salida de resultados [Data Views] para mostrar resultados en pantalla (ya sea de forma textual o gráfica).

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

¿QUE ES KNIME?



Data Access
MySQL, Oracle, ...
SAS, SPSS, ...
Excel, Flat, ...
Hive, Impala, ...
XML, JSON, PMML
Text, Doc, Image, ...
Web Crawlers
Industry Specific
Community / 3rd

Transformation
Row,
Column
Matrix
Text, Image
Time Series
Java
Python
Community / 3rd

Analysis & Mining
Statistics
Data Mining
Machine Learning
Web Analytics
Text Mining
Network Analysis
Social Media Analysis
R, Weka, Python
Community / 3rd

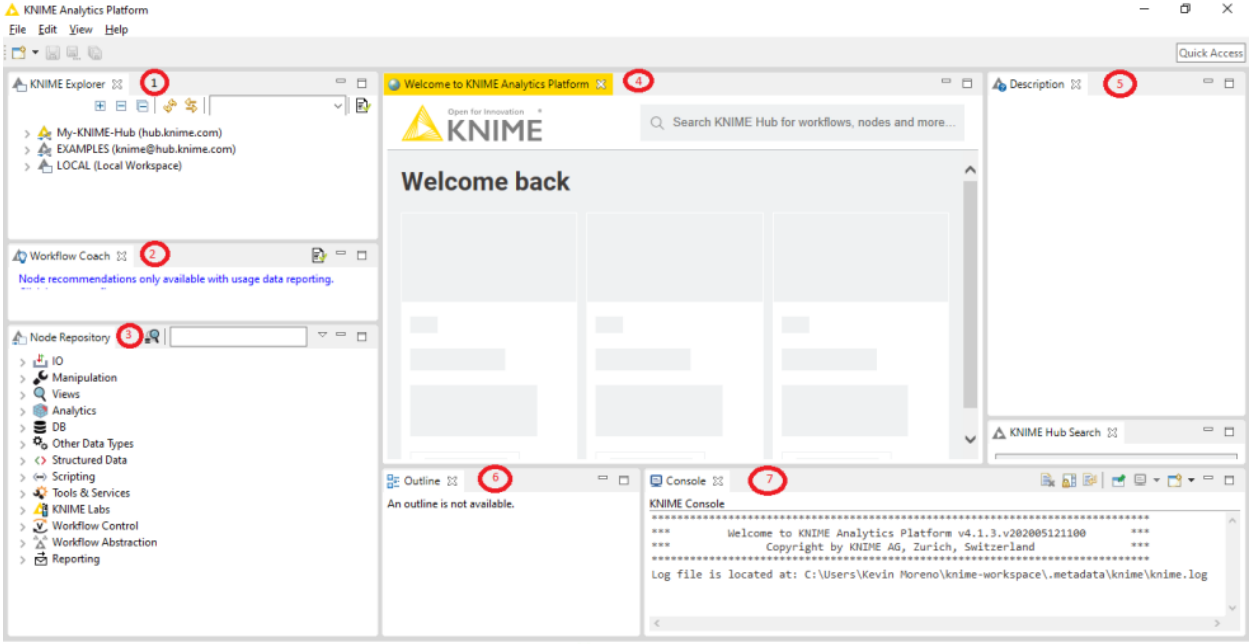
Visualization
R, Python
JFreeChart
JavaScript
Community / 3rd

Deployment
via BIRT
PMML
XML, JSON
Databases
Excel, Flat, etc.
Text, Doc, Image
Industry Specific
Community / 3rd

SERVICIOS INTEGRALES PARA EL DESARROLLO ONG
SOLUCIONES DE INGENIERÍA Lagit SAS
SISE Digital

CENTRO DE EDUCACIÓN CONTINUA DE ORIENTE
INSTITUCIÓN DE EDUCACIÓN TÉCNICA

INTERFAZ KNIME



SERVICIOS INTEGRALES PARA EL DESARROLLO ONG
SOLUCIONES DE INGENIERÍA Lagit SAS
SISE Digital