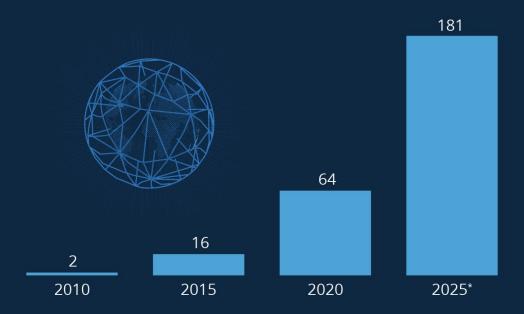


Evolución del Data Center Tradicional al Cloud Computing

El Big Bang del Big Data

Volumen estimado de datos digitales creados o replicados en todo el mundo, en zetabytes



Un zetabyte equivale a mil millones de gigabytes.

* Previsión de marzo de 2021.

Fuentes: IDC, Seagate, Statista









La creación de datos, a punto de explotar

Cantidad real y prevista de datos generados en todo el mundo (en zettabytes)

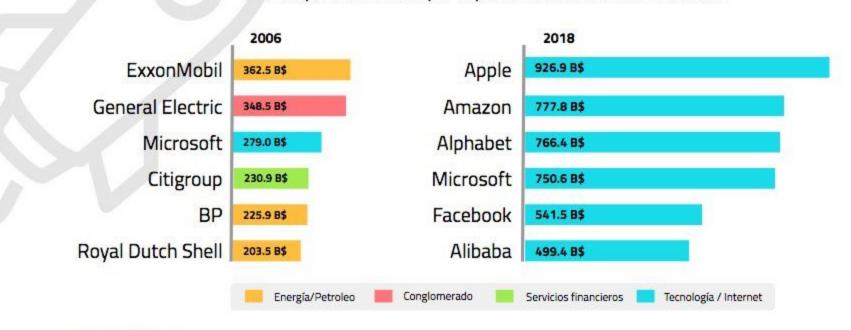






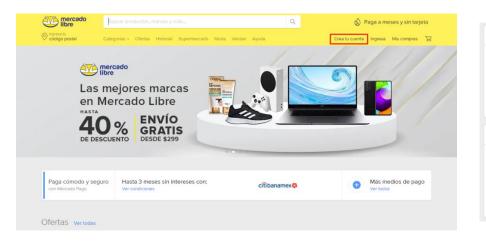
Los datos son el nuevo petróleo

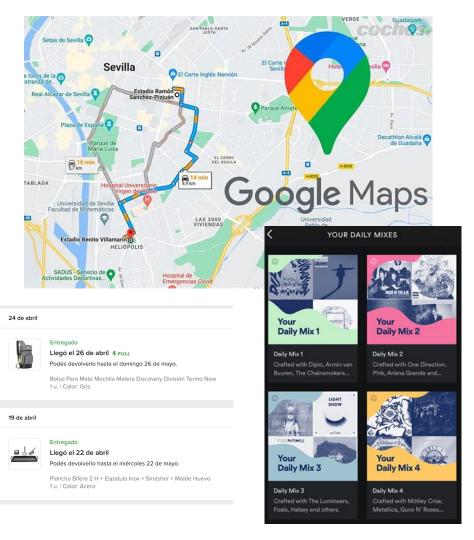
Las empresas con mayor capitalización bursátil del mundo



Fuente: Statista









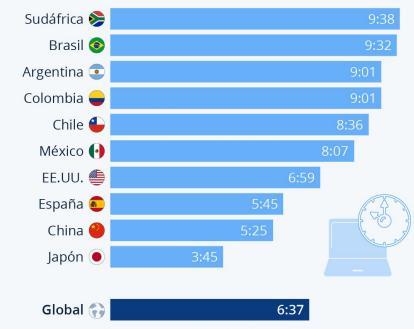






¿Cuántas horas al día pasamos en internet?

Tiempo promedio de uso diario de internet en países seleccionados (en horas y minutos)



Encuesta a usuarios de internet de 16 a 64 años. Datos del tercer trimestre de 2022. Fuente: GWI vía DataReportal

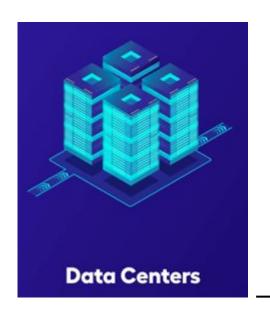








Objetivos







Lucas Trubiano

Consultor y Profesor en Data Engineering

Co-Founder & CDAO de 3ng.tech

Ing. en Mecatrónica UNCUYO y ex DJ

#Python #AWS #BigData















Innovación que transforma, conecta y simplifica

Eleva tu empresa: convertimos datos en decisiones estratégicas para potenciar tu negocio.

Potenciar mi negocio

Nuestros Servicios

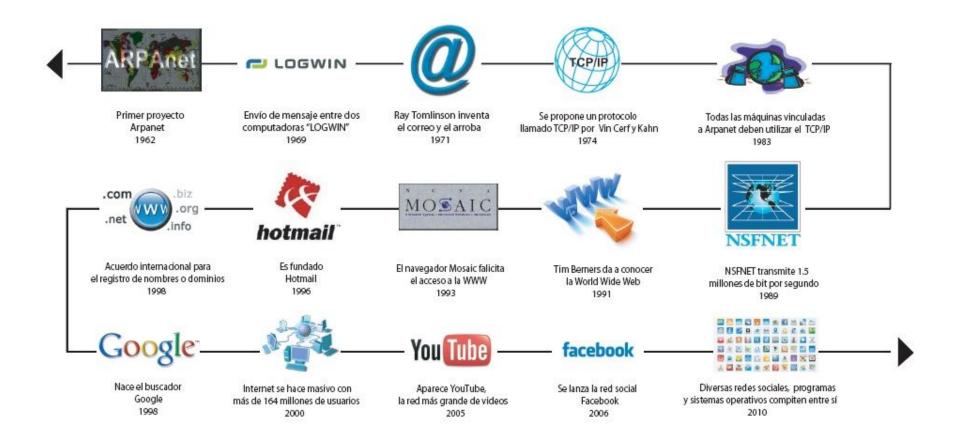
Centrado en los datos

Optimice su empresa:

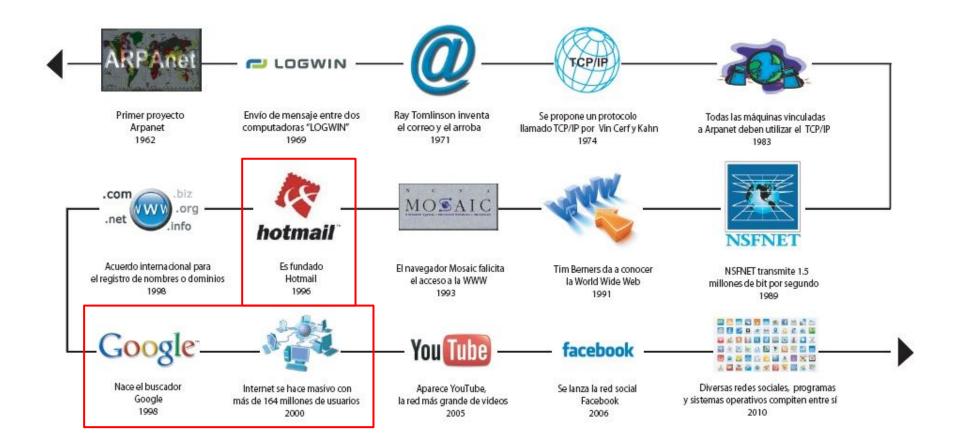


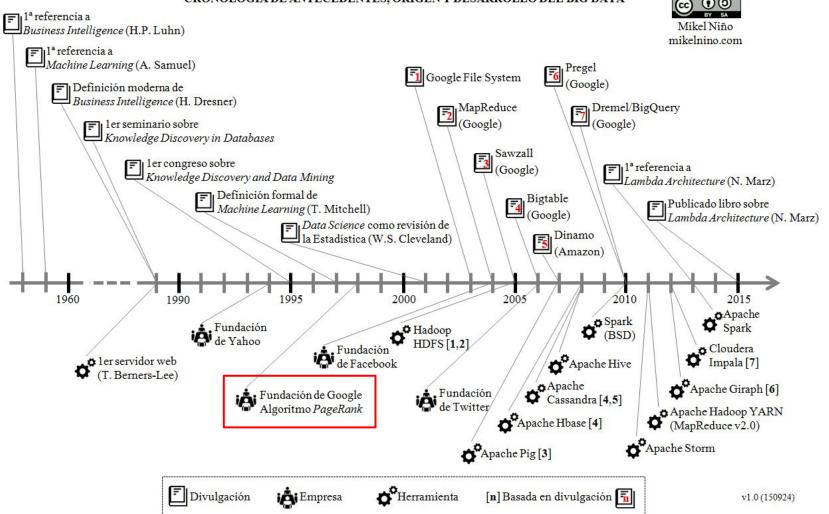


Grandes hitos del Internet

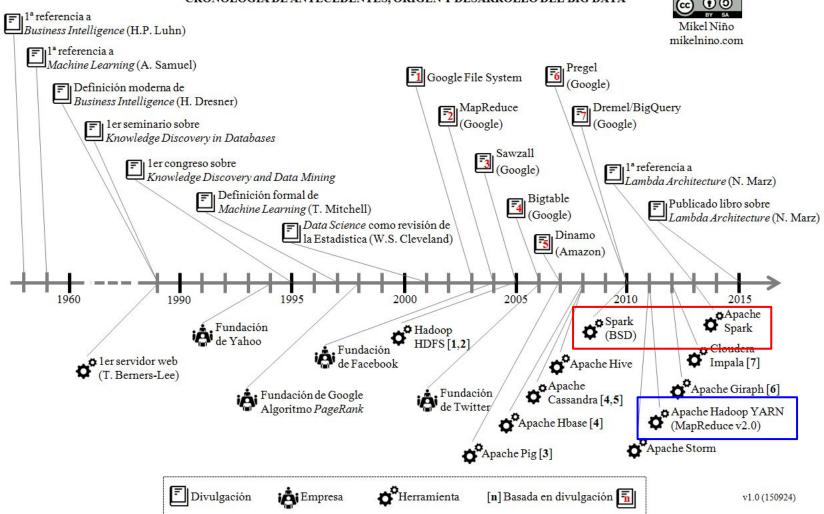


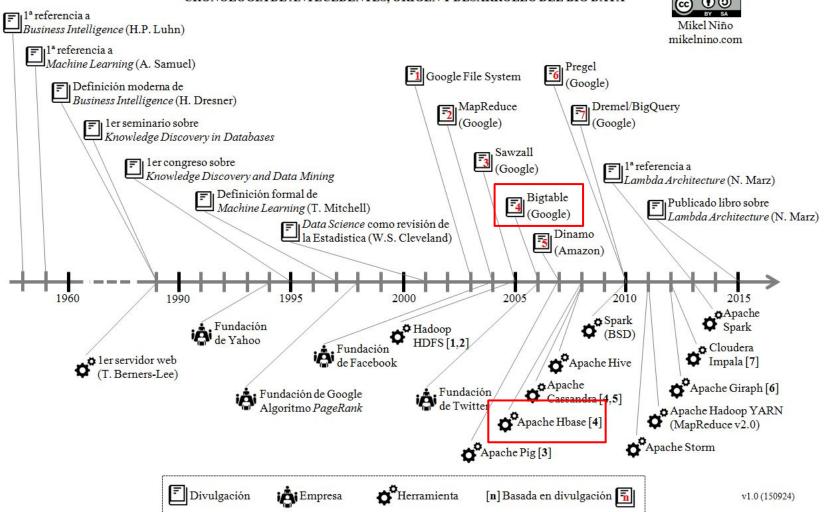
Grandes hitos del Internet

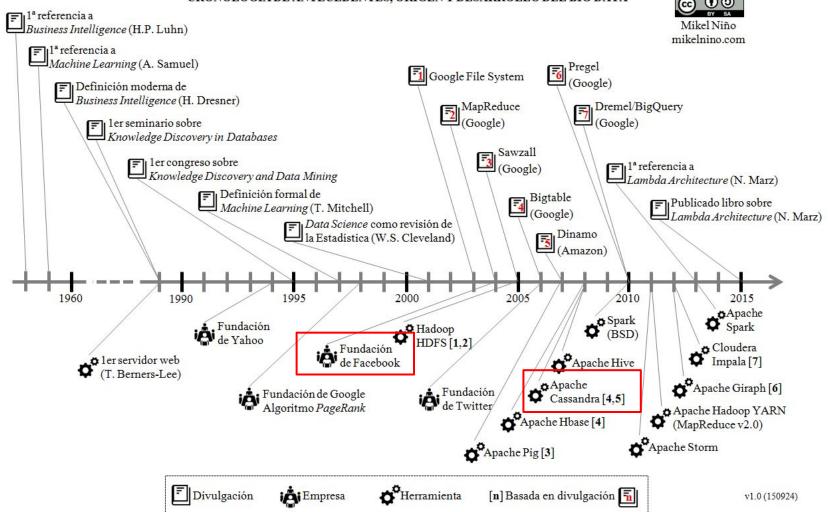


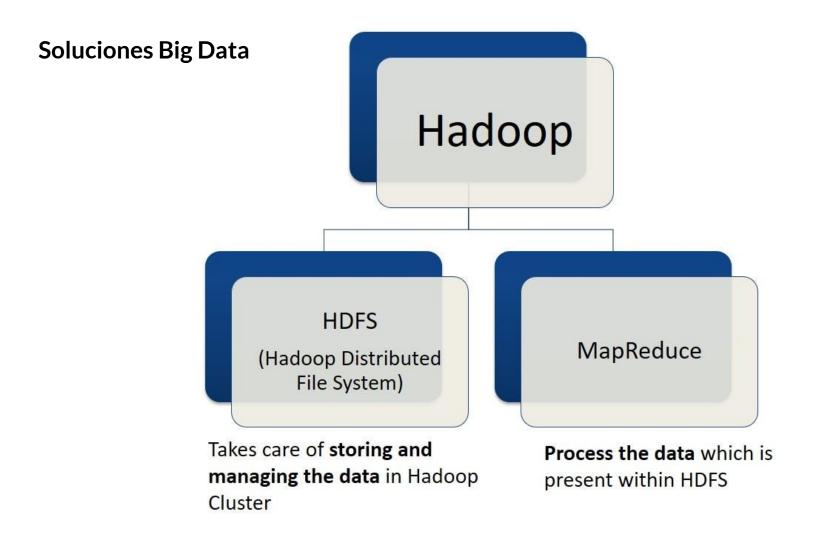


CRONOLOGÍA DE ANTECEDENTES, ORIGEN Y DESARROLLO DEL BIG DATA 1ª referencia a Mikel Niño Business Intelligence (H.P. Luhn) mikelnino.com F11ª referencia a Pregel (Google) Google File System Definición moderna de Business Intelligence (H. Dresner) MapReduce (Google) Dremel/BigQuery (Google) 1er seminario sobre Knowledge Discovery in Databases Sawzall (Google) ler congreso sobre 1ª referencia a Knowledge Discovery and Data Mining Lambda Architecture (N. Marz) Definición formal de Bigtable (Google) Publicado libro sobre Lambda Architecture (N. Marz) Machine Learning (T. Mitchell) Data Science como revisión de Dinamo (Amazon) la Estadística (W.S. Cleveland) 1960 1995 2005 2010 2015 1990 2000 Apache Spark (BSD) databricks Fundación Hadoop HDFS [1,2] Spark Fundación de Yahoo Cloudera Impala [7] ter servidor web Apache Hive de Facebook (T. Berners-Lee) Apache Apache Giraph [6] Fundación de Twitter Fundación de Google Algoritmo PageRank Cassandra [4,5] Apache Hadoop YARN (MapReduce v2.0) Apache Hbase [4] Apache Storm Apache Pig [3] Herramienta **Empresa** Divulgación [n] Basada en divulgación 🔚 v1.0 (150924)

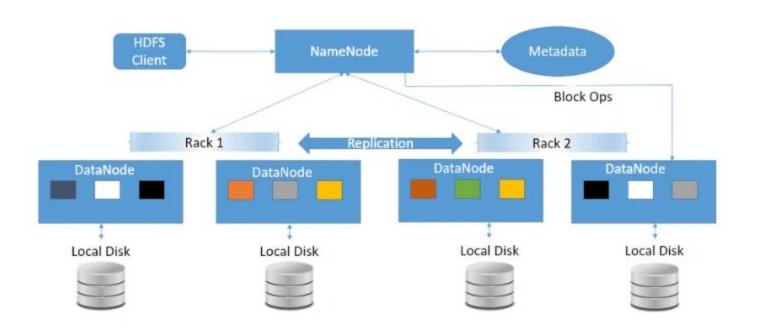




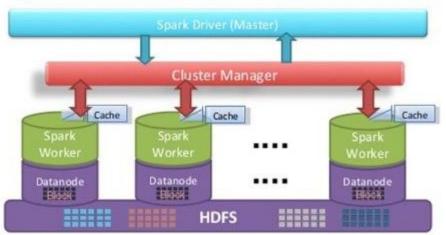


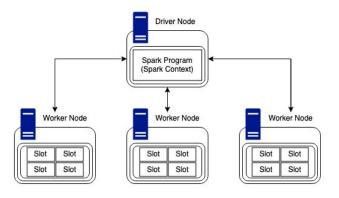


HDFS Architecture

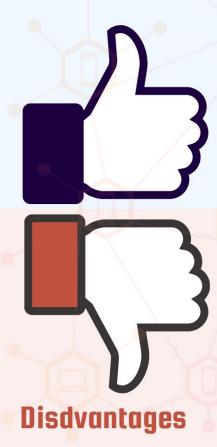




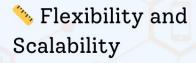




Advantages



Improved Security



2 Enhanced Performance

▶ Low Latency

Increased Reliability

Expanded Storage

\$ Cost effectiveness

Distributed Computing

§ Maintenance Fees

Component Failure

Bandwidth Restrictions

* Systems Complexity

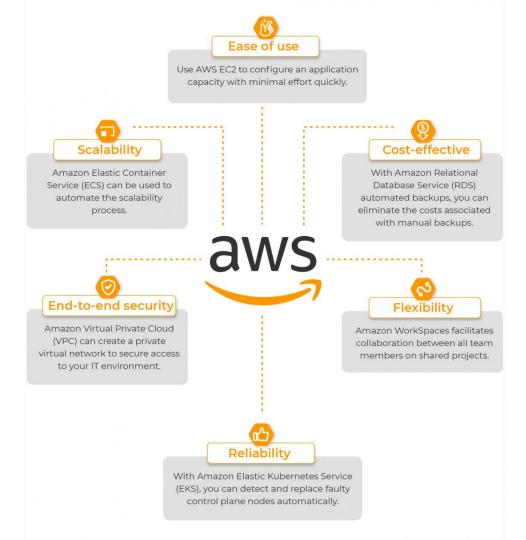
" Development Obstacles

Slow Network ? Issues with Transfers

Standardization



Ventajas



Pros and Cons of Cloud Computing

Lower the Infrastructure Price

Expand IT Resources

Environment Friendly

PROS

Reliability

Data Control

Data Backup & Recovery

Operating from Cloud to On-Premises

Cloud specialized skills

Downtime

Restricted Control

Rigid Contracts

Limited Control of Infrastructure





Desafíos



Data Security and Compliance Risks



Cost Management and Budgeting



Wrong Cloud Migration Strategy



No Internal Capacity to Perform the Cloud Migration



Migration Complexities



Application Compatibility



Data Transfer and Bandwidth Limitations



Cultural Resistance to Change



Legacy Systems Migration

Desafíos

Protocolo de red personalizado

AWS DataSync emplea un protocolo de transferencia diseñado por AWS, desacoplado del protocolo de almacenamiento, para acelerar el movimiento de datos. El protocolo lleva a cabo optimizaciones sobre cómo, dónde y qué datos se envían en la red. Las optimizaciones de red realizadas por DataSync incluyen transferencias incrementales, compresión en línea y detección de archivos dispersos, como así también la validación y el cifrado de datos en línea.

Las conexiones entre el agente local de DataSync y los componentes de servicio en la nube están en archivos de multiprocesos que potencian el rendimiento sobre una red de área amplia (WAN). Una sola tarea de DataSync es capaz de utilizar completamente 10 Gbps a través de un enlace de red entre su entorno en las instalaciones y AWS.

Varios años

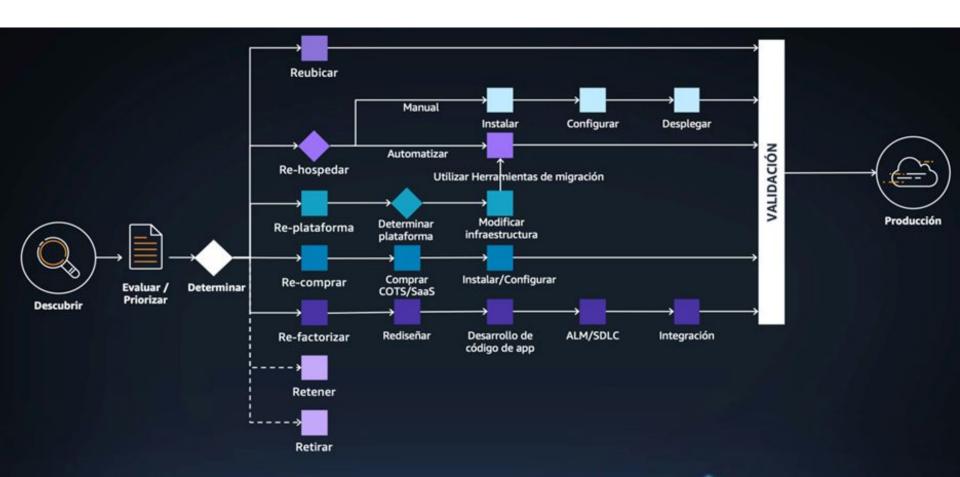
Desafíos

What is AWS Snowmobile?

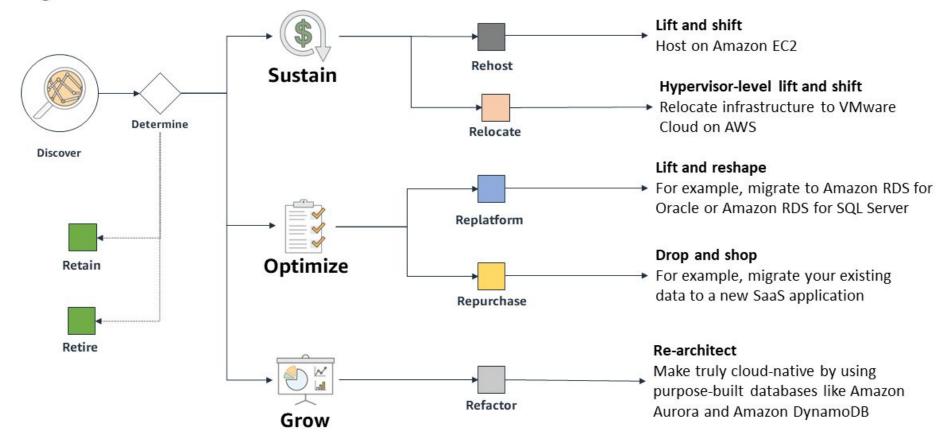
AWS Snowmobile is a 45-foot rugged shipping container that's typically hauled by a semi-trailer truck. It is used to transfer huge volumes of data (up to 100PB) to AWS. The AWS Snowmobile comes in handy when moving entire data centers, video libraries, image repositories, or other types of huge data that needs to be migrated to AWS. Before the introduction of AWS Snowmobile, clients would need years to migrate their data which was painfully slow and expensive.



Estrategias de Migración

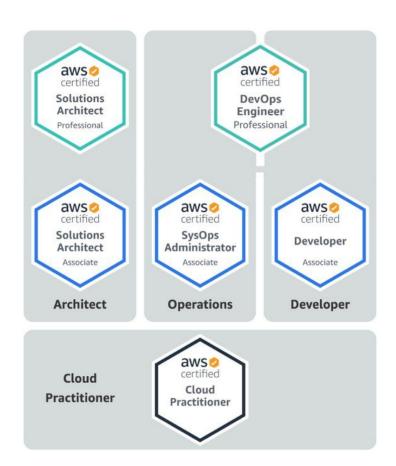


Migration Paths



¿Cómo hacer \$\$?

¡APRENDIENDO A RESOLVER LOS DESAFÍOS!







Estrategia y Gobierno de Datos



Arquitectura e Ingeniería de Datos



Visualización y Analytics

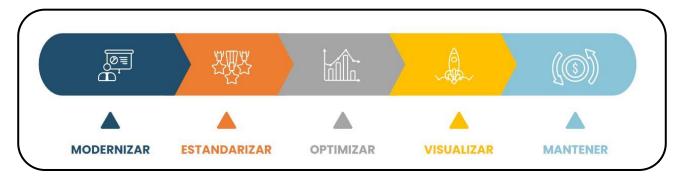


Ciencia, ML e IA



Productos de Datos

Cloud Services





iMuchas Gracias!

Preguntas | Dudas

Seguinos y contactanos:

