

MACHINE LEARNING EN LA NUBE PARA DATOS DESESTRUCTURADOS

HI THERE!



Sergio Gordillo

Cloud Architect en **Keepler Data Tech**

sergio.gordillo@keepler.io / @Sergio_Gordillo

The information contained in this document is property of KEEPLER DATA TECH and intended only for the person or entity to which it is sent. It may contain confidential and / or privileged material, the use of this information or any disclosure, copying or distribution is prohibited and may be unlawful. If you received this in error, please contact the sender and delete all copies.

WHAT?

Cloud Computing

Big Data

Artificial Intelligence

HOW?

TRANSFORMACIÓN AGILE

METODOLOGÍA / FRAMEWORK DEVOPS

ENFOQUE DATA PRODUCT

EXCELENCIA TÉCNICA

WHERE?

AGENDA

- | | |
|---|--|
| 1 | NUBE PÚBLICA Y ML EN LA NUBE |
| 2 | DEMO DE ML EN LA NUBE |
| 3 | DATOS DESESTRUCTURADOS Y EXTRACCIÓN DE VALOR |
| 4 | DEMOS EN LA NUBE CON ML Y DATOS DESESTRUCTURADOS |

PREGUNTAS

NUBE PÚBLICA Y MODELOS DE SERVICIO



SaaS

ANALYTICS AND BIG
DATA SERVICES

MACHINE LEARNING
SERVICES

DEVELOPER TOOLS

STORAGE AND DATABASES

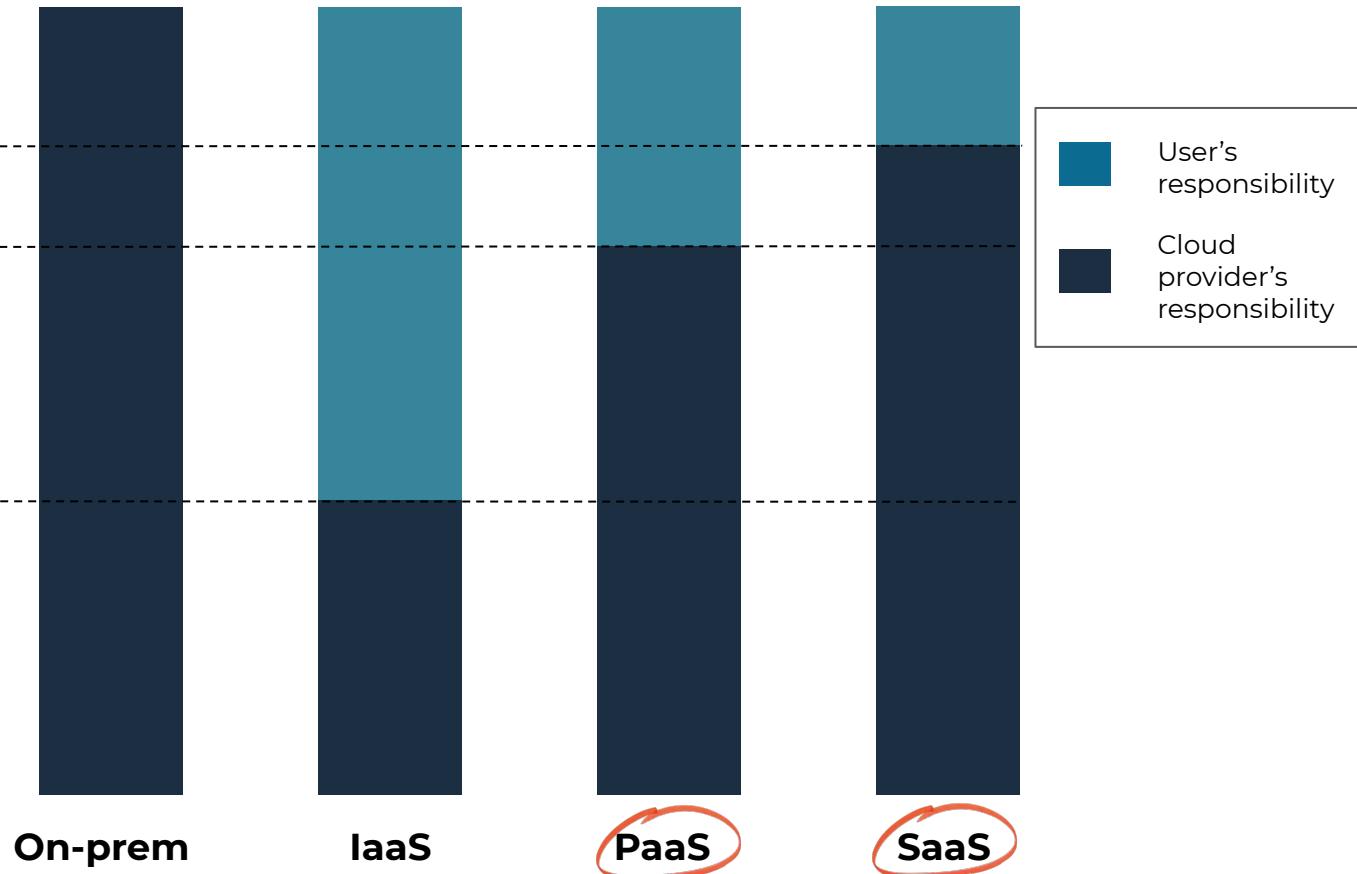
COMPUTE

NETWORKING

MODELOS DE SERVICIO Y RESPONSABILIDAD COMPARTIDA

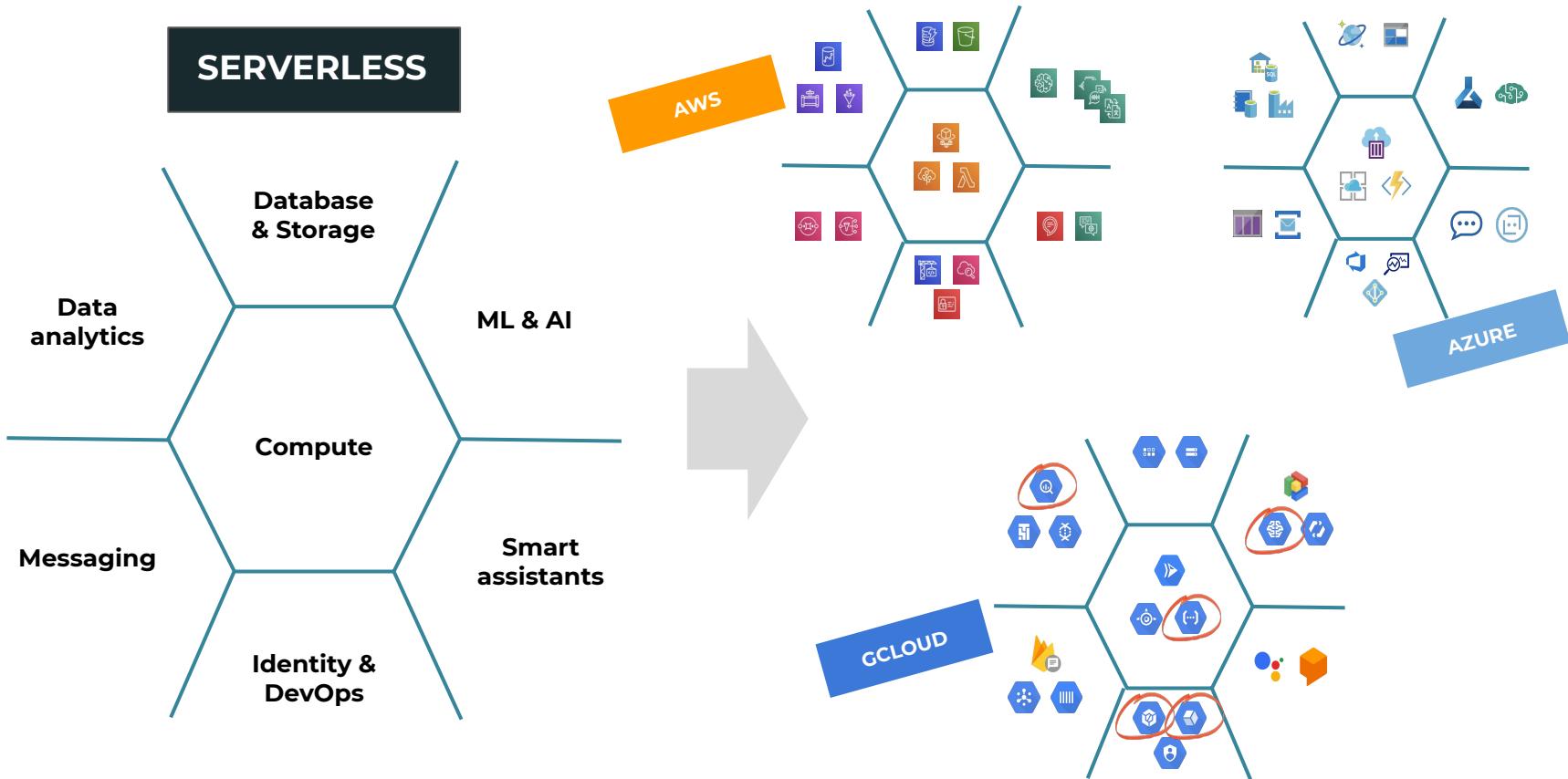
keeper

- Content
- Access policies
- Usage
- Deployment
- Web application security
- Identity
- Operations
- Access and authentication
- Network security
- Guest OS, data & content
- Audit logging
- Network
- Storage + encryption
- Hardened Kernel + IPC
- Boot
- Hardware

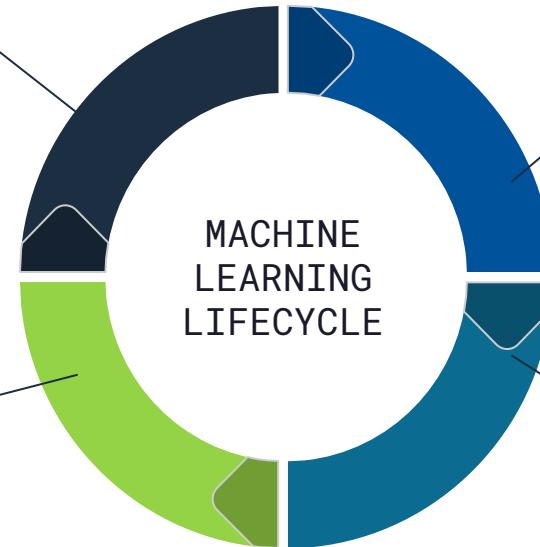


SERVERLESS Y PROVEEDORES CLOUD

keeper



MACHINE LEARNING EN LA NUBE



La fase de ingeniería del dato mediante la colecta, análisis, extracción y transformación del dato es la parte donde se suele dedicar entre el 70% y 80% del desarrollo y es que alimentar un modelo con los datos inapropiados conduce a modelos erráticos y con alto índice de ruido

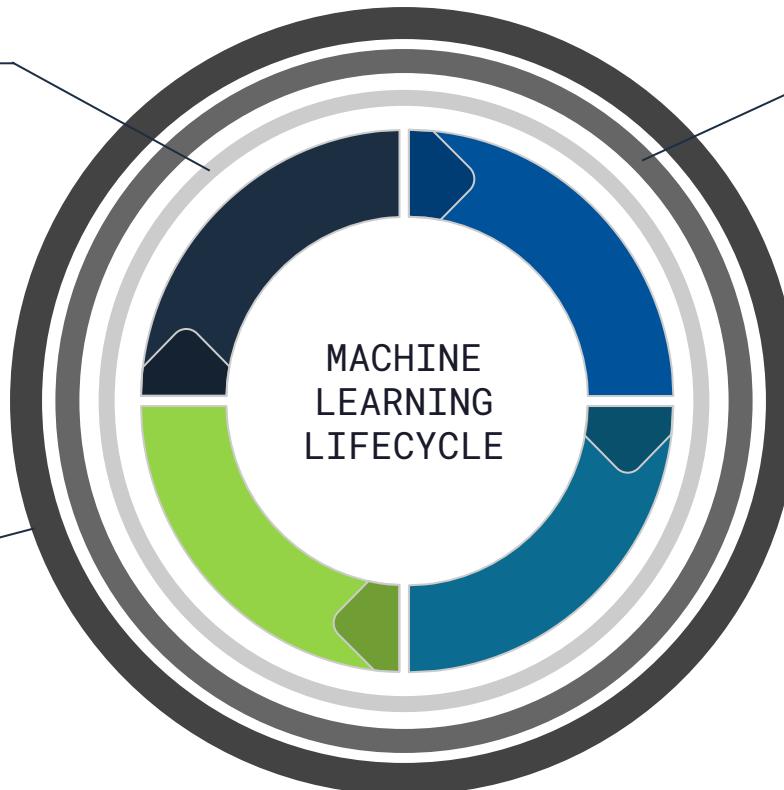
El modelado y construcción está gobernado por principios del propio ciclo de desarrollo del software. El concepto de “one-shot training” es sólo efectivo en casos de uso muy deterministas y por lo general simplifican en exceso la fase de modelado.

PRODUCTIVIZACIÓN Y DESPLIEGUE

El paso a la productivización de un modelo entrenado depende en gran medida de la solución técnica adoptada pero suele estar gobernado por el flujo actual de una pieza de software. Es por ello que la automatización y el uso de soluciones de integración continua juegan un papel fundamental

ENTRENAMIENTO

El entrenamiento de un modelo ML es una etapa recurrente sobre la cual se iterá tras los ajustes en la fase de modelado y es por ello esencial contar con este hecho a la hora de planificar los despliegues



CONTROL DE CALIDAD

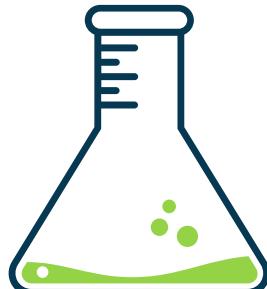
Sobre el ciclo de vida de los modelos ha de existir una capa de control de calidad que permita la medición de su factor de confianza en sus distintas versiones, la validación contra sets de datos de pruebas y en definitiva un capa que controle la calidad del modelo entregado de una forma similar a la que se tendría con piezas de software tradicionales

CAPA DE PLANIFICACIÓN

La capa de planificación permite unir el modelado y construcción de modelos con la fase de control de calidad de forma que permite tomar decisiones sobre cómo, cuándo y de qué forma aquellos modelos desplegados en entornos previos a producción y que hayan superado el control de calidad, están listos para ser consumidos mediante la interfaz estandarizada

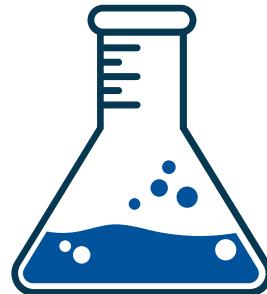
MONITORIZACIÓN

Todo el ecosistema ha de estar debidamente monitorizado de forma que tengamos el control de las iniciativas planificadas, estado de los entornos de trabajo, entrenamientos, versionado de modelos, evolutivos y pases a producción



**Uso de infraestructura
(GPU, TPU, CPU)
(computación)**

Uso de los **recursos** de **computación** para el modelado, entrenamiento o productivización de modelos de ML **sin uso completo** de servicios que cubren el ciclo de vida de ML. Datalabs, GPU, Compute/Container instances



Ciclo de vida de modelos de ML (plataforma)

Framework o plataforma de construcción de modelos de ML. Desde la etapa de la ingeniería de dato hasta su productivización y monitorización. **Basado en servicios serverless**



**Modelos pre-entrenados
(APIS)**

Cubrir necesidades o casos de uso específicos mediante el uso de modelos de ML **pre-entrenados o mediante reinforcement learning** para su consumo mediante **APIs** o **SDKs**. Servicios **gestionados**

← → C 🔒 cloud.google.com Guest :

Google Cloud Why Google Solutions Products Pricing Getting Started Contact Sales English ▾ Console 🌐

Featured Products

AI and Machine Learning → X

Compute Storage Databases Networking Operations Developer Tools Data Analytics AI and Machine Learning API Management Hybrid and Multi-cloud

See all products (100+)

Speech-to-Text Vision AI Text-to-Speech Cloud Translation

Speech recognition and transcription supporting 120 languages.

Custom and pre-trained models to detect emotion, text, more.

Speech synthesis in 180+ voices and 30+ languages.

Language detection, translation, and glossary support.

Natural Language AutoML AI Platform Video AI

Sentiment analysis and classification of unstructured text.

Custom machine learning model training and development.

Platform for training, hosting, and managing ML models.

Video classification and recognition using machine learning.

Cloud TPU Dialogflow AutoML Tables

Tensor processing units for machine learning applications.

Conversation applications and systems development suite.

Service for training ML models with structured data.

Not seeing what you're looking for? See all AI and machine learning products

infrastructure solutions for SAP, VMware, Windows, Oracle, data center migration, and other enterprise workloads.

your data, applications, and users, with advanced anti-malware and threat detection.

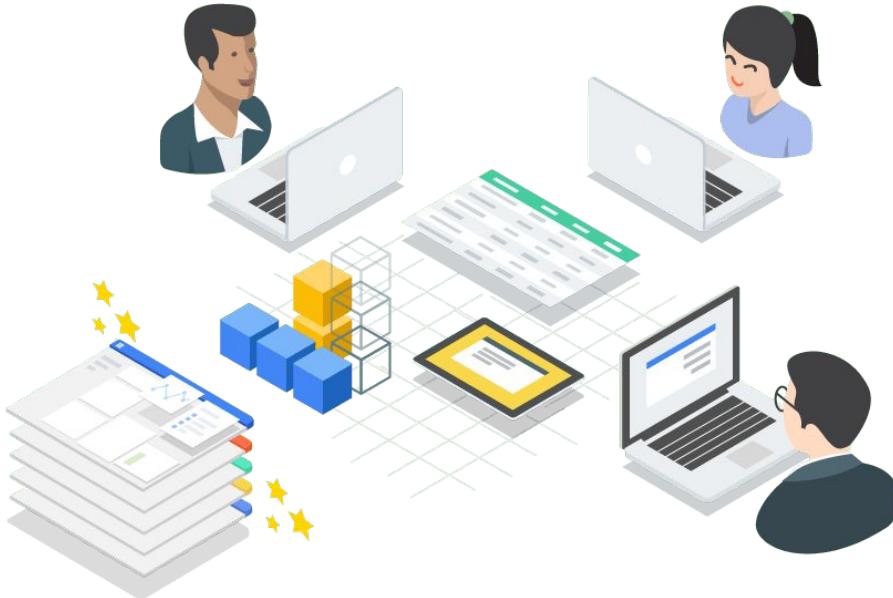
data, with a suite of scalable solutions for data warehouses, analytics, and AI and machine learning.

hybrid and multi-cloud environments with other cloud providers.

1

AI PLATFORM

"ML lifecycle strategy"



- Deep Learning VM
- AI Platform Notebooks
- AI Platform Training
- AI Platform Prediction
- Kubeflow
- Kubeflow pipelines



Kubeflow

2

AI HUB

"Plug-and-play AI components"

NEW

The screenshot shows the AI Hub interface on a computer monitor. At the top, there's a search bar and a 'Learn more' button. Below it, a banner says 'All of AI in one place' with a subtext: 'AI Hub is a platform to discover and deploy machine learning products, from experimentation to production'. A 'Scope' sidebar lists 'Public' and 'Restricted'. Under 'Asset type', it shows 'Kubeflow pipeline', 'Notebook', 'Service', 'Tensorflow module', 'VM image', and 'Trained model'. The main area displays 'Kubeflow pipelines' with three cards:

- Scikit-learn Trainer** by Google. It's a Kubeflow pipeline for model training using scikit-learn.
- Pipeline TFX0** by Google. It's a Kubeflow pipeline for creating a TensorFlow model on structured and image features, supporting classification and regression problems.
- Loan Delinquency Prediction** by Nitin Agrawal. It's a Kubeflow pipeline for predicting loan default and delinquency using TensorFlow and XGBoost models.

Below these, there's a 'View more Kubeflow pipelines' link. In the 'Notebooks' section, there's one card for 'Text generation using a RNN with eager execution'.

- **Hosted AI repository**
- **One-click** deployments for ML teams
- Access to AI components
- Share and deploy state-of-the-art solutions effortlessly

3

ML APIs

"Pre-trained models strategy"

Recommendations API

Deliver personalized recommendations that suit each customer's tastes



Video Intelligence API

Automatically recognize a vast number of objects, places, and actions in stored and streaming video



Cloud Vision

Assign labels to images and quickly classify them into millions of predefined categories



Natural Language

Natural Language uses machine learning to reveal the structure and meaning of text



4

AUTOML

"Reinforcement learning strategy"



AutoML Tables

Create supervised machine learning models with your tabular data



AutoML Video Intelligence

Train a custom video model using your own videos



AutoML Vision

Train a custom model to classify images or to detect objects



AutoML Natural Language

Train a custom video model using your own videos



DEMO: CICLO DE VIDA DE ML

EXTRACCIÓN DE VALOR DE LOS DATOS

EXTRAER VALOR DE LOS DATOS EN ENTORNOS ENTERPRISE

*“While **76%** said (company executives) their firms are aiming to **extract value from the data** they already have, **only 15%** said they currently **have achieved that goal**”*

Source: PwC, Trusted data optimization pulse survey, Feb 2019
(300 +500M-revenue US companies)





PRINCIPALES RETOS PARA EXTRAER VALOR DE LOS DATOS



CONFIANZA EN EL DATO

PLATAFORMAS IT NO SOPORTAN AI

SILOS DE DATOS E INACCESIBILIDAD

FALTA DE TALENTO ANALÍTICO

Source: PwC, Trusted data optimization pulse survey, Feb 2019

DATOS DESESTRUCTURADOS

Son aquellos datos que **no cuentan** con un *schema* fácilmente entendible por máquinas

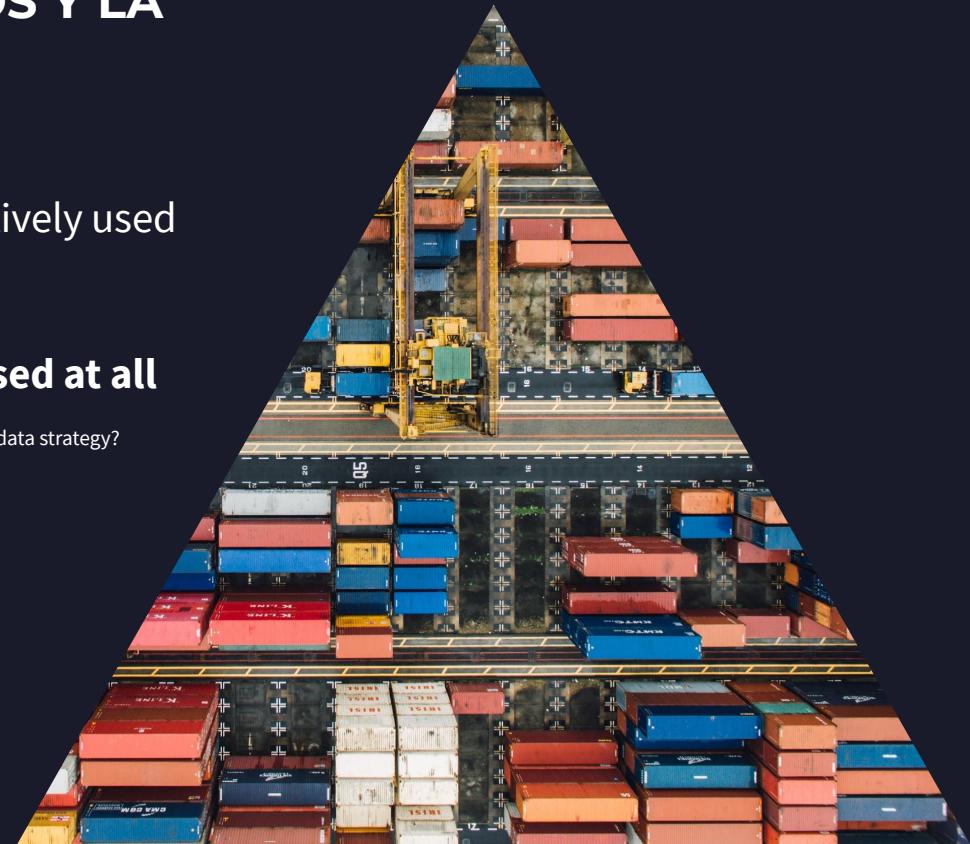


ESTRATEGIA EN EL USO DE DATOS Y LA TOMA DE DECISIONES

< 50% of an organization's structured data is actively used in making decisions

< 1% of its unstructured data is analyzed or used at all

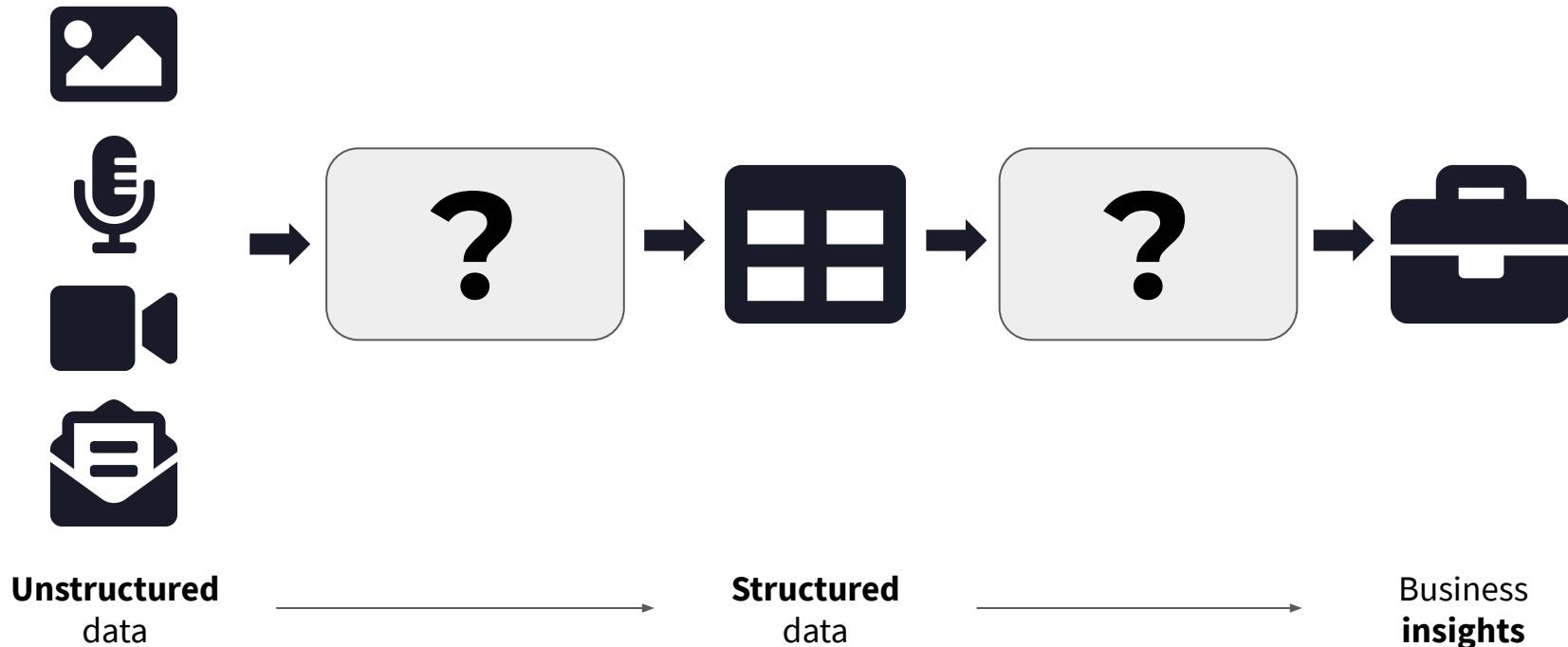
Source: Harvard Business Review (HBR), 2017/15 What's your data strategy?



EXTRACCIÓN DE VALOR DE LOS DATOS DESESTRUCTURADOS

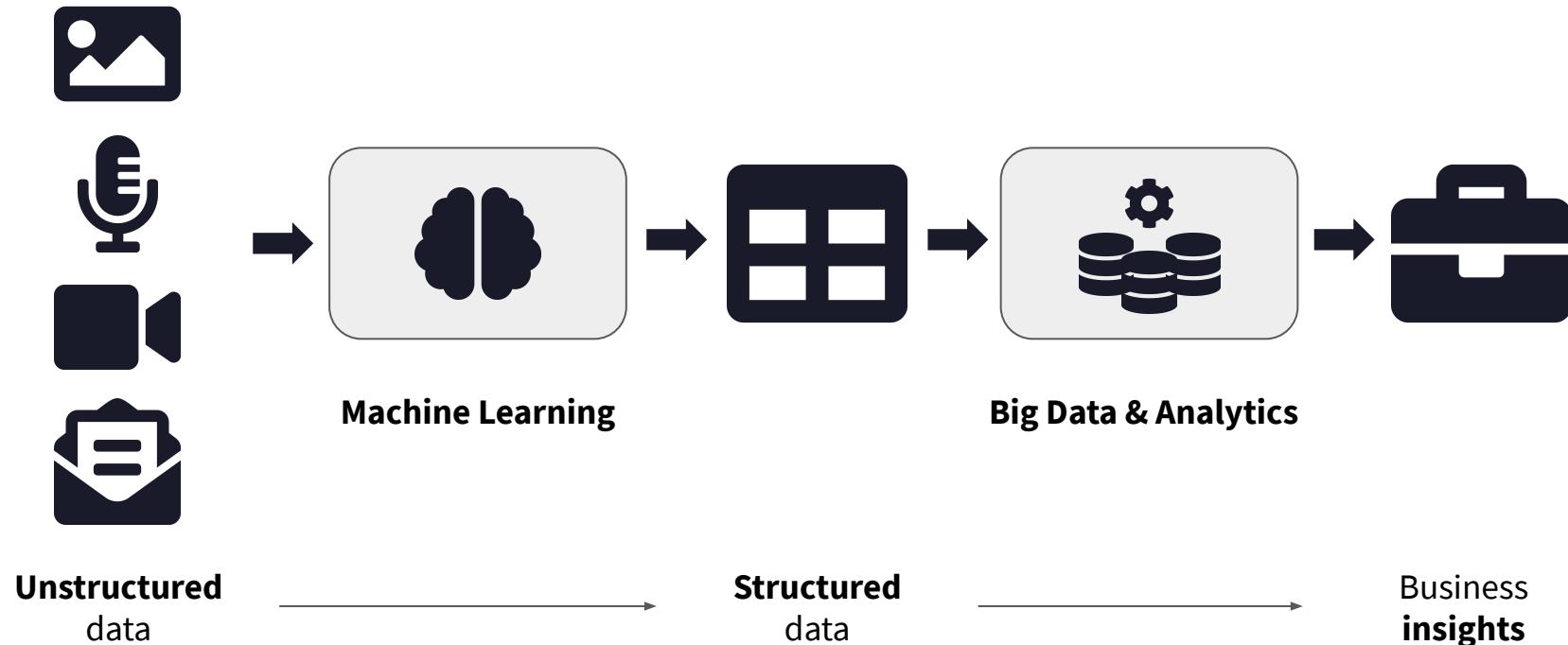
EXTRACCIÓN DE VALOR DE LOS DATOS DESESTRUCTURADOS

keepler



EXTRACCIÓN DE VALOR DE LOS DATOS DESESTRUCTURADOS

keepler



Objects & OCR



Módulo o servicio de ML para el reconocimiento de objetos y caracteres en imágenes

Speech-to-Text



Servicio de reconocimiento del habla y expresiones en distintos lenguajes, dialectos y acentos



NLP

Análisis de texto para la identificación de referencias, estructura de expresiones, análisis de sentimiento y extracción de claves



Recommendations

Sistema recomendadores basados en múltiples fuentes de datos y eventos

1

CONTACT CENTER

- Uso de servicios de **Speech-to-Text** para el reconocimiento de expresiones, **intenciones**, términos, sugerencias, **preguntas** u operaciones
- **Text-to-Speech** habilita la opción de generar respuestas automatizadas en base a criterios estáticos o de aprendizaje ML

2

GESTOR DOCUMENTAL Y DE EMAILS

- Gracias a servicios de **OCR** y de **NLP** se identifican **patrones**, términos, expresiones, **caracteres**, figuras que permiten la **clasificación automática de documentos**
- Construcción de plataformas de **reenvío, clasificación y tramitación de expedientes y correos electrónicos** mediante OCR, NLP, modelos de **ML** y reglas de **negocio**

3

RECOMENDADORES E INSIGHTS DE NEGOCIO

- Construcción de **motores de recomendación** basado en datos estructurados, históricos y su cruce con el aprendizaje de datos desestructurados como **reviews** de productos
- Descubrimiento y aprendizaje de **tendencias de mercado** de productos y servicios agregando **inteligencia de negocio** y fuentes de **datos desestructurados**

DEMO: CASO DEL YOGURT

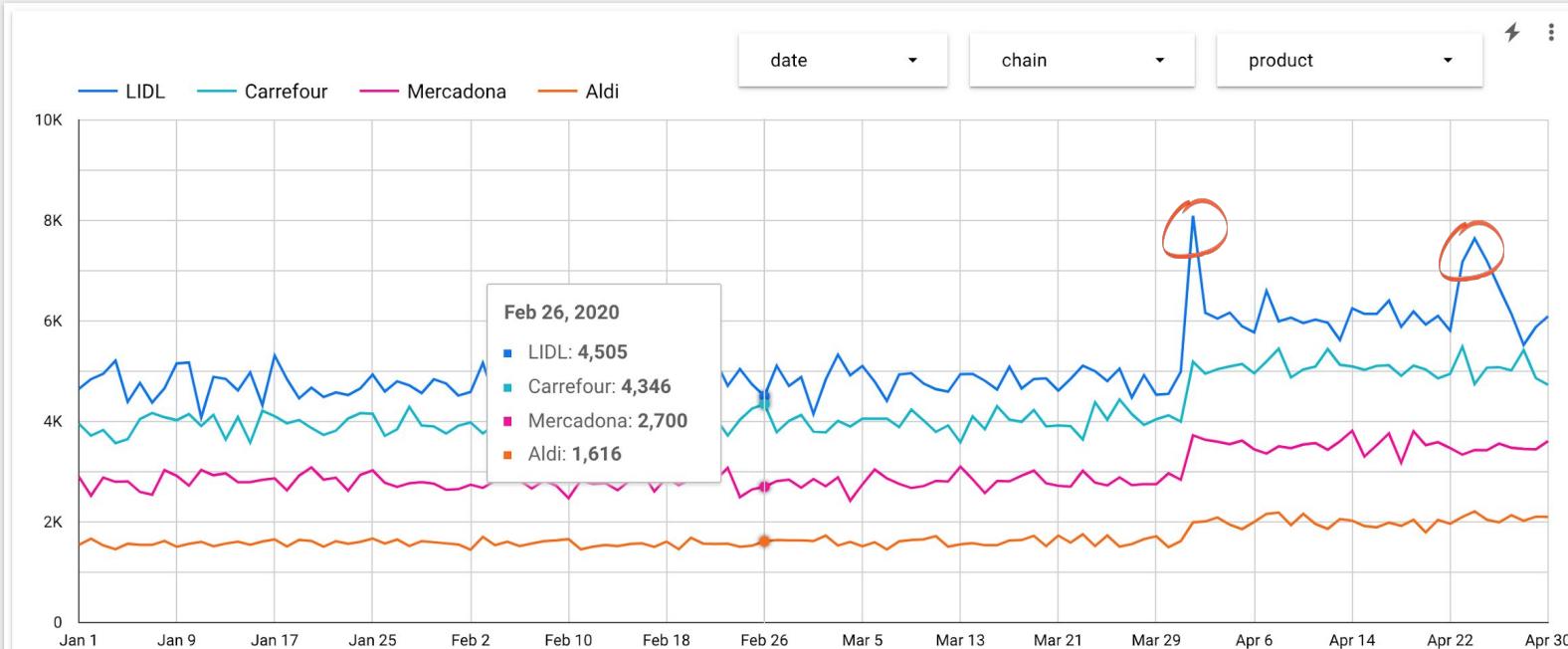
Cualquier parecido con la
realidad es pura
coincidencia

- Dar respuesta a las necesidades de negocio del vertical de yogures de una multinacional alimentaria
- Recientemente se ha **lanzado un nuevo producto al mercado** y quieres conocer:
 - La tendencia en ventas y **explicar los resultados** extrayendo valor de los datos
 - La acogida del producto y su **valoración entre los usuarios**

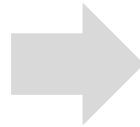


EXPLICACIÓN DE LA TENDENCIA

keeper



- Semana del 1 de abril de 2020
 - **Displays** en tienda
 - Promo inicial
 - Lanzamiento de **nuevo producto**

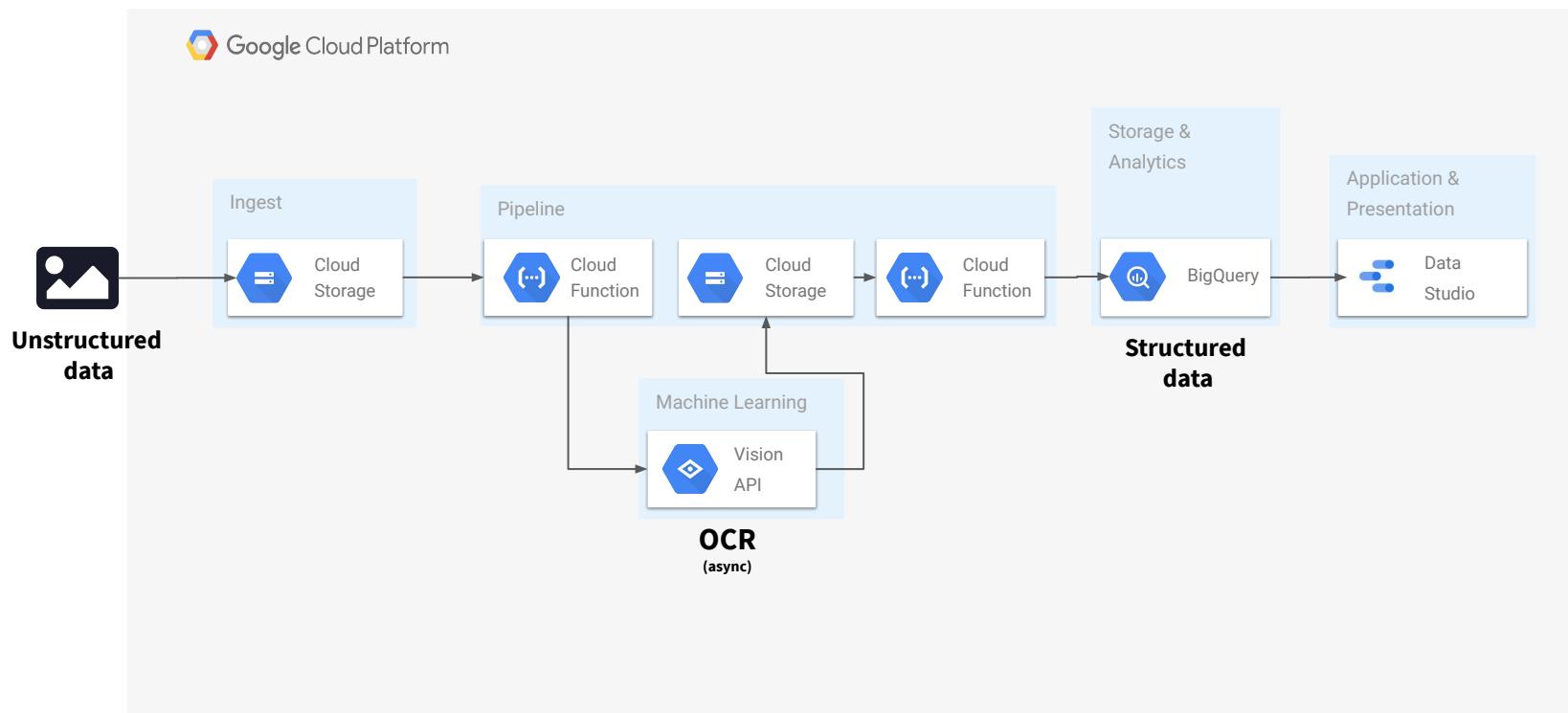


- Semana del 23 de abril de 2020
 - **Estudio de datos**
 - Análisis de folletos de la cadena líder

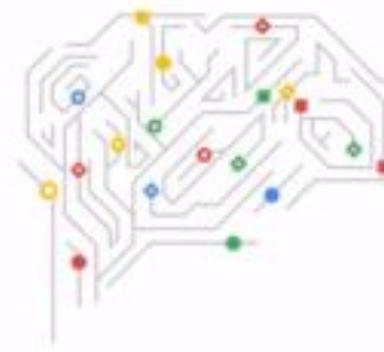
PIPELINE DE EXTRACCIÓN DE DATOS DE FOLLETOS

keeper

Architecture: Catalog OCR data



ARQUITECTURAS ORIENTADAS A EVENTOS EN LA NUBE





Kat L.
Fresno , CA
23 reviews



★★★★★
Purchased at: Walmart
March 7th 2020, 4:09 am

This product taste good and it was able to fill my craving for Something sweet. It was only 100 cal for one complete serving. They are reasonably price.DANNON® OIKOS® TRIPLE ZERO SALTED CARAMEL®

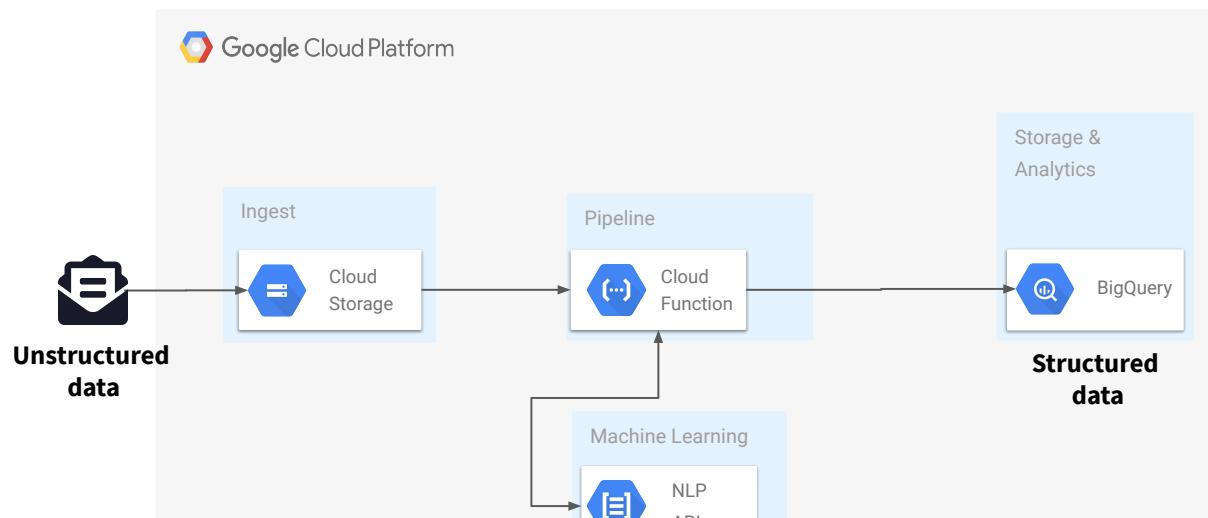
4  ...

- Conocer las **opiniones de los productos**
- Conocer los **términos** con los que se suele **asociar**
- Conocer el **sentimiento** general de las reviews al comprar el producto
- Analizar el texto para futuras **acciones comerciales**

PIPELINE DE ANÁLISIS DE REVIEWS

keeper

Architecture: Reviews NLP data



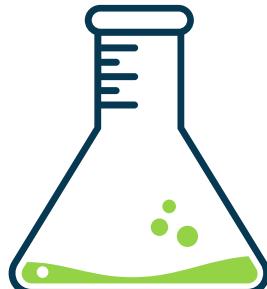
NLP
(sync)

- Usar esta info para anticipar/inferir el **resultado de campañas comerciales**
- Incluir servicios de reinforcement learning como **AutoNLP** y **AutoVision** para nuestros **casos de uso de clasificación concreta**
- Construcción de modelos ML para **score específico**



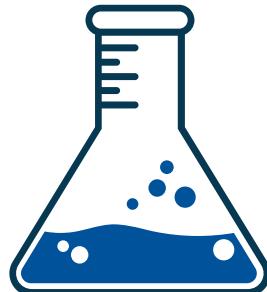


RECAP



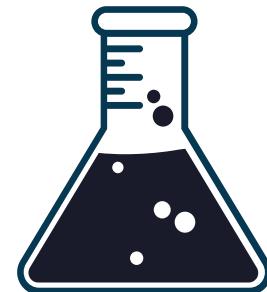
**Uso de infraestructura
(GPU, TPU, CPU)
(computación)**

Uso de los **recursos** de **computación** para el modelado, entrenamiento o productivización de modelos de ML **sin uso completo** de servicios que cubren el ciclo de vida de ML. Datalabs, GPU, Compute/Container instances



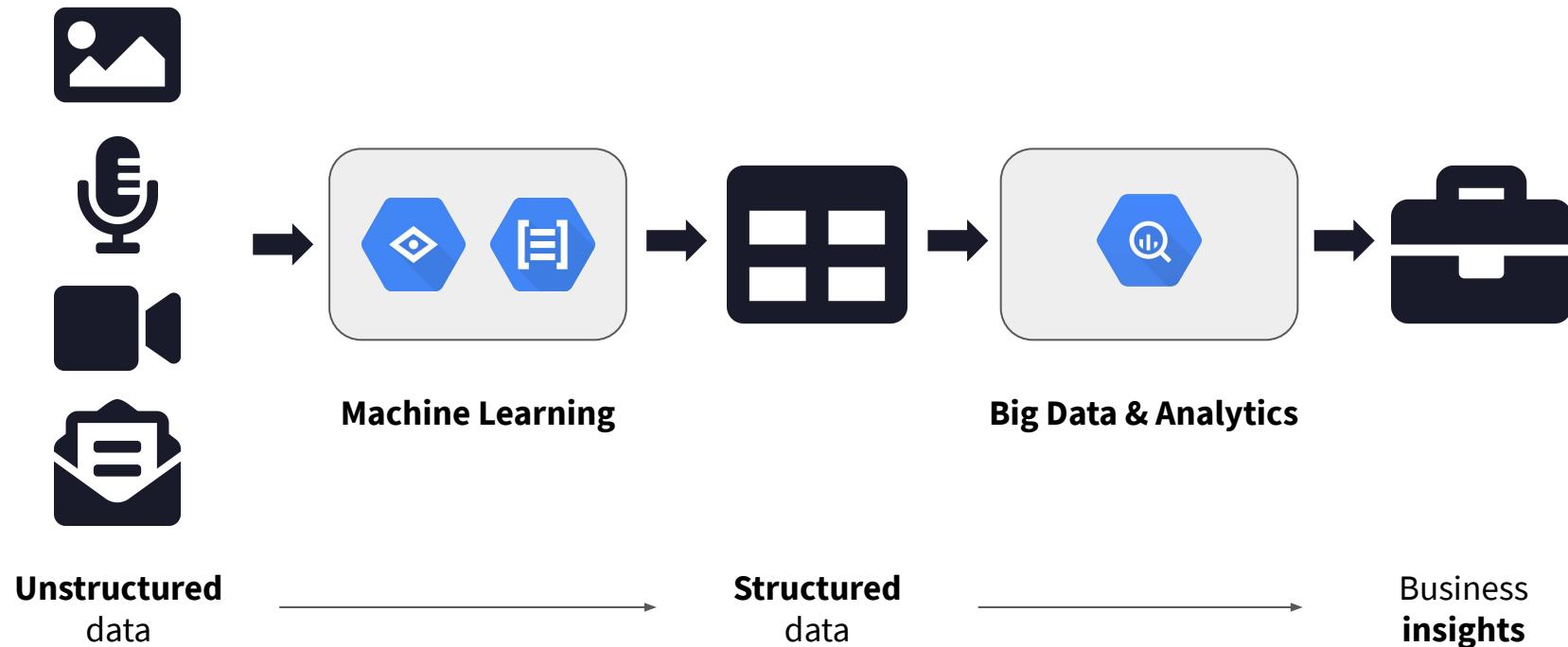
Ciclo de vida de modelos de ML (plataforma)

Framework o plataforma de construcción de modelos de ML. Desde la etapa de la ingeniería de dato hasta su productivización y monitorización. **Basado en servicios serverless**



**Modelos pre-entrenados
(APIS)**

Cubrir necesidades o casos de uso específicos mediante el uso de modelos de ML **pre-entrenados o mediante reinforcement learning** para su consumo mediante **APIs** o **SDKs**. Servicios **gestionados**



THANKS!



Sergio Gordillo

Cloud Architect en **Keepler Data Tech**

sergio.gordillo@keepler.io / @Sergio_Gordillo

WHAT?

Cloud Computing

Big Data

Artificial Intelligence

HOW?

TRANSFORMACIÓN AGILE

METODOLOGÍA / FRAMEWORK DEVOPS

ENFOQUE DATA PRODUCT

EXCELENCIA TÉCNICA

WHERE?

The information contained in this document is property of KEEPLER DATA TECH and intended only for the person or entity to which it is sent. It may contain confidential and / or privileged material, the use of this information or any disclosure, copying or distribution is prohibited and may be unlawful. If you received this in error, please contact the sender and delete all copies.