





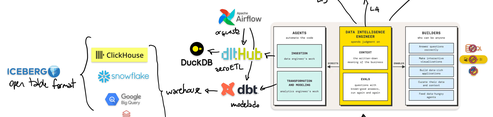
CDC
debezium → Broker kafka → Processing Flink → Streamhouse Fluss

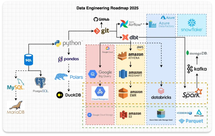
Streaming (L6)

Data Ops y FinOps

unstructured y RAGs

L5 L5 L4



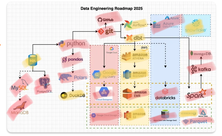


Empezar 2027 Aquí

L2 L3 L4 L5



8



9

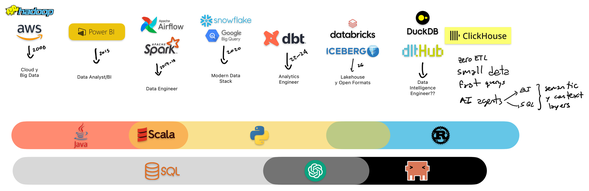




7

6

Historia



¿Qué skills valora el mercado IT?

La particularmente en datos

Ls y ahora que AI hace el código

1

Tipos de Empresas

- startup → producto y velocidad
- pyme → uso eficiente
- corp → control y calidad bi fiabilidad
- ≠ por industria → presupuesto, GI datos, regulación

5

líderes

equipos → chicos rápidos

negocio y todos codean



2

Tipos de Roles

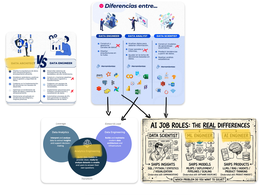


4

Necesidades



3





¿Qué skills valora el mercado IT?

1

Tipo
↳
↳
↳
↳

Ls particularmente en datos
Ls y ahora que AI hace el código

2

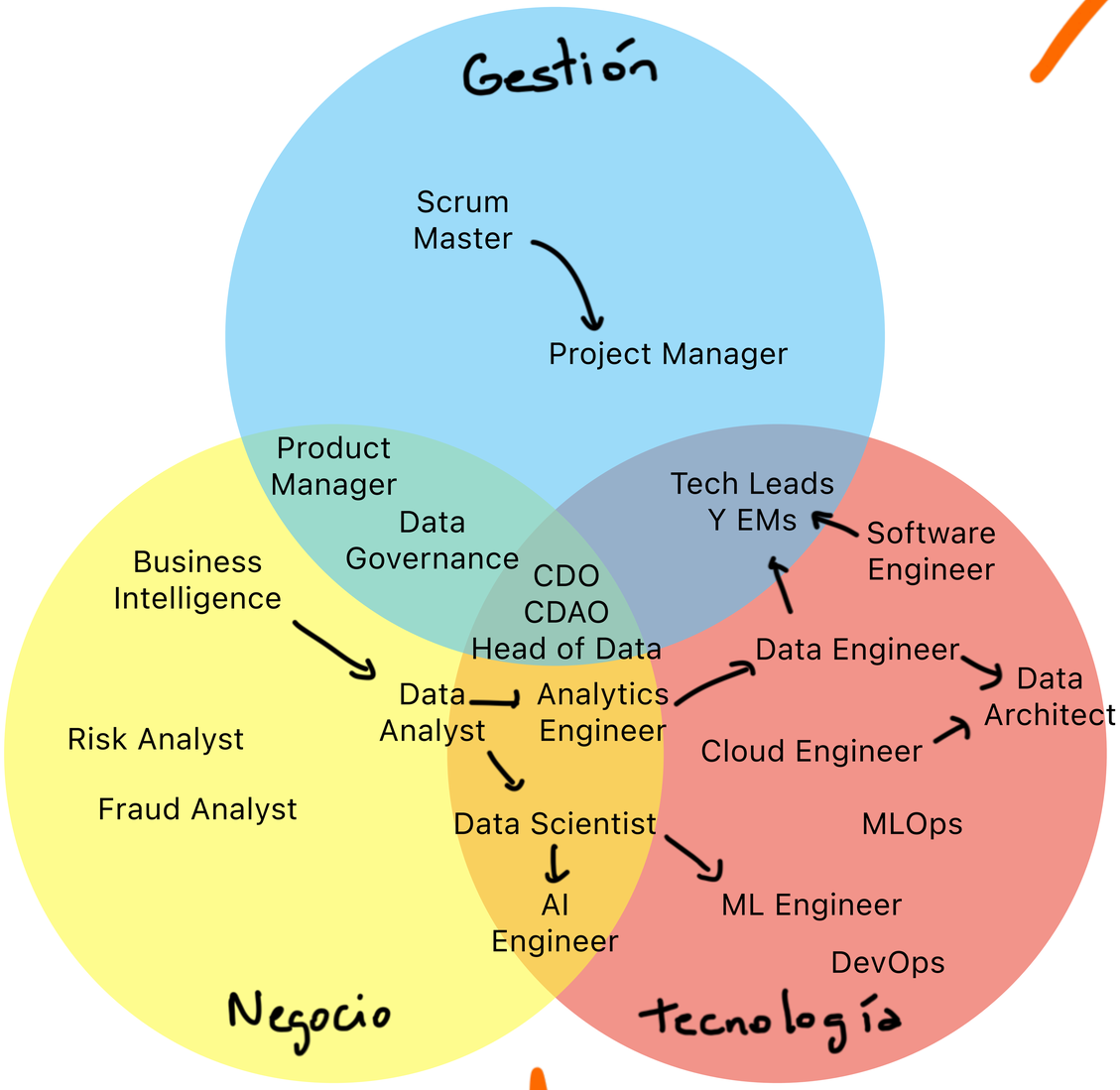
4

Tipos de Empresas

- startup → producto y velocidad
- ayne → versatilidad
- corpo → control y escalabilidad
- ≠ por industria → presupuesto
  - → Q datos
  - → regulación

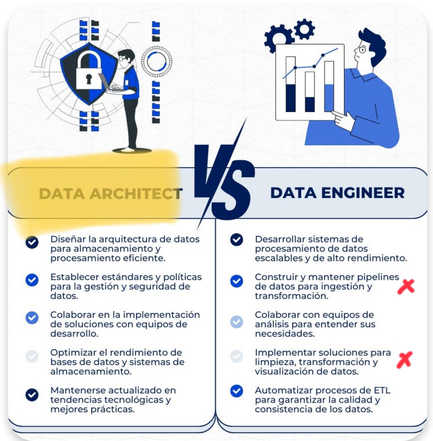
ódigo

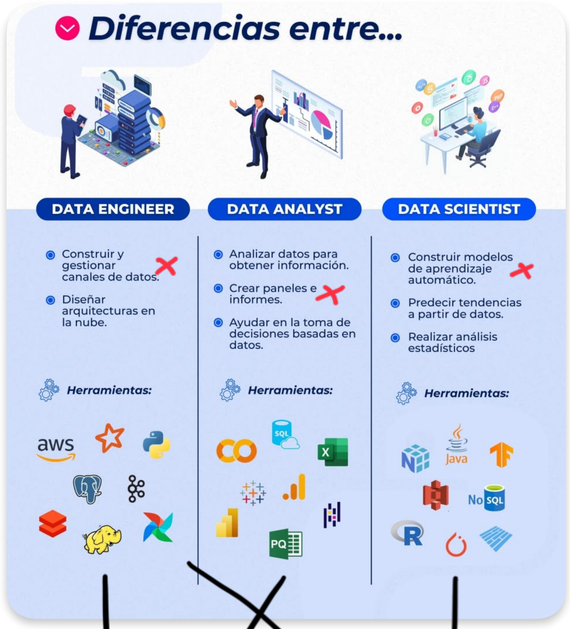
Tipos de Roles

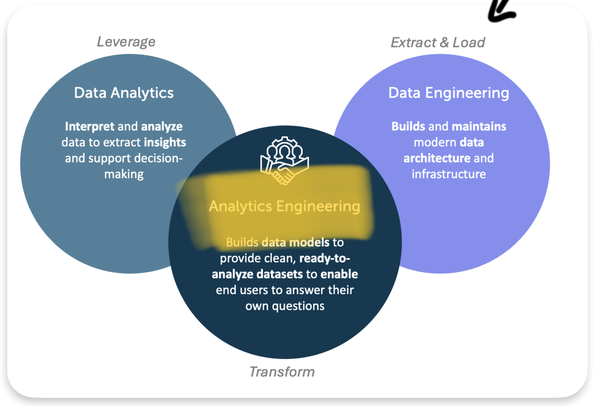


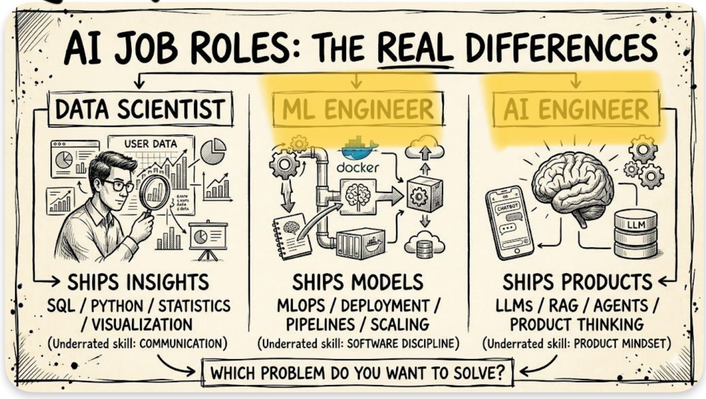
BI
SQL
} semantic y context layers

3


DATA ARCHITECT
VS
DATA ENGINEER
Diseñar la arquitectura de datos para almacenamiento y procesamiento eficiente.
Establecer estándares y políticas para la gestión y seguridad de datos.
Colaborar en la implementación de soluciones con equipos de desarrollo.
Optimizar el rendimiento de bases de datos y sistemas de almacenamiento.
Mantenerse actualizado en tendencias tecnológicas y mejores prácticas.
Desarrollar sistemas de procesamiento de datos escalables y de alto rendimiento.
Construir y mantener pipelines de datos para ingestión y transformación.
Colaborar con equipos de análisis para entender sus necesidades.
Implementar soluciones para limpieza, transformación y visualización de datos.
Automatizar procesos de ETL para garantizar la calidad y consistencia de los datos.


Diferencias entre...
DATA ENGINEER
DATA ANALYST
DATA SCIENTIST
Construir y gestionar canales de datos.
Diseñar arquitecturas en la nube.
Analizar datos para obtener información.
Crear paneles e informes.
Ayudar en la toma de decisiones basadas en datos.
Construir modelos de aprendizaje automático.
Predecir tendencias a partir de datos.
Realizar análisis estadísticos
Herramientas:
Herramientas:
Herramientas:


Leverage
Extract & Load
Data Analytics
Interpret and analyze data to extract insights and support decision-making
Data Engineering
Builds and maintains modern data architecture and infrastructure
Analytics Engineering
Builds data models to provide clean, ready-to-analyze datasets to enable end users to answer their own questions
Transform


AI JOB ROLES: THE REAL DIFFERENCES
DATA SCIENTIST
ML ENGINEER
AI ENGINEER
USER DATA
docker
LLM
SHIPS INSIGHTS
SQL / PYTHON / STATISTICS / VISUALIZATION
(Underrated skill: COMMUNICATION)
SHIPS MODELS
MLOPS / DEPLOYMENT / PIPELINES / SCALING
(Underrated skill: SOFTWARE DISCIPLINE)
SHIPS PRODUCTS
LLMs / RAG / AGENTS / PRODUCT THINKING
(Underrated skill: PRODUCT MINDSET)
WHICH PROBLEM DO YOU WANT TO SOLVE?



Necesidades

Startup / scale-up

1. **AI Engineer**: el producto mismo usa IA.
2. **Software / Product Engineer**: alguien que lleva una función de la idea a producción.
3. **Data Engineer generalista**: aparece en la scale-up, cuando los datos ya duelen.
4. **Product Manager**: decide qué construir cuando programar ya es barato.

- *Pesan poco*: Scrum Master, Project Manager y Governance, porque es estructura que todavía no necesitan.

Pyme

1. **Data Analyst / BI**: el primer valor está en los tableros y los números ordenados.
2. **Analytics Engineer generalista**: el "de datos" que hace todo.
3. **Especialista en automatización** (n8n, no-code, agentes livianos). *Te falta en el diagrama*: es donde más demanda real va a haber.
4. **Business Analyst**: traduce el negocio, porque no hay PM.

- *Pesan poco*: ML Engineer, MLOps y Data Architect, porque no hay escala que los justifique.

Corpo

1. **Data Architect / Platform**: tiene que integrar sistemas heredados con la IA.
2. **Data Governance**: sin linaje ni calidad de datos, la IA no pasa a producción.
3. **MLOps**: pasar de la prueba piloto a producción es su cuello de botella.
4. **Tech Leads / EMs y Project Manager**: coordinan muchos equipos.
5. **Seguridad y privacidad de datos/IA**. *También te falta*.

- *Pesan poco*: los generalistas que hacen todo, porque la corpo segmenta los roles.

Cómo pesa la industria

- **Presupuesto**: con plata se compran plataformas enterprise y se arman equipos especializados; sin plata, open source y generalistas.
- **Volumen de datos**: la IA necesita datos limpios, así que **más datos significa más valor para data engineering y plataforma**.
- **Regulación**: frena la IA generativa en producción y aumenta la demanda de linaje, explicabilidad y auditoría con una persona revisando. La habilidad escasa es **poder auditar lo que hizo la IA**.

Si la industria es regulada (banca, salud, seguros), suben **Governance, Risk/ Fraud Analyst, MLOps y Seguridad. Si es liviana**, suben **AI Engineer y Product Engineer**.

- *Pesan poco*: los generalistas que hacen todo, porque la corpo segmenta los roles.

Si la industria es regulada (banca, salud, seguros), suben **Governance, Risk/Fraud Analyst, MLOps y Seguridad. Si es liviana**, suben **AI Engineer y Product Engineer**.





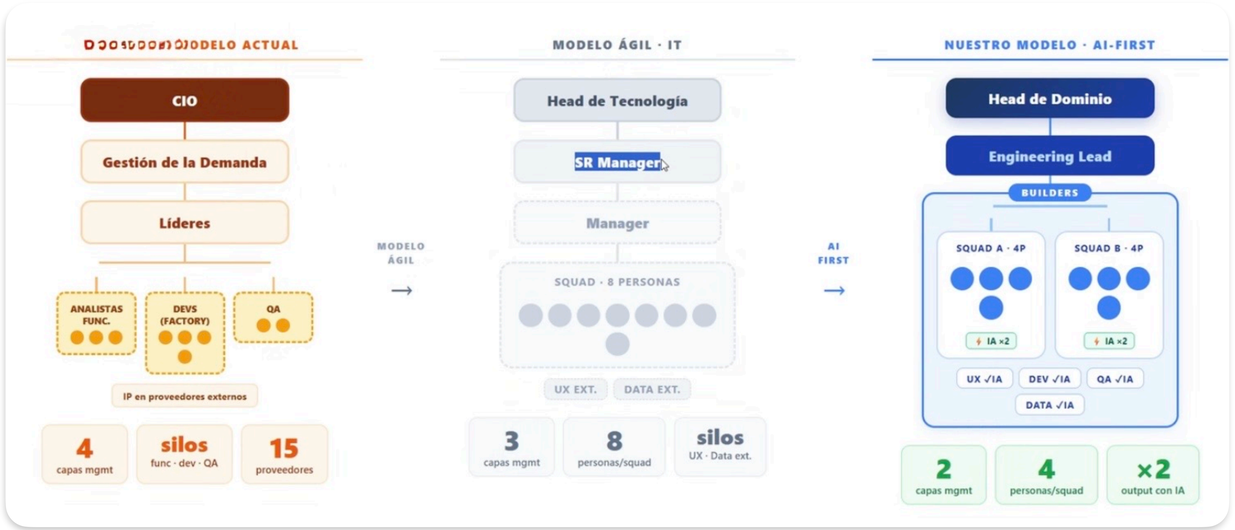
...sas

...y velocidad
...dad
...scalabilidad
→ presupuesto
→ Q datos
→ regulación

líderes

equipos → chicos
→ rápidos

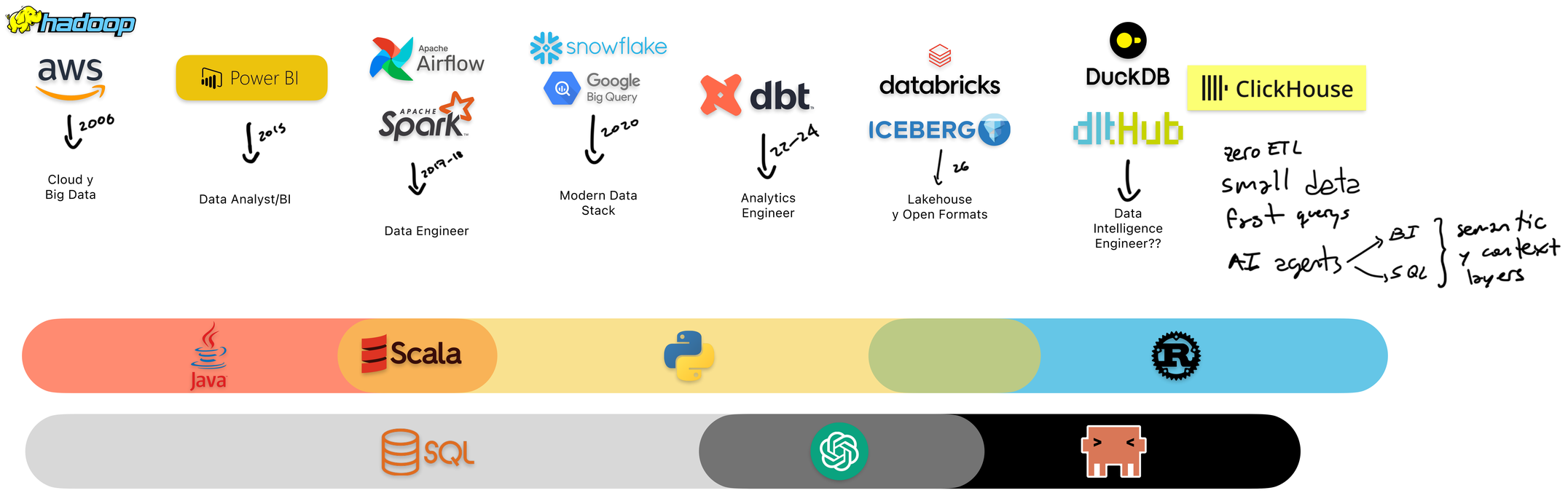
+negocio
y todos codean

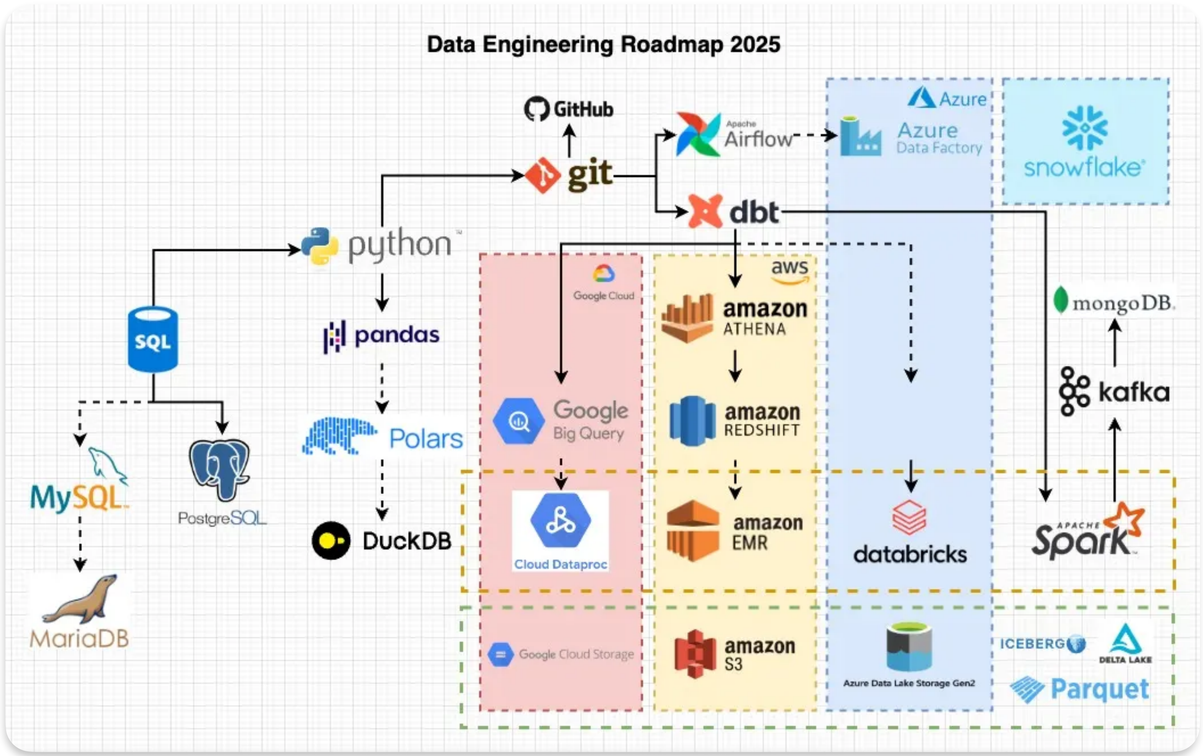


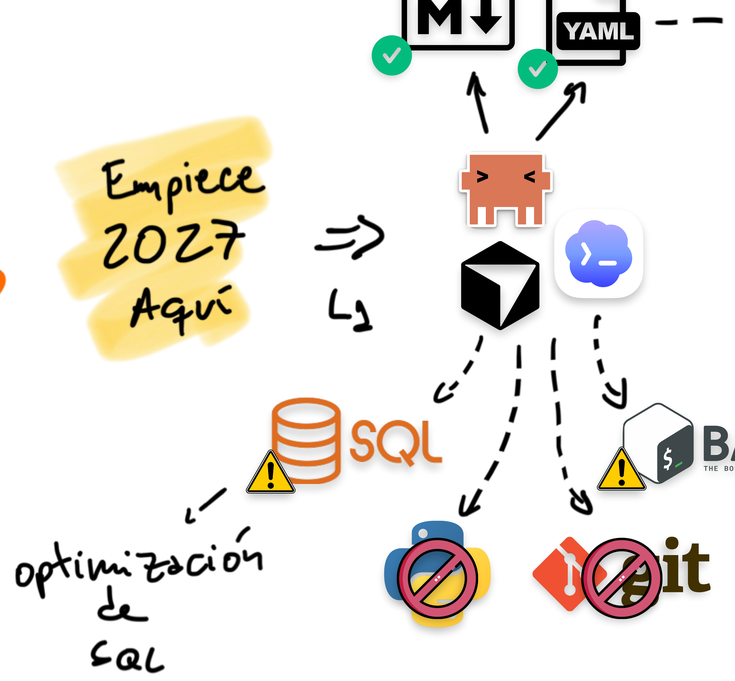
...sidades

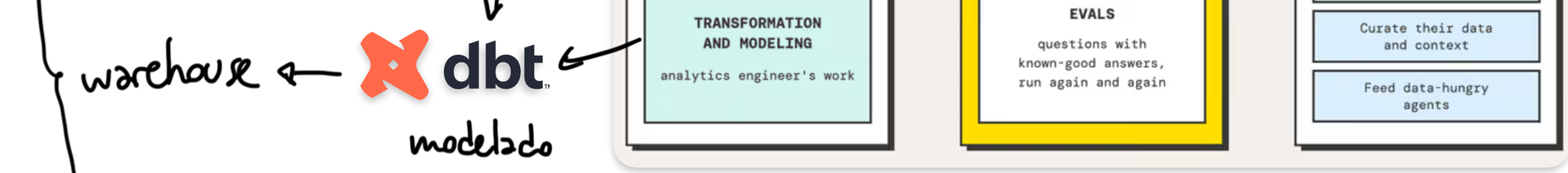
Pyme

Historia


hadoop
aws
2006
Cloud y
Big Data
Power BI
2015
Data Analyst/BI
Apache Airflow
APACHE Spark
2019-18
Data Engineer
snowflake
Google Big Query
2020
Modern Data
Stack
dbt
22-24
Analytics
Engineer
databricks
ICEBERG
26
Lakehouse
y Open Formats
DuckDB
dltHub
Data
Intelligence
Engineer??
ClickHouse
zero ETL
small data
fast queys
AI agents
BI
SQL
semantic
y context
layers
Java
Scala
SQL


Data Engineering Roadmap 2025
GitHub
git
Apache Airflow
Azure
Azure Data Factory
snowflake
dbt
python
SQL
pandas
Polars
DuckDB
MySQL
PostgreSQL
MariaDB
Google Cloud
Google Big Query
Cloud Dataproc
Google Cloud Storage
aws
amazon ATHENA
amazon REDSHIFT
amazon EMR
amazon S3
databricks
Azure Data Lake Storage Gen2
mongoDB.
kafka
APACHE Spark
ICEBERG
DELTA LAKE
Parquet


Empiece 2027 Aquí
YAML
SQL
optimización de SQL



warehouse ← dbt ← modelado

L2

why



8

Empiece 2027 Aquí

L1

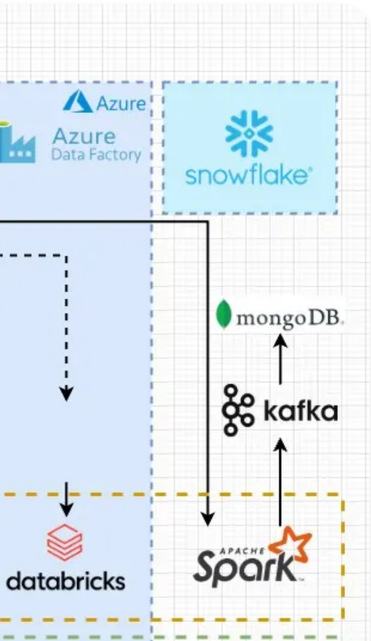
guardrails

SQL

BASH
THE BOURNE-AGAIN SHELL

PR Review

optimización de SQL



Apache Airflow

orquesta

DuckDB ← dltHub → zeroETL

ClickHouse

snowflake

Google Big Query

databricks

warehouse ← dbt → modelado

AGENTS	DATA INTELLIGENCE ENGINEER	BUILDERS
automate the code	spends judgment on	who can be anyone
INGESTION: data engineer's work	CONTEXT: the written-down meaning of the business	Answer questions correctly
TRANSFORMATION AND MODELING: analytics engineer's work	EVALS: questions with known-good answers, run again and again	Make interactive visualizations
		Build data-rich applications
		Curate their data and context
		Feed data-hungry agents

DIRECTS → ENABLES →

L2

why

Google Cloud

Open Knowledge Format (OKF)

Un formato abierto para organizar conocimiento en Markdown y YAML, comprensible para personas y agentes de IA.

Abierto y sin vendor lock-in · Basado en Markdown y YAML · Conectado y referenciable · Comprensible para personas y agentes de IA

Estandariza el conocimiento. Potencia agentes de IA más inteligentes.

Empiece 2027 Aquí

L1

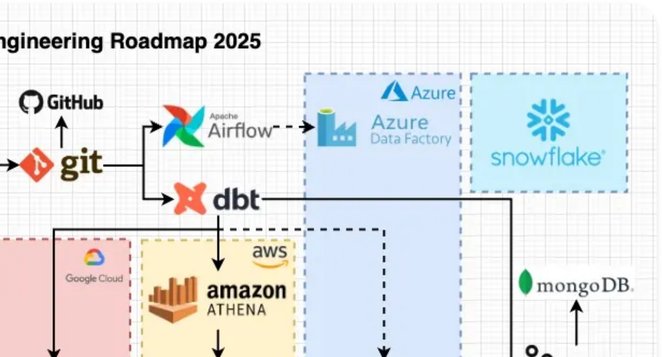
SQL

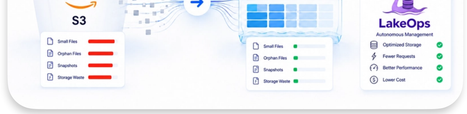
optimización de SQL

BASH

guardrails

PR Review





y FinOps

y RAG

L3

L4

Apache Airflow

orquesta

DuckDB

dltHub

zeroETL

ICEBERG

open table format

ClickHouse

snowflake

Google Big Query

databricks

warehouse

dbt

modelado

AGENTS
automate the code

INGESTION
data engineer's work

TRANSFORMATION AND MODELING
analytics engineer's work

DIRECTS

DATA INTELLIGENCE ENGINEER
spends judgment on

CONTEXT
the written-down meaning of the business

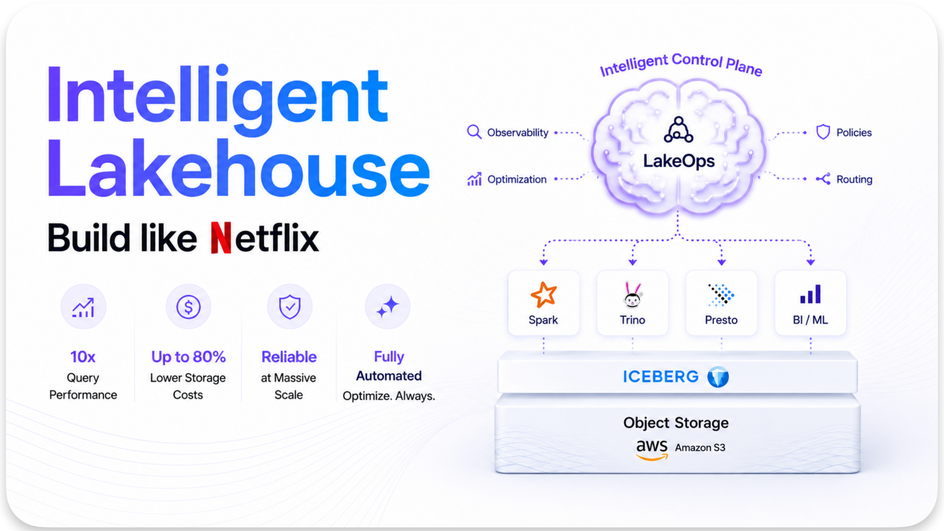
EVALS
questions with known-good answers, run again and again

L2

M↓

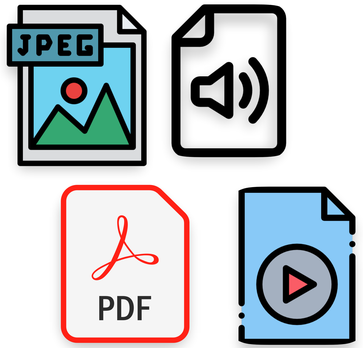
YAML

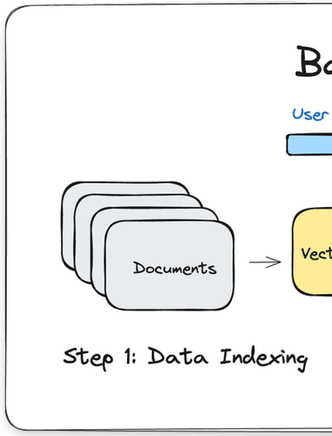
Empiece 2027











Data Ops

y FinOps

unstructu

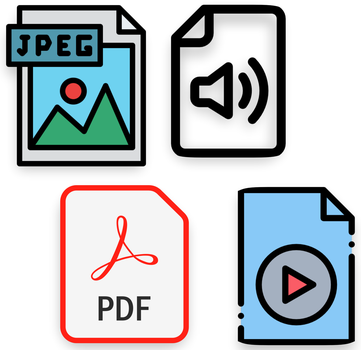
y RAG

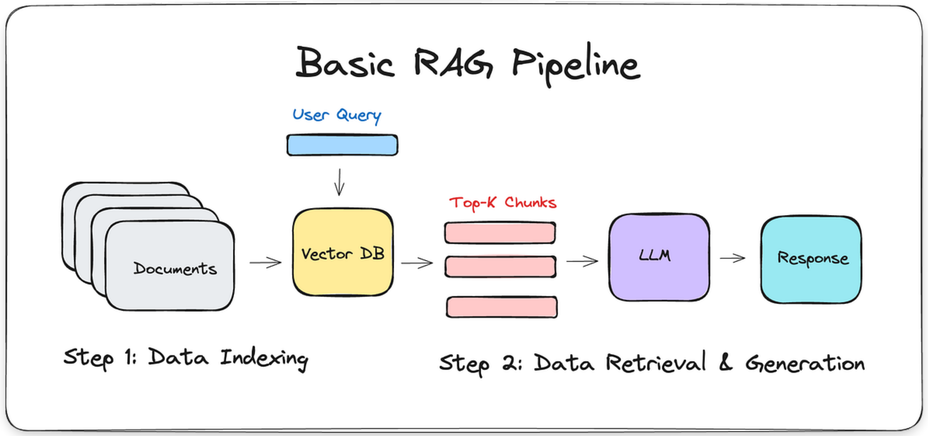
L3

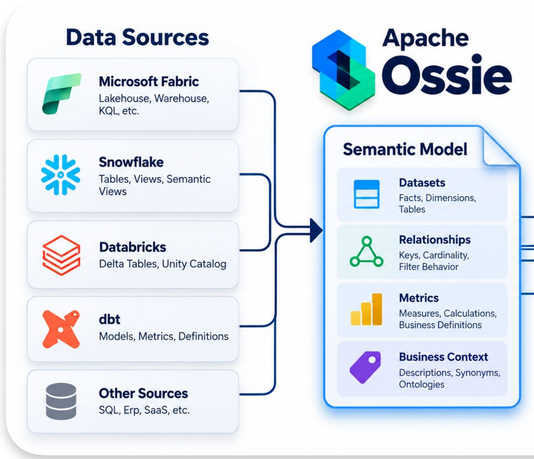
L4











LakeOps

Data Ops

y FinOps

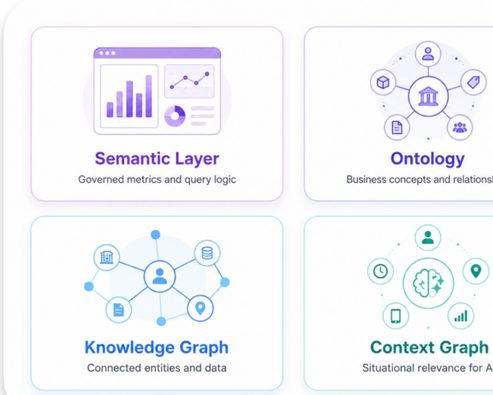
unstructured

y RAG

L3

L4

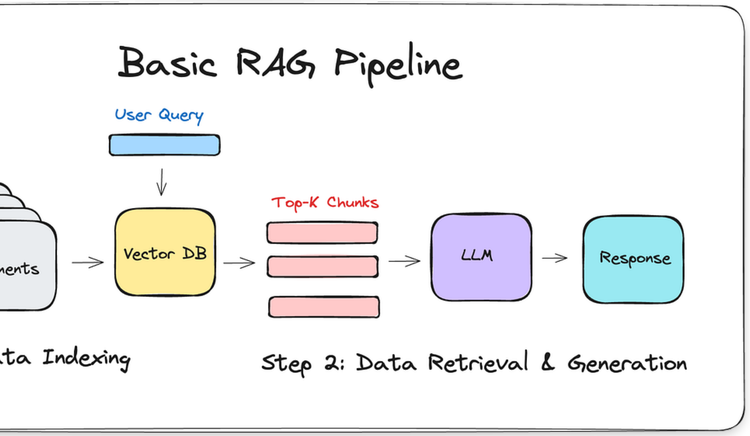
L5

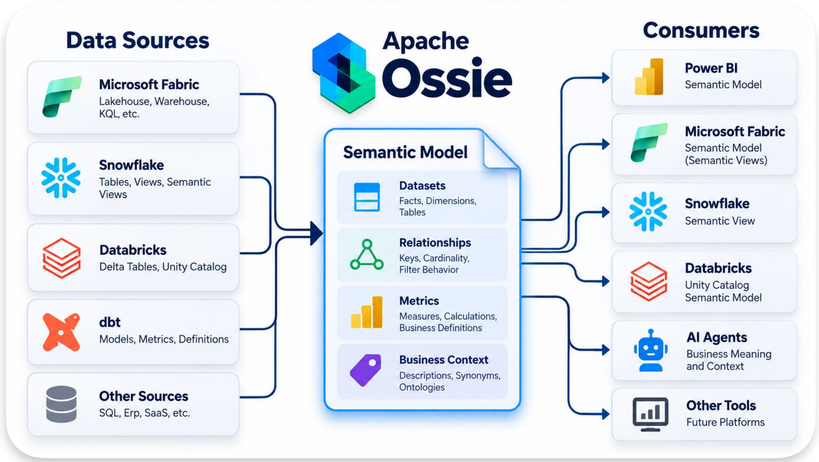


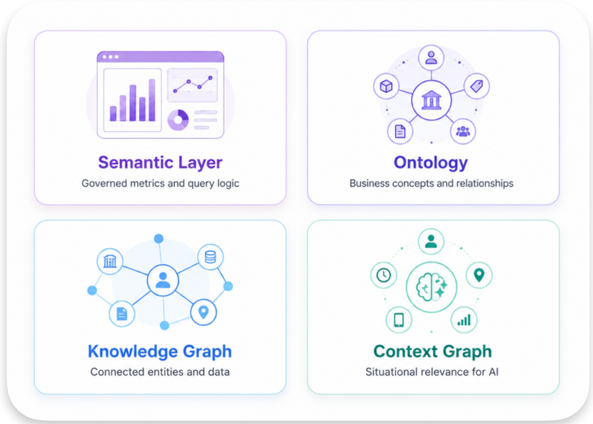
Otras Layers

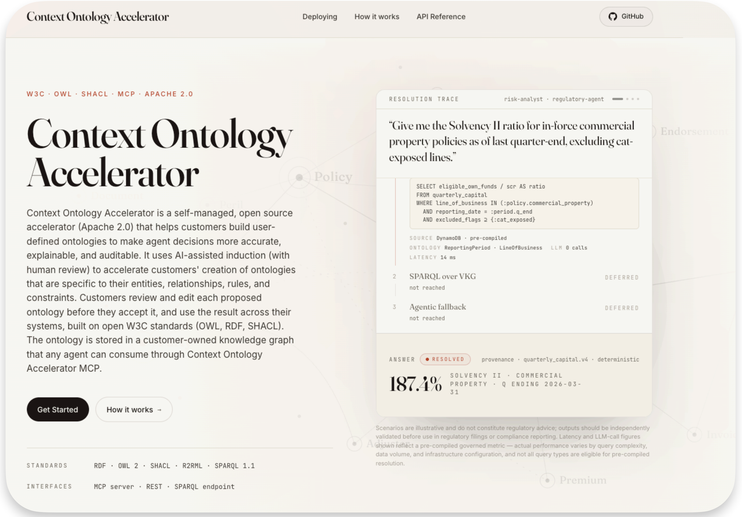
Apache Airflow



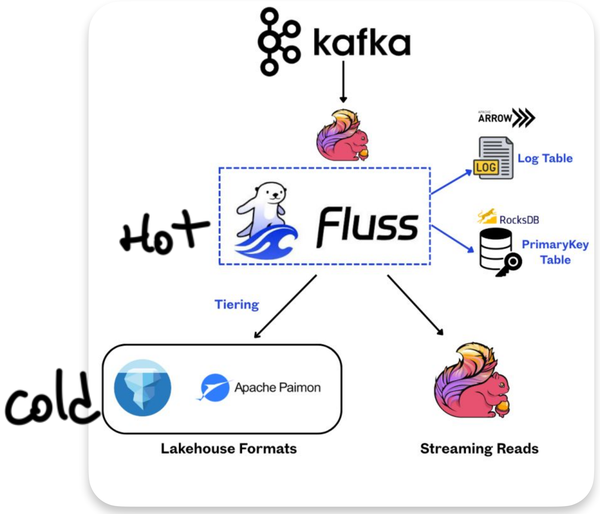

Basic RAG Pipeline
User Query
Top-K Chunks
Vector DB
LLM
Response
Step 2: Data Retrieval & Generation

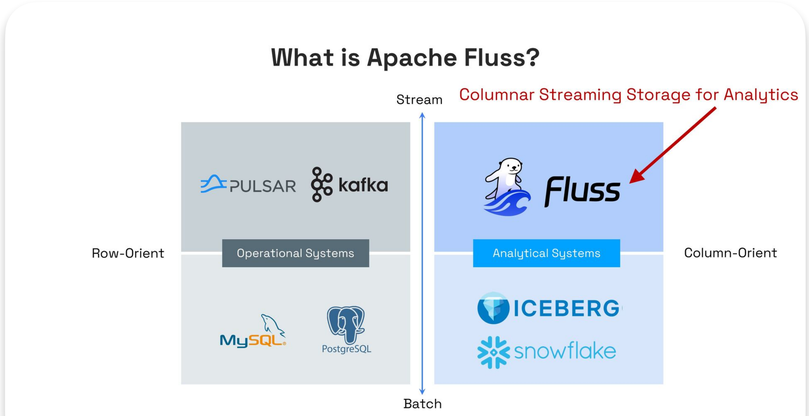

Data Sources
Microsoft Fabric
Lakehouse, Warehouse, KQL, etc.
Snowflake
Tables, Views, Semantic Views
Databricks
Delta Tables, Unity Catalog
dbt
Models, Metrics, Definitions
Other Sources
SQL, Erp, SaaS, etc.
Apache Ossie
Semantic Model
Datasets
Facts, Dimensions, Tables
Relationships
Keys, Cardinality, Filter Behavior
Metrics
Measures, Calculations, Business Definitions
Business Context
Descriptions, Synonyms, Ontologies
Consumers
Power BI
Semantic Model
Microsoft Fabric
Semantic Model (Semantic Views)
Snowflake
Semantic View
Databricks
Unity Catalog Semantic Model
AI Agents
Business Meaning and Context
Other Tools
Future Platforms


Semantic Layer
Governed metrics and query logic
Ontology
Business concepts and relationships
Knowledge Graph
Connected entities and data
Context Graph
Situational relevance for AI


Context Ontology Accelerator
Deploying
How it works
API Reference
GitHub
W3C · OWL · SHACL · MCP · APACHE 2.0
Context Ontology Accelerator
Context Ontology Accelerator is a self-managed, open source accelerator (Apache 2.0) that helps customers build user-defined ontologies to make agent decisions more accurate, explainable, and auditable. It uses AI-assisted induction (with human review) to accelerate customers' creation of ontologies that are specific to their entities, relationships, rules, and constraints. Customers review and edit each proposed ontology before they accept it, and use the result across their systems, built on open W3C standards (OWL, RDF, SHACL). The ontology is stored in a customer-owned knowledge graph that any agent can consume through Context Ontology Accelerator MCP.
Get Started
How it works →
RESOLUTION TRACE
risk-analyst · regulatory-agent
"Give me the Solvency II ratio for in-force commercial property policies as of last quarter-end, excluding cat-exposed lines."
SELECT eligible_own_funds / scr AS ratio
FROM quarterly_capital
WHERE line_of_business IN (:policy.commercial_property)
AND reporting_date = :period.q_end
AND excluded_flags ≥ (:cat_exposed)
SOURCE DynamoDB · pre-compiled
ONTOLOGY ReportingPeriod · LinedOfBusiness
LLM 0 calls
LATENCY 14 ms
2 SPARQL over VKG not reached DEFERRED
3 Agentic fallback not reached DEFERRED
ANSWER RESOLVED
provenance · quarterly_capital.v4 · deterministic
187.4%
SOLVENCY II · COMMERCIAL PROPERTY · Q ENDING 2026-03-31
Scenarios are illustrative and do not constitute regulatory advice; outputs should be independently validated before use in regulatory filings or compliance reporting. Latency and LLM-call figures shown reflect a pre-compiled governed metric — actual performance varies by query complexity, data volume, and infrastructure configuration, and not all query types are eligible for pre-compiled resolution.
STANDARDS
RDF · OWL 2 · SHACL · R2RML · SPARQL 1.1
INTERFACES
MCP server · REST · SPARQL endpoint









CDC

Broker

Processing

Streamhouse



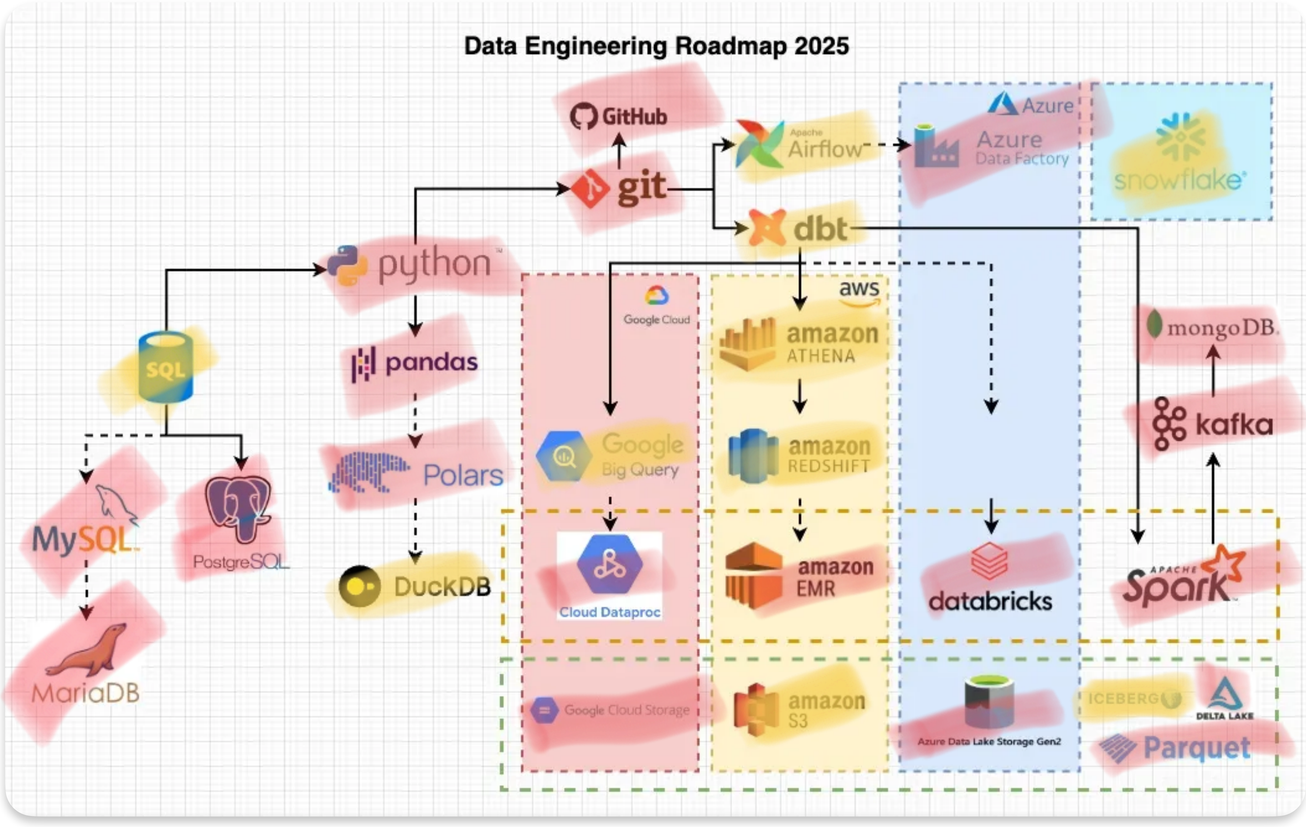
Streaming (L6)

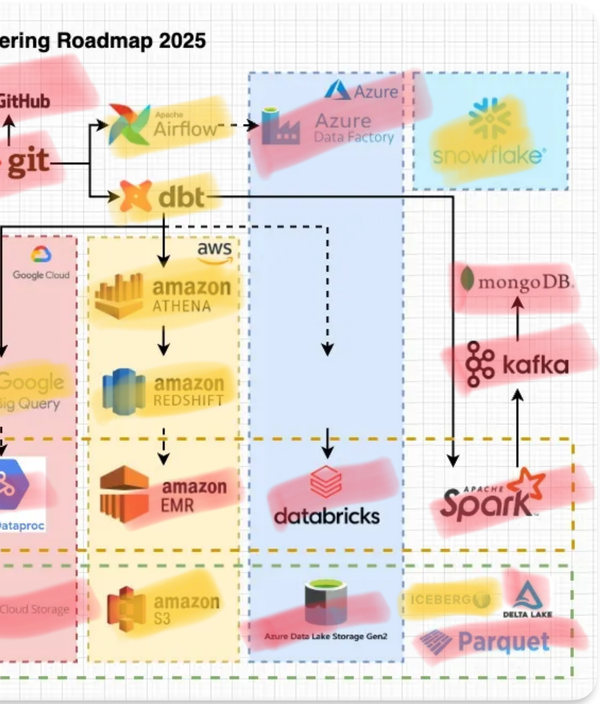




drails

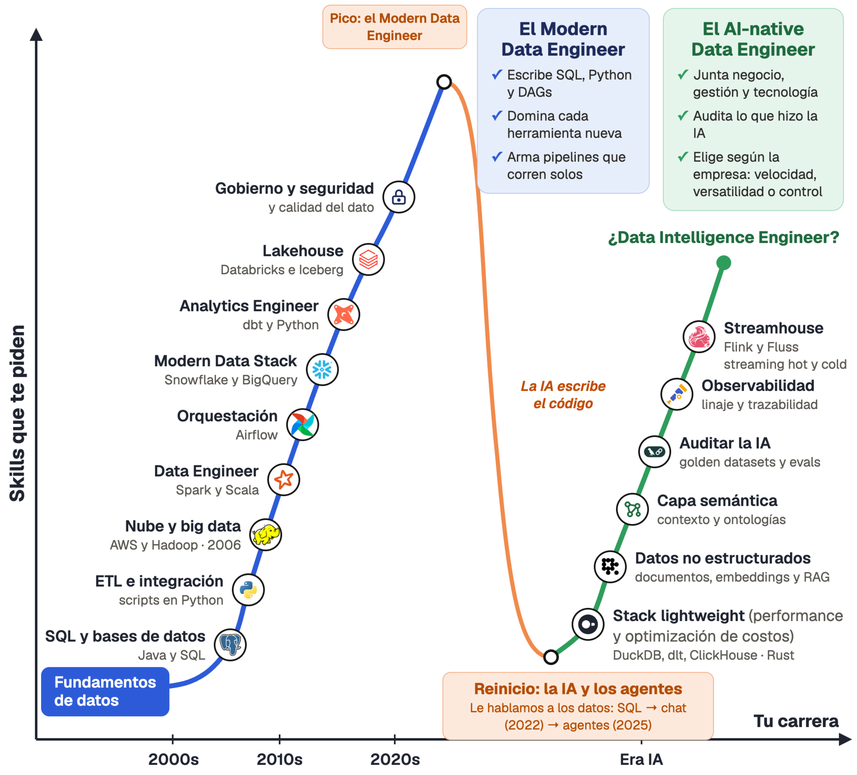
PR Review





Evolución de un Data Engineer

Cada herramienta le dio forma a un rol. Y con la IA, la curva se reinicia



El valor pasa de programar a juntar negocio, gestión y tecnología.

Los fundamentos siguen. Lo que cambia es qué perfil busca cada empresa: la startup paga velocidad, la pyme versatilidad y la corpo control.

HOY Masterclass gratis · jue 19/11 · 19 h ARG · trubidata.com/es/masterclass trubi data

