

# Usos de bases NoSQL en Feature Stores

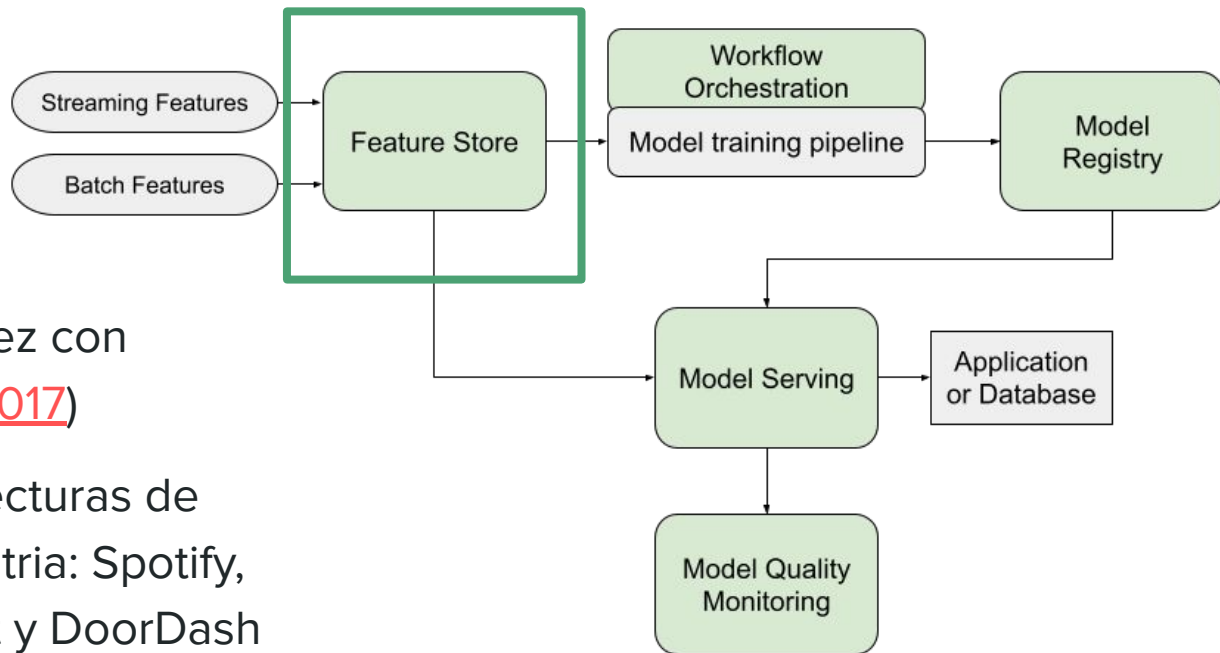
---

# Objetivos

## ¿Feature Stores? ↔ ¿NoSQL?

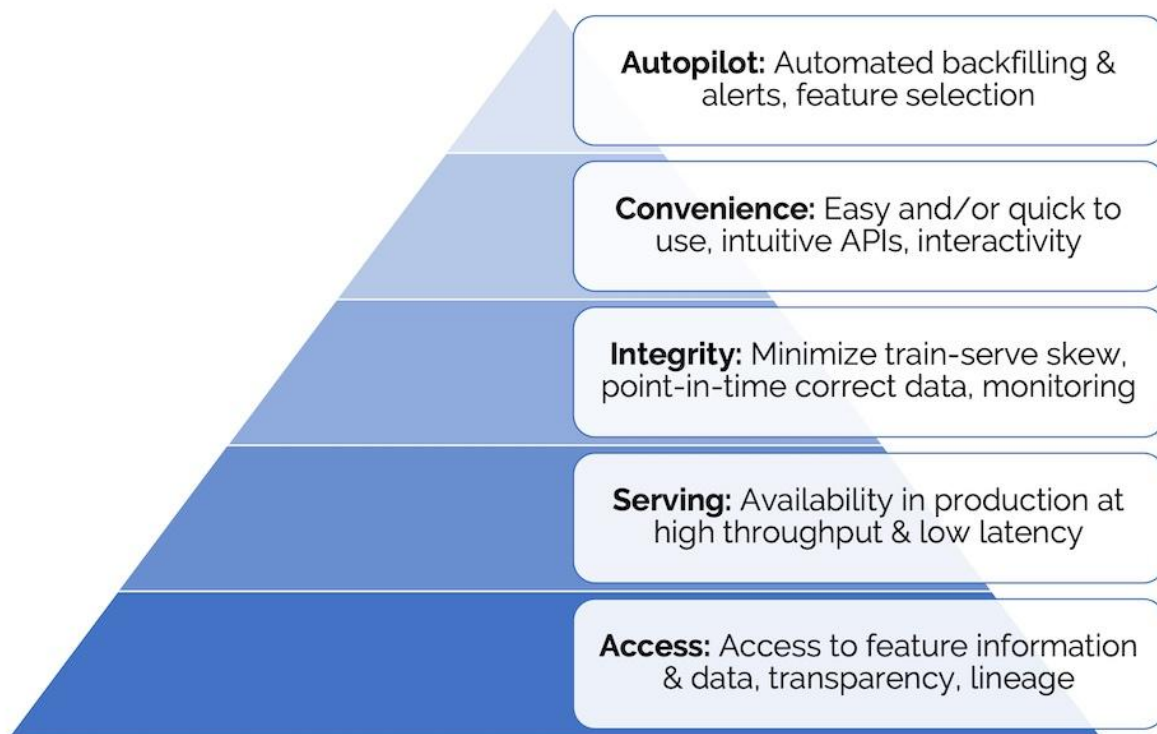
- Caracterizar Feature Stores
  - ¿Por qué surgen? ¿Qué problemas resuelven?
  - ¿Productos disponibles?
- Entender el uso de bases de datos NoSQL en este contexto
  - ¿Para que se usan? ¿Casos de uso dan soporte?
  - ¿Qué tipos de bases de datos NoSQL se utilizan? ¿Que productos?

# Feature Stores - Inicios

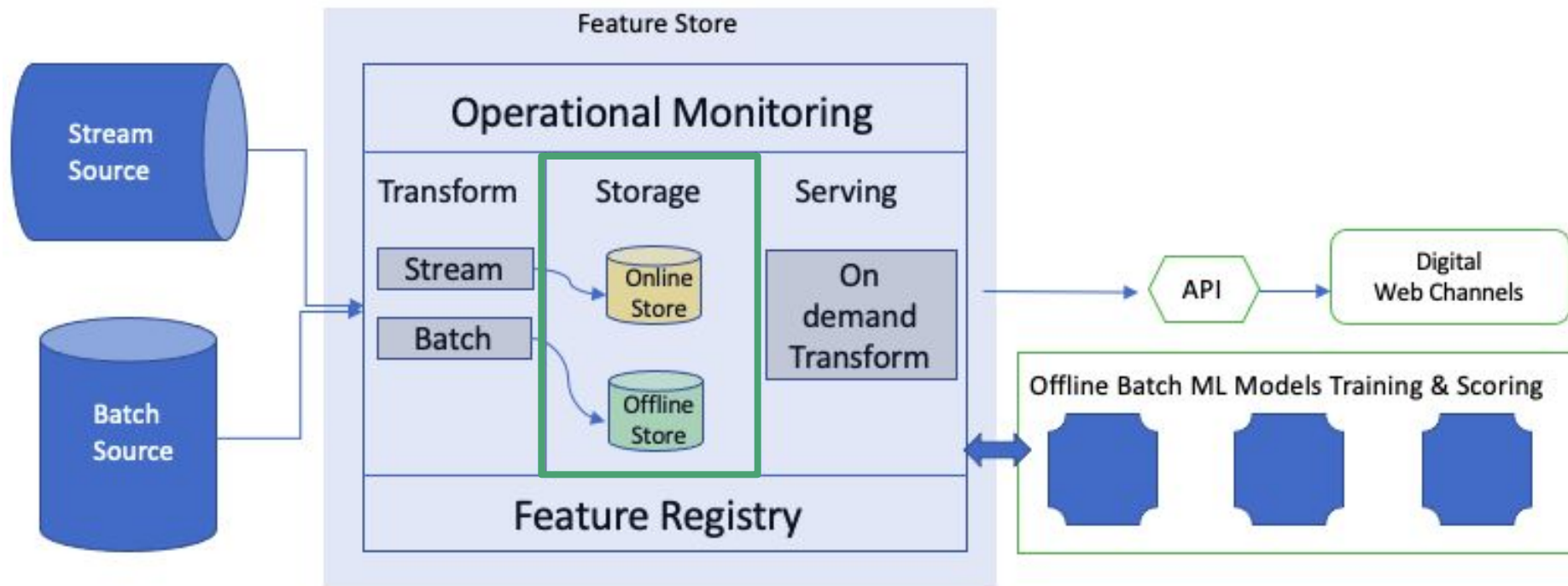


- Surgen por primera vez con Michelangelo ([Uber 2017](#))
- Se replican en arquitecturas de referentes de la industria: Spotify, Etsy, Paypal, Pinterest y DoorDash

# Feature Stores - Capacidades

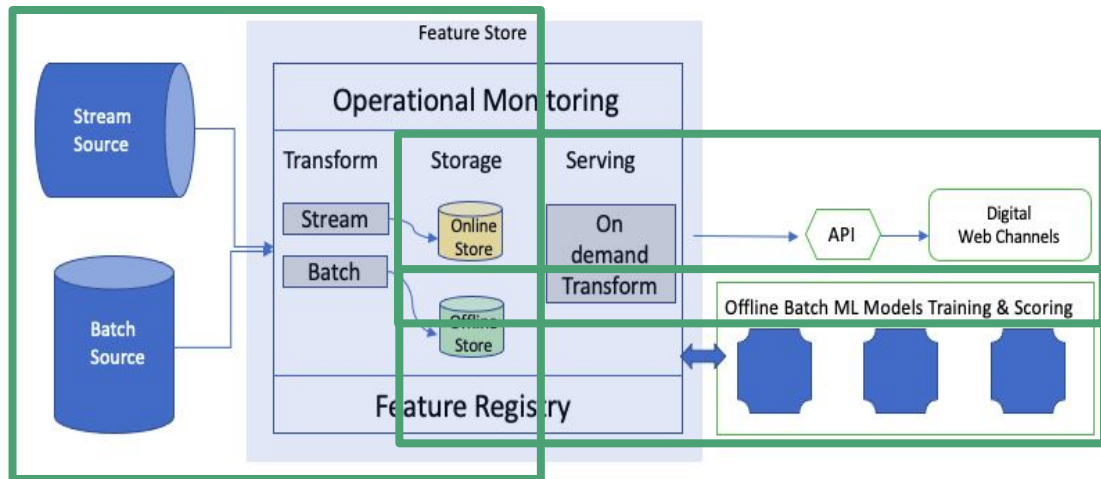


# Feature Stores - Componentes



# Uso de bases NoSQL → Storage de features

- Actualizar features
- Obtener features para training (offline)
- Obtener features para inferencia (online)



# Uso de bases NoSQL - Tipos y productos

## Repositorios Offline

- Soluciones de big data
- Bases de datos columnares

## Repositorios Online

- Key Value Stores
- Bases de datos columnares
- Bases de datos en memoria

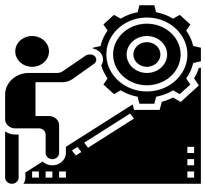
Plataforma	Offline	Online	Feature Eng
Hopsworks <sup>a</sup> (AGPL-V3)	Hudi/Hive	MySQL	PySpark, Python
Feast <sup>b</sup> (Apache V2)	BigQuery	BigTable, Redis	Beam, Python
Tecton <sup>c</sup> (Comercial)	Delta Lake	DynamoDB	SQL, PySpark, Python
StreamSQL <sup>d</sup> (Comercial)	S3	Cassandra, Redis, Coc- kroachDB	SQL, Python
Michelangelo <sup>e</sup> (Uber)	Hive	Cassandra	Spark, DSL
Conde Nast <sup>f</sup>	Kafka / Cas- sandra	Kafka / Cas- sandra	Librerías compartidas
Zipline <sup>g</sup> (Airbnb)	Hive	KV Store	Flink, Spark, DSL
Comcast <sup>h</sup> (NBC)	HDFS, Cas- sandra	Kafka / Re- dis	Flink, Spark
Metaflow <sup>i</sup> (Netflix)	Kafka, S3	Kafka, Mi- croservices	Spark, Librerías compartidas
Twitter <sup>j</sup>	HDFS	Strato / Manhattan	Scala DSL, Scalding
Pinterest Galaxy <sup>k</sup>	S3/Hive	?	DSL (Linchpin), Spark
Wix <sup>l</sup>	S3	HBase/Redis	SparkSQL

# Uso de bases NoSQL - Ventajas y Desafíos



Offline vs Online facilita uso de bases NoSQL

- Bases NoSQL específicas para cada caso
- Soporte para casos de uso (baja latencia, alto volumen de datos)



Consistencia

- Entre repositorios Offline y Online
- Temporal



# Conclusiones

- Feature Stores componentes clave en soluciones de ML
- Bases NoSQL claves para resolver repositorios Offline/Online
- Trabajo futuro
  - Probar Feature Stores
  - Comparar alternativas de bases NoSQL disponibles

# Referencias

- [Feature Store for ML](#)
- [Feature Store for ML - Docs](#)
- [Feature Stores for ML – Medium](#)
- [Lessons on ML Platforms — from Netflix, DoorDash, Spotify, and more](#)
- [Feature Stores - A Hierarchy of Needs](#)
- [5 Minimum Requirements of an Operational Feature Store](#)
- [List of Feature Store Architectures for Data Scientists and Big Data](#)
- [Meet Michelangelo: Uber's Machine Learning Platform](#)
- [Edge#77: How Feature Stores Were Started](#)
- [2021 — A year of ML Feature Stores Adoption? | by Shrey Shivam | Feature Stores for ML](#)

# ¿Preguntas?

# Gracias!