

Feature Store comparison: 4 Feature Stores - explained and compared.

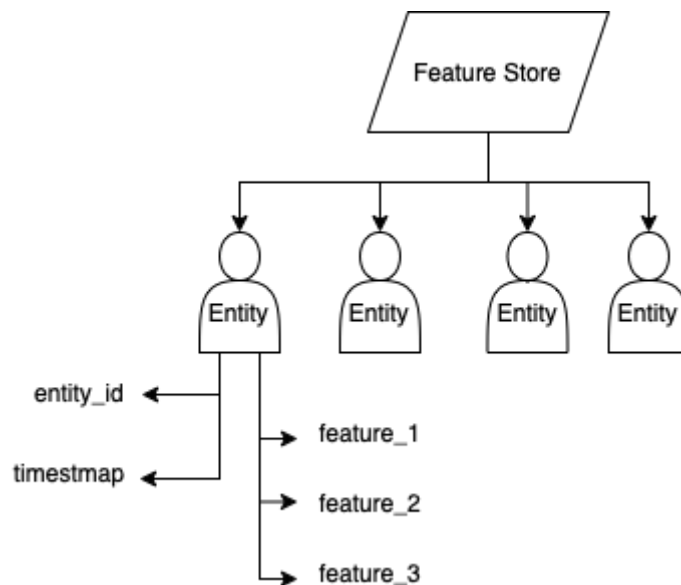
Tags	Feature Store comparison, Feature Store, MLOps, Machine Learning, Vertex AI Feature Store, FEAST Feature Store, AWS SageMaker Feature Store , Amazon SageMaker, Databricks Feature Store
Meta Description	A simple and clear comparison of 4 popular feature stores: Vertex AI Feature Store, FEAST, AWS SageMaker Feature Store and Databricks Feature Store
Intro KB	A simple and clear comparison of 4 popular feature stores: Vertex AI Feature Store, FEAST, AWS SageMaker Feature Store and Databricks Feature Store on one refcart.
Autor	Jakub Jurczak
LINK do grafik	

In this blog post, we will simply and clearly demonstrate the difference between 4 popular feature stores: **Vertex AI Feature Store, FEAST, AWS SageMaker Feature Store and Databricks Feature Store**. Their functions, capabilities and specifics will be compared on one refcart. Which feature store should you choose for your specific project needs? This comparison will make this decision much easier. But first:

Feature Store explained: What is a Feature Store?

A feature store is a data storage facility that enables you to keep features, labels, and metadata together in one place. We can use a feature store for training models and serving predictions in the production environment. Each feature is stored along with metadata information. **This is extremely helpful when working on a project, as every change can be tracked from start to finish, and each feature can be quickly recovered if needed.**

Before we go any further, let's look at the Feature Store data model in the diagram below.



A Feature Store contains the set of entities of a specified entity time. Each entity type defines fields like "entity_id", "timestamp" and a list of features like "feature_1", "feature_2" and so on.

So, we can think of a Feature Store as a centralized set of entities from the whole organization:

- **Business teams provide high-level business metrics with no noise or bias from low-level data.** For example, you don't want to build your fraud detection engine on data biased by the fraudulent activity of users.
- **Data scientists are interested in entities representing high-quality features to train their machine learning models.** Most of the time, these features are not business metrics but rather very granular values computed from the raw data of your application (for example, how many times the user X logged in within the last hour). These high-quality features are computationally expensive to derive and hard to maintain. The last thing you want is to have every machine learning model recomputing those features at each run.

The machine learning platform needs to access those features at scale when running your models in production.

The Feature Store can solve business problems, which I mentioned in this article: [MLOps 5 Machine Learning problems resulting in ineffective use of data.](#)

Still, before that, I would like to briefly introduce the solutions available on the market.

Feature Store compared

Below in the refcart, you will find a very specific comparison of the basic differences of the four most popular Feature Stores: **Vertex AI Feature Store**, **FEAST**, **AWS SageMaker Feature Store** and **Databricks Feature Store**.

Vertex AI Feature Store	Feast	AWS Sage Maker Feature Store	Databricks Feature Store

<ul style="list-style-type: none"> • The managed platform is offered by GCP. • Currently supports only BigQuery and GCS as a features source. • Offers features monitoring out of the box. • Highly scalable features serving service on demand. • Offers join the operation between entities. • Currently doesn't offer post-processing features and point-in-time joins between entities. • Web UI. 	<ul style="list-style-type: none"> • Open Source. • Requires manual installation. • Currently doesn't offer features monitoring. • Offers point-in-time joins between entities. • Requires Kubernetes, DB and Cache set up for highly scalable features serving. • Offers Feast CLI, which can cache features on your local machine. • It provides simple features transformation, but you can't treat it as an ETL/ELT tool. 	<ul style="list-style-type: none"> • AWS offers a managed platform. • Enables the user to store, discover and share features to all users of the Sage Maker platform. • Supports data ingestions from batches and streams. • The platform offers offline and online access to features. • Batch and online features serving. 	<ul style="list-style-type: none"> • Databricks offers a managed platform. • Features monitoring. • Batch and online features serving. • Notebooks manage feature ingestion jobs. • Support for the Spark framework.
--	--	---	---

- | | | | |
|--|--|--|--|
| | <ul style="list-style-type: none">● Offers connectors for Snowflake, Redshift, BigQuery, Parquet files, Synapse, Hive, Postgres, Spark, Kafka, Trino, Pandas, DynamoDB, Redis, Datastore, SQLite, Kinesis.● Web UI is experimental. | | |
|--|--|--|--|

An internal feature store to manage and deploy features across different machine learning systems is key practice for MLOps. Feature stores help develop, deploy, manage, and monitor machine learning models. It allows you to improve the development lifecycle of your model and the flexibility and scalability of machine learning infrastructure. You can also use the feature store to provide a unified interface for access to features across different environments, such as training and serving.

We are in the process of completing the release of an ebook that will show you specifically step-by-step, **how to build a feature store from scratch by using the Vertex AI platform, and how to resolve business problems that can occur in the Machine Learning process**. We will also point out the differences between BigQuery and Snowflake, a cloud-native data warehouse. Furthermore, we will demonstrate how to use dbt to build highly scalable ELT pipelines in minutes.

Release coming soon. To make sure you don't miss out, sign up for our newsletter to be notified when the ebook is ready.

If you have any questions or concerns in the area of Machine Learning and MLOps we encourage you to contact us. We have experience in the implementation and optimization of Machine Learning processes. We also develop original solutions in niche areas. We will be happy to serve you with our expertise.