FEATURE STORE
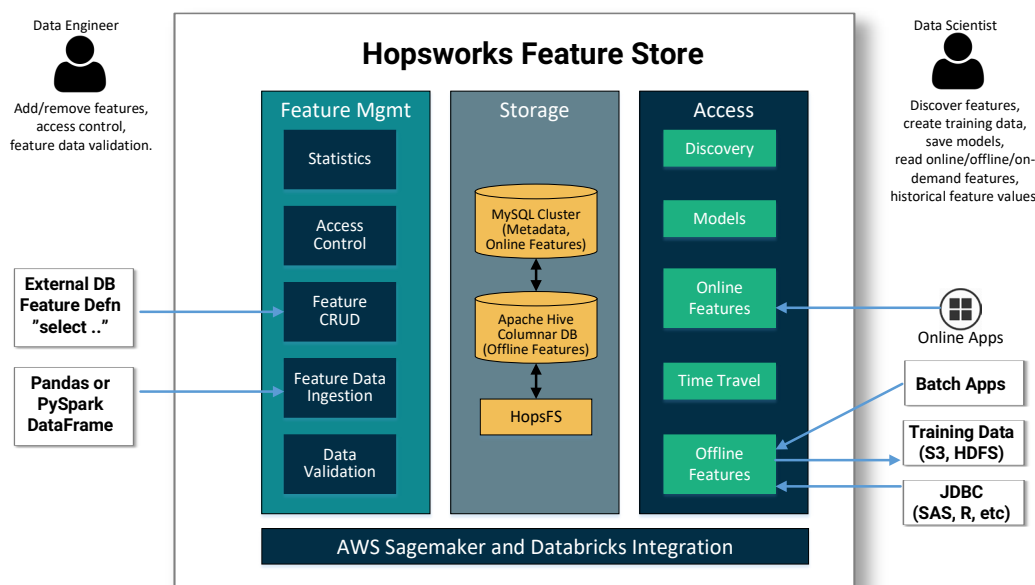
LOGICAL CLOCKS

Hopsworks Feature Store is a data warehouse for your ML features. The Feature Store acts as an API between Data Engineering and Data Science, enabling improved collaboration between Data Engineers, who engineer the features, with Data Scientists, who use the features to train models. The Feature Store enables features to be registered, discovered, and used as part of ML pipelines, thus making it easier to transform and validate the training data that is fed into machine learning systems.

For Data Scientists

The Feature Store frees Data Scientists from writing Feature Pipelines that pull and transform data coming from backend systems. Instead, Data Scientists browse the Feature Store to find features to use in their models, and then generate clean, validated Train/ Test data for those features in the file format of their choice (.tfrecords for TensorFlow, .numpy for PyTorch, .csv for Scikit-Learn, etc). Data Scientists can also inspect Features visually to understand their data distributions, preview the feature data, quickly see the max/min/mean values for features.The Feature Store can be used from Jupyter.

For Data Engineers

Data Engineers typically have responsibility for adding new features. A Feature could be computed by anything from a simple SQL query on an external datastore to complex graph embeddings computed from large amounts of graph data. Features can be created either in notebooks or programs (Python/Scala/ Java APIs) or in the Hopsworks UI (on-demand features using SQL queries on external databases). Hopsworks Feature Store supports access control, and versioning for both features and training datasets. Feature data can be validated before ingestion using our Data Validation API.

For ML Engineers

Hopsworks Feature Store supports time travel queries to lookup the value of a feature at a point in time in the past (used to generate new training data for late observations). Hopsworks Online and Offline Feature Store are both horizontally scalable and highly available. Hopsworks Feature store enables more flexible ML pipelines, breaking them up into feature engineering pipelines that ingest feature data into the Feature Store, and training pipelines after the Feature Store to deploy models into production. Connectors to HDFS, S3, and external databases are supported.

Silicon Valley Office

470 Ramona St
Palo Alto,
94301 California
USA

Stockholm Office

Box 1263,
Isafjordsgatan 22
164 40, Kista
Sweden

UK Office

IDEALondon,
69 Wilson St
EC2A2BB, London
UK

logicalclocks.com