Nexus Feature Store powering Disney Magic

Dustin Hamerla, Manager ML Engineering, Disney Streaming

FEATURE STORE SUMMIT 2022

Organized by HOPSWORKS

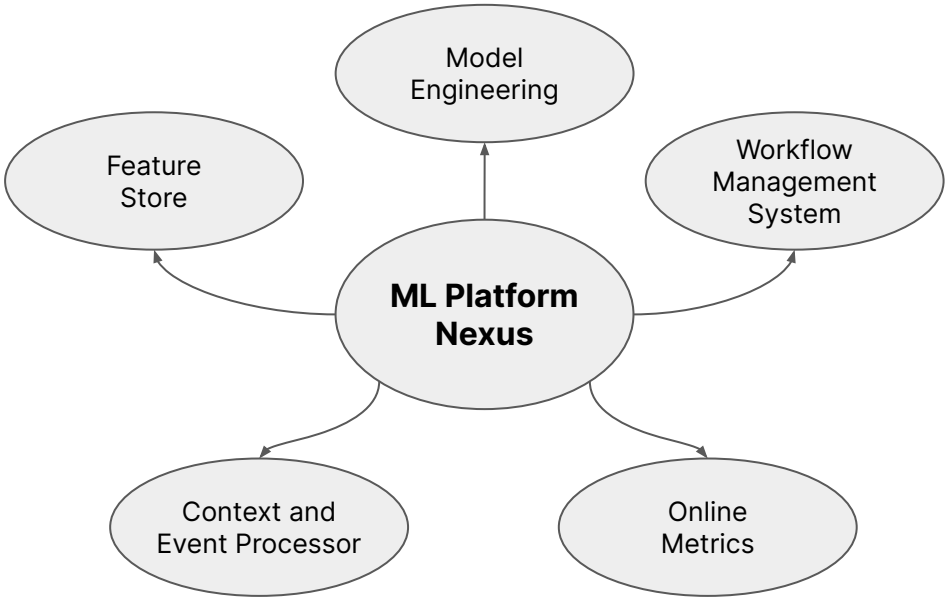
Agenda

- Nexus Feature Store
  - Overview & Motivation
  - Capabilities
  - Architecture

Overview & Motivation

ML Platform overview

Democratize & accelerate scaleable ML development across Disney Streaming

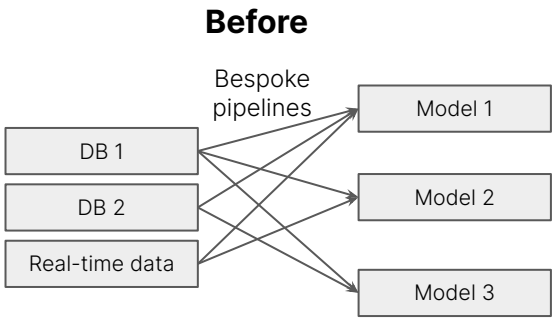


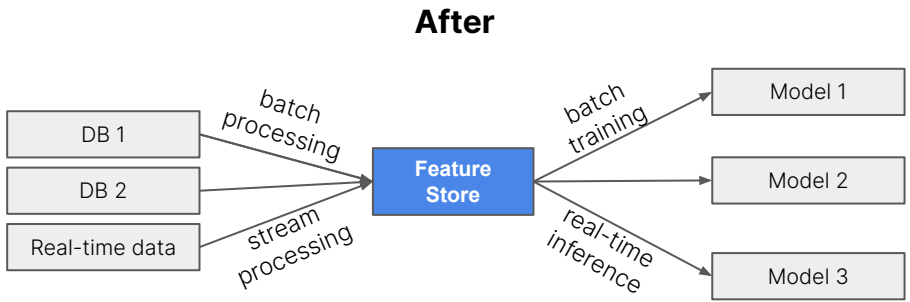
Why do we need a Feature Store?

Improve feature creation (Data Engineers) and

feature consumption (Data Scientists) for various ML use cases

- **Break-up bespoke** pipelines
- **Decouple** feature engineering from model development
- Support **various computing environments**
- **Discover & share** features



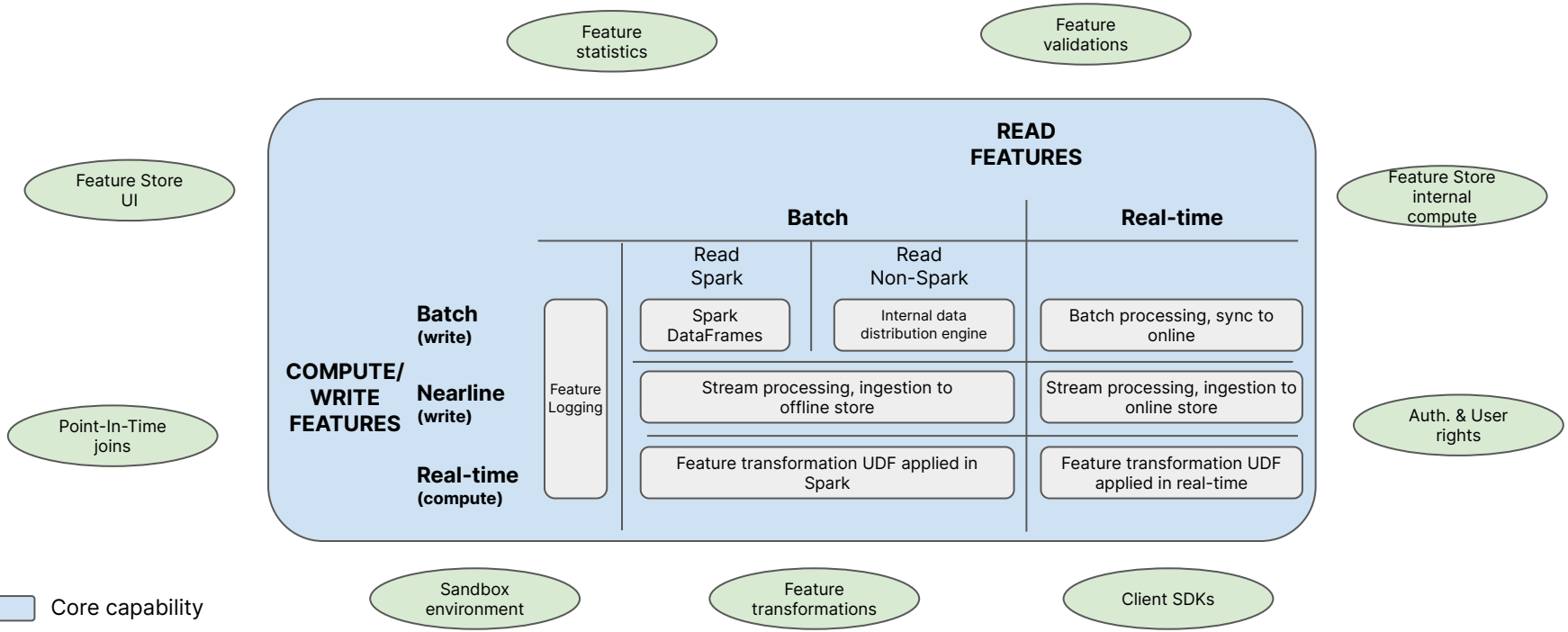


Why do we need to build our own Feature Store?

- Global scale batch training & scoring
  - **Region-specific** data
  - Focus on **experimentation**
  - Feeding **multiple datasets** – not just “one training dataset”
  - Bridge between **Spark based dataprep** and **single-node training & scoring**
- Customized data model
  - **Compatibility** with larger ML platform
- Existing ML-platform infrastructure and teams

Nexus Feature Store Capabilities

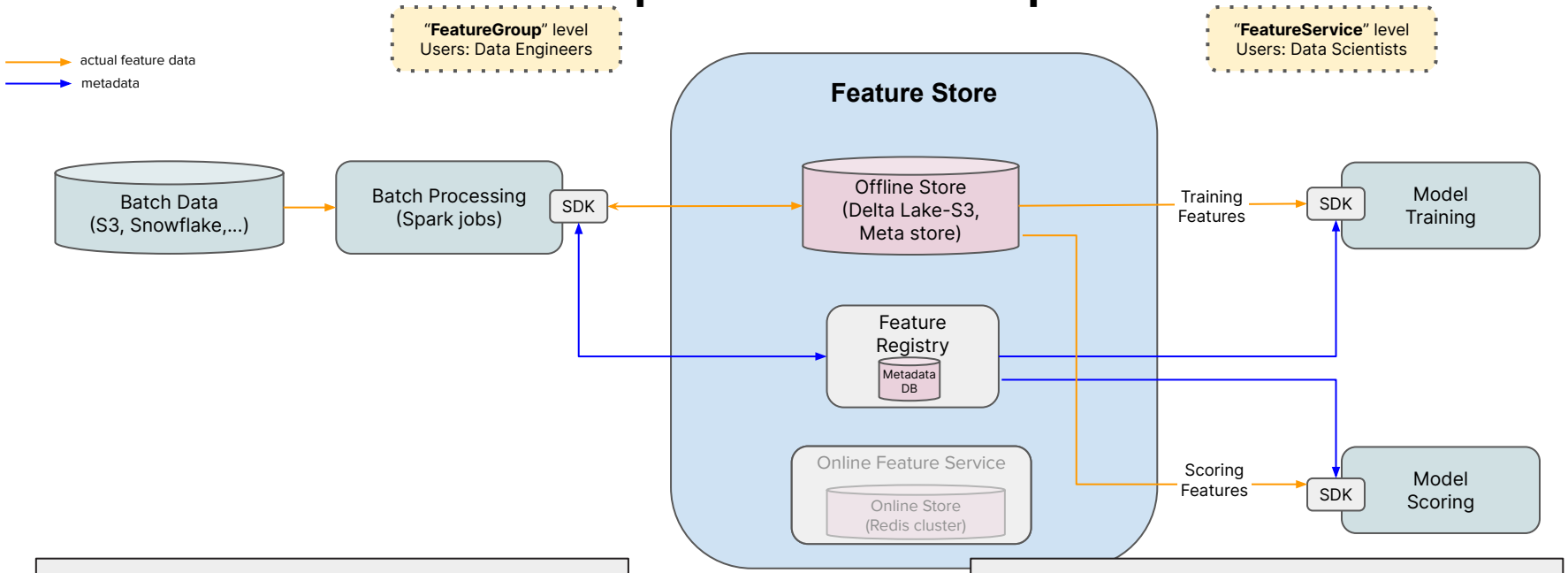
Nexus Feature Store capability map



Nexus Feature Store Architecture

FEATURE STORE SUMMIT 2022

Nexus FS Architecture - Spark write & Non-Spark read



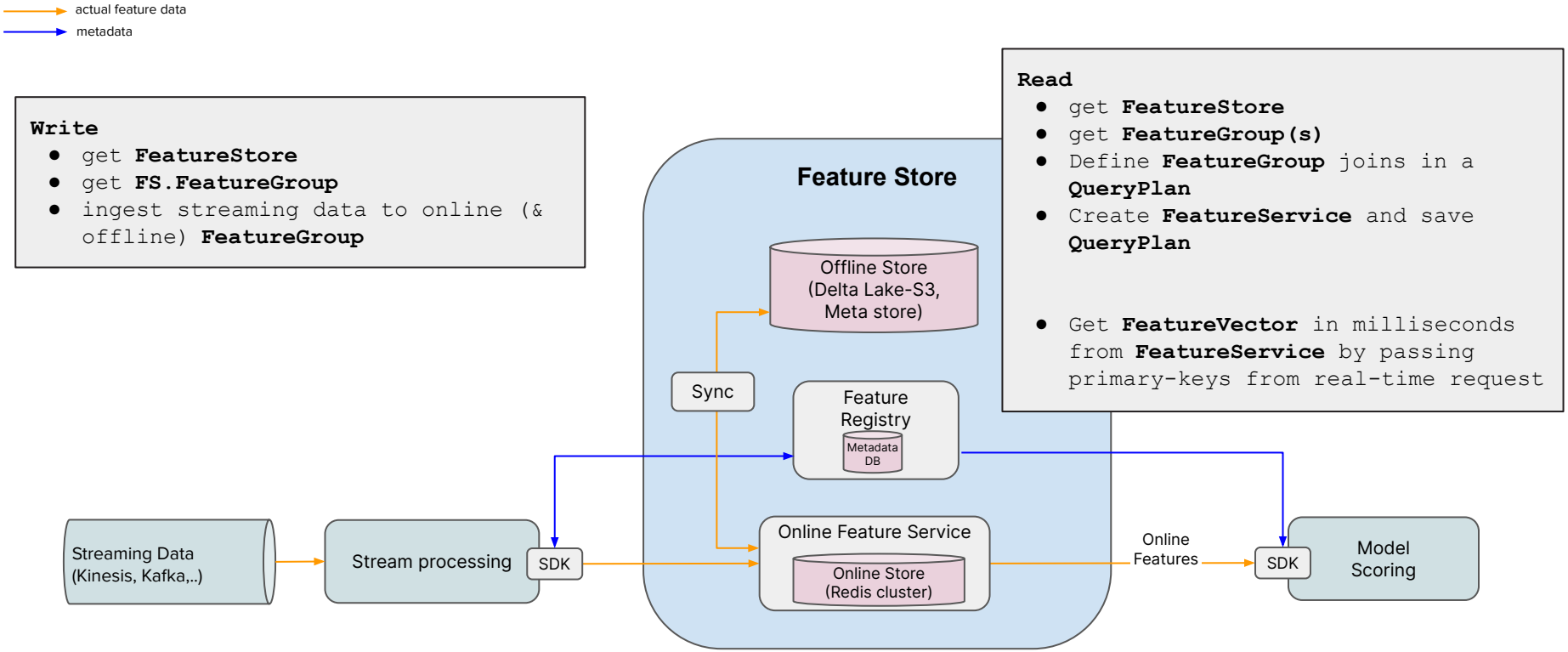
Write

- get **FeatureStore**
- get **FS.FeatureGroup(s)**
- create **FS.FeatureService**
- ingest data from one or many **FeatureGroups** to a **FeatureService**

Read

- get **FS.FeatureStore**
- get **FS.FeatureSerivce**
- load subsets of data from **FeatureService** (e.g sub-sampled, sub-partitioned etc.)

Nexus FS Architecture - Streaming write & real-time read



Thank you!