

Supercharging Intuit Devs with Feature Store Service

Vijay Kulkarni, Engineering Manager, Intuit
Harish Nagu Sana, Software Engineer, Intuit

INTUIT

 turbotax  creditkarma  quickbooks  mailchimp

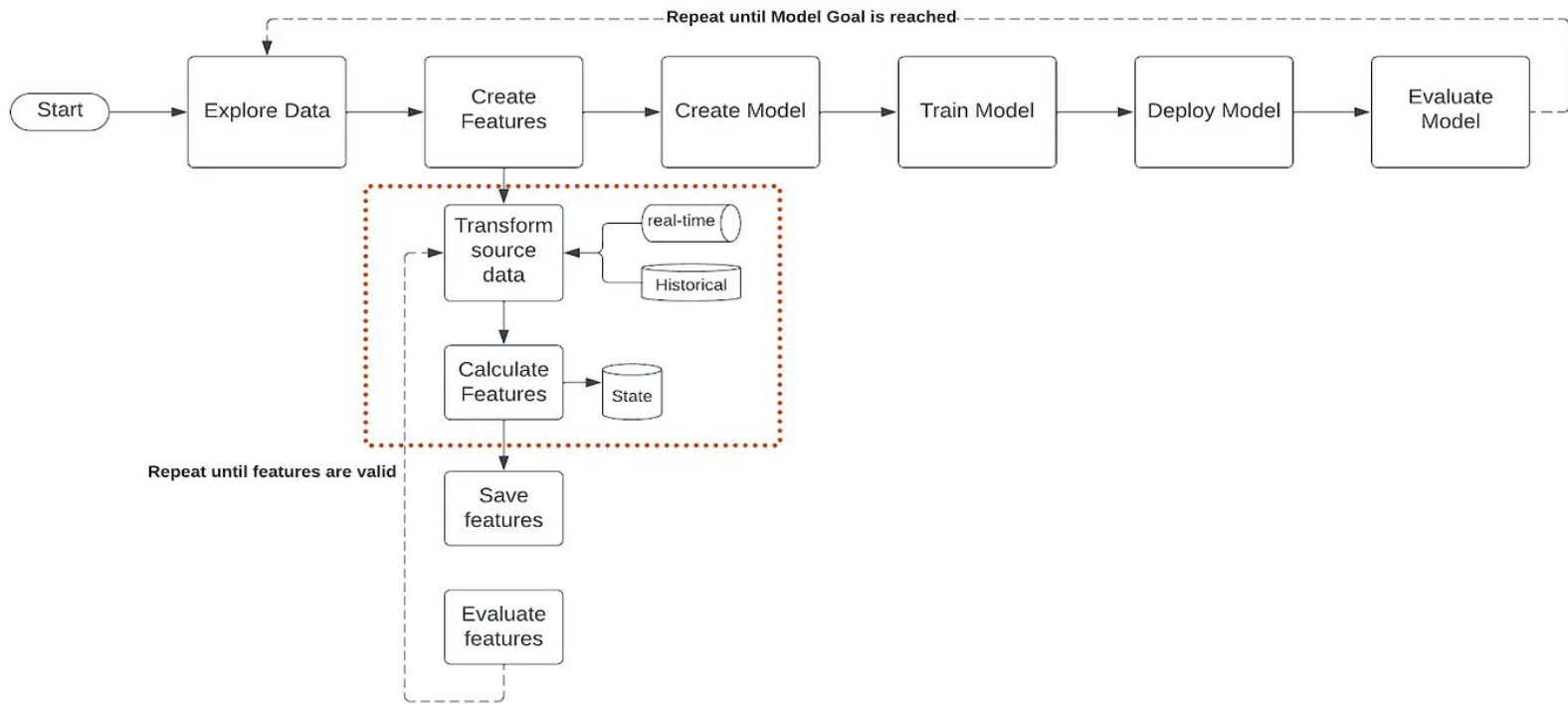
Organized by  **HOPSWORKS**

FEATURE STORE SUMMIT 2024

DATA FOR AI:
REAL-TIME, BATCH, AND LLMS

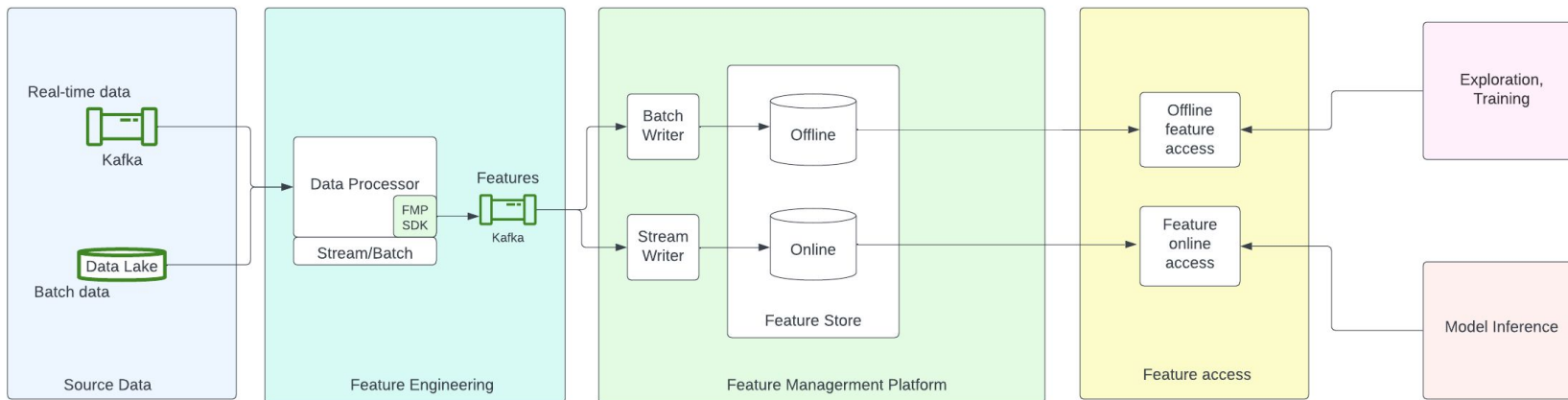


ML Model Development Lifecycle





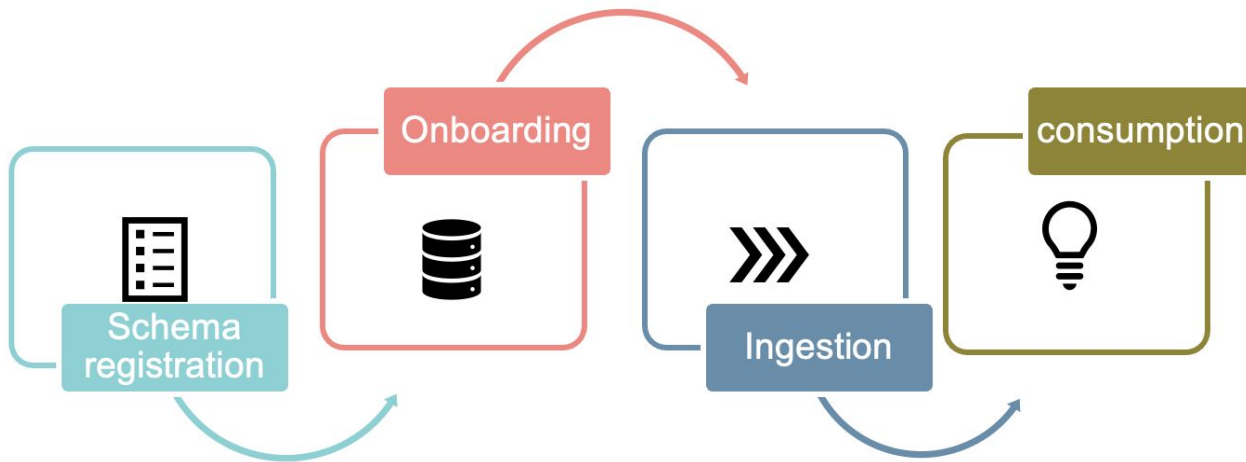
Feature Management Platform





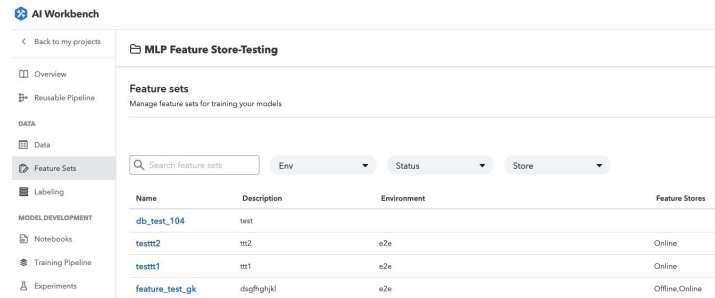


Streamlined Feature Management Lifecycle



Feature store console on AI Workbench

- **AI Workbench:** A new dedicated development environment for end-to-end AI/ML application development
- **Real-Time Status:** Stay informed with live pipeline and infrastructure updates.
- **Effortless Feature Management:** Add or update features
- **Seamless Deployment:** Test in pre-prod, promote to prod seamlessly.
- **Self-Service Simplicity:** Register, onboard, and manage features in minutes.
- **Rapid Start:** Get up and running quickly, no manual intervention needed!





Schema Registration

- **Define Feature schema:** Schema is stored in a central repository. Framework validates published features against predefined schema
- **Manage Features:** Manage feature metadata, add/update features
- **Discovery:** Once registered, the feature sets can be looked up in Data Catalog service. This is key to make sure features are discoverable and reusable.
- **Reuse:** Typically, users start with exploring existing feature sets in the offline store to discover and re-use any features that they find fit for their models.

Create New Feature Set

Basic Information

Essential details and settings for creating a feature set.

🕒 Feature set details

🕒 Define features

Onboard to Feature Store

Add feature set to the Feature Store for model fetching

1 Store inputs

2 Read/Write Configuration

Feature definition

Specify the features for the feature set

Define Features

```
[
  {
    "name": "feature_name1",
    "type": "INTEGER",
    "description": "Some description",
    "dataClassification": "PUBLIC",
    "isNullable": true
  },
  {
    "name": "feature_name2",
    "type": "INTEGER",
    "description": "Some description",
    "dataClassification": "PUBLIC",
    "isNullable": true
  }
]
```

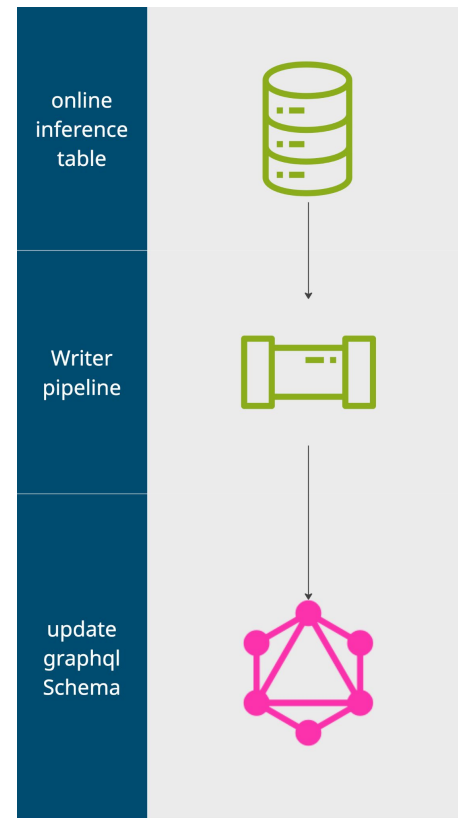


Onboarding

- **Online Inference:** An online inference table will be provisioned to enable real-time data processing.
- **Data Pipeline Integration:** A dedicated data pipeline will be implemented to listen to the feature stream and efficiently insert data into the online inference table.
- **Offline Job:** The existing offline job will be updated to retrieve data from the feature stream and generate an S3 Hive for offline training.
- **GraphQL Retrieval:** Data retrieval will be facilitated through a GraphQL service, ensuring efficient access and schema updates for corresponding features.

Benefits:

- **Speed & Efficiency:** Fast data retrieval and insertion processes enable real-time data processing capabilities.
- **Data Accuracy & Consistency:** Improved data accuracy and consistency are ensured through streamlined processes and real-time updates.





Ingestion

- **SDK Provisioning:** Users will be provided with an SDK to inject the data, making the process simple and unique for publishing data.
- **SDK Options:** Users will be provided with both Java and Python SDKs to accommodate their preferred programming language.
- **Feature Engineering:** Users can perform feature engineering and convert the rows into the SDK format, allowing for customization and flexibility.
- **Pipeline Processing:** The pipeline will understand and process the data, ensuring seamless integration and efficient data handling.





Consumption

- **Online inference** : A GraphQL service created to retrieve the data. Integrated with model execution platform
 - **Low Latency** : Real-time feature inference for low-latency applications, ensuring timely decision-making.
 - **Auto Scaling** : Automatic scaling to handle changing workloads, ensuring high performance and reliability.
 - **Integration with Model Execution Service** : Integration with model execution services, enabling streamlined model deployment and inference.
 - **Parallel Read Capability** : Ability to read features in parallel, reducing latency and increasing throughput.
- **Offline store**:
Users can access and retrieve data from Amazon S3, Apache Hive to train their machine learning models





Supercharge ML Model Development

- **Faster Onboarding:** AI Workbench provides a self-serve environment for users to quickly get started with feature management.
- **Increased Productivity:** The platform promotes feature discoverability and reuse, saving developers time and effort.
- **Simplified Data Publishing:** SDKs for Java and Python streamline the process of publishing data to the feature store.
- **Flexible Feature Serving:** The platform supports both real-time and offline feature serving, catering to diverse application needs.
- **Scalable and Performant:** The platform is highly scalable and ensures low-latency feature retrieval for real-time model inference.
- **Built-in Security and Reliability:** Security and reliability are integral to the platform's design, safeguarding valuable data and ensuring consistent performance.

Q&A

Thank you

INTUIT

✓ turbotax

ck creditkarma

qb quickbooks

mailchimp



FEATURE STORE SUMMIT 2024

DATA FOR AI:
REAL-TIME, BATCH, AND LLMS