databricks

Service Description

MLOps Optimized

This Service Description describes the scope of services (including associated Customer requirements) for the Databricks Advisory Service identified above (the "**Service**"), and applies to the Service under Customer's accepted Service Order.

Service Overview

Accelerate time to production for your machine learning pipelines by developing a repeatable, production-grade MLOps framework that enables MLOps best practices at scale.

Objective

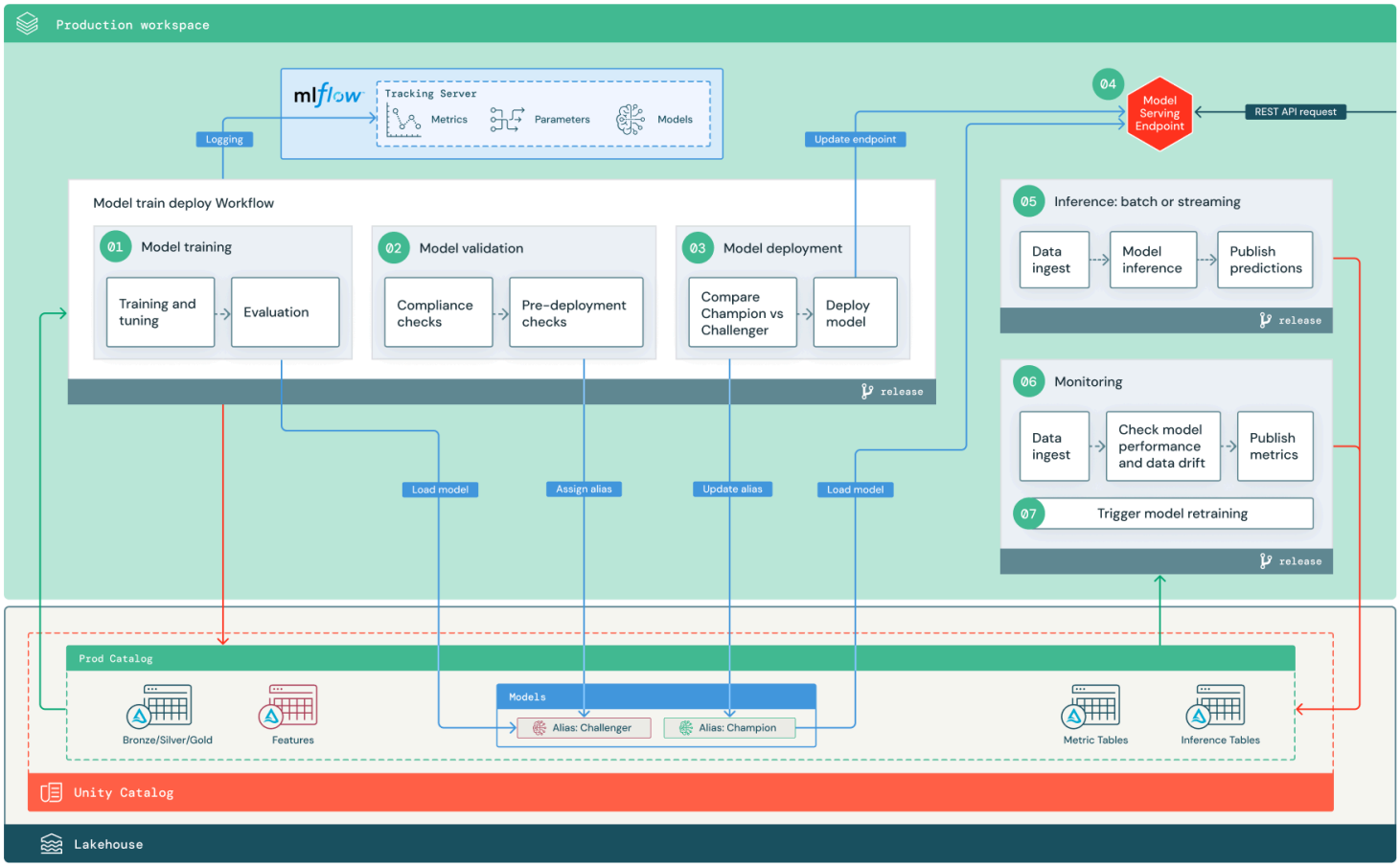
Develop a production-ready repository set up with infrastructure as code, CI/CD workflows, and referenceable ML pipeline, customized for your use case. This includes developing an end-to-end ML pipeline production architecture, setting up batch, streaming and/or real-time inference, and establishing data and model governance with Databricks Unity Catalog and MLflow.

Description of Services

Databricks will provide Services from the Technical Focus Areas and Representative Activities described below in assisting Customer on MLOps. **Specific activities performed will vary, depending on Customer-specific objectives.**

Technical Focus Area	Representative Activities
MLOps Strategy *Develop a comprehensive approach for MLOps*	• Kickoff project • Develop end-to-end ML pipeline production architecture (e.g. promote code or promote model approaches, inference requirements (batch, streaming, or real time), technology choices, etc.) • Configure MLOps Stacks for Customer's given ML use case, including proper permissions, etc.
MLOps Development *Automated, reusable CI/CD process for ML pipelines*	• Set up model lifecycle management with Databricks Unity Catalog and MLflow • Apply and customize MLOps Stacks to perform CI/CD with existing ML code • Set up batch, streaming, or real-time inference • Integrate solution with Unity Catalog • Configure Lakehouse Monitoring (optional)
Knowledge Transfer *Next steps*	• Document solution • Enable team and conduct knowledge transfer with a focus on system maintenance and extending to future ML use cases

The visual below illustrates a reference high-level MLOps production workflow.



Prerequisites

Throughout the engagement, Customer will assure that the following requirements are met to enable the Services:

- Customer must (before the start of the engagement) have developed an ML model/pipeline on Databricks
- Customer must have obtained licensed access to necessary Databricks Platform features, including enrolling in applicable preview features if not generally available yet, and meet all the specified prerequisites
- Customer must be willing to follow either promote model or promote code paradigm. Implementing both promote model and promote code will require additional implementation time.
- During the MLOps Stacks setup, Customer's infrastructure team must be made available to ensure proper permissions are granted
- Customer's technical, business, and domain experts to be reasonably available throughout the engagement to answer questions and provide necessary context
- Customer's technical resources who will own the pipelines after this engagement is complete for appropriate knowledge transfer
- Customer to provide access to Databricks environment (staging, development, and production), CI/CD tooling (eg Github Actions or Azure DevOps), data, information, and artifacts necessary to successfully complete the project

Out of Scope

- Large Language Model (LLM) deployment such as vector stores, especially large models, or LLM-specific optimizations can only be included as time permits
- Any modifications (eg hyperparameter tuning, feature engineering) to the existing ML solution
- Configuration & integration of non-Databricks products & systems
- Data cleansing associated with building broader data lake
- ETL non-related to ML
- Major modifications to the MLOps Stacks template

Resources and Schedule

Services consist of **up to 14 Days** of Data Scientist time, and up to **2 Days** of Project Management time, typically across a continuous **3-4** week period, applied against the Representative Activities in the Description of Services above.

Databricks will work with you to mutually agree to a project schedule as part of the Project Management phase. Resourcing assignments require a minimum 4-weeks advance request (while Databricks makes reasonable efforts to accommodate scheduling requests, personnel availability is subject to Databricks resourcing and discretion). Accordingly, Databricks recommends Customer coordinate with Databricks Services at least a month before placing its Service Order.

Additional Definitions and Terms

- "**Agreement**" means your agreement with Databricks providing general terms for our Services.
- **"Day"** means 8 working hours during local business days, excluding holidays.
- "**Services Order**" may be any of these mutually-accepted formats placed under your Agreement: an Order, Success Credit redemption request, written statement of work, or similar document
- "**we**", "**us**" or "**our**" means Databricks, Inc. or its Affiliates.
- "**you**" or "**your**" means the Customer organization that placed the Services Order