

Machine Learning Package

Summary

The potential for machine learning (ML) and deep learning practitioners to make breakthroughs and drive positive outcomes is unprecedented. But how do you take advantage of the myriad data and ML tools now available? What's the most efficient way to streamline processes, speed up discovery and scale implementations for real-life scenarios?

Based on your use case and where you are in your data science journey, we can help you with the following:

1. Migrate an existing ML model from central/parallel computing models to distributed modeling on Apache Spark™, enabling data discovery, model training and inference at scale
2. Build ML solutions to address business problems by
 - Translating business problems to ML solutions
 - Performing exploratory data analysis and feature engineering
 - Building ML workflows customized to your needs
 - Enabling internal team with regard to data science and ML
3. Optimize machine learning pipelines
4. Perform advisory role to productionize machine learning models

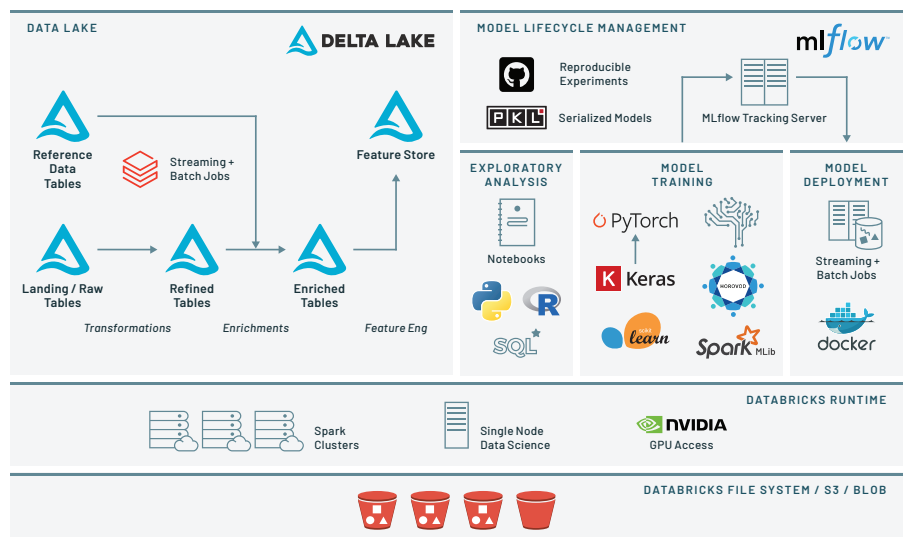
Key outcomes

- Build a reference implementation for one ML pipeline jointly determined by the customer and Databricks
- Apply machine learning best practices at scale
- Optimize machine learning pipelines
- Extend options for ML pipeline automation and MLOps by leveraging MLflow for experiment tracking, reproducibility and rollbacks

Strategy

The package offers two tiers: ML Model MVP and MLOps Optimized. The milestones and outcomes for each tier are produced by our prescriptive methodology, and each tier can be chained to have a greater impact on bolstering your enterprise ML initiatives and adoption.

See the **Resources and schedule** section for details.



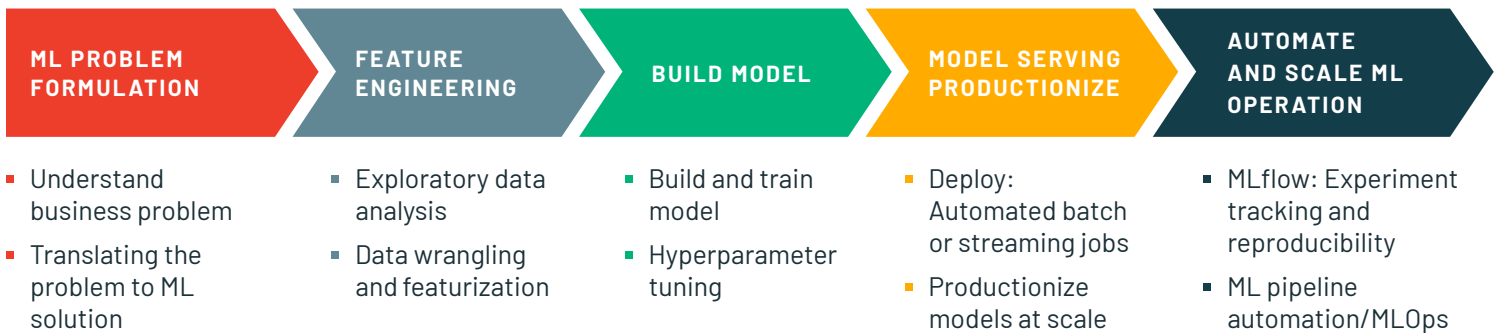
Challenges of building and deploying ML models

- Diversity and number of ML tools
- Transitioning from experiments to production without rewriting the code
- Data and model drift
- Translating business problems to ML problem statements
- Fully understanding ML limitations and where ML can be applied

Key benefits

- Data discovery and machine learning at scale
- Optimized machine learning pipelines
- Best machine learning practices
- Increased data science productivity
- Repeatable ML

Databricks ML pipeline workflow



Resources and schedule

ML Model MVP – 3 weeks

Pick one option:

- Reference implementation of one ML pipeline
- Migration of one ML pipeline from single node to Spark

MLOps Optimized – 3 weeks

ML pipeline automation/MLOps

- Machine learning best practices
- Optimize machine learning pipeline
- Repeatable ML with MLflow and Delta
- Productionize machine learning models

1 data scientist supporting the activity over a 3-week sprint

Prior to kickoff, be sure to review the readiness checklist and complete required tasks

Out of scope

- Configuration and integration of non-Databricks products and systems
- Data cleansing associated with building broader data lake
- ETL unrelated to ML