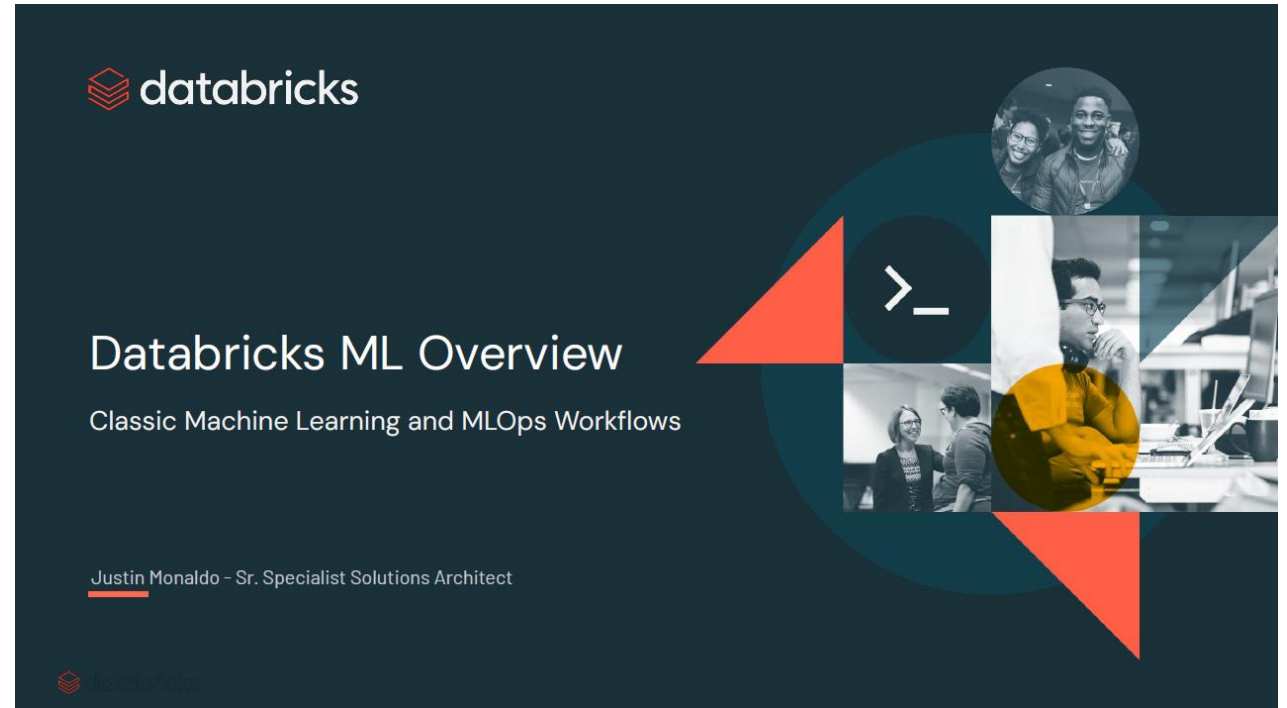




# Databricks ML Overview

Slide Deck



carahsoft.

For more information, contact Carahsoft or our reseller partners:

[Databricks@carahsoft.com](mailto:Databricks@carahsoft.com) | 703-581-6693



# Databricks ML Overview

Classic Machine Learning and MLOps Workflows

Justin Monaldo - Sr. Specialist Solutions Architect



# Agenda

- Overview
- MLflow
- Model Registry
- MLOps
- Feature Store
- AutoML
- Model Serving

# Databricks Data Intelligence Platform

100% serverless

Disaster recovery

Cost controls

Enterprise security

Mosaic AI  
Artificial intelligence



Databricks SQL  
Data warehousing



Workflows/DLT  
Ingest, ETL, streaming



AI/BI  
Business intelligence



## Lakehouse

 **Unity Catalog**

 **DELTA LAKE**

**ICEBERG**

 **Parquet**

# Databricks Machine Learning User Experience

Provide a collaborative environment for Unified Machine Learning & Data Analytics

Multi-Language  
Scala, SQL, Python, R:  
All in one notebook

Visualizations  
Built-in visualizations and  
support for the most popular  
visualization libraries  
(e.g. matplotlib, ggplot)

Experiment Tracking  
Built-in tracking of Data  
Science and ML experiments,  
with metrics, parameters,  
artifacts, and more

**Machine Learning** Provide Feedback

**Notebook**  
Create a notebook for querying, data processing, and ML.  
[Create a notebook](#)

**AutoML** Updated  
Quickly train ML models for discovery and iteration.  
[Start AutoML](#)

**Guide: Training**  
Get started with a tutorial on training and tuning ML models.  
[Start guide](#)

**Reference Solutions**  
Learn from notebooks that tackle common ML problems.  
[View reference solutions](#)

**Recently viewed**

Name	Type	Last viewed
crobison-milflow-wine	Endpoint	2 days ago
ML Flow Demo - UC Davis Wine Data	Notebook	2 days ago
Feature Store Taxi example	Notebook	4 days ago
crobison-milflow-wine	Model	16 days ago
ML Flow Demo - UC Davis Wine Data	Experiment	24 days ago

**Documentation**  
[Getting started with machine learning on Databricks](#)  
Targeted tutorials for different machine learning settings  
[Machine learning and deep learning guide](#)  
Documentation for model training and inference on Databricks  
[MLflow guide](#)  
Managing the machine learning lifecycle on Databricks  
[More documentation](#)

**Release Notes**  
[Runtime Release Notes](#)  
[MLflow Release Notes](#)  
[Platform Release Notes](#)  
[More release notes](#)

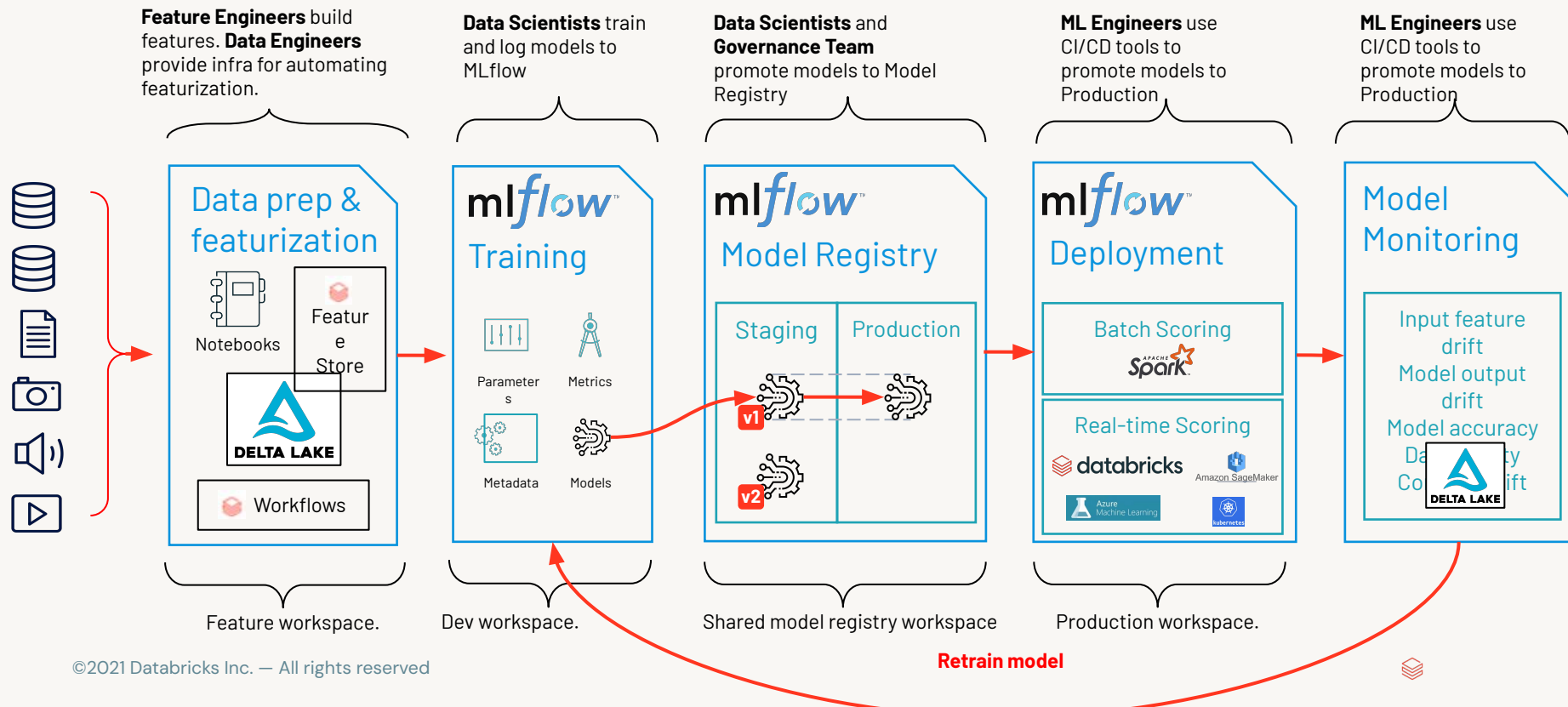
**Blog Posts**  
[Moneyball 2.0: Real-time Decision Making With MLB's Statcast Data](#)  
The Oakland Athletics baseball team in 2002 used data analysis and quantitative modeling to identify undervalued players and create a competitive lineup on a limited budget...  
[GPU-accelerated Sentiment Analysis Using Pytorch and Huggingface on Databricks](#)  
Sentiment analysis is commonly used to analyze the sentiment present within a body of text, which could range from a review, an

Reproducible  
Auto-logged revision history  
and Git integration for  
version control

Collaborative  
Realtime co-editing, with  
sharing and permissions

Enterprise Ready  
Enterprise grade access  
controls, identity pass-through,  
and auditability

# Databricks Machine Learning Pipeline



# mlflow Auto-Logging for Reproducibility

Auto-Logging of Cluster Specification and Environment Dependencies

**Source:** MLflow Model P

**Tags**

Name	Value
sparkDataSourceInfo	path=dbfs:/delta/clemens_w...

**Artifacts**

- model
  - data
    - MLmodel
    - conda.yaml
    - cluster\_spec.json
    - conda.yml
    - model\_summary.txt

**Reproducibility Checklist:**

- Code versioning
- Data versioning
- Cluster configuration
- Environment specification

**Reproduce Run Feature:**

Default Experiment > Run **cb55fd06262448ad99b08cb5e9dd2e98** - Clone Run

Date: 202...  
User: test

**Clone Run**

You can clone a run to recreate and iterate on the original notebook, cluster, and environment used to execute the run. [Learn more](#)

**Param**

- Clone notebook  
Clone Risk Model Training to /users/myfolder [View Spec](#)
- Attach to the cluster Shared autoscaling [View Spec](#)

**Metric**

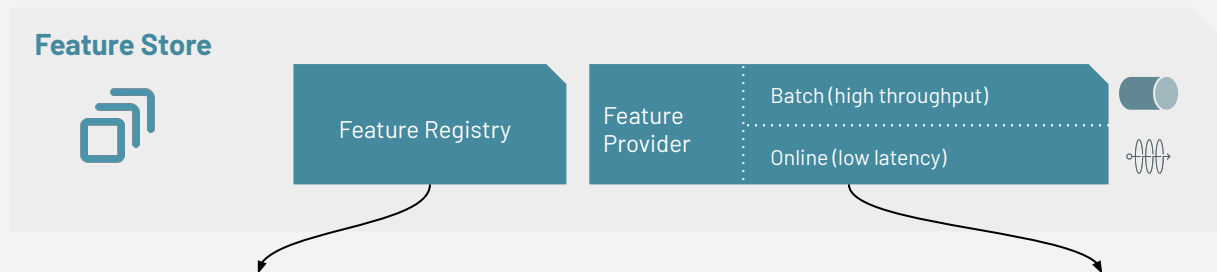
- Inject the environment set-up into the notebook [View Environment](#)

**Tags**

Name	Value	Copy	Share
run_origin	sklearn_git	<input type="checkbox"/>	<input type="checkbox"/>
version.sklearn	0.20.2	<input type="checkbox"/>	<input type="checkbox"/>

# Databricks Feature Store

The first Feature Store codesigned with a data and MLOps platform



## The Feature Registry allows reuse

- Allows reuse across projects
- Discoverability and Reusability
- Versioning
- Upstream and downstream Lineage

## The Feature Provider

- Prevents offline/online skew
- Batch and online access to Features
- Feature lookup packaged with Models
- Simplified deployment process

Co-designed with



- Open format
- Built-in data versioning and governance
- Native access through PySpark, SQL, etc.

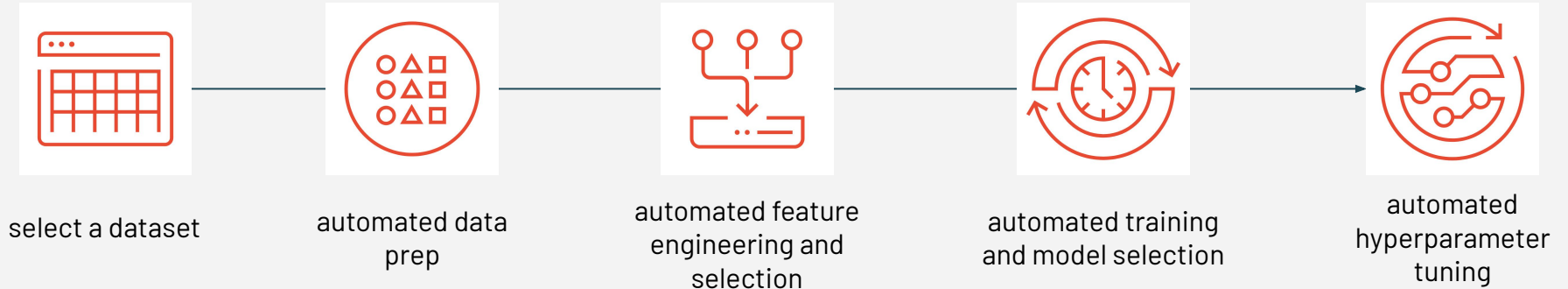
Co-designed with **mlflow**

- Open model format that supports all ML frameworks
- Feature version and lookup logic hermetically logged with Model



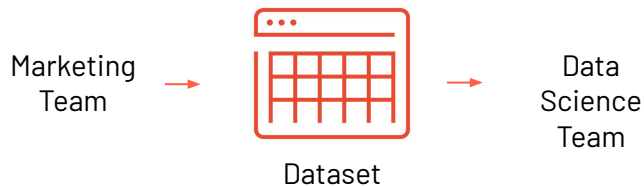
# What is AutoML?

**Automated machine learning (AutoML)** is a fully-automated model development solution seeking to “democratize” machine learning. While the scope of the automation varies, AutoML technologies usually automate the ML process from data to model selection.



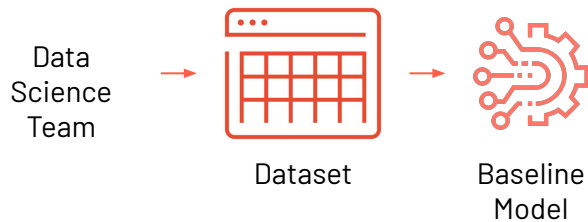
# AutoML solves two key pain points for data scientists

## Quickly Verify the Predictive Power of a Dataset



*“Can this dataset be used to predict customer churn?”*

## Get a Baseline Model to Guide Project Direction



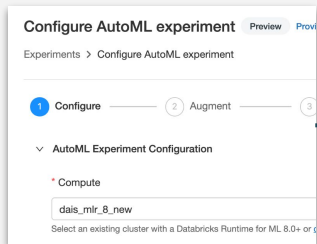
*“What direction should I go in for this ML project and what benchmark should I aim to beat?”*



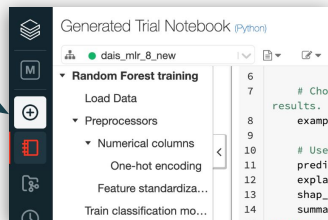
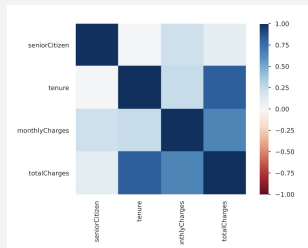
# Databricks AutoML

A glass-box solution that empowers data teams without taking away control

UI and API to start AutoML training



<input type="checkbox"/>	Start Time	Run Name	User	Source
<input type="checkbox"/>	2021-05-05 1	logistic_r...	kase...	Noteb...
<input type="checkbox"/>	2021-05-05 1	logistic_r...	alkis...	21-05...
<input type="checkbox"/>	2021-05-05 1	logistic_r...	alkis...	21-05...
<input type="checkbox"/>	2021-05-05 1	logistic_r...	kase...	Noteb...
<input type="checkbox"/>	2021-05-05 1	logistic_r...	kase...	Noteb...
<input type="checkbox"/>	2021-05-05 1	logistic_r...	kase...	Noteb...
<input type="checkbox"/>	2021-05-05 1	decision_...	kase...	Noteb...
<input type="checkbox"/>	2021-05-05 1	random_f...	kase...	Noteb...



## MLflow experiment

Auto-created MLflow Experiment to track models and metrics



Easily deploy to Model Registry

## Data exploration notebook

Generated notebook with feature summary statistics and distributions



Understand and debug data quality and preprocessing

## Reproducible trial notebooks

Generated notebooks with source code for every model

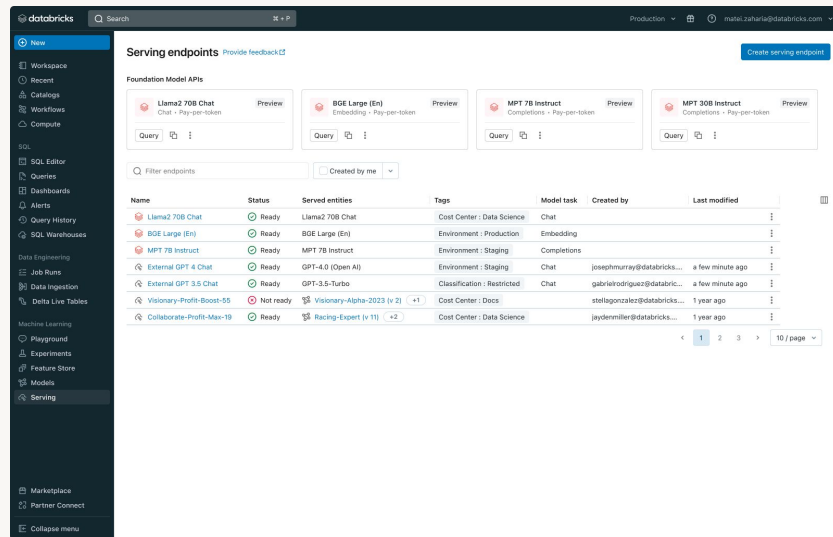


Iterate further on models from AutoML, adding your expertise

# Databricks Model Serving

## Access, Govern, and Monitor any AI Models with Databricks Model Serving

- Build AI apps faster with Unified Model Serving:** Deploy or query any model through a unified interface, including Custom Models, Databricks-managed Foundations Models (like base/fine-tuned Llama3), or externally-hosted models (like OpenAI GPT 3.5)
- Reduce TCO with Serverless Serving:** Highly available and scalable serving with LLM-specific optimization that reduces latency and cost
- Scale deployments with Data Intelligence Platform:** Automatic feature/vector lookups, monitoring, and unified governance that scales deployments and reduces errors



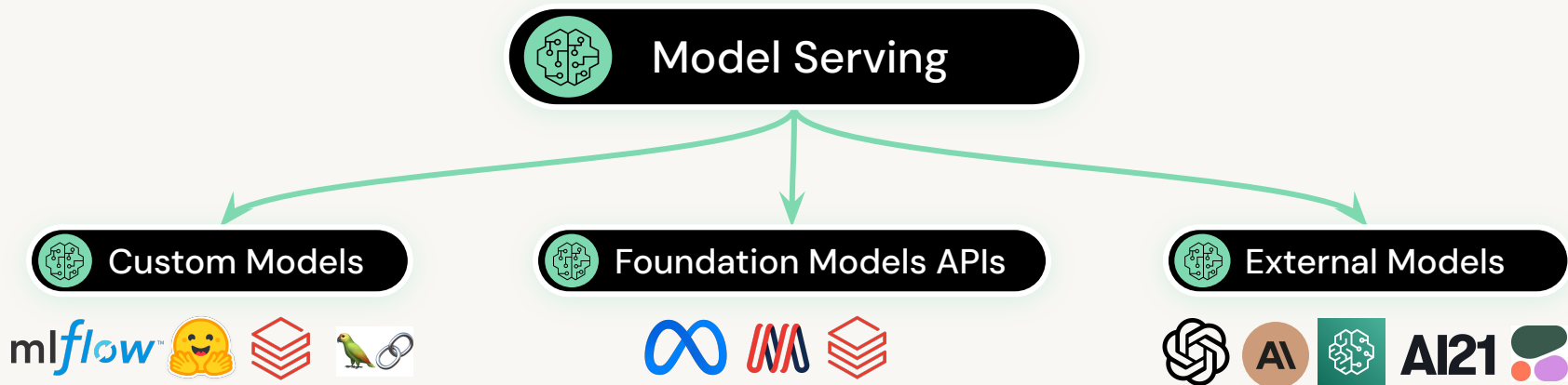
The screenshot displays the Databricks Model Serving interface. The top navigation bar includes the Databricks logo, a search bar, and the user's name 'nishi.zahara@databricks.com'. The left sidebar contains navigation options such as 'New', 'Workspace', 'Recent', 'Catalogs', 'Workflows', 'Compute', 'SQL', 'SQL Editor', 'Queries', 'Dashboards', 'Alerts', 'Query History', 'SQL Warehouses', 'Data Engineering', 'Job Runs', 'Data Inspector', 'Data Live Tables', 'Machine Learning', 'Playground', 'Experiments', 'Feature Store', 'Models', 'Serving', 'Marketplace', and 'Partner Connect'. The main content area is titled 'Serving endpoints' and features a 'Provide feedback' link and a 'Create serving endpoint' button. Below this, there are four preview cards for different models: 'Llama2 70B Chat', 'BGE Large (En)', 'MPT 7B Instruct', and 'MPT 30B Instruct'. Each card includes a 'Query' input field. A table below the cards lists the serving endpoints with columns for Name, Status, Served entities, Tags, Model task, Created by, and Last modified. The table contains several rows of data, including 'Llama2 70B Chat', 'BGE Large (En)', 'MPT 7B Instruct', 'External GPT 4 Chat', 'External GPT 3.5 Chat', 'Visionary-Profit-Boost-05', and 'Collaborate-Profit-Max-10'. The bottom right corner of the table shows pagination controls for 10 pages.

Name	Status	Served entities	Tags	Model task	Created by	Last modified
Llama2 70B Chat	Ready	Llama2 70B Chat	Cost Center : Data Science	Chat		
BGE Large (En)	Ready	BGE Large (En)	Environment : Production	Embedding		
MPT 7B Instruct	Ready	MPT 7B Instruct	Environment : Staging	Completions		
External GPT 4 Chat	Ready	GPT-4.0 (Open AI)	Environment : Staging	Chat	josephmurray@databricks...	a few minute ago
External GPT 3.5 Chat	Ready	GPT-3.5-Turbo	Classification : Restricted	Chat	gabrielrodriqez@databric...	a few minute ago
Visionary-Profit-Boost-05	Not ready	Visionary-Alpha-2023 (v 2)	Cost Center : Docs		stelegonzalez@databricks...	1 year ago
Collaborate-Profit-Max-10	Ready	Racing-Expert (v 1) +2	Cost Center : Data Science		jaydemiller@databricks...	1 year ago



# Databricks Model Serving

Unified UI, API & SDK for managing all types of AI Models



**Access any AI Model**, be it a fully custom model, an agent/chain model, a Databricks-managed Foundation Model, or a 3rd-party Foundation Model.

**Govern and Monitor all Models** in one place

**Query Models via unified interface** (Single API, SDK and UI)



# Foundation Model APIs

Fully Featured, Enterprise Grade and Secure API endpoints

**Instantly access popular Foundation Models** securely within Databricks

**Flexible Pricing:** Start with pay-per-token pricing or deploy fine-tuned models with performance guarantees on Provisioned Throughput

**Reduce TCO** with latency and throughput optimization

The screenshot displays the 'Serving endpoints' page in Databricks. At the top, there are three preview cards for different models: 'BGE Large (En)' (Embeddings - Pay-per-token), 'MPT 30B Instruct' (Completions - Pay-per-token), and 'MPT 7B Instruct' (Completions - Pay-per-token). Below these is a search bar and a filter for 'Owned by me'. A table lists the following serving endpoints:

Name	State	Served entities	Tags	Task	Created by	Last modified
quinn-bot-d...	Ready	quinn... (v 6008)			quinn.leng@data...	56 minutes ago
all-MiniLM-L...	Ready	all-MiniLM... (v 1)			sheng.zhan@dat...	2 hours ago
model-using...	Ready	feature_... (v 1)			mingyang.ge@d...	5 hours ago
table_valida...	Ready	ml.anl.te... (v 1)			anirudh.achal@d...	6 hours ago
table-validat...	Ready	ml.anl.te... (v 1)			anirudh.achal@d...	6 hours ago
table-validat...	Ready	ml.anl.te... (v 1)			anirudh.achal@d...	6 hours ago
table-validat...	Ready	ml.anl.te... (v 1)			anirudh.achal@d...	6 hours ago
ian-test-2	Ready	leon-sle... (v 1)			ian.rodney@data...	6 hours ago
ian-test	Ready	leon-sle... (v 1)			ian.rodney@data...	6 hours ago
glean_rag_c...	Ready	ml.glea... (v 31)			chengzu.ou@dat...	7 hours ago
fs-e2e-brick...	Ready	online_L... (v 4)			tommy.xiang@d...	8 hours ago



# Monitoring with Inference Tables

## Automatically capture request/responses in a Lakehouse Delta table

- **Lakehouse support**  
Query inference logs as first-class SQL data and easily integrate with DLT and DBSQL dashboards
- **Monitor model quality**  
Join ground-truth labels and alert on model performance
- **Debug models quickly**  
Investigate errors from raw requests/responses
- **Create training datasets**  
Use data from live endpoint traffic to retrain models
- **Meet compliance requirements**  
Track prediction responses to streamline compliance workflows

The screenshot displays the Databricks SQL interface. On the left, a sidebar shows navigation options like SQL Editor, Workspace, Queries, Dashboards, Alerts, Data, SQL Warehouses, Workflows, and Query History. The main area is split into a Schema browser and a New query editor. The Schema browser shows a hierarchy of tables under 'hive\_metastore > default', including various test tables and inference logs. The New query editor contains a SQL query: `select * from delta.`dbfs:/paylod-logging/flower-classification/Production``. Below the query, a table of inference logs is displayed with columns: #, inference\_id, date, http\_status, timestamp\_ms, endpoint\_url, model\_name, mode, sampling, and request. The table contains 14 rows of data, with the first row highlighted. The interface also shows a 'Run' button with a 'Run limit 1000' dropdown, and a 'Refreshed 4 minutes ago' status at the bottom right.

#	inference_id	date	http_status	timestamp_ms	endpoint_url	model_name	mode	sampling	request
1	d7d25499-12c3-494b-a8bf-8124d6d6e374	10/04/22	200	1664920968698	Production	test-model-1	1	1.00	{ "dataframe_recor
2	62b95237-2f0d-43b6-ab1c-9395a4a92a2a	10/04/22	200	1664920969479	Production	test-model-1	1	1.00	{ "dataframe_recor
3	69337192-8828-4f6a-ae9c-5f351ac44c4d	10/04/22	200	1664920969701	Production	test-model-1	1	1.00	{ "dataframe_recor
4	cbdf05b6-1371-4cac-9391-c9d4f8e3dbe3	10/04/22	200	1664920970011	Production	test-model-1	1	1.00	{ "dataframe_recor
5	35ff11fc-f65a-45c3-9968-ea0a3636991b	10/04/22	200	1664920970172	Production	test-model-1	1	1.00	{ "dataframe_recor
6	fa05c27b-7445-4161-afde-3f176f33d7a3	10/04/22	200	1664920971518	Production	test-model-1	1	1.00	{ "dataframe_recor
7	222581f5-8aca-4318-8b09-716c8662f8e5	10/04/22	200	1664920971620	Production	test-model-1	1	1.00	{ "dataframe_recor
8	6af0d483-ec7a-46b7-9d8c-3930b33bf1d8	10/04/22	200	1664920971687	Production	test-model-1	1	1.00	{ "dataframe_recor
9	547ac067-a355-43b8-b989-31c366d588d1	10/04/22	200	1664920972193	Production	test-model-1	1	1.00	{ "dataframe_recor
10	1f15957d-5b50-45e9-9166-24a333de5ebb	10/04/22	200	1664920972236	Production	test-model-1	1	1.00	{ "dataframe_recor
11	1426d059-bf21-4dd7-945f-ed5d6ce3c579	10/04/22	200	1664920972700	Production	test-model-1	1	1.00	{ "dataframe_recor
12	0b40e4ef-c35e-4132-8c5c-4262a52476a9	10/04/22	200	1664920972836	Production	test-model-1	1	1.00	{ "dataframe_recor
13	862290a0-9eb6-403b-9574-dfecd8e8501a	10/04/22	200	1664920973203	Production	test-model-1	1	1.00	{ "dataframe_recor
14	990a0536-3963-4cd2-99fa-9ecb98a843c4	10/04/22	200	1664920973259	Production	test-model-1	1	1.00	{ "dataframe_recor





---

Thank you for downloading this Databricks presentation! Carahsoft serves as the Master Government Aggregator® and Distributor for Databricks, offering expertise in government procurement processes and practices with purchasing available via GSA, SEWP V, ITES-SW and other contract vehicles.

To learn how to take the next step toward acquiring Databricks' solutions, please check out the following resources and information:



For additional resources, please visit [carah.io/DatabricksResources](https://carah.io/DatabricksResources)



For additional solutions, visit [carah.io/DatabricksSolutions](https://carah.io/DatabricksSolutions)



To speak with our team directly, email [Databricks@carahsoft.com](mailto:Databricks@carahsoft.com) or reach out at 703-581-6693.



To view our upcoming Databricks events, visit [carah.io/DatabricksEvents](https://carah.io/DatabricksEvents)



For additional Open Source solutions, visit [carah.io/OpenSourceSolutions](https://carah.io/OpenSourceSolutions)



To purchase, check out the contract vehicles available for procurement at [carah.io/DatabricksContracts](https://carah.io/DatabricksContracts)

**carahsoft**

For more information, contact Carahsoft or our reseller partners:  
[Databricks@carahsoft.com](mailto:Databricks@carahsoft.com) | 703-581-6693