

# Product Safe Harbor Statement

This information is provided to outline Databricks' general product direction and is for informational purposes only. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward-looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all.

# Enable Production ML with Databricks Feature Store



**Aakrati Talati**

Software Engineer, Databricks



**Avesh Singh**

Software Engineer, Databricks

# The Talk

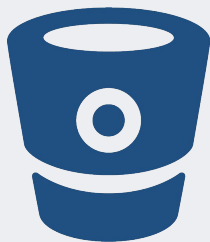
Thesis:

*Databricks Feature Store enables your ML projects to reach production.*

## Agenda

- What makes productionizing ML projects hard?
- 4 Production ML Challenges
- Databricks Feature Store
- Demo
- Conclusion & Questions

# Production Machine Learning



JUnit



Production Machine Learning

Production Software



PagerDuty

Production Machine Learning  
= Production Software + Data

# Data Challenges in Production Machine Learning

# Data Challenges in ML

Raw Data

Featurization  
Joins, Aggregates, Transforms, etc.

Training



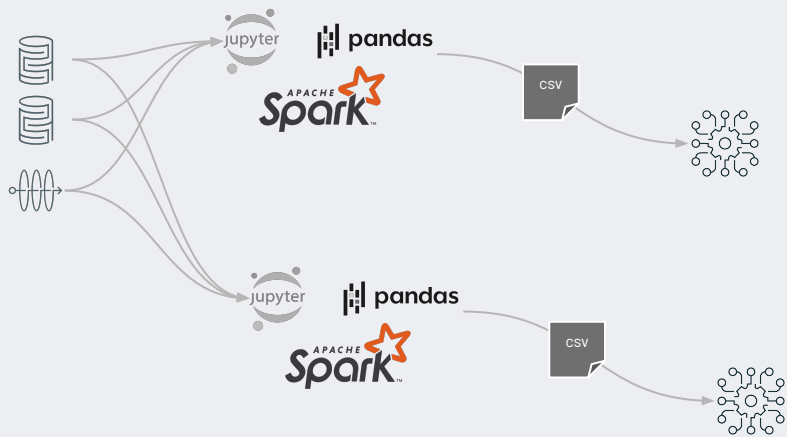


# Data Challenges in ML

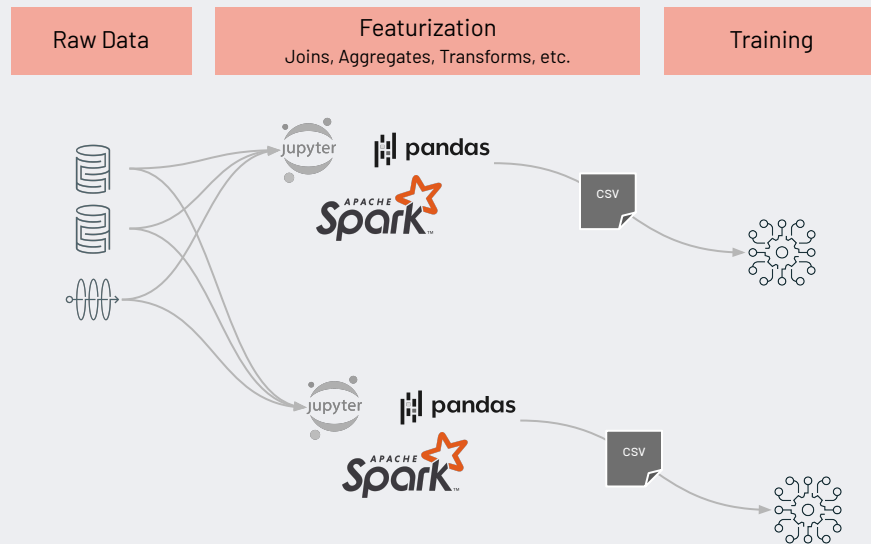
Raw Data

Featurization  
Joins, Aggregates, Transforms, etc.

Training



# Data Challenges in ML



Challenge 1: Data silos

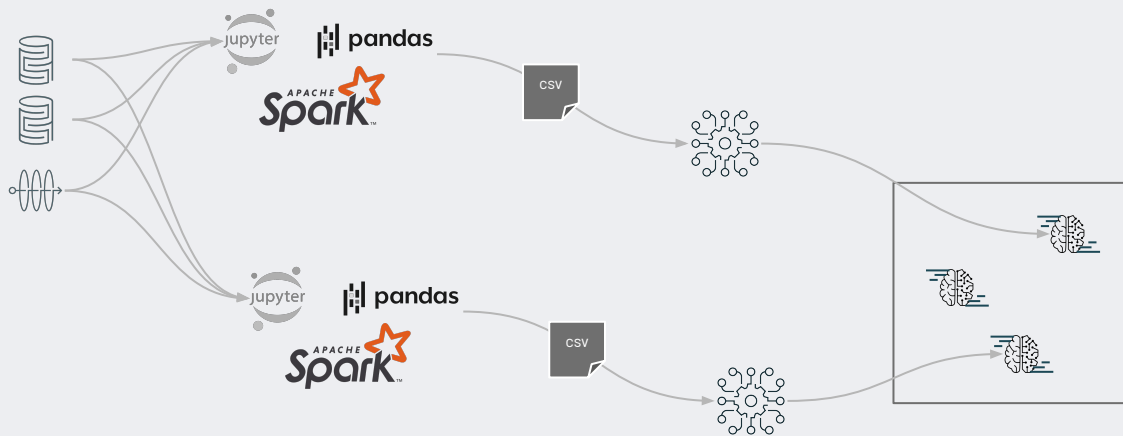
# Data Challenges in ML

Raw Data

Featurization  
Joins, Aggregates, Transforms, etc.

Training

Serving



# Data Challenges in ML

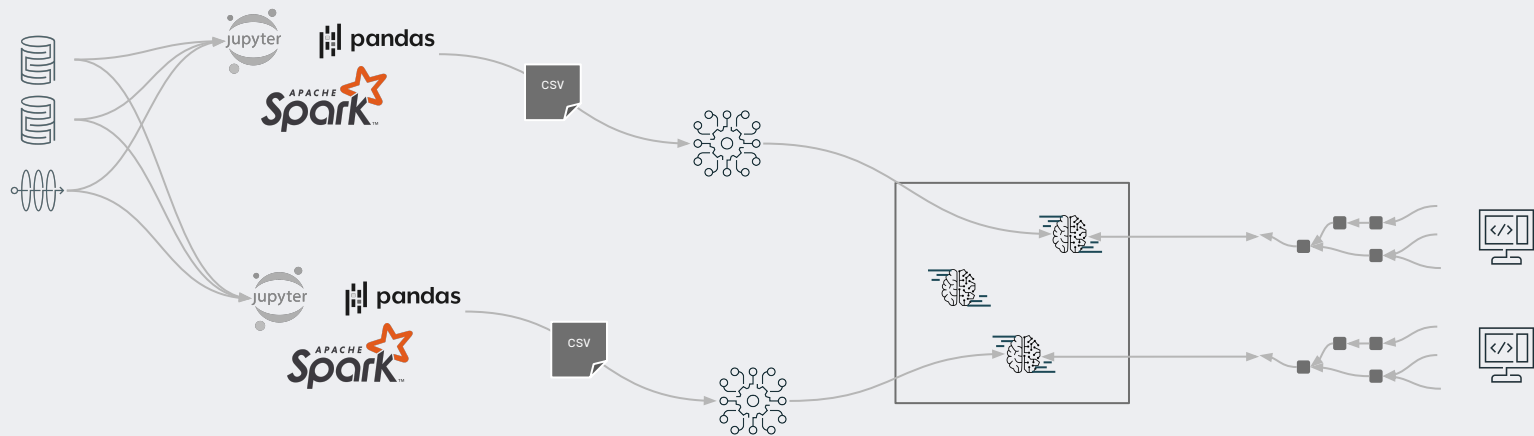
Raw Data

Featurization  
Joins, Aggregates, Transforms, etc.

Training

Serving

Client



# Data Challenges in ML

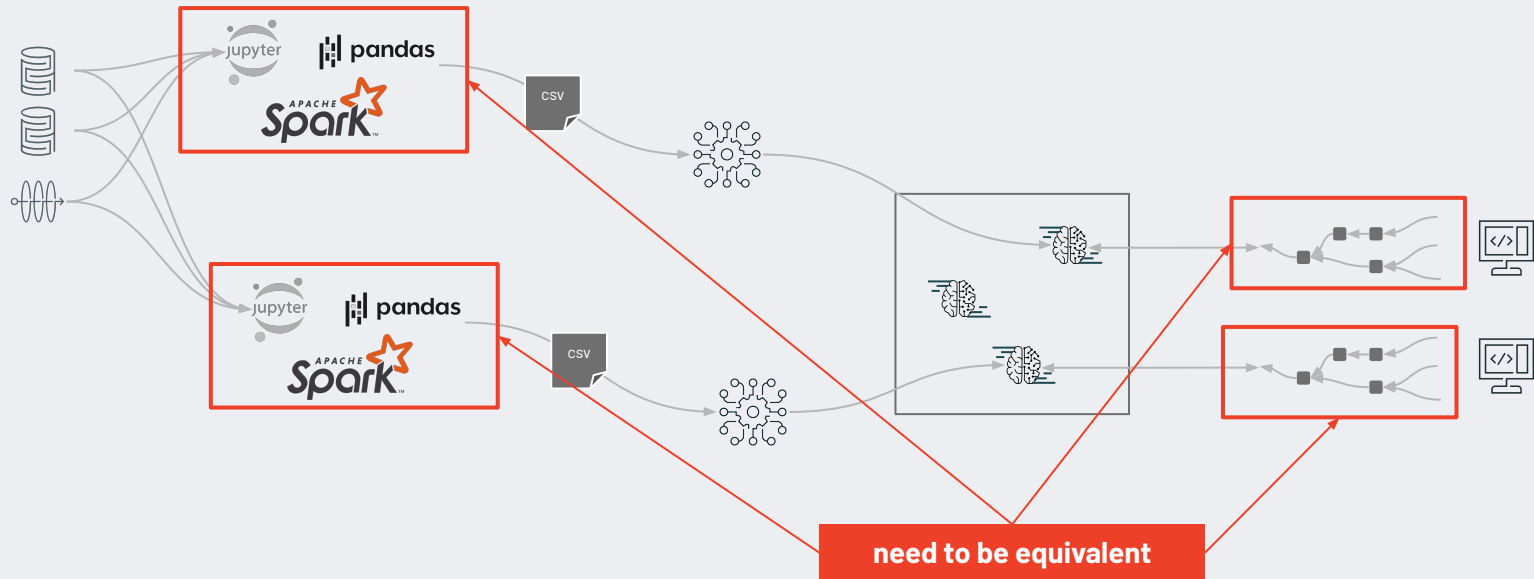
Raw Data

Featurization  
Joins, Aggregates, Transforms, etc.

Training

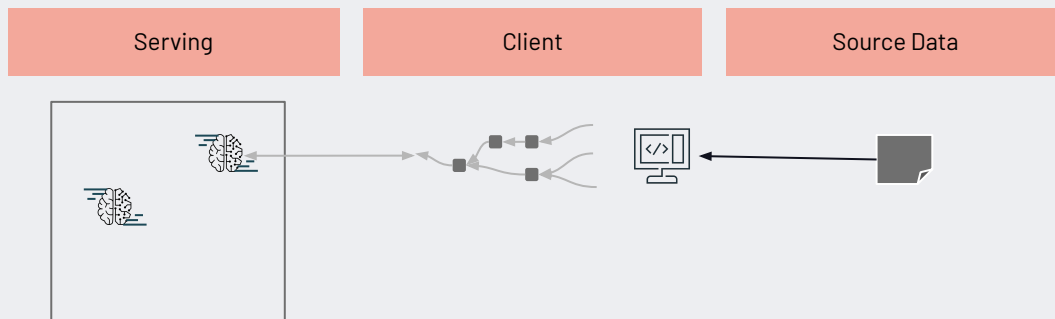
Serving

Client

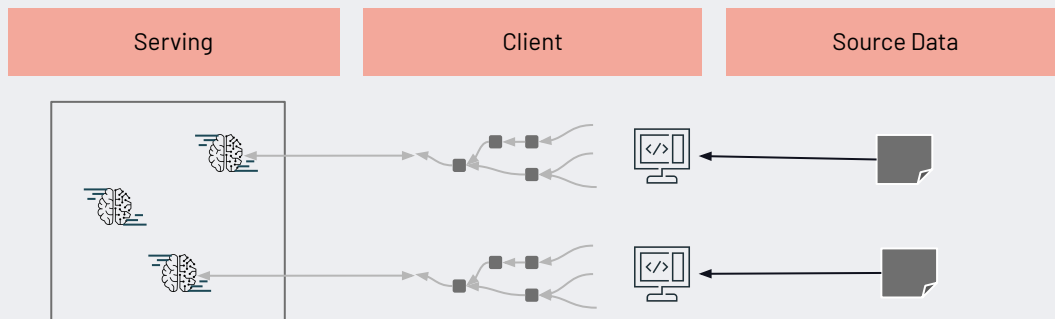


## Challenge 2: Online/offline skew

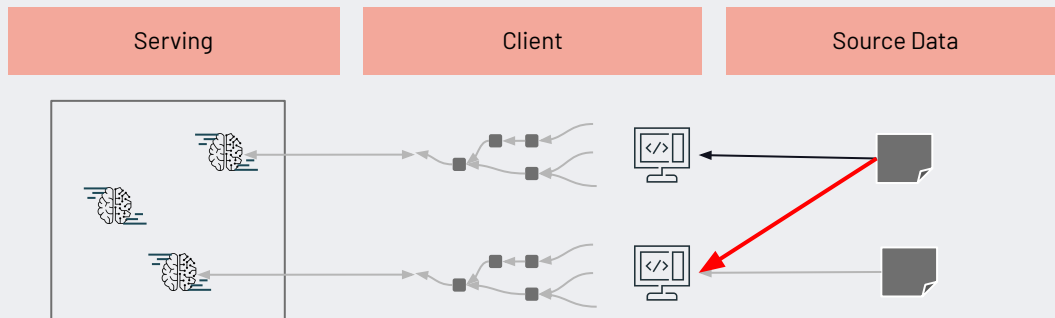
# Data Challenges in ML



# Data Challenges in ML



# Data Challenges in ML



## Challenge 3: Client Configuration



# Data Challenges in ML

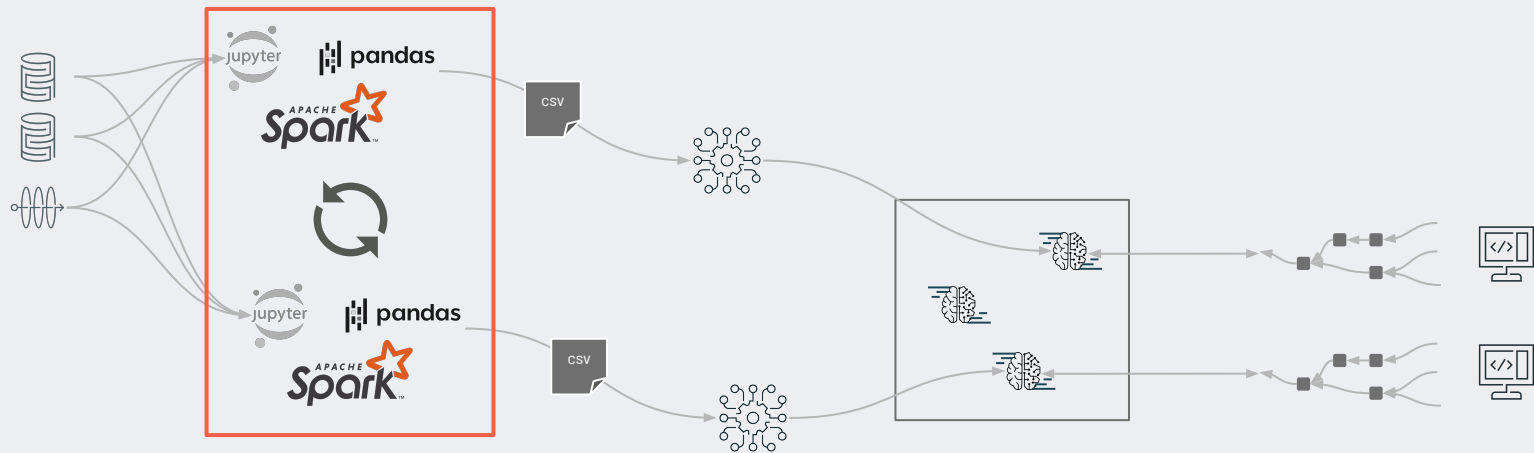
Raw Data

Featurization  
Joins, Aggregates, Transforms, etc.

Training

Serving

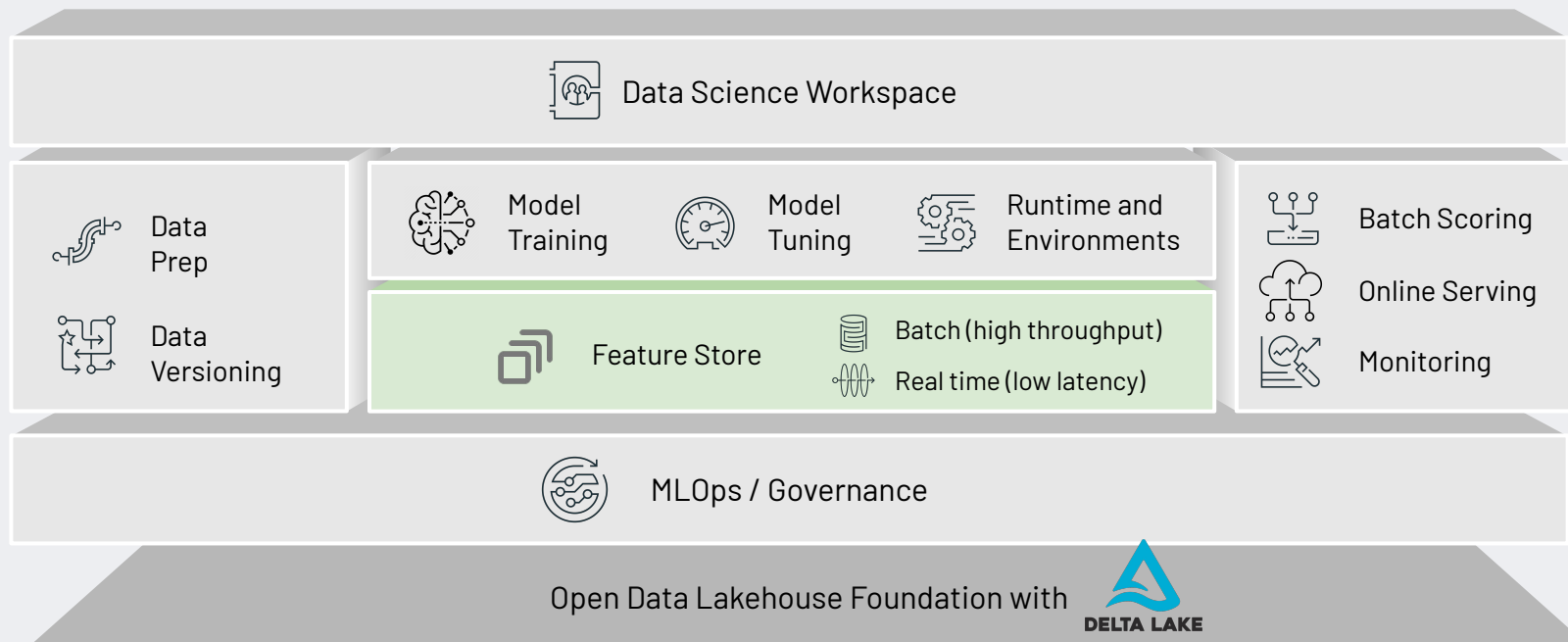
Client



## Challenge 4: Data Freshness

# Databricks Feature Store

A feature store built on top of the data lakehouse.



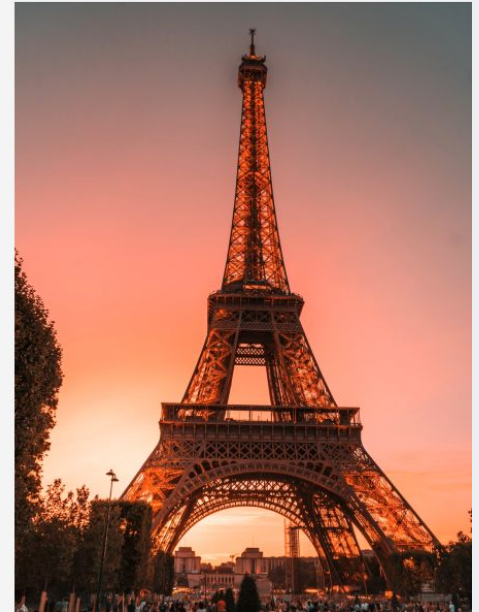
# Demo

# Bon Voyage Demo

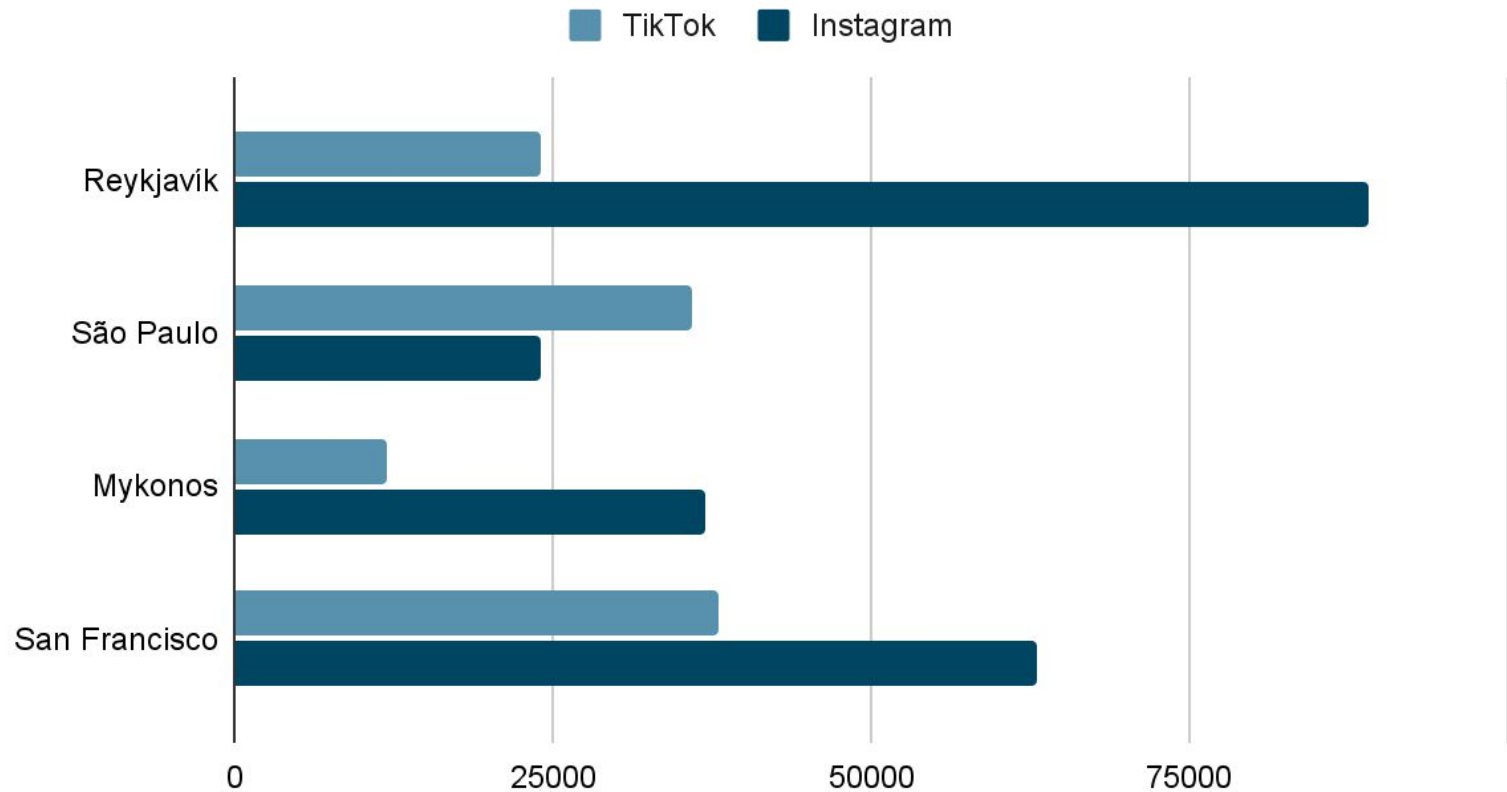


# Vacation Recommender

*Top picks for you*



# Top Travel Destination Mentions



# Summary

Solve Data challenges in  
Production ML with  
Databricks Feature Store



# Summary: Solve Data challenges in Production ML with Databricks Feature Store

- Remove data silos with native lakehouse integration
- DLT + Feature Store for reliable pipelines
- Eliminate offline/online skew
- On-demand features with MLflow pyfunc
- Keep your data fresh with Spark Structured Streaming
- Automatic MLflow integration





# Thank you!

## Questions?

Get Started – [databricks.com/try](https://databricks.com/try)

Contact us

- Aakrati Talati <[aakrati.talati@databricks.com](mailto:aakrati.talati@databricks.com)>
- Avesh Singh <[avesh@databricks.com](mailto:avesh@databricks.com)>
- [databricks.com/contact](https://databricks.com/contact)

# Additional slides



## Social Media Destination Popularity Pipeline

Development

Production

Delete

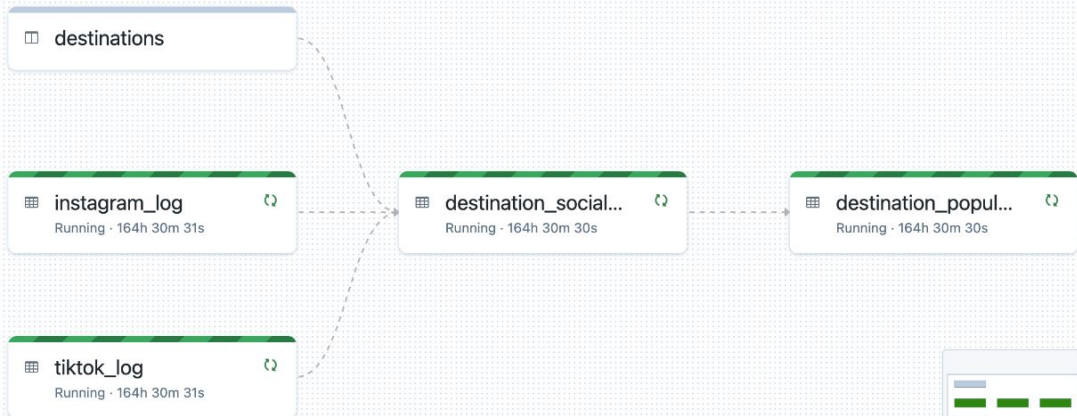
Permissions

Settings

Schedule

Stop

6/21/2022, 9:53:34 PM · Running ⌵



### Pipeline details

**Name** Social Media Destination Popularity Pipeline  
**Pipeline ID** aaccb7a3-eddc-4912-9de1-5f843dfc942c  
**Paths** [/Users/avesh.singh@databricks.com/DAIS 2022/Social Media Destination Popularity](#)  
**Run as** avesh.singh@databricks.com

### Update details

**Update ID** 048a6766-92df-4e7d-a3c2-c625fa854578  
**Status** Running  
**Creation time** 6/21/2022, 9:53:34 PM  
**Duration** 164h 30m 45s  
**Start time** 6/21/2022, 9:53:36 PM  
**Run time** 164h 30m 44s

| All | Info       | Warning       | Error | Filter...  | × |
|-----|------------|---------------|-------|--|---|
| 1/3 | 7 days ago | flow_progress |       | Flow 'instagram_log' is RUNNING.                 |   |
|     | 7 days ago | flow_progress |       | Flow 'destination_social_aggregate' is STARTING. |   |
|     | 7 days ago | flow_progress |       | Flow 'destination_social_aggregate' is RUNNING.  |   |
|     | 7 days ago | flow_progress |       | Flow 'destination_popularity' is STARTING.       |   |
|     | 7 days ago | flow_progress |       | Flow 'destination_popularity' is RUNNING.        |   |

Primary Keys: destination\_id (STRING)

Timestamp Keys: impression\_ts (TIMESTAMP)

▼ Description [Edit](#)

### Social media popularity signals for destination popularity.

▼ Tags (0)

| Name                              | Value                              | Actions                            |
|-----------------------------------|------------------------------------|------------------------------------|
| No tags found.                    |                                    |                                    |
| <input type="text" value="Name"/> | <input type="text" value="Value"/> | <input type="button" value="Add"/> |

▼ Producers (1)

| Name   | Schedule    | Status | Last run | Last written        |
|--|-------------|--------|----------|---------------------|
|  Social Media Destination Popularity Pipeline | No schedule | -      | -        | 2022-06-22 10:47:30 |

▼ Online Stores (1)

| Online Store  | Region    | Cloud | Storage  | Last published      | Last published by   |
|---|-----------|-------|----------|---------------------|---|
| <a href="#">Features, plans, and pricing</a><br><a href="#">Contact us</a><br><a href="#">Feedback</a><br><a href="#">Help</a><br><a href="#">Privacy</a><br><a href="#">Terms</a><br><a href="#">Security</a><br><a href="#">Status</a><br><a href="#">Partners</a><br><a href="#">Press</a><br><a href="#">Careers</a><br><a href="#">About</a> | us-east-2 | aws   | DYNAMODB | 2022-09-07 15:00:00 | xhr: 0, ws: 226, bytes: 244559, xhr-open: 0, ws-open: 0<br><a href="#">[OPS] Debug Metrics Requests</a> |



Permissions

Primary Keys: destination\_id (STRING)



Timestamp Keys: ts (TIMESTAMP)

Data Sources: <dbfs:/databricks-datasets/retail-org/products/products.csv>





▼ Description [Edit](#)

Aggregated destination features, updated daily.

▼ Tags (1)

| Name                              | Value                              | Actions   |
|-----------------------------------|------------------------------------|---|
| Owner                             | ds-demand-forecasting@bv.com       |   |
| <input type="text" value="Name"/> | <input type="text" value="Value"/> | <input type="button" value="Add"/>  |

## ▼ Producers (3)

| Name   | Schedule                          | Status  | Last run            | Last written        |
|--|-----------------------------------|---|---------------------|---------------------|
|  Destination Features Refresh | At 01:00 AM (America/Los_Angeles) |  Success | 2022-06-28 01:00:53 | 2022-06-28 01:01:12 |
|  Destination Features Refresh | No schedule                       | -   | -                   | 2022-06-21 23:44:01 |
|  Compute Destination Features | No schedule                       | -   | -                   | 2022-06-21 18:50:34 |



▼ Features (11)

xhr: 0, ws: 250, bytes: 256397, xhr-open: 0, ws-open: 0



# travel\_recommendations.destination\_availability [Provide Feedback](#)

⋮ Permissions

Created: 2022-06-27 15:00:51

Last written ⓘ: 2022-06-27 15:01:44

Last modified ⓘ: 2022-06-27 15:01:44

Primary Keys: destination\_id (STRING)

Created by: aakrati.talati@databricks.com

Last written by: aakrati.talati@databricks.com

Last modified by: aakrati.talati@databricks.com

Timestamp Keys: ts (TIMESTAMP)

Data Sources:

▼ Description [Edit](#)

Availability of hotels and attractions at the desitnation.

▼ Tags (0)

| Name                              | Value                              | Actions                            |
|-----------------------------------|------------------------------------|------------------------------------|
| No tags found.                    |                                    |                                    |
| <input type="text" value="Name"/> | <input type="text" value="Value"/> | <input type="button" value="Add"/> |

▼ Producers (2)

| Name                              | Schedule ⓘ  | Status | Last run ⓘ | Last written ⓘ      |
|-----------------------------------|-------------|--------|------------|---------------------|
| Travel Recommendation Model       | No schedule | -      | -          | 2022-06-27 15:01:44 |
| Destination Availability Features | No schedule | -      | -          | 2022-06-27 15:01:21 |

▼ Online Stores (1)

| Online Store  | Region    | Cloud | Storage  | Last published      | Last published by             |
|---|-----------|-------|----------|---------------------|-------------------------------|
| feature_store_travel_recommendations.destination_availability | us-west-2 | AWS   | DYNAMODB | 2022-06-27 15:02:47 | aakrati.talati@databricks.com |

xhr: 0, ws: 861, bytes: 506051, xhr-open: 0, ws-open: 0



Feature Store > travel\_recommendations.destination\_availability > feature\_store\_travel\_recommendations.destination\_availability

**feature\_store\_travel\_recommendations.destination\_availability** [Provide Feedback](#)

Created: 2022-06-27 15:02:47

Created by: aakrati.talati@databricks.com

Last published: 2022-06-27 15:02:47

Last published by: aakrati.talati@databricks.com

Cloud: AWS

Storage: DYNAMODB

Region: us-west-2

Table ARN: arn:aws:dynamodb:us-west-2:189836941263:table/feature\_store\_travel\_recommendations.destination\_availability

▼ Published Features (8)

| Feature                      |
|------------------------------|
| attractions_availability_1d  |
| attractions_availability_30d |
| attractions_availability_7d  |
| destination_id               |
| hotel_availability_1d        |
| hotel_availability_30d       |
| hotel_availability_7d        |
| ts                           |