

From Data Engineering to Predictive Analytics: A Journey with Azure Databricks and Azure ML

-SQL konferenz 2025

Julia Førde & Sai Neelakantam

First, let’s get a lay of the land

- Have you worked with  ?
- Have you worked with  ?



TODAY'S FOCUS

GENERATING SYNTHETIC DATA	How, we created the data? And a look..
TEMPLATES	Why and how do you implement templates on Lakehouse solution. What are the core building blocks
MEDALLION + WORKFLOWS	We will focus on naming, SCD2 & templates. We will also take a look at Unity Catalog.
DATABRICKS + AZURE ML CONFIG	Configuring Databricks to integrate with Azure ML for seamless model training, tracking, and deployment.
MODEL INFERENCE + PREDICTIONS	Performing model inference and generating predictions using Azure ML for real-time processing.

JULIA FØRDE

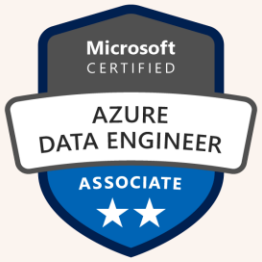
- Databricks MVP
- Lead consultant – Solutions Architect
- Microsoft on-premises → Azure → Azure Databricks
- Focus: Lakehouse's & migrations
- Based in Oslo, Norway



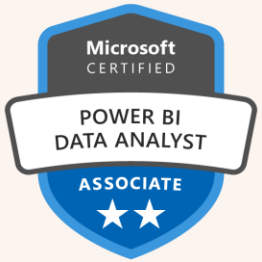
SAI NEELAKANTAM

- Data Engineer by ❤️
- Consultant @Evidi
- Microsoft Fabric, Power BI, Azure, Azure AI, Databricks
- Based in Oslo, Norway











PRODUCTS SPECIALS MYLIDL WALLET ABOUT HELP & SUPPORT

Search Products & Recipes

Find Stores
Culpeper, VA

SIGN IN

Recipes Weekly Ad Grand Openings Quality Standards Home Delivery Explore

recipes

unique recipes for you and your family.

browse by

- all recipes
  - featured
  - quick & easy
  - 30 mins or less
  - breakfast & brunch
  - lunch
  - dinner
  - healthy
  - vegan & vegetarian
  - fall
  - holiday
  - spring
  - summer
  - salad
  - seafood
  - back to school
  - beverages

featured



Tilapia Piccata



Prosciutto Rose Skewers





PRODUCTS SPECIALS MYLIDL WALLET ABOUT HELP & SUPPORT

Search Products & Recipes

Find Stores
Culpeper, VA

SIGN IN

Recipes Weekly Ad Grand Openings Quality Standards Home Delivery Explore



30 mins or less featured seafood holiday healthy dinner

share recipe:

Tilapia Piccata

4	20 min	10 min	30 min
servings	prep	cooking	total time

Piccata is an Italian dish, typically made with veal or chicken, where a meat is thinly sliced, dredged in flour, pan-fried and served with a delicious tangy sauce of lemon, butter and parsley. We've kept the exciting, homey flavors of a typical piccata and mixed it with the addition of lean tilapia protein.

ingredients

- [ ] 1 pound tilapia fillets
- [ ] ½ cup flour
- [ ] 1 teaspoon dried oregano
- [ ] 1 teaspoon garlic powder
- [ ] ¼ cup olive oil
- [ ] 3 tablespoons butter, divided

directions

1. Spread tilapia fillets on a tray, pat dry with a paper towel.

2. In a medium bowl, mix flour, oregano and garlic powder.

3. Dredge fish on both sides in flour mixture, shake off excess.

4. In a large cast iron skillet, heat olive oil and 1 tablespoon butter over medium-high heat. Add fish, cook for 3 minutes per side or until firm. Transfer to a paper towel-lined tray.

5. Add remaining butter, lemon juice and zest, capers and chicken broth to the pan. Boil and reduce slightly.

GENERATING SYNTHETIC DATA

Realistic, dynamic user profiles and preferences to drive recipe recommendations and personalization

Why Generate Fake Data?

- To create an enterprise scenario to predict user ratings for recipes
- Enables controlled experiments and testing in a scalable manner.
- Mimics real-world user behaviours and interactions.

  *(when controlled)

How We Created the Data?

- User Data:

  Generate realistic user profiles with Faker (name, email, location, etc.)

  → User Changes
  - Track user data updates: Last name, email, etc.
  - 5% probability for data change (ensuring variability).
- User preferences:

  Customized based on dietary restrictions, skill levels, and meal types.

  → User Preference Changes
  - Track changes to user preferences, such as cuisine or dietary restrictions.
  - 10% chance of preference updates each day.
- Recipe ratings:

  Ratings are influenced by the category of the recipe and match to user preference.

OUR TAKE ON THE MEDALLION ARCHITECTURE

Landing

- Main purpose: mirror of the source
- Same columns, only some datatype alignment
- Converts data to Delta format

Bronze

- Same columns
- Adding of technical columns
- Infers SCD2 on tables
- Datatype alignment

Silver

- Business logic is applied here
- Lookups to other tables(foreign keys)
- Technical naming(prefix_name_type)
- Filter away historic records if not used

Gold

- Business can decide names
- Aggregations on data

WHAT IS A TEMPLATE?

A template for us is a single notebook that allows you to run multiple tables by simply adjusting a few parameters to specify the table and run type

WHY TEMPLATES?

Although it takes more time to develop, the benefits far outweigh the initial effort:

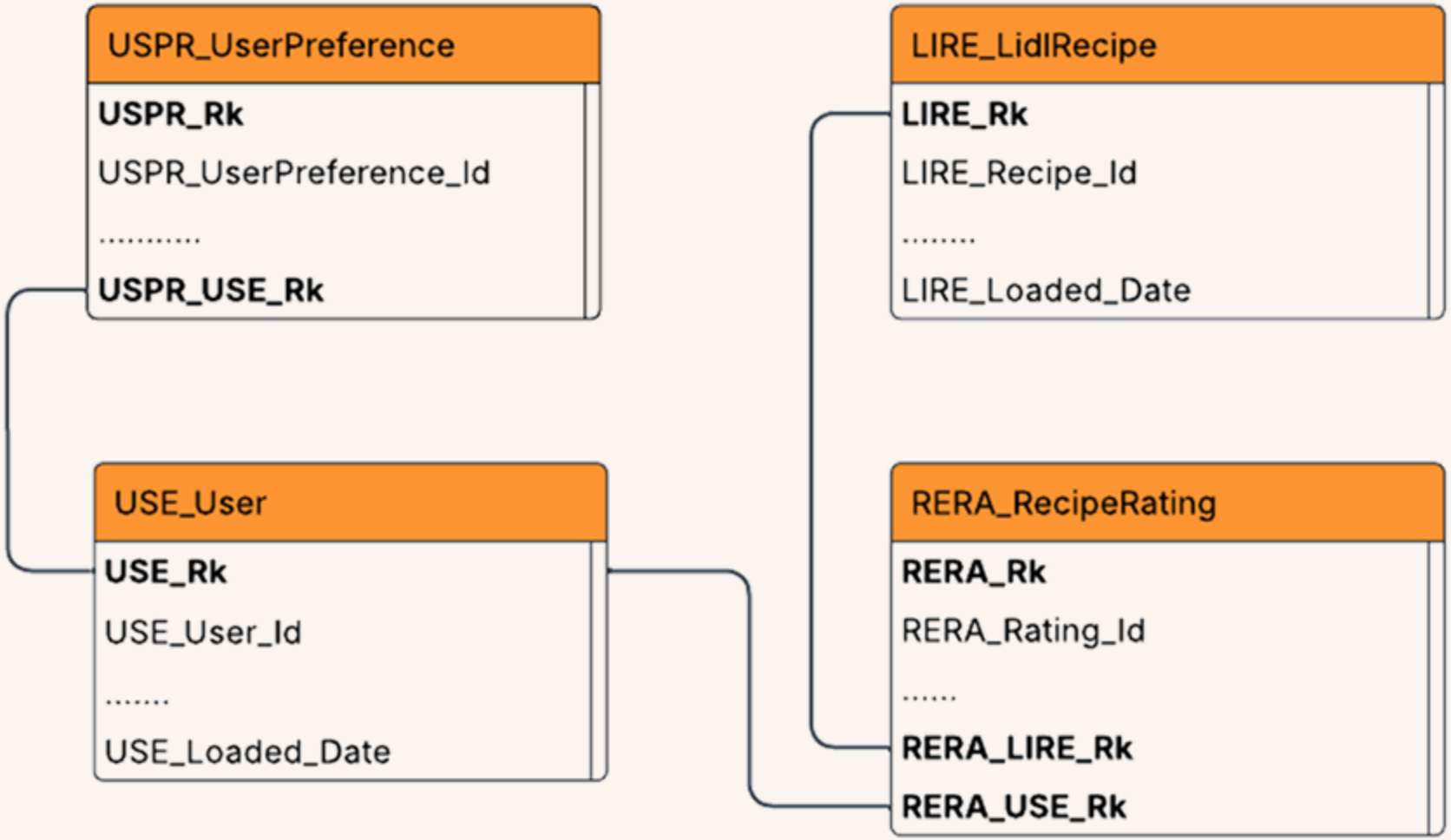
- ❖ Significant decreased development time per table.
- ❖ Provides a more standardized and consistent solution.
- ❖ Easier to debug for everyone on the team.
- ❖ Less dependent on individual developers
- ❖ Simplifies onboarding

AT THE CORE OF THE TEMPLATES

- Naming: Consistency is key
- Technical columns(Rk, Sk + timestamps)
- Parametrization
- Predefined strings for SQL and or pyspark statements
- Documentation
- Proper catalog and data storage structure
- Logging

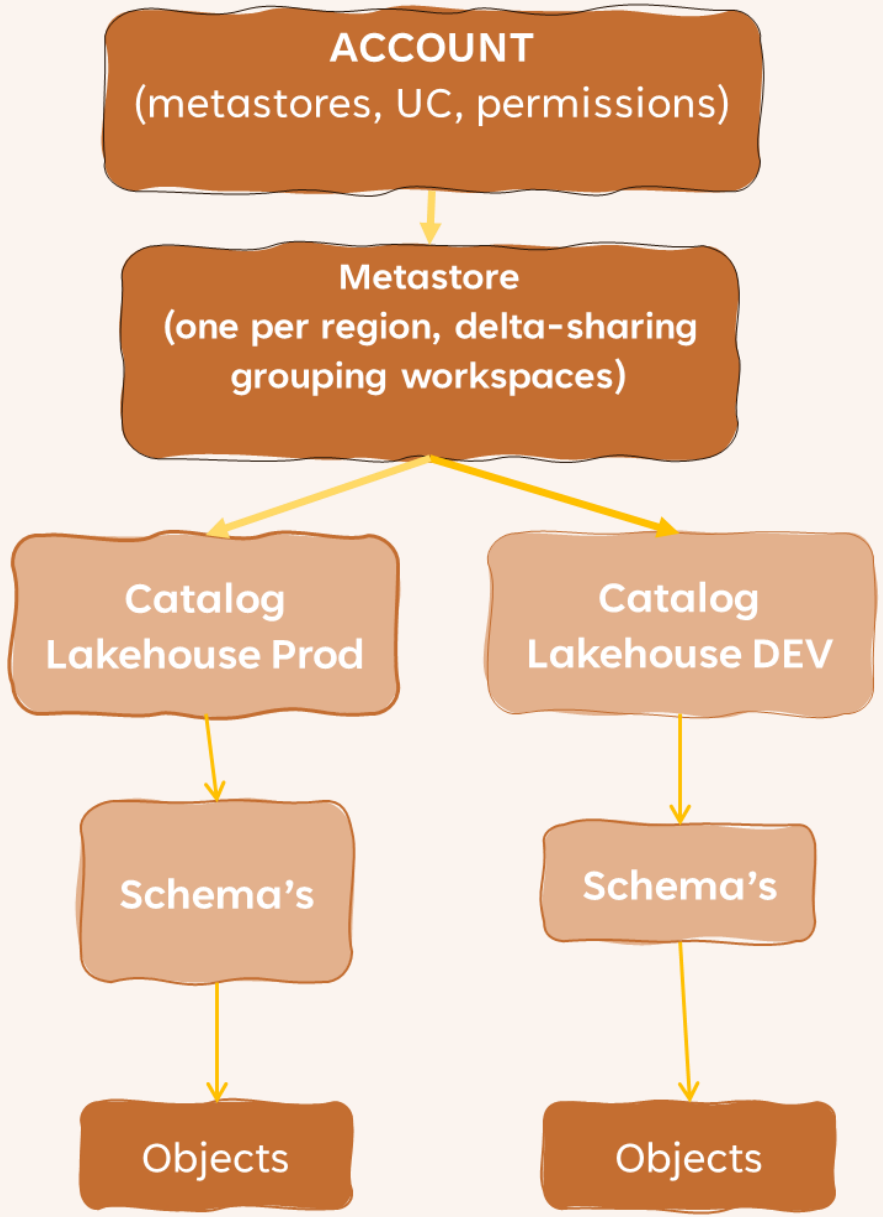
Naming example
USPR_Rk
USPR_UserPeference_Id
USPR_DietaryRestrictions_Str
USPR_Cuisine_Type
USPR_CokingTime_No
USPR_CookingLevel_Type
USPR_Quick_Flag
USPR_Healthy_Flag
USPR_Loaded_Date
USPR_USE_Rk

The magic of proper naming and keys:



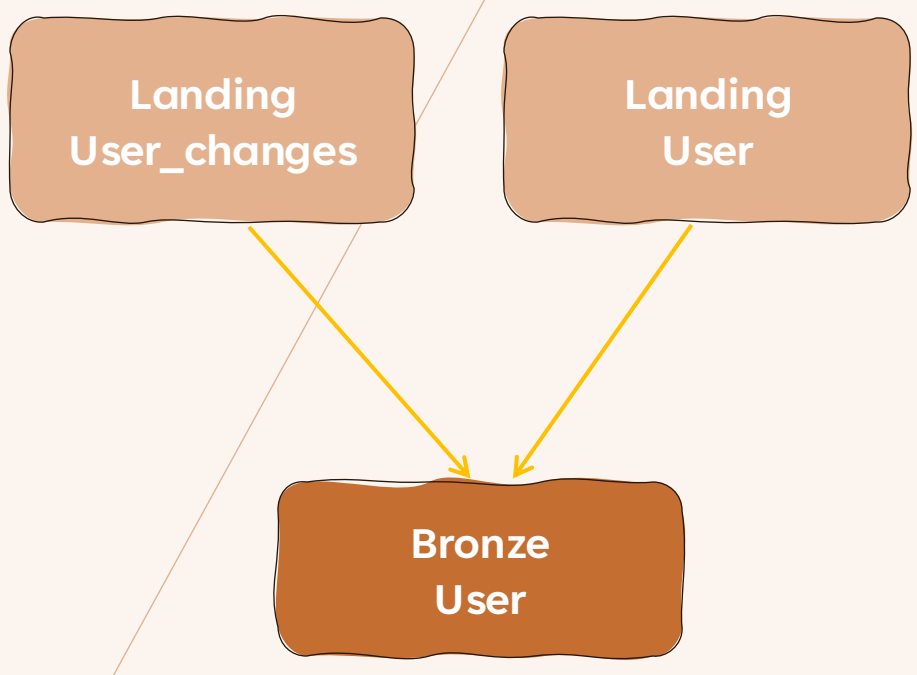
UNITY CATALOG

- More than a catalog
- Column level lineage, up & downstream
- Insights
- History → easy time travel
- Insights
- Access controls → row-level security
- Extensive data sharing capabilities(protocol, federation, clean rooms, api, market place)



OUR DATA

- Two types of tables: SCD1 and SCD2
- Main table/files
- Change table/file
- Keep them as they are landing, merging them in Bronze



LET’S TAKE A LOOK

ML OPS ARCHITECTURE WITH DATABRICKS AND AZURE ML

Leveraging the Best of Both Platforms for Scalable Data Processing, Model Training, and Deployment

STAND-ALONE ML OPS

databricks

Data Preparation

- Playground
- Experiments
- Features
- Models
- Serving

(OR)

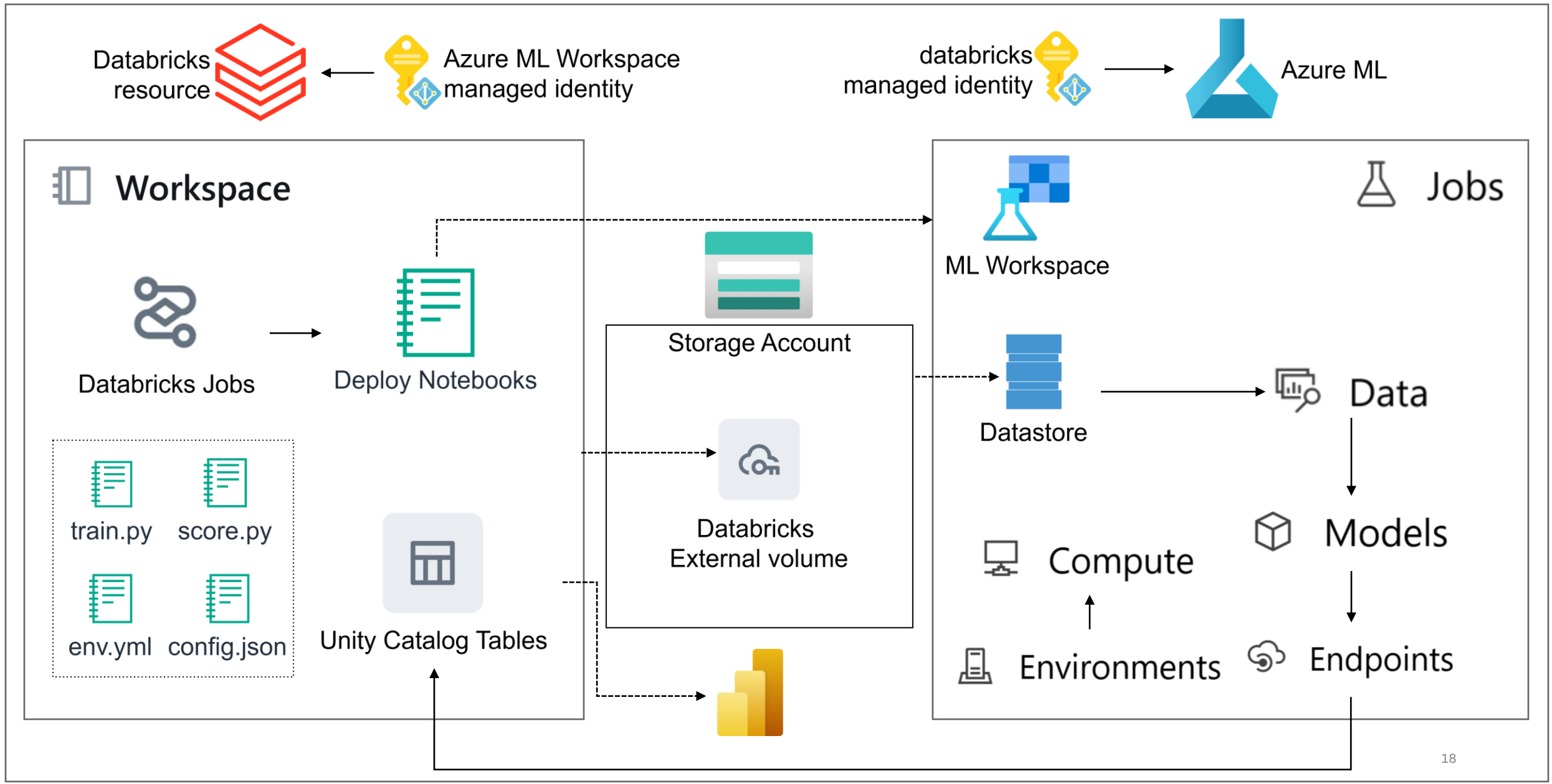
Azure ML

- Notebooks
- Automated ML
- Designer
- Prompt flow
- Compute
- Monitoring
- Data Labeling
- Linked Services
- Data
- Jobs
- Components
- Pipelines
- Environments
- Models
- Endpoints

HYBRID ML OPS



ML Ops Architecture with Databricks and Azure ML



LET’S TAKE A LOOK

THANK YOU!

Julia Førde

julia.forde@evidi.com

Sai Neelakantam

sai.neelakantam@evidi.com

Session feedback