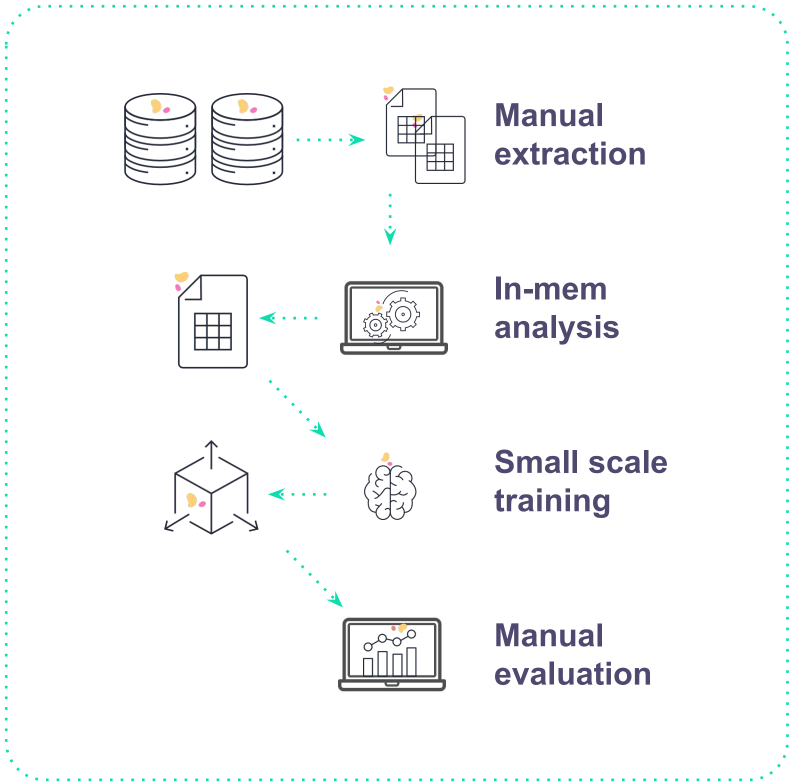
Feature Store: The Heart of Your Operational ML Pipeline

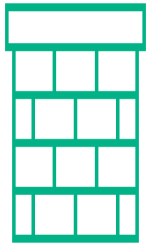
Yaron Haviv
CTO and Co-Founder
Iguazio

Most AI Projects Never Make it to Production

Research Environment

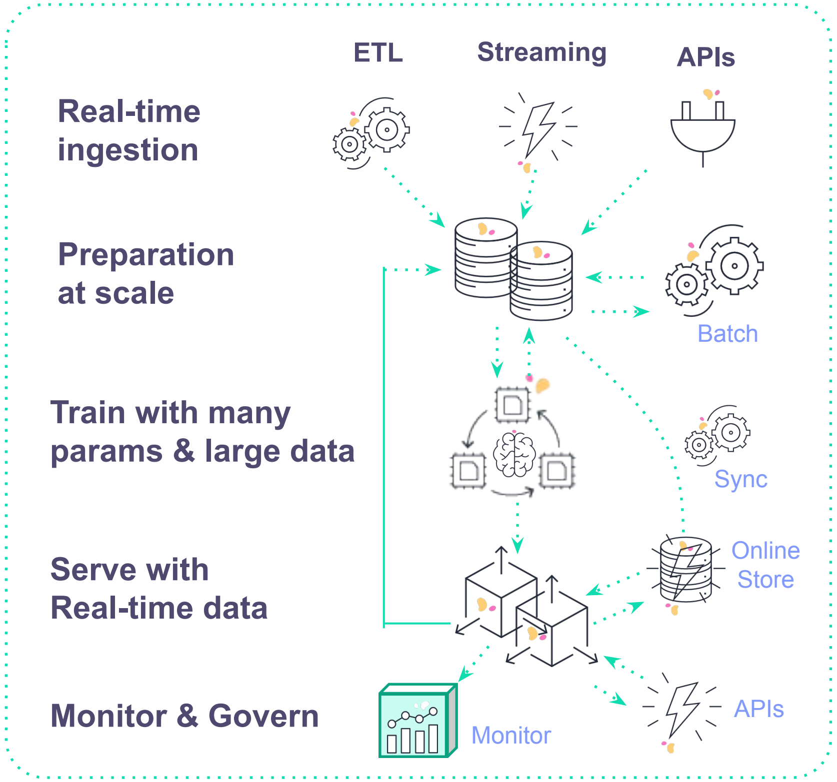


Data & Models



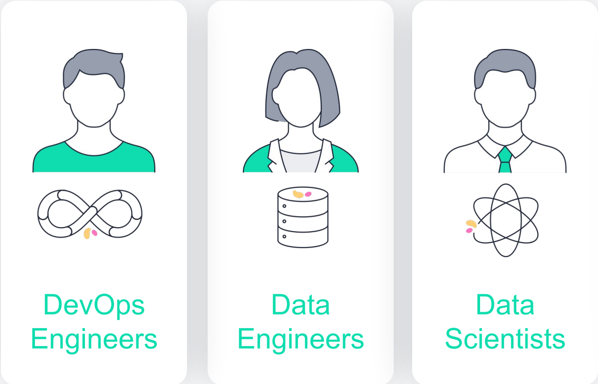
"Thrown over the wall"

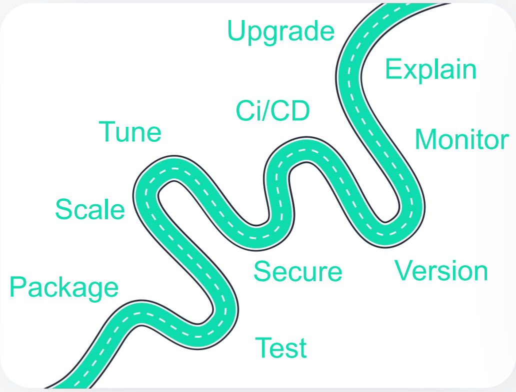
Production Pipeline

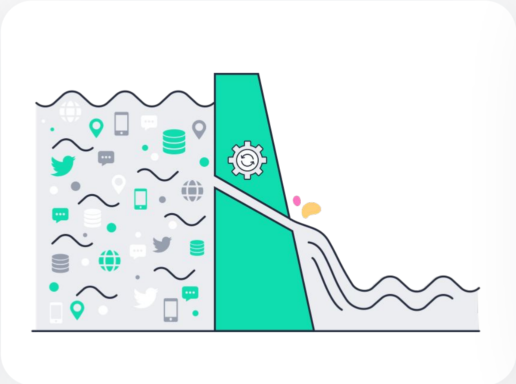


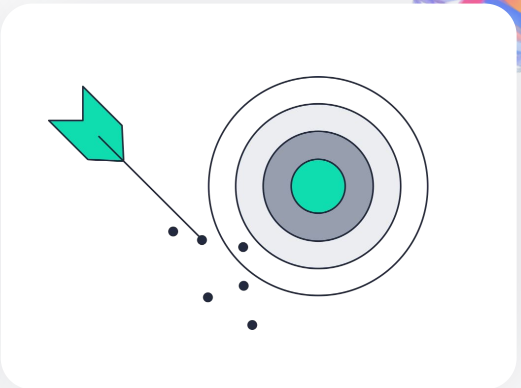
iguazio

Operationalizing Machine Learning is Challenging









Siloed Work

Re-implementation and lack of collaboration due to silos

Lengthy Process

Resource and time-consuming route from lab to production

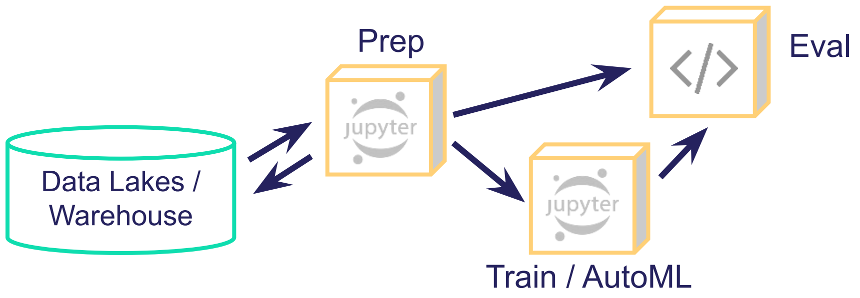
Access to Features

Accessing and preparing real-world features at scale

Model Accuracy

Tracking, maintaining and explaining model accuracy

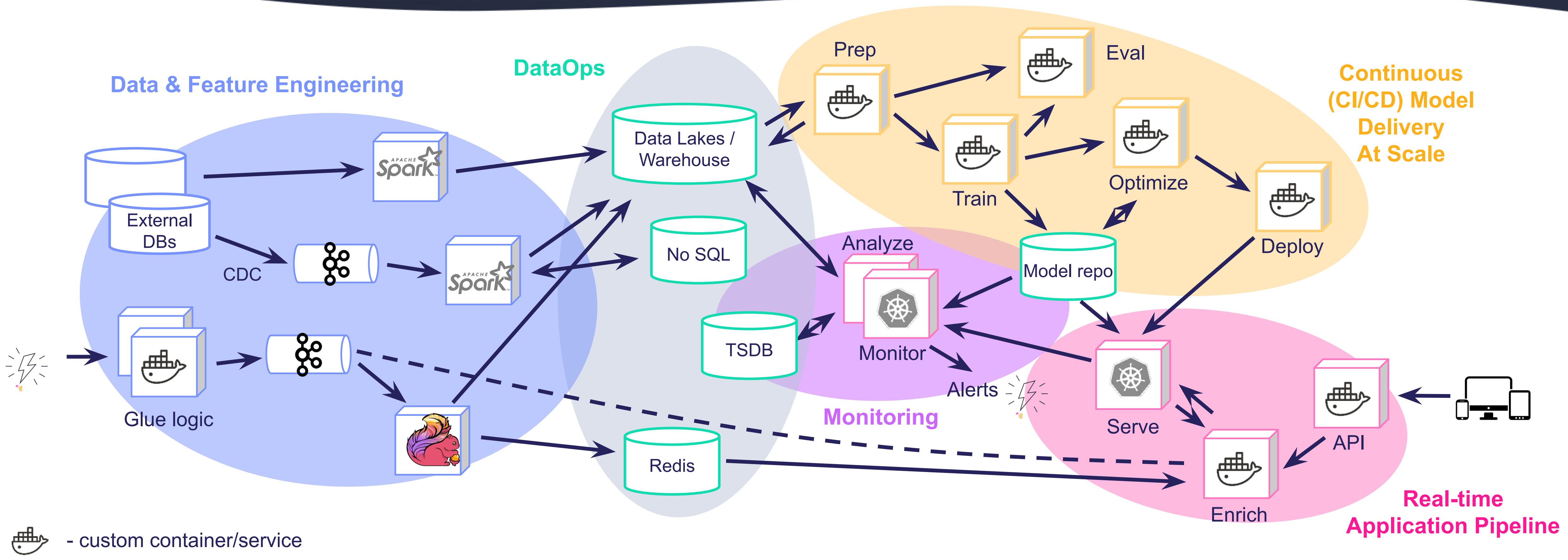
ML/AI Research Projects



Interactive / Iterative model development

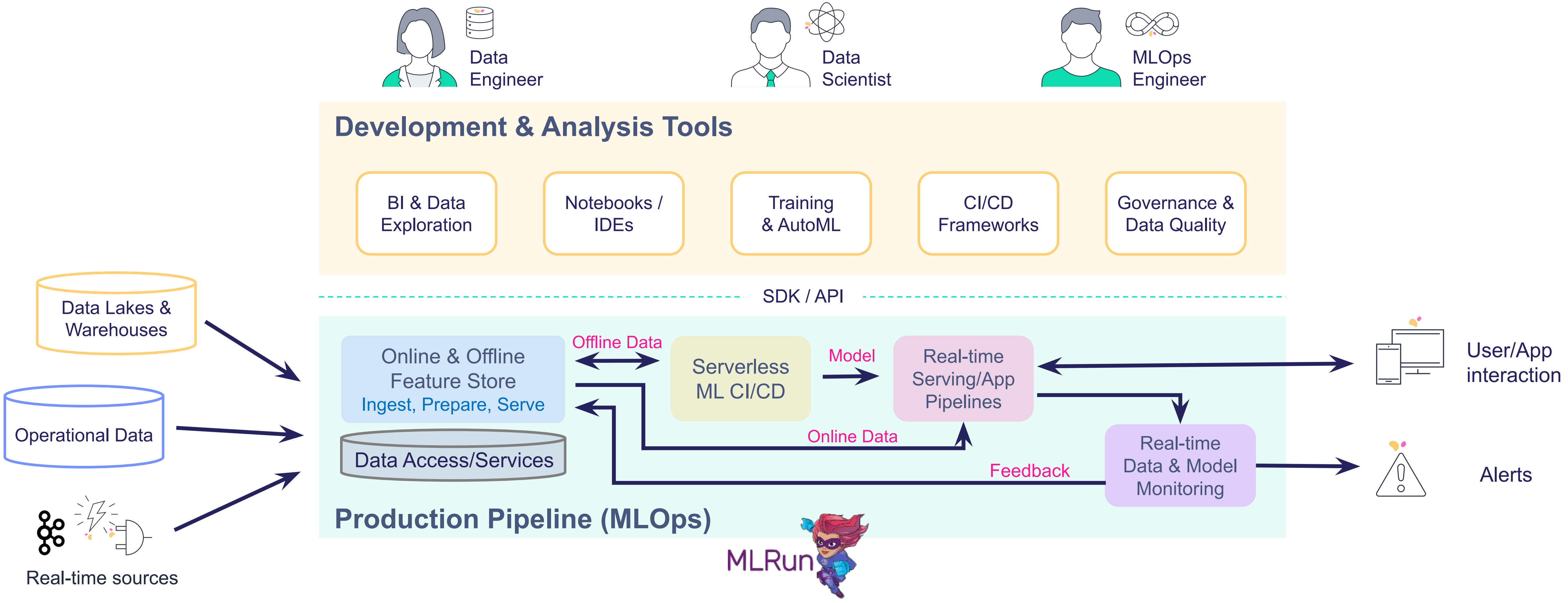
ML/AI Projects start with focus on building models
With a small data science team

But Productizing ML Is Exponentially Harder



Productizing AI/ML takes ages and requires an army of engineers

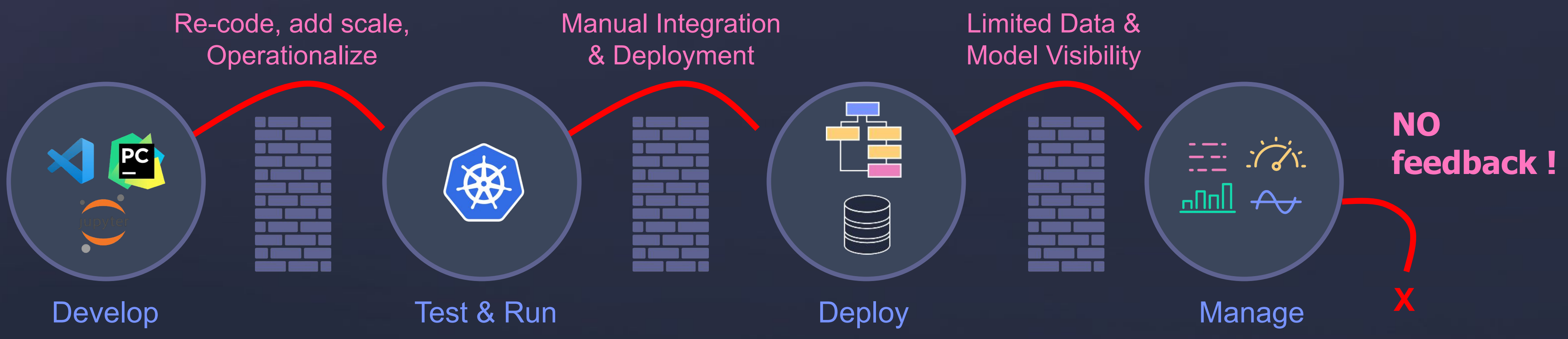
Accelerate Data Science to Production By Adopting Automation and a Production-first Mindset



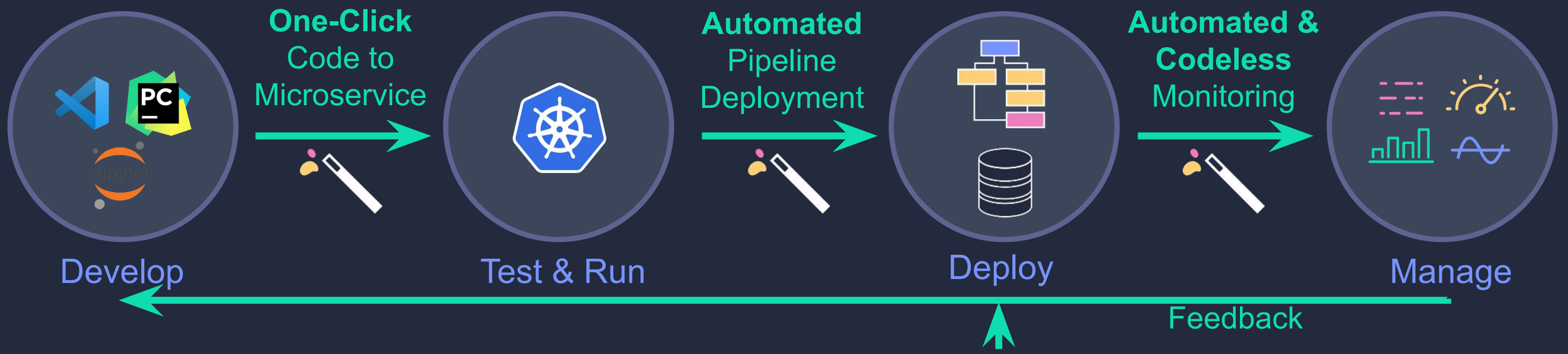
Operationalize ML/AI 12X Faster

iguazio

Before: Siloed, Complex and Manual Process



With iguazio : Automated, Fast, and Continuous



12X Faster

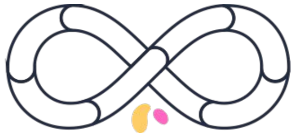
MLRun’s Key Components For MLOps Acceleration

#1

Feature Store

Automated offline & online feature engineering for real-time and batch data

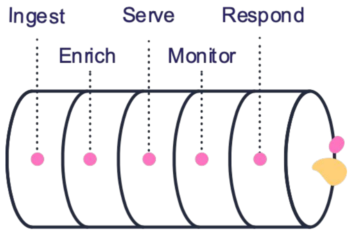
#2



CI/CD for ML

End to end MLOps automation. Integrated with mainstream ML, Git & CI/CD Frameworks

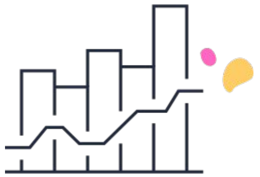
#3



Real-Time Serving Pipeline

Rapid deployment of scalable data and ML pipelines using real-time serverless technology

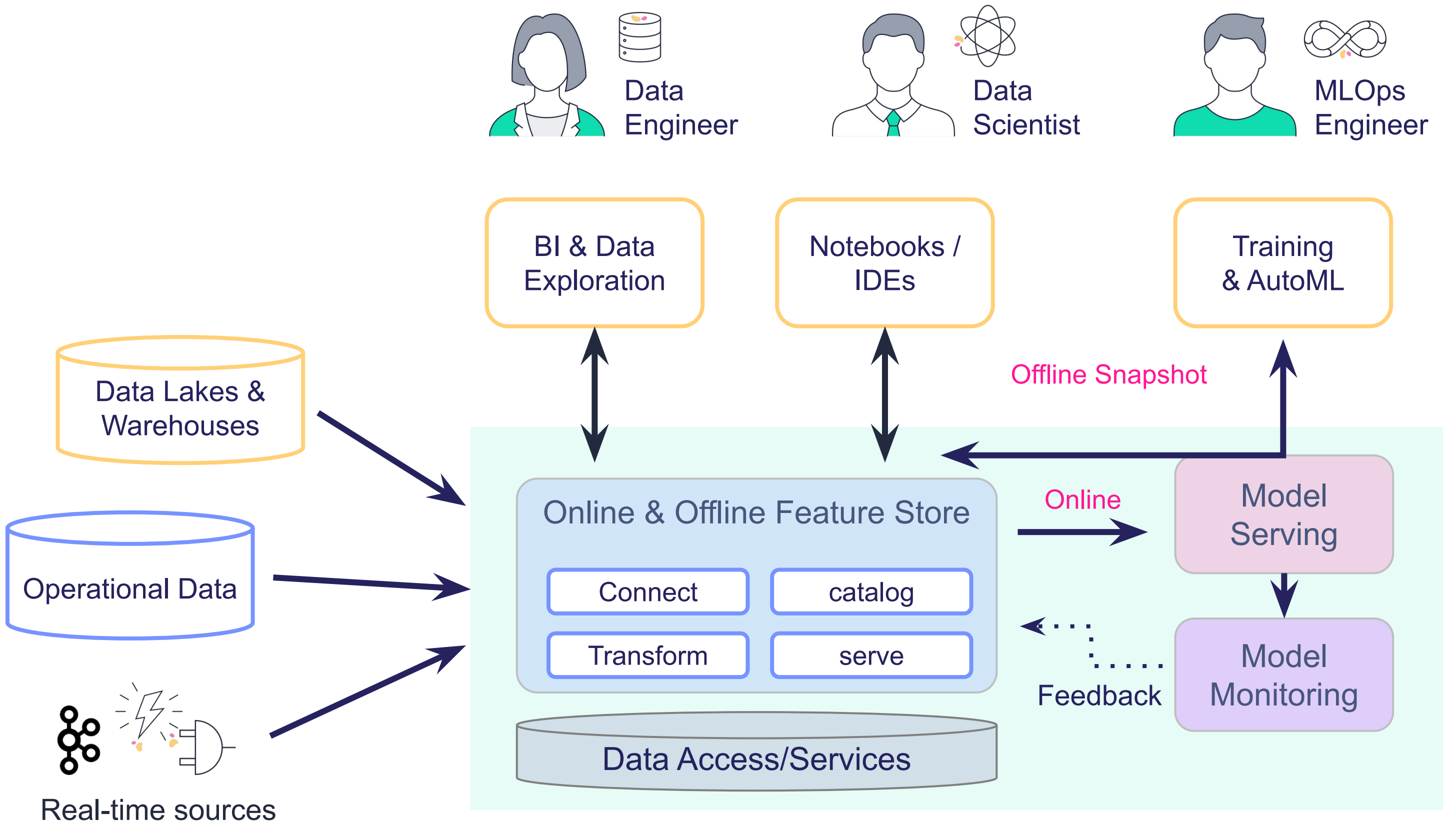
#4



Monitoring & Re-Training

Codeless data & model monitoring, drift detection & automated remediation/re-training

Feature Store - Automated Offline & Online Feature Engineering at Scale



Operational Pipeline

Benefits:

- Fast, simple and scalable way to build features from production data
- Implement once, use in training, real-time serving and monitoring
- Share and re-use features across teams and projects
- Glue-less integration with data and model monitoring
- Enable re-training directly from production data

Implementing A SINGLE Feature Using SQL

```
CREATE TABLE recency_feature_group_1 AS
(
    SELECT *,
            CASE WHEN std_interval_between_group_1_in_days IS NULL OR std_interval_between_group_1_in_days = 0
               THEN NULL
            ELSE mean_interval_between_group_1_in_days/std_interval_between_group_1_in_days
            END as cv_interval_group_1
     FROM
     (
        SELECT user_id,
               AVG(time_interval)/(3600*24) as mean_interval_between_group_1_in_days,
               stddev_pop(time_interval)/(3600*24) as std_interval_between_group_1_in_days
        FROM
        (
            SELECT user_id,
                   event_timestamp,
                   extract(epoch from event_timestamp)
                     - lag(extract(epoch from event_timestamp))
                     over (PARTITION BY user_id order by event_timestamp)
                     as time_interval
            FROM  "My_big_transactional_table"
            WHERE event_timestamp::timestamp <=
                    (cast( ' "2014-09-01 00:00:00" ' as date) - INTERVAL '7' DAY)
               AND
                  event_timestamp::timestamp >=
                     (cast( ' "2014-09-01 00:00:00" ' as date)
                       - INTERVAL '7' DAY - INTERVAL '6' MONTH)
        ) as table_layer_2
        GROUP BY user_id
     ) as table_layer_3
)
```

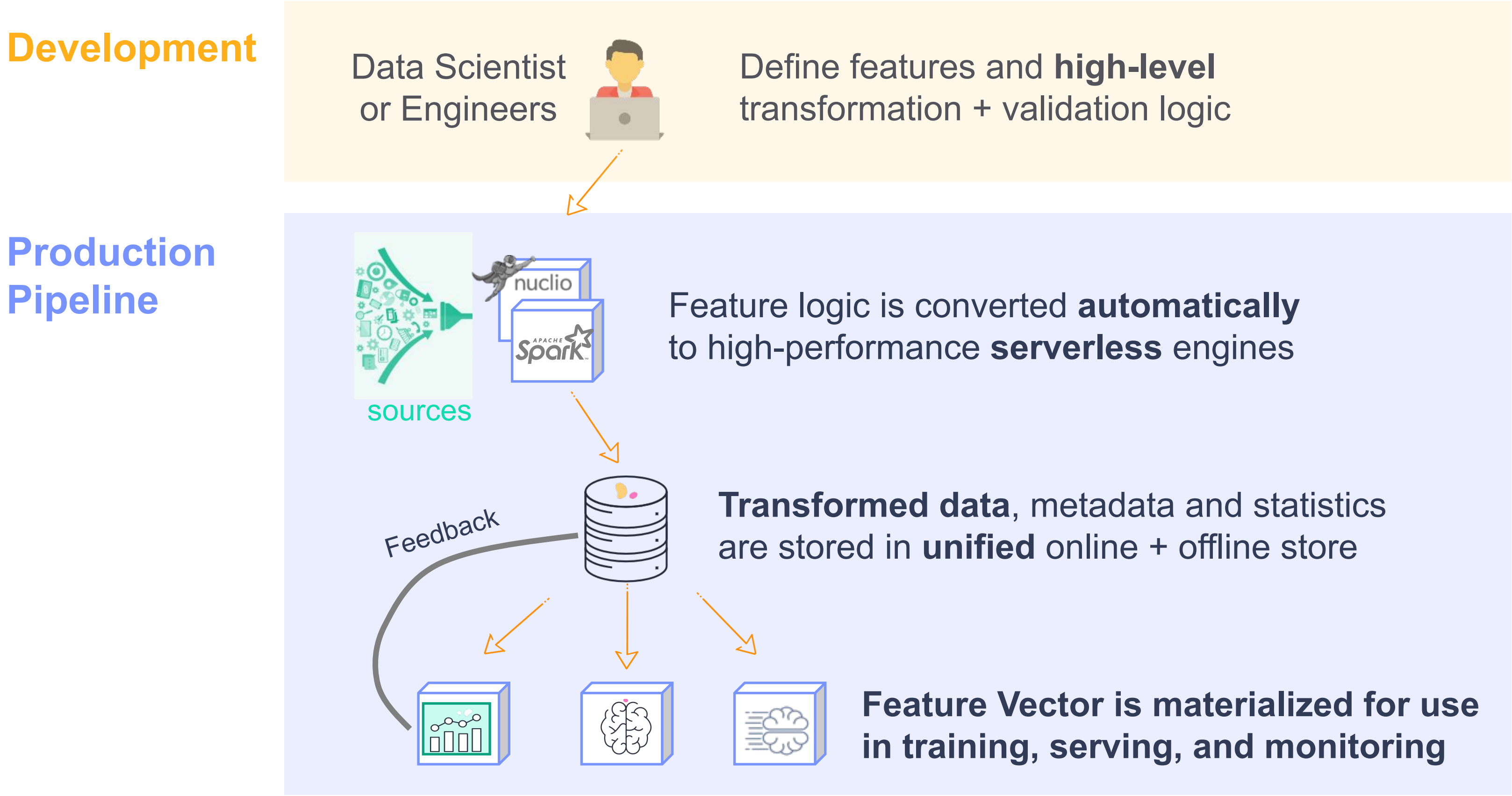
Very Complex

Slow and Resource Intensive

Won’t work in real-time

iguazio

MLRun Feature Store - How Does It Work?



- ✔ **Abstract**
- ✔ **Developed once**
- ✔ **Central** and **Consistent**
- ✔ **Glue-less** integration
- ✔ **High-performance**

Save Time And Resources !

Key Challenges for Online Feature Engineering

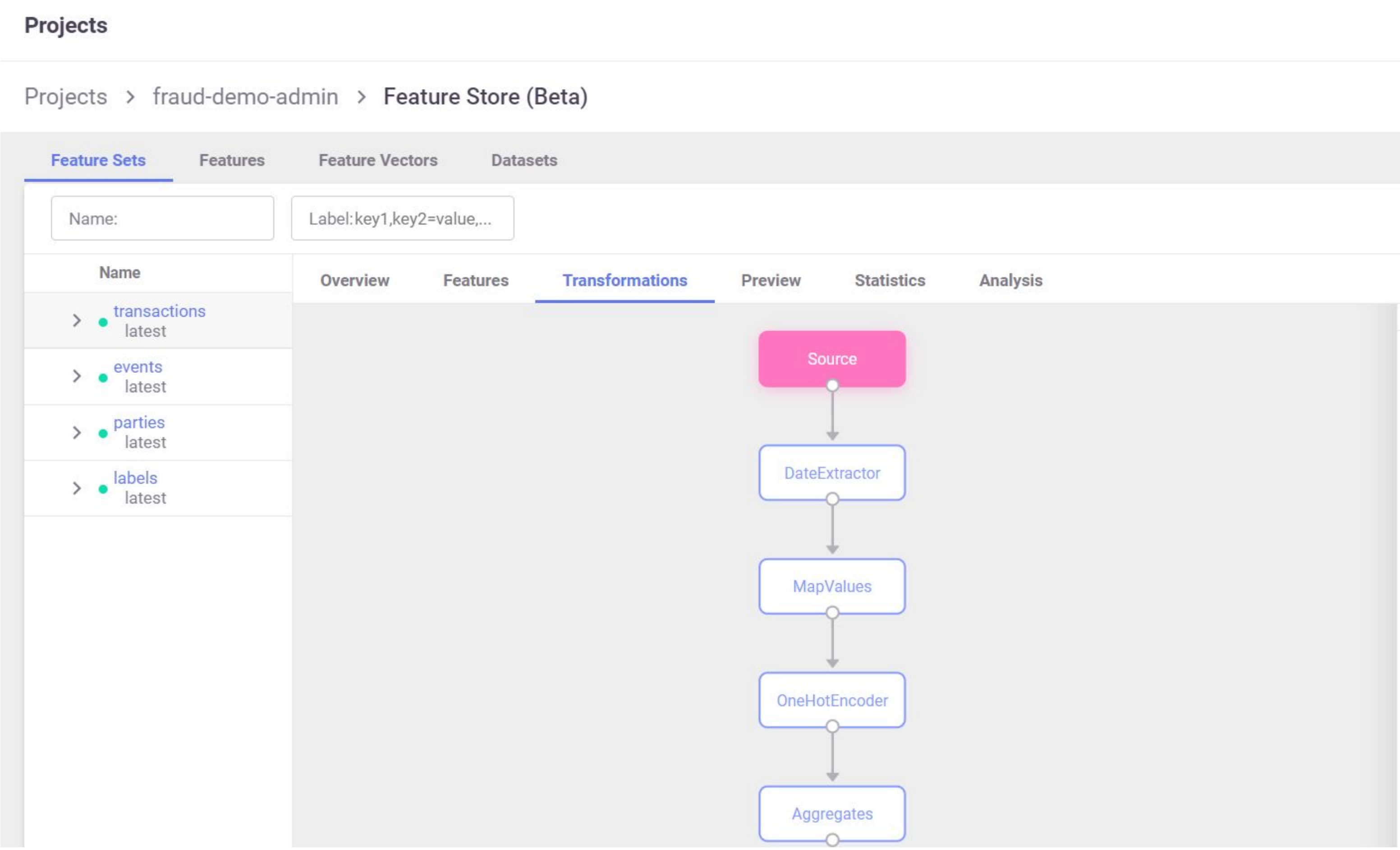
- **Data scientists are not data engineers**
  - Re-written code is needed to deploy it in production
  - Working with streaming sources as opposed to parquet files
- **Performance** - Calculate features in real time on live data at scale
- **Robust transformation –** e.g. aggregations on sliding windows
- **Enrichment** – Enrich real time events with historical / operational data
- **Consistency -** between training and serving
- **Data drift** - based on feature drift
- **Feature reuse –** use features for many projects
- **Feature versioning –** aligning feature and model version in production

Simple SDK for Creating a Real-Time Transformation

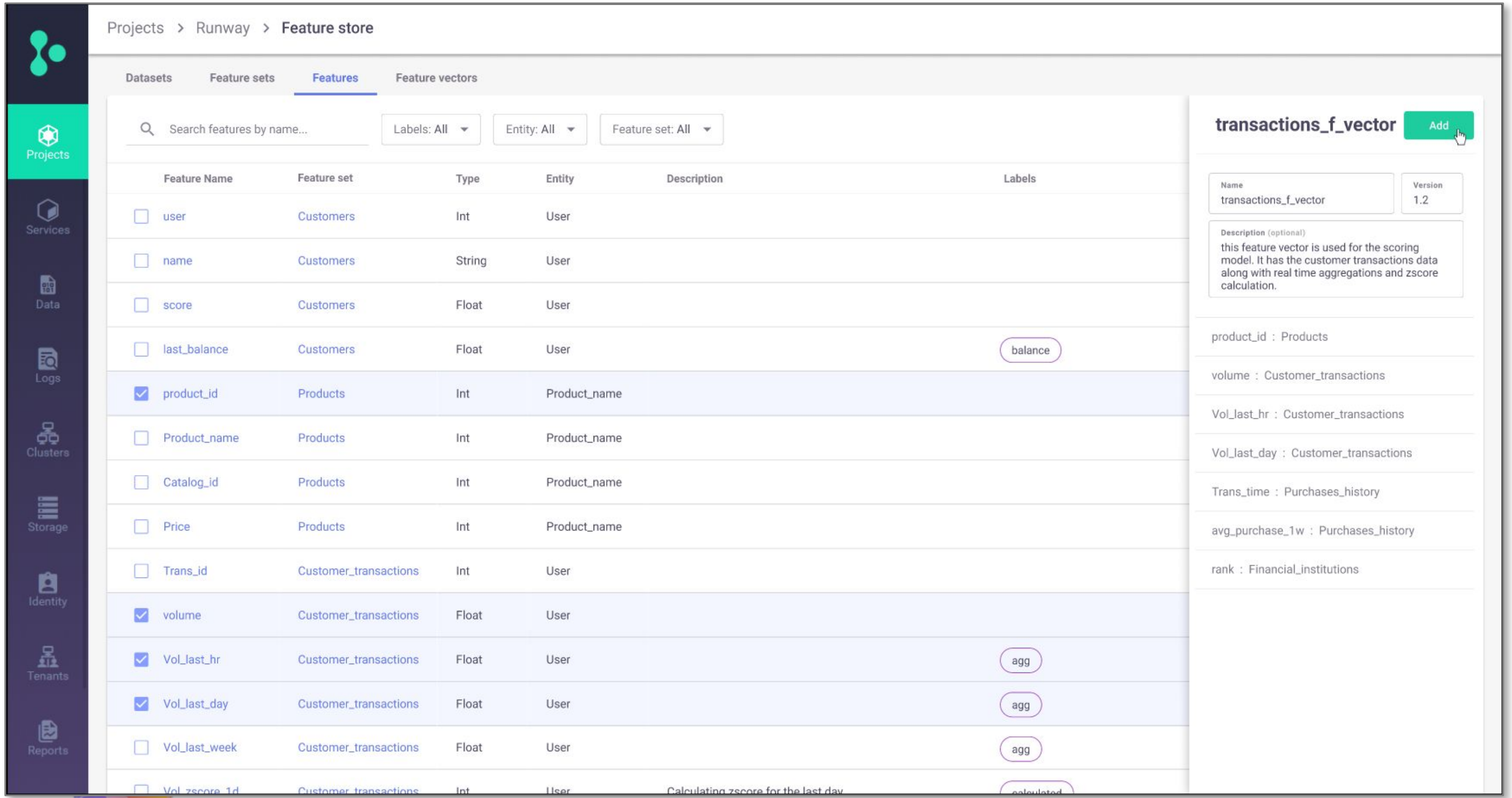
Quickly develop ML/DL features for offline and online/real-time use

```python
transaction_set.graph\
    .to(DateExtractor(parts = ['hour', 'day_of_week'], timestamp_col = 'timestamp'))\
    .to(MapValues(mapping={'age': {'U': '0'}}, with_original_features=True))\
    .to(OneHotEncoder(mapping=one_hot_encoder_mapping))


# Add multiple aggregations on multiple time windows
windows = ['2h', '12h', '24h']
transaction_set.add_aggregation(name=f'amount',
                                column='amount',
                                operations=['avg','sum', 'count','max'],
                                windows=windows,
                                period='1h')
```



Assemble online & offline features from catalog



Trivial Access APIs

Offline (Training & Exploration)

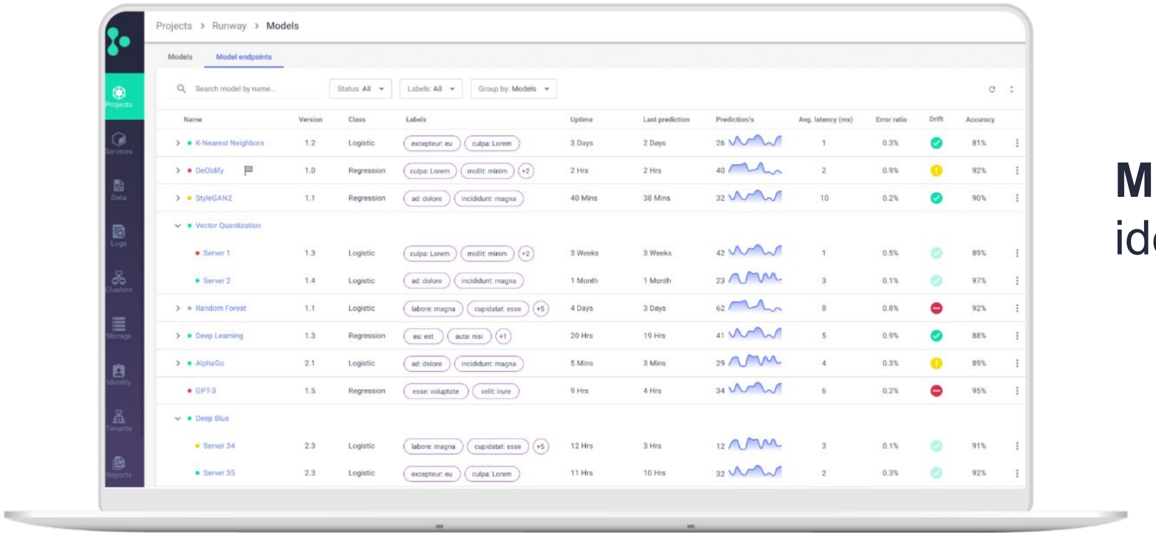
```
resp = client.get_offline_features(vector)
df = resp.to_dataframe()
```

Real-time (Serving & monitoring)

```
service = client.get_online_feature_service(vector)
service.get([{"patient_id": "838-21-8151"}])
```


iguazio

Integration with Model Monitoring
Drift Detection & Auto-Retraining

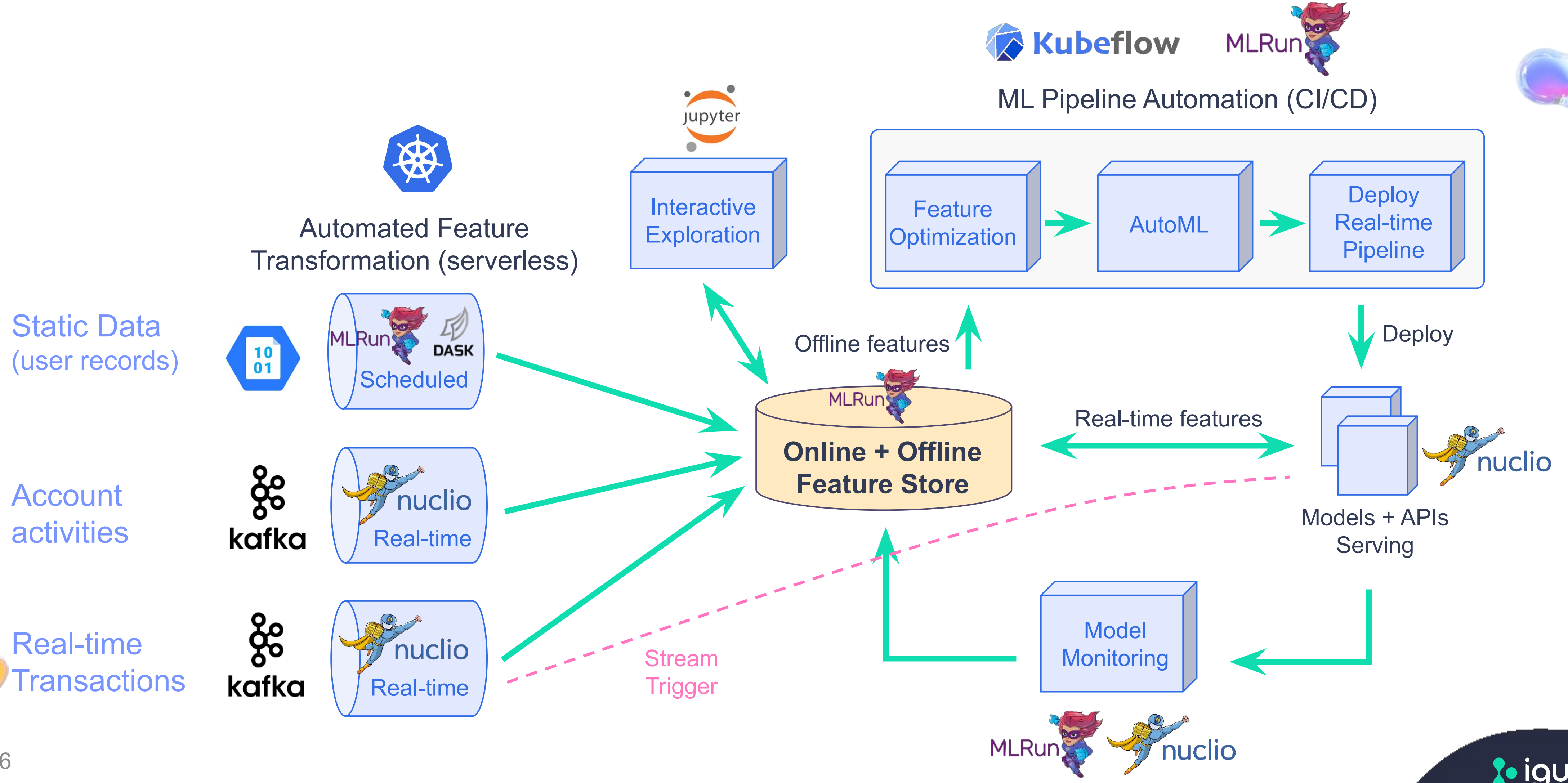


Monitor your models in production, identify and mitigate **drift** on the fly

Machine Learning Pipeline

Detect **model drift** based on **feature drift** via the integrated feature store and start retraining

Live Demo:
ML Production Pipeline - Real-Time Fraud Predictions

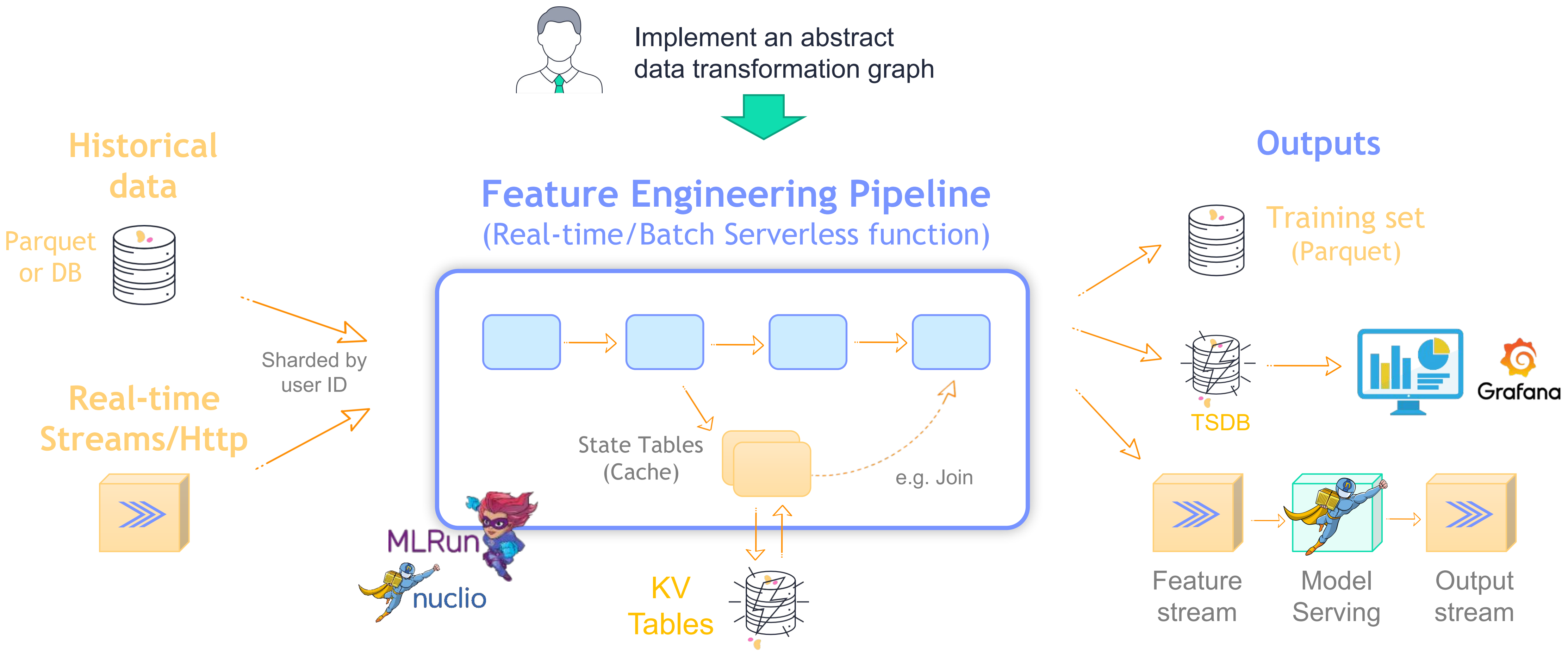


iguazio

iguazio

Demo

Serverless Stream Processing For Real-Time & Batch



Simplicity, Performance, and Scale

Thank you!

Do you have any questions?

Yaron Haviv
yaronh@iguazio.com

www.linkedin.com/in/yaronh/

@yaronhaviv