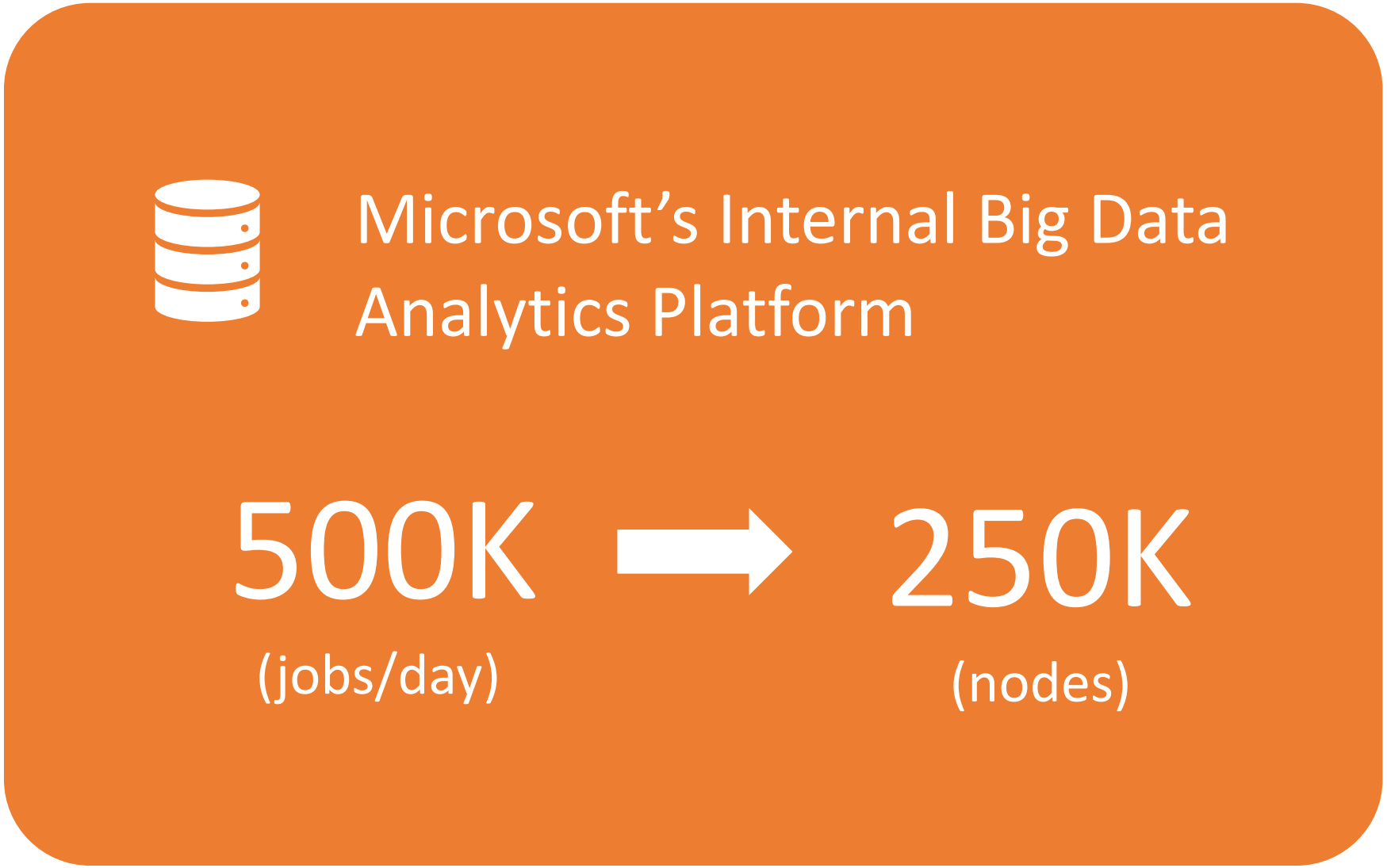
# Griffon: Reasoning about Job Anomalies with Unlabeled Data in Cloud-based Platforms

**Liqun Shao, Yiwen Zhu**, Siqi Liu*, Abhiram Eswaran, Kristin Lieber, Janhavi Mahajan, Minsoo Thigpen, Sudhir Darbha, Subru Krishnan, Soundar Srinivasan, Carlo Curino, Konstantinos Karanasos

Microsoft, *University of Pittsburgh


Microsoft’s Internal Big Data Analytics Platform
500K
(jobs/day)
250K
(nodes)


https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwjv9uXU0__lAhWtIDQIHaU0ABwQjB16BAgBEAM&url=https%3A%2F%2Fwww.intellectualtakeout.org%2Farticle%2Fkanye-wests-private-firefighting-force-good&psig=AOvVaw2pinteqP1A7uhZRdXBfq0J&ust=1574575139344414

# My job is SLOW …

# My job is SLOWER…

# On-Call Support Engineer Workflow


57 mins

88 mins

Identify job slowdown causes

End-to-End deployed and used

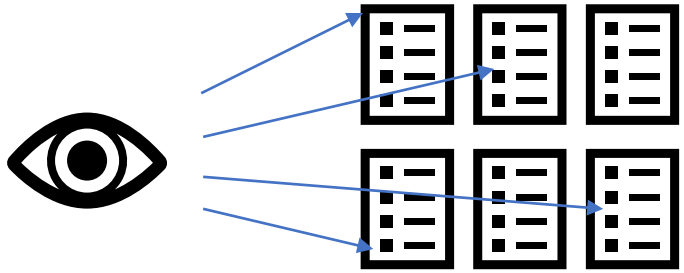
Consistent results validated by domain experts

Drops the investigation time

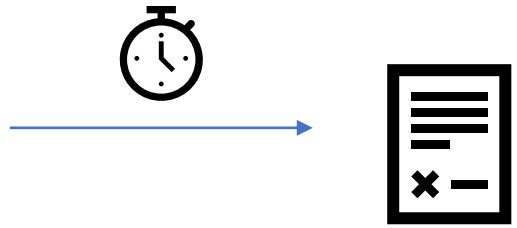
# Griffon: Before and After

## Before Griffon

A job goes out of service-level objectives (SLO) and the engineer is alerted



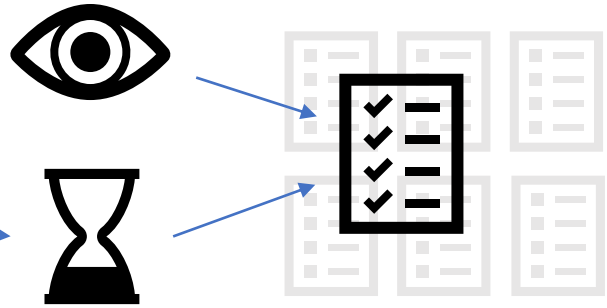
An Engineer spends hours of manual labor looking through hundreds of metrics



After 2-3 days of investigation, the reason for job slowdown is found.

## After Griffon

A job goes out of SLO and the engineer is alerted



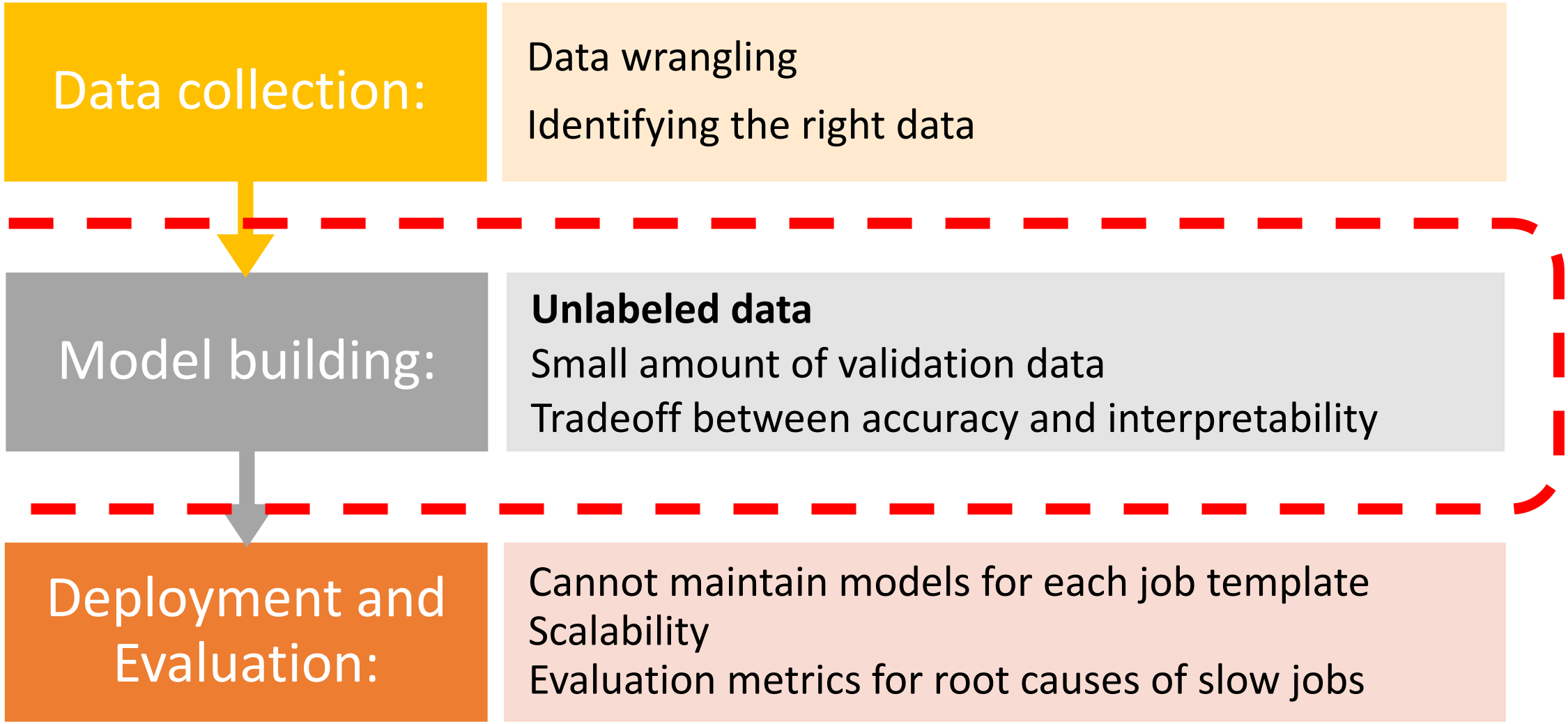
The Job ID and VC is fed through Griffon and the top reasons for job slowdown are generated automatically

The reason is found in the top five generated by Griffon.

All the metrics Griffon has looked at can be ruled out and the engineer can direct their efforts to a smaller set of metrics.

# Griffon

- ML Methodology
- System Architecture


Data collection:
Data wrangling
Identifying the right data
Model building:
Unlabeled data
Small amount of validation data
Tradeoff between accuracy and interpretability
Deployment and Evaluation:
Cannot maintain models for each job template
Scalability
Evaluation metrics for root causes of slow jobs

# Challenges

# Identify Job Slowdown Reasons

Job Runtime Predictor

Feature Contributions

# Job Runtime Prediction

Job Runtime Predictor

| MARE | LR | RF | GBT | DNN |
|---|---|---|---|---|
| Per-Template Model | 0.186 | 0.116 | 0.124 | 0.146 |
| Global Model | 0.235 | ***0.121*** | 0.277 | 0.353 |

# Feature Contributions

Reformulate decision tree models to linear models:

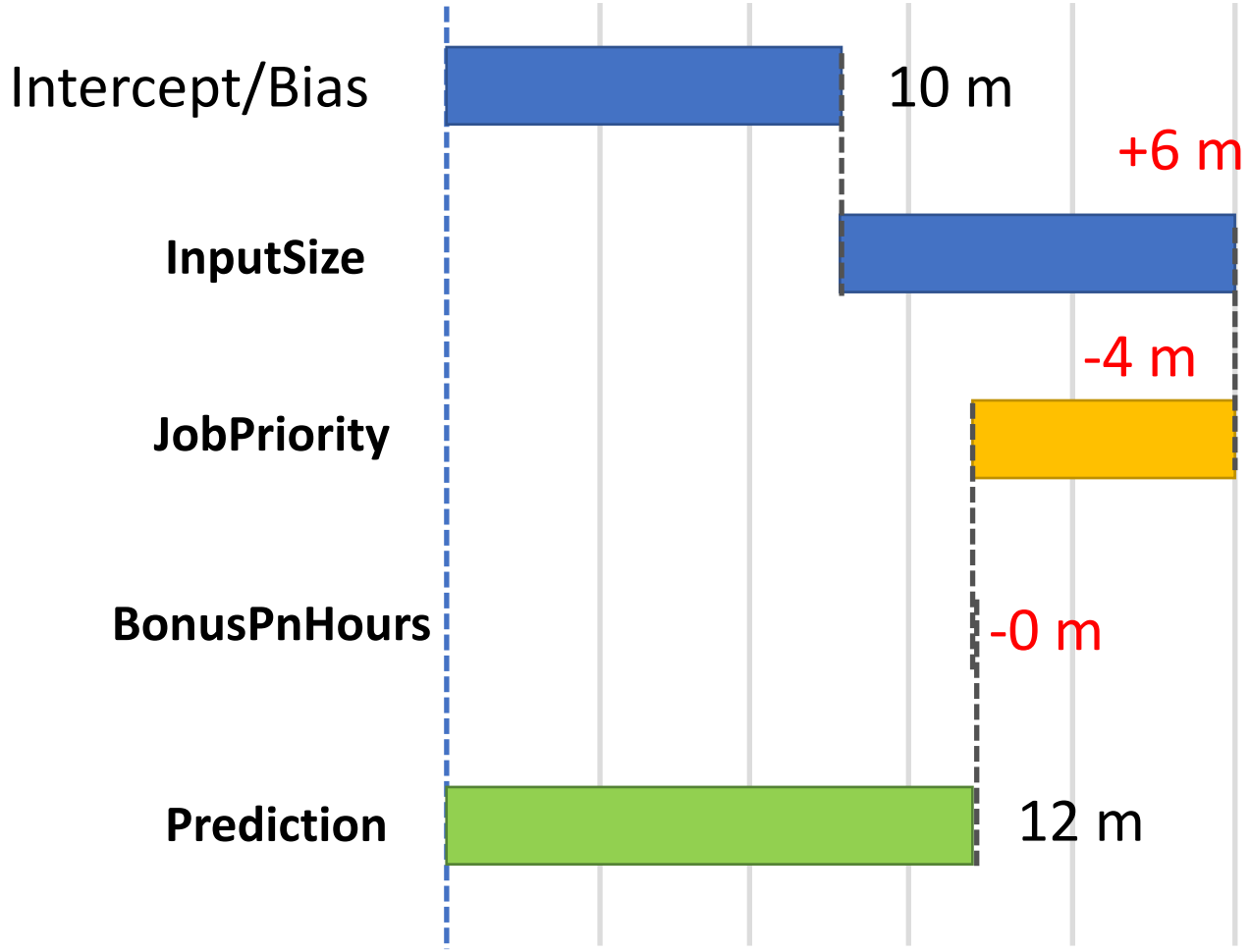
$$\hat{y} = c + \sum_{k=1}^{K} f c_k$$

Compare feature contributions to baseline predictions:

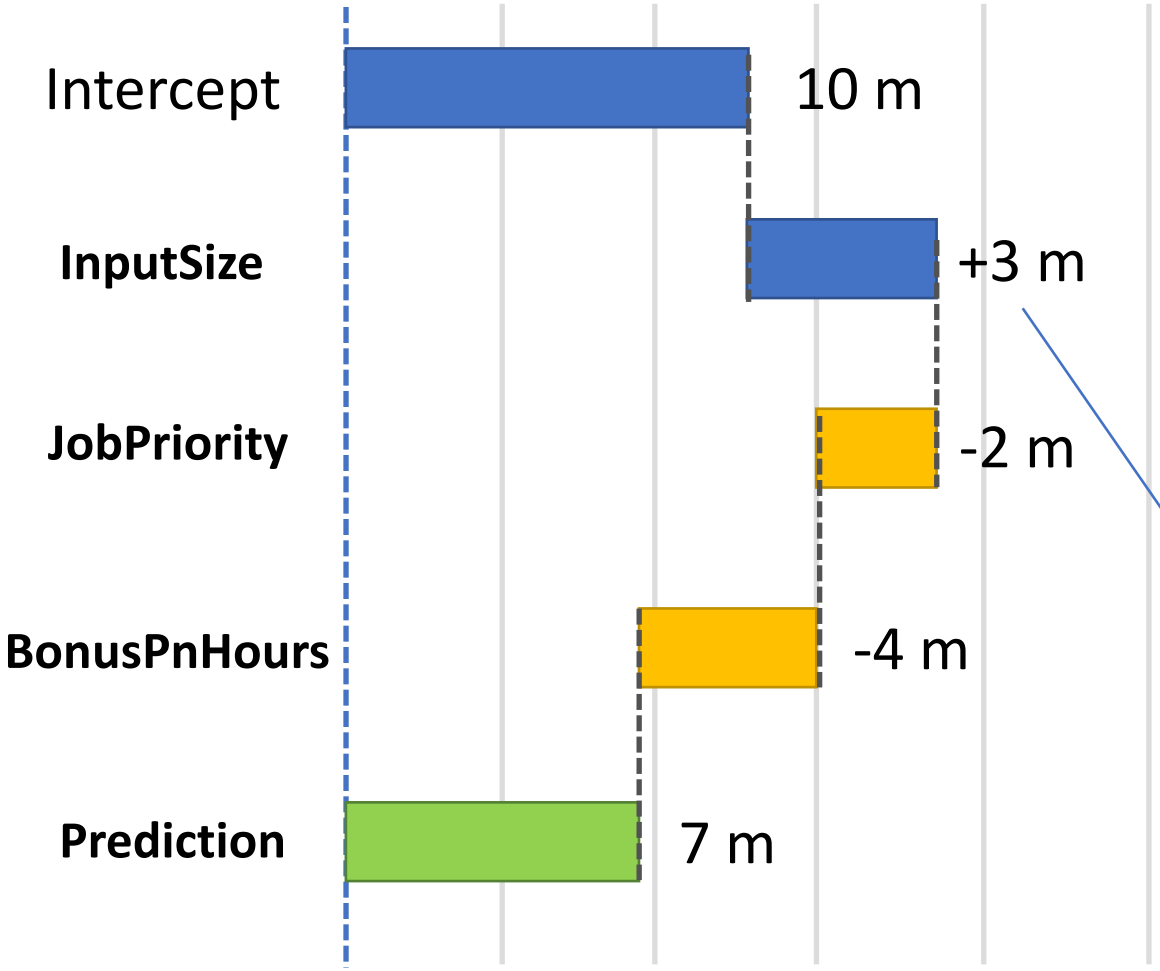
$$\hat{y} - \bar{\hat{y}}^{\beta} = \sum_{k=1}^{K} (f c_k - \bar{f} c_k^{\beta}) = \sum_{k=1}^{K} \boldsymbol{\Delta} f c_k$$

# Feature Contributions

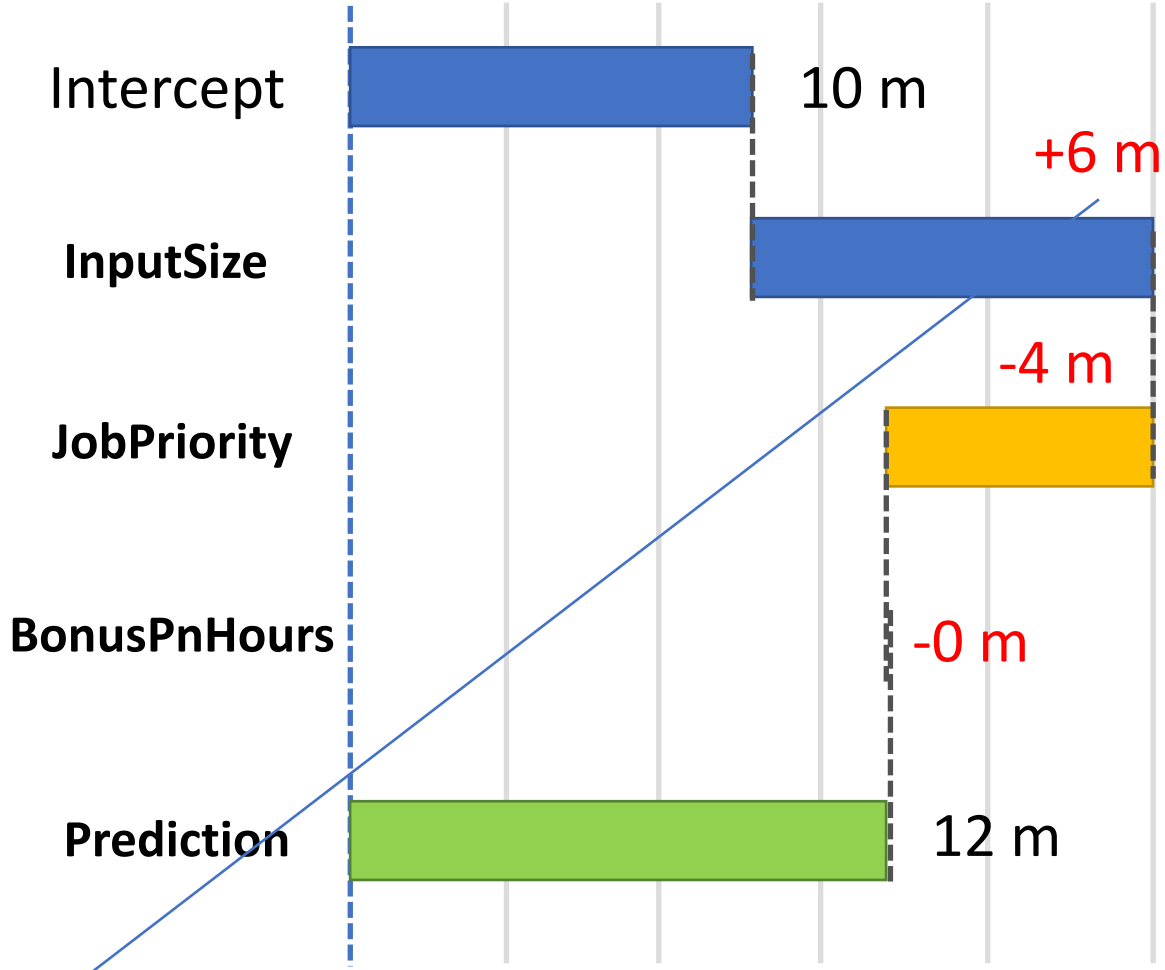
$$\hat{y} = c + \sum_{k=1}^{K} f c_k$$



$$\hat{y} - \bar{\hat{y}}^{\beta} = \sum_{k=1}^{K} (fc_k - \bar{f}c_k^{\beta}) = \sum_{k=1}^{K} \Delta fc_k$$




*Baseline Job*




*Slow Job*

**InputSize: 6-3 = 3**

**JobPriority: 4 -2 = 2**

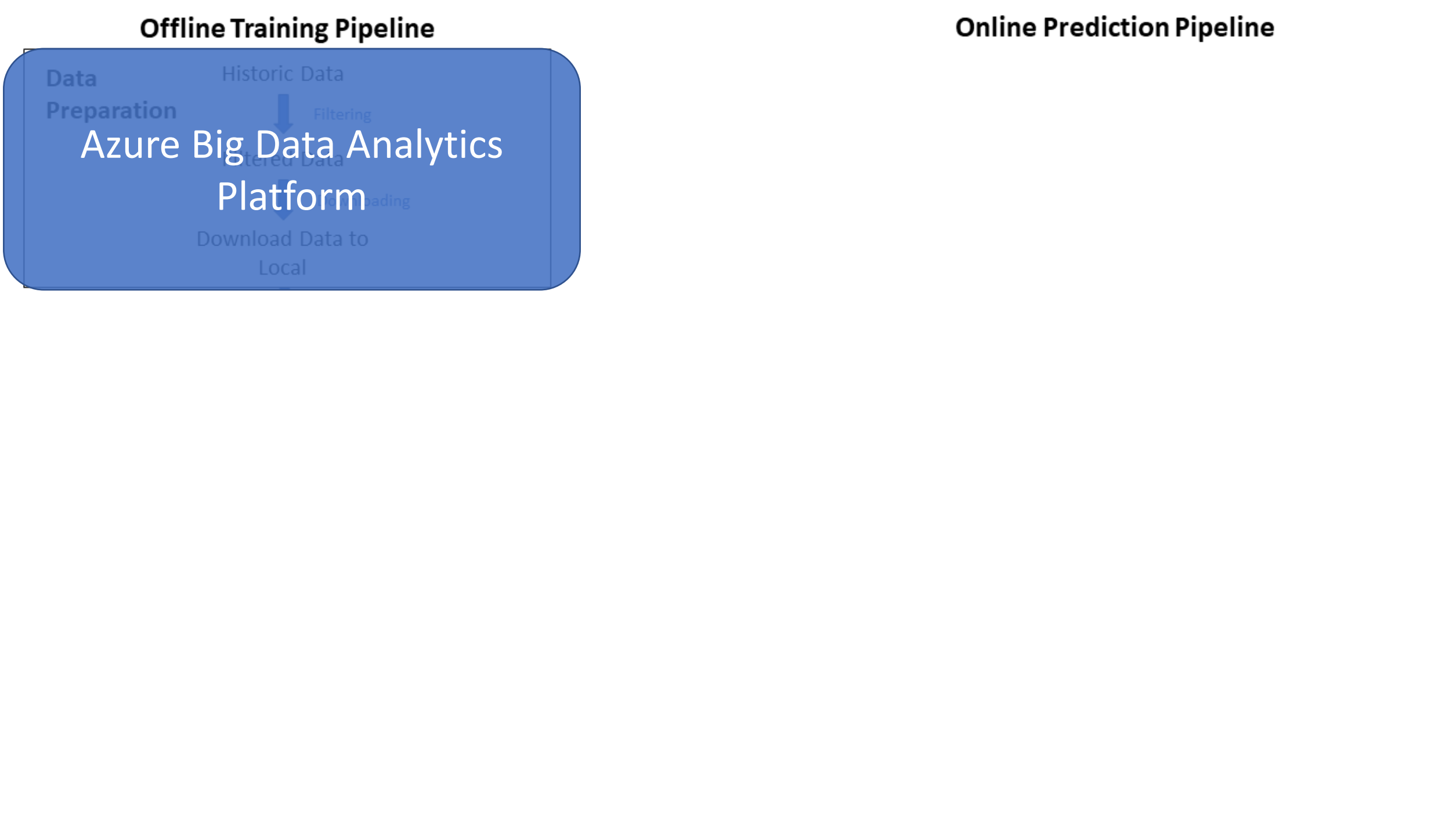
**BonusPnHours: 4 – 0 = 4**

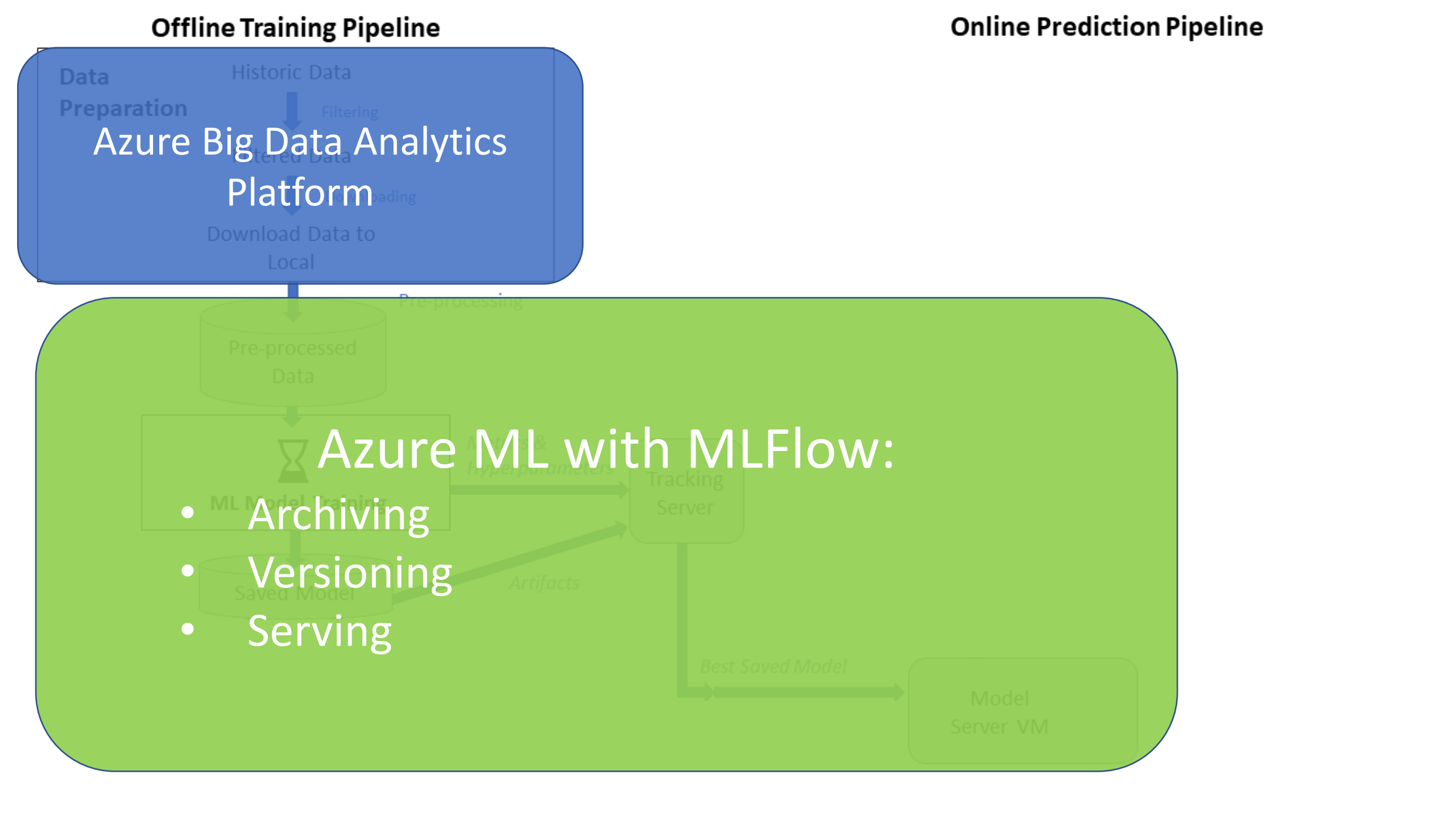
Offline Training Pipeline

Online Prediction Pipeline



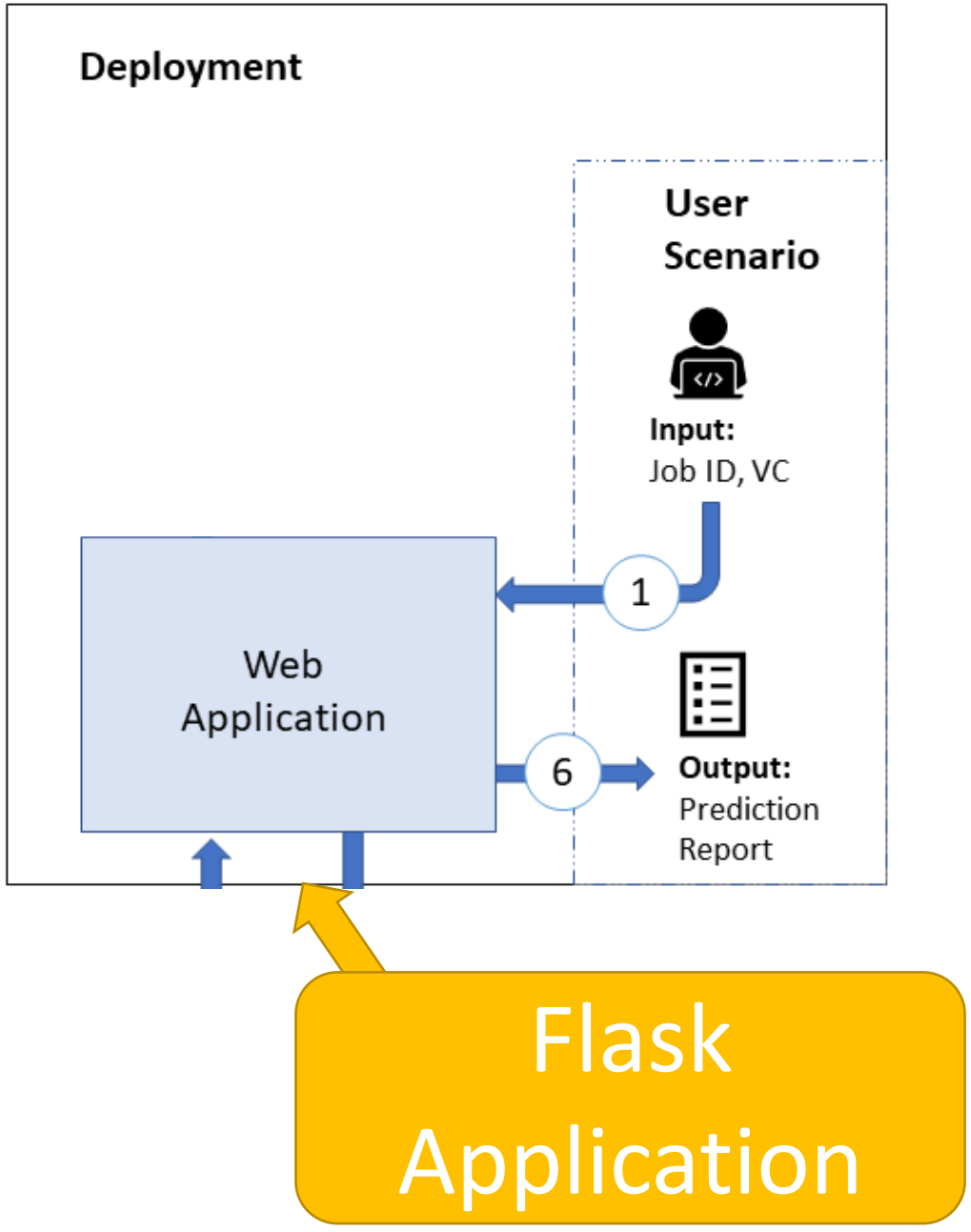
# Architecture


Offline Training Pipeline
Online Prediction Pipeline
Data Preparation
Historic Data
Filtering
Azure Big Data Analytics Platform
Download Data to Local


Offline Training Pipeline
Online Prediction Pipeline
Data Preparation
Historic Data
Filtering
Filtered Data
Download Data to Local
Pre-processing
Pre-processed Data
Tracking Server
Artifacts
Best Saved Model
Model Server VM
Azure Big Data Analytics Platform
Azure ML with MLFlow:
Archiving
Versioning
Serving

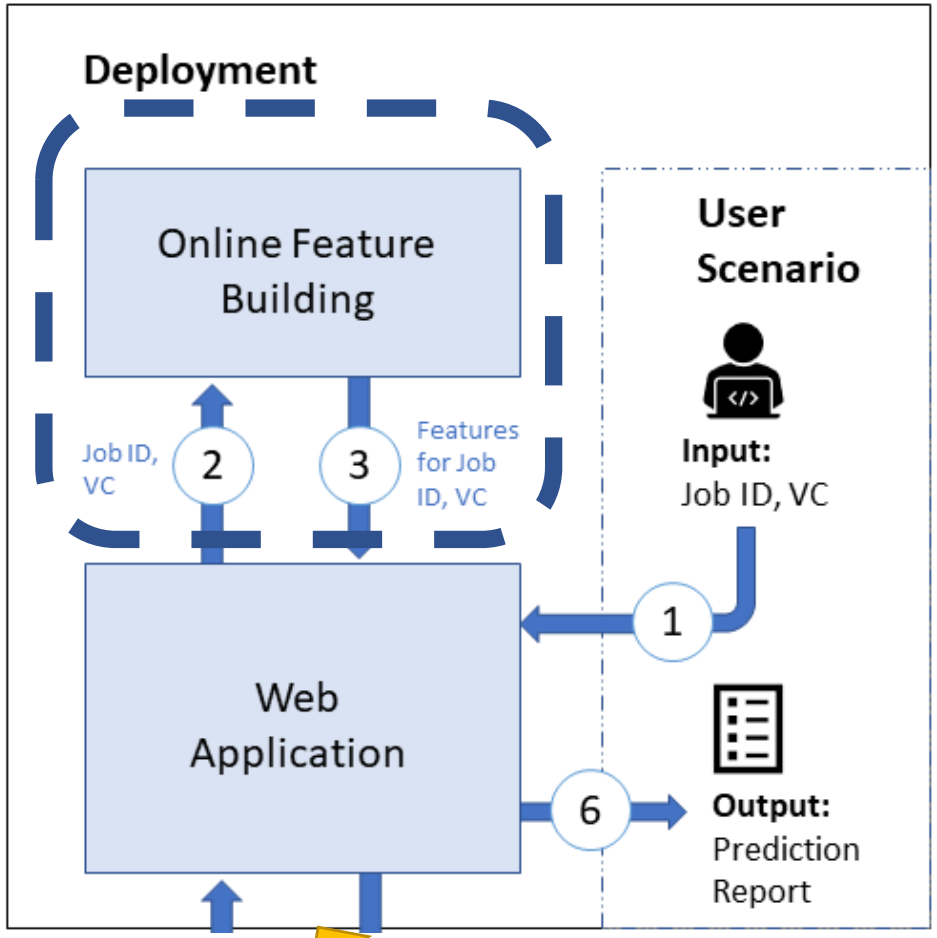
## Offline Training Pipeline

## Online Prediction Pipeline

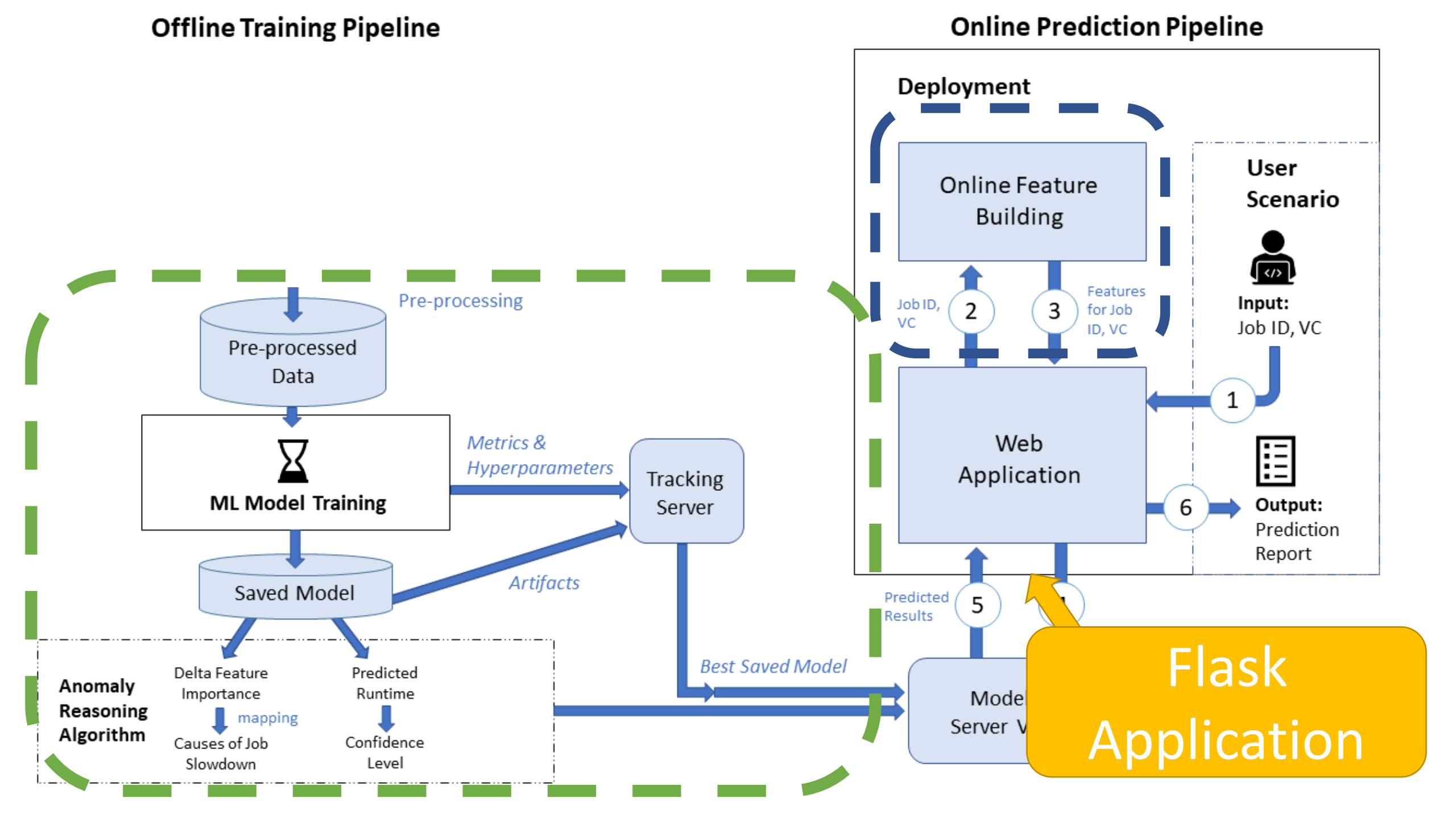


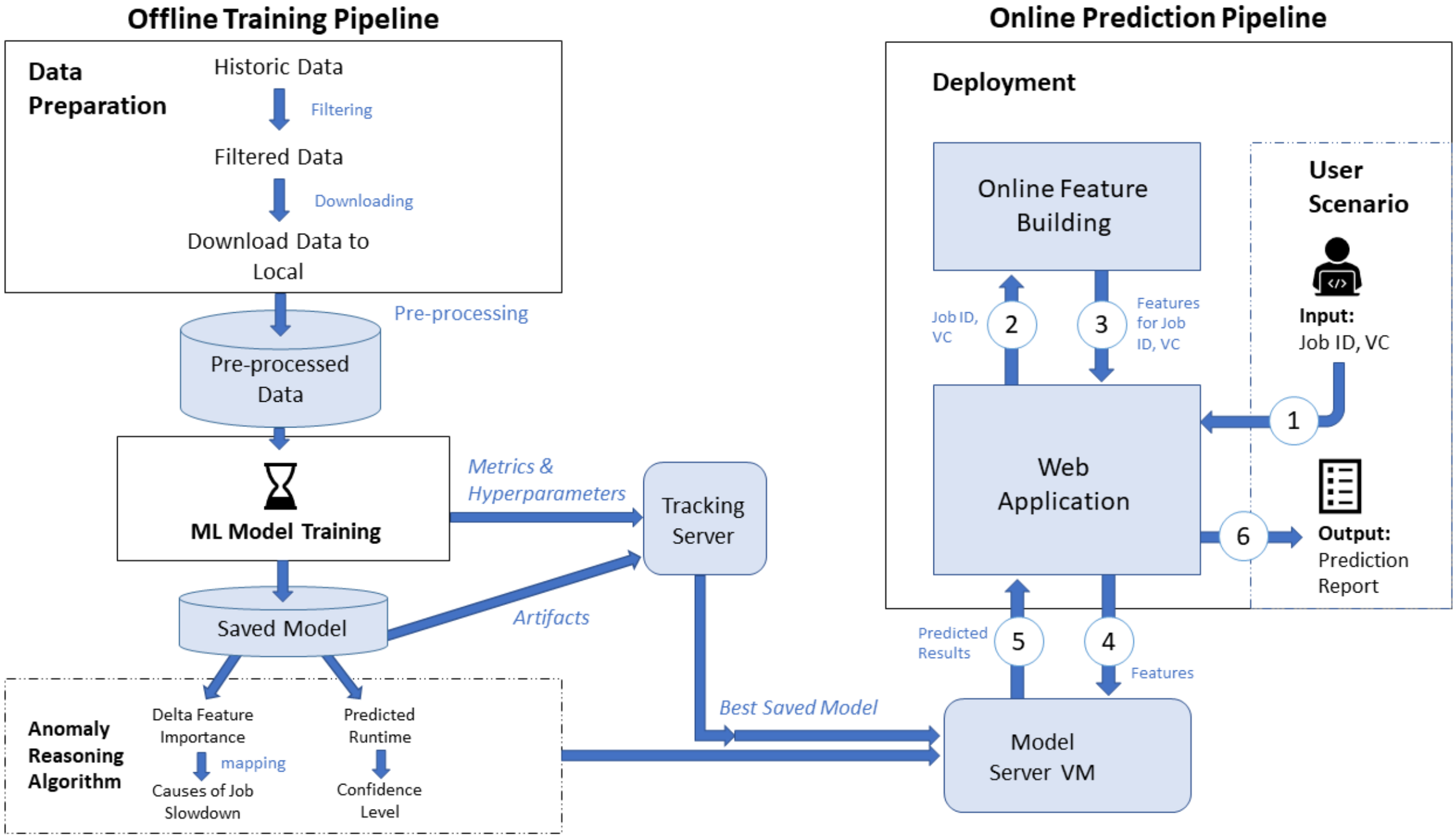
## Offline Training Pipeline

## Online Prediction Pipeline

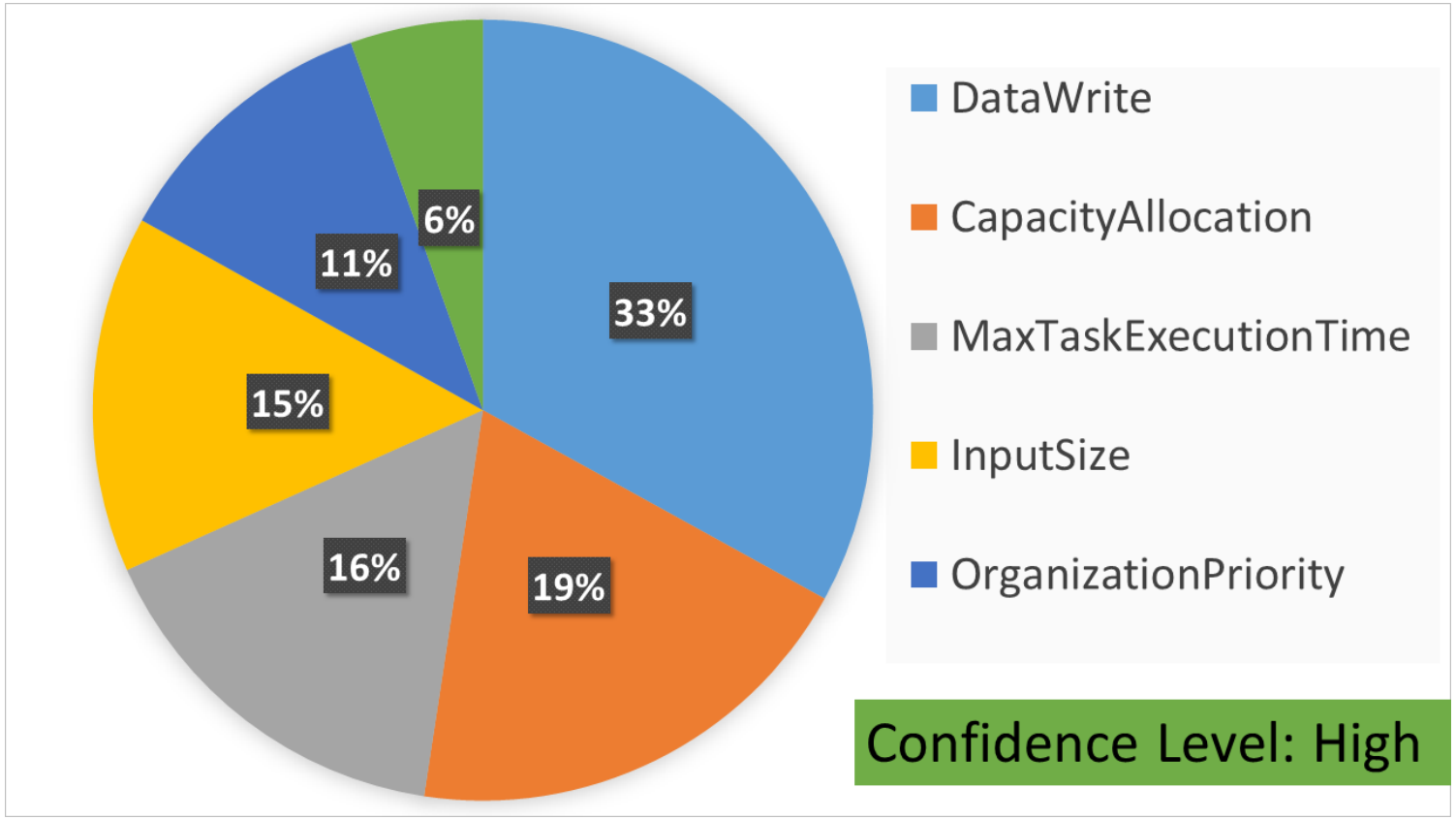



Flask Application


Offline Training Pipeline
Online Prediction Pipeline
Deployment
Online Feature Building
User Scenario
Input:
Job ID, VC
Job ID, VC
2
3
Features for Job ID, VC
1
Web Application
6
Output:
Prediction Report
Predicted Results
5
Pre-processing
Pre-processed Data
ML Model Training
Metrics & Hyperparameters
Tracking Server
Saved Model
Artifacts
Anomaly Reasoning Algorithm
Delta Feature Importance
mapping
Causes of Job Slowdown
Predicted Runtime
Confidence Level
Best Saved Model
Flask Application


Offline Training Pipeline
Data Preparation
Historic Data
Filtering
Filtered Data
Downloading
Download Data to Local
Pre-processing
Pre-processed Data
ML Model Training
Metrics & Hyperparameters
Tracking Server
Saved Model
Artifacts
Anomaly Reasoning Algorithm
Delta Feature Importance
mapping
Causes of Job Slowdown
Predicted Runtime
Confidence Level
Best Saved Model
Online Prediction Pipeline
Deployment
Online Feature Building
Job ID, VC
2
3
Features for Job ID, VC
User Scenario
Input:
Job ID, VC
1
Web Application
6
Output:
Prediction Report
Predicted Results
5
4
Features
Model Server VM

# Griffon Output



# Validation of Griffon Predictions

| Job Id | Predicted Reason | Engineer Validated Reason | Rank | Confidence Level |
|---|---|---|---|---|
| 9182 | Input size | Input size | 1 | High |

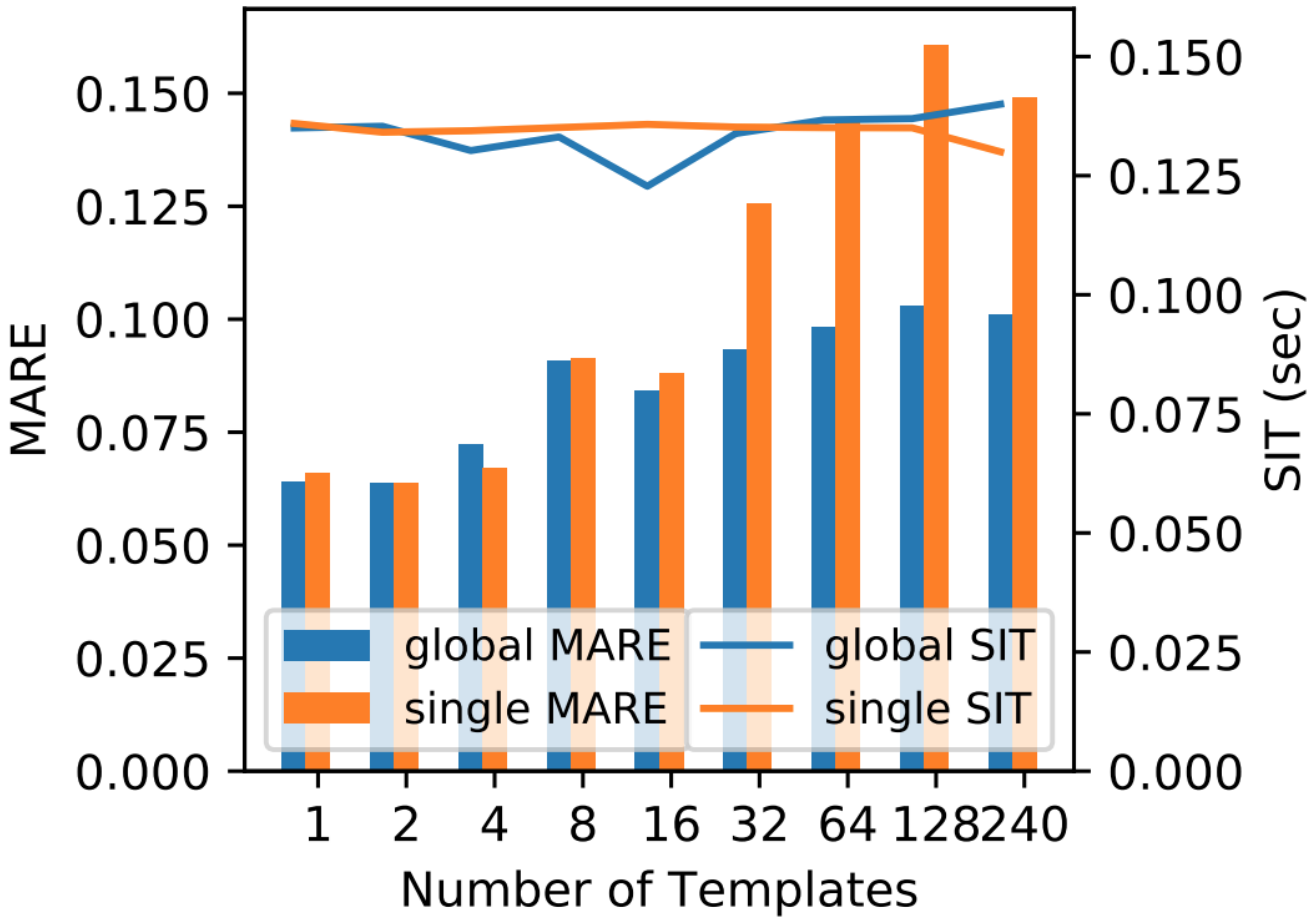
# Validation of Griffon Predictions

| Job Id | Predicted Reason | Engineer Validated Reason | Rank | Confidence Level |
|---|---|---|---|---|
| 9182 | Input size | Input size | 1 | High |
| 8578 | Revocation | Revocation | 4 | Medium |

# Validation of Griffon Predictions

| Job Id | Predicted Reason | Engineer Validated Reason | Rank | Confidence Level |
|---|---|---|---|---|
| 9182 | Input size | Input size | 1 | High |
| 8578 | Revocation | Revocation | 4 | Medium |
| 4414 | Yarn or cluster issue | Yarn or cluster issue | - | Low |
| 6170 | PN hours | PN hours | 5 | Medium |
| 7588 | Time skew | Time skew | 1 | High |
| 3798 | PN hours | PN hours | 1 | High |
| 1590 | PN hours | PN hours | 1 | High |
| 2560 | Usable machine count | Usable machine count | 2 | High |

# Scalability & Generalization



# Conclusions

- **End-to-end interpretable ranking system** to identify the root causes of job slowdowns
- **No human labeled reasons** needed
- **Highly consistent results** validated by on-call engineers
- Our model **generalizes well** by testing on job templates not included in the training set

# Thank you!

Please see our poster for more details ☺!