Microsoft

Microsoft Future Decoded

How Adatis helped Rank Gaming productionise & automate the management of machine learning models in Azure

Terry McCann – Adatis

Chris Conroy – Rank Group

Microsoft

Rank Gaming Group

Rank was established in 1937 when they were known for making films.

Rank now owns two of the biggest brands in gambling in the UK; Mecca Bingo and Grosvenor Casinos.

GROSVENOR CASINOS

MECCA

Club | Online | Mobile

Microsoft

Adatis Consulting

Adatis Consulting Limited is a Microsoft Gold Partner specialising in Data Engineering and Data Science.

Established 12 Years ago we are now a team of ~70. We have 3 Data Platform MVPs.

adatis

Microsoft

Rank Data Science

Rank has a team of 7 Data Scientists.

They hold advanced degrees in a range of disciplines.

Wide variety of skills (Python/Scala/R etc.)

Rank Data Science had big ambitions to use Azure to the fullest, to create a truly polyglot architecture decoupled from their existing data processes.

Limited engineering capability on the team.

Microsoft

The Problem?

Microsoft

Rank Data Science Problems

Existing data processes were tightly coupled with legacy processes, all new data had to go via a Data Warehouse.

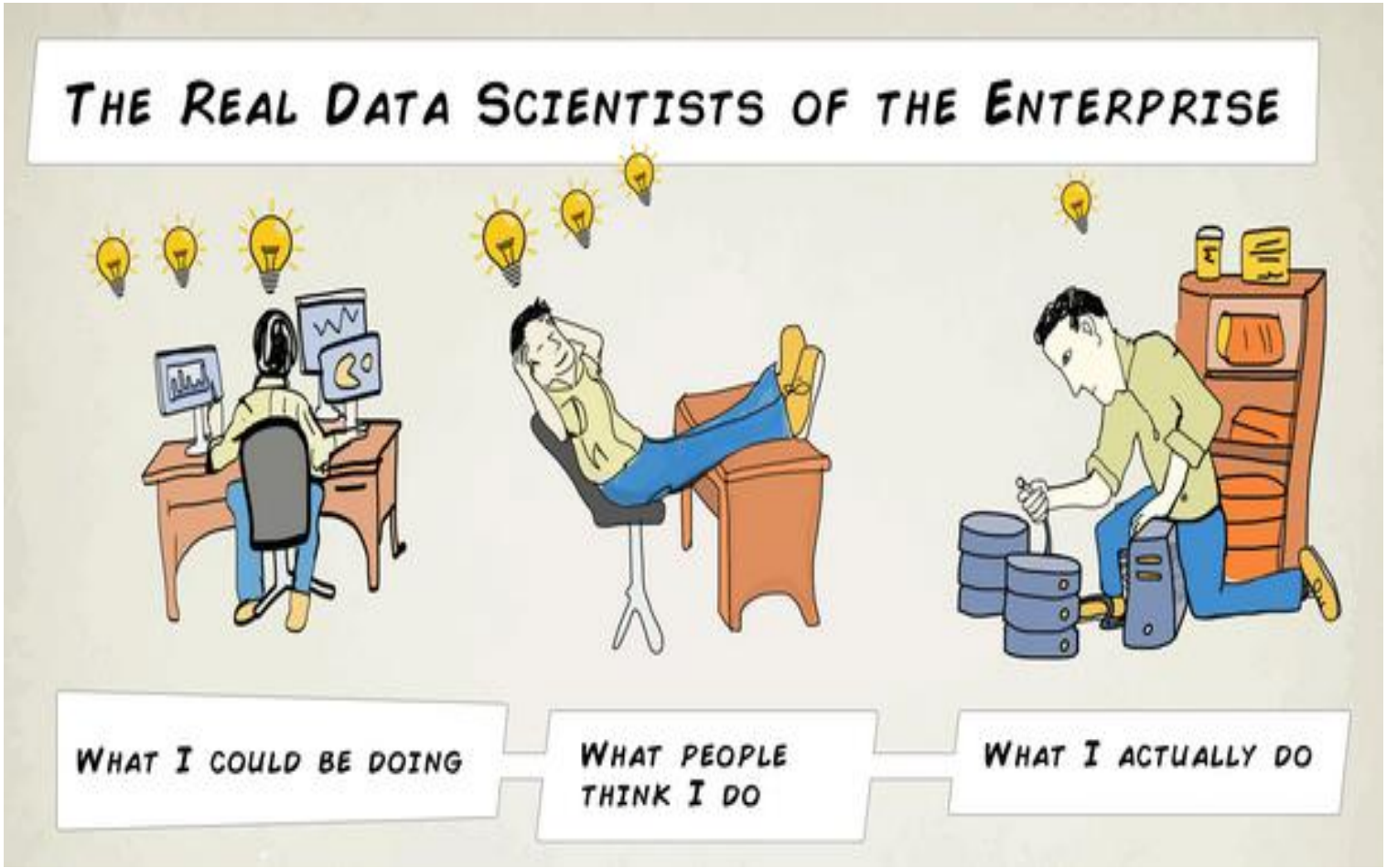
Access to new data could take anywhere up to 7 months to obtain.

Machine Learning models were in ‘manual production’, however not enough. Rank’s roadmap was long and ambitious.

Very limited engineering resource to call on.

When models were in production, Rank faced a problem with challenger/incumbent model replacement.

Microsoft


THE REAL DATA SCIENTISTS OF THE ENTERPRISE
WHAT I COULD BE DOING
WHAT PEOPLE THINK I DO
WHAT I ACTUALLY DO

Microsoft

1. Access to Data was too slow
2. Machine Learning Model Management

Microsoft

The Solution?

Microsoft

Leveraging a Data lake for access to data

A traditional Rank Analytical data process

1. Requirements are gathered
2. Requirements are mapped to the data
3. Data in acquired
4. Data is profiled
5. Data is loaded in to a stage process
6. Data is cleaned and transformed
7. Data is loaded in to a data warehouse
8. Data is loaded in to a semantic layer
9. Data is reported to the business

adatis

Microsoft

Where does the data scientist come in to this?

adatis

Microsoft

80% of Data Science is Cleaning and Preparing Data

adatis

https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#49611bd56f63

Microsoft

1. Requirements are gathered
2. Requirements are mapped to the data
3. Data in acquired
4. Data is profiled
5. Data is loaded in to a stage process
6. Data is cleaned and transformed
7. Data is loaded in to a data warehouse
8. Data is loaded in to a semantic layer
9. Data is reported to the business

data scientist

adatis

Microsoft

1. Requirements are gathered
2. Requirements are mapped to the data
3. **Data in acquired**
4. **Data is profiled**
5. **Data is loaded in to a stage process**
6. **Data is cleaned and transformed**
7. Data is loaded in to a data warehouse
8. Data is loaded in to a semantic layer
9. Data is reported to the business

data scientist

adatis

Microsoft

adatis

Microsoft

1. A central store of data (Good/bad/ugly)
2. Decoupled from existing processes
3. Automate the scoring of *batch* Machine Learning Models
4. Should apply DevOps (Source control/CI/CD)
5. A fully-automated cleaning process

adatis

Microsoft

adatis

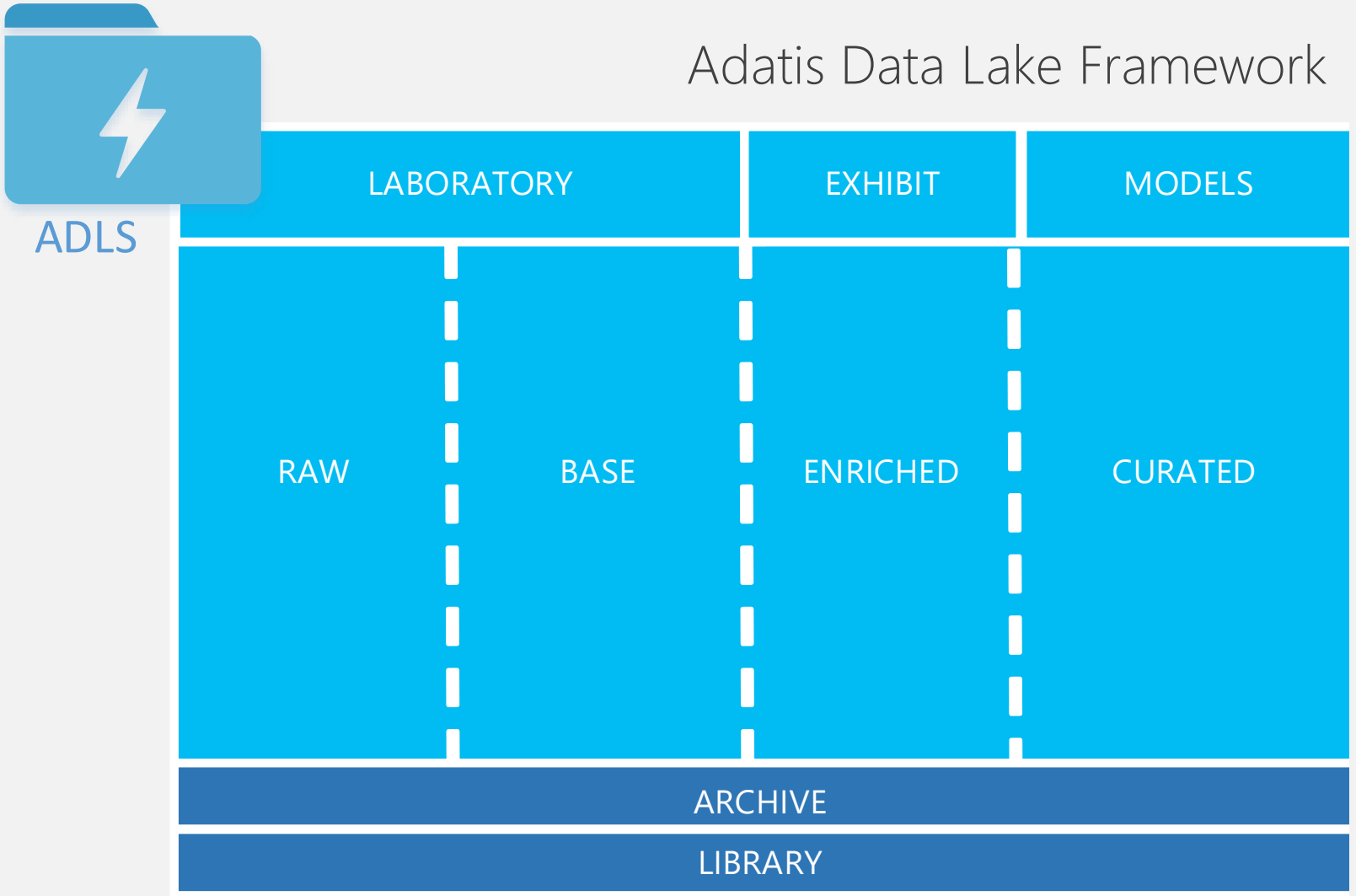
Microsoft

DATA LAKE

adatis

Microsoft

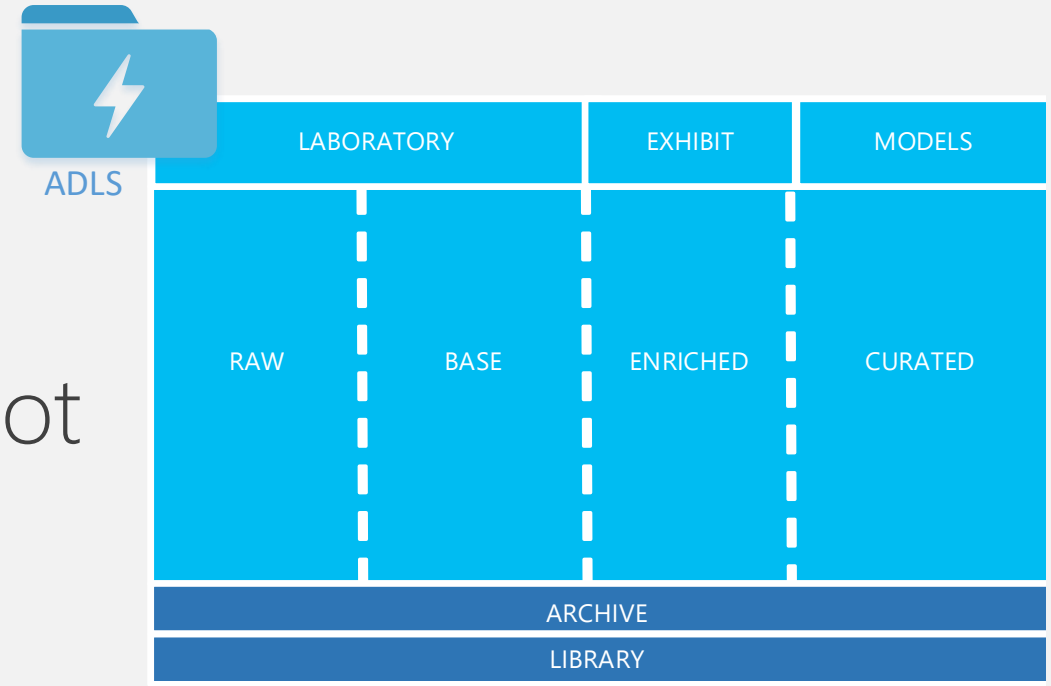
Adatis Data Lake Framework



adatis

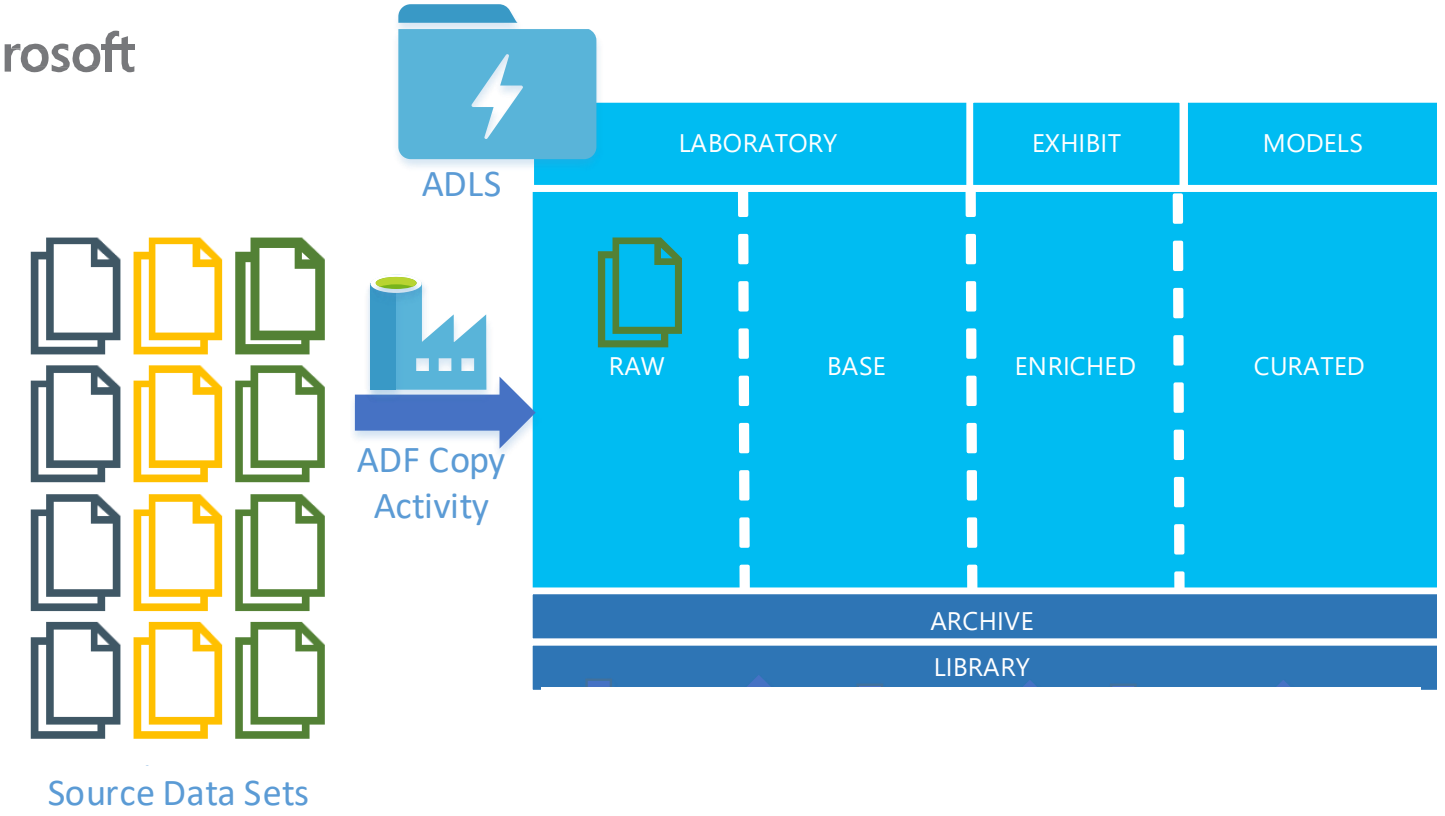
Microsoft

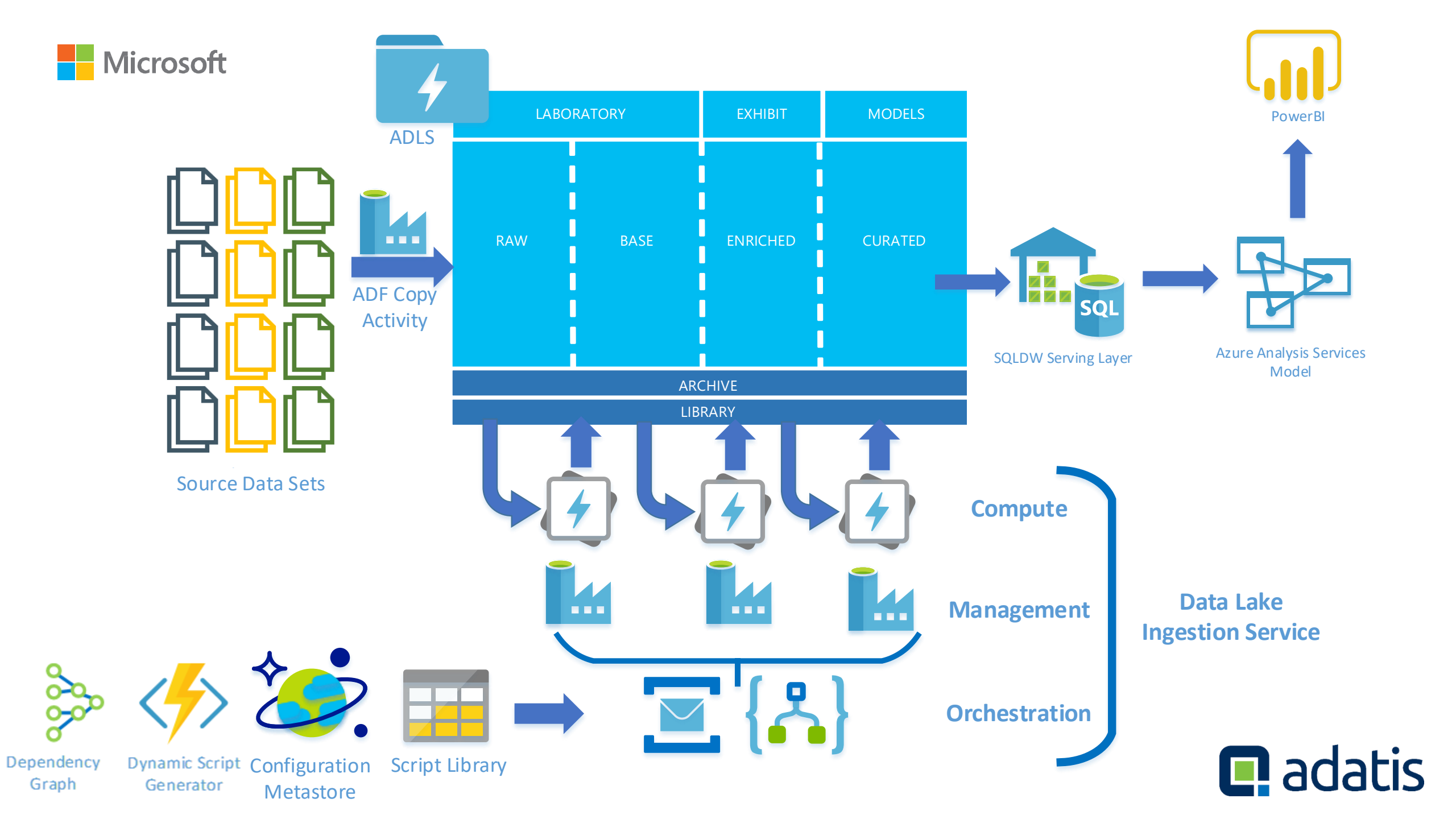
1. Ensure that the file is correct
2. Ensure that the schema has not drifted
3. Apply a series of quality rules
4. Enrich the data with new columns
5. Combine multiple files in to one.



adatis

Microsoft


ADLS
LABORATORY
EXHIBIT
MODELS
RAW
BASE
ENRICHED
CURATED
ADF Copy Activity
ARCHIVE
LIBRARY
Source Data Sets


Microsoft
ADLS
LABORATORY
EXHIBIT
MODELS
RAW
BASE
ENRICHED
CURATED
ARCHIVE
LIBRARY
PowerBI
ADF Copy Activity
SQL
SQLDW Serving Layer
Azure Analysis Services Model
Source Data Sets
Compute
Management
Orchestration
Data Lake Ingestion Service
Dependency Graph
Dynamic Script Generator
Configuration Metastore
Script Library
adatis

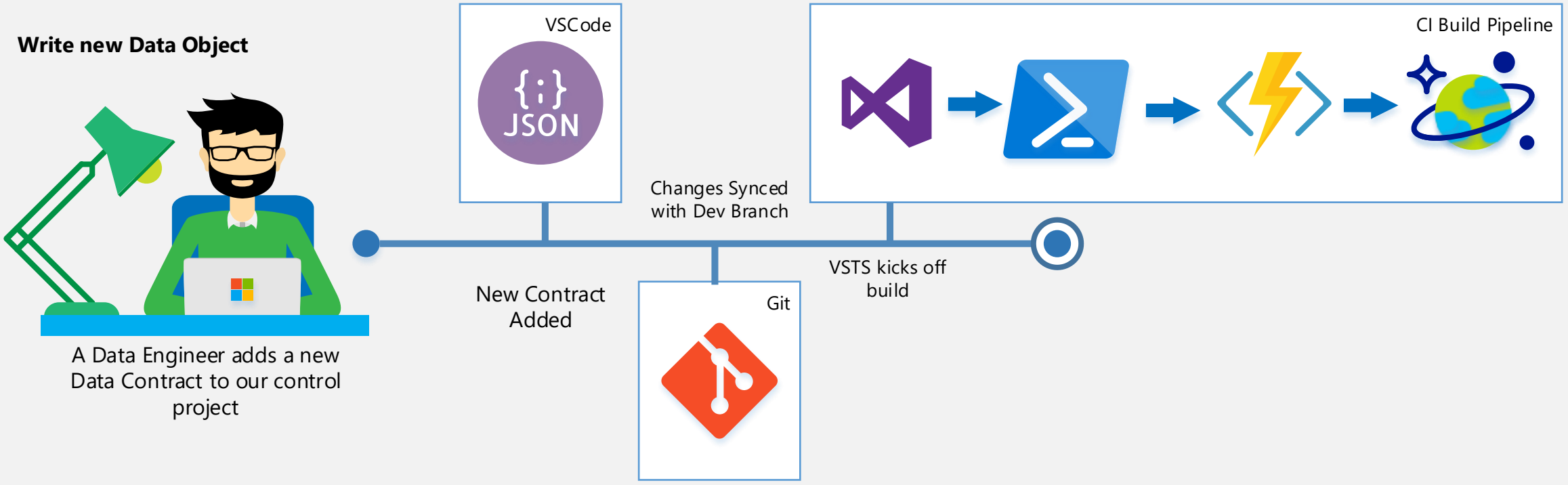
Microsoft

Benefits

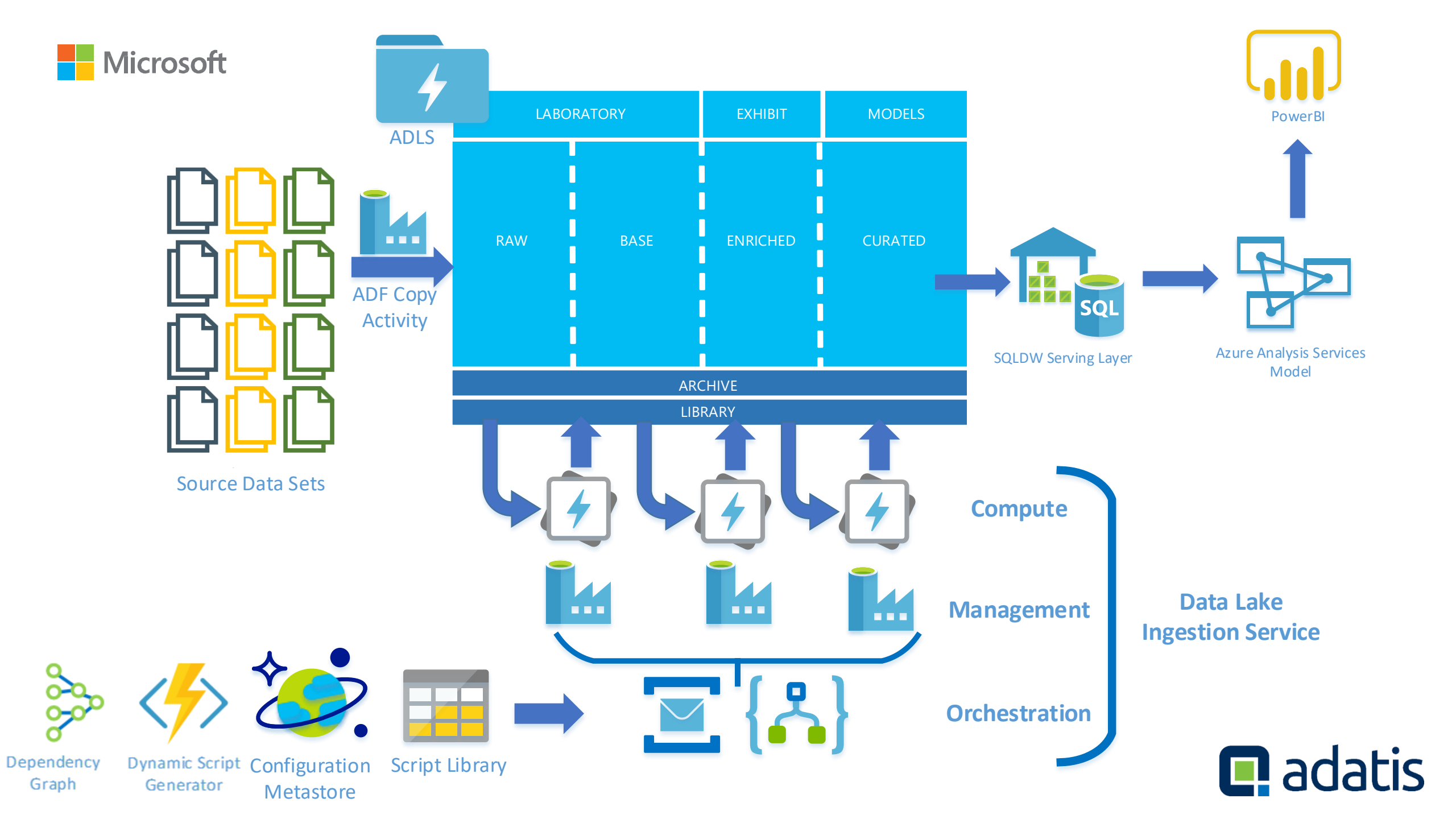
1. Time to onboard data has been reduced from months to minutes
2. Data is in a consistent format and level of quality for both Data Science and Traditional analytics
3. The lake serves as an immutable store of all data in the business.
4. Batch model scoring is fully automated
5. DevOps automates the deployment of new data processes

adatis

Solution Build Pipelines



adatis


Microsoft
ADLS
LABORATORY
EXHIBIT
MODELS
RAW
BASE
ENRICHED
CURATED
ARCHIVE
LIBRARY
PowerBI
ADF Copy Activity
SQL
SQLDW Serving Layer
Azure Analysis Services Model
Source Data Sets
Compute
Management
Orchestration
Data Lake Ingestion Service
Dependency Graph
Dynamic Script Generator
Configuration Metastore
Script Library
adatis

Microsoft

Machine Learning Model Management (Interactive)

Microsoft

Rank Data Science Problems

Existing data processes were tightly coupled with legacy processes, all new data had to go via a Data Warehouse.

Access to new data could take anywhere up to 7 months to obtain.

Machine Learning models were in ‘manual production’, however not enough. Rank’s roadmap was long and ambitious.

Very limited engineering resource to call on.

When models were in production, Rank faced a problem with challenger/incumbent model replacement.

Microsoft

Rank Data Science Problems

~~Existing data processes were tightly coupled with legacy processes, all new data had to go via a Data Warehouse.~~

~~Access to new data could take anywhere up to 7 months to obtain.~~

Machine Learning models were in ‘manual production’, however not enough. Rank’s roadmap was long and ambitious.

Very limited engineering resource to call on.

When models were in production, Rank faced a problem with challenger/incumbent model replacement.

Microsoft

The Machine Learning Process

1. Gathering data
2. Preparing that data
3. Select an algorithm
4. Train a model (data + algorithm)
5. Evaluate the model
6. Hyperparameter tuning
7. Test the prediction (new data + model)
8. Deploy the model (“Productionised”)

https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e

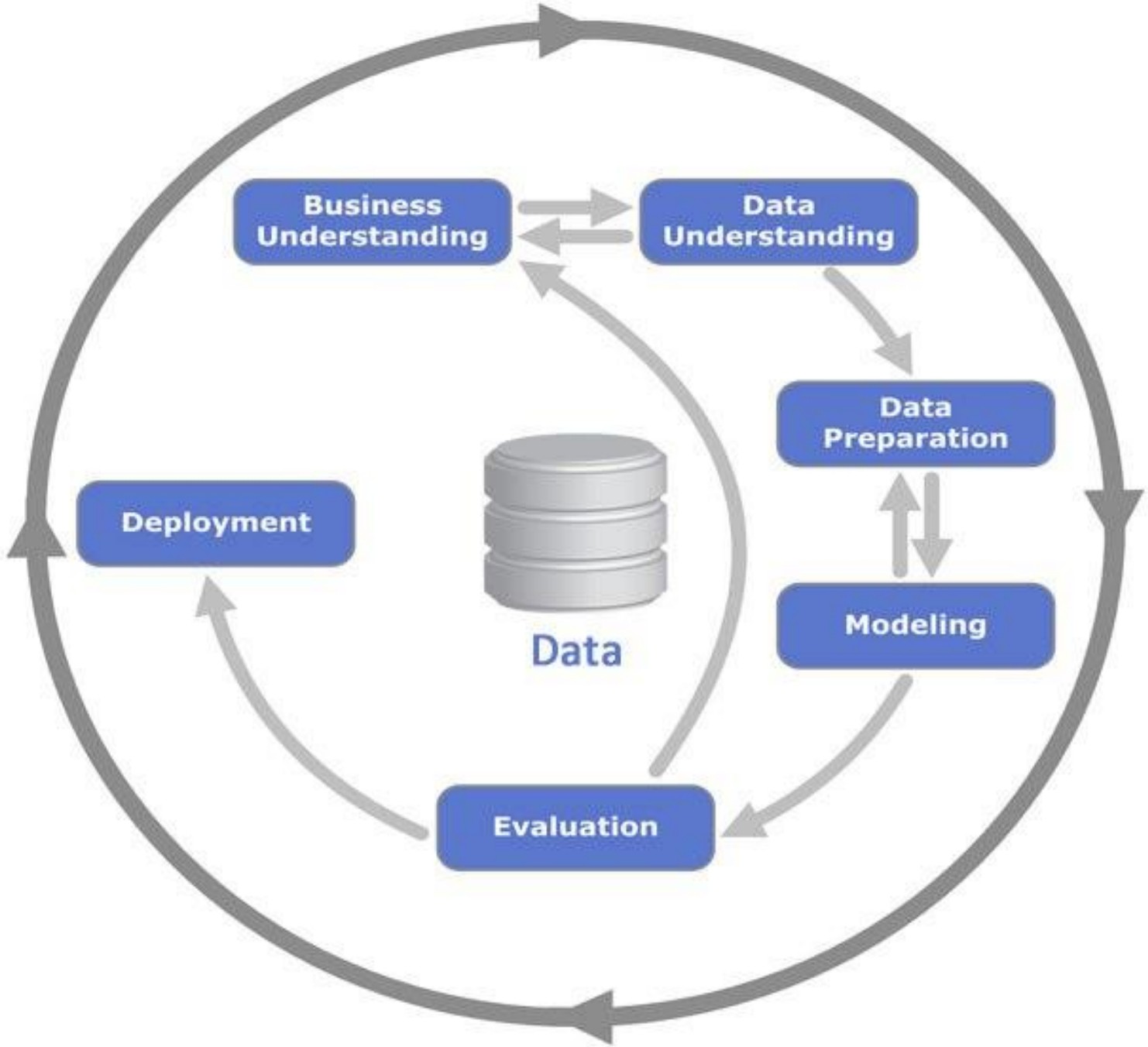
adatis

Microsoft

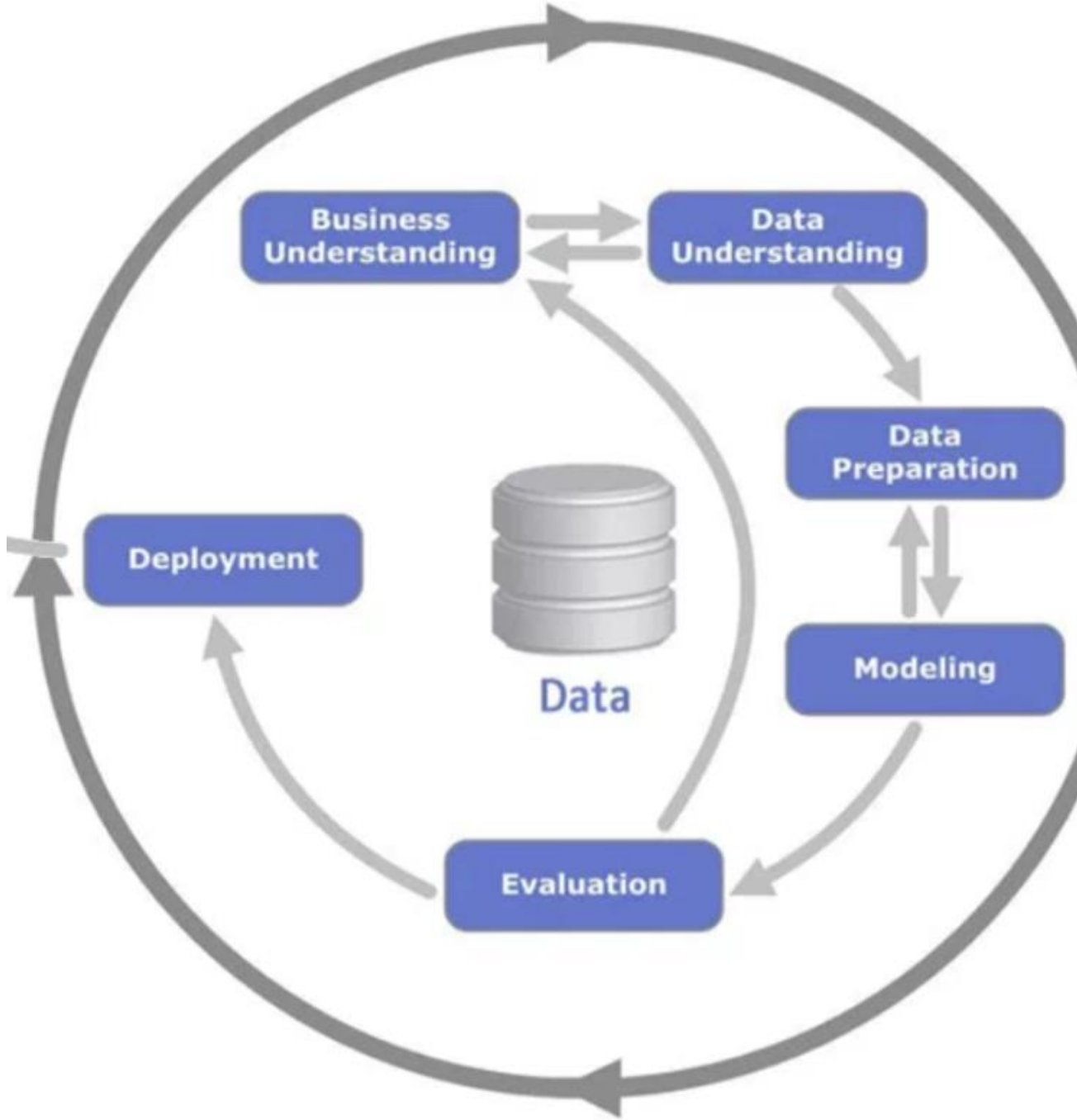
"Productionisation of models is the *TOUGHEST* problem in data science"

Schutt, R & O'Neill C, 2014. Doing Data Science Straight From the Front Line, O'Reilly Press. California

CRISP-DM Process Diagram



Source: Kenneth Jensen


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

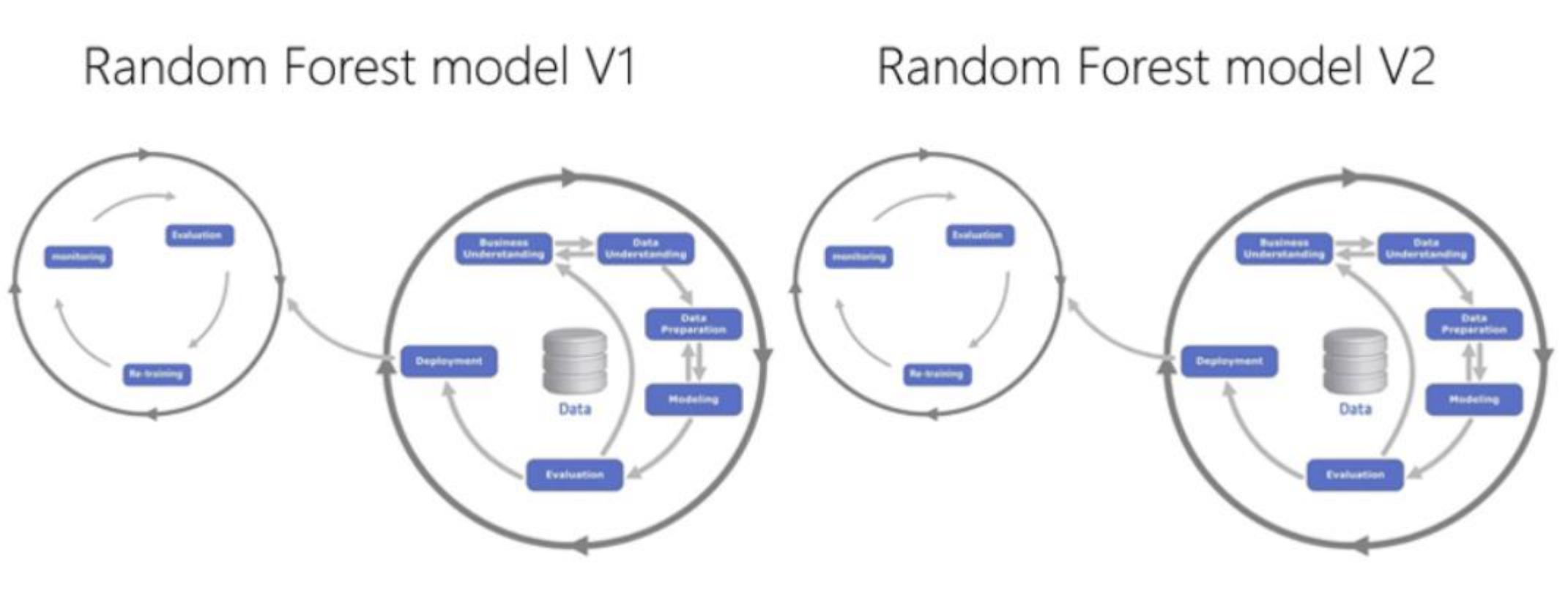
Microsoft

Machine Learning: Challenger model Vs Incumbent model

adatis

Microsoft

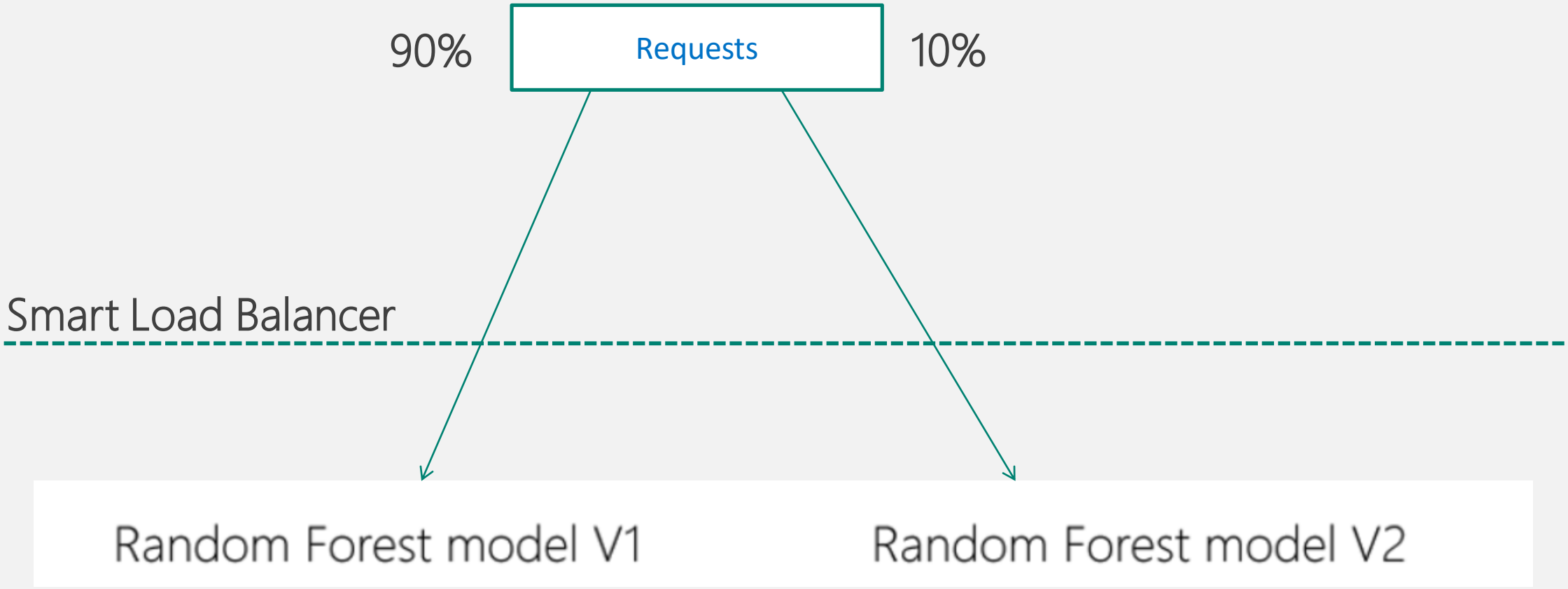
AB Testing scenario



adatis

Microsoft

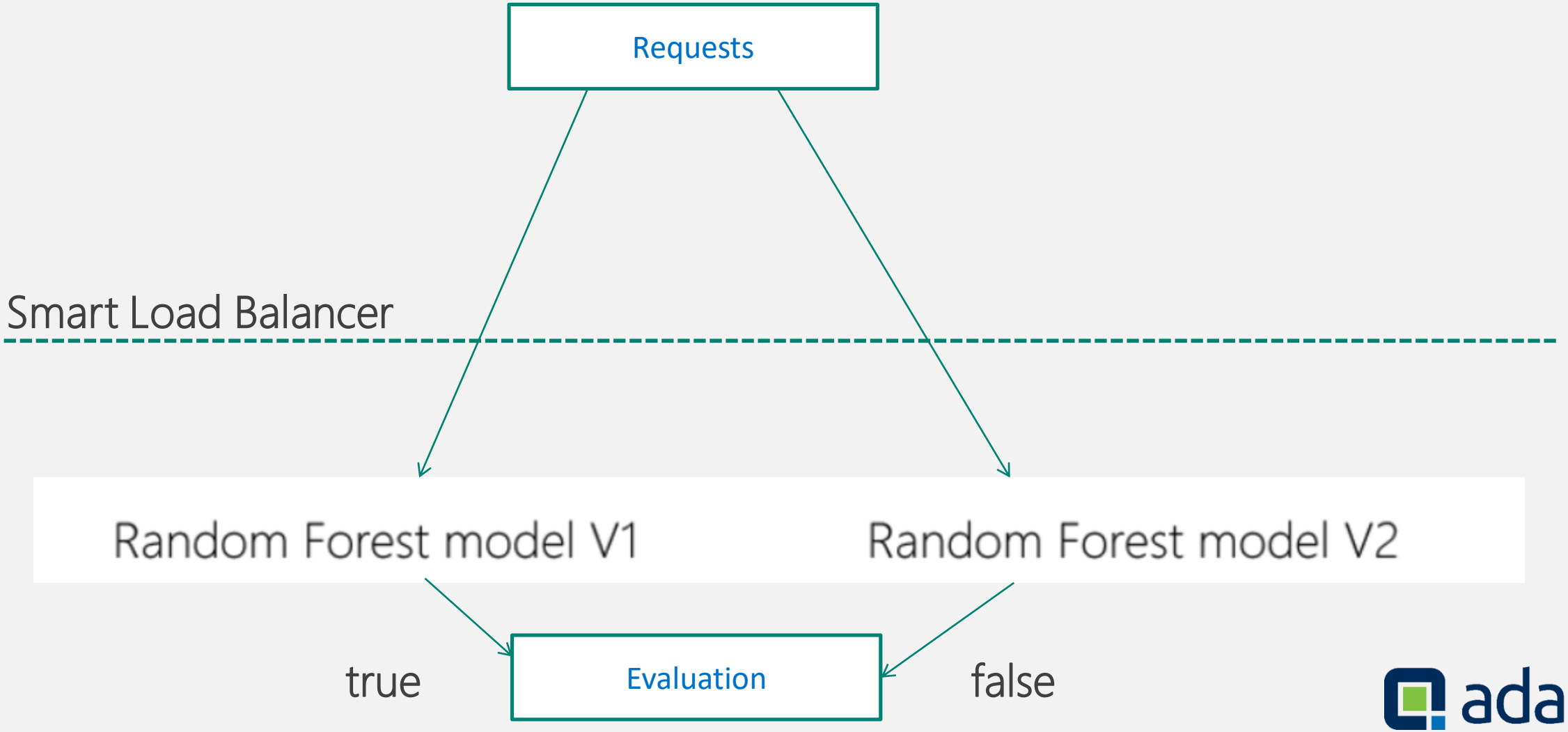
AB Testing scenario



adatis

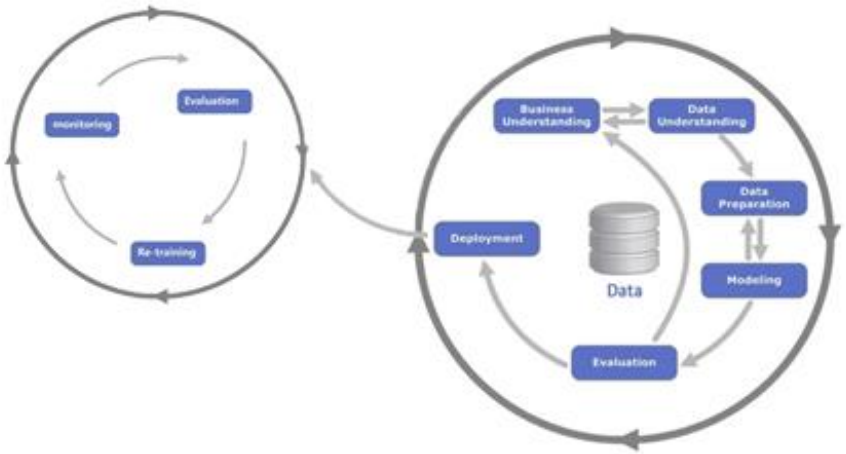
Microsoft

AB Testing scenario



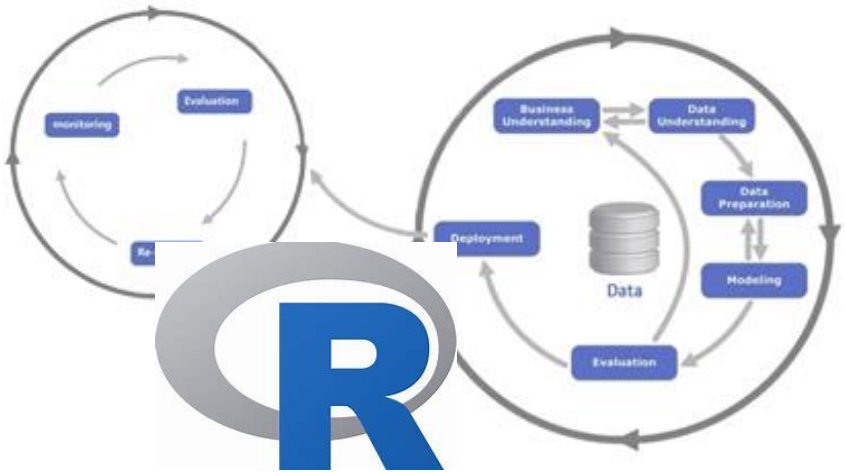
adatis

Random Forest model V1

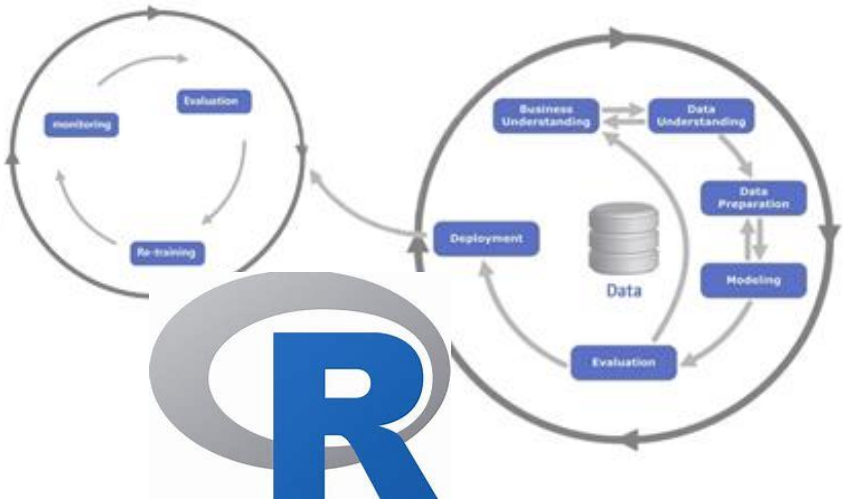


Random Forest model V2

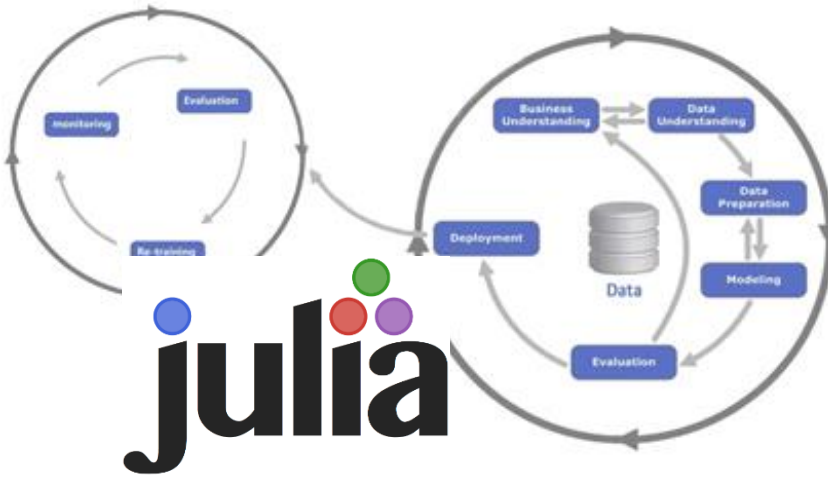
Random Forest model V1



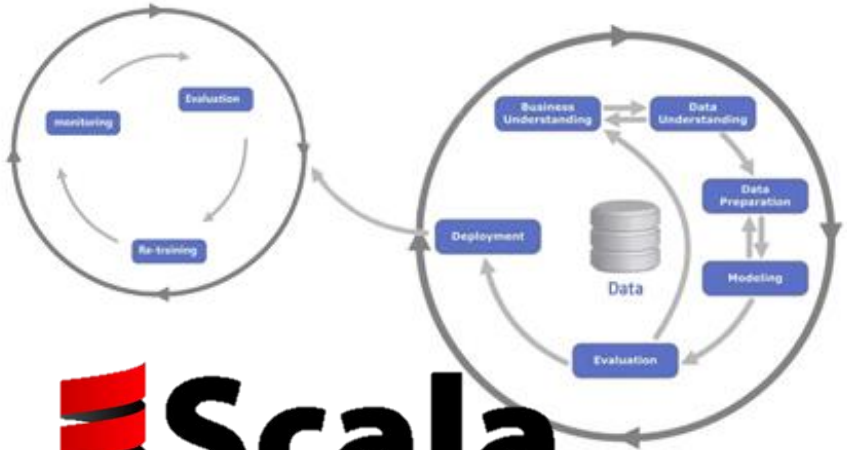
Random Forest model V2



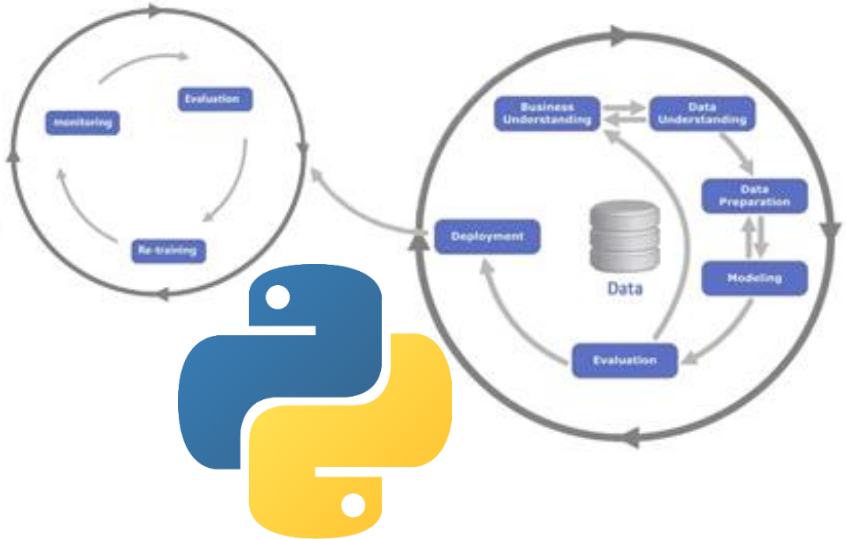
K-Nearest Neighbours



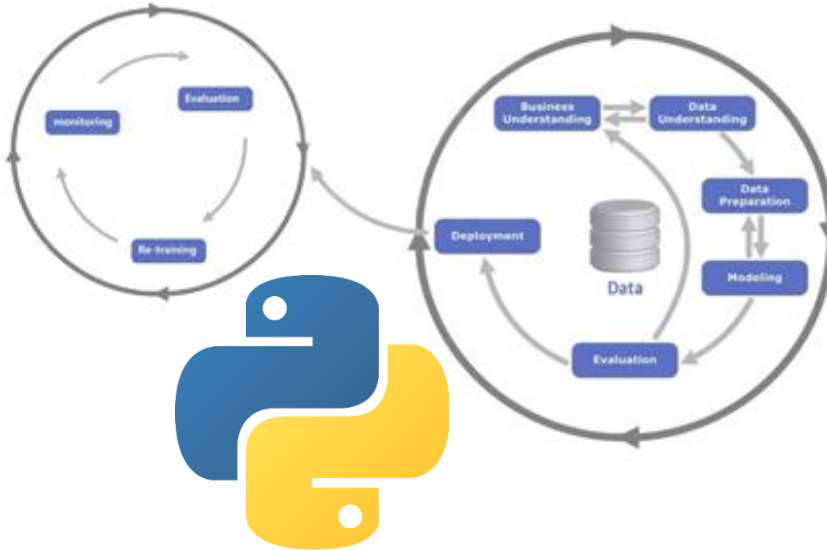
Ada Boost



Naive Bayes

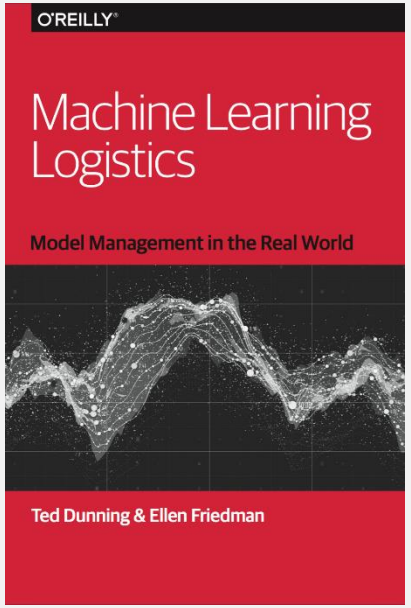


Decision Tree



Microsoft

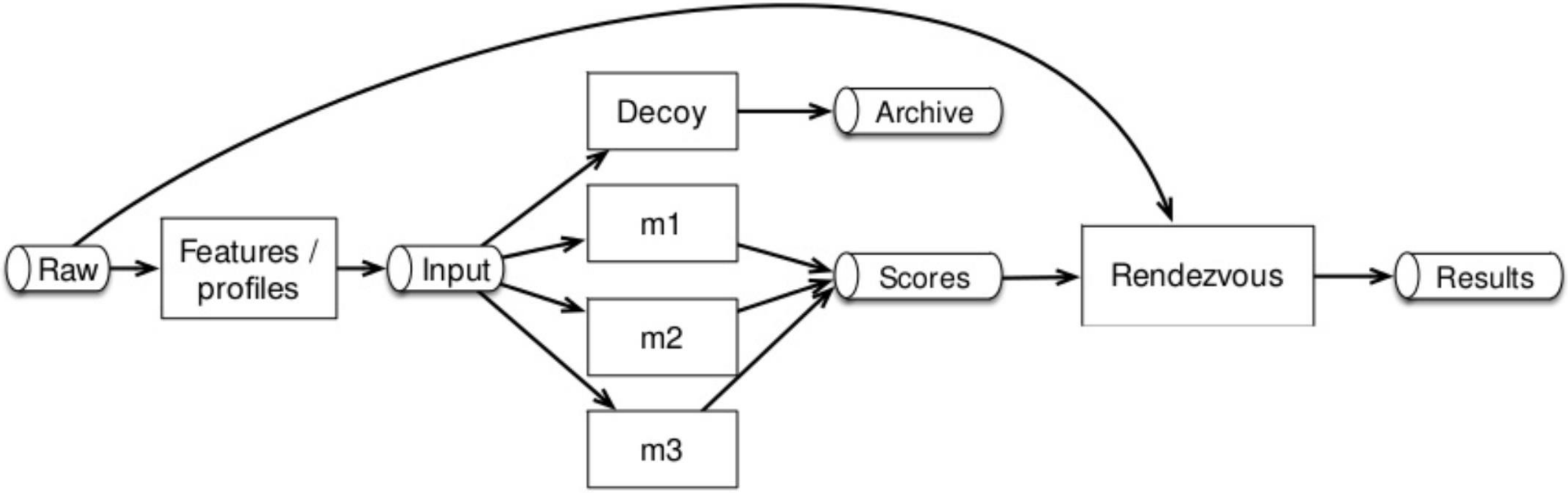
Rendezvous Architecture



https://mapr.com/ebook/machine-learning-logistics/

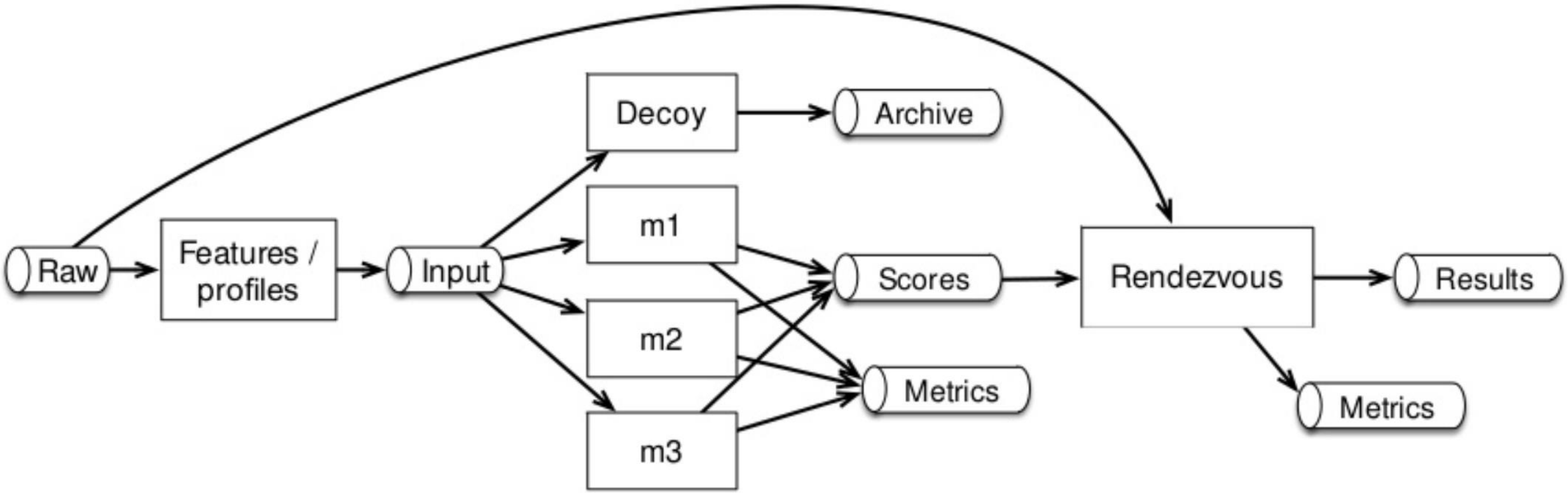
adatis

Microsoft



adatis

Microsoft

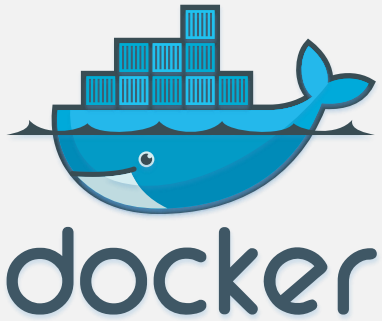


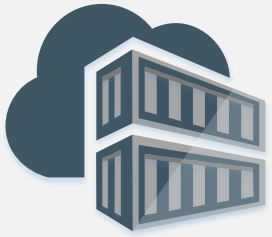
adatis

Microsoft

Key Components?

adatis





Azure Container Registry

Java Build
(to trigger a ML model)

Machine Learning Model
Build (Any language)

Java Image
(to trigger a ML model)

V1

Machine Learning Mode
(Any language)

V6

Java Image
(to trigger a ML model)

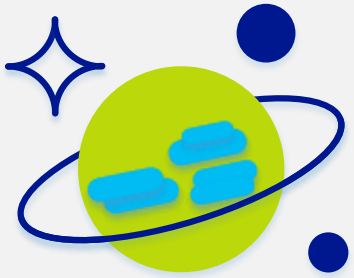
Machine Learning Model
(Any language)

Azure Functions

REST API

Service Bus

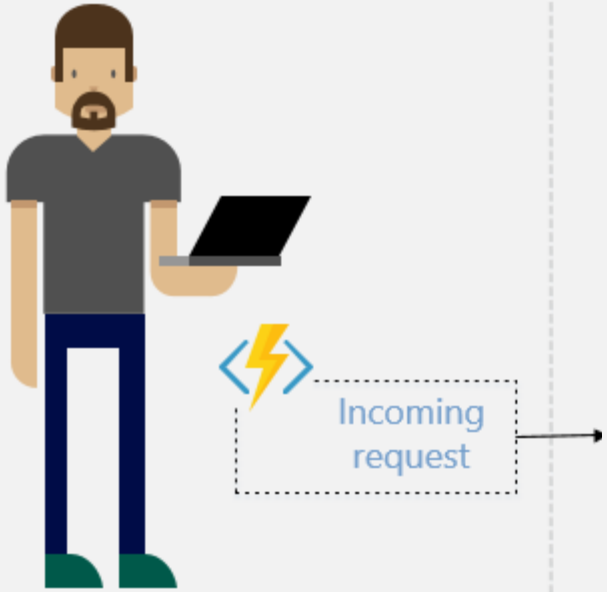
Topics and Queues to trigger models



CosmosDB

Operational Monitoring

Rendezvous



adatis

Microsoft

Do our Data Scientist's really need to learn all of that?

adatis

Introducing Azure DevOps

Azure Boards

Plan, track, and discuss work across teams, deliver value to your users faster.

Azure Repos

Unlimited cloud-hosted private Git repos. Collaborative pull requests, advanced file management, and more.

Azure Pipelines

CI/CD that works with any language, platform, and cloud. Connect to GitHub or any Git provider and deploy continuously to any cloud.

Azure Test Plans

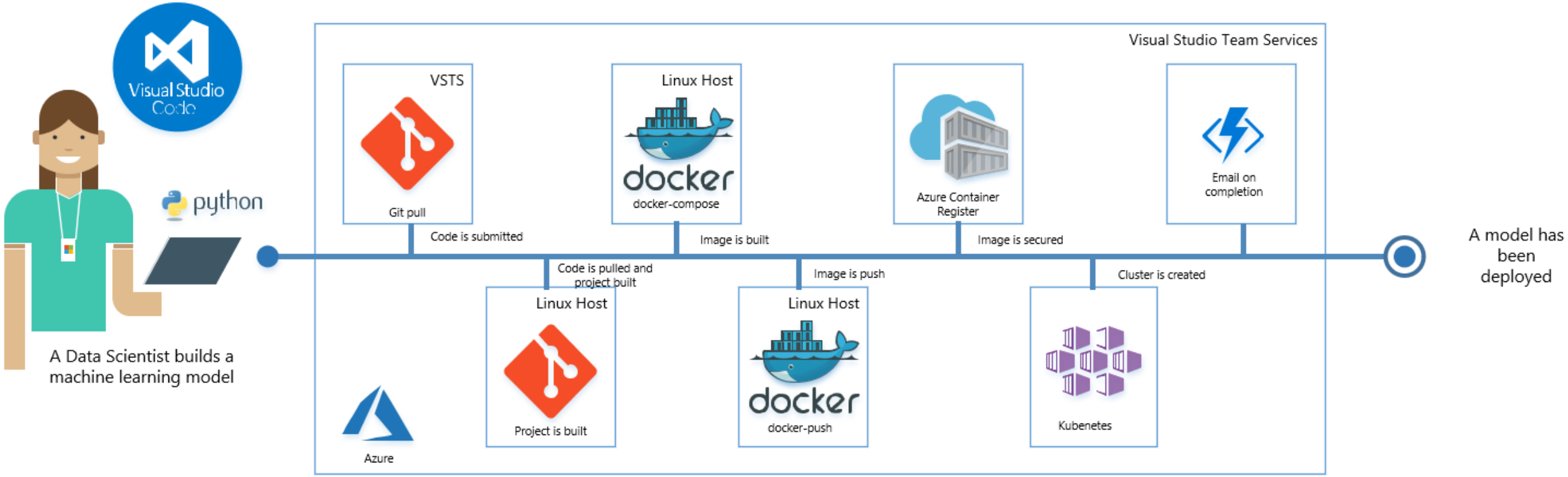
The test management and exploratory testing toolkit that lets you ship with confidence.

Azure Artifacts

Create, host, and share packages. Easily add artifacts to CI/CD pipelines.

adatis

Microsoft



adatis

Microsoft

Benefits

1. DevOps removes the Engineering requirement
2. Able to support any language in production
3. Supports both batch and interactive.
4. Automated Model Management
5. Operational statistics all fully logged
6. Elastic scale
7. Future proof solution

adatis

Microsoft

Looking to the future

adatis

Microsoft

Chris Conroy

Head of Data Science –
Rank Group
www.rank.com

Terry McCann

Head of Data Science –
Adatis Consulting
www.adatis.co.uk
tpm@adatis.co.uk

Microsoft

Things to do next

Session Feedback

Please rate this session in the Future Decoded app!

Microsoft UK AI Research Report

Download the AI Report at http://aka.ms/UKAIreport

Azure Marketplace Consulting Services

Find Microsoft partners to help at http://aka.ms/UKMarketplaceService

Enhance your Digital Skills

Find great guidance and resources at http://aka.ms/fdskills

Microsoft

Questions

Microsoft

Thank You