Anomaly Detection in Wind Turbine using Data Science and Big Data

Anish Mahapatra | 19 October, 2021

Agenda

01 Introduction

02 The Data Lifecycle

03 Production Machine Learning Architecture

04 Predictive Maintenance Dashboard

05 The Data Science Process Flow

06 Q&A Session

Introduction

Senior Data Scientist

LinkedIn: www.linkedin.com/in/**anishmahapatra**

Medium: **anishmahapatra**.medium.com

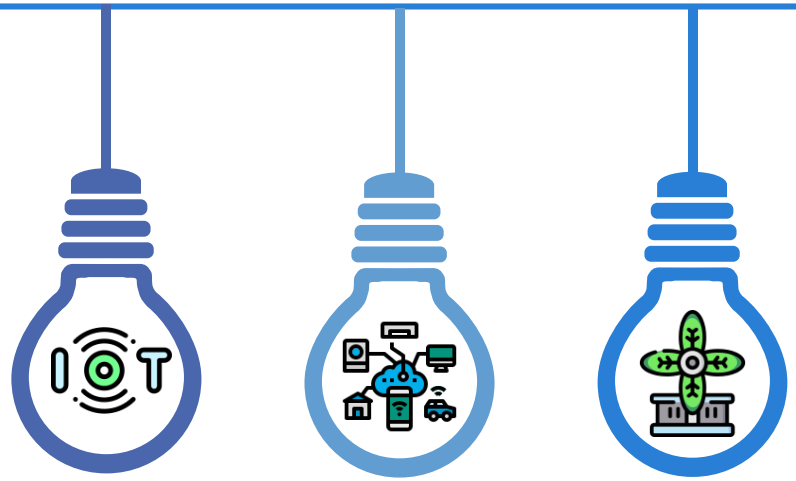
Website: www.**anishmahapatra**.com

GitHub: www.github.com/**anishmahapatra**

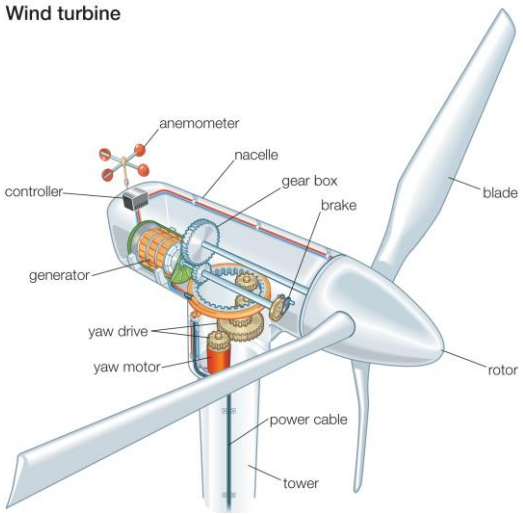
Q. How to get into Data Science?

A. Simple, ask a real-world Data Scientist.

The Energy Industry



Industrial IoT | **External Data** | **Turbine Data**



$882 Billion

Energy Industry Valuation 2021

25%

Cost savings using Predictive Maintenance

The Data Lifecycle

01

Understand the Data

Source the data,
Create Data Dictionary,
View Data Summary,
View Data Distribution

02

Data Exploration

Exploratory Data Analysis,
Make ER Diagram,
Main Analytical Data frame

03

Data Roll-up

Aggregation and roll-up of data,
Merging of Data on relevant keys,
Unit testing,
Building pipelines to experiment,

04

Feature Engineering

Building new features,
Combining external data,
Building a feature store

05

Model Interpretability

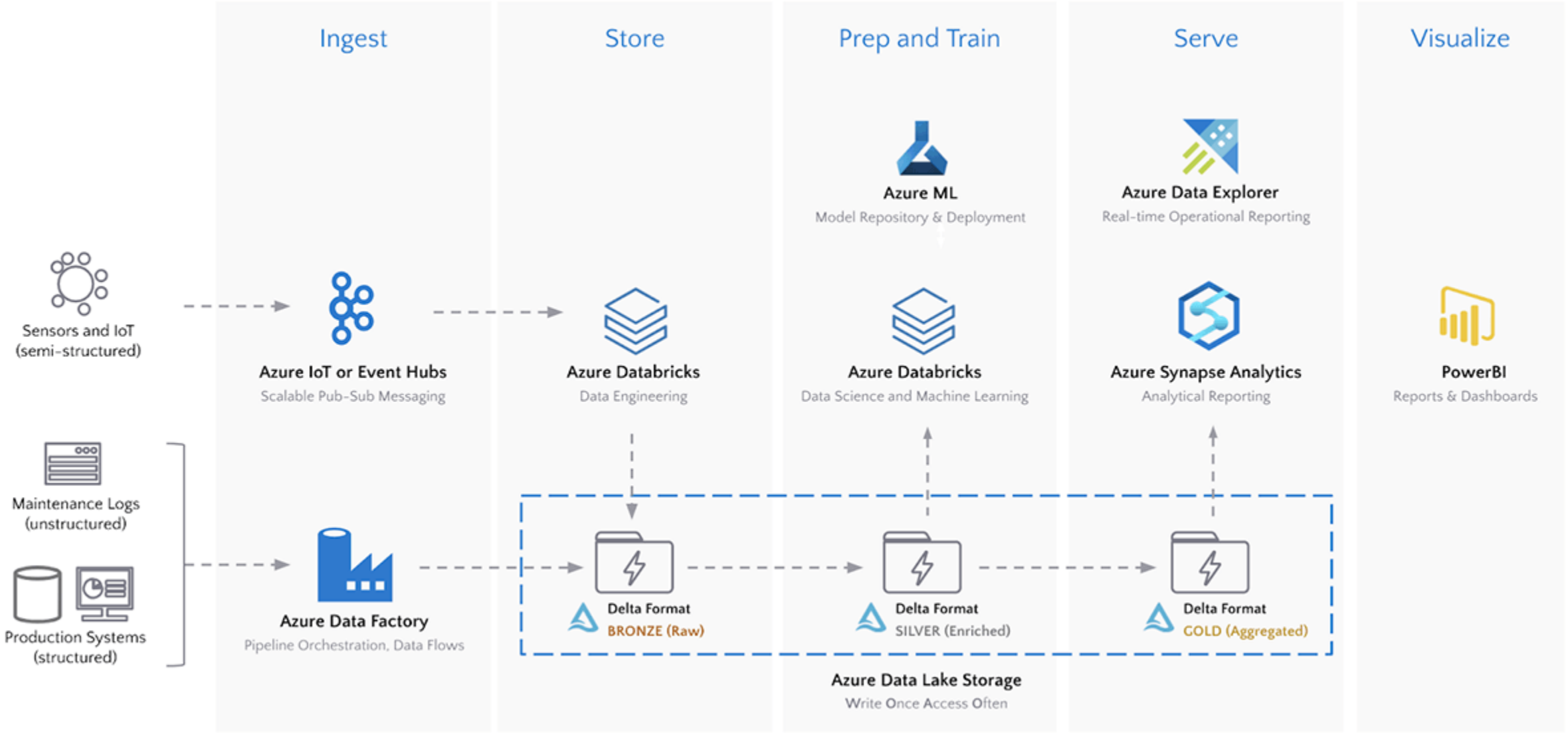
Model Interpretability using SHAP,
Machine Learning interpretation,
Business Explainability

06

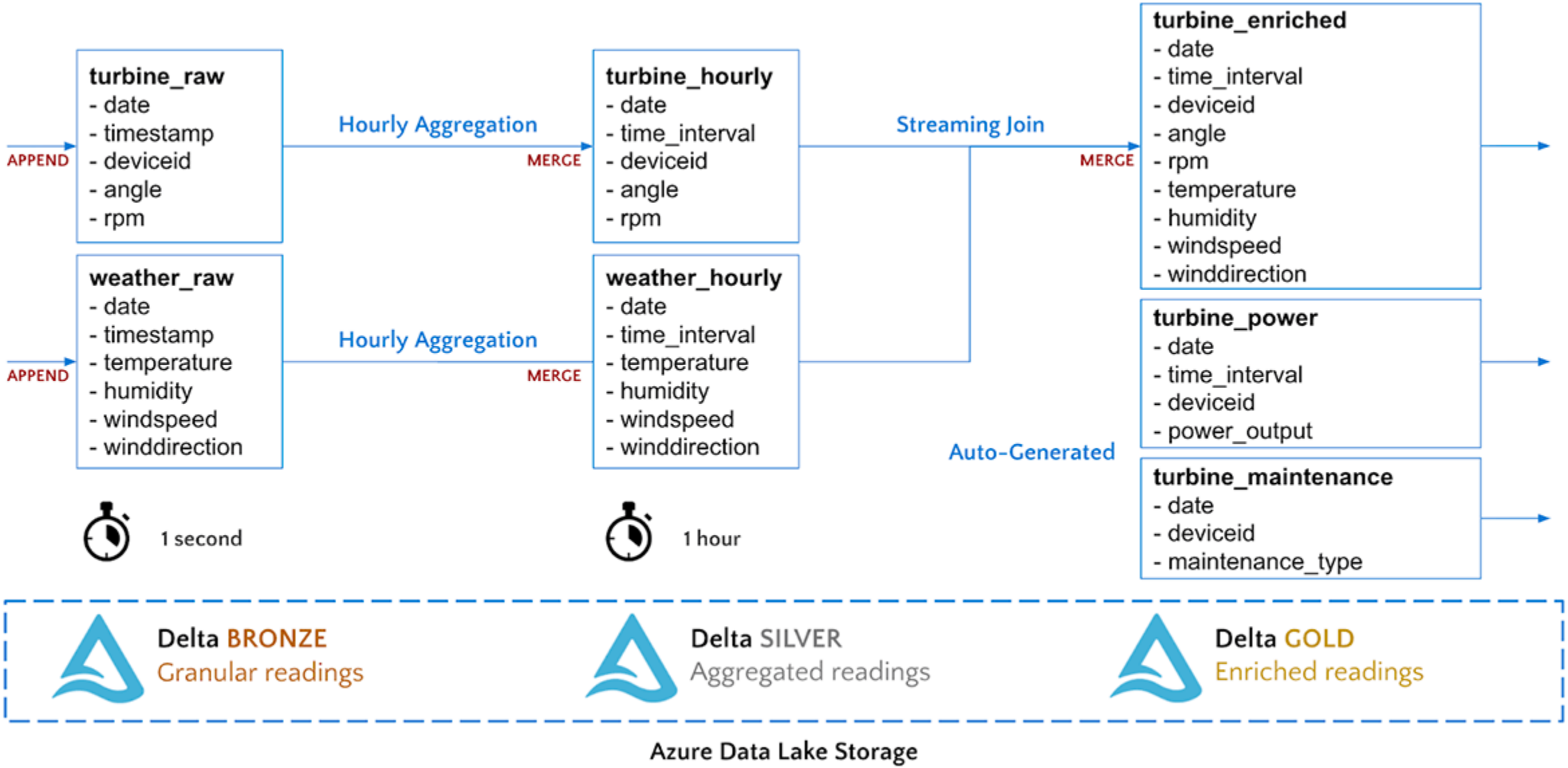
Model Evaluation

Model Evaluation,
Monitoring Model Performance and Drift,
Data Governance Policies

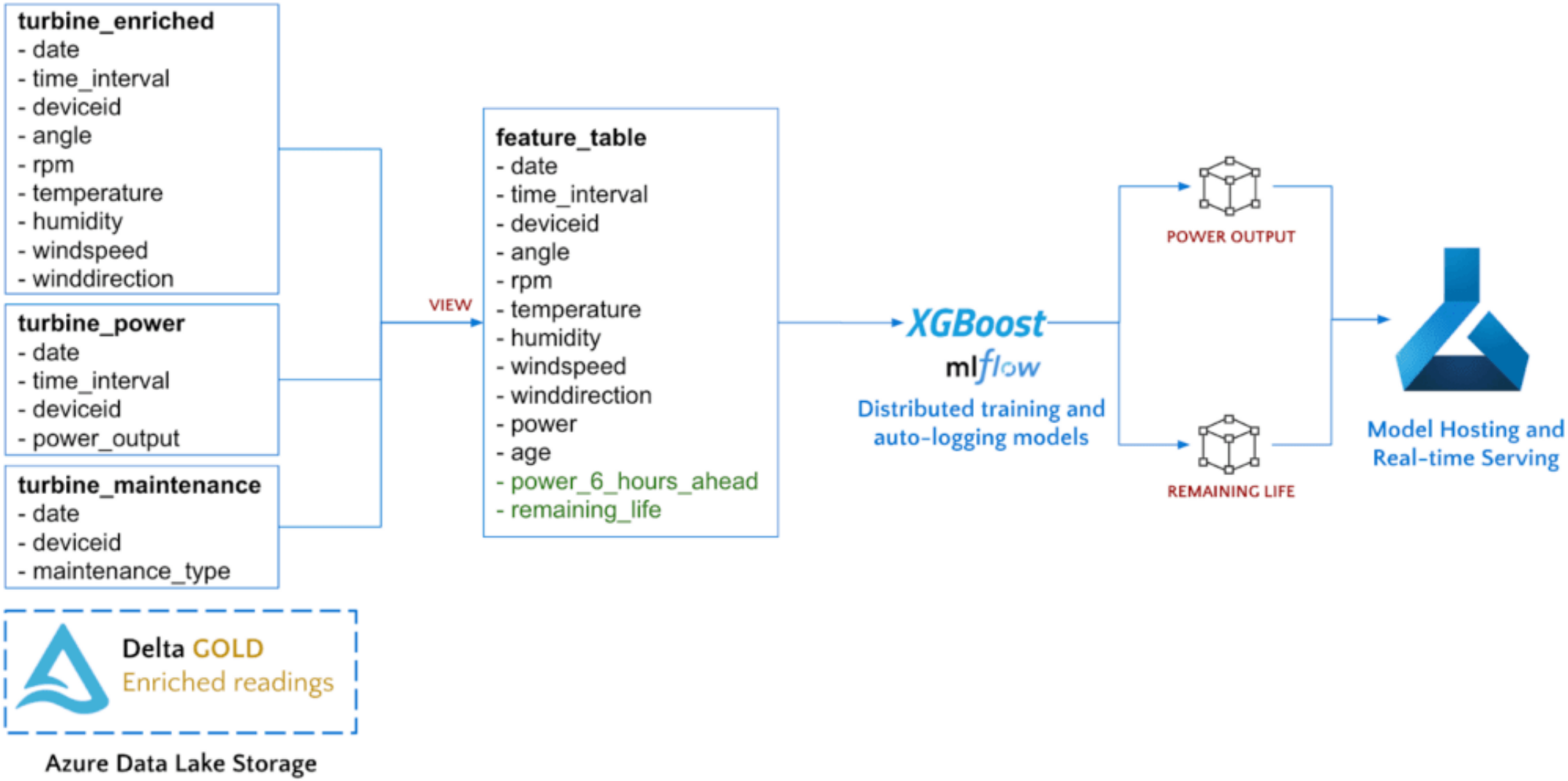
Production Machine Learning Architecture



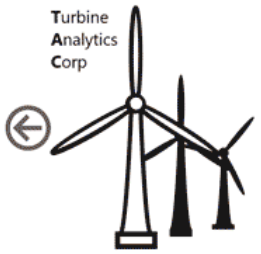
Implementation in Data Lake



Implementation in Data Lake



Predictive Maintenance Dashboard



date

6/1/2020 6/22/2020

Wind Farm

- Bethel Wind Farm
- Mesquite Creek Wind
- Miami Wind Energy Center
- Palo Duro Wind Facility
- Panhandle Wind Farm
- Panther Creek Wind Farm
- Papalote Creek Wind Farm
- Peñascal Wind Farm
- Rattlesnake Den Wind Farm
- Roscoe Wind Project

Turbine ID

- WindTurbine-0
- WindTurbine-1
- WindTurbine-2
- WindTurbine-3
- WindTurbine-4
- WindTurbine-5
- WindTurbine-6
- WindTurbine-7

Total Power Generated

36.52K

Predicted Power Generated Over Next 6 Hours

37.17K

Watts

Last Maintenance Event

15

Days

Predicted Next Maintenance

22

Days

Wind Direction

W

Wind Speed

0.00 67.89 135.78

Temperature

0.00 26.16 52.33

Humidity

0.00K 678.89 1.36K

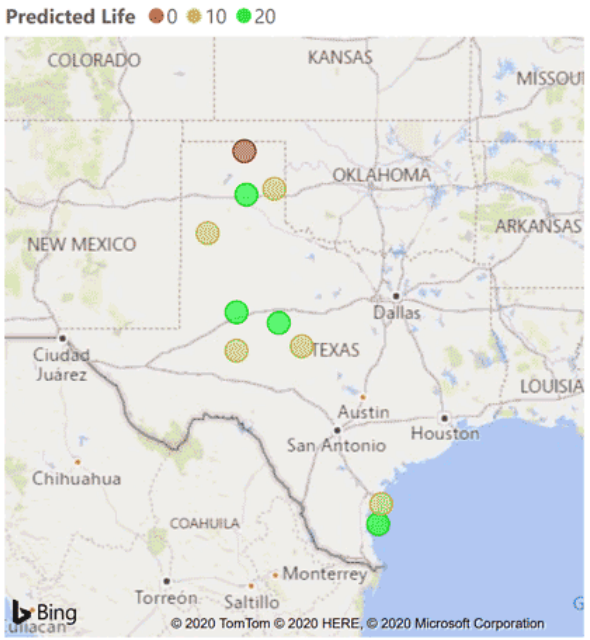
rpm

0.00 73.31 146.62

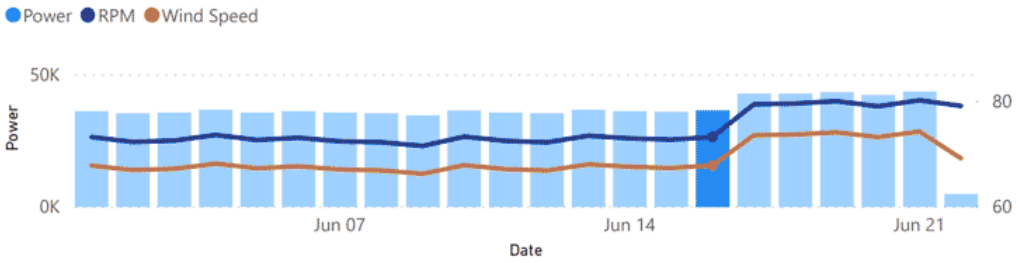
angle

0.00 59.08 118.17

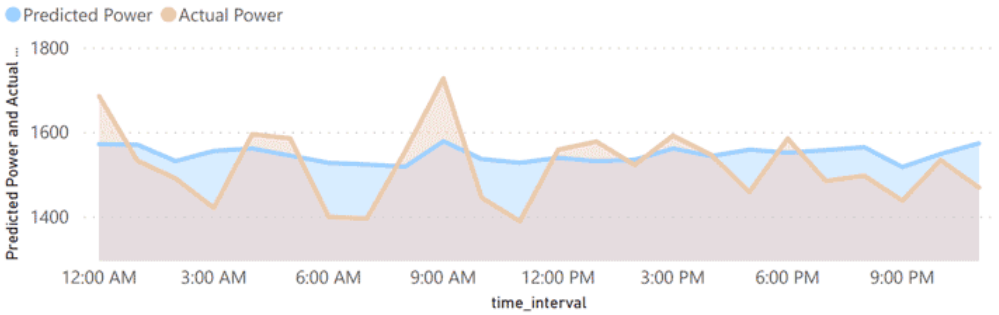
Wind Farms by Predicted Life



Daily Power, RPM and Wind Speed by Date



Hourly Actual vs. Predicted Power



Predictive Maintenance Dashboard

The Data Science Process Flow

1 Get Data from Data Silos

Bring data from data silos to data warehouses (ETL)

MySQL Excel

2 Understand what the business does

The business already does things a certain way. First, match the business to show business impact

3 Data Engineering

Orchestrate and build pipelines to automate the process

Google Cloud Microsoft Azure amazon web services

4 Machine Learning

Models are of three main types: Heuristics-based, Statistical or AI/ML

python Apache Spark R

5 Present to the Business Stakeholders

Analytics is only helpful to the business if it helps them make money. Data Science is Business, Mathematics & Technology.

tableau Power BI Qlik

6 Impact Analysis

Data Science is experimental. Fail fast and build faster to maximize profit

Key Takeaways

01

Predictive Maintenance in the Engineering Industry

02

Real-world Machine Learning (MLOps)

03

Data Science Process Flow

Anomaly Detection in Wind Turbine using Data Science and Big Data

Anish Mahapatra

Q&A Session

LinkedIn: www.linkedin.com/in/**anishmahapatra**

Email: anishmahapatra01@gmail.com

Anomaly Detection in Wind Turbine using Data Science and Big Data

Anish Mahapatra

Q. How to get into Data Science?

A. Simple, ask a real-world Data Scientist.

Thank you.

Email: anishmahapatra01@gmail.com