OCTOBER 26, 2023

Bringing Industrial Operations Data into your Analytics Platform with AVEVA™ Data Hub Data Views

Alicia Coppock and Stephen Christian

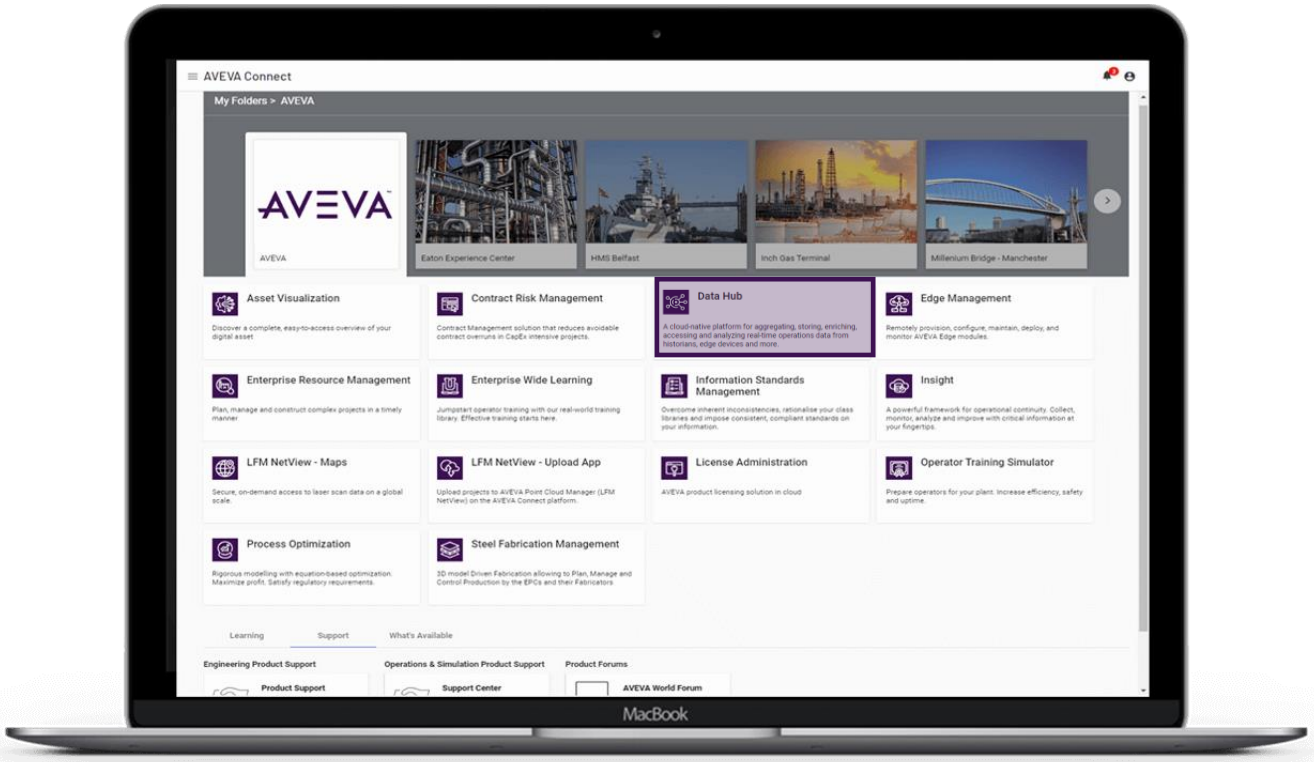
AVEVA

Agenda

- Intro to Data Hub
- Challenges in Data Science
- Enabling Data Science with AVEVA Data Hub
- Demo
- Q&A

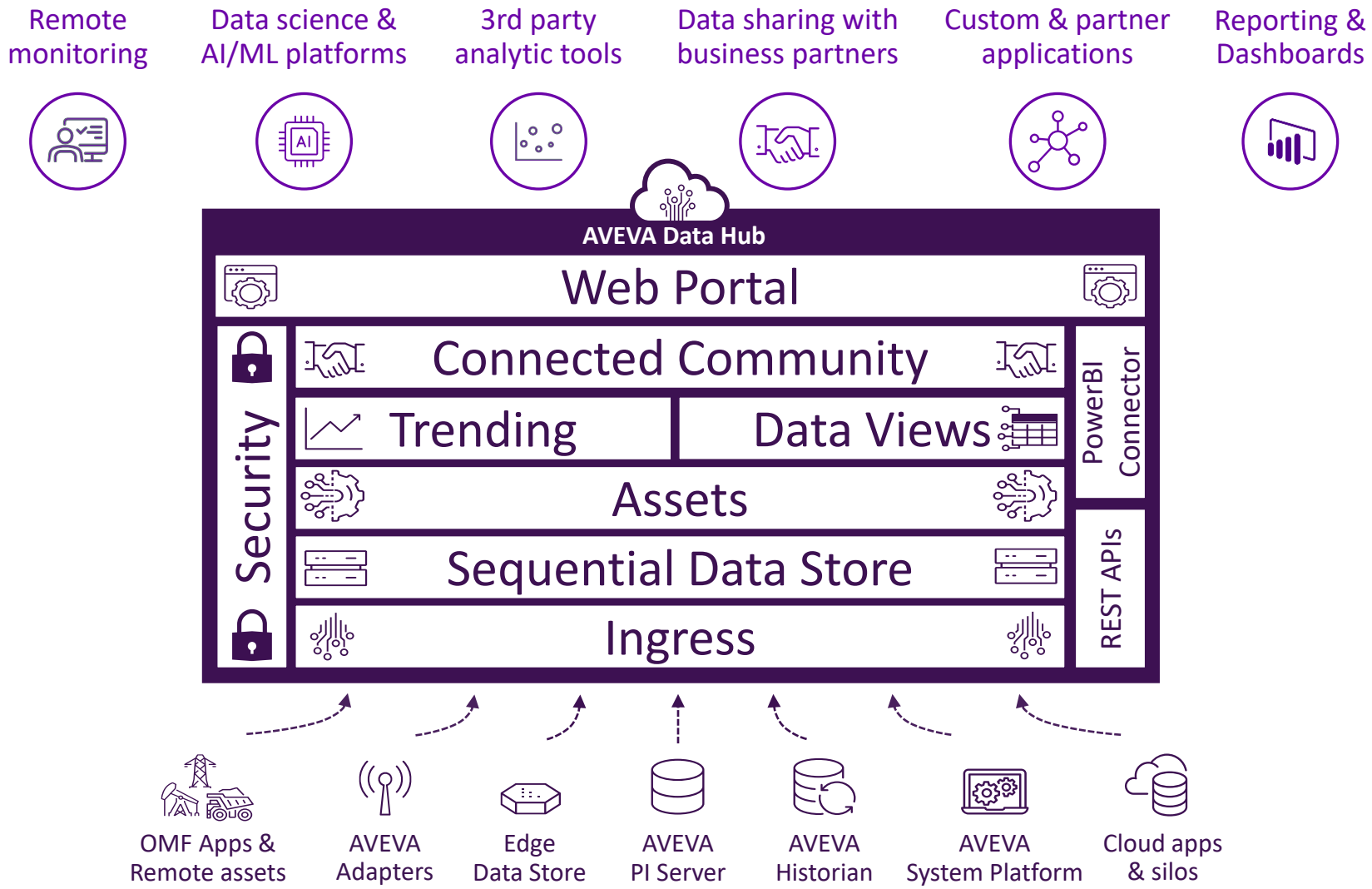
AVEVA Connect: Our industrial cloud platform

Our industrial cloud platform is your digital transformation hub



- ✓ Transform faster
- ✓ Enable new ways of work
- ✓ Connect and collaborate
- ✓ One single source of truth
- ✓ Move from CapEx to OpEx
- ✓ Lower dependency on IT
- ✓ Always the latest and greatest features
- ✓ Easily scale up and down
- ✓ Built with security, resiliency, high-availability

AVEVA Data Hub



A cloud-native industrial platform designed for aggregating, storing, enriching, accessing, analyzing, and securely sharing real-time operations data from historians, edge devices, and more

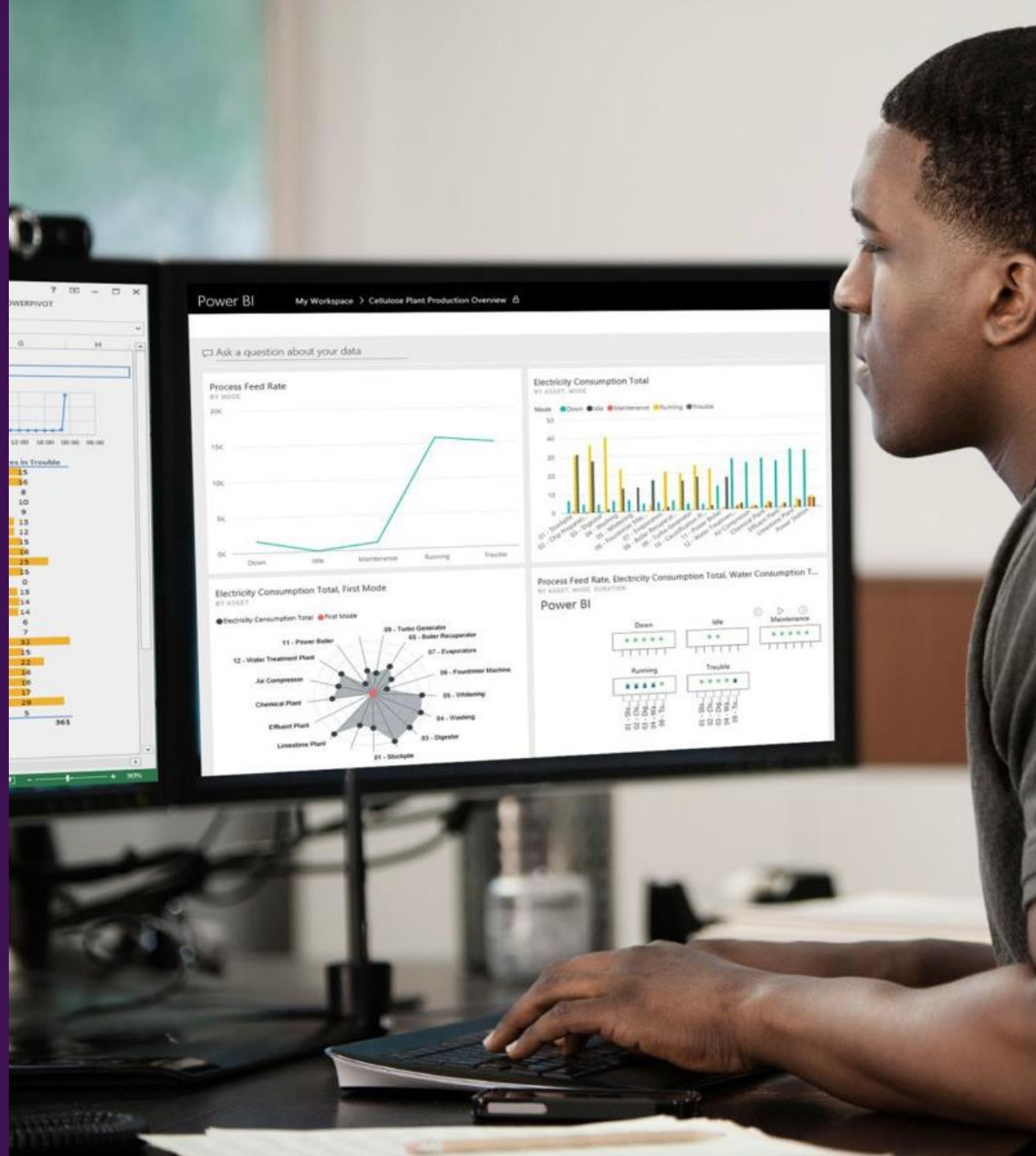
- Managed, secure, multi-tenant platform
- Operated & maintained by AVEVA
- High speed, scalable, elastic, & resilient
- Modern, secure REST APIs
- Built & deployed on Microsoft Azure

Supported Regions
West US (California)
North Europe (Ireland)
Australia East (New South Wales)

AVEVA

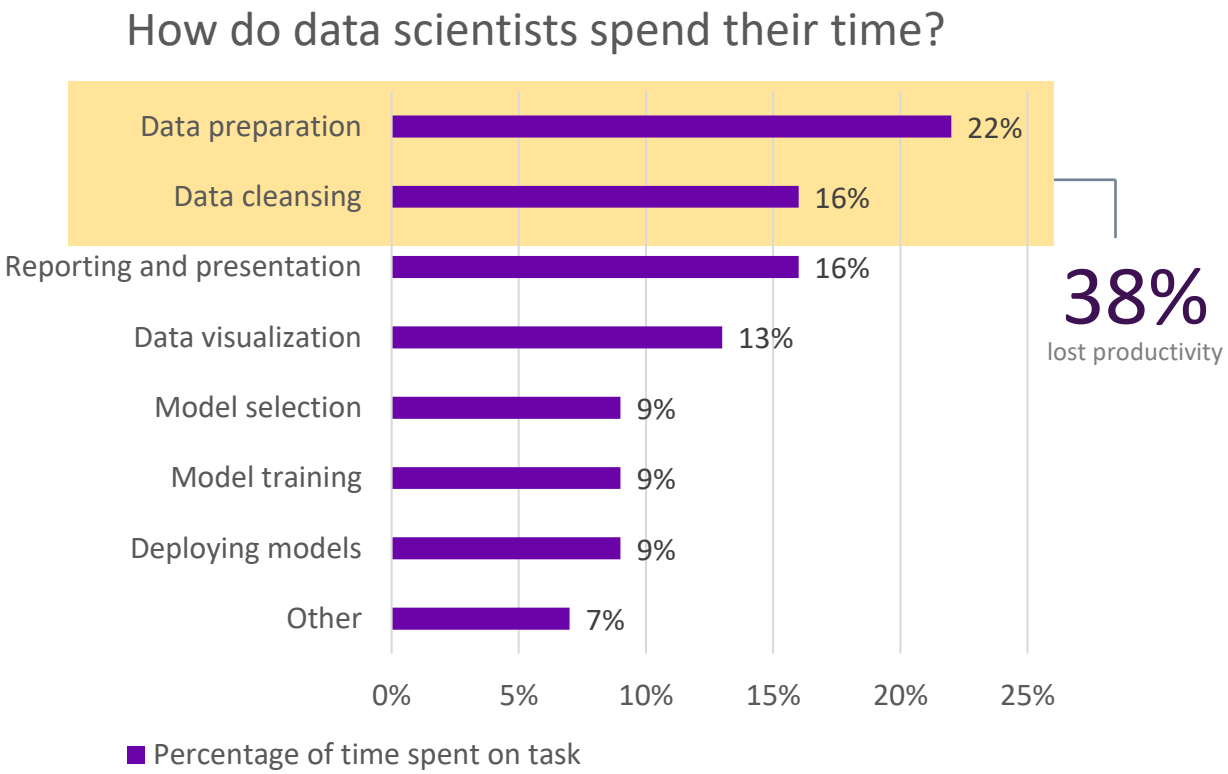
Customers face challenges to realizing the promise of data science

- Easy, secure access to **trustworthy data**
- **Flexible integration** enabling different users and applications
- Support for **enterprise-wide scale**
- **Curated, decision-ready data** that's contextualized, aggregated, and shaped for consumption



Preparing data for analytics

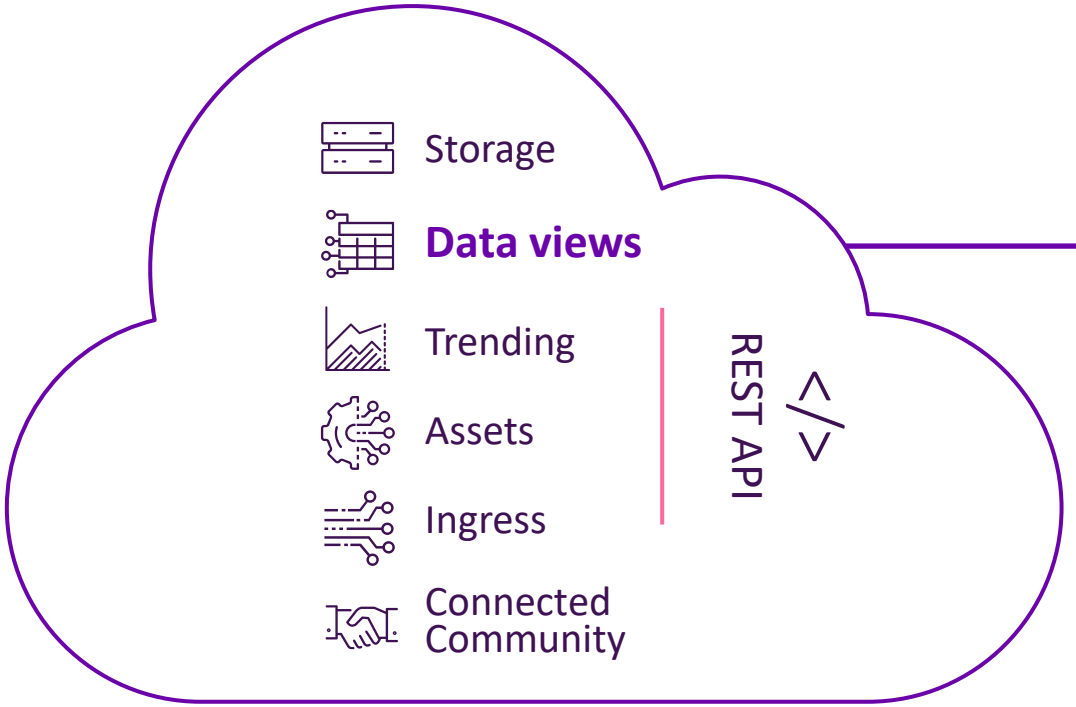
It’s time consuming, even for data scientists



Source: Anaconda’s State of Data Science 2022 Survey

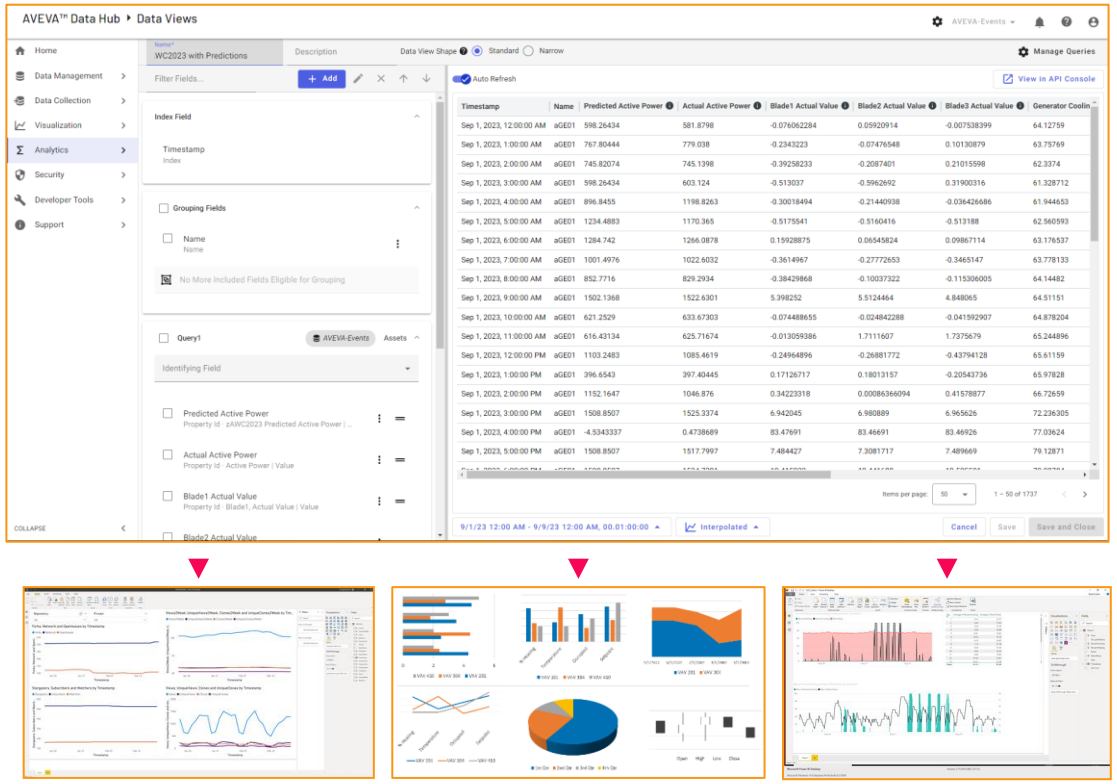
AVEVA

Data Views curate operational data for external consumption

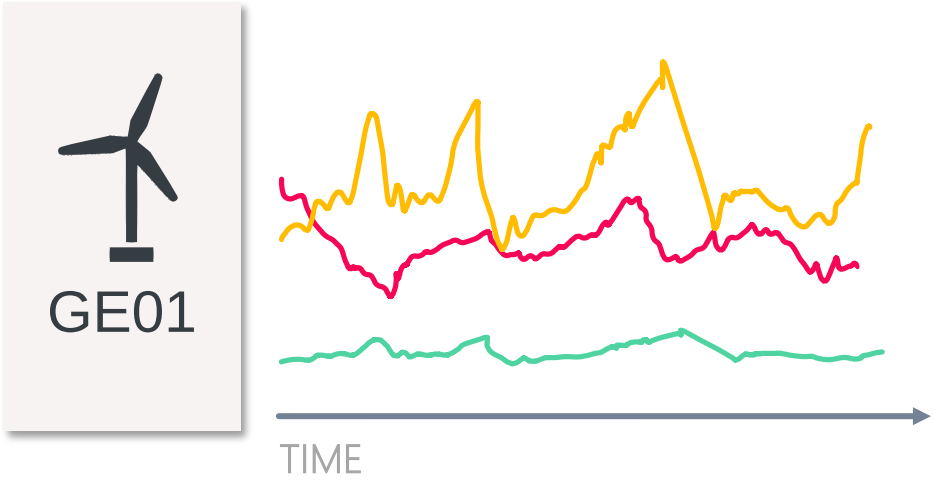


AVEVA Data Hub

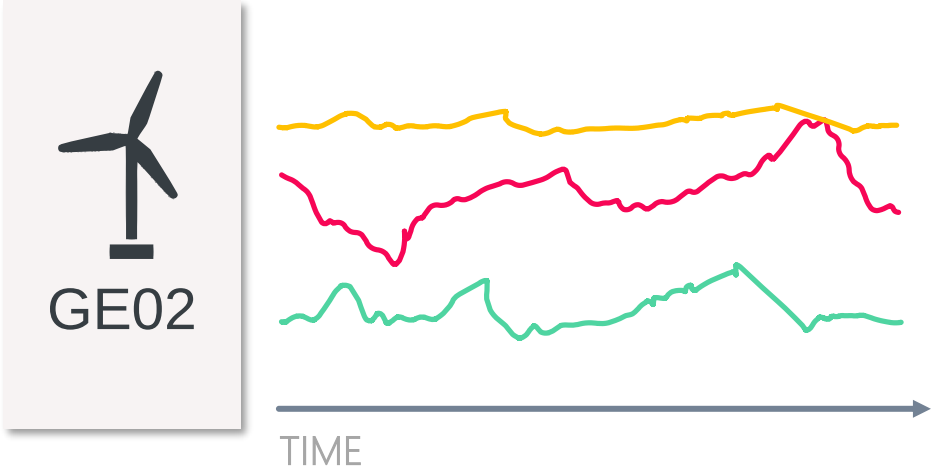
Enabling data exploration and integration



Data is aggregated and context-aware

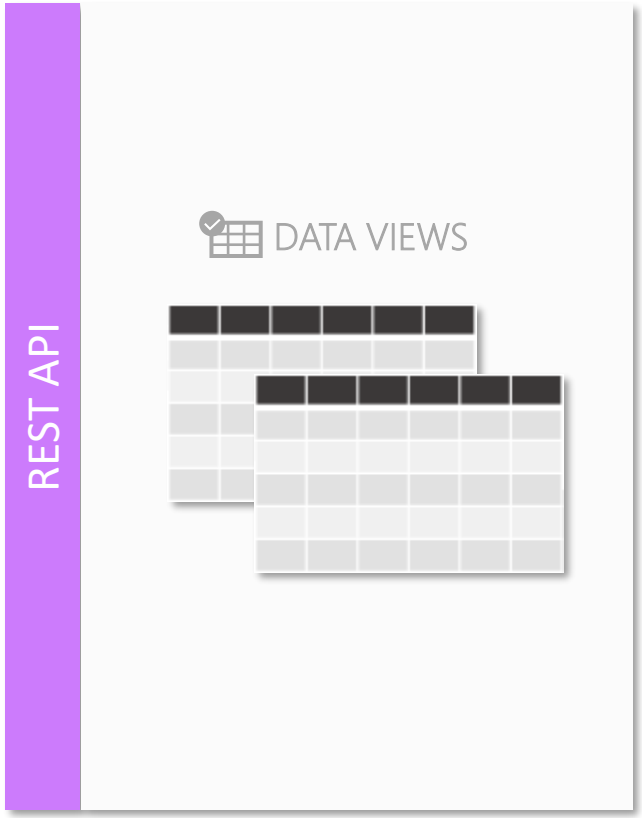


STREAMS	ASSETS + METADATA	
GE01_BL1.ACT_PV	*Turbine Measurement*	**GE01 Blade 1, Actual**
GE01_BL2.ACT_PV	*Turbine Measurement*	**GE01 Blade 2, Actual**
GE01_P.ACT_PV	*Turbine Measurement*	**GE01 Active Power**

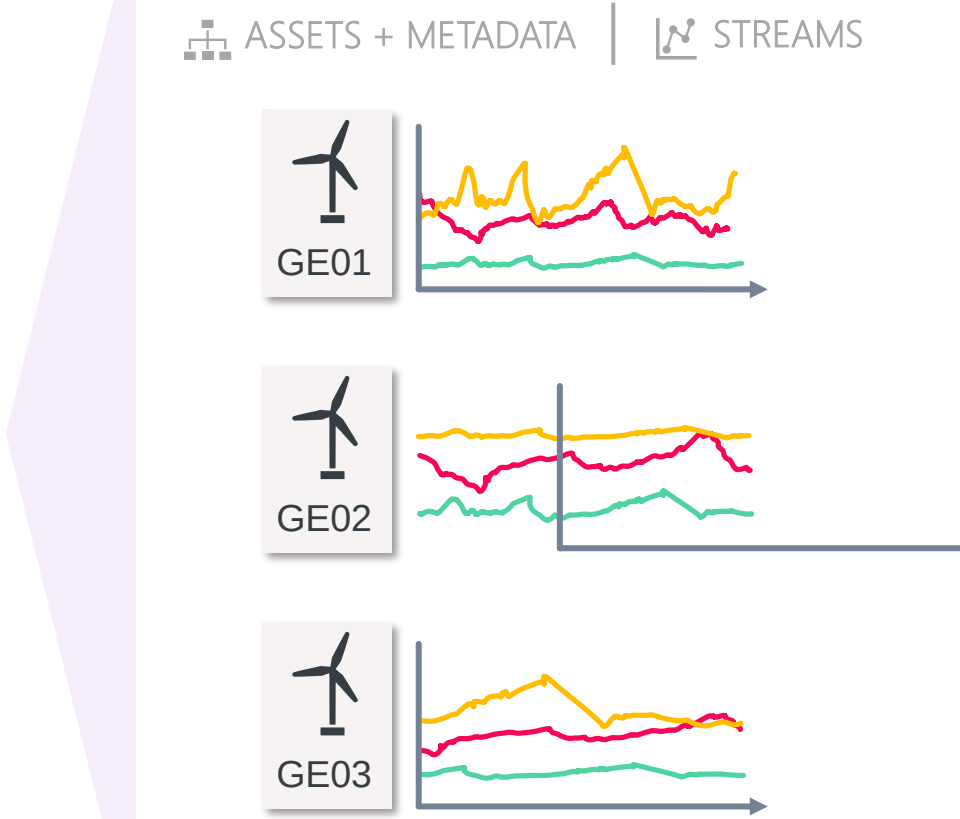


STREAMS	ASSETS + METADATA	
GE02_BL1.ACT_PV	*Turbine Measurement*	**GE02 Blade 1, Actual**
GE02_BL2.ACT_PV	*Turbine Measurement*	**GE02 Blade 2, Actual**
GE02_P.ACT_PV	*Turbine Measurement*	**GE02 Active Power**

Data is shaped for consumption



- Dynamic asset and stream queries
- Data selection & grouping
- Summary calculations
- Streams shared through AVEVA Data Hub Communities



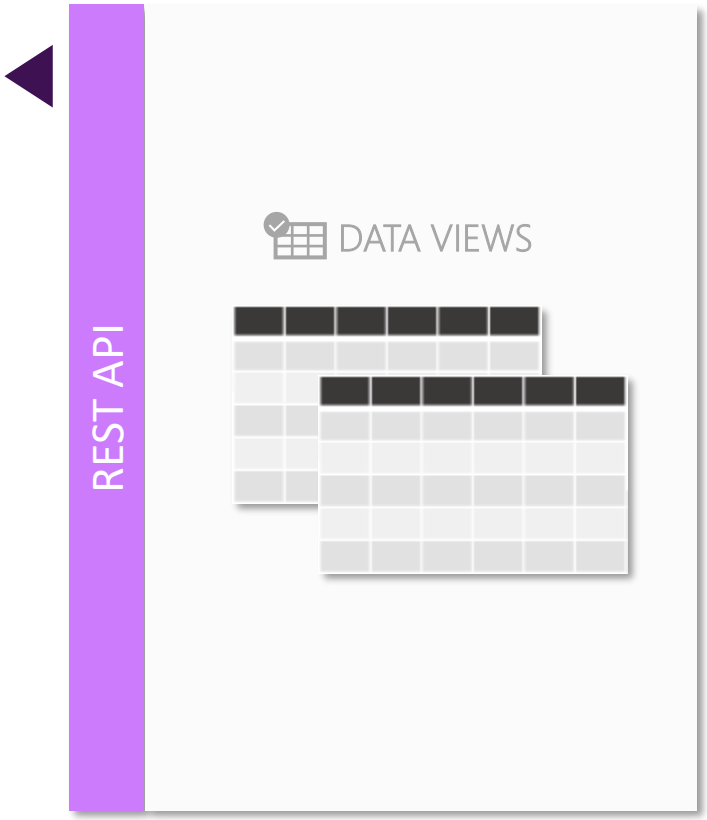
AVEVA

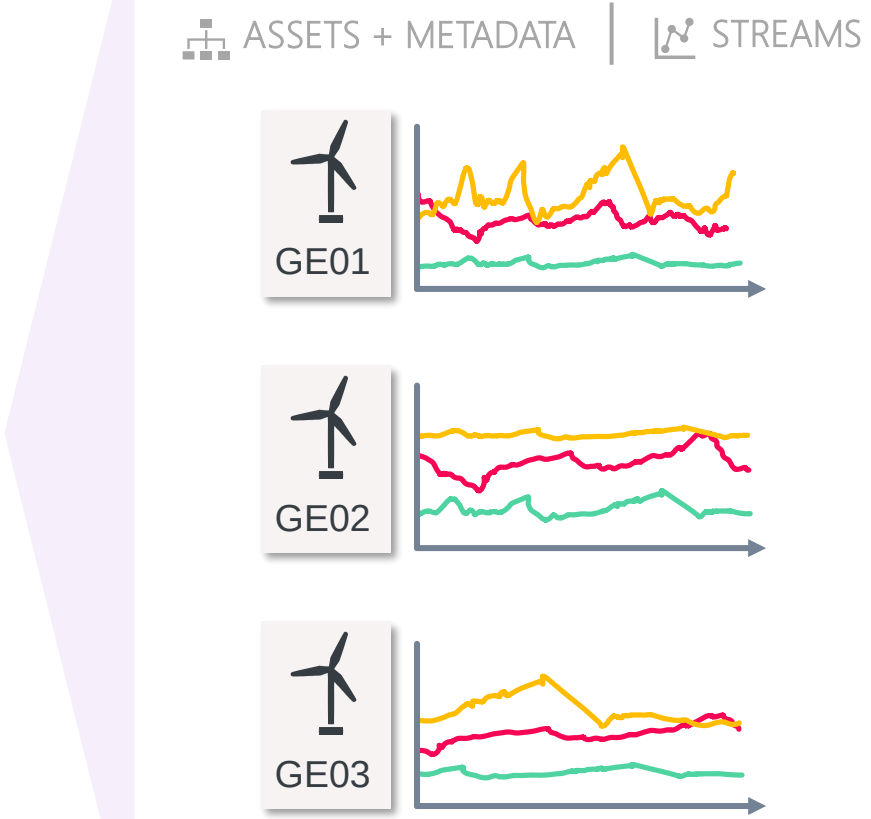
Curated data on-demand

Defined at runtime

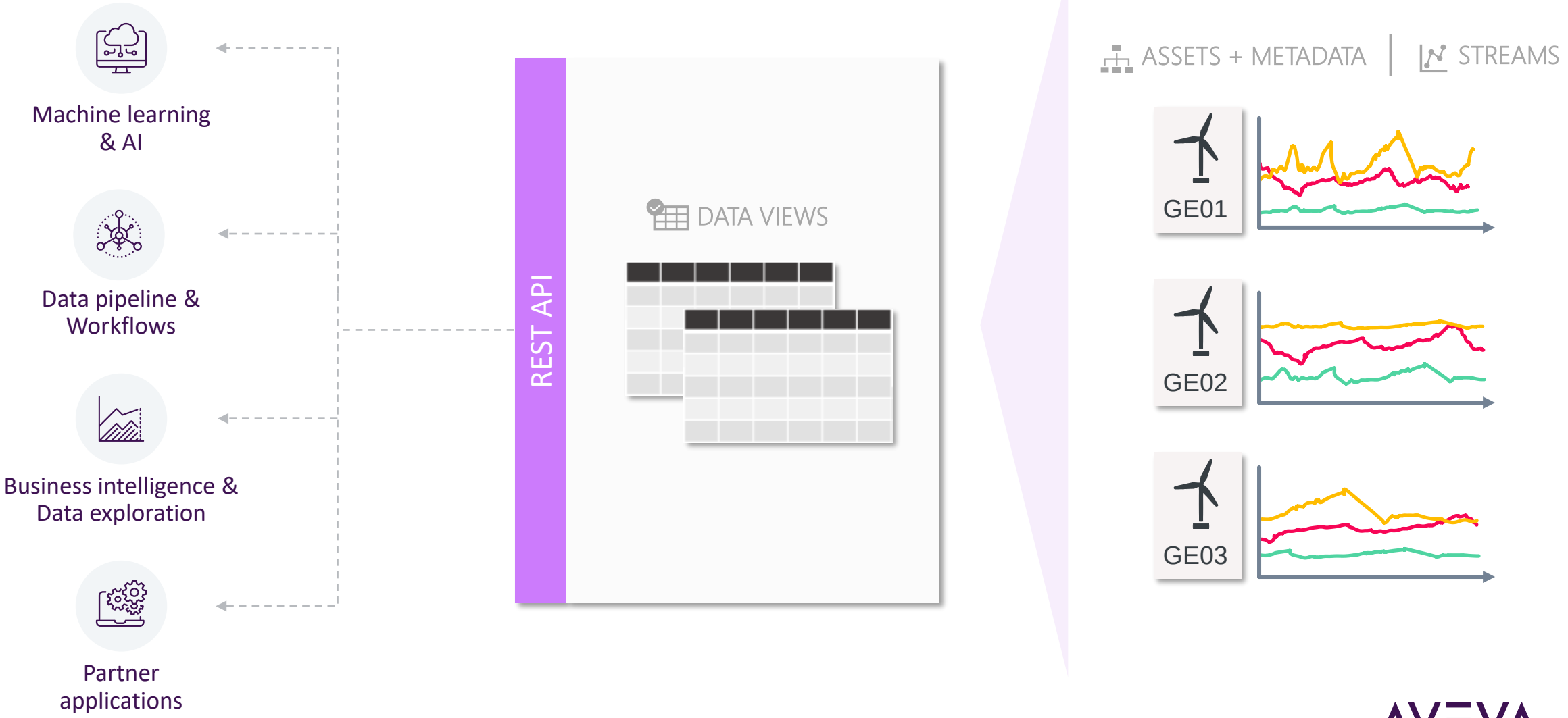
- Time range
- Stored or Interpolated data
- Interpolation interval
- Data format
  - JSON
  - CSV
  - Table

New Parquet File

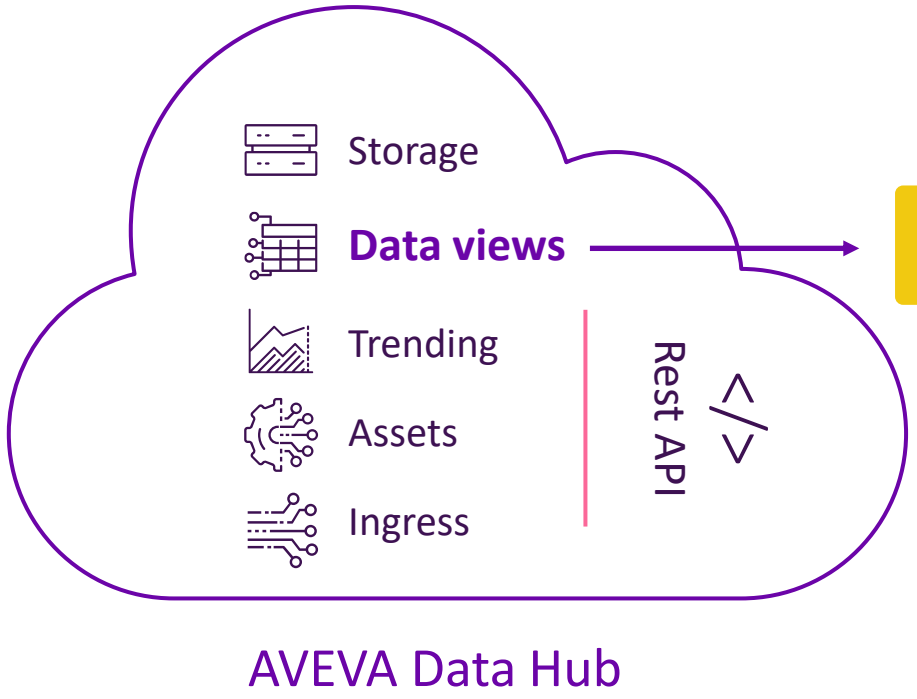




Data is available for many analytics applications



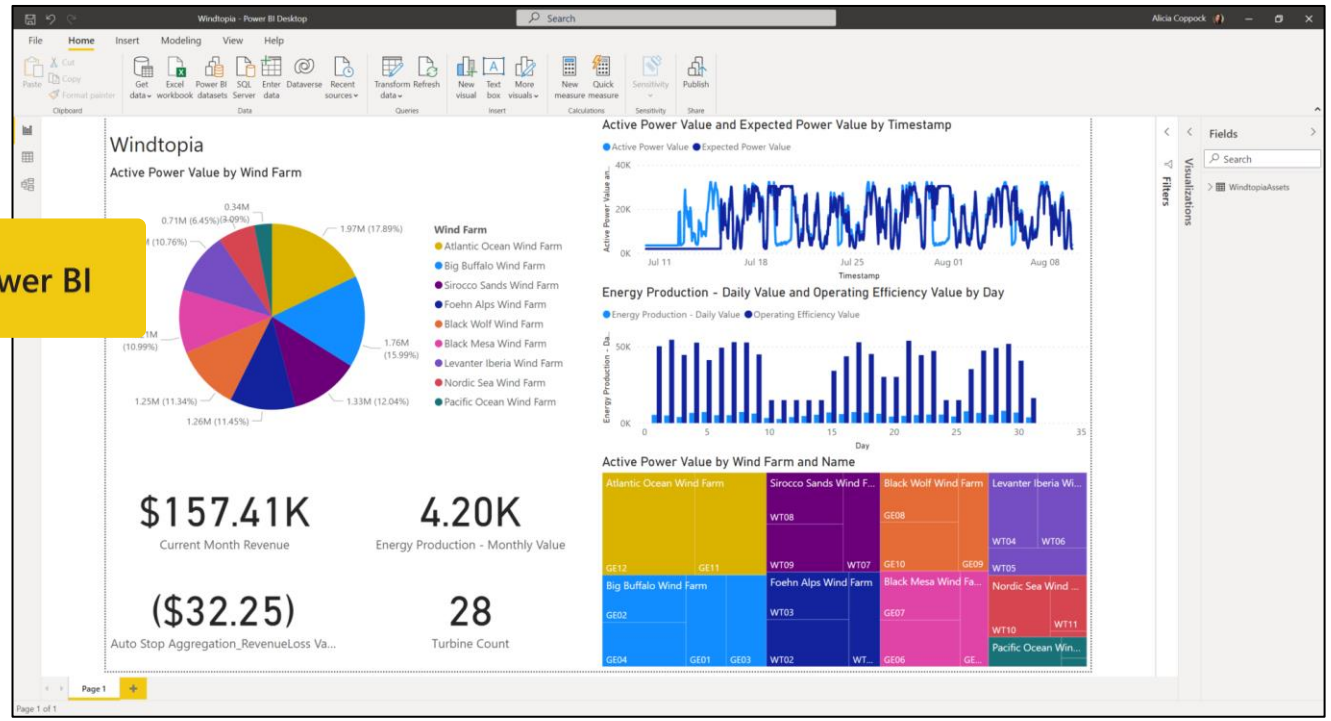
Easily slice & dice your data in Microsoft Power BI



AVEVA Data Hub

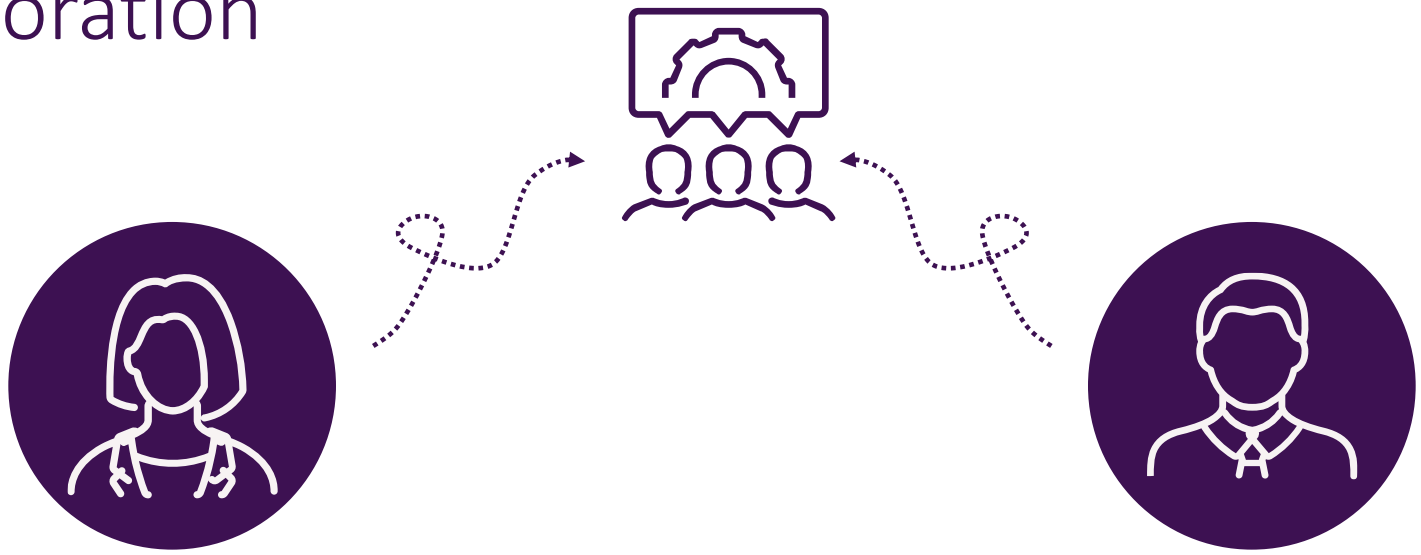
Power BI

AVEVA Data Hub Power BI Connector



New Authentication with Client Credentials

OT/IT Collaboration



Engineer

Data Scientist

- Operations subject matter expert
- Organize and shape data for analysis
- Add meaningful context to data

- Access Data Views
  - REST API
  - AVEVA Data Hub Power BI Connector
- Use AI / ML and visual analytics to enable new insights

Demo: Powering analytics with curated data from AVEVA Data Hub

Stephen Christian

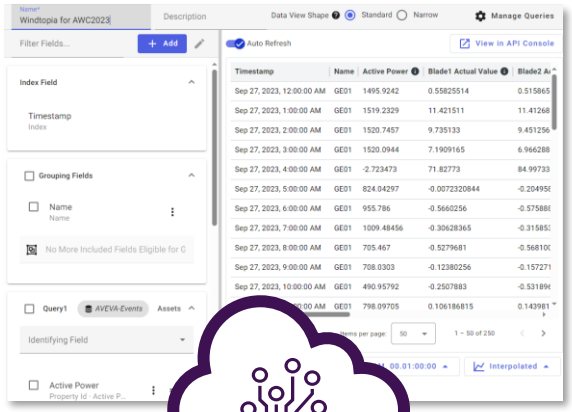
Demo: Forecasting wind farm power production

GE01

GE02

GE03

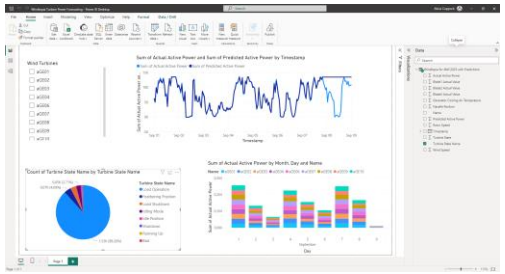
Aggregate and curate data for analysis



AVEVA Data Hub

databricks

Forecast power with machine learning



Power BI

Visualize power forecast report

AVEVA

Home
Data Management
Data Collection
Visualization
Analytics
Security
Developer Tools
Support

COLLAPSE

Name*
Windtopia AWC2023

Description

Data View Shape (Standard) (Narrow)

+ Add

Query1 | AVEVA-Events | Streams

Query Id*
Query1

Query Source
AVEVA-Events

Query Type
(Streams) (Assets)

Query Value
Search...

Search for Streams

Enter a Query Value above to search for streams. For more information see the search documentation.

Items per page: 50 | 0 of 0

Cancel | Save

Microsoft Azure | databricks | Search data, notebooks, recents, and more... CTRL + P | WindtopiaDatabricks | schristian@osisoft.com

New
Workspace
Recents
Catalog
Workflows
Compute
SQL
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
Features
Models
Serving
Marketplace
Partner Connect
Collapse menu

Wind Turbine Analysis

Python

File Edit View Run Help Last edit was 7 days ago Provide feedback

Run all | Terminated | Schedule | Share

Cmd 1

```
!python -m pip install --upgrade pip
!python -m pip install adh_sample_library_preview --upgrade
!python -m pip install mlflow --upgrade
```

```
Requirement already satisfied: pip in /local_disk0/.ephemeral_nfs/envs/pythonEnv-6d9f2831-51c6-4920-8c18-e9fd0c62daf4/lib/python3.9/site-packages (21.2.4)
Collecting pip
  Downloading pip-23.2.1-py3-none-any.whl (2.1 MB)
     |████████████████████████████████| 2.1 MB 30.8 MB/s eta 0:00:01
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 21.2.4
    Uninstalling pip-21.2.4:
      Successfully uninstalled pip-21.2.4
Successfully installed pip-23.2.1
Collecting adh_sample_library_preview
  Downloading adh_sample_library_preview-0.9.13rc0-py3-none-any.whl (116 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.5/116.5 kB 6.0 MB/s eta 0:00:00
Collecting jsonpatch>=1.32 (from adh_sample_library_preview)
  Obtaining dependency information for jsonpatch>=1.32 from https://files.pythonhosted.org/packages/73/07/02e16ed01e04a374e644b575638ec7987ae846d25ad97bcc9945a3ee4b0e/jsonpatch-1.33-py2.py3-none-any.whl.metadata
  Downloading jsonpatch-1.33-py2.py3-none-any.whl.metadata (3.0 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in /databricks/python3/lib/python3.9/site-packages (from adh_sample_library_preview) (2.8.2)
Collecting requests>=2.28.2 (from adh_sample_library_preview)
  Obtaining dependency information for requests>=2.28.2 from https://files.pythonhosted.org/packages/70/8e/0e2d847013cb52cd35b38c009bb167a1a26b2ce6cd6965bf26b47bc0bf44/requests-2.31.0-py3-none-any.whl.metadata
```

Command took 34.63 seconds -- by schristian@osisoft.com at 9/21/2023, 8:58:48 AM on Stephen Christian's Cluster

Cmd 2

```
import mlflow
import mlflow.sklearn
import requests
import json
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from datetime import date, timedelta
from pyspark.sql.functions import col, to_timestamp
from pyspark.pandas import DataFrame
from adh_sample_library_preview import ADHClient

mlflow.sklearn.autolog()
```

New
Workspace
Recents
Catalog
Workflows
Compute
SQL
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
Features
Models
Serving
Marketplace
Partner Connect
Collapse menu

Wind Turbine Analysis

Python | File | Edit | View | Run | Help | Last edit was 7 days ago | Provide feedback | Run all | Terminated | Schedule | Share

Command took 0.81 seconds -- by schristian@osisoft.com at 9/21/2023, 9:05:09 AM on Stephen Christian's Cluster

Cmd 4

```python
def retrieve_data_view(client, namespace_id, data_view_id, start_index, end_index, interval):
  data_page, next_page, first_page = client.DataViews.getDataInterpolated(
    namespace_id, data_view_id, start_index=start_index, end_index=end_index, interval = interval)
  df = spark.read.json(sc.parallelize([json.dumps(data_page)]))

  # Iterate through each subsequent page of results until there are no more pages
  while next_page != None:
    data_page, next_page, first_page = client.DataViews.getDataInterpolated(url=next_page)
    df = df.union(spark.read.json(sc.parallelize([json.dumps(data_page)])))

  return df

df = retrieve_data_view(client, namespace_id, data_view_id, start_index, end_index, interval)
display(df)
```

(17) Spark Jobs

df: pyspark.sql.dataframe.DataFrame = [Active Power: double, Blade1 Actual Value: double ... 9 more fields]

Table +

	Active Power	Blade1 Actual Value	Blade2 Actual Value	Blade3 Actual Value	Generator Cooling Air Temperature	Nacelle Position	Name	Rotor Speed	Timestamp	Turbine State
1	1380.4193	-0.15358979	-0.116629474	-0.110596865	65.60718	132.68161	GE01	19.844124	2023-09-01T00:00:00Z	16
2	990.1805	-0.18358225	-0.03936434	-0.16621605	65.60101	133.75717	GE01	19.729055	2023-09-01T00:01:00Z	16
3	1524.5835	4.0551815	4.009701	4.047217	65.59485	134.83272	GE01	19.954031	2023-09-01T00:02:00Z	16
4	1228.2428	-0.28279698	-0.27758697	-0.25653037	65.588684	135.90828	GE01	19.820591	2023-09-01T00:03:00Z	16
5	1113.4883	0.08640607	0.11865816	0.10602036	65.58252	140.33691	GE01	19.764425	2023-09-01T00:04:00Z	16
6	1265.8788	0.5062732	0.5142392	0.44296938	65.576355	141.93181	GE01	19.988466	2023-09-01T00:05:00Z	16

10,000 rows | Truncated data | 25.20 seconds runtime — Refreshed 8 days ago

Command took 25.20 seconds -- by schristian@osisoft.com at 9/21/2023, 9:05:13 AM on Stephen Christian's Cluster

Cmd 5

```python
# Remove the GE05 turbine rows from the data frame because it has a lower rating relative to all the other turbines
df_filter = df.where(col('Name') != 'GE05')

# Remove any rows with nulls - these will interfere with plots and the machine learning below
```

New
Workspace
Recents
Catalog
Workflows
Compute
SQL
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
Features
Models
Serving
Marketplace
Partner Connect
Collapse menu

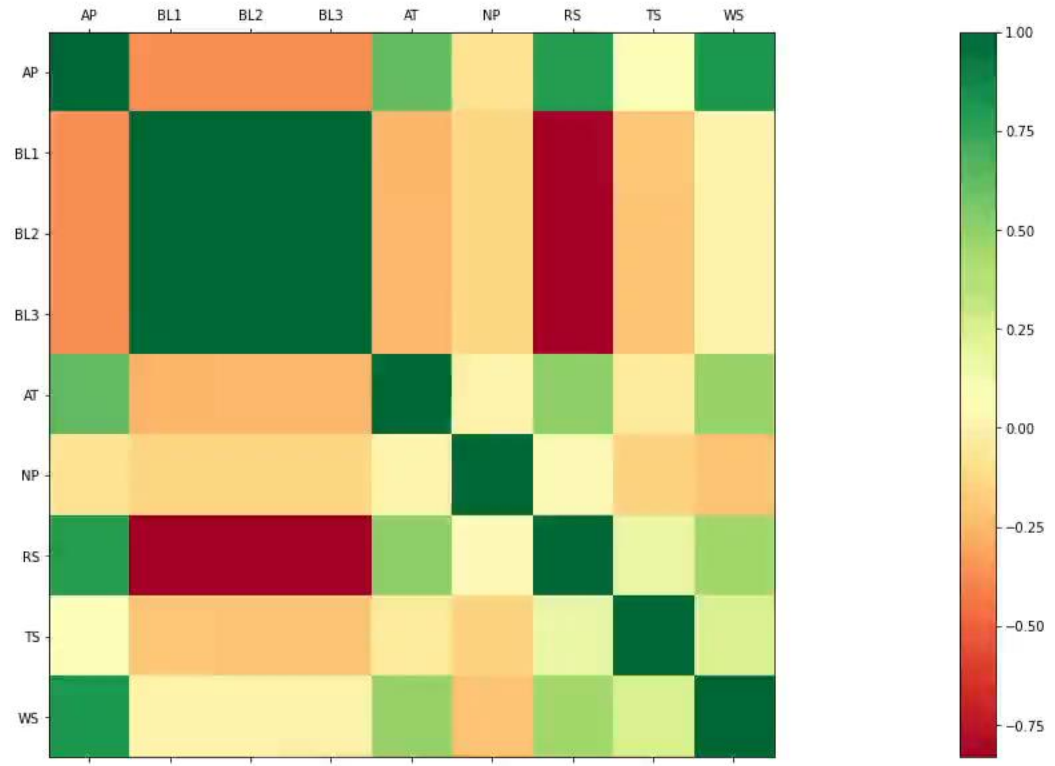
Wind Turbine Analysis

Python

File Edit View Run Help Last edit was 7 days ago Provide feedback

Run all | Terminated | Schedule | Share

```
16
17  # Make sure to show all the labels
18  ax.set_xticks(np.arange(len(labels)))
19  ax.set_yticks(np.arange(len(labels)))
20
21  # Setting labels for the x and y axes of the correlation plot
22  ax.set_xticklabels(labels)
23  ax.set_yticklabels(labels)
24
25  plt.show(block=False)
```



Command took 0.72 seconds -- by schristian@osisoft.com at 9/21/2023, 9:07:11 AM on Stephen Christian's Cluster

Cmd 8

New
Workspace
Recents
Catalog
Workflows
Compute
SQL
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
Features
Models
Serving
Marketplace
Partner Connect
Collapse menu

Wind Turbine Analysis

Python

File Edit View Run Help Last edit was 7 days ago Provide feedback

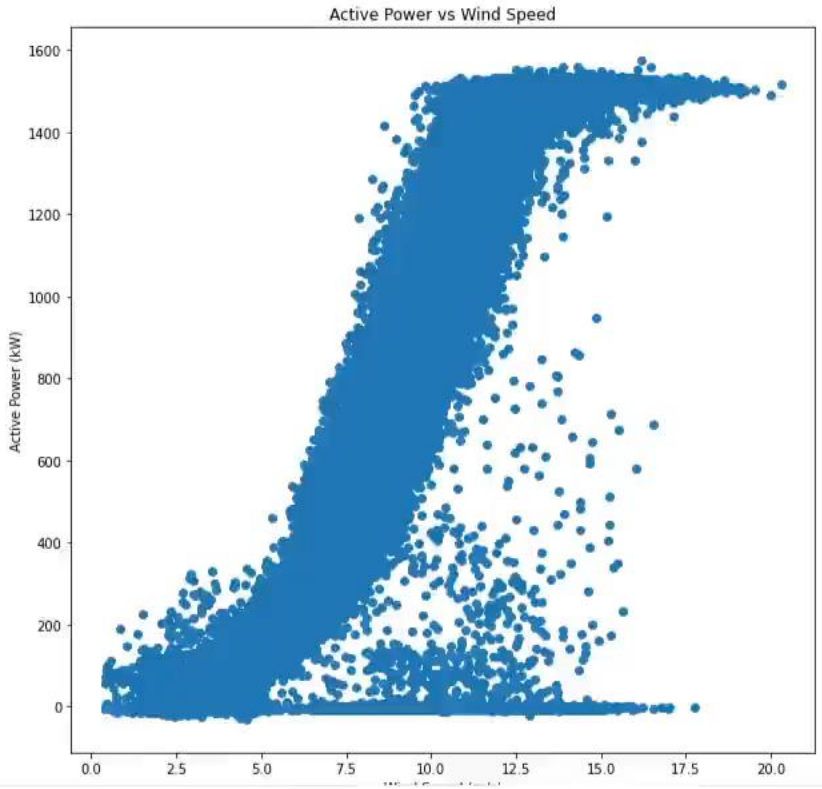
Run all | Terminated | Schedule | Share

Command took 0.72 seconds -- by schristian@osisoft.com at 9/21/2023, 9:07:11 AM on Stephen Christian's Cluster

Cmd 8

```
# Plotting Active Power versus Wind Speed
fig = plt.figure(figsize=(10,10))
ax = fig.add_subplot(111)
ax.scatter(df3['WS'].to_numpy(), df3['AP'].to_numpy())
ax.set_xlabel('Wind Speed (m/s)')
ax.set_ylabel('Active Power (kW)')
ax.set_title('Active Power vs Wind Speed')

plt.show(block=False)
```



Microsoft Azure | databricks   Search data, notebooks, recents, and more... CTRL + P   WindtopiaDatabricks   schristian@osisoft.com

New
Workspace
Recents
Catalog
Workflows
Compute
SQL
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
Features
Models
Serving
Marketplace
Partner Connect
Collapse menu

Wind Turbine Analysis

Python

File Edit View Run Help Last edit was 7 days ago Provide feedback

Run all Terminated Schedule Share

Cmd 9

```python
from sklearn.model_selection import train_test_split

X = df3.filter(items=['BL1', 'BL2', 'BL3', 'AT', 'NP', 'RS', 'TS', 'WS'])
y = df3.filter(items=['AP'])

# Split the dataset randomly into test and train sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
```

Command took 0.13 seconds -- by schristian@osisoft.com at 9/21/2023, 9:07:21 AM on Stephen Christian's Cluster

Cmd 10

```python
# Use the Decision Tree Regression Machine Learning model from scikit-learn
from sklearn.tree import DecisionTreeRegressor

# Try multiple depths of decision tree to see which is best
depths = list(range(1, 21))

regressors = [DecisionTreeRegressor(max_depth=depth) for depth in depths]
for regressor in regressors:
    regressor.fit(X_train, y_train)
```

```
2023/09/21 13:44:04 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:04 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:05 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:05 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:05 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:06 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:06 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:06 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:07 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:08 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/python_shell/scripts/db_ipykernel_launcher.py configure`
2023/09/21 13:44:08 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: You haven't configured the CLI yet! Please configure by entering `/databricks/pyt
```

New
Workspace
Recents
Catalog
Workflows
Compute
SQL
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
Features
Models
Serving
Marketplace
Partner Connect
Collapse menu

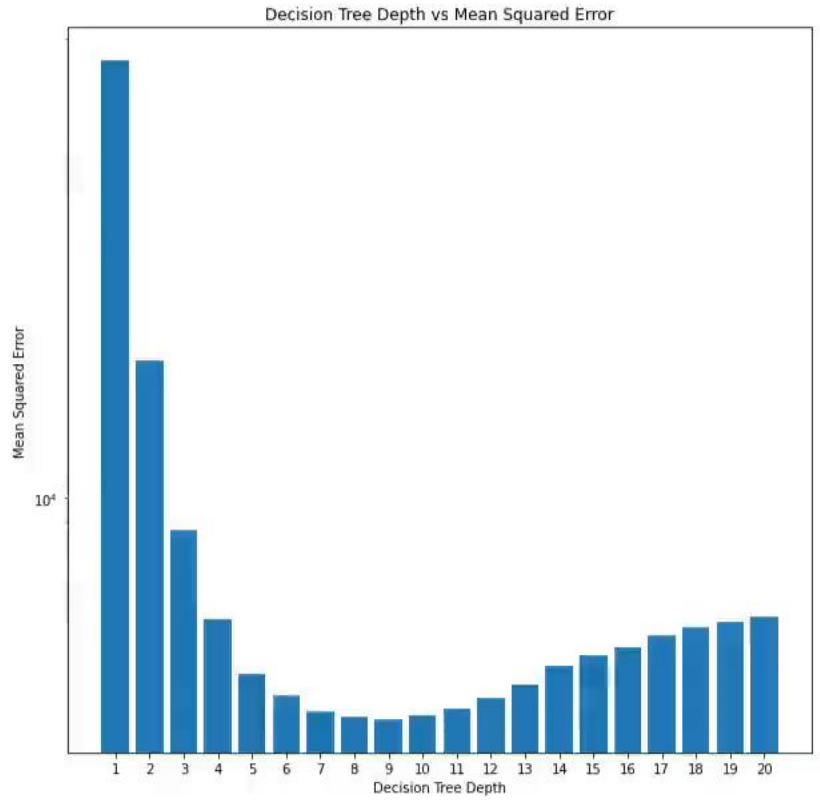
Wind Turbine Analysis

Python

File Edit View Run Help Last edit was 7 days ago Provide feedback

Run all | Terminated | Schedule | Share

```
13 fig = plt.figure(figsize=(10,10))
14 ax = fig.add_subplot()
15 ax.bar([str(depth) for depth in depths], errors)
16 ax.set_yscale('log')
17 ax.set_xlabel('Decision Tree Depth')
18 ax.set_ylabel('Mean Squared Error')
19 ax.set_title('Decision Tree Depth vs Mean Squared Error')
20
21 plt.show(block=False)
```



Command took 0.63 seconds -- by schristian@osisoft.com at 9/21/2023, 9:44:23 AM on Stephen Christian's Cluster

Cmd 12

New
Workspace
Recents
Catalog
Workflows
Compute
SQL
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Machine Learning
Experiments
Features
Models
Serving
Marketplace
Partner Connect
Collapse menu

Wind Turbine Analysis

Python

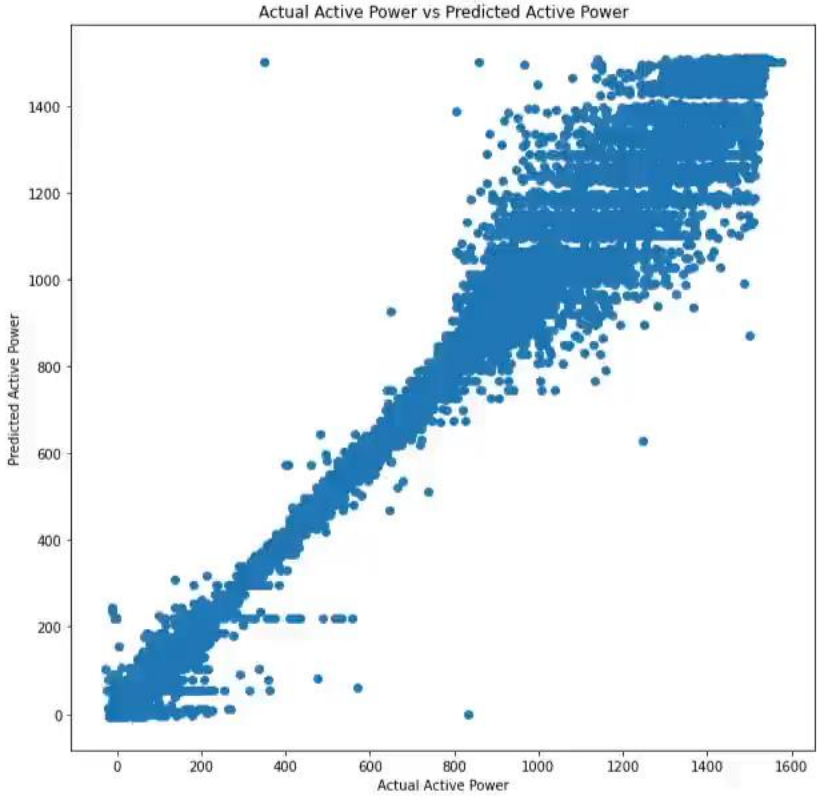
File Edit View Run Help Last edit was 7 days ago Provide feedback

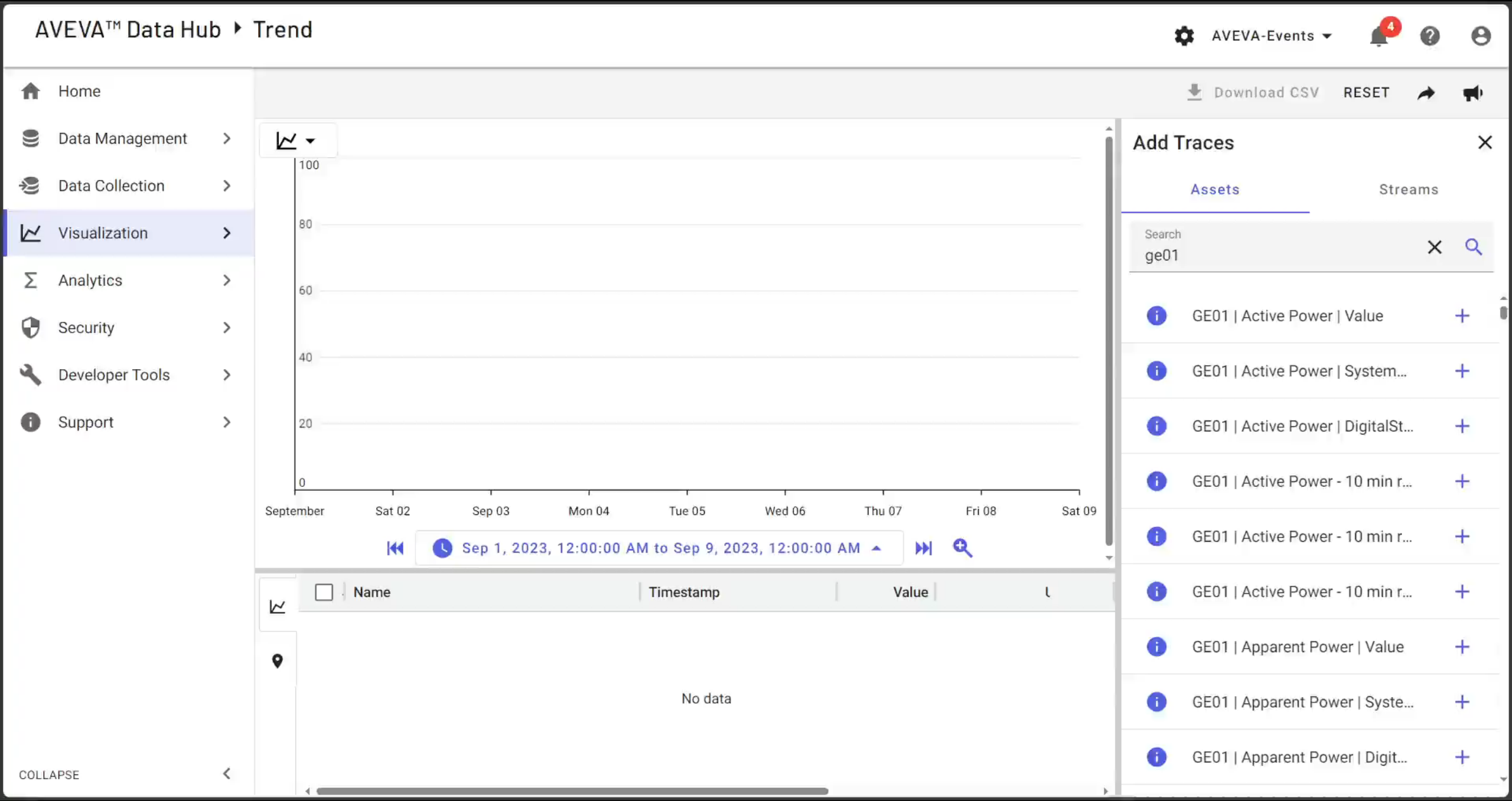
Run all | Terminated | Schedule | Share

```
best_index = np.array(errors).argmin()
best_regressor = regressors[best_index]
y_predict = best_regressor.predict(X_test)

# Plotting Actual Value versus Predicted Value
fig = plt.figure(figsize=(10,10))
ax = fig.add_subplot(111)
ax.scatter(y_test, y_predict)
ax.set_xlabel('Actual Active Power')
ax.set_ylabel('Predicted Active Power')
ax.set_title('Actual Active Power vs Predicted Active Power')

plt.show(block=False)
```




AVEVA™ Data Hub ▸ Trend
AVEVA-Events
4
Home
Data Management
Data Collection
Visualization
Analytics
Security
Developer Tools
Support
COLLAPSE
Download CSV
RESET
100
80
60
40
20
0
September
Sat 02
Sep 03
Mon 04
Tue 05
Wed 06
Thu 07
Fri 08
Sat 09
Sep 1, 2023, 12:00:00 AM to Sep 9, 2023, 12:00:00 AM
Name
Timestamp
Value
No data
Add Traces
Assets
Streams
Search
ge01
GE01 | Active Power | Value
GE01 | Active Power | System...
GE01 | Active Power | DigitalSt...
GE01 | Active Power - 10 min r...
GE01 | Active Power - 10 min r...
GE01 | Active Power - 10 min r...
GE01 | Apparent Power | Value
GE01 | Apparent Power | Syste...
GE01 | Apparent Power | Digit...

Windtopia Turbine Power Forecasting - Power BI Desktop

Search

Alicia Coppock

File Home Insert Modeling View Optimize Help

Paste Cut Copy Format painter — Clipboard

Get data Excel workbook OneLake data hub SQL Server Enter data Dataverse Recent sources — Data

Transform data Refresh — Queries

New visual Text box More visuals — Insert

New measure Quick measure — Calculations

Sensitivity — Sensitivity

Publish — Share

Add data to your report

Once loaded, your data will appear in the **Data** pane.

Import data from Excel

Import data from SQL Server

Paste data into a blank table

Try a sample dataset

Get data from another source →

Filters

Visualizations

Data

Search

You haven't loaded any data yet. Get data

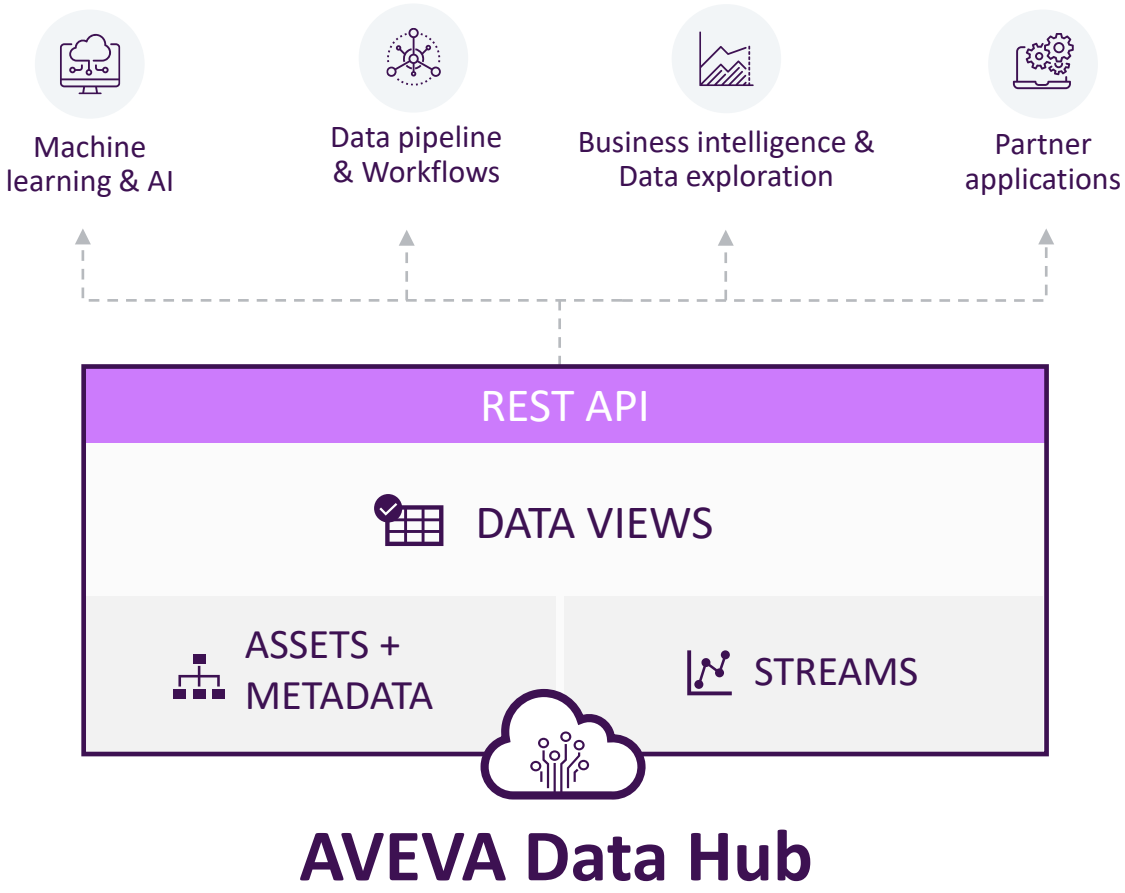
Page 1 Page 2

Page 2 of 2

113%

Enable analytics with AVEVA Data Hub

- Data is aggregated and in context
- Data is curated and shaped for consumption
- Data is fresh and trustworthy
- Data is Secure
- Scalable and resilient
- Flexible integration with 3rd party platforms



Alicia Coppock

Sr Technical Product Manager

- AVEVA
- alicia.coppock@aveva.com

Stephen Christian

Staff Software Developer

- AVEVA
- stephen.christian@aveva.com

This presentation may include predictions, estimates, intentions, beliefs and other statements that are or may be construed as being forward-looking. While these forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could result in actual outcomes differing materially from those projected in these statements. No statement contained herein constitutes a commitment by AVEVA to perform any particular action or to deliver any particular product or product features. Readers are cautioned not to place undue reliance on these forward-looking statements, which reflect our opinions only as of the date of this presentation.

The Company shall not be obliged to disclose any revision to these forward-looking statements to reflect events or circumstances occurring after the date on which they are made or to reflect the occurrence of future events.

linkedin.com/company/aveva

@avevagroup

ABOUT AVEVA

AVEVA is a world leader in industrial software, providing engineering and operational solutions across multiple industries, including oil and gas, chemical, pharmaceutical, power and utilities, marine, renewables, and food and beverage. Our agnostic and open architecture helps organizations design, build, operate, maintain and optimize the complete lifecycle of complex industrial assets, from production plants and offshore platforms to manufactured consumer goods.

Over 20,000 enterprises in over 100 countries rely on AVEVA to help them deliver life's essentials: safe and reliable energy, food, medicines, infrastructure and more. By connecting people with trusted information and AI-enriched insights, AVEVA enables teams to engineer efficiently and optimize operations, driving growth and sustainability.

Named as one of the world's most innovative companies, AVEVA supports customers with open solutions and the expertise of more than 6,400 employees, 5,000 partners and 5,700 certified developers. The company is headquartered in Cambridge, UK.

Learn more at www.aveva.com

AVEVA