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

Alicia Coppock and Stephen Christian
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

Annotations:
- Remote monitoring
- Data science & AI/ML platforms
- 3rd party analytic tools
- Data sharing with business partners
- Custom & partner applications
- Reporting & Dashboards

Description:
AVEVA Data Hub is 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.

Key Features:
- 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)
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

How do data scientists spend their time?

- Data preparation: 22%
- Data cleansing: 16%
- Reporting and presentation: 16%
- Data visualization: 13%
- Model selection: 9%
- Model training: 9%
- Deploying models: 9%
- Other: 7%

Source: Anaconda’s State of Data Science 2022 Survey
Data Views curate operational data for external consumption

Enabling data exploration and integration

Storage
Data views
Trending
Assets
Ingress
Connected
Community

AVEVA Data Hub

© 2022 AVEVA Group plc and its subsidiaries. All rights reserved.
Data is aggregated and context-aware

<table>
<thead>
<tr>
<th>STREAMS</th>
<th>ASSETS + METADATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE01_BL1.ACT_PV</td>
<td><strong>Turbine Measurement</strong></td>
</tr>
<tr>
<td>GE01_BL2.ACT_PV</td>
<td><strong>Turbine Measurement</strong></td>
</tr>
<tr>
<td>GE01_P.ACT_PV</td>
<td><strong>Turbine Measurement</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STREAMS</th>
<th>ASSETS + METADATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE02_BL1.ACT_PV</td>
<td><strong>Turbine Measurement</strong></td>
</tr>
<tr>
<td>GE02_BL2.ACT_PV</td>
<td><strong>Turbine Measurement</strong></td>
</tr>
<tr>
<td>GE02_P.ACT_PV</td>
<td><strong>Turbine Measurement</strong></td>
</tr>
</tbody>
</table>
Data is shaped for consumption

• Dynamic asset and stream queries
• Data selection & grouping
• Summary calculations
• Streams shared through AVEVA Data Hub Communities
Curated data on-demand

Defined at runtime

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

New REST API

DATA VIEWS

ASSETS + METADATA | STREAMS

GE01

GE02

GE03
Data is available for many analytics applications

- Machine learning & AI
- Data pipeline & Workflows
- Business intelligence & Data exploration
- Partner applications

[Diagram showing REST API and data views with ASSETS + METADATA and STREAMS]
Easily slice & dice your data in Microsoft Power BI

AVEVA Data Hub Power BI Connector

Storage
Data views
Trending
Assets
Ingress

Rest API

New Authentication with Client Credentials
OT/IT Collaboration

- 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

Aggregate and curate data for analysis

Forecast power with machine learning

Visualize power forecast report
def retrieve_data_view(client, namespace_id, data_view_id, start_index, end_index, interval):
    data_page, next_page, first_page = client.DataViews.getDataInterpolated(url=url_interpolated)
    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)

# Columns: (Active Power, double; Blade1 Actual Value, double; Blade2 Actual Value, double; Blade3 Actual Value, double; Generator Cooling Air Temperature, double; Nacelle Position, double; Name, string; Rotor Speed, number; Timestamp, string; Turbine State, double)

# Remove the GEBS turbine rows from the data frame because it has a lower rating relative to all other turbines
df = df.filter(df['Name'] != 'GEBS')

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

import matplotlib.pyplot as plt

import numpy as np

labels = ['AP', 'BL1', 'BL2', 'BL3', 'AT', 'NP', 'RS', 'TS', 'WS']

ax = plt.subplots(figsize=(10, 10))

# Make sure to show all the labels
ax.set_sticks(np.arange(len(labels)))
ax.set_yticks(np.arange(len(labels)))

# Setting labels for the x and y axes of the correlation plot
ax.set_labels(labels)
ax.set_yticklabels(labels)

plt.show(block=True)
# Plotting Active Power versus Wind Speed

```python
fig = plt.figure(figsize=(10, 10))
ax = fig.add_subplot(111)
ax.scatter(x, y, c='blue', s=10, alpha=0.5)
ax.set_xlabel('Active Power (kW)')
ax.set_ylabel('Wind Speed (m/s)')
ax.set_title('Active Power vs Wind Speed')
plt.show()
```
from sklearn.model_selection import train_test_split
X = df1.filter(items=['BL1', 'BL2', 'BL3', 'AT', 'NP', 'RS', 'TS', 'WS'])
y = df1.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 0.12 seconds -- by schristian@oss.com on 9/21/2023, 9:01:21 AM on Stephen Christian's Cluster

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)
```python
fig = plt.figure(figsize=(18,10))
ax = fig.add_subplot()
ax.barh(str(depth) for depth in depths, errors)
ax.set_xlabel('Decision Tree Depth')
ax.set_ylabel('Mean Squared Error')
ax.set_title('Decision Tree Depth vs Mean Squared Error')
plt.show(block=False)
```
```python
best_index = np.argmax(errors)
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)
```
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 →
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.
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

© 2023 AVEVA Group plc and its subsidiaries. All rights reserved.