

PHEMI-Semios Architecture Integration Guide

Contents

- Introduction..... 3**
- Proposed Solution Architecture..... 4**
- Configuration..... 6**
 - Users..... 6
 - Data Sources..... 6
 - Data Processing Functions..... 7
 - Data Ingestions..... 7
 - Data Aggregations..... 7
 - Data Source User Interface..... 8

Introduction

This document specifies the high level architecture for Semios system integration with PHEMI Central. It describes the Semios - PHEMI Central system integration points and their use, and the deployment and configuration of PHEMI Central for Semios.

M2M integration with PHEMI Central is achieved with the PHEMI REST interface, which is also used internally by PHEMI Central's Web and Command Line interfaces. The PHEMI REST interface provides the capability to write data into the system (ingest) and read data from the system (query).

The Semios system uses PHEMI Central in a headless configuration utilizing the PHEMI REST interface. Semios raw sensor data is ingested into the system using this interface. Queries are conducted on Semios raw and derived data (e.g. sensor aggregation) using the same interface. For example, the Semios App Server queries PHEMI Central to provide analytical information back to users to satisfy the Semios “Read Logs” and “Drill Down” use cases.

During ingestion the sensor data are stored, processed and aggregated. The parsing, processing and aggregation tasks are done by custom Data Processing Functions (DPFs). The resulting data structures support the following calculations at scale by providing appropriate sensor aggregations:

- Degree Days
- Frost
- Fireblight

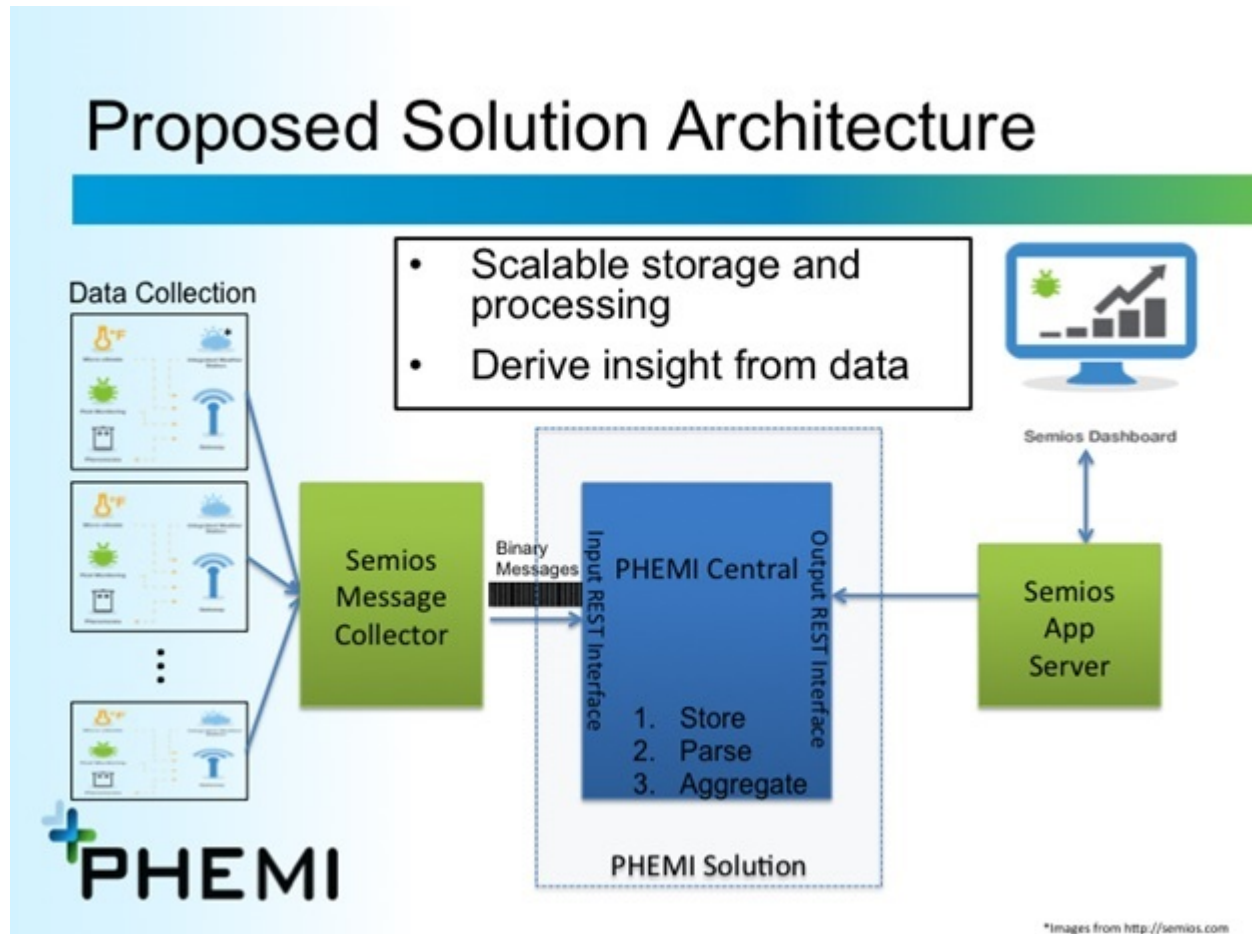
PHEMI Central provides efficient and scalable methods utilizing aggregations to calculate running averages, summations, minimums and maximums of the Semios data. To describe sets of aggregated data, sets are tagged with user defined labels, such as ‘Block’, ‘Property’ or ‘Crop Type’. By using the aggregations, the task of calculating combinatory values is aided and accelerated by PHEMI Central, simplifying the Semios application.

The following diagram illustrates the high level architecture of the Semios-PHEMI integration. The flow of data between PHEMI Central and the Semios components is managed by the PHEMI REST interface.

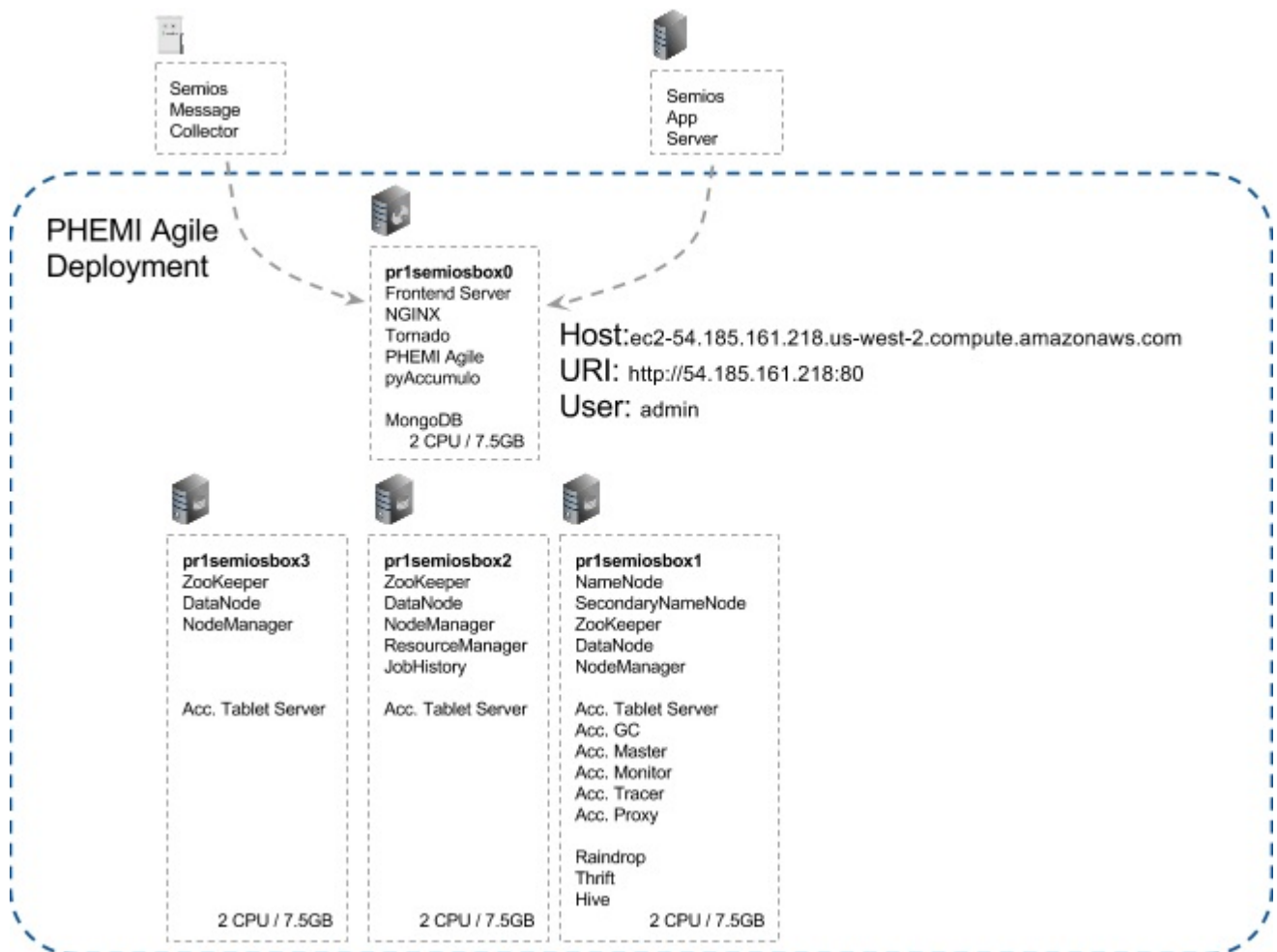
Proposed Solution Architecture

Architecture, deployment, and infrastructure of proposed solution

The following diagram illustrates the high level architecture of the Semios-PHEMI integration.



This architecture is deployed with the following infrastructure details



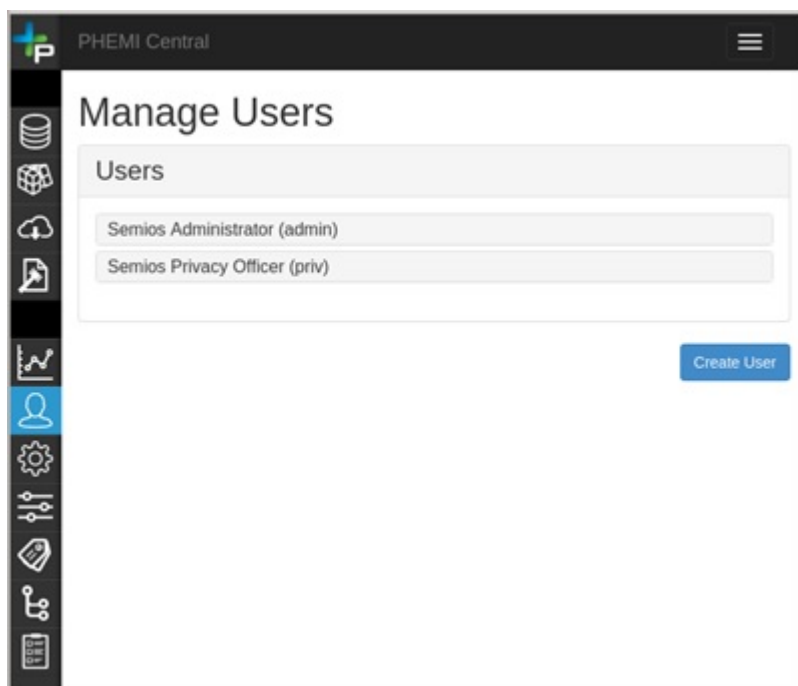
Configuration

asdf

Users

PHEMI Central is configured for Semios with a PHEMI administrator user **admin** that has Administrator and Analyst authorizations. A privacy officer **priv** also exists.

The following is a screenshot of the user interface showing the pre-populated Semios users.



Data Sources

A **Semios Sensors** data source is created and configured with a category of **Semios**. Its Institutional Owner and Source owner are set to the **admin** user. The privacy officer for the data source is set to the **priv** user. The document format is **Binary** and the data source retention rule is **Do not delete**.

- **When new data arrives:** When data is ingested, the DPF is run immediately.
- **Manually:** When data is ingested, the DPF is not run until the user manually runs the DPF. This is done by hitting the “Execute” button on the Data Processing Function tab of the Datasource page.
- **Periodic:** The DPF is run on a periodic basis as specified by the periodic value, regardless if data has been ingested or not; the DPF processes ingested data starting at the first incomplete file.

The trigger setting for the Semios data processing function is set to ***Periodic for 5 Minutes***. This means that when Semios data is ingested, derived data will not necessarily be generated immediately, but generated in 5 minute intervals.

Data Processing Functions

Data Processing Functions (DPFs) are instructions for parsing a document (e.g. a sensor log message) into digital assets such as a temperature reading. DPFs are discussed in more detail in another document.

The PHEMI-developed DPFs read and process Semios sensor log messages to produce digital assets that support Semios use cases. The DPFs perform additional calculations to facilitate efficient retrieval of data assets at scale by aggregating sensor readings.

To improve efficiency for many small files, it is preferred that DPFs process many ingested files before terminating. DPFs can be configured to run in different ways:

- When new data arrives: When data is ingested, the DPF is run immediately.
- Manually: When data is ingested, the DPF is not run until the user manually runs the DPF. This is done by hitting the “Execute” button on the Data Processing Function tab of the Datasource page.
- Periodic: The DPF is run on a periodic basis as specified by the periodic value, regardless if data has been ingested or not; the DPF processes ingested data starting at the first incompleting file.

The trigger setting for the Semios data processing function is set to **Periodic for 5 Minutes**. This means that when Semios data is ingested, derived data will not necessarily be generated immediately, but generated in 5 minute intervals.

Data Ingestions

The process of inputting data into the system is referred to as *ingestion*. The system will ingest and process input documents. The document ingested by the system is considered the *raw data*.

As part of document ingestion, the system will annotate the document with *meta data* which describes the document. Meta data is not generated by examining the contents of the ingested document, but rather by looking at the document from an external point of view. For example, meta data includes descriptive information such as the data source the document was ingested into, the timestamp when the document was ingested, the name of the document and more. The full list of descriptive meta data is provided in the [DPF Meta Data](#) section in the Appendix.

The ingested document can be processed further with a Data Processing Function (DPF). The data generated by the DPF is considered *derived data*. The DPF generated data is done by analyzing the ingested document and producing additional information. Any information that requires analyzing the contents of the document is considered derived data such as data mined information, calculated information and inferred information. For a list of derived data and their types for Semios, please see the [DPF Processed Tags](#) and [DPF Processed Data](#) sections in the Appendix.

The DPF can also generate information that system will aggregate over specific time based *bins*. The data generated by the system based on the information provided by the DPF is considered *aggregated data*. The bin types specified in [Query Aggregate Data](#). The aggregated data is specified in [Data Aggregations](#).

To summarize, there are four types of data:

- Raw Data: The input document provided by an external entity
- Metadata: Descriptive information about the input document gathered without examining the contents of the input document
- Derived Data: Data mined, calculated or inferred data generated by a DPF by analyzing the contents of the input document
- Aggregated Data: Calculated time-based data

Data Aggregations

PHEMI Central supports aggregations over multiple contiguous readings from the same sensor and use them to compute the following derived values:

- number of data points in the aggregation (count)
- minimum reading
- maximum reading

- sum of readings
- sum of squares for the readings

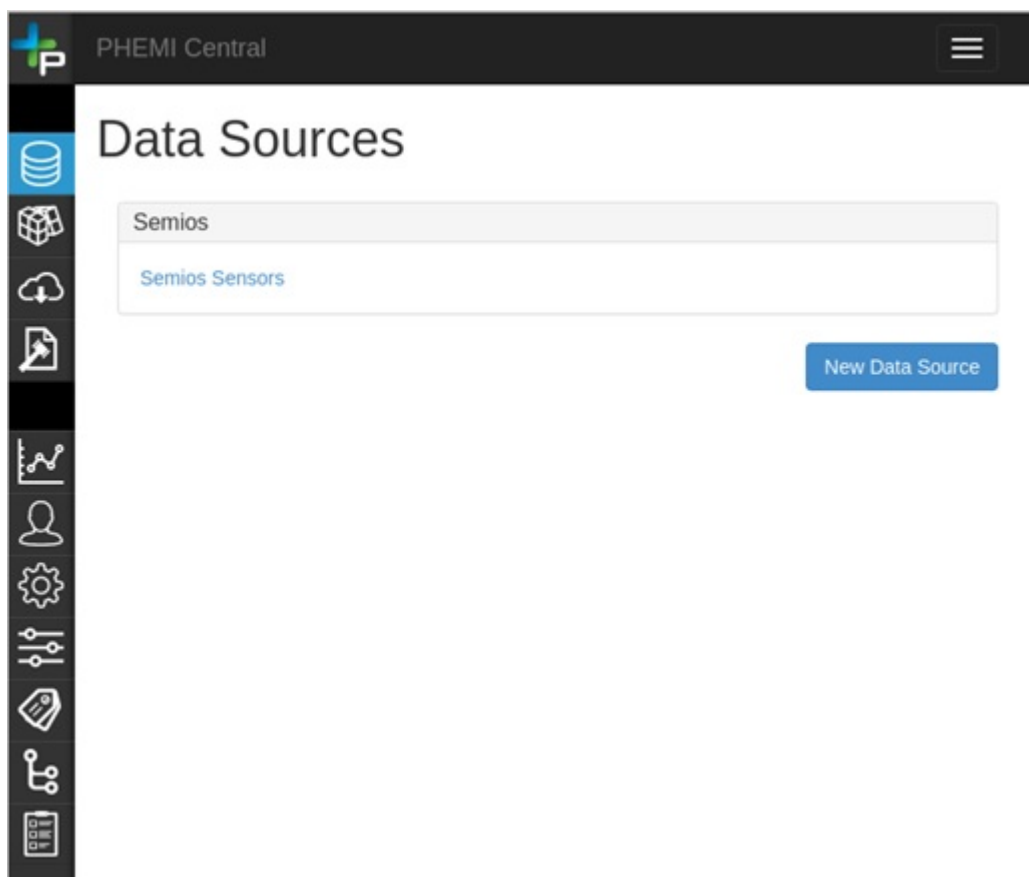
The aggregations are currently computed on a hourly, daily, monthly and yearly basis, per sensor. Several separate aggregations are computed for each type of sensor, depending on the specific sensor.

Wherever PHEMI Central returns an aggregated value (e.g., average hourly temperature), the number of data points used to calculate the aggregated value is also provided. The Semios web application can use this count to determine how many data points were not received by the Semios aggregator, fetching replacement data from an external web service if necessary.

The sum and the sum of squares for the readings are used to calculate the mean and the standard deviation.

Data Source User Interface


PHEMI provides a graphical user interface to configure and use a data source. The following screenshot shows a list of data sources:














When a data source is selected, a screen is displayed showing:

- **Data Policy** - Provide the information such as ownership, privacy, data sharing and other policy information
- **Ingest Data** - Provides a method of ingesting data by uploading a single or set of files
- **Data Processing Function** - The DPF to be used to process the ingested data to generate a set of derived data

The following is a screenshot for the Data Policy Details page:

 PHEMI Central Logout



Semios: Semios Sensors

Data Policy Ingest Data Data Processing Function

Source Description

Name	Semios Sensors
Source Category	Semios
Institutional Owner	Semios Administrator
Privacy Officer	Semios Privacy Officer
Source Owner	Semios Administrator
Document Format	Binary
Definition	Sample of Semios Sensors
Notes	Notes
Data Sharing Agreement	<div>Choose file No file chosen</div> <div>Start dd/mm/yyyy End dd/mm/yyyy</div>

Retention Rules

Do not delete

The following is a screenshot for ingesting data to a data source:

PHEMI Central Logout

Semios: Semios Sensors

[Data Policy](#) [Ingest Data](#) [Data Processing Function](#)

Manual

Files for Ingest Choose Files No file chosen

☐ Extract archives and ingest contents (.tar.gz, .tar.bz2, .tar, and .zip files only)

Ingest Files

Scheduled

Hourly

API-JSON

Configure

The following is a screenshot for uploading a DPF to the data source so it can process the Semios log information:

The following is a screenshot for uploading a DPF to the data source so it can process the Semios log information:

PHEMI Central Logout

Semios: Semios Sensors

[Data Policy](#) [Ingest Data](#) [Data Processing Function](#)

Name

Trigger

Status

Command

Description

Delete Replace Save Changes

Code Library [Show 1 Files](#)

Digital Asset Output

Name	Attributes	Definition	Data Type	Unit	Data Validation Rules
report_date	NON_IDENTIFIED	Date of report	DATE		Date in YYYY-MM-DD format
temperature	NON_IDENTIFIED	Temperature value	DOUBLE	C	Between -40.0 and 40.0