Submitting Data to PHEMI Central

Contents

Submitting Data to PHEMI Central..... 3
Using the RESTful API.....3
Using Manual Ingest.....3
Using Bulk Ingestion.....3
Using ETL Tools.....3

Submitting Data to PHEMI Central

Data can be submitted to PHEMI Central using the PHEMI RESTful API, by manually ingesting it, by using FTP or SSH batch ingestion, or by using extract, transform, and load (ETL) tools.

Using the RESTful API

If the data source is able to publish data, the system can be programmed to publish to PHEMI Central using the PHEMI RESTful API.

In REST-based ingestion, the client (that is, the data source or submitting system) sends an HTTP or HTTPS POST request. The POST request contains valid user credentials in JSON format in the payload body.

When the credentials are authenticated, PHEMI Central returns the session ID and URI for the session, as well as a session cookie. Once the session is established, the client can POST data to the appropriate data collection by referencing the data collection ID .

PHEMI Central listens for REST queries on port 80 (for HTTP) and port 443 (for HTTPS).

REST-based ingestion is useful in situations where a system submits smaller pieces of data very frequently. Since PHEMI Central always listens on the port, the client system can be set up with a scheduled task to submit the data as often as needed.

Using Manual Ingest

You can use the Management and Governance Console to manually ingest data objects into PHEMI Central.

Manual upload is a good method when you have very large amounts of data such that HTTP/REST is not suitable (for example, gigabytes or terabytes of data), and/or data that needs to be ingested relatively infrequently.

How do I manually ingest files?

Using Bulk Ingestion

Batch ingest of data is extremely fast. Configure a secure FTP or an SSH connection to allow a system to write data to a temporary landing space within PHEMI Central. PHEMI Professional Services will help you get this set up.

You can trigger the bulk ingest process remotely or you can use a scheduled task such as a cron job. Triggering the process launches a MapReduce job that inserts the bulk data into PHEMI Central at a very fast rate. The temporary files are then purged from the system.

Using ETL Tools

Some data collections (for example, some databases) are not able to submit data directly to PHEMI Central. For such systems, extract, transform, and load (ETL) tools can be used to extract data from the source system and then use either REST-based ingestion or bulk ingestion, depending on the requirements.