

OHDSI on Databricks: A Complete Guide to Implementing OHDSI on Databricks

John Gresh, Brad Rechkemmer

Background

Databricks is an increasingly popular tool for managing large datasets including instances of the OHDSI Common Data Model (CDM). Databricks provides a unified and collaborative environment and tools for data engineers, data scientists, and analysts. We have created an end-to-end automated solution for standing up an OHDSI stack on Databricks using either Broadsea or a standalone Tomcat instance of WebAPI and Atlas.

The OHDSI on Databricks implementation guide provides an end-to-end solution for connecting an existing Common Data Model (CDM) to OHDSI. This guide is based on an automated process implemented in the Ponos project. The Ponos tool can be used to perform all the steps required to OHDSI enable a CDM instance in Databricks. The Ponos tool also includes a solution to create an instance of the Broadsea DEMO_CDM in Databricks. This guide is intended to provide the following.

1. *An automated build*

The Ponos tool is provided to automate the process of getting an OHDSI instance set up in Databricks. This tool can be used to create an instance of the Broadsea DEMO_CDM in Databricks. This tool can be used to connect any instance of the CDM in Databricks to OHDSI including development, test, and production instances.

2. *A reference implementation*

The information provided here can be used as a reference implementation. There are other ways the work done by the Ponos tool can be implemented. The Ponos tool represents a known working example of how to create an OHDSI instance from a CDM in Databricks.

3. *Testing/Validation*

The Ponos tool creates a working OHDSI instance in Databricks and thereby provides a successful test and validation of the underlying tools used to do so.

4. *Insight into the process*

The code used by Ponos is available in GitHub. The code can be run from an IDE such as Eclipse and can be reviewed and stepped through to gain insight into the process and tools used here to create an instance of OHDSI using Databricks.

Methods

The implementation guides and other documentation created by the OHDSI Databricks Users Group will be used to guide participants through the simplified and automated process we have created to stand up an OHDSI stack using Databricks as the underlying database for the CDM and Achilles results tables.

Results

By the end of this demonstration, participants will be able to stand up an OHDSI stack in their own environment using Broadsea and Databricks using either their own CDM instance or using a demo instance of the CDM created as part of the demonstration.