OHDSI 2023 Abstract

Utilizing ARACHNE Data Node and
Execution Engine for Network Study execution

Gregory Klebanov, Alexey Manoylenko
Odysseus Data Services, Inc.

Background

The Observational Health Data Sciences and Informatics (OHDSI) collaborative is a global community that conducts ongoing observational network research to enable better healthcare. OMOP common data model (CDM) is a representation of biomedical data that standardizes entities, attributes, and relationships across multiple sources. The use of CDM-formatted data permits researchers across the OHDSI network to generate meaningfully comparable inferences from the same analyses. Over the past couple of years, OHDSI also made a significant investment and made great improvements in creating a more standardized framework and approach to the data analytics through HADES. Still, having a consistent, easy and reproducible execution of studies across the network remains one of the major challenges.

As a part of the large ARACHNE Platform, Odysseus engineers have developed two standalone components - the ARACHNE Data Node and Execution Engine – that can be deployed at the data partner site to address this challenge.

Methods

The ARACHNE Platform was designed and developed utilizing modular design and has three core components:

- ARACHNE Central is a centralized collaborative environment where study can be managed as a project, including security and sharing the data analytics code across the network. The ARACHNE Central can be configured to connect to one or more Data Nodes installed at data partners thus forming a data network.
- ARACHNE Data Node is installed at the data partner site and can be used – among other things - to execute study data analytics code against the local data set(s) through ARACHNE Execution Engine.
- ARACHNE Execution Engine is a pre-configured machine that comes with all standard OHDSI R libraries pre-installed and created a clean execution environment for each R execution package request.

Each of the components is dockerized and can be installed and utilized separately. The purpose of this demo is to demonstrate how ARACHNE Data Node and Execution Engine components can be effectively used by the OHDSI / EHDEN Data Partners to execute OHDSI studies against a local database(s), including with Strategus JSON or legal R codebase.
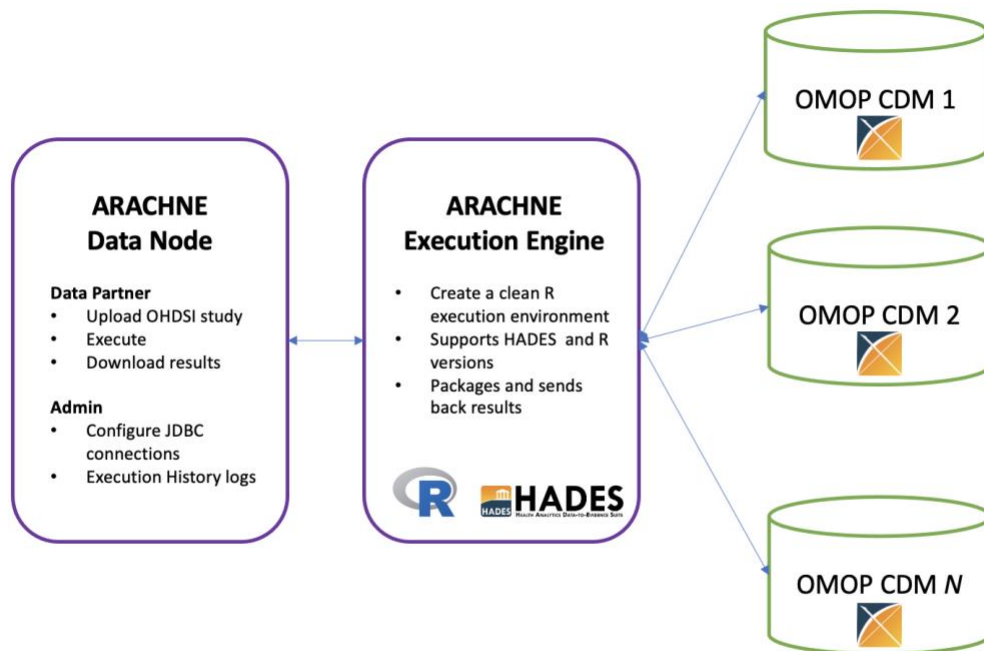
The following are the key ARACHNE Data Node and Execution Engine features:

- Support for OHDSI Strategus framework
- Support for OHDSI HADES, including automatic or manual versioning of the execution environment to support different HADES releases for HADES-based studies.
- Ability to upload Strategy JSON or a legal R code and submit execution request against a local database(s)
- Monitor job executions.
- Receive a full result package, including:
    o Files submitted.
    o Results generated.
    o Debugging and execution logs.
- Preview results as raw files
- Download results and share with OHDSI study team as a zip file.

In practically all mature Pharma and Healthcare organizations, IT security teams disable an ability to automatically download external executable packages from any code, including R code and HADES libraries, which present an additional challenge. To resolve this challenge, the ARACHNE Execution Engine comes with all required libraries pre-installed and was designed to determine and spin off a clean execution environment with correct versions as required by each OHDSI study codebase.

Results

By enabling a dockerized study execution tool that provides full support for OHDSI Strategus as well as R packages (legacy) that comes pre-packaged with versioned R and HADES packages, we are enabling a

simple, effective and scalable approach to execution of OHDSI Network Studies across the network. This approach does not require data partners to have any knowledge of R programming or even having a need to perform a complex and error prone installation of an internal R environment with correct versions of packages installed.

Conclusions

The ARACHNE Data Node and Execution Engine Components provide an easy, effective and scalable way to execution network studies.