



Abstract

The OHDSI cloud provides shared computing and collaboration services for the global OHDSI community. It has been developed iteratively over a one year time frame based on expanding community needs. It encompasses several capabilities ranging from community collaboration tools through to continuous integration tools for OHDSI application software development. It will continue to be enhanced as the community expands and new OHDSI applications are introduced.

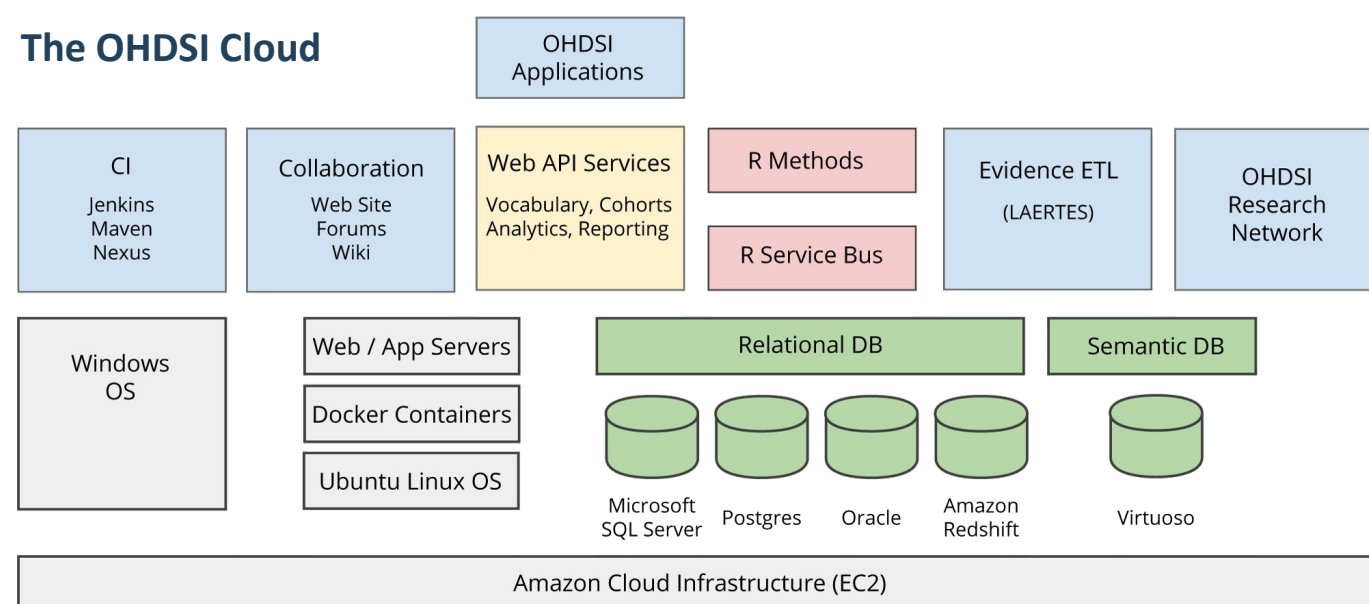
Capabilities

The OHDSI cloud supports the following capabilities:

- OHDSI Community collaboration tools - website, wiki, forums
- Continuous integration server to build & deploy OHDSI applications
- Sandbox/demo environment running the OHDSI WebAPI & Applications
- OHDSI LAERTES knowledge base ETL operational server
- OHDSI research network study aggregate summary results repository

Infrastructure

The OHDSI Cloud



OHDSI LAERTES knowledge base ETL Server

The LAERTES evidence knowledge base is refreshed quarterly. The LAERTES ETL server processes multiple data feeds including SemMedDB, PubMed, UMLS meta-thesaurus, Standard Product Labels and FAERS adverse events. The data sources are downloaded as files from websites onto Amazon EC2 servers where the data is loaded into MySQL or PostgreSQL databases, transformed via Python programs and then loaded into the LAERTES PostgreSQL and Virtuoso semantic databases. Virtuoso provides a semantic SPARQL interface to LAERTES knowledge and PostgreSQL provides counts and statistics.

OHDSI research network aggregate results

A PostgreSQL database hosted in Amazon RDS is configured with SSL certificates to identify the database server to research database clients and ensure that all traffic between the database clients and the server is securely encrypted. Only de-identified aggregated summary results of OHDSI research network studies is stored in the database. PostgreSQL password protected read-only database accounts are provided to OHDSI community researchers who participate in the studies, so they can review and analyze the study results.

WebAPI

The WebAPI provides a set of web services that are used by the OHDSI applications and also provides a general purpose Application Programming Interface for access to OHDSI standard information.

The WebAPI services provides access to the following information:

- Standardized Vocabularies
- Common Data Model
- LAERTES Evidence Knowledge base
- Cohort definition / analysis
- Study feasibility
- Therapy path reporting

OHDSI applications make SQL calls that are dynamically translated to the required DBMS SQL dialect through a tool called SQLRender which is available as a WebAPI service.

Continuous Integration

A Jenkins continuous integration server automatically builds and deploys the OHDSI applications to Tomcat and Apache servers running within Docker containers on the OHDSI sandbox environment. Docker volumes are shared with the Amazon EC2 host servers for persistent storage. Multiple DBMSs are supported, including Oracle, PostgreSQL, SQL Server and Amazon Redshift. Multiple operating systems are supported through the use of Java and the Spring java framework. Software artifacts are built and managed using Maven and Nexus.

Future Road-map

- Enhance the OHDSI sandbox environment with an interactive SQL translation tool to support the design & testing of new OHDSI research analyses that will run on a variety of DBMSs.
- Expand the capability to upload and manage the storage of aggregate summary results shared by research sites to the Columbia central coordinating center.
- Integrate the R Methods library analysis tools with the OHDSI web applications using the R Service Bus in the OHDSI cloud.
- Introduce flexible access and authentication through the deployment of the Shiro Java security framework.

Conflicts of Interest

LTS Computing LLC is a commercial IT projects & services company focused on Life Sciences.