Name:	Yuriy Khoma
Affiliation:	ODYSSEUS DATA SERVICES INC
Email:	yuri.khoma@odysseusinc.com
Presentation type :	Collaborator Demonstration

Arachne: Automated Execution of distributed OHDSI Network Research

Nikolai Puntikov[1)], PhD, Yuriy Khoma[1,2)], PhD
[1)]ODYSSEUS DATA SERVICES INC, Cambridge, MA
[2)]Lviv Polytechnic National University, Lviv, Ukraine

Abstract

We will be presenting and demonstrating a life prototype of ARACHNE, a system for distributed observational research. It will make the participation of Data Nodes in Studies conducted through the OHDSI Research Network very efficient and independent of local technical expertise, while at the same time providing a strict system of governance and data protection.

Introduction

OHDSI's Vision is to create an open and public network of scientific collaborators for the generation of evidence about healthcare. These collaborators may contribute data, clinical expertise or technical methodology. To realize this vision, OHDSI needs to bring together Data Providers, Clinical Investigators, and Data Scientists (for the necessary statistics, software, etc.), without the burden of each participating OHDSI member having to have all those resources available locally. Instead, OHDSI as a research collaborative would share and utilize the collective expertise, while still be based on a large network of data, which will enable robust and fast research.

Demonstration of Use Case

ARACHNE automates the process of evidence generation through distributed Research Studies. The ARACHE demonstration will illustrate the entire use case:

- Contriving a new study by the Chief Investigator;
- Placing the execution code in R which can access data in OMOP CDM and calculates the necessary statistics;
- Distributing the study to Data Nodes of the OHDSI Research Network;
- The acceptance to participate in the Study by the Data Node;
- The automated execution of the R code locally at the Data Node;
- The return of the aggregate study results to the Chief Investigator, who then can analyze them, draw conclusions and prepare the publication, as well as error messages that may be generated in the process;
- The governance of this process.

ARACHNE is implemented in two separate software applications:

ARACHNE Central is responsible for registering and initializing new research projects, followed by creation of a package of all necessary R scripts required to complete the project. The Installation of R distribution requires considerable effort in setup and database connection at the Data Nodes. Using the Docker technology, we are reducing this burden. Docker containers wrap the software with everything necessary for execution: code, runtime, a complete filesystem, system tools, system libraries. This guarantees that it will always run the same, regardless of the environment it is running in.

ARACHNE Daemon is installed at the Data Nodes and communicates with ARACHNE Central to perform the required actions. It runs the R scripts and executes the embedded SQL against the CDM database. The resulting tables are sent back to the ARACHNE Central, but each of these steps is subject to approval by the Data Node administration.

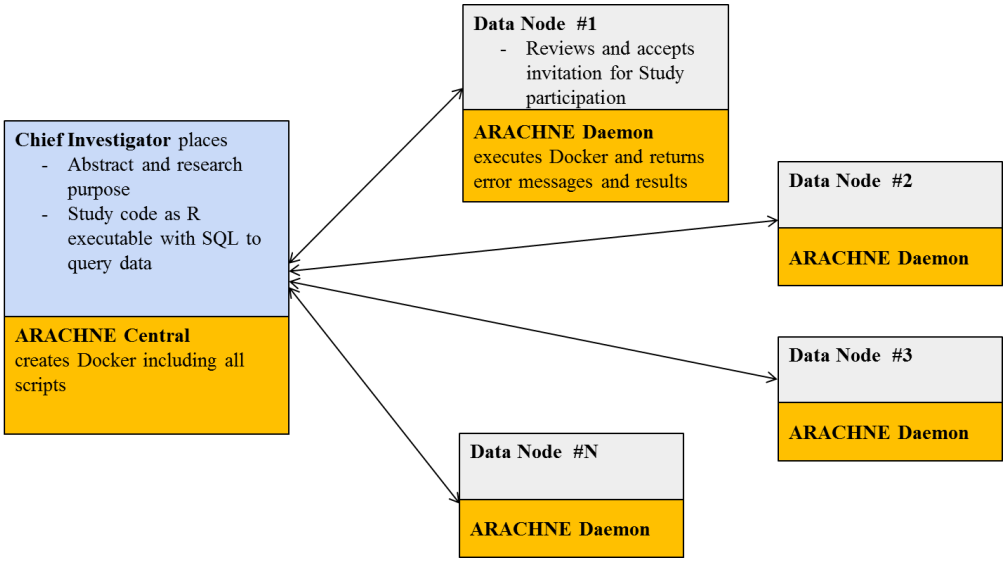

Figure 1. Arachne: System Architecture.

Technology

The system back-end is written on Java8 using the Spring Framework (Boot, Security) and Hibernate. Spring Boot is used to save all setting and dependencies within executable file, which allows deploying and running the application automatically. Spring Security is responsible for authorization and access management to system components. Hibernate provides simplified data access based on object relational model. The front-end is based on angular JS and the Bootstrap framework. Currently, the user interface includes the following features: login screen, main page with research list and research constructor. As mentioned above, the system uses Docker for R scripts containerization. Apache Cassandra, as BigData storage, collects and aggregates research results, while PostgreSQL saves user data.

Future Developments

We are planning on implementing the following additional features, which should enhance the utility of ARACHE substantially:

1. Implementation of the blockchain technology for enhanced data protection and governance, including logging and post-hoc auditing.
2. Sandbox for development and testing of Study scripts, including test data, before these scripts are being placed into distribution.
3. Module for announcing and collecting feedback on potential research studies amongst the network.

Conclusion

We believe that implementation of Arachne will become a key factor in enabling efficient observational research. OHDSI already achieved great results by providing:

- Standardization of data formats and coding (vocabularies);
- Standardization of methods;
- Standardization of cohort definitions; and
- Collaborative network of renowned scientists in the areas of clinical research, statistics, method development and scaling.

Arachne aims at removing a remaining obstacle of technical execution of studies across a network of distributed databases: each contributor to a study shall be able to focus on the scientific problem, rather than dealing with SQL queries and compatibility challenges (SQL dialects, technical infrastructure, security, etc.).

References

1. OHDSI: The research community. Accessed June 20, 2016. Available from: http://www.ohdsi.org/
2. PopMedNet.org. Accessed June 20, 2016. Available from: http://www.popmednet.org/.
3. Nikolai Puntikov, Gleb Malikov, “Vision for Technical Infrastructure to Facilitate OHDSI Network Research”: *2015 OHDSI Symposium*. Accessed June 20, 2016. Available from: http://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=resources:arachneposterabstract.pdf