|  |  |
| --- | --- |
| Name: | Michael Ames, MBI(Contact responsible for submission; not an author.) |
| Affiliation: | University of Colorado Anschutz Medical Campus |
| Email: | Michael.ames@ucdenver.edu |
| Presentation type (s):  | **Lightning Talk and 10-minute video** |

**Google BigQuery and the OHDSI Ecosystem**

**Melanie Philofsky, RN, BSN, MS1, Matthew Young-Lai, PhD2, Paul Church MS2, Ilia Tulchinsky2, Ming Jack Po, MD, PhD2, Michael G. Kahn, MD, PhD1**

**1University of Colorado Anschutz Medical Campus, Aurora, CO, USA**

**2Google Cloud, Mountain View, CA, USA**

**Abstract**

*Health Data Compass has adopted Google BigQuery as the analytical database engine for its multi-institutional enterprise health data warehouse. BigQuery is a massively scalable, easy-to-use, cloud-based data warehouse solution. While BigQuery offers ODBC connectivity and SQL SELECT compatibility, it has not previously been compatible with the various tools of the OHDSI ecosystem, e.g., ATLAS and ACHILLES. This year, Google has adapted many components of these tools to work with BigQuery, enabling Health Data Compass to take advantage of the scalability and performance of BigQuery with the richness of the OHDSI ecosystem. The talk will review Health Data Compass’s architecture in the cloud, their experience using OHDSI tools with BigQuery, and recommendations for other OHDSI institutions wishing to use OHDSI tools with Google’s cloud-based data warehouse engine. The video will demonstrate how to deploy an OMOP model to BigQuery, how to connect to it with select OHDSI tools, and how to perform common profiling, mapping, and analytical functions.*

**Introduction**

Health Data Compass (Compass) is an enterprise health data warehouse headquartered on the University of Colorado Anschutz Medical Campus in Aurora, CO. Compass integrates data from two hospital systems, a physician billing plan, several University of Colorado Denver research labs, and other sources. This data is made available to researchers, clinicians, and administrators from across the University and our partner institutions for a variety of purposes, including clinical, translational, and personalized medicine research, and operational excellence across the clinical and research enterprise.

In 2016, Health Data Compass embarked upon a major transformation, transitioning its data warehouse infrastructure from a traditional on-premises technology stack to an entirely new suite of technologies provided by the Google Cloud Platform. This transition was completed in January 2017 with Google BigQuery as the core analytical query engine – the “beating heart” of Compass’s new data warehouse on the cloud.



**Figure 1.** High-level view of Health Data Compass’s data warehouse architecture on Google Cloud Platform

BigQuery provides significant advantages over traditional data warehouse systems in terms of speed, scalability, ease-of-use, and cost. It is SQL SELECT compatible and can be queried from any ODBC/JDBC client. However, many of the commonly-used tools within the OHDSI ecosystem, such as ATLAS and ACHILLES, were not compatible with BigQuery out of the box.

Google Cloud is working closely with the healthcare industry to provide technology and tools that help create better biomedical research and clinical care. To that end, Google Cloud is adapting the OHDSI tools to work seamlessly with Google BigQuery, committing several open-source updates during the spring and summer of 2017. These updates include adding missing language support to JDBC drivers used by BigQuery, allowing [RStudio](https://www.rstudio.com/) running in a [BroadSea](https://github.com/OHDSI/Broadsea) docker container the ability to connect BigQuery with [DatabaseConnector](https://github.com/OHDSI/DatabaseConnector), and letting BigQuery function with all of the [Achilles](https://github.com/OHDSI/Achilles) queries including [Achilles\_v5.sql](https://github.com/OHDSI/Achilles/blob/master/inst/sql/sql_server/Achilles_v5.sql) , [AchillesHeel\_v5.sql](https://github.com/OHDSI/Achilles/blob/master/inst/sql/sql_server/AchillesHeel_v5.sql) , and [export\_v5](https://github.com/OHDSI/Achilles/tree/master/inst/sql/sql_server/export_v5) executed from [R scripts](https://github.com/OHDSI/Achilles/tree/master/R). In addition, all the changes to support the above are now contributed back to the OHDSI codebase.

Health Data Compass is now using these tools within its data warehouse on Google BigQuery. In addition to the technical advantages provided by BigQuery, Compass is also combing the OHDSI CDM with BigQuery’s pre-populated and freely available [public datasets](https://cloud.google.com/bigquery/public-data/) such as weather, air quality, census, RxNorm, CMS utilization summaries and other high value health related datasets.

**Observations and Conclusions**

Specific observations and conclusions drawn from Compass’s testing and use of OHDSI tools on BigQuery are ongoing, and will be presented in the talk.