Moving OMOP to the Cloud
With DBT and Snowflake

Roger Carlson
APRIL 29, 2024
Moving OMOP to the Cloud With DBT and Snowflake

Background
- Spectrum Health has been evaluating the use of modern cloud computing for its analytical processing needs.
- Cloud-computing Advantages
  - Improved speed, scalability, security, data-sharing capabilities.
  - Unique opportunity to invest in modern, open-source tools and methodologies through the adoption of a unified tool set.
- Envisioned Platform (Proof-of-concept)
  - Amazon Web Services (AWS)
  - All of Us® Research Program (NIH), which transforms EPIC® Clarity EHR data into the OMOP format.
  - Relatively small-scale project (~12,000 patients)
  - Complex and robust ETL process.

Methods
- Legacy System
  - Microsoft SQL Server database (on-prem)
- Modern system
  - Snowflake database on AWS platform
  - Reduced toolset: Snowflake, DBT, GitHub, REDCap, VSCode, DBeaver, and R Studio.

Timeline

<table>
<thead>
<tr>
<th></th>
<th>MSQL</th>
<th>SNOWFLAKE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIS</td>
<td>1h:50m</td>
<td>20m</td>
</tr>
<tr>
<td>DBT</td>
<td>~50 m</td>
<td>5m</td>
</tr>
</tbody>
</table>

Run Times

Development
- Phase 1: Move database to cloud-based database (MSSQL → Snowflake)
  - Timeframe: Apr 2021 – Dec 2021
  - Scope: 60 tables, 35 views, 262 queries
  - OMOP v5.2 to v5.3.1 upgrade
- Phase 2: Convert SSIS project to work with Snowflake
  - Timeframe: Sept 2021 – Dec 2021
  - Scope: 19 packages, 171 tasks, 262 queries
- Phase 3: Move workflow process to open-source tool (SSIS → DBT)
  - Timeframe: Jan 2022 – May 2022
  - Scope: 578 steps, 416 models (347 views, 29 tables), 202 tests, 433 macros, 95 sources
  - Full Production Run:
    - July 7, 2022 (21 person months total)
- Phase 4: Full Integration and automation, i.e., delivery of OMOP files from AWS to Google Bucket.
  - Timeframe: Jun 2022 – Nov 2022 (completed)

Workflow Comparison

<table>
<thead>
<tr>
<th>Workflow Process</th>
<th>SSIS</th>
<th>DBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>API download from REDCap®</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Extract data from EPIC® Clarity</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Transform Clarity data into OMOP</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Built-in and custom testing features</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Referential Integrity</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Curation Reporting</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Automated data export</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Automated SMTP transfer</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>Transfer from S3 bucket to Google®</td>
<td>☑</td>
<td>☑</td>
</tr>
</tbody>
</table>

Conclusion
- Conversion from SQL Server using SSIS to Snowflake using DBT was timely and effective. Our result is a more robust platform featuring a collaborative workflow built on modern toolsets.
- Snowflake/DBT significantly outperforms MSSQL/SSIS.
- Snowflake is effectively unlimited in terms of scalability and complies with a wide range of compliance standards including HIPAA/HITRUST, SOC 1 Type II and SOC 2 Type II.
Questions?

Roger.Carlson@corewellhealth.org

1. The Snowflake Platform
2. Cloud computing with AWS
3. What is dbt?
4. Snowflake Security