Corewell Health™

**OHDSI DevCon 2024
Open Source Developers Conference**

Moving OMOP to the Cloud With DBT and Snowflake

Roger Carlson

APRIL 29, 2024

Moving OMOP to the Cloud With DBT and Snowflake

Corewell Health™

Roger Carlson - Corewell Health
Matthew Phad - Corewell Health
Sam Martin - Corewell Health
Southern Lower Michigan

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)
  - Tools used: SQL Server Management Studio, SQL Server Integration Services, Visual Studio, Redcap, Oracle SQL Developer, R-Studio, Tortoise SVN, and Microsoft Access.
- Modern system
  - Snowflake database on AWS platform
  - Reduced toolset: Snowflake, DBT, GitHub, REDCap, VSCode, DBeaver, and R Studio.

Timeline
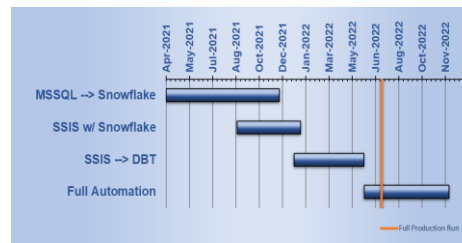
Run Times

	MSSQL	SNOWFLAKE
SSIS	1h:50m	20m
DBT	~50 m	5m

Workflow Comparison

Workflow Process	SSIS	DBT
API download from REDCap®	☒	☑
Extract data from EPIC® Clarity	☑	☑
Transform Clarity data into OMOP	☑	☑
Built-in and custom testing features	☒	☑
Referential Integrity	☑	☑
Curation Reporting	☒	☑
Automated data export	☑	☑
Automated SMTP transfer	☒	☒
Transfer from S3 bucket to Google*	☒	☑

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)

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