

Design and Implementation of an Automated OMOP Database Migration Tool from PostgreSQL to SQL Server

Background

The Observational Medical Outcomes Partnership (**OMOP**) common data model (CDM) standardises observational health data across institutions. However, its use differs across Relational Database Management System (**RDBMS**) platforms such as PostgreSQL and Microsoft SQL Server, making database migration complex and inconsistent. A reliable, reproducible framework is therefore needed to accurately migrate schemas and data while preserving OMOP CDM integrity.

Objectives

To design and implement a modular, transparent, and maintainable database migration tool for transferring OMOP CDM databases across different RDBMS environments, using PostgreSQL to Microsoft SQL Server as a case study.

Methods

The proposed modular framework for transferring OMOP CDM databases across RDBMS platforms as shown in **Figure 1**.

1. **Extraction:** Batch scripts export tables to CSV files.
2. **Loading:** T-SQL bulk insert exported data into staging tables.
3. **Schema Creation:** OMOP CDM DDL scripts from Observational Health Data Sciences and Informatics (OHDSI) create tables and objects.
4. **Constraints:** Keys and indexes are applied post-load
5. **Automation:** Execution is fully automated in Batch.
6. **Configuration:** Parameters are ready for multi-environment use.
7. **Logging:** Execution status and errors are recorded.

Results

Implementation of the proposed migration approach demonstrated the successful recreation of the OMOP CDM databases from PostgreSQL to Microsoft SQL Server through extraction, transfer and reconstruction. Schema objects, primary keys, indexes and foreign keys were recreated on the target system with no structural inconsistencies observed during the database reconstruction. Execution logs confirmed the completion of each migration stage without errors, demonstrating suitability for repeatable and maintainable OMOP CDM migrations across heterogeneous RDBMS environments.

Conclusion

The proposed migration framework addresses a key interoperability challenge faced by institutions deploying OMOP CDM databases across RDBMS. The staged bulk-loading strategy effectively eliminates possible issues related to constraint enforcement and data type incompatibilities during migration by separating data ingestion from schema enforcement. This approach supports systematic recreation of schemas, tables, primary keys, indexes and foreign keys in the target SQL Server environment.

This design choice improves transparency by making structural transformations explicit and reviewable, while also supporting version control and long-term maintainability. Limitations of the current framework include the small number of RDBMS used. Future enhancements will focus on extending the framework across other RDBMS, optimising performance, and automating the execution and integration of RDBMS-specific DDL artefacts provided by OHDSI.

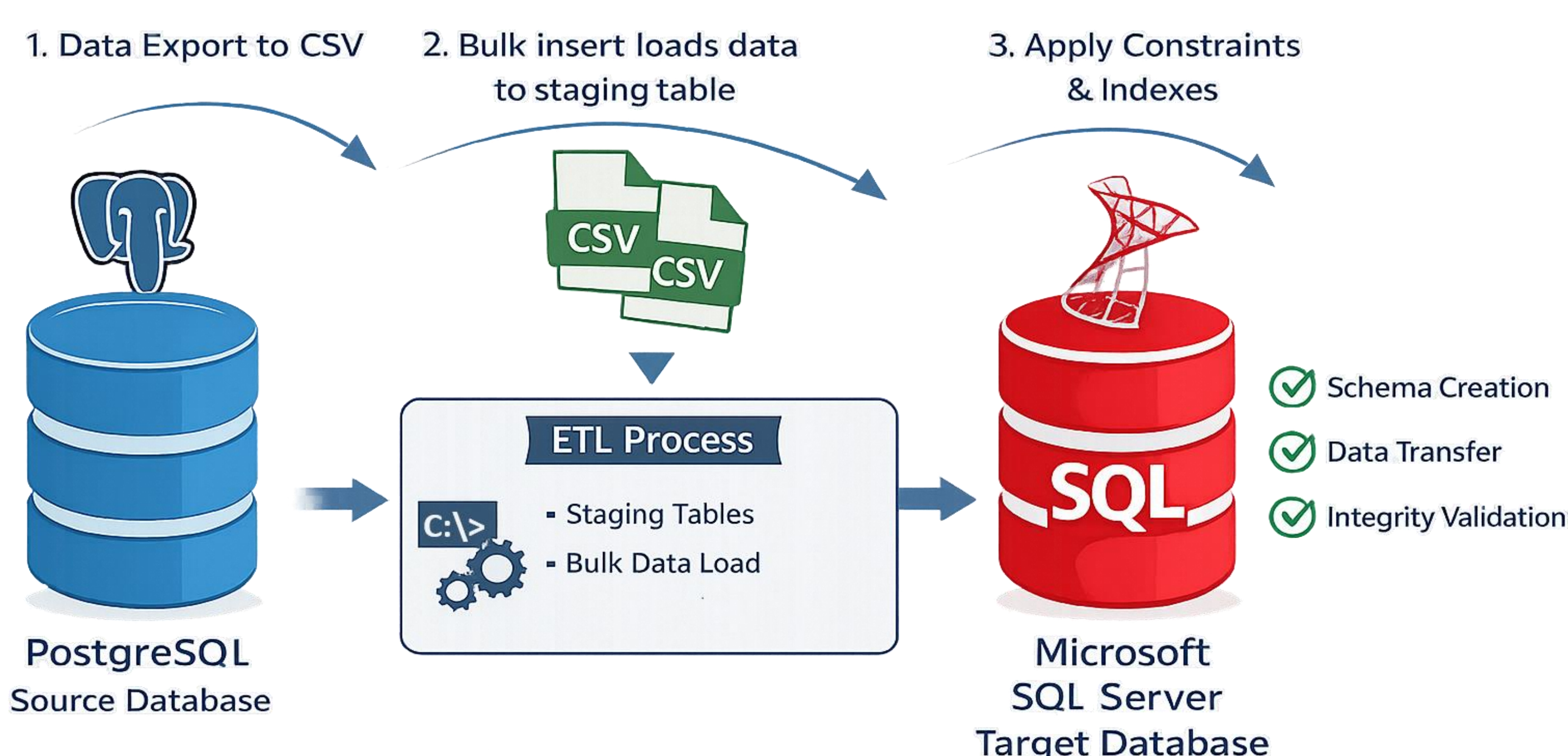


Figure 1. OMOP CDM migration workflow from PostgreSQL to Microsoft SQL Server



GitHub



Mandickel Kamtengeni¹, Antonella Delmestri¹
¹Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
University of Oxford, Oxford, United Kingdom
Email: mandickel.kamtengeni@ndorms.ox.ac.uk



SUPPORTED BY



National Institute for Health and Care Research