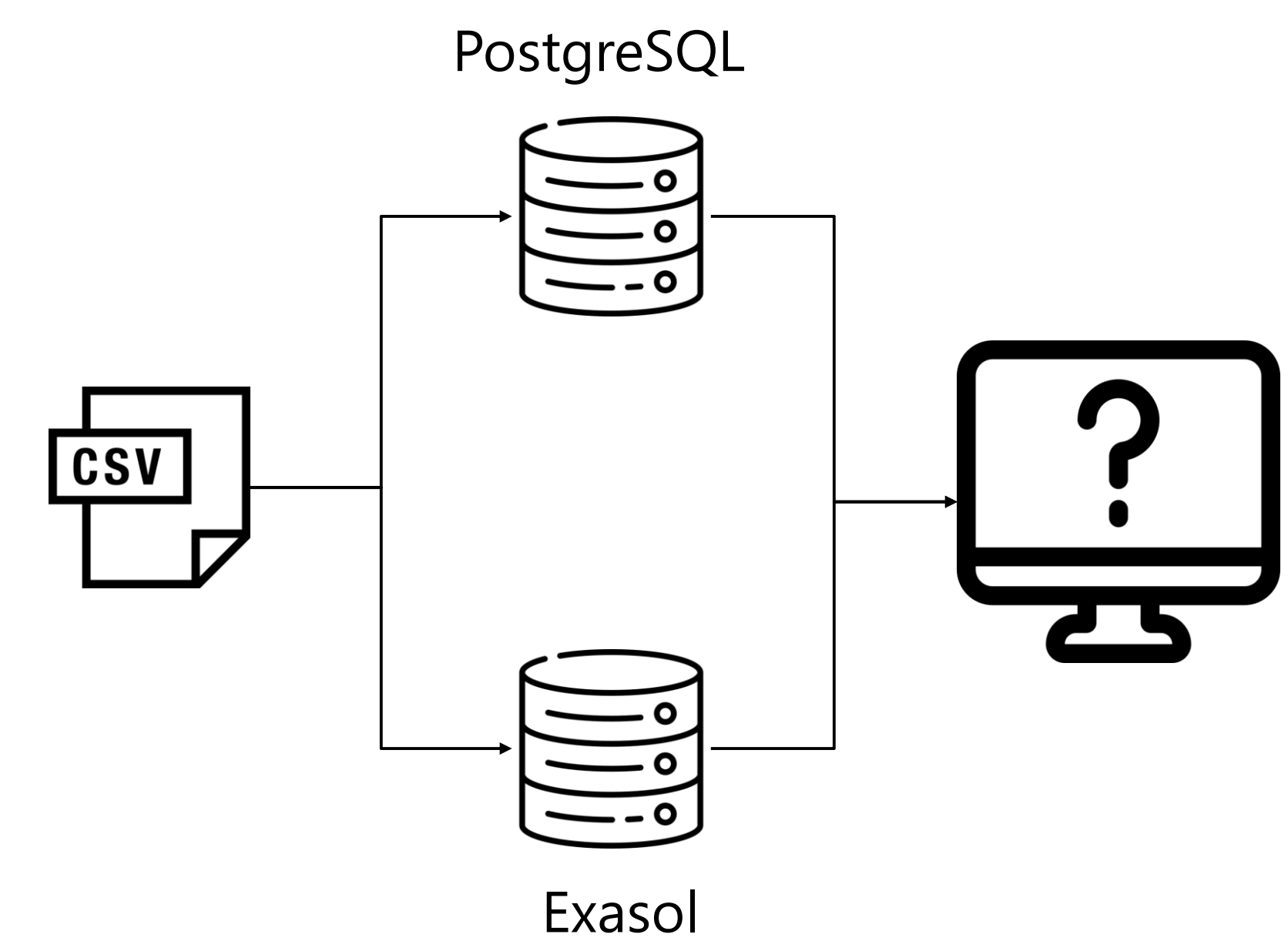


# Vergleich der Performance von Exasol und PostgreSQL in der Verwendung von OMOP CDM

 Yuan Peng<sup>1</sup>, Ines Reinecke<sup>1</sup>, Martin Sedlmayr<sup>1</sup>
<sup>1</sup>Institut für Medizinische Informatik und Biometrie, Medizinische Fakultät Carl Gustav Carus der Technischen Universität Dresden

## Hintergrund

- Die Analyse von großen Datenmengen aus der Routineversorgung für die klinische Forschung gewinnt zunehmend an Bedeutung.
- Die Schaffung von Datenintegrationszentren durch die Medizininformatik-Initiative (MII) schafft die Voraussetzungen für die Verwendung der Daten über unterschiedliche Standorte hinweg. [1]
- OHDSI OMOP ist eine Forschungsdatenrepository, welches Common Data Model (CDM) nutzt. CDM ist eine standardisierte Form der Darstellung und Speicherung von klinischen Daten. OMOP enthält einheitlichen Terminologien für die standardisierte Repräsentation von Daten sowie die Möglichkeit für deren standardisierte Auswertung.
- OMOP CDM ist in MIRACUM momentan auf PostgreSQL aufgesetzt, welche eine zeilenorientierte Datenbank ist, d.h. PostgreSQL hat niedrige Leistung, um eine große Datenmenge zu bearbeiten oder analysieren. [2]
- Exasol ist eine spaltenorientierte in-memory Datenbank, die sich für analytische Anfragen als Alternative anbietet. [3]



## Methode

Die unterschiedliche Performance von PostgreSQL und Exasol wurden durch die folgende vier Kriterien bewertet

- **OMOP Vokabulare importieren**
  - Alle Vokabulare sind von OHDSI Athena heruntergeladen [4]
- **Import von Daten**
  - 521 Daten von 2004 bis 2019 wurden als Importdaten genutzt
- **Abfrage von Daten(Lesen)**
  - Einfache Suchabfrage: sucht nach einem Patienten
  - Komplexere Suchabfrage: sucht nach mehreren Patienten (Kohorte)
- **Aktualisierung von Daten(Schreiben)**
  - Eine Spalte für einen Patienten aktualisieren
  - Eine Spalte für 231 Patienten aktualisieren

## Ergebnisse

	Exasol	PostgreSQL
Datengröße (Patientenanzahl)	471.164	471.164
Datenimport	1h29m	1h10m
Einfache Suchabfrage	0,08s	0,48s
Komplex Suchabfrage	10,6s	52,76s
Ein Patient aktualisieren	0,058s	0,008s
231 Patienten aktualisieren	0,2703s	0,4567s

## Diskussion & Ausblick

- Die Exasol Datenbank hat einen erheblichen Performance Vorteil bei der Verarbeitung oder Veränderung von mehreren Daten die bereits in der Datenbank gespeichert sind (Schreiben).
- Da bisher die auf OMOP aufsetzende WebAPI nicht mit Exasol kompatibel ist, muss in einem nächsten Schritt eine Anpassung der WebAPI stattfinden. Dazu kann die in der WebAPI verwendete Flyway Komponente angepasst werden oder alternativ eine eigene Anpassung des Quellcodes der WebAPI ohne die Verwendung von Flyway erfolgen.
- Im nächsten Schritt wird Exasol erneut mit denselben Methoden getestet. Jedoch mit viel mehr Daten, um festzustellen, ob Exasol bei der Verarbeitung von Millionen Daten immer noch so effizient sein kann.

[1] Prokosch H et al. MIRACUM: Medical Informatics in Research and Care in University Medicine: A Large Data Sharing Network to Enhance Translational Research and Medical Care. Methods of Information in Medicine. 2018; S01:e82-e91

[2] Vandana Bhagat and Arpita Gopal; Comparative Study of Row and Column Oriented Database. 2012 Fifth International Conference on Emerging Trends in engineering and Technology, Himeji, pp. 196-201

[3] Exasol: <https://www.exasol.com/de/>

[4] OHDSI Athena: <http://athena.ohdsi.org/search-terms/terms>

Bild Quelle: <https://www.flaticon.com/de/autoren/freepik>