

Towards Automated Schema and Data Migration Between Different Data Stores



André Conrad, Uta Störl
University of Hagen

Introduction

Do I use the optimal data store or should I **migrate** to another?



Why to migrate to another data store:

- > Emergence of new technologies!
- > Changed or new requirements



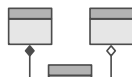
Why is this service so slow?

Migration without schema optimizations can lead to poor performance!

Challenges



Heterogeneity between NoSQL data stores

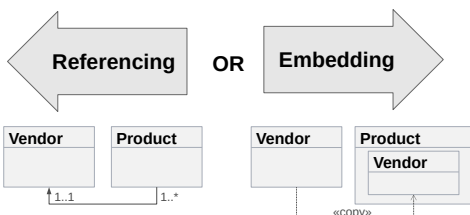


No standard processes for designing optimal NoSQL schemas

Usually **no joins** in NoSQL data stores



In simple terms, the question for **schema optimization** of aggregate-oriented models is:



A compromise between read performance (possible redundancy) and write performance!

A big challenge is the **optimization of read AND write operations!**

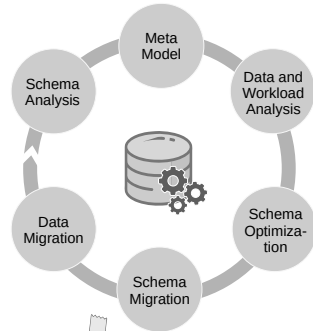
Project Aim

- An **automatic** process of **schema and data migration** between different data stores including **schema optimizations**
- **Extensibility** in terms of supported data stores from and to which migration is possible



Methodology

Core Components:



1. NoSQL stores are often schema-flexible, which can make **schema analysis** a very complex task



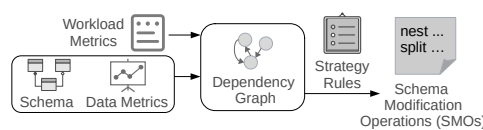
2. A platform Independent **meta model** enables the flexible extension of the supported data stores



3. A prerequisite for the optimization of the schema is an analysis of the **data and workload metrics**



4. An important aspect and a great challenge is the automated **optimization process of the schema**

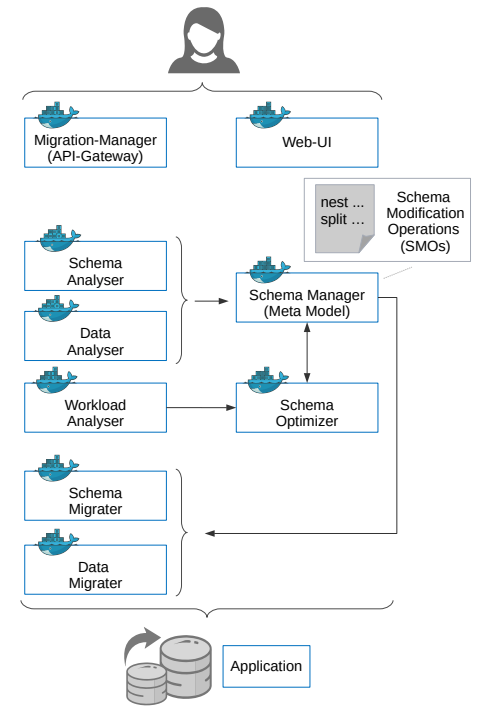


- **Dependency Graph:** Describes performance critical "relations" with respect to the query workload
- **Strategy Rules:** The structure of the optimized schema is generated based on target store specific rules

Architecture

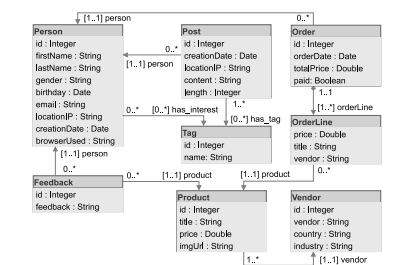
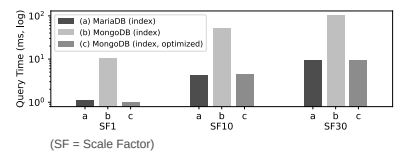
Implemented as a **containerized application**:

- Manage the **complexity** of larger projects
- **Flexibility** in replacing different components
- Horizontally **scalable**



Measurements

First measurements to prove an increase in performance with a migration from MariaDB to MongoDB (single node, single query)



Next Step: Measurements with a multi-node database cluster and multiple queries (read and write)

References

A. Conrad, S. Gärtner, U. Störl. *Towards Automated Schema Optimization*, in ER 2021.

