Declarative Querying of Heterogeneous NoSQL Stores

Nikolaos Koutroumanis
Dept. of Digital Systems
University of Piraeus
Piraeus, Greece
koutroumanis@unipi.gr

Nikolaos Kousathanas
Dept. of Digital Systems
University of Piraeus
Piraeus, Greece
nikolaos.kousathanas@gmail.com

Christos Doulkeridis
Dept. of Digital Systems
University of Piraeus
Piraeus, Greece
cdoulk@unipi.gr

Akrivi Vlachou
Dept.of Inf. & Com.Syst.Engineering
University of Aegean
Karlovasi, Greece
avlachou@aegean.gr

ABSTRACT

Nowadays, large quantities of data reside in different and heterogeneous NoSQL stores that accommodate the individual requirements of each application, such as scalability, efficiency and flexibility to schema changes. In contrast to the well-established relational model, NoSQL stores are still non-standardized and use heterogeneous languages and APIs for data access. In consequence, big data developers and data analysts need to write customized code for data access, exploration and analysis over different NoSQL stores. We present a solution to this problem that allows seamless access to different NoSQL stores using a common programming API. Moreover, we show that we can exploit this API in order to provide declarative access to NoSQL stores using a SQL-like language.

Reference Format:
Nikolaos Koutroumanis, Nikolaos Kousathanas, Christos Doulkeridis, and Akrivi Vlachou. Declarative Querying of Heterogeneous NoSQL Stores. In the 2nd Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores (SEA Data 2021).

1 MOTIVATION & RESEARCH CHALLENGES

Despite their popularity in the development of scalable, big data applications, NoSQL stores [1] still rely on heterogeneous data models, languages and APIs. Even though this is considered as a positive feature for modern, data-intensive applications (as we know nowadays that "one size does *not* fit all" when it comes to DBMS [6]), it also poses important problems. In particular, developers need to learn different query languages to access different NoSQL stores, a fact that also hinders portability of applications when a different storage system is chosen.

Existing solutions to this problem include polystores [2], database engines that use different systems (including NoSQL) for storage of different data types. However, polystores comprise yet another query engine (with components for query execution, optimization, etc.) that needs to interact with existing storage systems that include their own query engines. Another relevant approach is Facebook's Presto [5] (recently known as Trino), which is an SQL-compliant query engine that operates on a wide variety of different data sources. Again, the general idea is to put a new query engine in order to unify query processing on top of existing systems that already provide native support for query processing. Although this is meaningful for certain applications, it is not necessarily appealing for developers that need to use popular NoSQL stores in their big data architectures and query them using the same language.

Instead, we envision a unified approach for declarative querying of heterogeneous NoSQL stores using the same query language. Yet, our objective is to support this without building a new query engine. Our solution to this problem is a lightweight, unified API, called NoDA [3, 4] (https://github.com/the-noda-project), that consists of simple data access operators, such as `filter`, `project`, `sort`, `limit` and `aggregate`. Inspired by the ODBC/JDBC paradigm in relational databases, NoDA defines data access operators that are implemented for different NoSQL stores. Using NoDA, developers can express their queries in the same language, but target different NoSQL stores by simply changing only the connection to the underlying store. Perhaps most importantly, NoDA's data access operators have enabled the provision of an SQL interface which takes as input a SQL statement, translates it to NoDA data access operators, which can be directed to any of the supported NoSQL stores. Currently, we have implemented NoDA [3] for diverse NoSQL stores: MongoDB (document store), HBase (wide-column store), Redis (key-value store) and Neo4J (graph database).

2 FUTURE RESEARCH DIRECTIONS

Several interesting research directions can be followed in the future:

- How our approach can be exploited to fetch data stored across multiple NoSQL stores and retrieve the combined results.
- Handling more complex data types is also challenging; currently, we work on spatio-temporal data, but other complex types are of interest, such as trajectories, graphs and textually annotated spatial data.
- Our approach focuses on analytical queries, so extending it towards supporting updates is also of interest.
- How to efficiently support joins of distributed data collections is another challenging direction, even more across different NoSQL stores.

Published in the Proceedings of the 2nd Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores, co-located with VLDB 2021 (August 16-20, 2021, Copenhagen, Denmark) on CEUR-WS.org.

ACKNOWLEDGMENTS

The research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: HFRI-FM17-81).

REFERENCES

[1] Ali Davoudian, Liu Chen, and Mengchi Liu. 2018. A Survey on NoSQL Stores. *ACM Comput. Surv.* 51, 2 (2018), 40:1–40:43.

[2] Jennie Duggan, Aaron J. Elmore, Michael Stonebraker, Magdalena Balazinska, Bill Howe, Jeremy Kepner, Sam Madden, David Maier, Tim Mattson, and Stanley B. Zdonik. 2015. The BigDAWG Polystore System. *SIGMOD Record* 44, 2 (2015), 11–16.

[3] Nikolaos Koutroumanis, Nikolaos Kousathanas, Christos Doulkeridis, and Akrivi Vlachou. 2021. A Demonstration of NoDA: Unified Access to NoSQL Stores. In *Proceedings of the 47th International Conference on Very Large Data Bases (VLDB'21), Copenhagen, Denmark, August 16-20, 2021.*

[4] Nikolaos Koutroumanis, Panagiotis Nikitopoulos, Akrivi Vlachou, and Christos Doulkeridis. 2019. NoDA: Unified NoSQL Data Access Operators for Mobility Data. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases, SSTD 2019, Vienna, Austria, August 19-21, 2019.* 174–177.

[5] Raghav Sethi, Martin Traverso, Dain Sundstrom, David Phillips, Wenlei Xie, Yutian Sun, Nezih Yegitbasi, Haozhun Jin, Eric Hwang, Nileema Shingte, and Christopher Berner. 2019. Presto: SQL on Everything. In *35th IEEE International Conference on Data Engineering.* 1802–1813.

[6] Michael Stonebraker. 2008. Technical perspective - One size fits all: an idea whose time has come and gone. *Commun. ACM* 51, 12 (2008), 76.