Towards Responsible Decentralized Data Architectures

Ilaria Battiston [1], Peter Boncz[1]

[1] CWI, Amsterdam, The Netherlands
{ilaria,boncz}@cwi.nl

We propose an alternative to the standard cloud-centralized architecture, leaving part of the application data under the control of the individual data owners in *decentralized personal data stores* [1]. Our primary goal is to increase **data minimization**, i. e. to enable more sensitive personal data to be under the control of its owners while providing a straightforward and efficient framework for designing such architectures. As a platform for our prototyping, we choose the open-source novel data management system DuckDB.

We focus on the design of a declarative framework in which information architects can use SQL to split a model between a *centralized* and *decentralized* part; we define specific keywords to establish the physical location of the tables as well as the column attributes, such as sensitivity and minimum level of aggregation for them to be privacy-preserving.

The centralized part of the schema contains aggregating views over this decentralized data, which can be seen as a buffer to keep intermediate or partial results until they satisfy the privacy requirements and can be permanently stored in centralized tables. Local updates need to be reflected in the centralized views; for this, we pursue the integration of distributed materialized view maintenance. We implement incremental computations in DuckDB [2], building a SQL-to-SQL compiler to easily port instructions across multiple components in our infrastructure.

However, query computations must also be hidden while results are being collected and processed. We investigate hardware-based security techniques such as Intel SGX and multi-party computation (MPC) to do so. We port DuckDB to SGX 2 [3], resulting in a maximum of 1.2x overhead for TPC-H, making it feasible to leverage enclaves for our use case. Future work includes investigating MPC and applying incremental maintenance techniques to it.

To evaluate its performance properties, we aim to implement and test this system, where personal data stores could live on mobile devices or in encrypted cloud storage.

References

[1] Ilaria Battiston and Peter Boncz. Improving data minimization through decentralized data architectures. In *VLDB 2023, PhD Workshop*, number 3452, August 2023.

[2] Ilaria Battiston, Kriti Kathuria, and Peter Boncz. Openivm: a sql-to-sql compiler for incremental computations. In *Companion of the 2024 International Conference on Management of Data*, SIGMOD/PODS '24, 2024.

[3] Ilaria et. al. Battiston. Duckdb-sgx2: The good, the bad and the ugly within confidential analytical query processing. In *Proceedings of the 20th International Workshop on Data Management on New Hardware*, SIGMOD/PODS '24, 2024.