Time Machine: Real-Time Analytics on Arbitrarily Fresh Data

Keywords: speculative execution, fresh data, HTAP, view maintenance, approximate query processing

Problem: As organizations and companies increasingly depend on real-time insights, they require timely and accurate analytical processing over the latest data. The increasing demand for supporting hybrid transactional and analytical (HTAP) workloads has led to novel DBMS architectures that frequently propagate fresh transactional data for analytical queries to process. However, efficiently processing fresh data is challenging. First, update propagation often requires data transfers, and possibly, maintenance operations on auxiliary structures such as views and indices, which increase the latency of analytics. Second, analytical queries can be long-running and hence limit the ability of DBMS to provide responses over seconds-old data. Therefore, the propagate-then-execute paradigm provides only limited freshness.

Project: The goal of this project is to empower DBMS to respond to analytical queries over arbitrarily fresh data. To this end, the DBMS should decouple fresh data propagation and analytical query processing. Recent work [1] uses speculations to decouple and parallelize inter-dependent sub-queries in complex query workloads. In this project, the student is expected to adapt the speculative framework to the HTAP use case. By using stale data to extrapolate newer data versions and by deferring the application of updates, the DBMS can update query results to reflect the state of the database at response time. This result has wide-ranging implications for a critical problem of data management.

Plan:  
1. Familiarize oneself with background work on HTAP, incremental view maintenance and speculative execution.  
2. Extend the theoretical framework of [1] to incorporate the concepts of data freshness and design a mechanism for deferred update propagation.  
3. Implement a prototype system that practically demonstrates the ability of the techniques. Ideally, the prototype will be built on top of a mature analytical engine (i.e., Presto) or an HTAP system.

References


Supervisor: Prof. Anastasia Ailamaki, anastasia.ailamaki@epfl.ch

Responsible collaborator(s): Dr. Ioannis Mytilinis, ioannis.mytilinis@epfl.ch  
Panagiotis Sioulas, panagiotis.sioulas@epfl.ch  
Viktor Sanca, viktor.sanca@epfl.ch

Duration: 1 semester