Applied Time Traveling with Temporal Data Management

Degree programme : Master of Science in Engineering | Specialisation : Computer Science
Thesis advisor : Prof. Marcel Pfahrer
Expert : Prof.Stefan F. Keller (OST Ostschweizer Fachhochschule)

Any software data application running in production for several years inevitably undergoes data modifications. Those typically come as corrections of wrongly inserted data (e.g. typos) and real-world changes that require data to evolve (e.g. population). This thesis examines how temporal data management techniques can enhance an existing software system to enable seamless navigation over several time axes. As seamless as traveling through time.

Eugene Kudryavtsev

In the context of this thesis, a particular data-heavy application from the agricultural domain is under focus: RISE (abbr. Response-Inducing Sustainability Evaluation). It's an indicator-based decision support system backed by a detailed survey. During a visit to the farm, the survey is filled out with farm data and an extensive tree of calculations produces a sustainability overview of the farm according to 10 themes (from „Soil use" to „Working Conditions" via „Animal Husbandry").

These calculations use survey answers, static regional reference data (e.g. dry matter in a particular cereal), and other (sub-) calculations to produce scalar or vectorial outputs. A collection of these outputs aggregated into the 10 themes, constitutes a sustainability report that is printed and archived. As time progresses, calculation formulas are corrected and adjusted, reference data is updated and answers might be subsequently modified. These temporal changes pose a challenge in reconstructing sustainability reports to compare them. In this comparison lies the real key to quantifying sustainability.

Goals of the project

The first goal of this thesis is to transform a „single temporal" and snapshot-based data architecture into a multi-temporal one. The „single temporal" approach has the obvious disadvantage of keeping only the latest value, and the snapshot-based makes all data modifications only local to the snapshot at hand. Moving an existing application from such data-handling paradigms requires merging snapshots and data archeology to find previous values.

The second is to enable an intuitive, time-traveling approach to navigate and use that data through time. Practically, that would allow for a given survey to perfectly reproduce a sustainability report at any point in time, with the values of the reality at that time, with or without corrections. This is the part that must be intuitive and feel like time traveling.

And finally, data-managerial aspects are tackled to enable the point-in-time data modification and extension of validity time ranges to avoid gaps in time axes. This should allow users to correct data without fearing overwriting or deleting values.

The takeaway

Once this data transformation is completed, setting data changes in the future becomes possible (e.g. setting a value that will be valid only starting next year with knowledge available now). And ultimately, with enough confidence in this data architecture, sustainability assessments can be dropped from storage altogether, as the calculation formula can reproduce any calculated result in a sustainability assessment without the need for relevant data to be present in the snapshot.