Contribution ID: **10**   Type: **Talk**

Tabular Data Synthesis for Data Management

Tuesday, March 12, 2024 11:15 AM (45 minutes)

The problem of generating synthetic data is almost as old as modern research itself. However, with the advent of generative AI, new possibilities for synthesizing tabular data have emerged that go far beyond the capabilities of traditional statistical or rule-based approaches. Most of this new research comes from the ML community, where ML models need to be fed with useful training data. Since many data management use cases also require synthetic data, it makes sense to adapt these research results. Nevertheless, those use cases, such as query optimization, have different requirements than ML use cases. Requirements that are currently not met by such modern synthesizers. In this talk, we will give an overview of the current state of the art in the field of tabular data synthesis and discuss open challenges in the context of generating synthetic tabular data for data management.

Type of Poster

Presenter: PANSE, Fabian (HPI)

Session Classification: Talks

Track Classification: Vortrag