Data science is mostly not about doing data science...

Data scientists spend about 80% of their time on data preparation (while perceiving this as the least enjoyable activity) [1]. Collecting, organizing, and cleaning are particularly time-consuming. Metacrate aims to facilitate these tasks in three steps...

Chart: Average time spent on tasks throughout data science projects [1].

Profile.

Data profiling tools, such as Metanome [2], can extract a wide array of data profiles; particularly, data dependencies and data summaries.

Organize.

Metacrate’s flexible data model organizes the various data profile types. It supports several storage backends, thereby providing scalability.

Analyze.

Metacrate allows to query, integrate, and analyze data profiles. Besides a set of common visualizations, Metacrate provides a metadata analytics library. This combination facilitates data anamnesis, data discovery, and data cleaning.