Procedures are named, reusable programs used to implement long-running batch operations with no return values. Procedures generally run over Datasets and can be configured via SQL expressions. The outputs of a Procedure can include Datasets and files. Procedures are used to:

- Transform or clean up data
- Train machine learning models
- Apply machine models in batch mode

MLDB provides a number of built-in Procedure types for various use-cases. Learn more about Procedures.
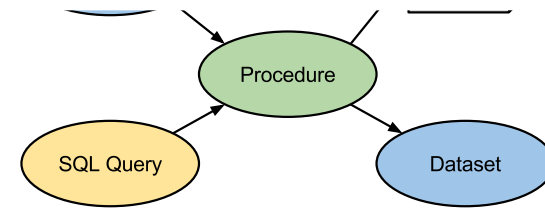
Datasets

Datasets are named sets of data points composed of (row, column, timestamp, value) tuples. Datasets can have millions of named columns and rows, and data points can take on multiple timestamped values which can be numbers or strings.

Datasets can be loaded from or saved to files, and can serve as the input to Procedures as well as be created by them. And of course, Datasets can be queried via SQL queries.

MLDB provides a number of built-in Dataset types for various use-cases. Learn more about Datasets.
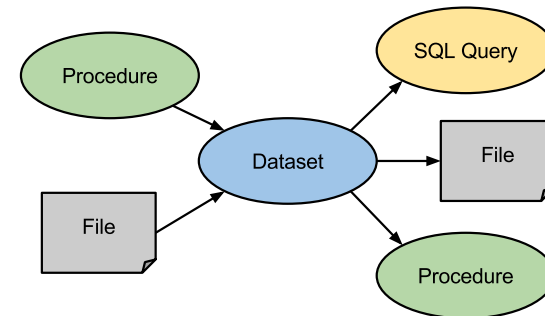
Files

MLDB gives users full control over where and how data is persisted. Datasets can be saved and loaded from files. Procedures can create files and Functions can load up parameters from files.