Big Data Architectures

Polystores

Paweł Guzewicz

École Polytechnique, Inria

M2 Data and Knowledge 2019/2020

Université Paris Saclay

So far during labs 1-3

- We have seen
  - Redis
  - MongoDB
  - Neo4j

- These systems differ in their:
  - Architecture design
  - Storage type
  - Query language
  - Database schema
  - Interface

Comparison of systems used in previous labs

System	Storage type	Database schema	Query language	Interface we used
Redis	Key-value	Dictionary mappings	set(), get(), …	Programmatic application interface (API)
MongoDB	JSON document collections	Semi-structured, JSON schema	MongoDB query language	Command-line interface (CLI)
Neo4j	Property graph	Graph schema	Cypher	Web user interface

Today: Apache Drill

- **Polystore**: modern incarnation of mediators
- **Apache Drill** is one open-source polystore
  - Querying heterogeneous sources using SQL: relational schema view
- Allows to read, e.g., CSV, Apache Parquet files, and to connect to RDBMs, MongoDB, Hive, HBase, etc. through *plug-ins*
- Internally implemented as a *column store*