Advanced Database and Information Systems

Dr. Fang Wei-Kleiner

16. April 2021

ADBIS: How to process semi-structured data efficiently?

Traditionally:

- SQL-Queries (Relational Databases) when data is structured in tables.
- Processing takes place on a single machine.

Weaknesses of the Relational Model (- critical for certain applications)

- Inadequate Representation of Data.
- Semantic overloading.
- Weak support for recursion.
- Homogeneity.
- Rigid schema.
- Limited throughput.

Traditionally:

- SQL-Queries (Relational Databases) when data is structured in tables.
- Processing takes place on a single machine.

Today:

- Data is organized in complex structures, only partially structured or even unstructured.
- Sparseness and schema independence (schemaless, data-first-schema last).
- Self-descriptive data.
- Large data volumes.
- Scalability by distributed processing on computer clusters.

Course Outline

Number of lectures: 13

(1) **Semistructured Data Models** (4)
XML, JSON and Graph Databases

(2) **Beyond Relational Database Systems** (1)
Columnar Stores and NoSQL Database Systems

(3) **Distributed Processing using Computer Cluster** (2)
Hadoop, MapReduce, PigLatin, Hive and Spark

(4) **Relational Query Languages leveraged by Recursion** (4)
SQL, Datalog

(5) **Advanced topics** (2)

Organization

- Lecture: Tuesday 10-12
- Tutorials: Thursday 10-12 (the first week there will be a lecture)
- ECTS: 6 Points
- Program of Study: Bachelor CS, Master CS/ESE, Lehramt Informatik
- Exam: oral; date tba

Tutorials: Fang Wei-Kleiner

Literature:

- *Advanced Data Management For SQL, NoSQL, Cloud and Distributed Databases*. Lena Wiese. Berlin/Boston De Gruyter 2015
- *Spark in Action*. Petar Zeceviuc and Marko Bonaci. Manning Shelter Island, 2016.
- more literature suggestions during the lectures.