Flexible Modeling and Execution of Data Integration Flows

Pascal Hirmer, University of Stuttgart, Institute of Parallel and Distributed Systems

**Thesis Goals and Research Problem**

- **Flexible** modeling and execution of data integration flows
- **Iterative and explorative** trial-and-error-like data integration
- Support of **various data sources**, structured and unstructured
- **Use case dependent** implementation and execution (efficient / robust / etc.)
- Enabling usage by non IT expert users

**Method for Modeling and Executing Data Integration Flows**

1. Domain-specific data source modeling using **business objects**. Support of heterogeneous data sources, e.g., relational databases, unstructured text, sensor data, etc.
2. Domain-specific modeling of various data operations using **modeling patterns**.
   - Result: domain-specific flow model
3. Transformation **Pattern selection** according to the use case scenario. Example patterns: Robust, Time-Critical, Big Data, Secure, ...
4. Transformation of **domain-specific** data flow into **technical, executable** data flow.
   - Format based on the selected patterns.
5. Data flow **execution** using a **suitable execution engine**.
6. **Storage** of the integrated result in a suitable data store.
7. Utilization for **visualization** purposes, **analysis** and other value-adding scenarios.

References: https://www.ipvs.uni-stuttgart.de?id=Pascal.Hirmer