# HighQ: Information Integration using the IBM Information Server

## Goal

Provide and evaluate mechanisms for monitoring and solving data quality issues

# **Duplicate Detection**

- Faster Duplicate Detection
  - Custom similarity measurements
  - Fewer comparisons with windowing strategy
  - Better results through multiple execution
- SAP related optimization
  - Default similarity measurements

#### **Data Fusion**

- Enhanced conflict resolution strategies
  - Representing further relations between data fields
- Subsumption
  - Effectively dealing with NULL-Values
  - No configuration required
  - Leveraging parallelization capabilities

## **Data Quality Monitoring**

- SAP-related Data Quality Monitoring
  - Integrated change history for SAP products
  - Predefined rules for defining data quality in terms of SAP
  - SAP-related attribute weighting

Data Fusion

HighQ

Duplicate Detection

**Data Quality Monitoring** 







Peter Faymonville, Dandy Fenz, Daniel Hefenbrock, Matthias Pohl, Florian Reinhart, Gerald Töpper, Sönmez Ulutürk

Advisors: Alexander Albrecht, Prof. Felix Naumann In association with IBM: Albert Maier, Martin Oberhofer, Yannick Saillet, Thomas Schwarz