Business focus:
• Data analytics & information delivery solutions with focus on geo-spatial data
• Implementation of Data Warehouses with spatial and non-spatial data

Product Cadenza:
• Our platform for geo-analytics & data discovery
• Several thousand users in Germany and Austria

Headquarters: Karlsruhe, Germany. 100 employees.
WIRE: Intelligent Methods for Integration and Quality Assurance of Geo Data

- German SME research project, Funded by BMBF, Grant # 01IS16039
- Duration: 01/2017 – 12/2018

Application scenarios:

- Smart agriculture
- Environment monitoring / water management

Project Partners

- Disy Informationssysteem GmbH
- FZI, Research Centre for Information Technology
**Motivation**

- Massive growth of application big geo data to be fused in real-time
- Data management and quality control needs more automation and tool support

**Some methods and tools developed:**

- Geospatial extension for Talend Data Integration workbench
- Machine-learning services for geo-data schema mapping
- Algorithms for detection of geo-data quality problems
- Automated corrections of geo-data errors
What is ETL?

- Extract
- Transform
- Load

Filter, Analyse, Visualize
Integrated quality data
What is Geo-ETL

„E“ Extract
+ (Formats, Datenbanse, Interfaces)

„T“ Transform
+ Geometric Operations
+ Validation und correction (geometric)
+ Combination of Geo- und Non-Geodata

„L“ Load
+ (Formats, Datenbanse, Interfaces)

Filter, Analyse, Visualize
Integrated quality data
Leading Platform for Big Data and Cloud and Data Integration

- User Interface for Job definition / code creation
- Hundreds of connectors, components and routines
- Repository management for Job-reuse and teamwork
- Monitoring and logging optimized for large installations
- Free and Open Source but still powerfull entry version

BUT: no Geodata and no Geo-Operations
GeoSpatial Integration for Talend
„Everything you need to extend ETL to Geo-ETL“

Data-Connectors for
- Oracle Locator/Spatial, PostGIS
- SpatiaLite, Shapefile, WKT, WKB
- GeoJSON

Components for
- Length and area calculation
- Geometry transformation
- Geometry validation
- Buffer, Bounding Box, Centroid.…

Routines
- More than 40 geometric Operations
- Like Visvalingam-Wyatt Algorithm (Simplification)
GeoSpatial Integration for Talend
Use drag & Drop with new components

GeoSpatial components are fully integrated in the Talend ETL process.
GeoSpatial Integration
For Talend Real-time Big Data

- Use Big Data Capabilities of Talend Platform for Geodata-Processing
- Create and Deploy Spark-Code with Geo-Routines Hadoop/Spark-Clusters

Parallele execution in a Spark-Cluster to process Big Data Streams coming from Sensors, Social Media etc.
Case Study
San Francisco

San Francisco Municipal Transport Authority

- All Vehicle send radio signals every minute
- Metro (MUNI), Busses, Car- and Bicycle-sharing
- Data are stored in a Hadoop Data Lake
- Talend Big Data Platform + GeoSpatial Plugin for Talend
- Currently: mainly data analysis.
- Vision: Real-time data services for other departments, the public & companies
Intelligent ETL-Processes for Geo-Data

Thank you for your attention

You can download and try the plug-in for free here…

Contact

Dr. Wassilios Kazakos
Disy Informationssysteme GmbH
Tel. +49 721 16006-000
wassilios.kazakos@disy.net

www.disy.net/spatial-etl