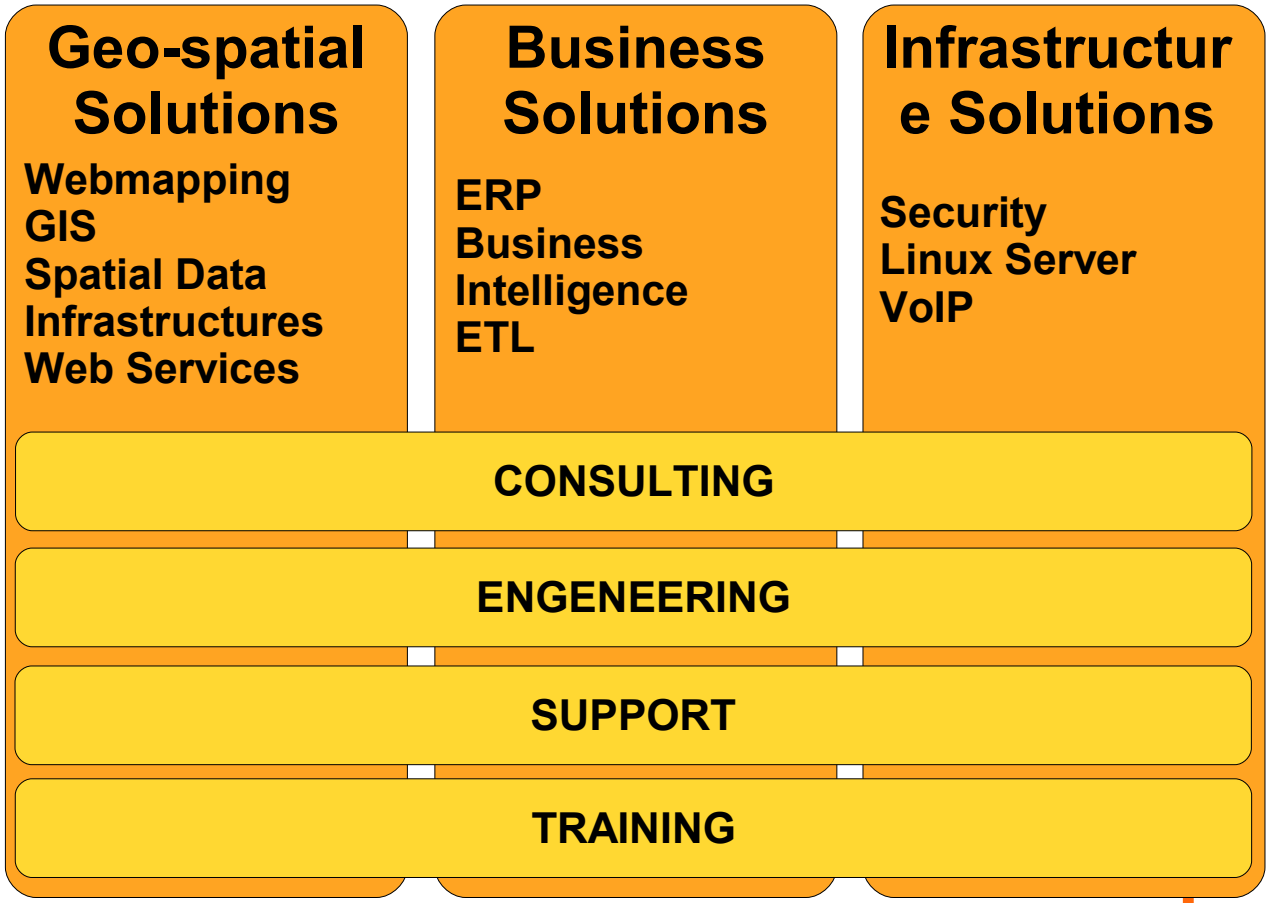
Spatial Data Integrator powered by talend

*open data solutions

Open Source Spatial ETL

Camptocamp, an Open Source Base Camp !

- **35 employees**
  - Switzerland & France
- **About 50 to 70 % of growth per year since 2002**
- **3 activity domains**
  - Spatial solutions
  - Business solutions
  - Infrastructure solutions
- **4 services poles**
  - Consulting
  - Engeneering
  - Supporting
  - Training



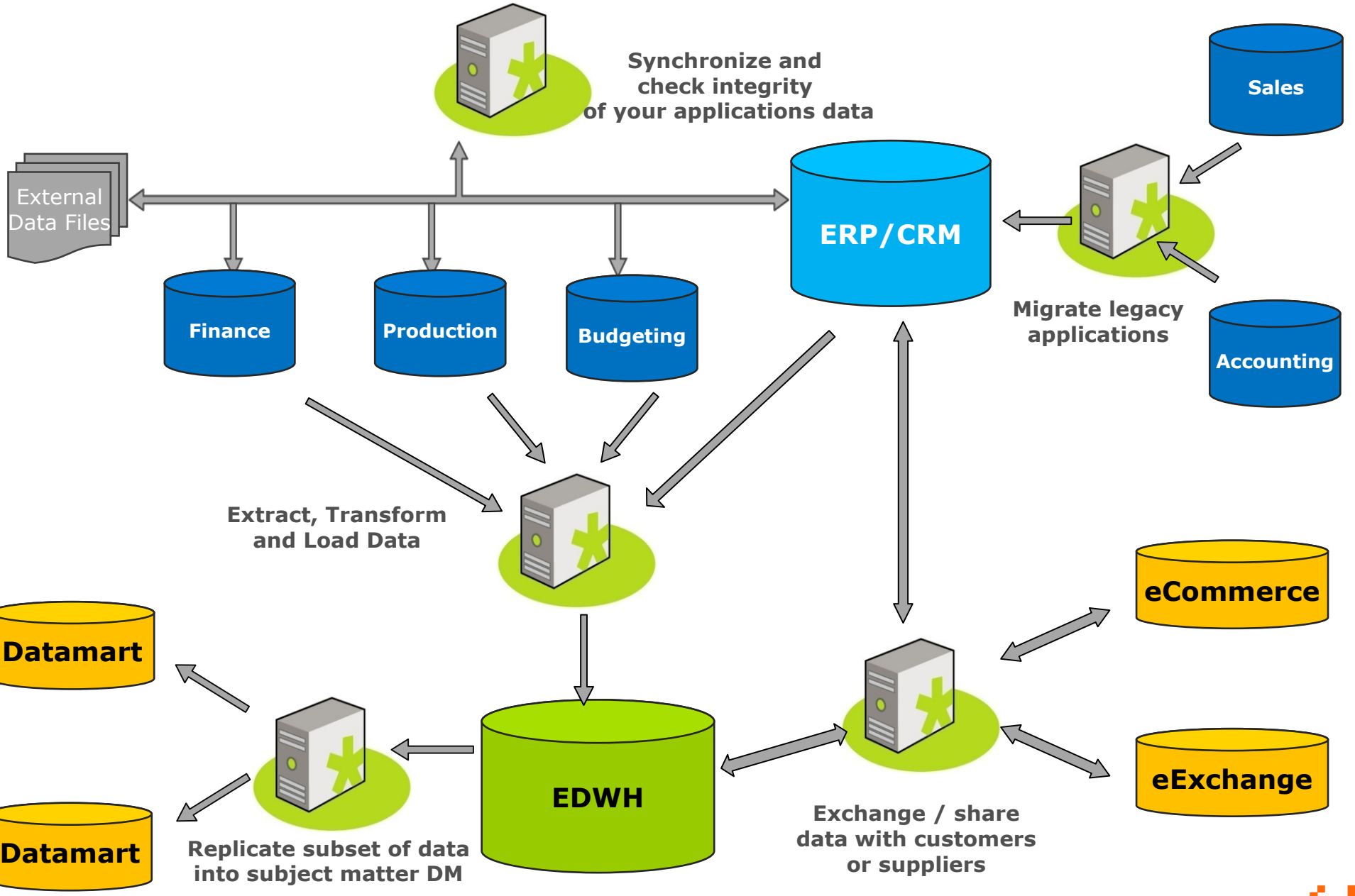
Talend overview

- Talend is the first provider of open source data integration software
- Located in France, USA, Germany, China
  - VC-funded
  - 50 employees
- First product release: 2006
- Leader in open source data integration
  - Rival large established proprietary players

Why Spatial Data Integration

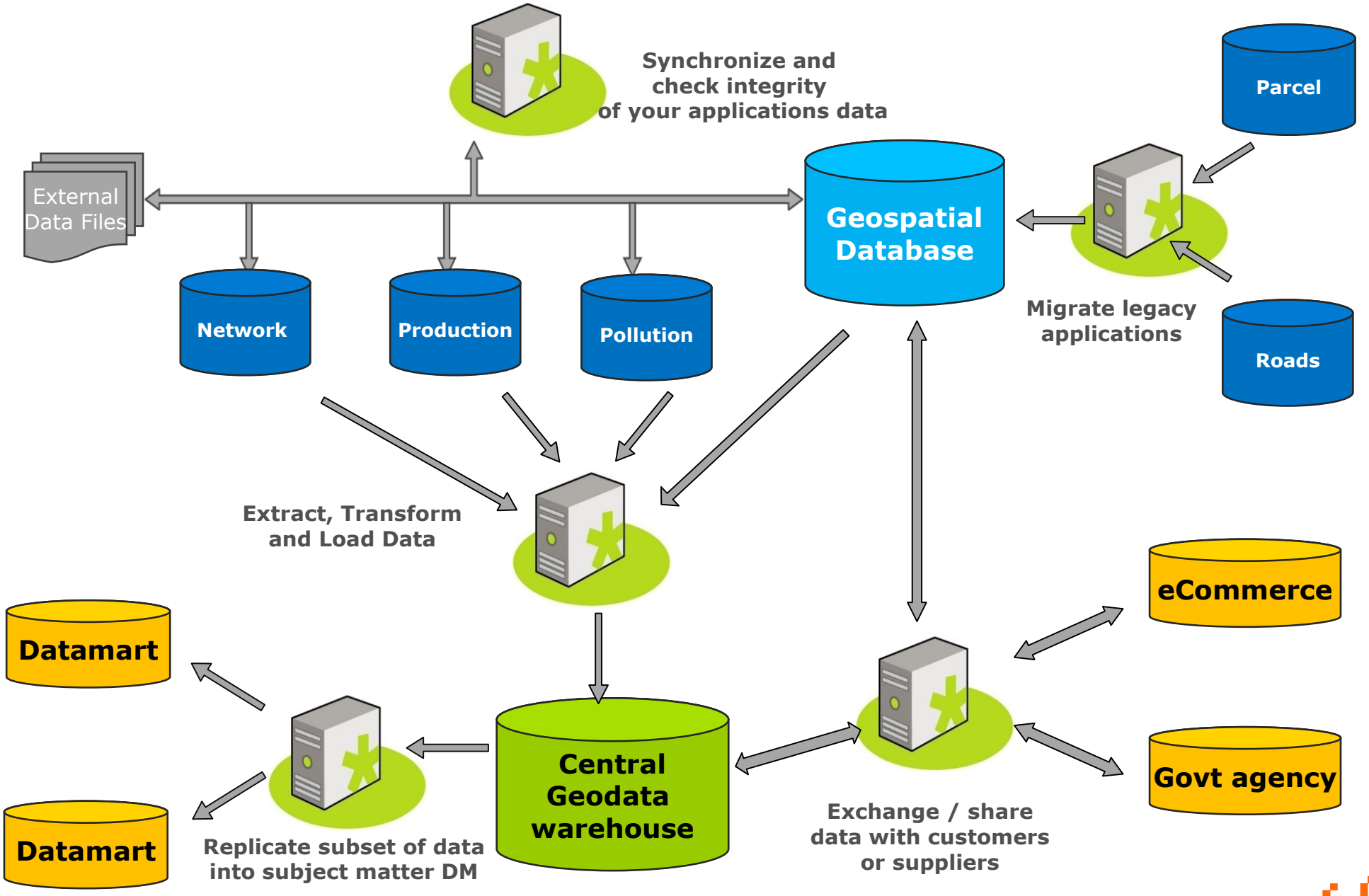
- Data integration is a key process
  - Data volumes in exponential growth
  - Diversity and heterogeneity of data sources
  - Data processing plays a major role in implementing GIS projects
  - Consolidating and aggregating spatial data with data from other sources is often required
- GIS data integration situation
  - Use command or hand-made script from various tools and libraries
    - gdal/ogr commands, fwtools, postgis command, ...
  - Proprietary Spatial ETL such as FME
  - Lack of Open Source global geo-spatial data integrator
- **Spatial Data Integrator,** *Powered by Talend* is now available!

Talend Data integration



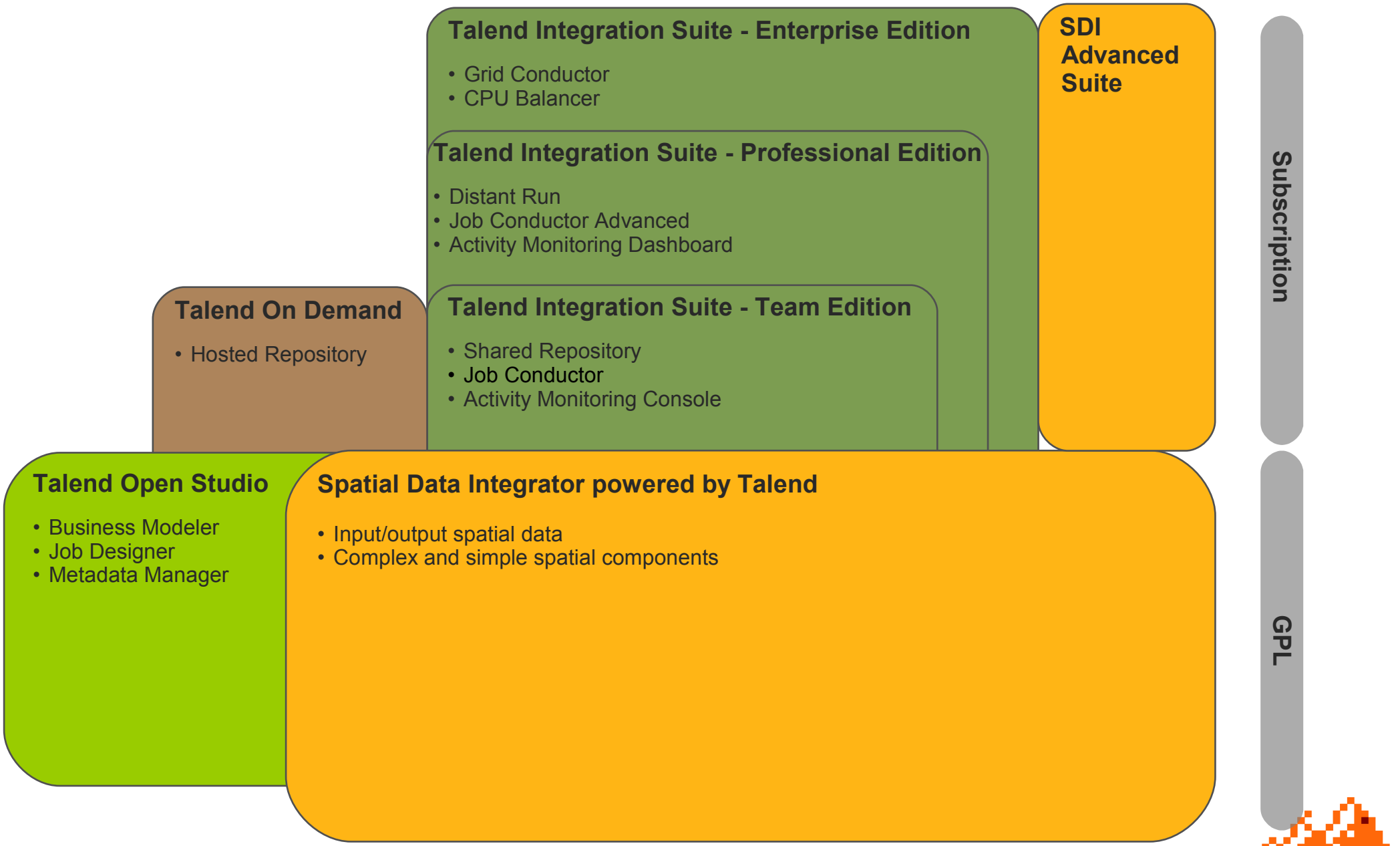
camptocamp

Spatial Data integration



camptocamp

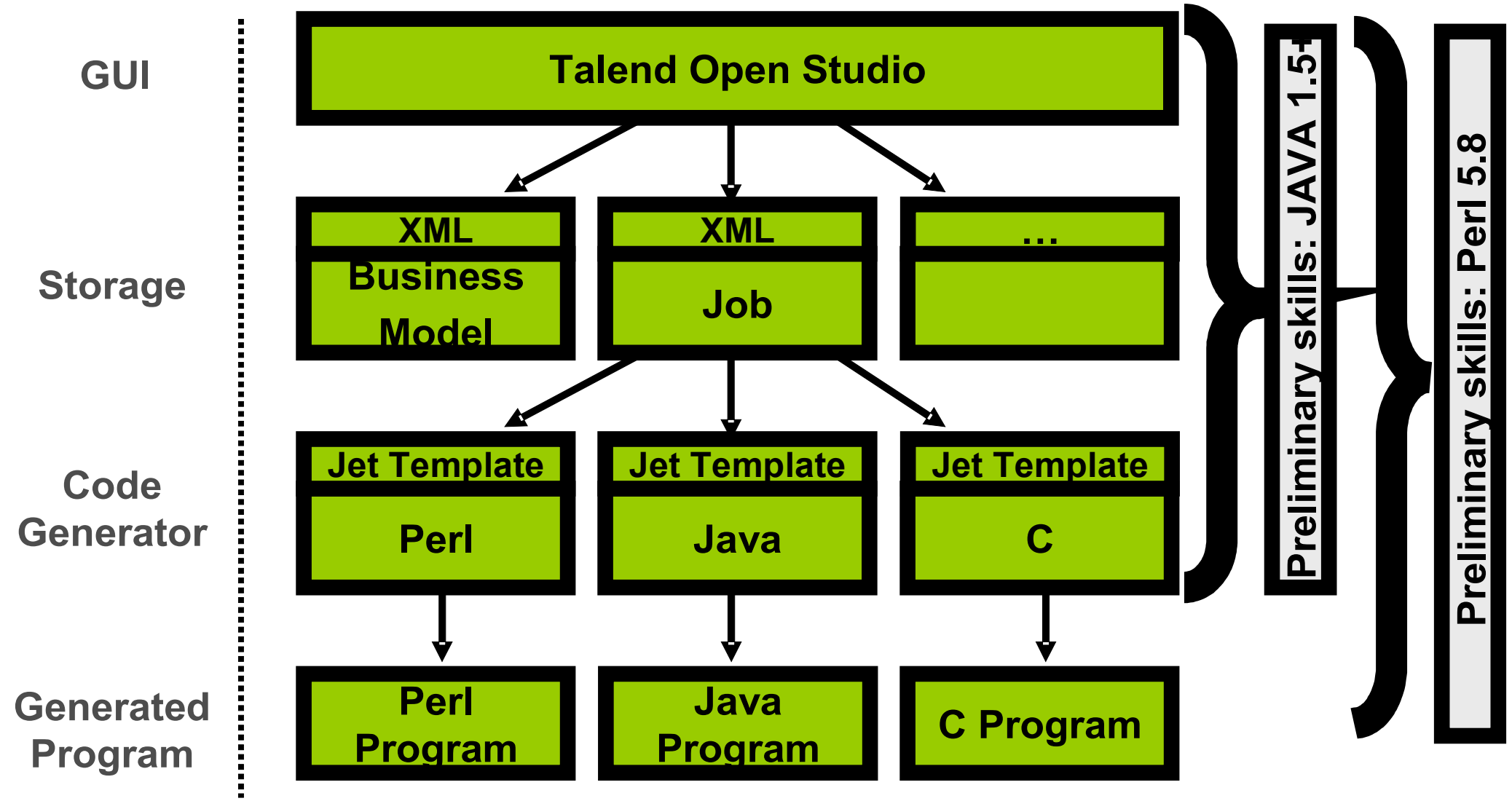
The Talend offering



Talend Open Studio

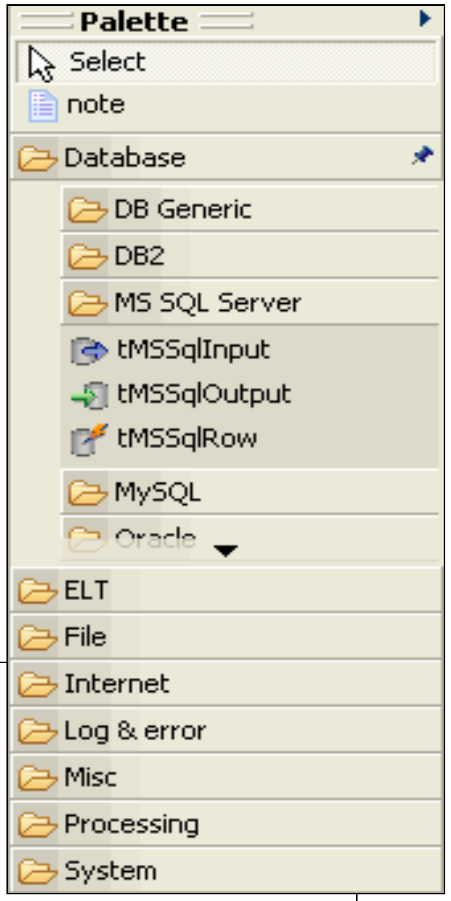
- Key features
  - Business-oriented process modeling
  - Graphical development
  - Robust and scalable execution
  - Broadest connectivity to support all systems
  - Project repository for design and execution
  - Real-time debugging
- A high adoption rate
  - 100,000 product downloads
  - 20% register as users
- Active community
  - 1,000 beta testers
  - 500 forum contributors

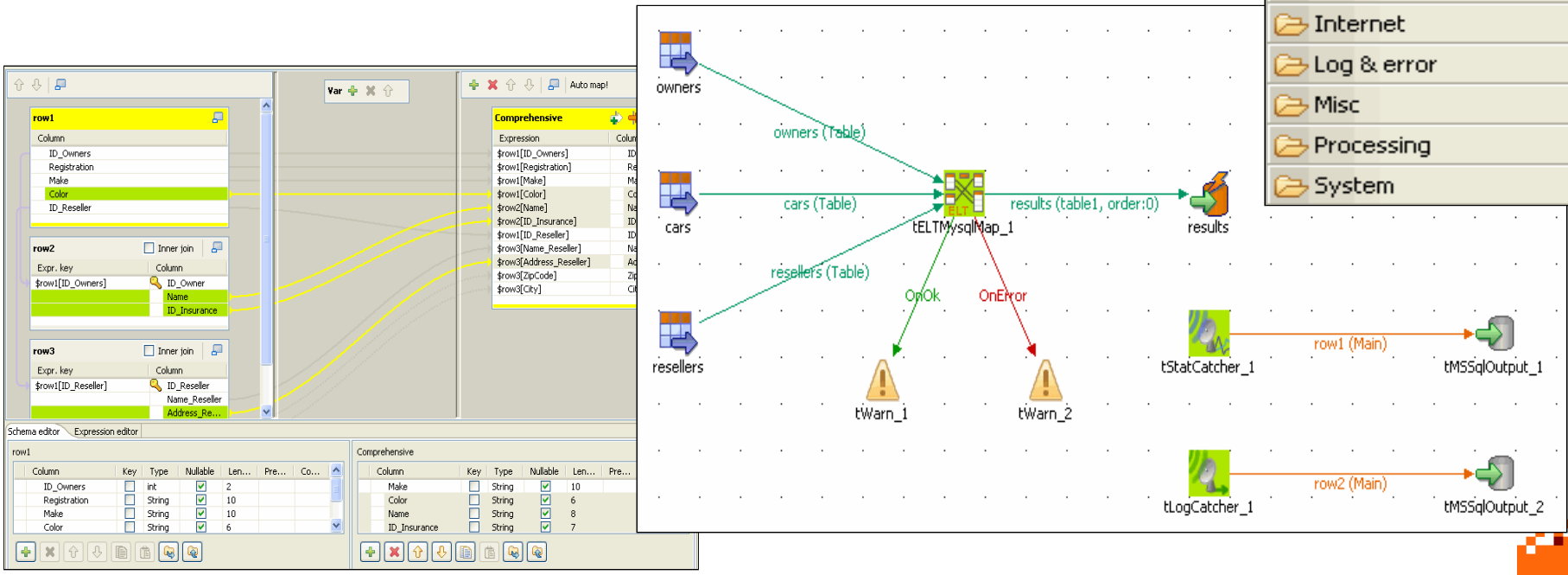
Talen Open Studio architecture



Productivity & Ease of Use

- Graphical development
  - Dramatically increased productivity & ramp up
  - Combined graphical & technical views
  - Drag-and-drop mapping interface
  - Large library of components & connectors
- Leverage industry-standard languages
  - Java, Perl, SQL

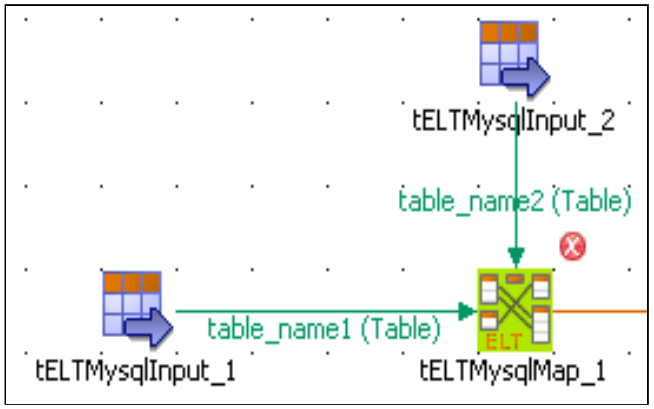


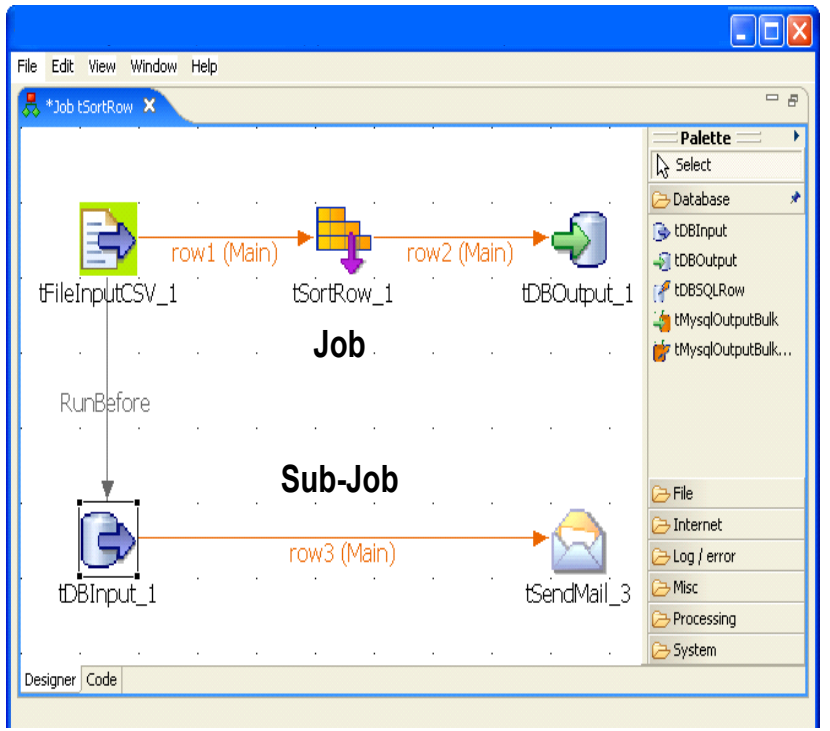


camptocamp

Performance and robustness

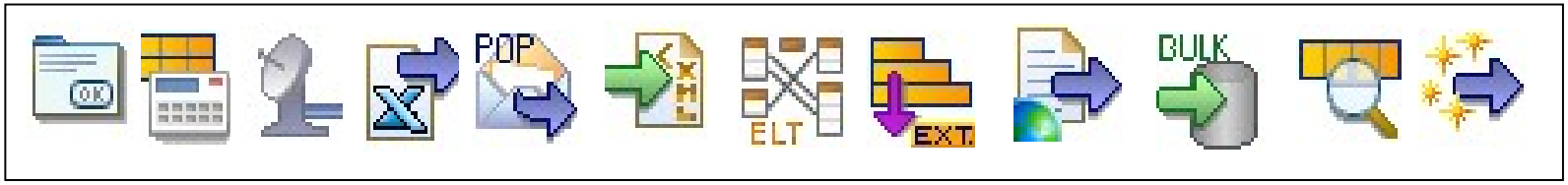
- Highest performance, robust and scalable execution
  - Grid-distributed processing
  - Industry-standard code generated (Java or Perl)
  - Leverage both ETL and ELT architectures
  - Process data closest to the source





Versatility through Connectivity

- Broadest connectivity to support all systems
  - 100+ connectors available out of the box
- Business Applications:
  - SugarCRM, SalesForce.com, LDAP…
- RDBMS:
  - Oracle, MySQL, DB2, SQL Server, Sybase, Ingres, PostgreSQL…
- Web:
  - Web Services, FTP, HTTP, POP, SMTP…
- Files:
  - Delimited, positional, XML, Excel…

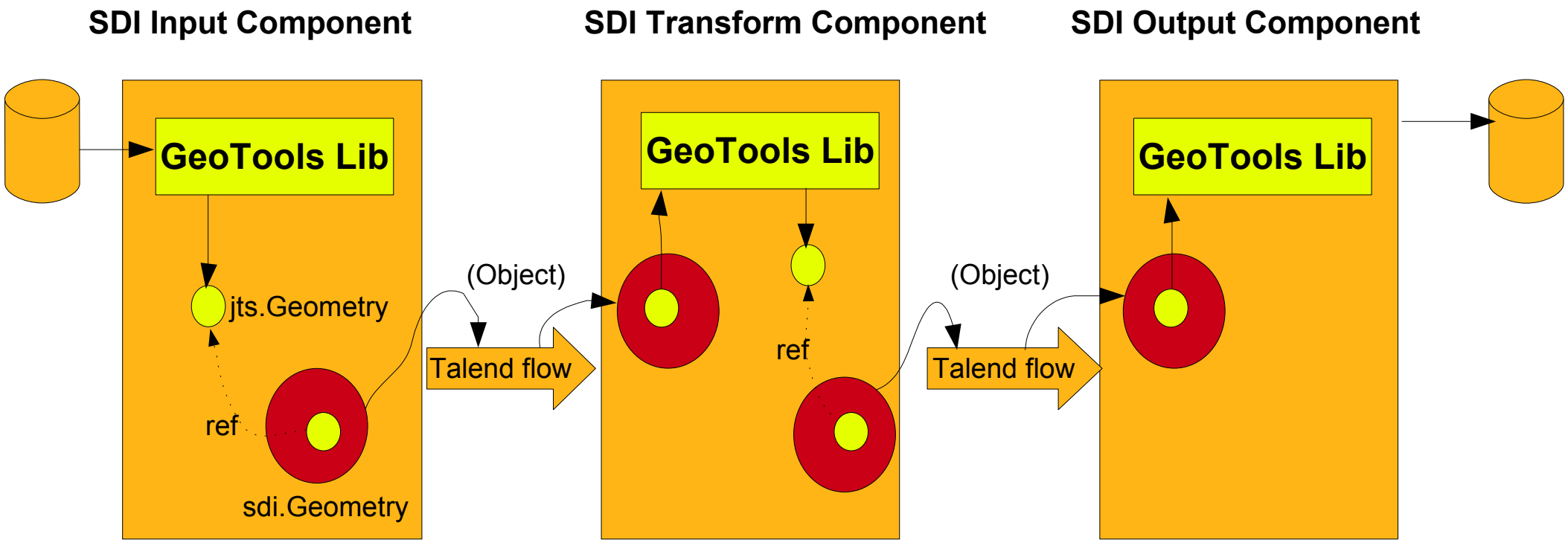


Spatial Data Integrator - SDI

- Talend Open Studio with geo-spatial extensions
- SDI integrates a new family of vector and raster geo components
- Based on reliable open source tools:
  - Java Topology Suite
  - GeoTools
  - GRASS

Spatial Data Integrator Architecture

- Uses GeoTools / Java Topology Suite librairies



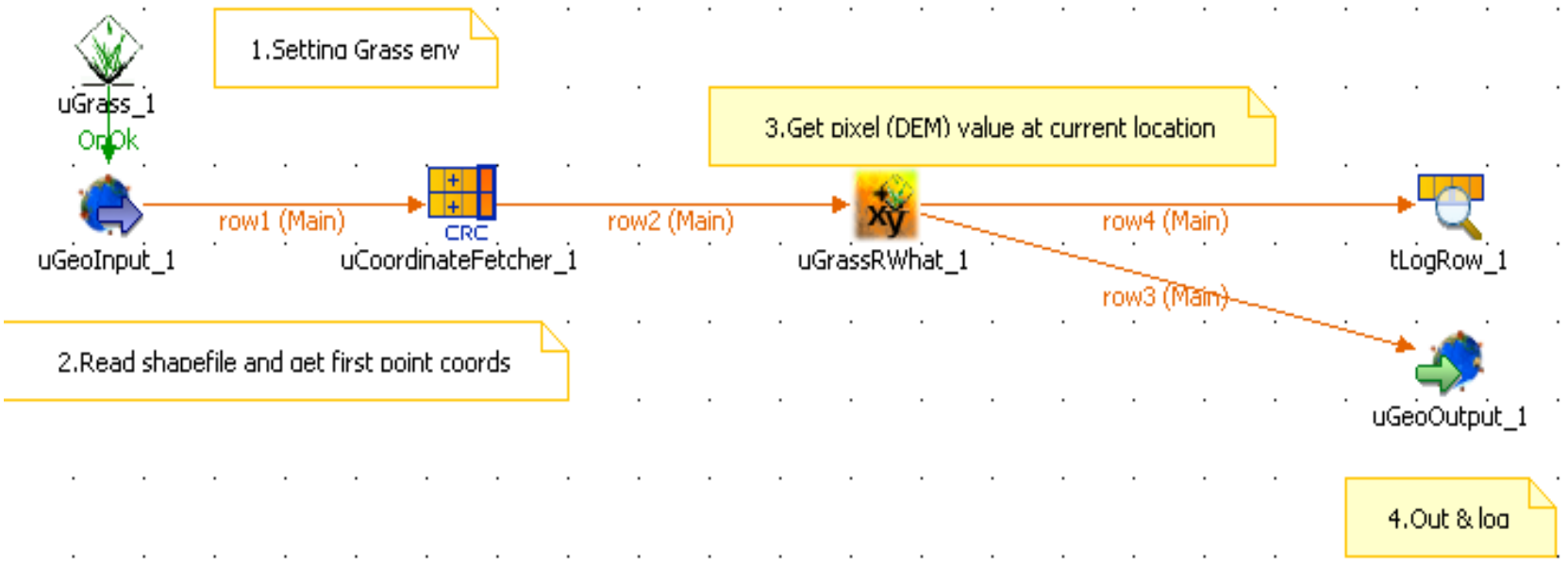
camptocamp

Spatial Data Integrator: Specific components

- File:
  - uMapinfoInput
  - uMapinfoOutput
  - uShapefileInput
  - uShapefileOutput
- Database:
  - uPostgisInput
  - uPostgisOutput
- Collectors:
  - uBoundingBoxAccumulator
  - uConvexHullAccumulator
  - uNeighborFinder
- Calculators:
  - uAreaCalculator
  - uBufferCalculator
  - uCentroidCalculator
  - uCoordinateFetcher
  - uDecimalDegreesCalculator
  - uEnvelopeCalculator
  - uLengthCalculator
- Manipulators:
  - u2DPointReplacer
- Geometric Operators:
  - uDissolver

Spatial Data Integrator: Raster components

- Raster components use GRASS tools
- GRASS components:
  - uGrass
  - uGrassRContour
  - uGrassRInGdal
  - uGrassRWhat
  - uGrassRWoutOgr



Spatial Data Integrator strengths

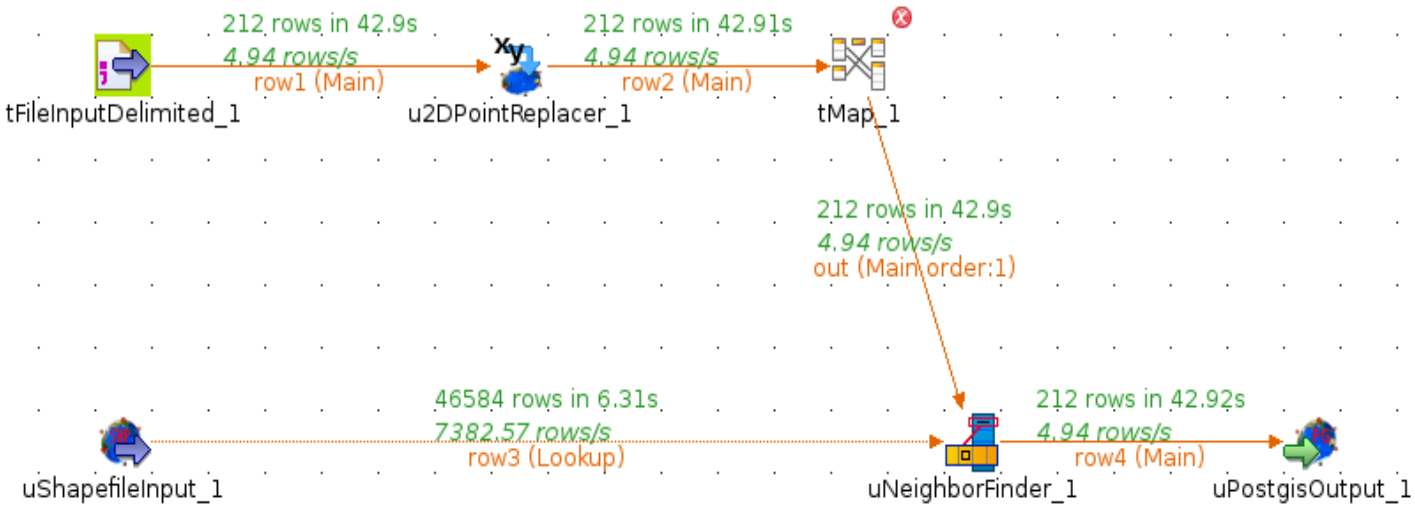
- Fast and efficient
- User-friendly Interface
- Easily customizable jobs (code generation)
- Benefits of « classical » ETL features
- Fully Open Source (GPL licence)
- Scalable
- High level of support by Camptocamp and Talend

Perspectives

- Development of new components:
  - Simple and complex components
  - New input and output formats
  - Community contribution
- Spatial data viewer (uDig)
- Raster components optimization (JGrass)
- Integration of high-end Talend features:
  - Load balancing, Job conductor, Grid conductor
- Integration in Entreprise Service Bus (ESB) systems (PEtALS)

Demo

- Input:
  - CSV file with x,y and attributes columns (monitoring station)
  - Shapefile (rivers)
- Output:
  - PostGIS table
- Process:
  - Find the closest river of each station and determine the distance to the river



Spatial Data Integrator project

- Community infrastructure is being set up.
- Register your interest to be informed:
  - www.camptocamp.com/sdi

camptocamp

Contacts

- Camptocamp:
  - David Jonglez, david.jonglez@camptocamp.com
  - Claude Philipona, claude.philipona@camptocamp
  - www.camptocamp.com/sdi