



Advanced Analytics for Extremely Large European Databases

26th November 2012

Version: 1.0

Kathryn Woodcock
kathryn@2ndquadrant.com
2ndQuadrant Limited

D1.10) Project Fact Sheet: Provision of a brief project Fact Sheet suitable for Web publishing, within one month from the start of the project. The Fact Sheet will outline the project's rationale and objectives, specify its technical baseline and intended target groups and application domains, and detail intermediate and final outputs. The Fact Sheet will be used by the Commission for its own dissemination and awareness activities throughout the project lifecycle, and will be published on EC and EC sponsored websites. The factsheet has to be maintained and updated until the end of the project; this will be documented in the regular reporting [month 1]

Dissemination level: Public (PU)

Rationale

AXLE focuses on automatic scaling of complex analytics, while addressing the full requirements of real data sets. Real data sources have many difficult characteristics. Sources often start small and can grow extremely large as business/initiatives succeed, so the ability to grow seamlessly and automatically is at least as important as managing large data volumes once you know you have them. Extremely large data stores have added concerns such as data quality, privacy, security and auditability.

Aspects of the project will include

- * Scalability Engineering - Autopartitioning, Compression
- * Security, Privacy and Audit Techniques
- * Visual Analytics
- * Advanced Architectures for Hardware and Software

Validation will be carried out by industrial consortium partners with access to large volumes of private medical data, as well as other wide ranging data from other interested parties.

Software features will be released as commercially-usable open source code, and submitted for wide use as core parts of the PostgreSQL database or Orange project, or pluggable extensions for those tools.

AXLE brings together a diverse group of researchers covering hardware, database kernel and visualisation experts all focused on solving the needs of extremely large data analysis.

Technical Baseline

The consortium includes top research and system integration organisations with nonoverlapping skills in the areas of computer architecture, databases, reconfigurable systems, runtime environments, programming models and benchmark design.

The academic partners are hardware and compilation/runtime experts. In addition, they are experts in accelerators and multi- and many-cores as well as reconfigurable computing. Therefore, they are in a position to provide the industry partners (who are experts in databases) with the necessary knowledge tools to develop database engines for future architectures, as well as for cutting edge multi- and many-core processors of today.

The project will include

- new features or extensions to PostgreSQL including
 - Row Level Security

- Optional auditing features for database objects, and for both read and write access.
- Autopartitioning
- Bitmap indexing
- Scalability Engineering: Dynamic Compilation and Optimization SQL Searches
Research into how dynamic compilation to execute code can be used to take advantage of, and to optimize the intra query parallelism.
- Advanced Architectures: how Advanced Hardware Architectures can accelerate and facilitate the design of Database Management Systems. The contributions of this WP will include advanced micro-architectural support, an evaluation of executing DBMS on top of several memory systems, and the acceleration of DBMS with specialized accelerator units.
- Visual Analytics: exploring the techniques for visual analysis of large volume data and working prototypes of a practical implementation.

Intended Target Groups and Domain

The AXLE project aims to greatly improve the speed and quality of decision making on real-world data sets and to make those improvements generally available through high quality open source implementations via the PostgreSQL and Orange products.

AXLE targets databases which are:

- Important
- Highly secure
- Complex
- Standardised/Widely used
- Extremely large

The improvements will focus on functionality and performance for use in business intelligence applications on very large datastores, especially with the proviso that transforming and re-formatting data into a data warehouse is not a viable option at very large data volumes.

Expected Results

AXLE will deliver:

- Advanced analytical hardware/software techniques that show significant measurable improvements in database processing speed over existing techniques when applied to extremely large and realistic data volumes.
- Advanced techniques for addressing the scalability challenge of

extremely large datasets, specifically the ability for many common queries to return in the same time no matter how large the data by using flexible proof-based approaches to query handling.

- Visual analytics capable of exploring extremely large data volumes without significant loss of speed or functionality as data volumes grow.
- A capability to measure and evaluate performance against extremely large volumes of real data, with a mechanism for more easily publishing and comparing results.
- More scalable data management with integrated security controls.
- High security database software capable of securing and auditing data in its application context, as well as pass external assessment as being suitable for Common Criteria for IT Security Evaluation. <http://www.commoncriteriaportal.org/>

At a glance

Project title: AXLE - Advanced Analytics for EXtremely Large European Databases

Project coordinator: Simon Riggs, 2ndQuadrant Limited (UK)

Partners: 2ndQuadrant Limited (UK), Barcelona Super Computers (ES), Portavita B.V. (NL), The University of Manchester (UK), The University of Ljubljana (SI)

Duration: 1st November 2012 - 30 October 2015

EU contribution: EUR2.9M

Further information: <http://axleproject.eu>