



CoherentPaaS

Coherent and Rich PaaS with a
Common Programming Model

ICT FP7-611068

Scalable SQL and SQL-like data stores

D6.1 Cloud SQL-like data stores extensions design

March 2014

Document Information

Scheduled delivery	31.03.2014
Actual delivery	06.05.2014
Version	1.2
Responsible Partner	MonetDB Solutions

Dissemination Level: PU

PU	Public
PP	Restricted to other programme participants (including the Commission)
RE	Restricted to a group specified by the consortium (including the Commission)
CO	Confidential, only for members of the consortium (including the Commission)

Revision History

Date	Editor	Status	Version	Changes
DD.MM.YYYY	<Name>	Draft, Final, Revised	0.1	<Short description of current version>
23.01.2014	Kersten	Draft	0.1	Sketch of structure
28.01.2014	Ricardo Vilaça Kersten	Draft	0.2	Add Derby initial system extensions. Input from Aveiro meeting
04.02.2014	Ricardo Vilaça	Draft	0.3	Add Derby overview and synopsis and extend system extensions.
25.02.2014	Kersten	Draft	0.4	Merged and updated
04.03.2014	Kersten	Draft	0.5	Sweep after QuartetFS
16.03.2014	Zhang	Revised	0.9	Incorporate reviews from INRIA and FORTH
21.03.2014	Nedev, Kersten, Zhang	Revised	1.0	Final internal review
07.04.2014	Zoupanos, Kersten, Zhang	Revised	1.1	Incorporated second round reviews.
06.05.2014	Zhang	Revised	1.2	Rename MdbQL into CloudMdsQL (cloud multi-datastore query language).

Contributors

Martin Kersten, Ying Zhang, Sjoerd Mullender, Dimitar Nedev, Martin van Dinther (MonetDB Solutions)
 Rui Oliveira (INESC TEC), Ricardo Vilaça (INESC TEC)
 Antoine Chambille, Romain Colle, Spyros Zoupanos (QuartetFS)

Internal Reviewers

INRIA, FORTH

Acknowledgements

Research partially funded by EC 7th Framework Programme FP7/2007-2013 under grant agreement n° 611068.

More information

Additional information and public deliverables of CoherentPaaS can be found at:

<http://coherentpaas.eu>

Executive Summary

This document describes the current state-of-the-art of the three SQL-like data stores, namely, a column-oriented data store, an in-memory analytical database, and an SQL engine, in the context of providing database management services within the CoherentPaaS project.

Subsequently, we provide a short overview of the CoherentPaaS system architecture as it pertains to the SQL storage engines' functionality. For each SQL-like data store we sketch the changes needed to satisfy the CoherentPaaS ultimate goal to handle large-scale distributed and heterogeneous database queries and transactions. Therefore each SQL-like data store will:

- Provide a concise technical description of the data store functionality and implementation.
- Identify and propose the changes required to adapt it to CoherentPaaS Holistic transaction processing model.
- Identify and propose the changes required to adapt it to handle the calls from the CoherentPaaS query processor.

Finally, a tentative roadmap for each relational store is given, which can be used to align the integration activities within the project.

Background

In CoherentPaaS we bridge three data modelling domains: SQL-based systems, NoSQL based systems, and graph-based systems. In WP6 we focus on the SQL-based approach and extend it towards the scalable Cloud processing. Both scale-up, i.e. more powerful processor nodes, and scale-out, i.e. larger clusters, are considered from the perspective of the individual product offerings.

The ambition of CoherentPaaS is to consolidate transaction processing and distributed heterogeneous query processing within a single software layer. WP6 will produce a set of SQL Cloud data stores to be integrated into the CoherentPaaS framework. These data stores will include: MonetDB, ActivePivot, and DQE:

MonetDB is the leading open-source SQL-based column-oriented data management system. It provides for the full set of programming interfaces. The system is optimized for multi-core processing, cache-conscious algorithms and read-dominant environments. Its query execution model is based on materialise-all-intermediates and its transaction processing based on lazy multi-version optimistic concurrency-controls (MVCC) scheme.

ActivePivot is a real-time, in-memory analytics platform, which provides instant insight into complex and dynamic data for timely business decisions. Going beyond traditional BI or OLAP solutions, which are bound by latency, ActivePivot fits in operational systems and provides real-time business intelligence for immediate decision making, when and where it is needed.

DQE is a scalable SQL engine developed in the context of the EU FP7 CumuloNimbo project and is a modified version of Apache Derby. DQE relies on HBase data store as its storage layer and transforms SQL statements into query plans that use specialised scan operators for HBase. Each instance preserves transactional contexts and coherence through the interface to the holistic transactional framework. Multiple instances of DQE can be used for both high availability and scalability, to cope with a workload composed by a large number of concurrent transactions.

CoherentPaaS requirements

The three SQL-based systems cover a wide-spectrum of modern database technology, ranging from pure main-memory (ActivePivot), to virtual memory processing (MonetDB), into distributed memory processing (DQE). Multi-core optimisations and compression are dealt with in ActivePivot and MonetDB, while DQE is focused on fast distributed scan processing. The underlying implementation use both C (MonetDB) and Java (ActivePivot, DQE). They also address a different part of the market, ranging from stream-based OLAP (ActivePivot) to data warehousing (MonetDB, DQE).

The CoherentPaaS project targets for all three systems are:

- Supporting the high-volume distributed transaction processing scheme, originally developed in the context of the CumuloNimbo project.
- Providing a Python/Java API library for the middleware query execution engine.
- Providing a Cloud version for access and experimentation.

The CoherentPaaS project gives an additional edge on the competition in the following exploitation directions:

- MonetDB expands into a Cloud service offering provisioning, JSON-store functionality, Python UDF-integration.
- ActivePivot expands its capabilities by supporting multi-version concurrency control and JSON-store functionalities.
- DQE expands its capabilities by supporting the CoherentPaaS Holistic Transaction Management and the external tables required by CloudMdsQL (cloud multi-datastore query language).