



CoherentPaaS

Coherent and Rich PaaS with a
Common Programming Model

ICT FP7-611068

Query Engine for the Common Query Language v1 (basic functionality)

D3.3

September, 2015

Document Information

Scheduled delivery	31.07.2015
Actual delivery	17.09.2015
Version	0.8
Responsible Partner	Sparsity

Dissemination Level:

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
17.07.2015	R.Pau	Draft	0.1	Skeleton
03.09.2015	J. Orlando	Draft	0.2	Wrappers section added
08.09.2015	B.Kolev	Reviewed	0.3	Grammatical and orthographic corrections
15.09.2015	F. Savary	Reviewed	1.0	-

Contributors

Boyan Kolev (INRIA), Raquel Pau (Sparsity), Ricard Gavaldà (UPC), Marta Perez (UPC), Josep Lluís Llorca, (Sparsity), Jose Orlando Pereira (INESC)

Internal Reviewers

Francois Savary (Active Pivot), Boyan Kolev (INRIA)

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>

1. Executive Summary

Providing massive data processing capabilities in the cloud is a major trend in the design of data management solutions deployed on the cloud. The experience of the latest years is that no single data management system is the silver bullet for data processing, where all the data needs can be mapped. Currently, companies are using a variety of data solutions ranging from relational databases to NoSQL data stores, which come in multiple flavors such as graph databases, key-value data stores, array data stores, analytical cloud frameworks, document databases, data stream systems, etc. CoherentPaaS provides a software infrastructure to allow efficient and easy to program communication among this multitude of data management systems. In this deliverable, we describe the internal design, the build and the execution process of the query engine that interconnects the data stores.

The query engine is central piece that coordinates the execution of queries in CoherentPaaS. This module executes queries in the CloudMdsQL language (see D3.1), which defines a syntax to mix the operations among the data repositories. The data in the query engine is modeled as tabular data: sets of tuples with a fixed number of attributes. This model is simple enough to allow importation and exportation of data from NoSQL data representations.

Clients connecting to the CoherentPaaS infrastructure will have the impression that all systems act as a single database. In order to provide such a feeling from the client perspective, we will provide a JDBC connector for the CoherentPaaS infrastructure of CoherentPaaS. JDBC is one of the standard methods to connect to a database system, and thus clients will connect to the database like a regular database system. However, user queries will be able to take full advantage of the different repositories in the CoherentPaaS infrastructure by introducing SQL and NoSQL statements.

The aim of this document is providing the implementation details of the first prototype of the Query Engine, the main responsible to orchestrate the query resolution between the involved datastores.