CoherentPaaS

Coherent and Rich PaaS with a
Common Programming Model

ICT FP7-611068

Query Compiler for the Common Query v1 (Basic Functionality)

D3.5

March, 2015

Document Information

Scheduled delivery	31.03.2015
Actual delivery	31.03.2015
Version	1.1
Responsible Partner	INRIA

Dissemination Level:

PU   Public

Revision History

Date	Editor	Status	Version	Changes
17.03.2015	Kolev	Draft	0.1	Initial draft
25.03.2015	Kolev	Revised	1.0	Revision after MonetDB review
28.03.2015	Kolev	Revised	1.1	Revision after UPM review

Contributors

Boyan Kolev, Patrick Valduriez

Internal Reviewers

Marta Patiño, Ricardo Jiménez UPM,
Jennie Zhang MonetDB

Acknowledgements

More information

Additional information and public deliverables of CoherentPaaS can be found at: http://coherentpaas.eu

1. Executive Summary

CloudMdsQL (Cloud Multidatastore Query Language) is a functional SQL-like language, capable of querying multiple cloud data stores (SQL, NoSQL, HDFS, etc.) within a single query that contains embedded invocations to each data store's native query interface. The query compiler parses a CloudMdsQL query and generates a query execution plan to be processed by the query operator engine.

In its current version, the compiler supports the full grammar for CloudMdsQL SELECT queries. It parses a query first into an abstract syntax tree (AST) kept in C++ data structures. Then the compiler transforms the AST into a query execution plan (QEP), represented as a directed acyclic graph, where leaf nodes are references to named tables and all other nodes represent relational algebra operations. Each named table is mapped to a sub-plan, which is represented as a relational algebra tree, where the leaf nodes may be references to other named tables (for nested SQL sub-queries), references to data store tables (for named table expressions against SQL data stores), or native/Python expression definitions (for native/Python named table expressions).