CoherentPaaS

CoherentPaaS aims at providing a full Atomicity, Consistency, Isolation and Durability (ACID) coherent, scalable and efficient environment integrating NoSQL, SQL and Complex Event Processing (CEP) data management technologies, while allowing application developers to program using a simple and powerful programming model and query language.

Why CoherentPaaS?

The increasing demand for efficiency has resulted in a wide diversity of different cloud data stores, each one specialised and optimal for specific applications, thus leading to a “no one size fits all” situation. The consequence refers to large proliferation of APIs, lack of a common programming framework and lack of coherence across different cloud data management technologies (in the cloud landscape ACID coherence has been totally lost). Performance and efficiency issues are also pertinent, being of major importance for Big Data analytics where emphasis is put on large queries over big cloud data stores. Furthermore, emerging Internet of Things applications highlight specific challenges with respect to data management, both in real-time (requiring enhanced CEP techniques) and offline (requiring application development on top of diverse resources types and data stores).

AT A GLANCE

Project title: Coherent and Rich PaaS with a Common Programming Model

Project number: 611068

Project coordinator: Ricardo Jimenez-Peris, Universidad Politecnica de Madrid (ES)
riimenez@fi.upm.es

Partners: Universidad Politecnica de Madrid (ES), Neurocom SA (GR), INRIA (FR), Foundation for Research and Technology - Hellas (GR), Institute of Engineering Systems and Computers (PT), Sparsity (ES), MonetDB (NL), QuartetFS (UK), Portugal Telecom (PT), Institute of Communication and Computer Systems (GR)


Total cost: € 6.53 M

Programme: FP7 ICT Call 10

Website: http://www.coherentpaas.eu
What makes CoherentPaaS unique?

CoherentPaaS will provide a rich PaaS with different “one size” data stores optimised for particular tasks, data, and workloads. CoherentPaaS will integrate NoSQL data stores such as key-value data stores and graph databases, SQL data stores such as in-memory and column-oriented databases, hybrid systems such as SQL engines on top of key-value data stores, and complex event processing data management systems.

What is more, CoherentPaaS will provide a common query language for all cloud services and data stores. This common query language provides a uniform way to access all data stores in an integrated manner. Its federated approach also enables to fully exploit efficiently the capabilities of each cloud data store. CoherentPaaS will enable the development of new cloud applications that exploit the performance and scalability of new cloud data management technology, while hiding the complexity of the underlying technology under a unified query language.

Additionally, a scalable transactional management system will provide holistic coherence across data stores. CoherentPaaS will design and implement within each data store the necessary functionality to enable the integration with the new transactional technology in a scalable way, providing a common transactional API to enable the holistic transactional management in order to orchestrate transactions across data management technologies transparently to the applications.

Value proposition

CoherentPaaS will enable the development of cloud applications that exploit the performance, coherence and scalability of an environment providing an integrated cloud data management technology; while hiding the complexity and diversity of the underlying technology under a simple and powerful programming model and query language. Thus, the added value of CoherentPaaS is summarised as follows:

- **Simplicity:** Programming with a single query language.
- **Scalability:** Across all data management technologies.
- **Efficiency:** Avoiding copying and translating TBs of data from one data store to another (i.e. ETL jobs).
- **Coherence:** Transactional semantics across data stores.

Demonstrators

The CoherentPaaS outcomes will be validated through four different application scenarios: Cloud Machine-to-Machine for rating, invoicing and analysis of M2M events, and fraud identification and management; Media Planning for analysis of social media, news or blogging data; Real-Time Network Performance Analysis for collection and real-time analysis of information in a telco environment; and Bibliographic Search for exploiting information from various data sources (e.g. Cordis, Wipo, DBLP, PubMed, ArXiv, Twitter, DBpedia, Yago).