



TYPHON will provide a methodology and an integrated technical solution for designing, developing, querying and evolving scalable architectures for persistence, analytics and monitoring of very large volumes of hybrid data, including relational, graph-based, document-based, natural language, and others, which will meet the growing scalability and heterogeneity requirements encountered by industry when developing Big Data applications within multiple European market sectors.

AT A GLANCE

Project Title

Polyglot and Hybrid Persistence Architectures for Big Data Analytics

Project Coordinator

The Open Group (UK)

Partners

Alpha Bank (GR)
 ATB (DE)
 Edge Hill University (UK)
 CLMS (GR)
 CWI (NL)
 GMV (ES)
 OTE (GR)
 SWAT Engineering (NL)
 University of L'Aquila (IT)
 University of Namur (BE)
 University of York (UK)
 Volkswagen (DE)

Duration

01.2018 - 12.2020

Total Cost

4.499.447 €

EC Contribution

4.499.447 €

Programme

H2020-ICT-2017-1

Further Information

www.typhon-project.org

Context and Motivation

Organisations are faced with the challenge of managing ever-growing volumes of data, which can vary significantly in terms of consistency and availability requirements. For example, e-commerce systems using data to provide recommendations for products need to be highly available as data is constantly retrieved and updated as users browse the system. Consistency of such data is not critical so a small loss of part of the data can be reasonably traded for a significant improvement in availability. On the other hand, for other subsets of data in the same system, such as recording customer orders and payments, compromising data consistency to improve availability is not acceptable.

Relational databases were once considered the *de facto* technology for persisting and managing large volumes of data. This has changed recently with the emergence of Google, Twitter, Facebook, Amazon and others, which were faced with extremely large data sets and unprecedentedly high availability requirements. The challenges involved in scaling such databases has led to the emergence of a new generation of purpose-specific databases grouped under the term NoSQL, which are designed with horizontal scalability as a primary concern, and that deliver increased availability and fault tolerance at a cost of having temporary inconsistency and reduced durability of data.

Challenge

To balance requirements for data consistency and availability, organisations increasingly are migrating towards hybrid data persistence architectures comprising both relational and NoSQL databases for managing different subsets of their data using ad-hoc architectures. At the same time, as the volume and the value of textual content constantly grows, built-in support for sophisticated text processing in data persistence architectures is becoming increasingly essential.

This introduces a number of challenges including ensuring the coherency of the overall design, the assembly and configuration of the different components of the architecture, and the consistency of the overlapping data. Also, in order to access the data, developers need to write application code against different types of persistence backends.

Unlike relational databases, NoSQL databases do not conform to a common set of standards (e.g. SQL, ODBC, JDBC) and application code is specific to the NoSQL database used, making it difficult to migrate. Undisciplined development of such data persistence architectures also introduces data evolution and migration challenges and complicates the development and maintenance of real-time analytics and monitoring capabilities.

Solution

TYPHON will provide an industry validated methodology and integrated technologies for designing, developing, querying, evolving, analysing and monitoring architectures for scalable persistence of hybrid data. The key scientific and technology innovations that will be developed include:

- Technologies and methodology for designing hybrid polystores taking into account the structure of the data, the availability, partitioning and consistency requirements of different subsets of the data and the available deployment resources.
- Novel algorithms for transforming hybrid polystore design models into preconfigured optimised virtual machines which can be deployed on cloud infrastructure.
- An extensible high-level language for querying and modifying data persisted in hybrid polystores, and facilities for translating high-level queries into efficient native queries.
- A high-performance framework for publishing and processing data access



and update events to facilitate real-time monitoring and predictive analytics.

- Technologies and methodology for evolving the organisation and distribution of data in hybrid polystores, along with tools for monitoring use of polystores for more optimised evolution.

The TYPHON technologies will be validated through four industrial Big Data applications that will involve datasets with different volume, variety and velocity characteristics from key European sectors of automotive, aerospace, banking, and telecom.

Expected Impact

The key industrial impacts targeted by the TYPHON project technologies and innovations are:

- Powerful Big Data processing tools and methods for hybrid data and demonstrations of their applicability in real-world settings.
- Significant increase in the speed of data throughput and access for hybrid data architectures that have been measured against industry validated benchmarks.
- Substantial increase in the definition and uptake of standards fostering data sharing, exchange and interoperability for hybrid data architectures.

TYPHON innovations are driven by Big Data requirements from four industrial user partners with Big Data applications for Smart Connected Vehicles, Earth Observation Data Management, Hybrid Bank Data Warehousing and Telecom Maintenance & Quality of Service Data, who will validate the expected project impacts are fulfilled using hybrid data architectures and industrial persistence technologies.