

CumuloNimbo: A Highly-Scalable Transaction Processing Platform as a Service

by Ricardo Jimenez-Peris, Marta Patiño-Martinez, Kostas Magoutis, Angelos Bilas and Ivan Brondino

One of the main challenges facing next generation Cloud platform services is the need to simultaneously achieve ease of programming, consistency, and high scalability. Big Data applications have so far focused on batch processing. The next step for Big Data is to move to the online world. This shift will raise the requirements for transactional guarantees. CumuloNimbo is a new EC-funded project led by Universidad Politécnica de Madrid (UPM) that addresses these issues via a highly scalable multi-tier transactional platform as a service (PaaS) that bridges the gap between OLTP and Big Data applications.

CumuloNimbo aims at architecting and developing an ultra-scalable transactional Cloud platform as a service (PaaS). The current state of the art in transactional PaaS is to scale by resorting to sharding or horizontal partitioning of data across database servers, sacrificing consistency and ease of programming. Sharding destroys transactional semantics since it is applied to only subsets of the overall data set. Additionally, it forces modifications to applications and/or requires rebuilding them from scratch, and in most cases also changing the business rules to adapt to the shortcomings of current technologies. Thus it becomes imperative to address these issues by providing an easily programmable platform with the same consistency levels as current service-oriented platforms.

The CumuloNimbo PaaS addresses these challenges by providing support

for familiar programming interfaces such as Java Enterprise Edition (EE), SQL, as well as No SQL data stores, ensuring seamless portability across a wide range of application domains. Simultaneously the platform is designed to support Internet-scale Cloud services (hundreds of nodes providing service to millions of clients) in terms of both data processing and storage capacity. These challenges require careful consideration of architectural issues at multiple tiers, from the application and transactional model all the way to scalable communication and storage.

CumuloNimbo improves the scalability of transactional systems, enabling them to process update transaction rates in the range of one million update transactions per second in a fully transparent way. This transparency is both syntactic and semantic. Syntactic transparency

means that existing applications will be able to run totally unmodified on top of CumuloNimbo and benefit automatically from the underlying ultra-scalability, elasticity and high availability. Semantic transparency means that applications will continue to work exactly as they did on centralized infrastructure, with exactly the same semantics and preserving the same coherence they had. The full transparency will remove one of the most important obstacles to migration of applications to the cloud, ie the need to heavily modify, or even fully rebuild, them.

CumuloNimbo adopts a novel approach for providing SQL processing. Its main breakthrough lies in the scalability of transactional management, which is achieved by decomposing the different functions required for transactional processing and scaling each of them separately in a composable manner (refer to

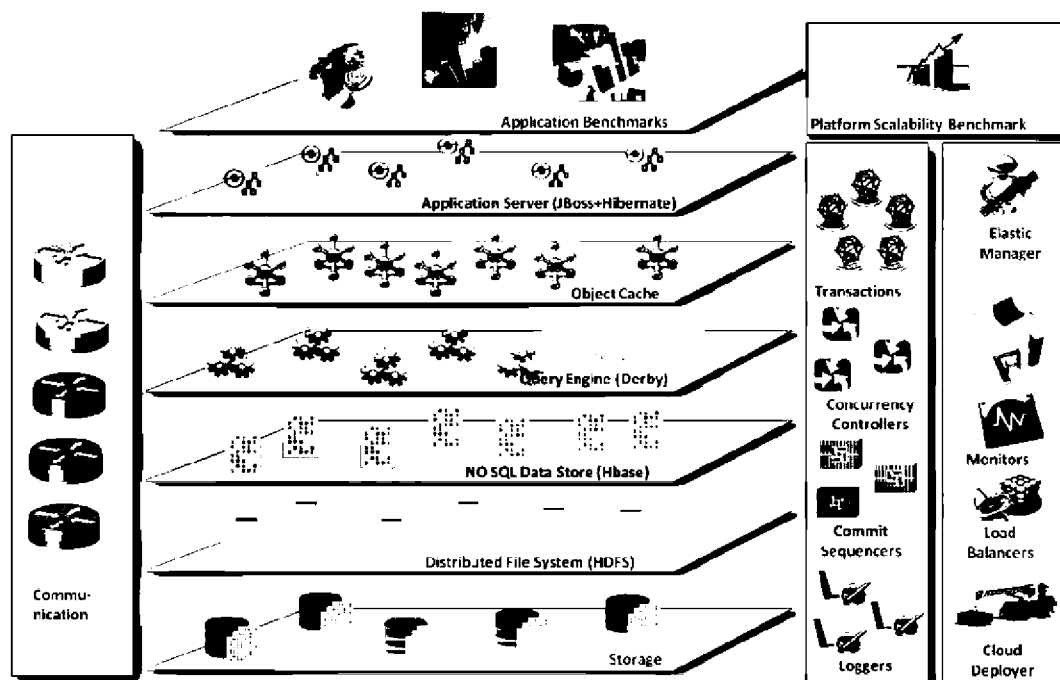


Figure 1: CumuloNimbo architecture

Figure 1). In contrast to many of the current approaches that constrain the query language to simple key-value stores, CumuloNimbo provides full SQL support based on the snapshot isolation transaction model and scaling to large update transaction rates. The SQL engines use a No-SQL data store (Apache HBase) as an underlying storage layer, leveraging support for scalable data access and version management. The project is optimizing this data store to operate over indexed block-level storage volumes in direct-attached or network-accessible storage devices.

Currently, the project has completed the specification of the architecture and the development of a first version of the core components, which have been successfully integrated. CumuloNimbo is expected to have a very high impact by enabling scalability of transaction processing over Cloud infrastructures without changes for OLTP and bridge the gap between Big-Data applications and OLTP. The project is carried out by Universidad Politécnica de Madrid (UPM), Foundation for Research and Technology – Hellas (FORTH), Yahoo Iberia, University of Minho, McGill University, SAP, and Flexiant. The

CumuloNimbo project is part of the portfolio of the Software & Service Architectures and Infrastructures Unit – D3, Directorate General Information Society (<http://cordis.europa.eu/fp7/ict/ssai>).

Link:

<http://cumulonimbo.eu>

Please contact:

Ricardo Jimenez-Peris
Universidad Politécnica de Madrid,
Spain
Tel: +34 656 68 29 48
E-mail: rjimenez@fi.upm.es