Highly Scalable Real-Time Analytics with CloudDBAppliance

Boyan Kolev, Oleksandra Levchenko, Florent Masseglia, Reza Akbarina, Esther Pacitti, Patrick Valduriez
(INRIA, France)

Work partially funded by the EU’s Horizon 2020 programme, grant agreement No. 732051
Motivation

• The cloud today
  • Cloud data infrastructures fail to provide:
    • Predictable performance
    • Support for high loads / strict SLAs

• Consequence
  • Data critical applications still use on-premise mainframe architectures instead of moving to the cloud

• The solution
  • Cloud appliance for providing database-as-a-service with predictable performance, robustness and reliability

Work partially funded by the EU’s Horizon 2020 programme, grant agreement No. 732051
Objectives

• **Innovations**
  - Powerful hardware enabling In-Memory databases
    - 32TB RAM
    - 1000+ CPU cores
  - Vertically scalable in-memory operational database
  - Vertically scalable in-memory analytics
  - Vertically scalable real-time streaming analytics
  - Operational Hadoop data lake

• **Characteristics**
  - Predictable performance
  - High availability

Work partially funded by the EU’s Horizon 2020 programme, grant agreement No. 732051
High-level Architecture

Work partially funded by the EU’s Horizon 2020 programme, grant agreement No. 732051
Real-time Streaming Analytics

- **Ultra scalable streaming engine**
  - Linear scale-up on many core (1000+) architectures
  - Algebraic and custom operators to incorporate data mining and machine learning tasks

- **Time series correlation mining approach**
  - Fast online discovery of correlations over sliding windows of time series data
  - Massively parallelizable approach
    - High scalability
  - Incremental algorithm
    - Near real-time response
  - Utilizes in-memory storage
    - Sharing intermediate data across streaming operators

Work partially funded by the EU’s Horizon 2020 programme, grant agreement No. 732051
CloudDBAppliance Use Cases

• Validated through five real industrial use application scenarios in three sectors

  • Finance/Banking
    • Real-time risk analysis
    • ATM optimization

  • Telco
    • Cell phone number portability

  • Retail
    • Proximity marketing
    • Real-time pricing

Work partially funded by the EU’s Horizon 2020 programme, grant agreement No. 732051
Highly Scalable Real-Time Analytics with CloudDBAppliance

Thank you!

Work partially funded by the EU’s Horizon 2020 programme, grant agreement No. 732051