FinTech and InsuranceTech case studies digitally transforming Europe’s future with BigData & AI driven innovation

INFINITECH

Pavlos Kranas, LeanXcale Spain, pavlos@leanxcale.com
Data: Data Movements Today.

• SQL data management technologies are targeted either for operations (operational databases) or for analytical purposes (data warehouses and data lakes).

• The weaknesses of SQL have resulted in the proliferation of NoSQL solutions for dealing with specific data management problems not handled well by SQL technologies.

• Data silos appear due to the usage of different data managers (operational vs analytical, SQL vs NoSQL) that prevent data from being queried across them.

• These data silos force to do ETLs, i.e., move data, from operational databases to data warehouses and/or data lakes to blend data together and enable to query them.

• These movements of data are performed on a daily basis.
Data: Avoidance of data movements by INFINITECH

• HTAP database: Infintechn is extending LeanXcale database with HTAP capabilities.
• HTAP (Hybrid Transactional Analytical Processing) lies in being able to handle operational data (i.e. support updates efficiently, support data coherence through ACID transactions) and answer analytical queries in short time.
• LeanXcale is being extended with intra-query parallelism (both inter-operator and intra-operator parallelism) to have analytical capabilities and be able to answer analytical queries with short response times.
• LeanXcale internal processing of updates is designed to support HTAP workloads. On one hand, it is able to handle massive data ingestion rates (as fast as key-value NoSQL technologies) and on the other is able to query this fastly ingested data very efficiently (as efficiently as SQL technologies. It does so thanks to a novel algorithm and data structure to process updates and queries.
• LeanXcale HTAP capabilities will make INFINITCH able to handle both an operational and analytical database, thus avoiding to move data between operational and analytical SQL databases.
Data: Avoidance of data movements by INFINITECH

• INFINITECH is also offering with polyglot capabilities.
• These polyglot capabilities enable to query NoSQL data stores such as key-value data stores (e.g., Hbase), document data stores (e.g. MongoDB or CouchBase) and graph databases (e.g. Neo4J) together with SQL data.
• The approach to these polyglot queries is quite novel, instead of forcing to put a schema over schemaless or semi-structured data, it allows to query NoSQL data with their native API or query language.
• These native subqueries materialize their resultsets as temporary SQL tables that are queried by an integration SQL query.
• Thus, it combines the power of the native NoSQL query capabilities in the subqueries with the ease of SQL queries for the integration query.
• Again this polyglot capabilities will avoid moving data across data silos created by the usage of different SQL and NoSQL technologies
Data Curation and Anonymization

- INFINITECH uses the state of art on data curation and anonymization techniques.

- It makes them accessible by the creation of specific testbeds for different areas in the financial and insurance sectors.

- Each testbed chooses the most appropriate algorithms for each task and handle these tasks in a fully automated way.
Workflows:
Approach for mastering the complexity and orchestration

• Infinitech approach to workflows lie in automating them for each specific subdomain with the financial and insurance domains.

• This domain specific customization of the workflows in the testbeds hides their complexity.

• These workflows are automated in custom testbeds for each subdomain including:
  • data cleaning,
  • data curation,
  • data anonymization,
  • enforcement of GDPR,
  • etc.
HPC/Cloud infrastructure to edge

- Infinitech uses HPC systematically for AI/ML tasks.
- Infinitech automates the usage of HPC within domain specific testbeds.
- This domain specific approach enables to customize the usage of HPC for optimal use for each subdomain of finance and insurance.
- Data sharing across organizations is handled through standardization of APIs and use of blockchain that enable to query across different organizations.
- A blockchain approach is used to share data across the organizations, again customized on a per testbed basis to use the optimal technology for each use case.
- Management of data on the edge is fulfilled by usage of data streaming technology that manages data locally and sends relevant data to a cloud database.