REPLICATION TECHNOLOGIES

L. LOBATO PARDAVILA

WHICH ARE OUR CHALLENGES?

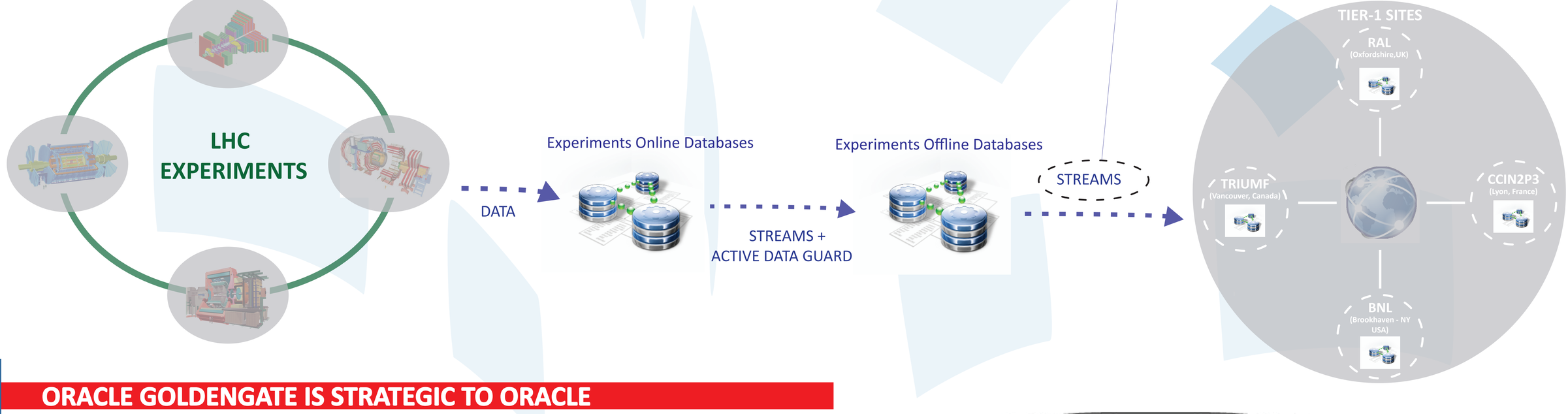
The mission of the Worldwide LHC Computing Grid Project (WLCG) is to develop, build and maintain a computing infrastructure for storage and analysis of the LHC data. Data will be distributed around the globe according to a four -tiered model. A primary backup will be recorded at CERN, the "Tier-0" centre of WLCG. After initial processing, these data will be distributed to a series of Tier-1 centers, large computer centers with sufficient compute and storage capacities as well as round-the-clock support for Grid operations.The Tier-1 centers will perform recurrent reconstruction passes and make data available to Tier-2 centers, each consisting of one or several collaborating computing facilities, which can store sufficient data and provide adequate computing power for specific analysis tasks. Individual scientists will access these facilities through Tier-3 computing resources. From CERN, the data are distributed to four sites around the world (part of Tier-1 sites) enabling a highly complex replication environment where the ongoing maintenance presents a variety of challenges., like real data replication, distributed monitoring or real time data consistency validation.

WHAT DO WE REPLICATE ?

- PVSS (Supervisory Control and Data Acquisition)
  - > Data from hw (or sw) devices in order to use it for their controls (DDL and DML operations)
  - > 4TB of data, 81% of source db, average workload : 694 LCRs/s

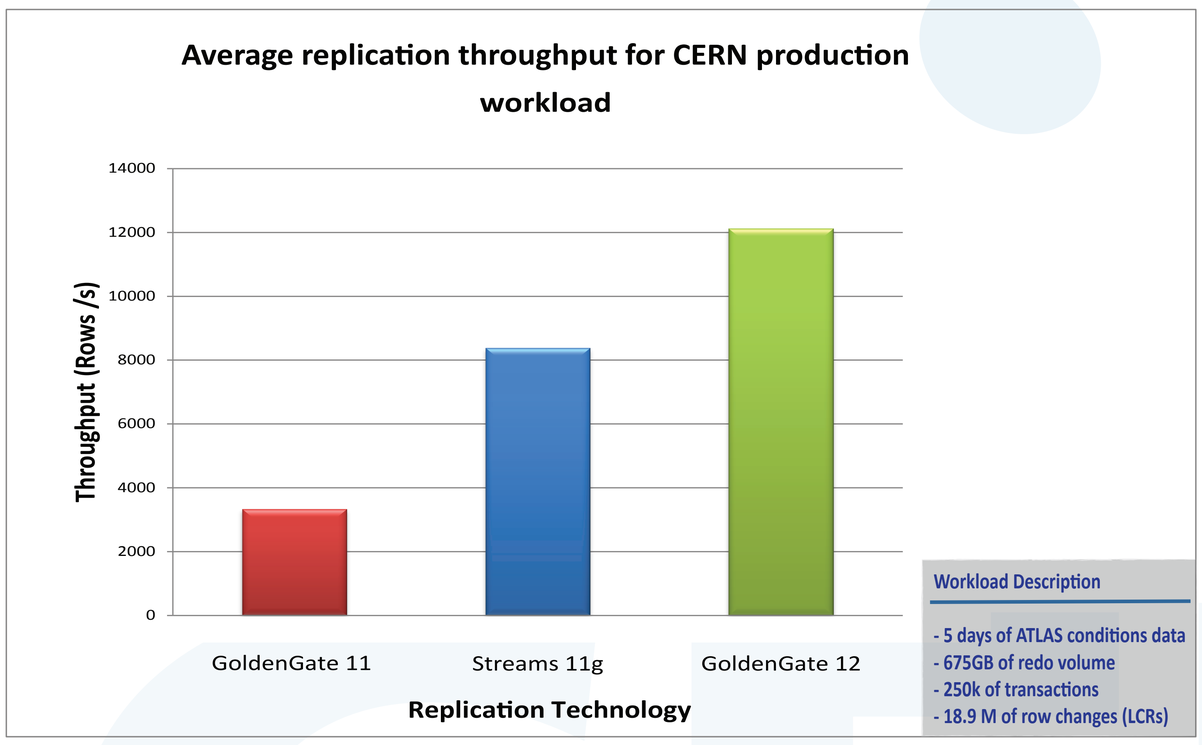
- Experiments CONDITIONS data
  - > Record the state of the detector: calibration, alignment, environmental parameters, … (DDL and DML operations)
  - > 900 GB of data, 8% of source db, avg workload 50 LCRs/s

- Other:
  - > Muon calibration data (DML & DDL); 72 GB
  - > ATLAS Metadata Interface (DML & DDL); 80 GB



ORACLE GOLDENGATE IS STRATEGIC TO ORACLE

Within the CERN openlab collaboration with ORACLE, extensive research has been performed in order to improve data replication solutions. With introduction of a new technology called GoldenGate, Oracle provided solution to CERN replication requirements.
Advantages of this approach include straightforward and out-of-the-box configuration, extended data type support, improved scalability and tighter integration with Oracle database software. Additionaly simplified management and in-database monitoring tools are provided to build a scalable solution fit for CERN needs.

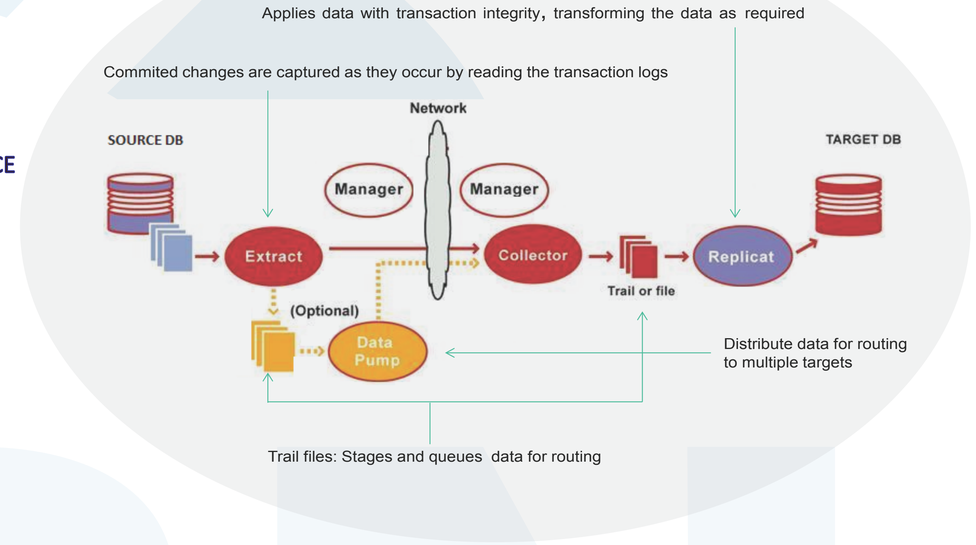


Workload Description

- 5 days of ATLAS conditions data
- 675GB of redo volume
- 250k of transactions
- 18.9 M of row changes (LCRs)

EVOLUTION PERFORMANCE at CERN

Architecture

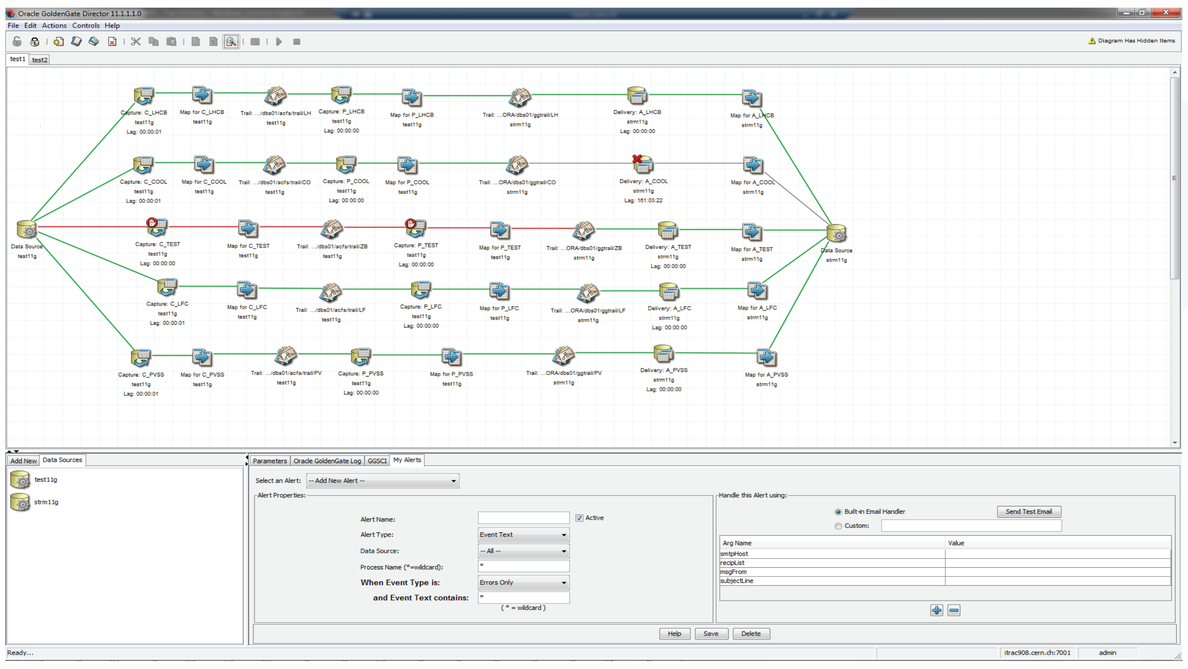


GoldenGate 12c combines the best of Streams technology and the best of the previous versions, fitting perfectly with CERN cases. It is a powerful tool to replicate data from one database to one or more databases while achieving high performance with accuracy. OGG can map, filter and transform transactional data changes between the same or different Oracle versions and even covers Oracle to Oracle as well as to hetereogeneous environments. Also OGG supports different configurations : Uni-DIrectonial, Bi-Directional, Peer to Peer, Broadcast, Consolidation and Cascading and due to support asynchronous replication, GoldenGate maintains availability and has not negative impact on source database from replicas downtime. In the last year, a dedicated testbed configuration was established and since then, we have been working on new tests and preformances in order to be more robust against user errors. Getting good results, we are working on the migration between Streams and GoldenGate for the replicats already existing to Tir 1s, achieving successfully this goal no longer than 2015.

MONITORING

ORACLE GoldenGate Director

Oracle GoldenGate Director is a multi tier client server application that enables the configuration and management of the GoldenGate environment from a remote client which includes a web browser based client.



CONSISTENCY VALIDATION

ORACLE GoldenGate Veridata

Oracle GoldenGate Veridata is a high-speed data-comparison solution that identifies and reports on data discrepancies between heterogeneous databases without interrupting ongoing business processes.