Contribution ID: **57**   Type: **presentation**

Scaling the EOS namespace

Tuesday 10 July 2018 15:00 (15 minutes)

The EOS namespace has outgrown its legacy in-memory implementation, presenting the need for an alternative solution. In response to this need we developed QuarkDB, a highly-available datastore capable of serving as the metadata backend for EOS. Even though the datastore was tailored to the needs of the namespace, its capabilities are generic.

We will present the overall system design, and our efforts in providing comparable performance with the in-memory approach, both when reading, through the use of extensive caching on the MGM, and when writing through the use of latency-hiding techniques involving a persistent, back-pressured local queue for batching updates to the QuarkDB backend.

We will also discuss the architectural decisions taken when designing our datastore, including the choice of consensus algorithm to maintain strong consistency between identical replicas (raft), the choice of underlying storage backend (rocksdb) and communication protocol (redis serialization protocol - RESP), as well as the overall testing strategy to ensure correctness and stability of this important infrastructure component.

Authors: SINDRILARU, Elvin Alin (CERN); BITZES, Georgios (CERN); PETERS, Andreas Joachim (CERN); MANZI, Andrea (CERN)

Presenter: MANZI, Andrea (CERN)

Session Classification: T4 - Data handling

Track Classification: Track 4 - Data Handling