

Session 3 - Scaling Cloud to Big Data

Dirk Duellmann, CERN

XLDB 2017: 10th Extremely Large Databases Conference
11 October 2017 | Clermont-Ferrand, France

One decade XLDB

- Since the first meeting at SLAC
 - formula still works
 - DB academics +
Data intensive science +
Data intensive industry

One decade XLDB

- Since the first meeting at SLAC
- formula still works
- DB academics +
Data intensive science +
Data intensive industry



also one decade...

One decade of data access “cloud style”

- 10 years since amazon dynamo paper
 - Scaling can be more important than single cluster with strong consistency
 - CAP theorem, KV-stores, Hadoop and friends
 - Scaling stable deployment can be more important than ultimate performance
- Good match for data analysis (stable, read-only)
 - additional challenge for other applications
- Cloud style data access has fully arrived in science infrastructures

3 different aspects

- Dan van der Steer, CERN
 - Building Scale-Out Storage Infrastructures with RADOS and Ceph
- Michael Mueller, Container Solutions
 - Cloud Native Databases
- Mike Hildreth, University of Notre Dame
 - Data, Software, and Knowledge Preservation of Scientific Results