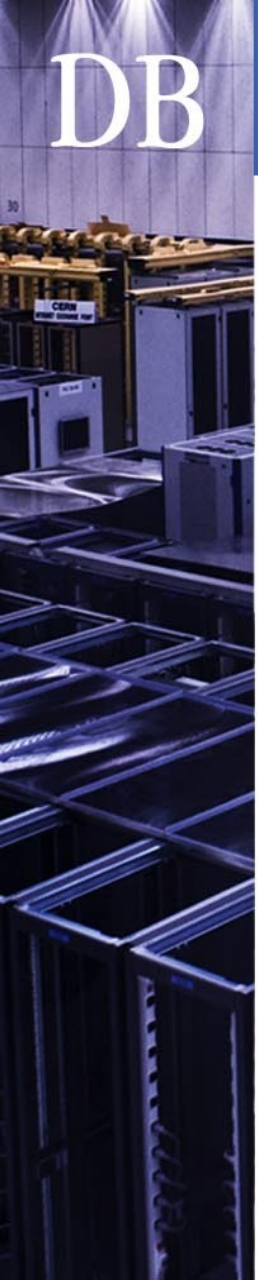
XLDB 2010 (Extremely Large Databases) conference summary

Dawid Wójcik

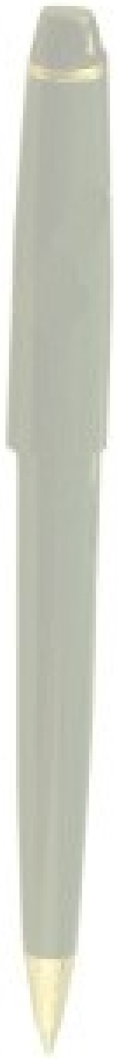
About XLDB conference

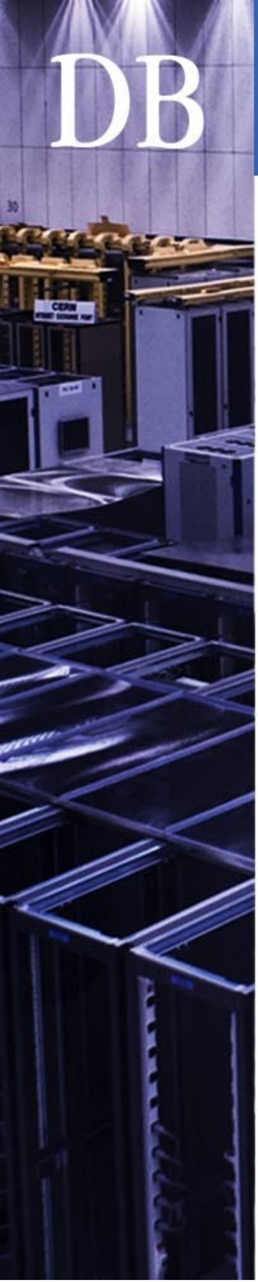
- Conference format
  - Invitational workshop (industry & science)
  - Main conference
- Participants
  - Industry: oil/gas (Chevron, Exxon, others), financial (Visa, NY exchange, banking), Medical/Bioinformatics, others including big names like eBay, Yahoo, Facebook, Amazon, IBM, EMC, HP, …; RDBMS vendors (Oracle, MS, Teradata)
  - Science: many laboratories and science institutes, representatives of different science projects in astronomy, astrophysics, bioinformatics; open-source communities, …



XLDB topics

- Main topics and challenges
  - How to store PBs of data and retrieve them efficiently
    - HEP community – 10-15PB/year now
    - Astronomy – 10PB/year
      - Large Synoptic Survey Telescope (in 2017)
    - Use of SSDs
  - How to analyze the data
    - Unpredictable query load (real-time vs. offline processing)
    - Full scans preferred over index access for some data (astronomical pixel data, genome, ...)
    - Complicated algorithms for data processing
      - Use of GPUs for offloading
    - Stream processing
  - SciDB, benchmarking





XLDB questions

- Data management
  - File systems vs. (R)DBMSs
  - Scientific tools and data formats
  - Online data and historical data challenges
    - Millisecond latency vs. PB analysis
- Data processing
  - How to build efficient processing systems
    - 2nd Amdahl's Law – number of bits of IO/sec per instruction/sec
  - Parallel processing

XLDB solutions

- Different scientific tools and data formats
  - ROOT, FTOOLS, DS9
  - dCache, CASTOR, Xrootd
  - netCDF, HDF5, fits, xtc
- SSDs used for data and caching
- Clusters with Amdahl number = 1 for under $40k (18GB IO/s)
  - Test – histogram of 544 million objects from 1.2TB of data – SQL executes in 100s
- SQL query offloading in GPUs
- Scalable share-nothing MySQL (Facebook)
  - Memcache, flsahcache
- Different RDBMS systems – most run Oracle, MSSQL or MySQL (industry also runs MySQL)
- **Move the processing to the data**
- **Hadoop** (Yahoo and many others in industry and science!)
  - Map reduce and extreme parallel processing

Hadoop

- Open-source project for reliable, scalable, distributed computing
  - Subprojects: Chukwa (monitoring), HBase (DB with Bigtable-like capabilities), HDFS (clsuter file system), Hive (parallel SQL data warehouse), MapReduce, Pig (high-level data-flow language), ZooKeeper (a high-performance coordination service)
  - Many use cases across different domains (industry and science)
  - Super parallel computing (60 seconds for 1TB sort with 1500 nodes – year 2009)
  - Yahoo – 3.7PB data processed daily, 120TB daily event data processed, >4000 nodes, 16 PB raw disk space
    - Streaming analytics
    - Warehouse solution

Hadoop - HDFS

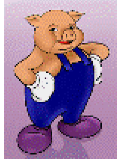
- Hadoop Distributed File System (HDFS)
  - Primary storage system used by Hadoop applications. HDFS creates multiple replicas of data blocks and distributes them on compute nodes throughout a cluster to enable reliable, extremely rapid computations.
  - HDFS used as storage layer for CMS Tier-2 at Nebraska – replaced dCache (see this link)

Hadoop – how to analyze data

- Hadoop provides massive scale out and fault tolerance capabilities for data storage and processing (using the map-reduce programming paradigm)
- Hadoop core – java programming required
- Hive – SQL like interface:

```
CREATE TABLE invites (foo INT, bar STRING) PARTITIONED BY (ds STRING);
LOAD DATA LOCAL INPATH './files/kv2.txt' OVERWRITE INTO TABLE invites PARTITION
    (ds='2008-08-15');
FROM invites a INSERT OVERWRITE TABLE events SELECT a.bar, count(*) WHERE a.foo >
    0 GROUP BY a.bar;
```

- Pig – Pig Latin language:



```
raw = LOAD ‘mylog.log' USING PigStorage('\t') AS (user, time, query);
clean = FOREACH raw GENERATE user, time, org.apache.pig.tutorial.ToLower(query) as
    query;
houred = FOREACH clean GENERATE user, org.apache.pig.tutorial.ExtractHour(time) as
    hour, query;
ngramed1 = FOREACH houred GENERATE user, hour,
    flatten(org.apache.pig.tutorial.NGramGenerator(query)) as ngram;
...
STORE res2 INTO '/tmp/tutorial-join-results' USING PigStorage();
```

XLDB summary

- XLDB – a very interesting conference with science and industry brought together
- Many questions asked and issues raised
- Still not many successful stories (mostly industry)
  - SKA (Square Kilometre Array) may change it – see:
    - http://www.skatelescope.org/video/SKA_Animation_2010.mov
- SciDB – a solution to science DB projects?
  - See http://www.scidb.org/
- Most presentations from XLDB 2010:
  - http://www-conf.slac.stanford.edu/xldb10/Program.asp