Database Technology Evolution

Vinayak Javaly

vjavaly@gmail.com

5/2/2013

Agenda

- Background
- DB Evolution
- NoSQL Movement
- NoSQL Taxonomy
- MapReduce
- MySQL
- MongoDB
- Download, Setup, Examples

Background

- 30+ years software development experience at IBM, Merrill Lynch, numerous startups

- 20+ years DB experience with Sybase, Oracle, MySQL, MongoDB, etc.

- 10+ years Adjunct Professor at New York Institute of Technology

DB Evolution

- ## 1960s
  - Hierarchical data structure (IBM IMS)
  - Network data structure (CODASYL)

- ## 1970s
  - Relational data model
    - *A Relational Model of Data for Large Shared Data Banks* – E. F. Codd [1970]
  - System R (IBM), Ingres (Berkeley)

DB Evolution

- 1980s
  - Commercialization of RDBMS
    - Oracle, Sybase, IBM DB2, Informix
  - SQL
  - ACID (**A**tomic, **C**onsistent, **I**solated, **D**urable)
- 1990s
  - PC RDBMS
    - Paradox, Microsoft SQL Server & Access
  - Larger DBs, driven by internet
  - Consolidation among commercial DB vendors

DB Evolution

- 2000s
  - Commercialization of Open Source RDBMS
    - MySQL, Postgres
  - Evolving requirements expose RDBMS limitations
    - Storing complex and dynamic objects
    - Processing increasing data volumes
    - Analyzing massive amounts of data

NoSQL Movement

- Eric Brewer's CAP (**C**onsistency, **A**vailability, **P**artition Tolerance) Theorem [2000]
  - Pick 2!
- Research
  - *MapReduce: Simplified Data Processing on Large Clusters* – Google [2004]
  - *Bigtable: A Distributed Storage System for Structured Data* – Google [2006]
  - *Dynamo: Amazon's Highly Available Key-value Store* – Werner Vogels, et. al. [2007]
  - *Pregel: A System for Large-Scale Graph Processing* – Google [2010]
- BASE (**B**asic **A**vailability, **S**oft-state, **E**ventually Consistent)

NoSQL Taxonomy

- ## Key-value
  - memcached, Redis, Riak, Tokyo Cabinet, Voldemort, Amazon SimpleDB

- ## Column-oriented (Bigtable clones)
  - Cassandra, HBase

- ## Document-oriented
  - MongoDB, CouchDB

- ## Graph
  - Neo4J, FlockDB, OrientDB, Pregel (Google)

MapReduce

- Framework for processing parallelizable problems across huge datasets using a large number of computers
  - What does this mean?

1. Prep Map input
2. Execute Mappers on input partitions
3. Shuffle Map outputs to prep Reducers
4. Execute Reducers on "shuffled" partitions
5. Produce final output

MySQL

- MySQL released Open Source [2000]
- Key features - pluggable storage engines, replication, transactions, event scheduler
- During 2000s
  - Increasing conversions from commercial RDBMS
  - Gradually added full RDBMS functionality
  - Maturing ecosystem
- Acquired by Sun [2008], then Oracle [2009]
- Large users – Amazon, Google, Facebook

MongoDB

- Created by founders of DoubleClick [2007]
- Developed (Open Source) by 10gen
- Key features – schema-free, aggregation, sharding, replica set, data expiration
- Extremely responsive to client requirements
- Great user community & ecosystem
- Large users – Foursquare, Bit.ly, SourceForge

Download, Setup, Examples

- brew install mysql
- brew install mongodb
- mysql.server start
- mongod --config </etc/mongod.conf>
- http://www.mysqltutorial.org/mysql-sample-database.aspx
- mysql –u root <sample_db_file>
- mongo

Questions

- ? ? ?