Team Presentations: Part I
Topic Selection

Due date: Tuesday, 14.

Topic Selection

It is time for each team to pick the research topic. During the course of the quarter, each team must deliver one 60-80 minute presentation discussing the architecture and the operation of one specific DBMS. For the most part, we are interested in exploring a diverse range of DBMS. As we are discussing traditional, row-oriented relational DBMS in class, instances of RDBMS (Oracle, MySQL InnoDB, MS SQL Server, DB2) are considered off-limits. Another "off-limits" DBMS is MarkLogic Server, as we are discussing its work in class as part of the course project.

Beyond that, all other DBMS are fair game. Specifically, we can identify the following categories of DBMS that are of interest to us:

- Key-Value stores. Simple disk-based or in-memory DBMS designed around put() and get() requests. E.g., memcached, CouchBase, Dynamo (early versions).
- Document stores. The bread-and-butter of NoSQL DBMS, these provide key-value storage facilities, but also allow some querying inside the stored objects. E.g., MongoDB.
- Column-based Relational DBMS. These RDBMS store data grouped by columns, not rows. This requires a significant change in how queries are processed at the low level. E.g. MonetDB.
- NoSQL Column DBMS. Built to facilitate distributed computations (MapReduce and such), these DBMS represent data with column families and columns. E.g. BugTable, Hive, Cassandra.
- Graph-oriented DBMS. Built to analyze relationships between different objects. E.g., Neo4j, OrientDB.
- In memory DBMS. A new class of DBMS concentrated on processing everything in RAM, without letting persistent storage play a key role in processing. E.g., SAP Hana, Oracle Coherence, H2, WebDNA.
- Niche DBMS. Some DBMS are built for a very specific purpose, such as managing large quantities of scientific data, or supporting specific types of queries. E.g., H-Store, VoltDB, SciDB, ASTERIX.

Figure out, what type of DBMS interests you, find out what DBMS are available, and conduct a preliminary literature search in order to determine how much information about a specific system is available to you. We are specifically looking for information related to the internal organization of the DBMS, which would allow us to do an apples-to-apples compare and contrast with the DBMS we are studying in class.

Deadline: Complete this part of the assignment by Tuesday, October 14. Presentations may commence the week after.

What to do: Put the name of the DBMS your team is studying, together with the bibliographic references to any and all materials about this DBMS you are planning to use for your presentation on your team's Wiki page in `The-Mothership` github repository for the course. Additionally, submit in hardcopy the same information (one page, with your team name, team members, name of the DBMS, bibliographic sources and a brief (2-3 sentences) statement why you chose this DBMS). Bring the hard copy submission to class next Tuesday.

What is next: I hope to approve all selected topics (approval is based on you selecting a real DBMS and providing sufficient evidence that you have material to prepare your presenation). After that, we will set aside seven lab periods, one per team, to deliver the presentations. We will collectively determine the presentation schedule and you will then start working on the presentations. The synergistic activities part of Stage 1.5 will outline specific expectations from the actual presentation.