DBMS: Challenges

What We Want From DBMS

- **Efficiency:** DB queries/updates processed in real-time.
- **Availability:** uniterrupted access to the stored data.
- **Multi-user**
- **ACID properties**:
  - **Atomicity**
  - **Consistency**
  - **Isolation**
  - **Durability**
- **Scalability:** DBMS is capable of processing vastly increasing quantities of data efficiently.
- **Integrity**
- **Rich queries**
- **Rich data models**

What else?

Challenges

These have been outlined in [1] and [2].

More data. Data volumes beyond what can be stored in a single location. Challenges:

- Scalability
- Efficiency

More velocity. Streaming data. Challenges:

- Scalability
- Availability
- Efficiency

More users. Users from multiple geographic locations. Challenges:

- Availability
- Isolation (as in: ability to perform so many transactions at the same time correctly)
- Consistency

More diverse data. Unstructured. Textual. Semi-structured. Challenges:

- Rich (?) data models.

More purposes. Data management for business analytics, Web 2.0 data, scientific data, workflow data, in the cloud, etc... Challenges:

- Rich queries (not always necessary)
- Scalability
- Consistency (not always necessary?)
- Isolation (not need for transactions?)

More uncertainty in data. Data from multiple sources. Integration. Data with uncertain values. Crowdsourcing. Challenges:

- Rich (?) data models
- Rich (?) queries

More processors. Challenges:

- ACID

More prying eyes. Challenges:

- Integrity

Big Data

Big Data. A collection of datasets so large and complex that it becomes hard to process it using traditional database management tools[1].

- Big data is huge in size.
- Big data is non-transactional in nature.
- Big data has complex structure.
- Big data needs to be stored and indexed.
- Big data needs to be analyzed.

Processing Big Data. Five steps and five challenges.

Steps:

- Acquisition
- Extraction/Cleaning
- Integration
- Analysis/Modeling
- Interpretation

Challenges:

- Hetergeity (see: "more diverse data")
- Scale (see: "more data", "more velocity")
- Timeliness (see: "more velocity", "more processors")
- Privacy (see: "more prying eyes")
- Human Collaboration (see "crowdsourcing")

Research Directions

New database engines. Special purpose DB engines (DBMS) for specific types of data management problems.

Data management in the cloud. Data management while sharing physical resources.

Data management for structured, semistructured and unstructured data.

Data management for mobile applications.

[1] http://en.wikipedia.org/wiki/Big_data

Declarative programming for emerging paradigms.

References

[1] R. Agrawal et al. (2009) The Claremont Report on the Database Research, *Communicatios of the ACM*, Vol. 52, No. 7, June 2009.

[2] D. Agrawal et al. (2012) Challenges and Opportunities with Big Data, *community white paper*.