No SQL Databases[IT413]

pm_jat @ daiict

SQL Databases

- SQL Databases primarily mean Relational Databases
- Relational databases are predominantly around since
  - its proposal by EF Codd [1970]
  - IBM System R first implementation of SQL [1974]
- In recent decades, the enterprise computing saw many changes in terms of programming languages, architectures, platforms, and processes; however one thing remained same was "relational databases", till the time
- "No SQL" emerged (typically since 2009) as a major challenger.

Relational "Reign"

- What has been great about Relational databases that they ruled for decades?
  - Shared Access of Integrated data
  - Data Independence [Three Schema Architecture]
  - Standardization [ANSI SQL, ODBC, JDBC, etc]
  - Concurrent Access of data
  - Query Optimization

NoSQL

- The term NoSQL first appeared in 2009 in name of meet "NoSQL Meet"
- Today, most commonly spelled as "Not Only SQL" – has some critic though
- NoSQL generally applied to a number of recent non-relational databases such as Cassandra, MongoDB, Neo4J, and Riak.
- Primary characteristics are
  - Run on clusters – can scale seamlessly well
  - Flexible Schema – can be even schema less
  - Easier to program

Course Content

- No SQL databases - Introduction and Motivation.
- Understand, and experience popular genres of NoSQL Databases
  - Map Reduce framework. Querying raw data without Schema- Pig, Hive, and Spark-SQL
  - Key value databases [Dynamo DB | Redis]
  - Document Databases. [CouchDB | MongoDB]
  - Column Oriented databases [MonetDB | Cassandra]
  - Graph Databases [Neo4J]
- Dive into some of NoSQL implementations

Pedagogy

- The book "NoSQL distilled" [text 1] gives complete width of the course and is first read for the course
- More depth is drawn from other texts or respective white/research papers
- There will be series of Lab exercises.
- Students are expected to get dive into one of topic/problem in the scope the course and deliver a term paper. It can be a database development and programming project as well.
- As a mid term deliverable of term paper, students will have to present a seminar

Evaluation

In semester examination (one)	20%
Seminar (will finish by Feb 28) – will be bonus marks upto 100% to early presenters	10%
Term Paper	30%
Lab Assignments	10%
End Semester examination	30%

Term Paper and Seminar

- Objective: to get more detailed insight in a technology, technique, relevant problem and its solution, and PRESENT to the class
- Can be
  - A research problem and its solutions [at least one good research paper is required]
    - Storgae, Sharding, Indexing, Performance,
  - It can be study of any of NoSQL database management systems [https://nosql-database.org/]
  - Database Design and Programming project
- Seminar is typically mid term presentation of your "term paper".

Texts

1. Sadalage, Pramod J., and Martin Fowler. *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. Pearson Education, 2013.
2. Perkins, Luc, Eric Redmond, and Jim Wilson. *Seven databases in seven weeks: a guide to modern databases and the NoSQL movement*. Pragmatic Bookshelf, 2018.
3. Sullivan, Dan. *NoSQL for mere mortals*. Addison-Wesley Professional, 2015.
4. Harrison, Guy. *Next Generation Databases: NoSQLand Big Data*. Apress, 2015.