

## CS3700

### Introduction to Database Systems

Prof P Sreenivasa Kumar  
Department of CSE, I I T Madras

---

---

---

---

---

---

---

---

## Introduction to Database Systems

### CS3700

Objectives:  
Introduce the learner to the fundamental ideas/ principles of relational databases and relational database management systems.

Next few slides:  
A perspective of the “data” space

---

---

---

---

---

---

---

---

Prof P Sreenivasa Kumar, Department of CS&E, IITM.

2

## Kinds of Data

- Various kinds of Data we encounter everyday:
  - Enterprise (business) data
    - Structured data
  - Documents / webpages – Text data
  - Objects – JSON; XML – (label, value) pairs data
    - Semi-structured data
  - Voice / music / image / video data
    - Unstructured data
- This course focuses on *structured data*

---

---

---

---

---

---

---

---

Prof P Sreenivasa Kumar, Department of CS&E, IITM.

3

## Data at Large Enterprises

- Large Enterprises (Businesses)
  - Need to deal with multiple types of data
    - E-commerce company
      - Product specs, images, demo videos, customer profiles ...
      - Transactions data
  - Multiples sources of data
    - Each division – its own data
    - External data – stock prices, media reports etc
  - Data Integration – a necessity
- Need multiple DBs and a Data Warehouse

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

4

---

---

---

---

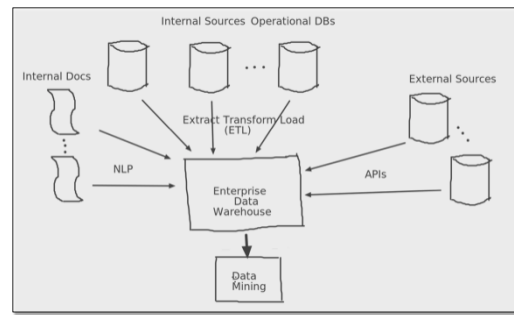
---

---

---

---

## Data Warehouses



Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

5

---

---

---

---

---

---

---

---

## Data Mining vs Querying

- Data Mining – Knowledge Discovery in Databases
  - Discover hither-to unknown 'knowledge' in data
  - Knowledge – associations / clusters / classifications ...
    - Custom-designed algorithms
- DB Querying
  - Ask for details present in data or can be derived
  - Use logic-based expression languages – SQL/OQL etc
  - Ask for aggregated information
    - How many 'fiction' books got issued to 'CS' students in the last quarter ?

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

6

---

---

---

---

---

---

---

---

## OLAP - OLTP

- OLAP (Online analytical processing) queries:
  - Ask for aggregated information from data
  - Needs summarization of large amount of data
    - How many 'cosmetics' products were sold in all the stores located in the 'southern' region of the country in the last 2 quarters ( Big Basket or Amazon )
  - Specialized architecting of data is needed
    - Done on warehouse data using "Data Cubes"
- OLTP( Online transaction processing)
  - Process a transaction ( sale / credit card charge / reservation ) quickly – concurrency needs attention
  - An important functionality of DB servers

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

7

## Text data - IR Systems

- IR – Information Retrieval Systems
  - Typified by Search Engines – Web/Enterprise
  - Query – list of keywords – (avg : 2.4 words/query)
    - Intention is imprecise
    - Relational queries are very precise ( to see later)
  - Results – a ranked list(s) of documents
    - Ordered by relevance to the Query
  - Hugely successful
  - Modern IR systems – not just a repository of docs!
  - Needs a separate course

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

8

## OO Languages and Databases

- Object-Oriented (OO) programming
  - Deeply structured objects – nesting is common
    - Rectangle – set of 4 point objects; Point – a pair
  - Objects are not 'persistent' !
- Data persistence is one of the strength of DBs
- Two approaches
  - Take persistence to OO Languages
    - Object Database Systems – OQL
  - Take 'rich structures' to databases
    - Object-Relational Databases (ORDBMSs)

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

9

## Knowledge Bases

- Knowledge Bases
  - Founded on mathematical logic based languages
  - Inferences / reasoning incorporated
    - *Every surgeon is a doctor; John is a surgeon and hence a doctor*
  - Data : modeled as collection of ‘statements’
  - Metadata: Terms/Vocabulary of the domain
    - Names of entity types of interest, properties of interest for these entities; binary relationships among entities
    - Modeled as Ontologies to capture domain knowledge
  - Can be viewed as graphs – ‘Knowledge Graphs’
  - More in the TAO course....

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

10

---

---

---

---

---

---

---

---

## NoSQL Databases

- Non-relational databases / Not only SQL
  - Offer flexibility – no rigid row/col structure
    - Semi-structured + other simpler data models
  - Need: faster response to certain queries, quick data model updates, quick app development, horizontal scale-out ..
  - System ‘availability’ over ‘consistency’ of reads
  - (-ve) No standard query language to access data
  - Multiple types of DBs
    - JSON / XML, key-value, columnar, graph etc

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

11

---

---

---

---

---

---

---

---

## Types of NoSQL DBs

- JSON / XML DBs
  - aka ‘Document-oriented’ / ‘document’ DBs
  - Semi-structured / non-homogenous data model
- Key-value Stores
  - Value – single or a collection of values (opaque)
  - Like hash-tables – give key, get value – simple
- Graph DBs – Nodes and labeled edges
  - Query language support – SPARQL etc
- Column-oriented DBs – fast access of col data

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

12

---

---

---

---

---

---

---

---

### Course Organization

- Tutorials – Problem sessions – 5 or 6
  - Problem sheet – start of the class or a day before
  - Students to solve - discussion, questions – Ok
  - No marks – no judging; answers posted later
- Moodle based assessment tests – TBD
- Design and Dev assignment – group/individual
  - Running assignment – with 3 or 4 stages
- End-Sem exam – at the campus or KVs (?)
- Attendance – Google Meet tool

Prof P Sreenivasa Kumar, Department  
of CS&E, IITM.

13

---

---

---

---

---

---

---