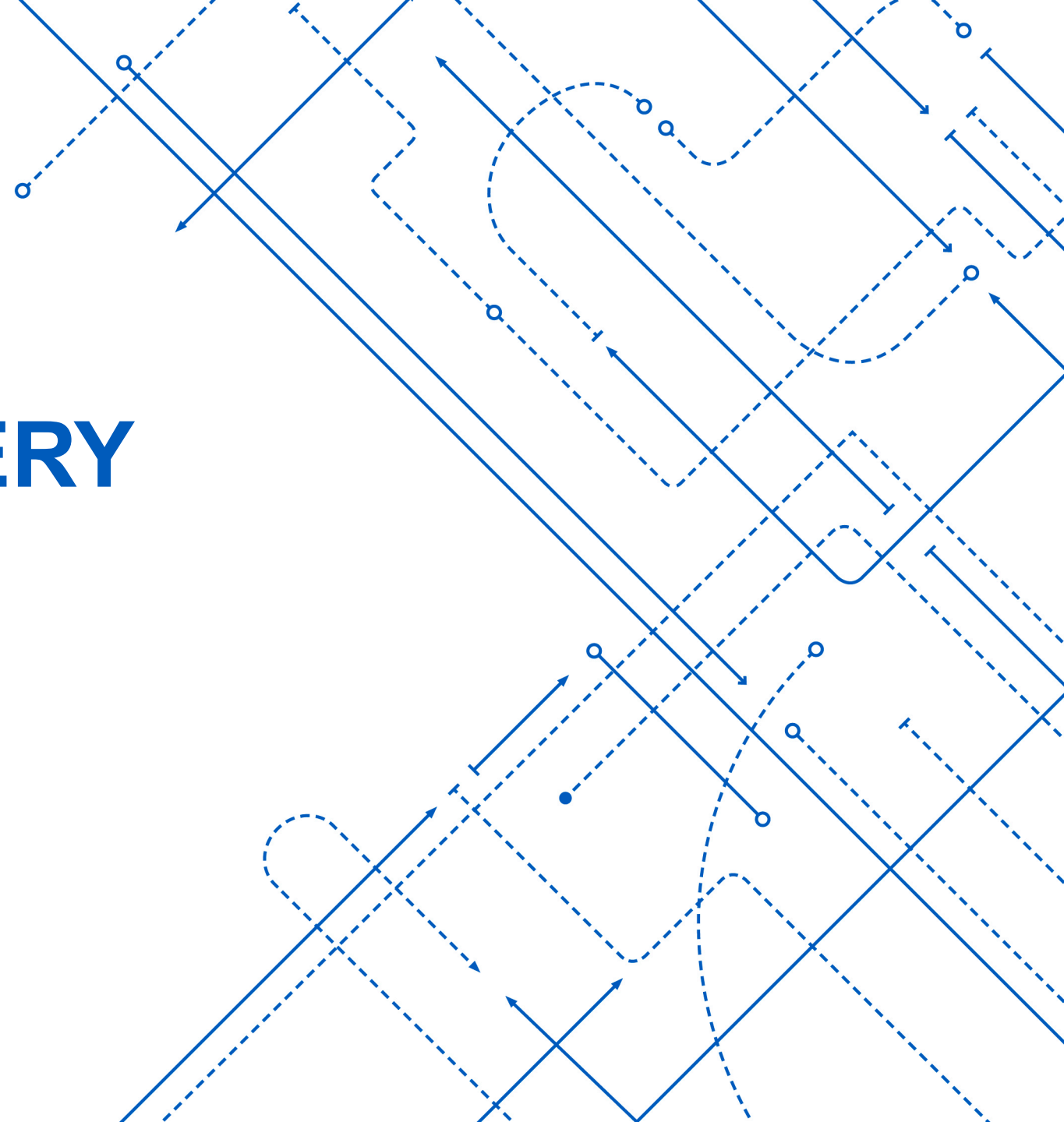


CSE460/560 DATA MODELS AND QUERY LANGUAGES

Introduction

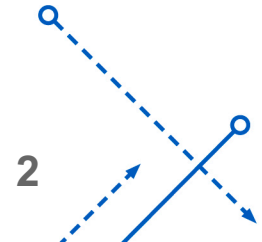
Cheng-En Chuang

(Slides Adopted from Jan Chomicki and Ning Deng)

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

Announcement

- About Force Registration
- TA: Zijian An

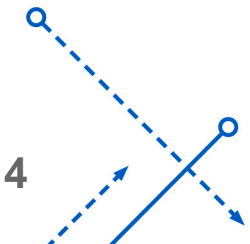


About Me

- A 2nd year PhD Student
- Worked at TrendMicro for 5 years
- TAed CSE460/560 for last two years
- Office: Capen 212
- Office Hour: 3:00 – 4:00

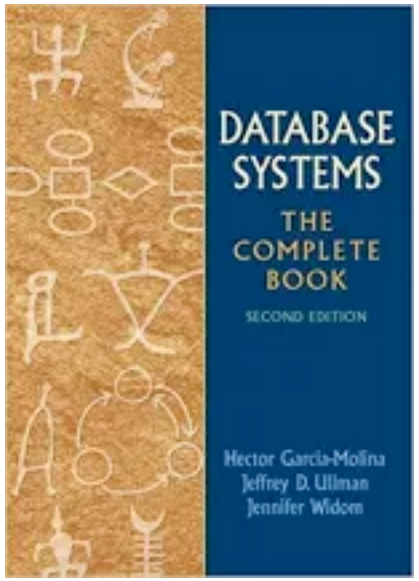
About CSE 460/560

- To build a database application
  - data models
  - query languages
- Topics
  - Design data models
  - Query languages
  - Relational Algebra
  - XML DB/XQuery
  - Views/Indexes
  - On-Line Analytic Processing



Textbook

- H. Garcia-Molina, J. Ullman, J. Widom.
  Database Systems: The Complete Book, 2nd Edition, Prentice Hall, 2008

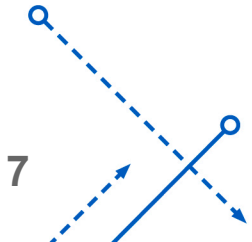


Course Webpage: Piazza

- https://piazza.com/buffalo/spring2020/cse460560
- Q&A
  - Form questions in a general manner
  - Do not cut/paste your code in public post
  - You can post anonymously
    - Only staffs know your identity
  - No stupid/wrong question when you are learning!
- Announcements
- Materials

Grades Breakdown

- Exams (55%)
  - In-class Midterm (15%)
    - Mar. 12 tentatively
  - Final Exam (40%)
- Three Projects (45%, 15% per each project)
- Final Grade Allocation (for CSE460)
  - A/A- (at least 80%)
  - B+/B/B-(at least 65%)
  - C+/C/C- (at least 50%)
  - D+/D (at least 40%)
  - F (below 40%)
  - CSE560 may be slightly adjust upwards

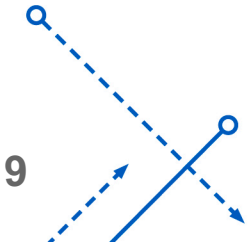


Policy

- Make-Up Exams
  - The request should be made at least one week before the test
  - with valid, documented reasons
- Late Submission
  - The submissions are due at midnight on the due date
  - No late submissions are accepted
- Regrading
  - within a week after the grades published

Academic Integrity

- Zero tolerance on cheating (projects or exams)
  - A violation on academic integrity will result in F of the course
- You should NOT copy others' work under any circumstance



Outline

1. Introduction
    1. What is database
    2. Database Management System (DBMS)

2. Data Models and Query Languages
    1. Data Models
    2. DBMS Languages

3. DBMS Components
    1. Major DBMS Components
    2. Other Components

Outline

1. Introduction
    1. **What is database**
    2. Database Management System (DBMS)

2. Data Models and Query Languages
    1. Data Models
    2. DBMS Languages

3. DBMS Components
    1. Major DBMS Components
    2. Other Components

What is Database

- Database
  - A computer representation of a collection of information
  - Persistent, Long-Lived
  - Typically shared by many users and programs
- What do we care when we manipulate and persistent data?
  - Correctness
  - Consistency
  - Parallelism
- What do we care when we ask and answer questions about data?
  - Accuracy
  - Efficiency

Databases are powerful

- Massive: terabyes of data every day
- Safey: ensure that data remains consistent
- Multi-User: concurrency control
- Efficient: efficiently data access and manipulation

Outline

1. Introduction
    1. What is database
    2. **Database Management System (DBMS)**
2. Data Models and Query Languages
    1. Data Models
    2. DBMS Languages
3. DBMS Components
    1. Major DBMS Components
    2. Other Components

Database Management System (DBMS)

- DBMS
  - A software system where the power of databases embodies in
    - Make it possible to build, modify and efficiently access large databases
    - Support one or more data models
    - Support multiple alnguages and interfaces for data definition, manipulatin, and querying
    - Application-independent
  - Additional DBMS functions
    - Integrity maintenance
    - Concurrency Control
    - Recovery
    - Access Control

Outline

1. Introduction
    1. What is database
    2. Database Management System (DBMS)

2. Data Models and Query Languages
    1. **Data Models**
    2. DBMS Languages

3. DBMS Components
    1. Major DBMS Components
    2. Other Components

Data Models

- A notation for describing data or information
  - Structure of the data
  - Operations on the data
  - Constraints on the data
- A view of data shared by the programs and users interacting with a database
- Supported by DBMS

Data Models

Data Model	Structure	Operations	Constraints
Relational	Relations	Relational Algebra	Key Constraints
Object-Oriented	Objects	Retrieve, Expand	Key Constraints
XML	Labelled trees	Traversal	Domain Constraints
Graph	Nodes, Edges	Traversal	Domain Constraints
NoSQL	Key-Value pairs	…	…

Schema vs. Instance

- Schema
  - Capture and describes the structure of the data
  - Time-independent
- Instance
  - Capture the crrent state of data
  - Conforms to the schema
  - Time-dependent
  - Only one exists at any given time

A Data model for Flight

- Which part is the schema? And Instances?

Schema →

Instances →

Number	Airline	Stock Price
63	American	26.11
556	Delta	56.83
1910	United	77.63

Outline

1. Introduction
    1. What is database
    2. Database Management System (DBMS)

2. Data Models and Query Languages
    1. Data Models
    2. **DBMS Languages**

3. DBMS Components
    1. Major DBMS Components
    2. Other Components

DBMS Languages

- Data Definition Language (DDL)
  - Define/Modify database schema
- Data Manipulation Language (DML)
  - Create and update instance
  - Various kinds of updates: incremental, bulk
  - Transactions
- Query Languages
  - Retrieve information from database

Example Queries

- Simple lookup query : What is the airline which number is 63?

```
SELECT FLIGHT.AIRLINE
FROM FLIGHT
WHERE FLIGHT.NUMBER = 63
```

- Complex lookup query (join) : What is the terminal for flight number 63?

```
SELECT TERMINAL.TNAME
FROM FLIGHT, TERMINAL
WHERE FLIGHT.NUMBER = 63
AND FLIGHT.AIRLINE = TERMINAL.AIRLINE
```

- Complex lookup query (aggregation) : How many flights for each company?

```
SELECT AIRLINE, COUNT(NUMBER)
FROM FLIGHT
GROUP BY AIRLINE
```

Outline

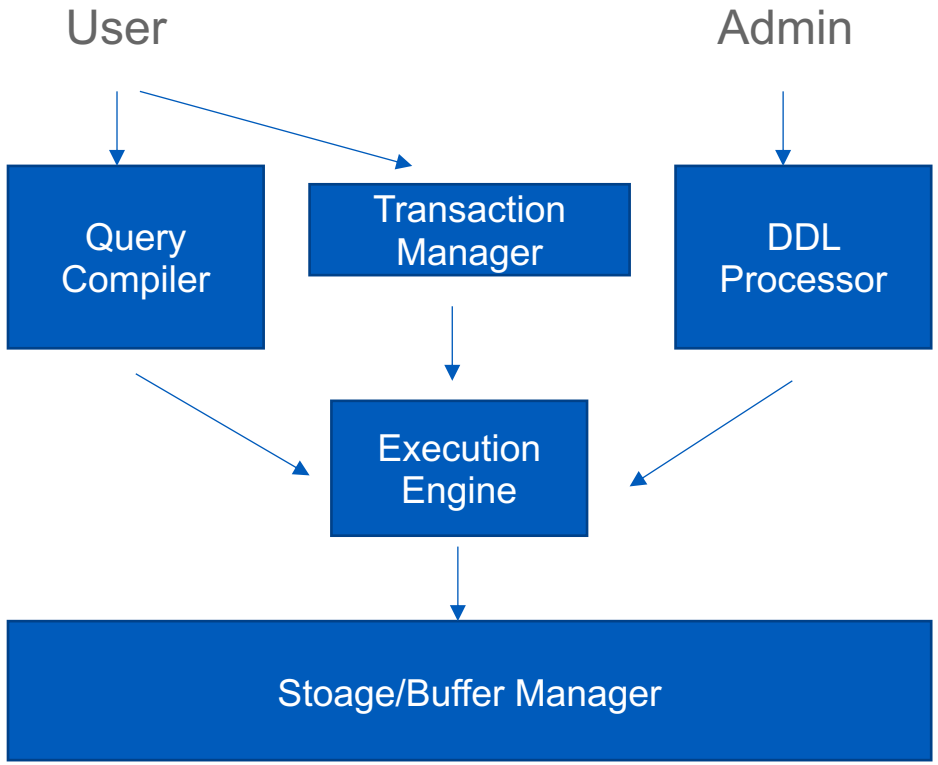
1. Introduction
    1. What is database
    2. Database Management System (DBMS)

2. Data Models and Query Languages
    1. Data Models
    2. DBMS Languages

3. DBMS Components
    1. **Major DBMS Components**
    2. Other Components

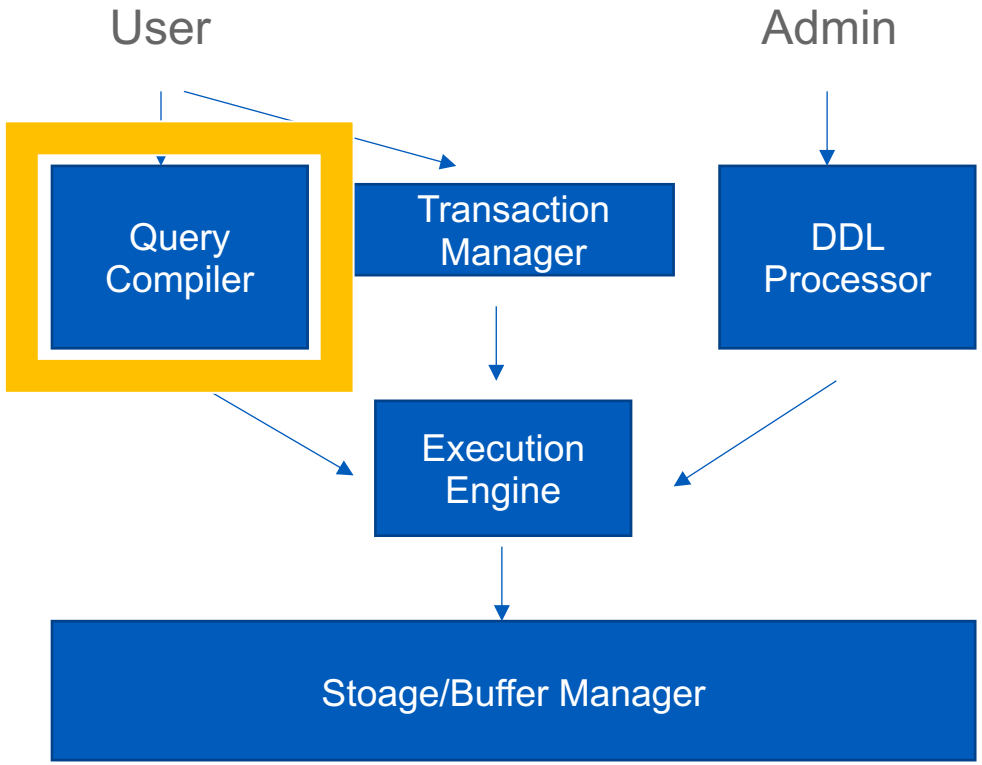
Major DBMS Components

- Types of Commands
  - Scheme change
    - Data-Definition Languages
  - Ask for data or modify data
    - Anwering the query (Query Compiler)
    - Transaction Processing (Transaction Manager)
  - Execution Engine
  - Storage and Buffer Management



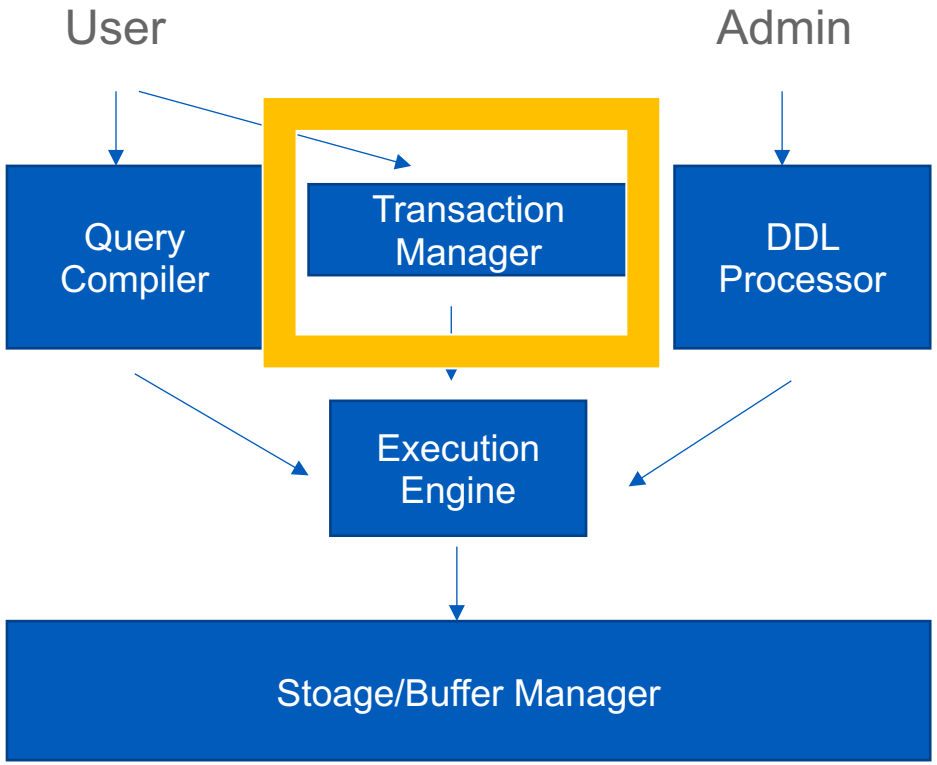
Major DBMS Components

- Query Compiler
  1. Query Parser
     - Build a tree structure from textual query
  2. Query Preprocessor
     - Perform semantic checks
     - Transform tree into algebraic operators form, the initial query plan
  3. Query Optimizer
     - Transform the initial query plan into best available sequence of operations
     - (x * y) + (x * z) = x * (y + z)
       1. (x * y) + (x * z) => 2 multiplications, 1 addition
       2. x * (y + z) => 1 multiplication, 1 addition



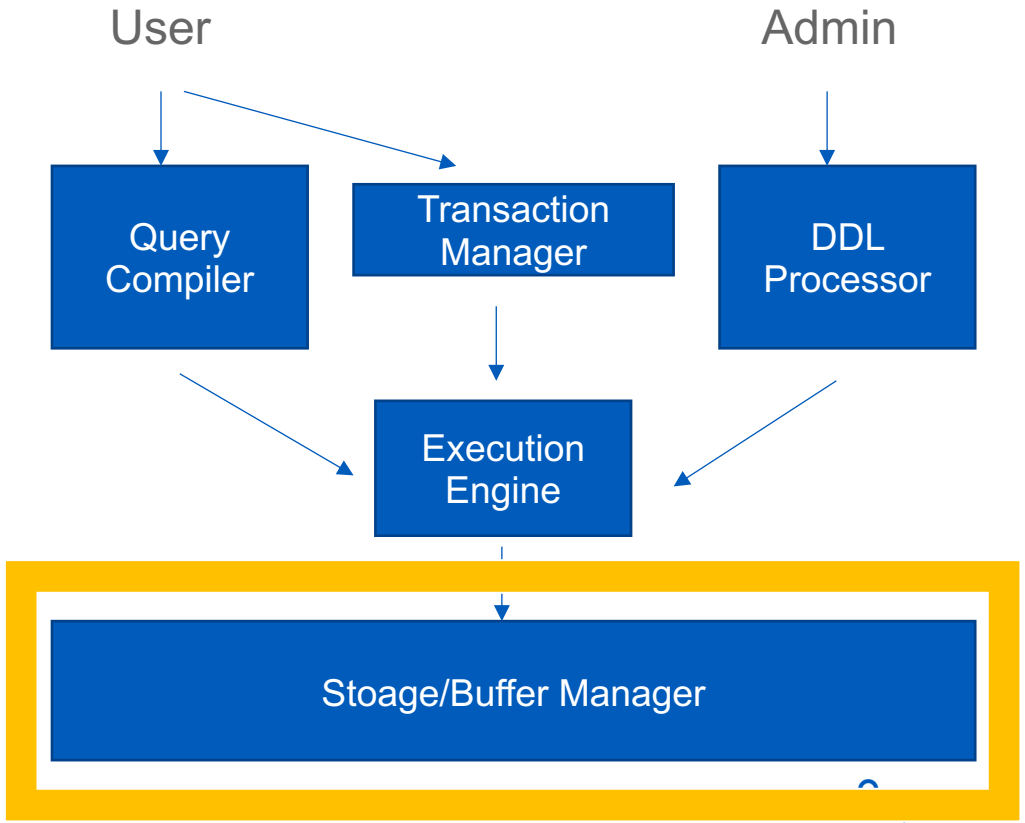
Major DBMS Components

- Transaction Processing (Transaction Manager)
  - ACID Principle
    - Atomicity: All of Nothing
      - Deadlock resolution (roll-back/abort if needed)
    - Consistency: Keep data constraints
    - Isolation: Execution as ther is no others
      - Concurrent Control
        Use *locks* to ensure isolation
    - Durability: Effect never be lost
      - Logging
        log every changes to assure durability



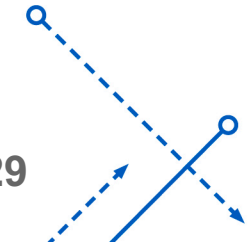
Major DBMS Components

- Storage/Buffer Manager
  - Data: the content of database
  - Metadata: the data describes the structure of, and contraints on, the databsae
  - Log Records: information about recent changes
  - Statistics: information gathered by DBMS about the database
  - Indexes: data structure that support efficient access
    - Example?



Client Communications Manager

- Architecture
  - Local and remote protocols
  - Two-tier: Client Server
  - Three-Tier: a middle tier between client and server
- Functions
  - Authentication
  - Connection state handling
  - Forwarding client requests
  - Shipping the result back



Outline

1. Introduction
    1. What is database
    2. Database Management System (DBMS)

2. Data Models and Query Languages
    1. Data Models
    2. DBMS Languages

3. DBMS Components
    1. Major DBMS Components
    2. **Other Components**

Other Components

- Process Manager
  - Allocating/Deallocating DBMS workers to clients
  - Processes and threads
- Shared Components and utilities
  - Catalog manager
  - Memory manager
  - …