



**cceHUB**  
cancer care engineering

[Login](#)[Register](#)

[HOME](#) [MY HUB](#) [RESOURCES](#) [CONTRIBUTORS](#) [EVENTS](#) [ABOUT](#) [SUPPORT](#)

# cceHUB: The Metadata Database

Ahmad Abusalah



# Database

Domain  
Knowledge

Database  
Normalization

Database  
Design

Administration  
Tasks

- Several Domains:
  - **D1:** Samples → Aliquots
  - **D2:** Tools and Instruments → Workflow
  - **D3:** Input and Output datasets → Tracking
  - **D4:** Patient Information → Activities, habits, diet, demographics, etc
  - **D5:** Treatments
- More than 439 attributes.



# Database

Domain  
Knowledge

Database  
Normalization

Database  
Design

Administration  
Tasks

• **Input:**

439 Attributes

• **Goal:**

Clean Database Design  
Organization  
No Redundancy  
Easy to modify

• **Process:**

1NF, 2NF, 3NF, BCNF

• **Output:**

Recommended Design



# Database

Domain  
Knowledge

Database  
Normalization

Database  
Design

Administration  
Tasks

- Recommended Design → Used Design:
  - Performance → some redundancy
- Add Constraints:
  - Null, Not NULL
  - Primary Key, Foreign Key
  - Data Type
- Select Storage Engine: InnoDB vs MyISAM
- Indexes: B-tree, FULLTEXT
- Scalable and Extendable.



# Database

Domain  
Knowledge

Database  
Normalization

Database  
Design

Administration  
Tasks

• **Scripts to perform tasks:**

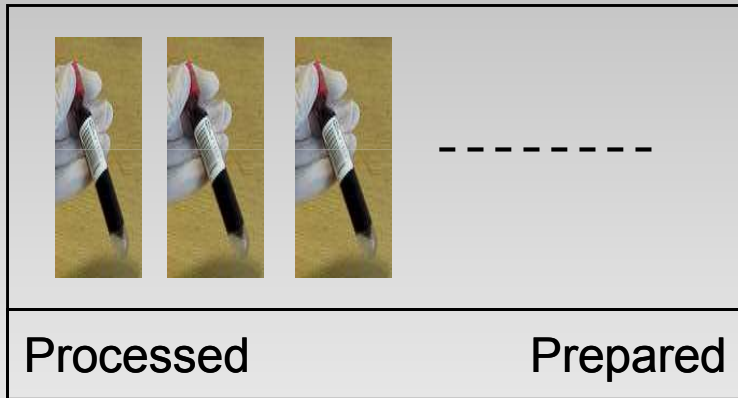
- Drop DB and recreate everything
- Populate base-data
- Populate test-data
- Create backups
- Restore backups



# Database – Domain Knowledge



**IU**

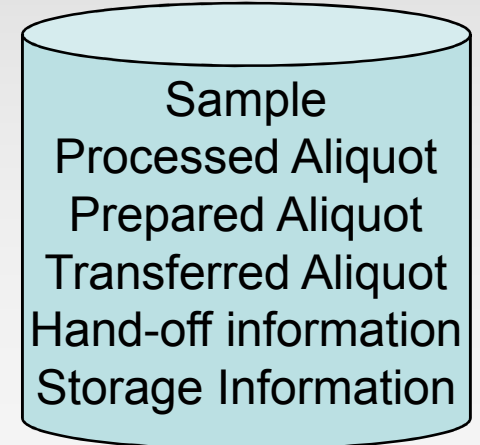
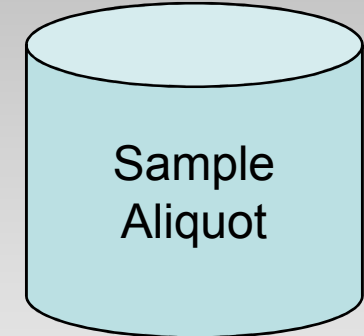
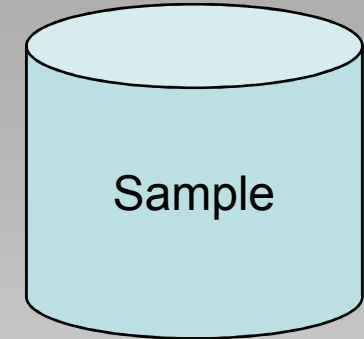


Transferred  
Hand-off  
Storage

**P**

16  
attributes

> 130  
attributes





# Database – Database Design

- Normalized vs. used Database Design
- SQL Statements may require some redundancy to support performance

- **B-tree index and FULLTEXT index**

**B-Tree:**

```
select * from table_name where column_name like '%value%';
```

**Full-Text:**

```
select * from table_name where match(column_name) against ('value');
```

- If column\_name is of type text or has long string, then search keywords inside this string is much faster through FULLTEXT.
- However, one limitation is that you need to search for exact matches.



# Database – Administration Tasks

- **Re-generate the database:**

- Shell Script that:
  - Drops the whole Database
  - Creates a database name “omics\_db”
  - Creates all tables using sql script
  - populate base-data and test-data using sql script

- **Backup script:**

- Shell Script that:
  - creates a backup using mysqldump
  - gzip it.
  - the created file name is: today's\_date\_db\_name.gz
    - For example (04012009\_omics\_db.gz)
  - cron job runs regularly

- **Restore script:**

- Shell Script that:
  - takes as argument the backupfile name
  - unzip it
  - drop current database
  - creates tables and populate data using the backup file



# Database - Questions

