1.What is incremental plan?

An **incremental plan** is a strategy used in project management and development that involves breaking a project into smaller, manageable components or increments. This approach allows teams to deliver parts of the project gradually, rather than waiting for the entire project to be completed.

**Example**

In software development, an incremental plan might involve:

1. **Increment 1**: Develop and deliver a basic version of the software with core functionalities.
2. **Increment 2**: Add additional features based on user feedback from Increment 1.
3. **Increment 3**: Improve performance and usability based on the next round of feedback.

**2.What is hdfs files format , hadoop files format for delimiter fiels**

In Hadoop, various file formats can be used to store data in HDFS (Hadoop Distributed File System), especially when dealing with delimited files.

**Example of Writing Delimited Files**

In frameworks like Apache Hive or Spark, you can specify the delimiter when creating tables or data frames. For instance, in Hive:



This creates a table that uses a comma as the delimiter for each record in a text file format.

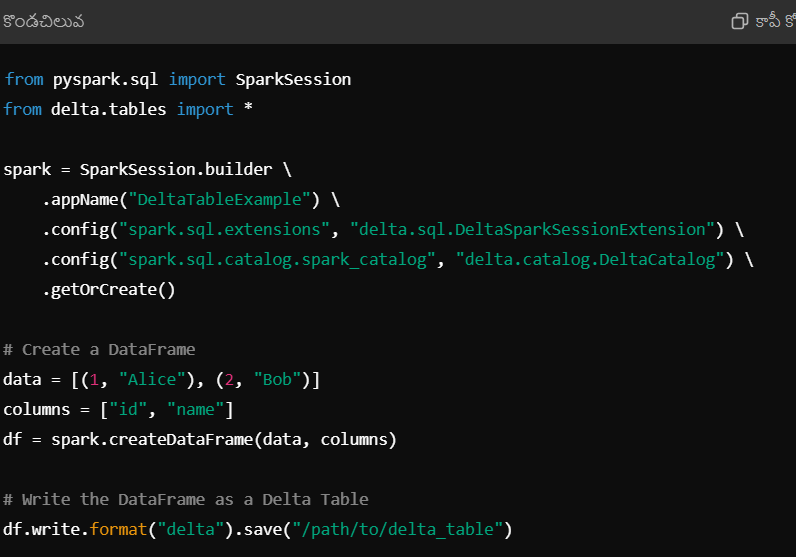
Output:



2.what is delta tables in hadoop?

**Delta Tables** in Hadoop refer to a specific type of data management strategy often associated with Delta Lake, an open-source storage layer that brings ACID transactions to big data workloads. Delta Tables allow for efficient handling of large datasets with the ability to manage updates, deletes, and merges in a more structured way compared to traditional data lakes.

Example



3.how to archive files in Hadoop?

Archiving files in Hadoop typically involves moving older or less frequently accessed data to a separate location for storage. This helps to optimize performance and manage storage costs. Here’s a step-by-step guide on how to archive files in Hadoop:

1.Identify Data to Archive.

2.Create an Archive Directory:

3.Move Files to Archive:

4.Use Compression (Optional):

5.Automate the Process (Optional).

6.Create External Table in Hive (Optional).

7.Delete Original Files (Optional):

Example:



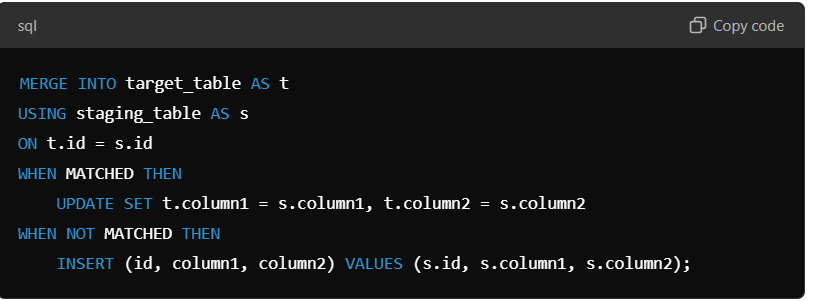
5.incremental load in hadoop table?

Incremental loading in Hadoop refers to the process of loading only the new or changed data into a target table, rather than reloading the entire dataset. This approach optimizes data processing and reduces resource usage. Here’s a high-level overview of how to perform incremental loads in Hadoop:

**Steps for Incremental Load in Hadoop**

1. **Identify the Source Data**:
   * Determine the source system and the method to identify new or changed records (e.g., timestamps, version numbers).
2. **Create a Staging Area**:
   * Set up a temporary staging area in HDFS where the incremental data will be loaded.
3. **Extract New Data**:
   * Use a suitable method (e.g., querying a database or reading files) to extract only the new or updated records from the source.
4. **Load Data into Staging**:
   * Use tools like Apache Sqoop, Apache NiFi, or custom scripts to load the incremental data into the staging area.
5. **Transform Data (if needed)**:
   * Apply any necessary transformations to the data before loading it into the target table.
6. **Merge Data**:
   * Use tools like Apache Hive or Apache Spark to merge the incremental data from the staging area into the target table.
   * This can involve:
     + **Inserts** for new records.
     + **Updates** for modified records.
     + **Deletes** for records that are no longer valid.

**Example using Hive**:



Listing instances

The following command is used to list instances:

db2ilist

This command lists all the instances that are available on a system.

**Syntax:**

db2ilist

**Example:**[To see how many instances are created in DB2 copy]

db2ilist

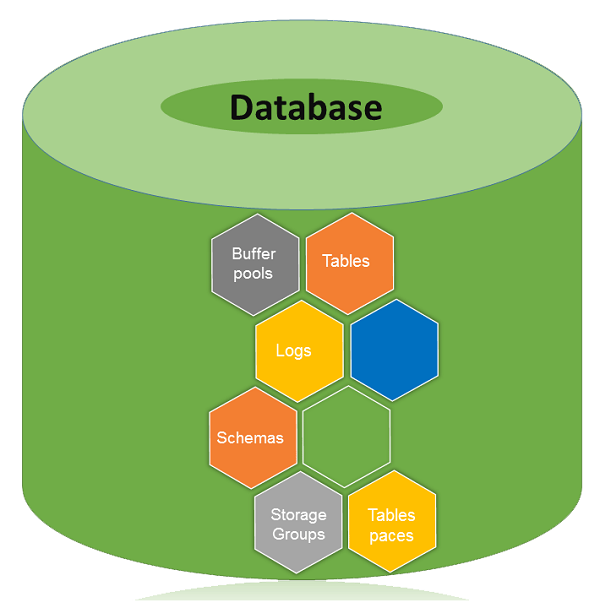
**Output:**

db2inst1

db2inst2

db2inst3

Database architecture



A database is a collection of Tables, Schemas, Bufferpools, Logs, Storage groups and Tablespaces working together to handle database operations efficiently.