Hello Friend, Let's finish what we started about How DBMS store and interact with Database.

We said in last lecture that DBMS store database in files and entire those files we store tuples in pages But we have another technique to store tuples.

Instead of storing tuples in pages, the DBMS can stores tuples in Log Records.

#The system appends log records to the file of how the database was modified.

#To read a record, the DBMS scans the log backwards and "recreate" the tuple to find what it's needs.

#Build indexes to allow it to jump to locations in log.

#Periodically compact log.

Compaction coalesces larger log files into smaller files by removing unnecessary records

Tuples:

   A tuple is essentially a sequence of bytes.

   The DBMS's job is to interpret these bytes into attribute type and values

The DBMS catalog contain the schema information about tuples that the system uses to figure out the tuples layout.

Large values: most DBMS don't allow a tuple to exceed the size of a single page.

To store values that are larger than a page, the DBMS uses separate overflow storage pages.

Some systems allow you to store larger values in an external file, treated as a Blob type.

Note: the DBMS cannot manipulate the content of an external file.

SYSTEM CATALOG

A DBMS stores meta-data about a database in it's internal catalogs

Note: almost every DBMS stores the database's catalog inside itself.

You can query the DBMS's internal by command Information_schemes

To get info about database.

Database Workloads: have three types

1-online-transaction processing (OLTP)

Fast operation that only read/update a small amount of data each time.

2-online-Analytical processing (OLAP)

Complex queries that read a lot of data to compute aggregates.

3- Hybrid transaction + Analytical processing

OLTP + OLAP together in the same database instance.

Data storage models

The DBMS can storage tuples in different ways that are better for either OLTP or OLAP workloads.

We have been assuming N-array storage model (row storage).

N-Array storage model (NSM):

The DBMS stores all attributes for a single tuple contiguously in a page

Their advantages:

Fast insert, updates and deletes.

Good for queries that need the entire tuple.

Not good for scanning large portion of the table and/or a subset of the

attributes.

Decomposition storage model (DSM):

The DBMS store values of a single attribute for all tuples contiguously in a

page (column storage).

Advantages:

Reduce amount wasted I/O

Better query processing and data compression.

Dis advantages:

Slow for point query , updates and deletes.

Tuple identifications:

#choice1: Fixed Length Offsets:

Each value is the same length for an attribute

#choice2: Embedded Tuple ID :

Each value is stored with it's tuple id in the column.

Conclusion:

The storage manager is not entirely independent from the rest of the DBMS.

It's important to choice the right storage model for the target workload:

OLTP = Row store

OLAP = column store