

Data Science & Artificial Intelligence



Warehousing

Measures - Categorization
and Computation

ONE SHOT



Astha Singh Ma'am

Recap of Previous Lecture



Topic

One topic

Topic

Two topic

Topics to be Covered



Topic

IMS Database

Topic

Measures

Topic

Houstic

Topic

Distributive

Topic

Algebric



Topic : The IMS Database System

- Information Management System — first developed in the late 1960s; historically among the largest databases.
- Issue queries through embedded calls which are part of the IMS database language DL/I.
- Allows the database designer a broad number of options in the data-definition language.
 - Designer defines a physically hierarchy as the database schema.
 - Can define several subschemas (or view) by constructing a logical hierarchy from the record types constituting the schema
 - Options such as block sizes, special pointer fields, and so on, allow the database administrator to tune the system



Topic : Record Access Schemes

- Hierarchical sequential-access method (HSAM) — used for physically sequential files (such as tape files). Records are stored physically in preorder.
- Hierarchical indexed-sequential-access method (HISAM) — an index sequential organization at the root level of the hierarchy.
- Hierarchical indexed-direct-access method (HIDAM) — index organization at the root level with pointers to child records.
- Hierarchical direct-access method (HDAM) — similar to HIDAM, but with hashed access at the root level.



Topic : IMS Concurrency Control

- Early versions handled concurrency control by permitting only one update application program to run at a time. Read-only applications could run concurrent with updates.
- Later versions included a program-isolation feature
 - Allowed for improved concurrency control
 - Offered more sophisticated transaction-recovery techniques (such as logging); important to online transactions.
- The need for high-performance transaction processing led to the introduction of IMS Fast Path. *



Topic: IMS Fast Path



- Uses an alternative physical data organization that allows the most active parts of the database to reside in main memory.
- Instead of updates to disk being forced at the end of a transaction, update is deferred until a checkpoint or synchronization point.
- In the event of a crash, the recovery subsystem must redo all committed transactions whose updates were not forced to disk.
- Allows for extremely high rates of transaction throughput.
- Forerunner of main-memory database systems.



Topic : Measures Their Categorization and Computation

Measures can be categorized or arranged into three categories: **holistic**, **distributive**, and **algebraic**. The said classification or division of measures is based on which type of aggregate functions is being used in them.

Three

Holistic:

If there is no defined constraint or limit on the storage amount needed to define the sub-aggregate, any given aggregate function is said to be holistic. It can be described as an algebraic function with n arguments.

For example, `median()`, `rank()`, and `mode()` are holistic measures. If any measure uses the holistic aggregate function then it can be said to be holistic. The majority of cube applications that work with big amounts of data demand quick computations of distributive and algebraic measurement.

Examples \rightarrow Mean() Median() Mode()

Distributive:

If any function is calculated in a delivered manner as listed then it is said to be a distributive function.

For example, `count()` for a data cube can be calculated by dividing or partitioning the cube into a group of sub-cubes of the same size. We can calculate `count()` for each sub-cube and then add them to get the total so we can conclude that the `count()` function is a distributive aggregate service.

A y measure is said to be distributive if it can be obtained by using the distributive aggregate service.

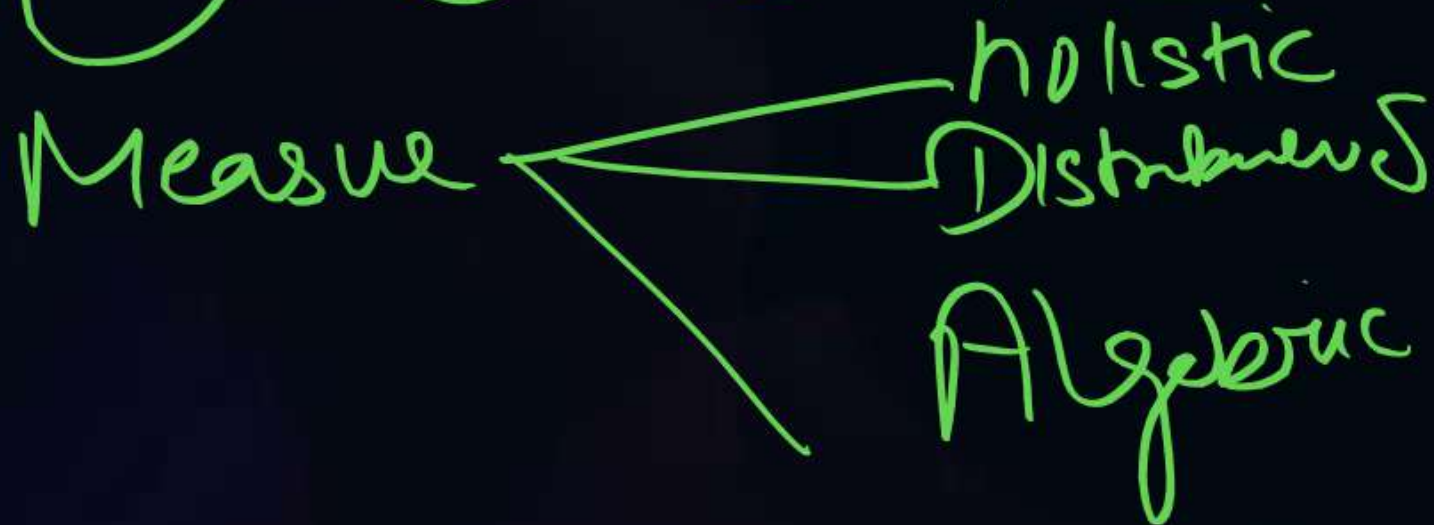
Examples: `Sum()`, `Count()`, `Minimum()`

Example

Algebraic:

If any aggregated function can be calculated by using an algebraic service then it is said to be algebraic. It is calculated by an algebraic function of N arguments where N is a positive integer.

Example: Average(), Max(), Min(), CenterofMass()



① Data warehouse

→ What is data warehouse

→ Data Mart

→ OLAP

→ Schemas

→ Data Normalization

→ Data transformation

→ Measure

Snowflake schema

Star schema



2 mins Summary



Topic

One

Topic

Two

Topic

Three

Topic

Four

Topic

Five



THANK - YOU