Data Models

Conceptual

Highest-level, Important entities and the relationships among them.

No attribute is specified.	No primary key is specified.

Logical

Data in as much detail as possible, without regard to how they will be physical implemented in the database

{{Steps:}}

1, Specify primary keys for all entities.

2, Find the relationships between different entities.

3, Find all attributes for each entity.

4. Resolve many-to-many relationships.	separate the two entities and create two one-to-many (1:n) relationships between them with a third intersect entity.

5, Normalization.

Physical

How the model will be built in the database.

A physical database model shows all table structures, including column name, column data type, column constraints, primary key, foreign key, and relationships between tables.

Convert entities into tables.

Convert relationships into foreign keys.

Convert attributes into columns.

Modify the physical data model based on physical constraints / requirements.

Partitioning Tables

Use Case:

As our data scales, queries/updates may not always fit into memory

Horizontal Partitioning:

Splitting a table by rows

Pros:	Indices of heavily-used partitions fit in memory; Supports OLAP and OLTP
Cons:	Partitioning existing table can be a hassle; Some constraints can not be set

Horizontal partitioning

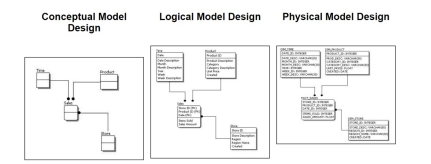
```
ATE TABLE sales (
...
timestamp DATE NOT NULL
)
PARTITION BY RANGE (times tamp);
CREATE TABLE sales_ 2019_q1
PARTITION OF sales
FOR VALUES FROM ('2019 -01 -01')
TO ('2019 -03 -31');
...
CREATE TABLE sales_ 2019_q4
PARTITION OF sales
FOR VALUES FROM ('2019 -09 -01')
TO ('2019 -12 -31');
CREATE INDEX ON sales ('time -
stamp')
```

Sharding

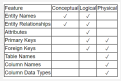
sharding implies the data is spread across multiple computers while partitioning does not	Partitioning is about grouping subsets of data within a single database instance

Sharding and partitioning are both about breaking up a large data set into smaller subsets.

Models: Diagramitic Overview



Models: Tablular Overivew



Feature	Conceptual	Logical	Physical
Entity Names	✓	✓	
Entity Relationships	✓	✓	
Attributes		✓	
Primary Keys		✓	✓
Foreign Keys		✓	✓
Table Names			✓
Column Names			✓
Column Data Types			✓

SSAS

Has a three-tier architecture

RDBMS: The data from different sources like Excel, Database, Text, others can be pulled with the help of ETL tool into the RDBMS.

SSAS: Aggregate data from RDBMS is pushed into SSAS cubes by using analysis services projects. The SSAS cubes will create an analysis database, and once the analysis database is ready, it can be used for many purposes.

Client: Clients can access data using Dashboards, Scorecards, Portals etc.

A cube is a basic unit of storage	It is a collection of data which has been aggregated to allow queries to return data quickly.

The MOLAP is made of data cube which contains of measures and dimensions

Dimension Table	contains dimensions, or characteristics (who, what, where), of a fact.
	Joined to fact table with a foreign key
	de-normalized tables.

SSAS (cont)

A fact table is the most important table in a dimens-ional model.	A Fact Table contains Measurements/fact and Foreign key to the dimension table. For example, payroll operations.

Two SSAS model types

Multi-Dim-ensional Data Model	consists of a data cube. It is a group of operations which allows you to query the value of cells by using cube and dimension members as coordinates.
	rules which decide the way that measure values are rolled up within hierarchies
Tabular modeling organizes data into related tables.	The table doesn't designate as "dimensions" or "facts" and development time is less with tabular because of all related tables able to serve both roles.

No SQL databases

Key - Value Stores	Document	Columnar	Graph

OLAP

On-Line Analytical Processing.

OLAP is Multidime-nsional	Providing ability to analyze metrics in different dimensions such as time, geography, gender, product, etc

For example, sales for the company are up. What region is most responsible for this increase? Which store in this region is most responsible for the increase?

For example, sales for the company are up:

What region is most responsible for this increase?	Which store in this region is most responsible for the increase?

SSRS

Microsoft SQL Server has three types of SQL Services

Microsoft SQL Server Integration services which integrate data from different sources.

Microsoft SQL Server Analytical service which helps for the analysis of the data

Microsoft SQL Server Reporting service allows for generating a visual report of the data.

SSRS data sources

Retrieve data from managed, OLE ODBC, and DB connections

Display data in a variety of formats which includes tabular, free-form, and charts

Reporting Life Cylce

Authoring: In this phase, the report author defines the layout and syntax of the data. The tools used in this process are the SQL Server Development Studio and SSRS tool.

SSRS (cont)

Management: This phase involves managing a published report which is mostly part of the websites. In this stage, you need to consider access control over report execution.

Delivery: In this phase, you need to understand when the reports need to be delivered to the customer base. Delivery can be on-demand or pre-defined schedule. You can also add an automation feature of subscription which creates reports and sends to the customer automatically

By **datamansam**

cheatography.com/datamansam/

Published 25th May, 2022.
Last updated 25th May, 2022.