

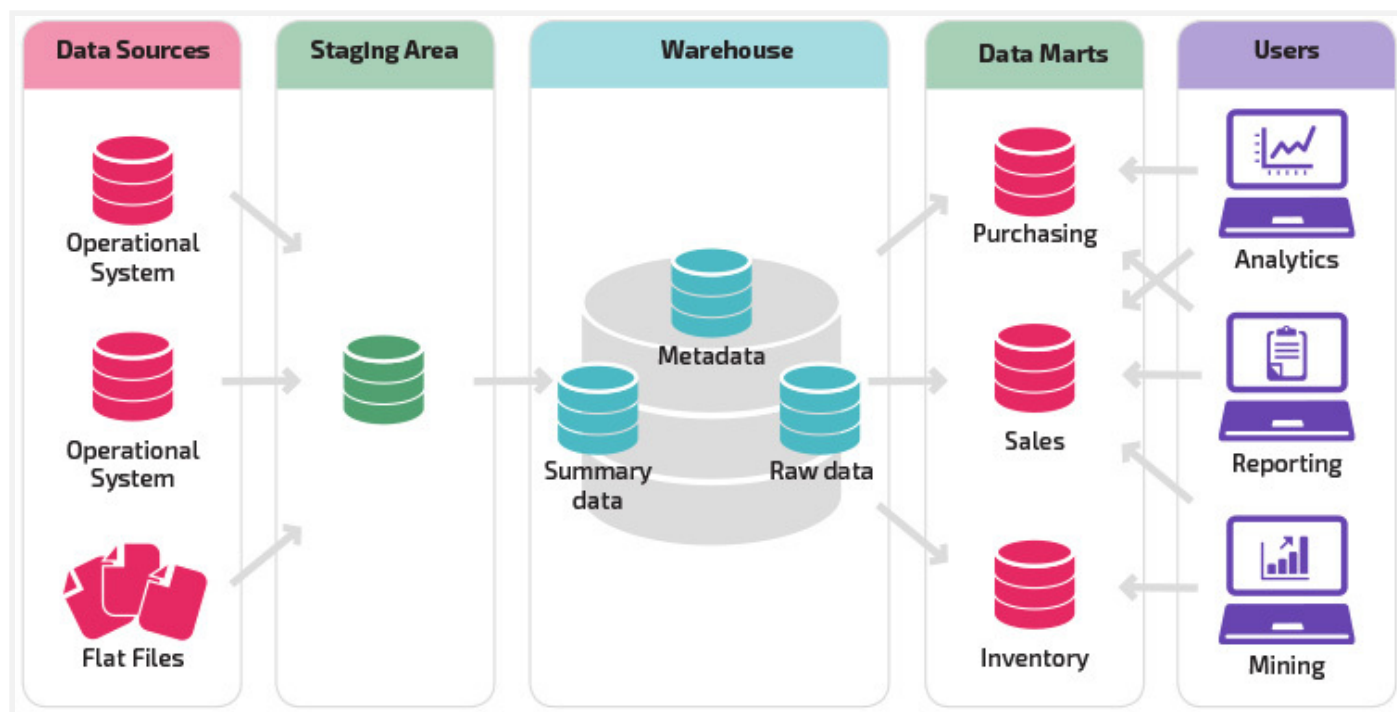
Data Warehouse Guide

Data Mart vs. Data Warehouse

Data mart vs. data warehouse—what is the difference? Discover why the old question of how to structure the data warehouse is no longer relevant.

A data mart is a subset of a data warehouse oriented to a specific business line. Data marts contain repositories of summarized data collected for analysis on a specific section or unit within an organization, for example, the sales department.

A data warehouse is a large centralized repository of data that contains information from many sources within an organization. The collated data is used to guide business decisions through analysis, reporting, and data mining tools.



Data Mart and Data Warehouse Comparison

Data Mart

- **Focus:** A single subject or functional organization area
- **Data Sources:** Relatively few sources linked to one line of business
- **Size:** Less than 100 GB
- **Normalization:** No preference between a normalized and denormalized structure
- **Decision Types:** Tactical decisions pertaining to particular business lines and ways of doing things
- **Cost:** Typically from \$10,000 upwards
- **Setup Time:** 3-6 months
- **Data Held:** Typically summarized data

Data Warehouse

- **Focus:** Enterprise-wide repository of disparate data sources
- **Data Sources:** Many external and internal sources from different areas of an organization
- **Size:** 100 GB minimum but often in the range of terabytes for large organizations
- **Normalization:** Modern warehouses are mostly denormalized for quicker data querying and read performance
- **Decision Types:** Strategic decisions that affect the entire enterprise
- **Cost:** Varies but often greater than \$100,000; for cloud solutions costs can be dramatically lower as organizations pay per use
- **Setup Time:** At least a year for on-premise warehouses; cloud data warehouses are much quicker to set up

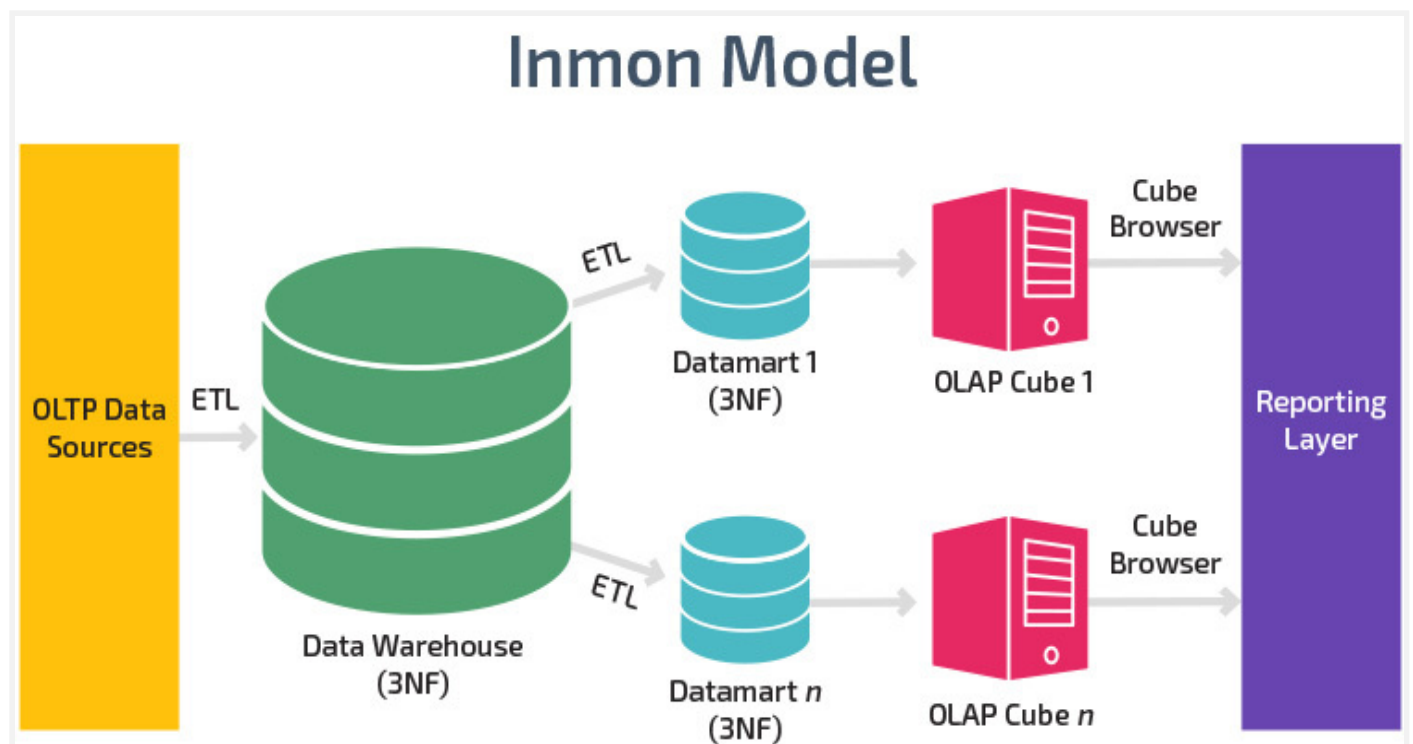
- **Data Held:** Raw data, metadata, and summary data

Inmon vs. Kimball

Two data warehouse pioneers, Bill Inmon and Ralph Kimball differ in their views on how data warehouses should be designed from the organization's perspective.

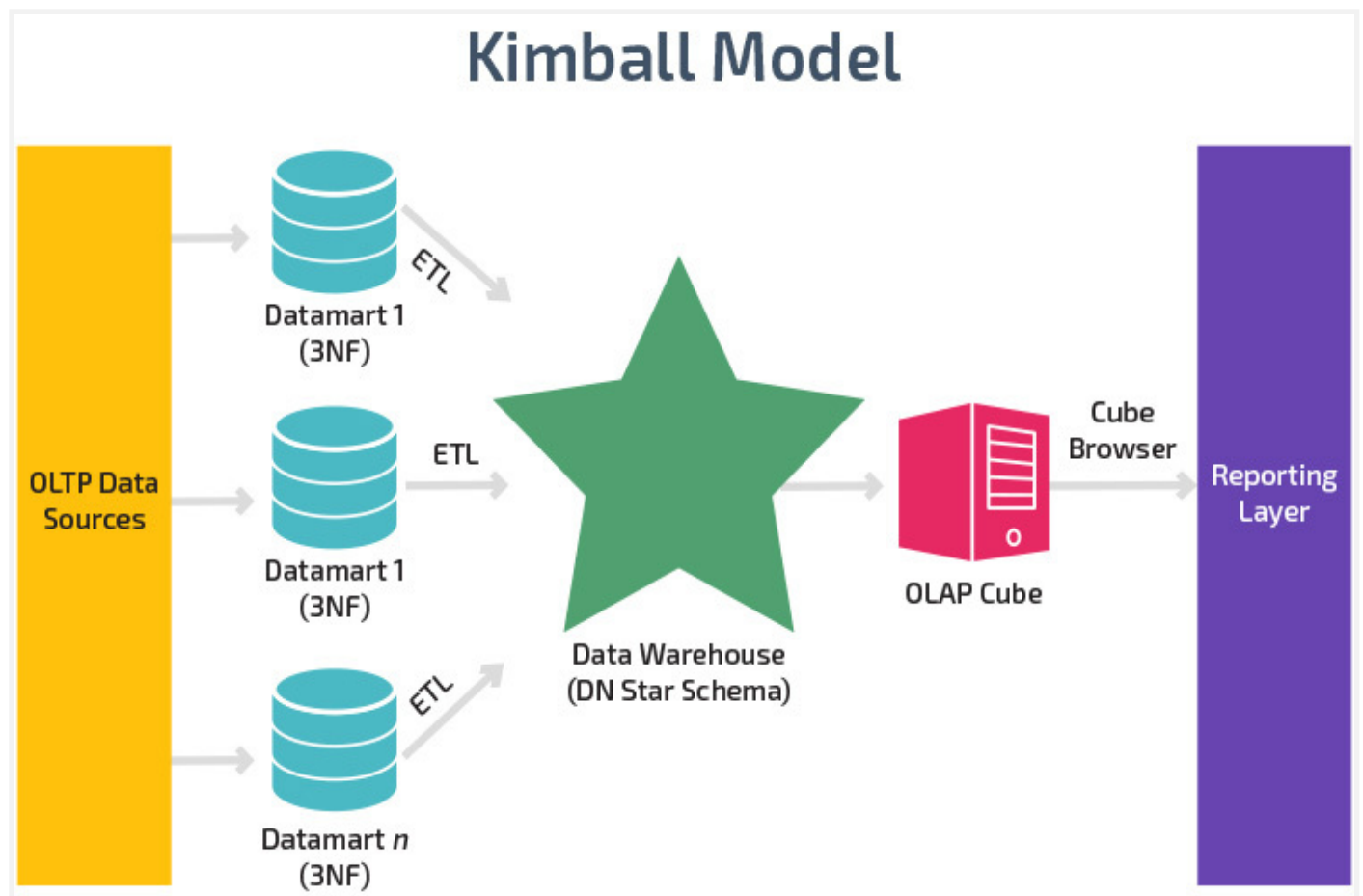
Bill Inmon's approach favours a top-down design in which the data warehouse is the centralized data repository and the most important component of an organization's data systems.

The Inmon approach first builds the centralized corporate data model, and the data warehouse is seen as the physical representation of this model. Dimensional data marts related to specific business lines can be created from the data warehouse when they are needed.



In the Inmon model, data in the data warehouse is integrated, meaning the data warehouse is the source of the data that ends up in the different data marts. This ensures data integrity and consistency across the organization.

Ralph Kimball's data warehouse design starts with the most important business processes. In this approach, an organization creates data marts that aggregate relevant data around subject-specific areas. The data warehouse is the combination of the organization's individual data marts.



With the Kimball approach, the data warehouse is the conglomerate of a number of data marts. This is in contrast to Inmon's approach, which creates data marts based on information in the warehouse. As Kimball said in 1997, "the data warehouse is nothing more than the union of all data marts."*

* Quoted from Kimball's book, "The Data Warehouse Lifecycle Toolkit".

Data Marts vs. Centralized Data Warehouse: Use Cases

The following use cases highlight some examples of when to use each approach to data warehousing.

Data Marts Use Cases

- Marketing analysis and reporting favor a data mart approach because these activities are typically performed in a specialized business unit, and do not require enterprise-wide data.
- A financial analyst can use a finance data mart to carry out financial reporting.

Centralized Data Warehouse Use Cases

- A company considering an expansion needs to incorporate data from a variety of data sources across the organization to come to an informed decision. This requires a data warehouse that aggregates data from sales, marketing, store management, customer loyalty, supply chains, etc.
- Many factors drive profitability at an insurance company. An insurance company reporting on its profits needs a centralized data warehouse to combine information from its claims department, sales, customer demographics, investments, and other areas.

Are Data Marts Still Relevant in a Cloud Architecture?

Organizations that want to make data-driven decisions are faced with a challenge—when should they use data marts versus data warehouses to analyze and report on the data they collect?

Data marts can guide tactical decisions at a departmental level while data warehouses guide high-level strategic business decisions by providing a consolidated view of all organizational data.

There are two approaches to this challenge that reflect the classic Bill Inmon versus Ralph Kimball debate:

- The first approach, based on Bill Inmon's opinion, is to build the data warehouse as the centralized repository of all enterprise data, from which data marts can be created later on to serve particular departmental needs.
- The second approach, in line with Ralph Kimball's thoughts, is to initially create separate data marts that hold aggregate data on the most important businesses processes, before merging these data marts as a data warehouse later on.

Data warehouses provide a convenient, single repository for all enterprise data, but the cost of implementing such a system on-site is much greater than building data marts. On-premise data warehouse systems also take a significant length of time to build.

However, cloud-based data warehouse services have made data warehouses much easier and quicker to set up, and cheaper to run, which negates the need for a “start small” approach that recommends starting with data marts and merging them later on into a data warehouse.

Since cloud-based data warehouse services are cost-effective, scalable, and extremely accessible, organizations of all sizes can leverage cloud infrastructure and build a centralized data warehouse first.

Learn More about Data Warehouses

- [Data Warehouse Architecture: Traditional vs. Cloud](#)
- [Data Warehouse Concepts: Traditional vs. Cloud](#)
- [Database vs. Data Warehouse](#)

- [Amazon Redshift Architecture](#)
-



Built by Panoply