



Data Warehouse Architecture

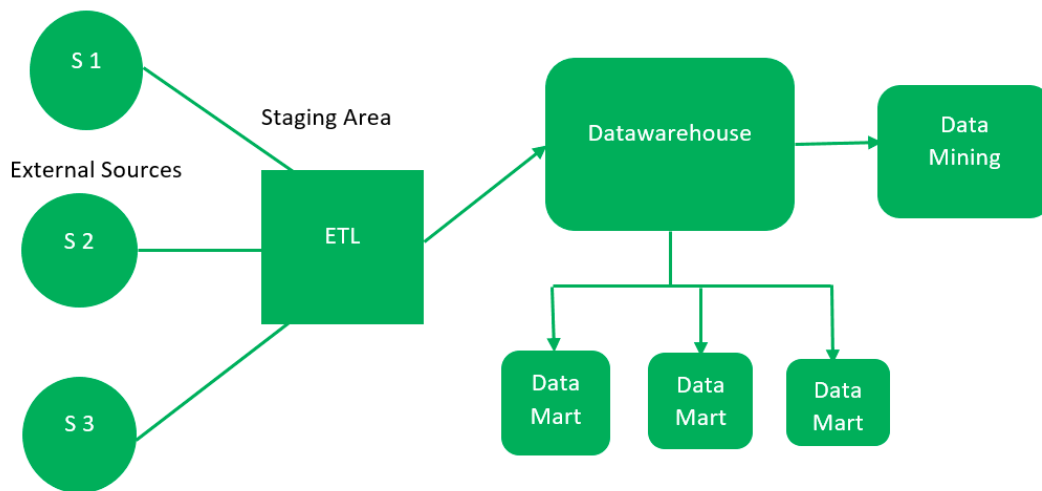
Read

Discuss

Courses

A **data-warehouse** is a heterogeneous collection of different data sources organised under a unified schema. There are 2 approaches for constructing data-warehouse: Top-down approach and Bottom-up approach are explained as below.

1. Top-down approach:



The essential components are discussed below:



1. External Sources –

External source is a source from where data is collected irrespective of the type of data. Data can be structured, semi structured and unstructured as well.



2. Stage Area –

Since the data, extracted from the external sources does not follow a particular format, so there is a need to validate this data to load into datawarehouse. For this purpose, it is recommended to use **ETL** tool.

- **E(Extracted)**: Data is extracted from External data source.
- **T(Transform)**: Data is transformed into the standard format.
- **L(Load)**: Data is loaded into datawarehouse after transforming it into the standard format.

3. Data-warehouse –

After cleansing of data, it is stored in the datawarehouse as central repository. It actually stores the meta data and the actual data gets stored in the data marts. **Note** that datawarehouse stores the data in its purest form in this top-down approach.

4. Data Marts –

Data mart is also a part of storage component. It stores the information of a particular function of an organisation which is handled by single authority. There can be as many number of data marts in an organisation depending upon the functions. We can also say that data mart contains subset of the data stored in datawarehouse.

5. Data Mining –

The practice of analysing the big data present in datawarehouse is data mining. It is used to find the hidden patterns that are present in the database or in datawarehouse with the help of algorithm of data mining.

This approach is defined by **Inmon** as – datawarehouse as a central repository for the complete organisation and data marts are created from it after the complete datawarehouse has been created.

Advantages of Top-Down Approach –

1. Since the data marts are created from the datawarehouse, provides consistent dimensional view of data marts.
2. Also, this model is considered as the strongest model for business changes. That's why, big organisations prefer to follow this approach.

3. Creating data mart from datawarehouse is easy.
4. Improved data consistency: The top-down approach promotes data consistency by ensuring that all data marts are sourced from a common data warehouse. This ensures that all data is standardized, reducing the risk of errors and inconsistencies in reporting.
5. Easier maintenance: Since all data marts are sourced from a central data warehouse, it is easier to maintain and update the data in a top-down approach. Changes can be made to the data warehouse, and those changes will automatically propagate to all the data marts that rely on it.
6. Better scalability: The top-down approach is highly scalable, allowing organizations to add new data marts as needed without disrupting the existing infrastructure. This is particularly important for organizations that are experiencing rapid growth or have evolving business needs.
7. Improved governance: The top-down approach facilitates better governance by enabling centralized control of data access, security, and quality. This ensures that all data is managed consistently and that it meets the organization's standards for quality and compliance.
8. Reduced duplication: The top-down approach reduces data duplication by ensuring that data is stored only once in the data warehouse. This saves storage space and reduces the risk of data inconsistencies.
9. Better reporting: The top-down approach enables better reporting by providing a consistent view of data across all data marts. This makes it easier to create accurate and timely reports, which can improve decision-making and drive better business outcomes.
10. Better data integration: The top-down approach enables better data integration by ensuring that all data marts are sourced from a common data warehouse. This makes it easier to integrate data from different sources and provides a more complete view of the organization's data.

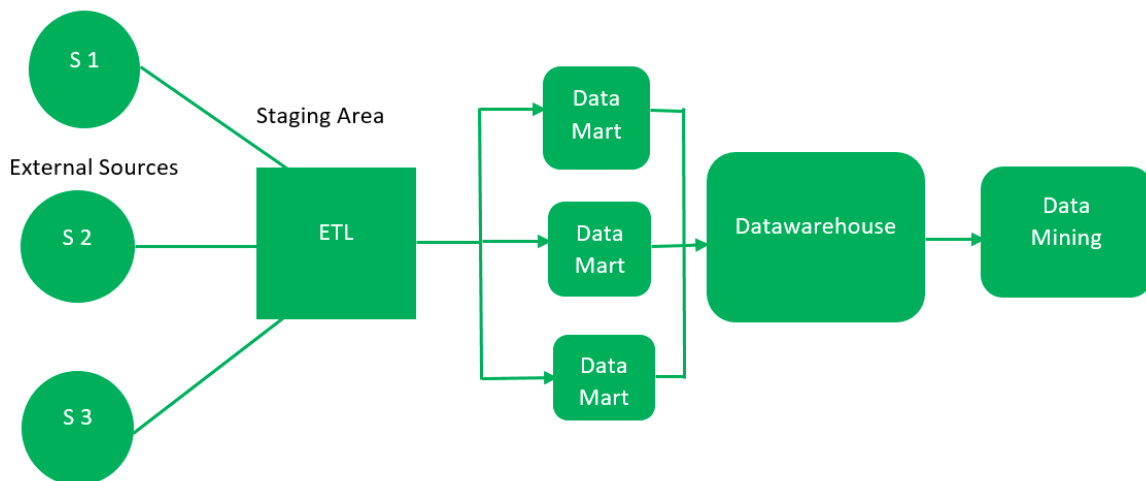
Disadvantages of Top-Down Approach –

1. The cost, time taken in designing and its maintenance is very high.
2. Complexity: The top-down approach can be complex to implement and maintain, particularly for large organizations with complex data needs. The design and implementation of the data warehouse and data marts can be time-consuming and costly.
3. Lack of flexibility: The top-down approach may not be suitable for organizations that require a high degree of flexibility in their data reporting and analysis. Since the

design of the data warehouse and data marts is pre-determined, it may not be possible to adapt to new or changing business requirements.

4. Limited user involvement: The top-down approach can be dominated by IT departments, which may lead to limited user involvement in the design and implementation process. This can result in data marts that do not meet the specific needs of business users.
5. Data latency: The top-down approach may result in data latency, particularly when data is sourced from multiple systems. This can impact the accuracy and timeliness of reporting and analysis.
6. Data ownership: The top-down approach can create challenges around data ownership and control. Since data is centralized in the data warehouse, it may not be clear who is responsible for maintaining and updating the data.
7. Cost: The top-down approach can be expensive to implement and maintain, particularly for smaller organizations that may not have the resources to invest in a large-scale data warehouse and associated data marts.
8. Integration challenges: The top-down approach may face challenges in integrating data from different sources, particularly when data is stored in different formats or structures. This can lead to data inconsistencies and inaccuracies.

2. Bottom-up approach:



1. First, the data is extracted from external sources (same as happens in top-down approach).
2. Then, the data go through the staging area (as explained above) and loaded into data marts instead of datawarehouse. The data marts are created first and provide reporting capability. It addresses a single business area.

3. These data marts are then integrated into datawarehouse.

This approach is given by **Kinball** as – data marts are created first and provides a thin view for analyses and datawarehouse is created after complete data marts have been created.

Advantages of Bottom-Up Approach –

1. As the data marts are created first, so the reports are quickly generated.
2. We can accommodate more number of data marts here and in this way datawarehouse can be extended.
3. Also, the cost and time taken in designing this model is low comparatively.
4. Incremental development: The bottom-up approach supports incremental development, allowing for the creation of data marts one at a time. This allows for quick wins and incremental improvements in data reporting and analysis.
5. User involvement: The bottom-up approach encourages user involvement in the design and implementation process. Business users can provide feedback on the data marts and reports, helping to ensure that the data marts meet their specific needs.
6. Flexibility: The bottom-up approach is more flexible than the top-down approach, as it allows for the creation of data marts based on specific business needs. This approach can be particularly useful for organizations that require a high degree of flexibility in their reporting and analysis.
7. Faster time to value: The bottom-up approach can deliver faster time to value, as the data marts can be created more quickly than a centralized data warehouse. This can be particularly useful for smaller organizations with limited resources.
8. Reduced risk: The bottom-up approach reduces the risk of failure, as data marts can be tested and refined before being incorporated into a larger data warehouse. This approach can also help to identify and address potential data quality issues early in the process.
9. Scalability: The bottom-up approach can be scaled up over time, as new data marts can be added as needed. This approach can be particularly useful for organizations that are growing rapidly or undergoing significant change.
10. Data ownership: The bottom-up approach can help to clarify data ownership and control, as each data mart is typically owned and managed by a specific business unit. This can help to ensure that data is accurate and up-to-date, and that it is being used in a consistent and appropriate way across the organization.

Disadvantage of Bottom-Up Approach –

1. This model is not strong as top-down approach as dimensional view of data marts is not consistent as it is in above approach.
2. Data silos: The bottom-up approach can lead to the creation of data silos, where different business units create their own data marts without considering the needs of other parts of the organization. This can lead to inconsistencies and redundancies in the data, as well as difficulties in integrating data across the organization.
3. Integration challenges: Because the bottom-up approach relies on the integration of multiple data marts, it can be more difficult to integrate data from different sources and ensure consistency across the organization. This can lead to issues with data quality and accuracy.
4. Duplication of effort: In a bottom-up approach, different business units may duplicate effort by creating their own data marts with similar or overlapping data. This can lead to inefficiencies and higher costs in data management.
5. Lack of enterprise-wide view: The bottom-up approach can result in a lack of enterprise-wide view, as data marts are typically designed to meet the needs of specific business units rather than the organization as a whole. This can make it difficult to gain a comprehensive understanding of the organization's data and business processes.
6. Complexity: The bottom-up approach can be more complex than the top-down approach, as it involves the integration of multiple data marts with varying levels of complexity and granularity. This can make it more difficult to manage and maintain the data warehouse over time.
7. Risk of inconsistency: Because the bottom-up approach allows for the creation of data marts with different structures and granularities, there is a risk of inconsistency in the data. This can make it difficult to compare data across different parts of the organization or to ensure that reports are accurate and reliable.

Unlock the Power of Placement Preparation!

Feeling lost in OS, DBMS, CN, SQL, and DSA chaos? Our [Complete Interview Preparation](#) Course is the ultimate guide to conquer placements. Trusted by over 100,000+ geeks, this course is your roadmap to interview triumph.

Ready to dive in? Explore our Free Demo Content and join our [Complete Interview Preparation](#) course.