

Manager's Guide to Data Warehousing Tools

By [SearchCIO.in](https://searchcio.in)

Data warehousing tools help organizations build an information warehouse, which in turn, provides the base to perform refined reporting and analytics using means such as business intelligence (BI).

This Manager's Guide to [data warehousing tools](#) has been divided into seven easy to digest topics as below:

- **Data warehousing tools: Definition and scope**
- **Trends in the data warehousing tools market**
- **Benefits of deploying data warehousing tools**
- **Challenges of deploying data warehousing tools**
- **DW tools: Key evaluation considerations**
- **DW vendors and the tools they offer**
- **Further reading**

Data warehousing tools: Definition and scope

Organizations generate two types of data:

- **Operational data:** Generated for carrying out day-to-day operations.
- **Historical data:** Over time, data generated regularly in the past needs to be retained for reporting and analysis without impacting the operational data systems. Such data is, therefore, held separately in a data warehouse. Data warehousing tools help organizations build such information warehouses.

A data warehouse can be useful for applications such as forecasting, trend analysis, analysis of compliance issues, inventory management, competition analysis, etc.

A data warehouse comprises three layers:

- **Staging layer** is used to store raw data.
- **Integration layer** integrates data.
- **Access layer** makes data available to end-users.

It typically receives operational data by one of the following means:

- **By using multiple tools developed in-house or**

purchased off the shelf: This approach creates the possibility of 'key' business rule being lost. For example, if data transformation occurs at multiple locations, developer attempting to modify existing transformation routine might overlook one or more location(s).

➤ **By using ETL (extract, transform, and load) tools:** These tools ensure that data with equivalent interpretation from different operational systems has a uniform representation before being stored in the data warehouse.

For instance, "Full-Time" and "Part-Time" might be used by one operational system to represent status of an employee while another may use "FT" or "PT". ETL tools remove such inconsistencies.

Trends in the data warehousing tools market

- **Data warehousing appliances**: Vendors are increasingly offering appliances to help organizations leverage their existing hardware instead of procuring proprietary hardware. This may reduce the deployment time and cost.
- **Fusion of structured and unstructured data**: DW tool vendors have been improving their text tagging and annotation mechanisms that integrate unstructured and structured data.

This trend must be viewed in the backdrop of the regulatory compliance pressure on the organizations with respect to dealing with and accountability of unstructured data. Unstructured data typically comprises the data found in e-mails, email attachments, and loose files.

- **Cloud-based data warehousing tools** have been gaining popularity, specially in the SMB segment.

Benefits of deploying data warehousing tools

A DW tool helps create a data warehouse, which once built, offers the following benefits:

- **Central repository with 'consistent' data:** Data from multiple data sources is pooled into a central repository to create a standard data representation. Inconsistencies are flagged and resolved before it enters data warehouse. As a result, reporting and analysis become robust.

- **Faster data retrieval:** Since data warehouses and operational systems are different entities, faster data retrieval from the former is doable without impacting performance of the latter.

- **Lower TCO in case of a cloud-based DW tool:**
 - i. **Handling of sudden, short term needs:** Data used for this type of analysis may not have to be retained from a historical perspective. Organizations can then adopt the cloud model to create data marts, use these data marts for the duration needed, and then cancel the subscription without worrying about hardware or software costs.

 - ii. **Pay-as-you-go model:** Business units can pay monthly usage fees from their op-ex instead of going through cumbersome cap-ex approvals.

Challenges of deploying data warehousing tools

Data warehousing projects can be ridden with the following issues:

- Initial 'go live' time is long.
- Hardware costs are high. They may be marginally lower if the chosen vendor also offers a DW appliance.
- Software licensing costs are high.
- In case of cloud-based DW tools, subscription fee is incurred but corresponding benefits are not seen if the DW tool in question is not designed properly to handle analysis-heavy requirements.

DW tools: Key evaluation considerations

Before commencing on a DW tool evaluation, ensure that there's a top-level agreement on the need for a data warehouse and accompanying costs.

Further, there needs to be a reasonable degree of standardization amongst business units with respect to data and associated policies and procedures. This will ease the overall process of choosing the right tool for your data warehousing needs. Once that's in place, consider the following 'key' aspects while evaluating a tool.

➤ **Hardware requirements –**

i. In case you are building a data warehouse tool in-house, buy a storage server with sufficient space for current and future storage needs; also pay attention to aspects such as high processing speed and memory.

ii. Assess if the data warehousing tool vendor offers an appliance that allows you to use your existing hardware. This can be a significant cost buster.

➤ Following **data warehousing architecture options** may be evaluated based on what you expect the tool to deliver:

i. Central data warehouse which handles the analytical

workload as well.

ii. Central data warehouse but analytical workloads can be performed in small, logical data marts.

iii. Operational data store which receives information from operational systems and passes it to the actual data warehouse which stores the data in a normalized way to reduce data redundancy. Data marts may be created for analytics.

A few other important aspects to evaluate include:

- Can the data warehousing tool obtain data from multiple operational sources seamlessly in an ongoing manner instead of starting from ground zero?
- Evaluate the ETL (extract, transform and load) abilities of the data warehousing tool. Set ETL rules in such a way that only the good quality and consistent data goes into the data warehouse.
- Does the data warehousing tool allow easy extendibility?
- How easy is it to customize the DW tool for your domain?
- Are there built-in data auditing capabilities offered?

DW vendors and the tools they offer

Data warehousing vendor	DW tool	What the tool features
IBM	InfoSphere Warehouse	Offered in five flavors ranging from as broad as enterprise-level to as granular as developer-level. Warehouse packs are offered separately to reduce deployment cost.
Informatica	Informatica Enterprise Data Warehousing solution	Data integration platform that helps deliver trusted, relevant data on-premise or in the cloud. It supports both business and IT roles involved in the data integration and management process.
Microsoft	SQL Server Integration Services (SSIS)	Component of Microsoft SQL Server that has replaced Data Transformation Services. Provides a platform for data integration, data warehouse management and workflow applications.
Oracle	Oracle Warehouse builder	Built into every Oracle Database 11g, it's a solution that facilitates data warehousing and integration. Offers cube loading, partitioning, metadata management, integration with third-party data cleansing products.

Further reading

Definition: [Data warehouse or information warehouse](#)

NEWS: [Mega-vendors dominate Gartner's Magic quadrants](#)

News: [DW technologies to evolve rapidly, says Gartner](#)

Tip: [Data warehousing best practices: Part I](#)

Tip: [Data warehousing best practices: Part II](#)

Tip: [Avoid data warehouse project failures with these tips](#)

Tip: [Making the business case for an enterprise DW](#)

Tip: [6 data warehouse design mistakes to avoid](#)

News: [2010 data warehouse appliance guide](#)