Data Warehouse

A Data Warehouse (DW) is a relational database that is designed for query and analysis rather than transaction processing. It includes historical data derived from transaction data from single and multiple sources.

A Data Warehouse provides integrated, enterprise-wide, historical data and focuses on providing support for decision-makers for data modelling and analysis.

A Data Warehouse is a group of data specific to the entire organization, not only to a particular group of users.

It is not used for daily operations and transaction processing but used for making decisions.

Characteristics and Functions of Data warehouse

Data warehouse can be controlled when the user has a shared way of explaining the trends that are introduced as specific subject.
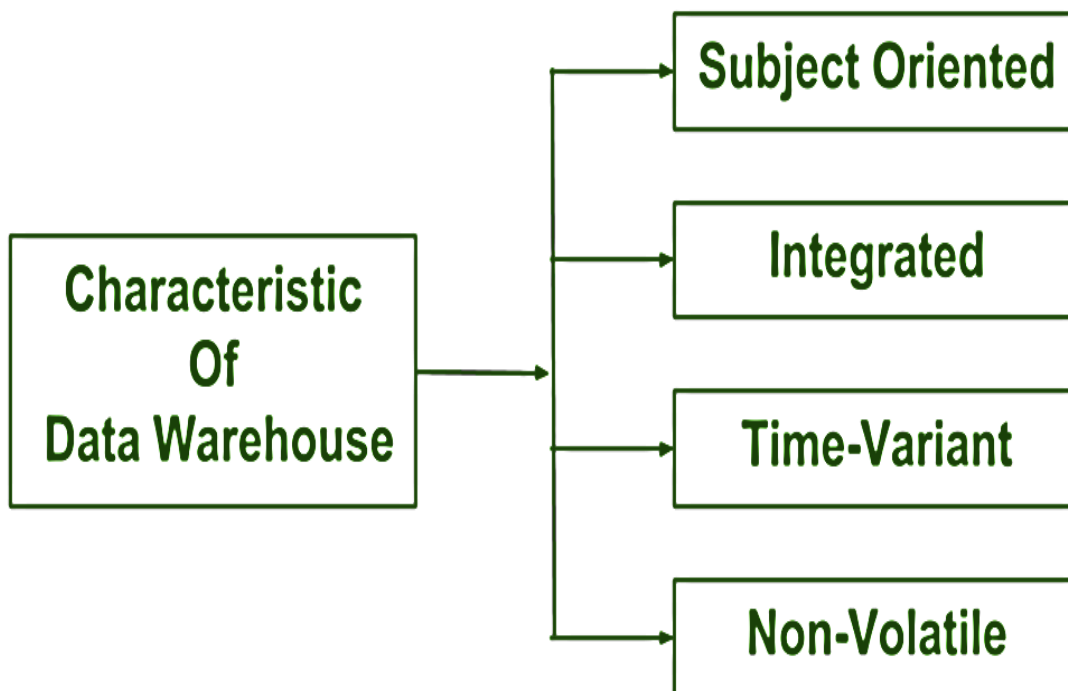
Subject-oriented –

- A data warehouse is always a subject oriented as it delivers information about a theme instead of organization's current operations.
- It can be achieved on specific theme.
- That means the data warehousing process is proposed to handle with a specific theme which is more defined.
- These themes can be sales, distributions, marketing etc.

- A data warehouse never put emphasis only current operations.
- Instead, it focuses on demonstrating and analysis of data to make various decision.
- It also delivers an easy and precise demonstration around particular theme by eliminating data which is not required to make the decisions.

Integrated :

It is somewhere same as subject orientation which is made in a reliable format. Integration means founding a shared entity to scale the all similar data from the different databases. The data also required to be resided into various data warehouse in shared and generally granted manner.

A data warehouse is built by integrating data from various sources of data such that a mainframe and a relational database. In addition, it must have reliable naming conventions, format and codes. Integration of data warehouse benefits in effective analysis of data. Reliability in naming conventions, column scaling, encoding structure etc. should be confirmed. Integration of data warehouse handles various subject related warehouse.

Time-Variant

Historical information is kept in a data warehouse. For example, one can retrieve files from 3 months, 6 months, 12 months, or even previous data from a data warehouse. These variations with a transactions system, where often only the most current file is kept.

this data is maintained via **different intervals of time such as weekly, monthly, or annually etc**. It founds various time limit which are structured between the large datasets and are held in online transaction process (OLTP). The time limits for data warehouse is wide-ranged than that of operational systems. The data resided in data warehouse is predictable with a specific interval of time and delivers information from the historical perspective. It comprises elements of time explicitly or implicitly. Another feature of time-variance is that once data is stored in the data warehouse then it cannot be modified, alter, or updated.

Non-Volatile

As the name defines the data resided in data warehouse is permanent. It also means that data is not erased or deleted when new data is inserted. It includes the mammoth quantity of data that is inserted into modification between the selected quantity on logical business. It evaluates the analysis within the technologies of warehouse.

In this, data is read-only and refreshed at particular intervals. This is beneficial in analysing historical data and in comprehension the functionality. It does not need transaction process, recapture and concurrency control mechanism. Functionalities such as delete, update, and

insert that are done in an operational application are lost in data warehouse environment. Two types of data operations done in the data.

Components of data warehouse:

- ❖ Load manager
- ❖ Query manager
- ❖ Warehouse manager
- ❖ End-user access tool
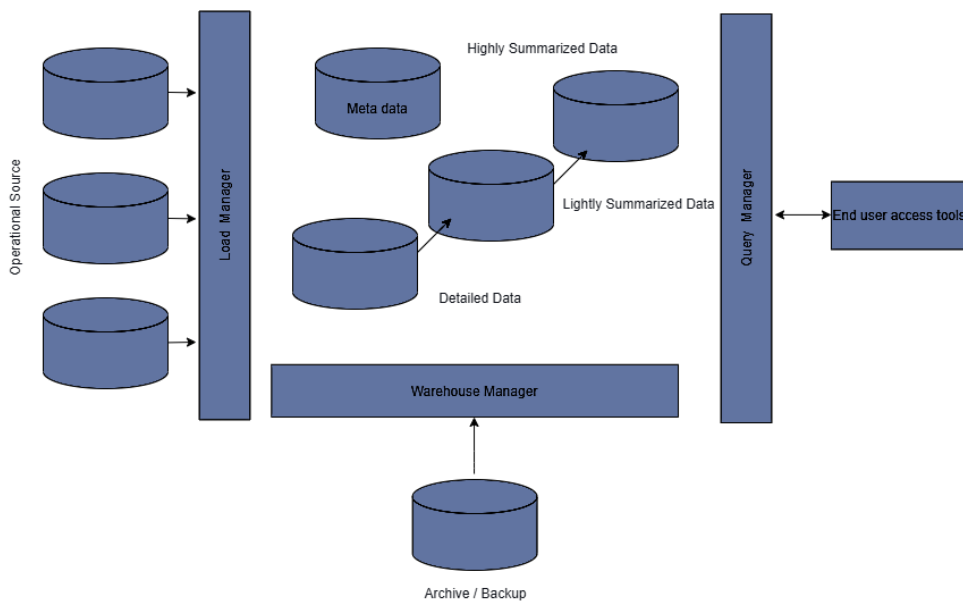
Development of data warehouse approach

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:
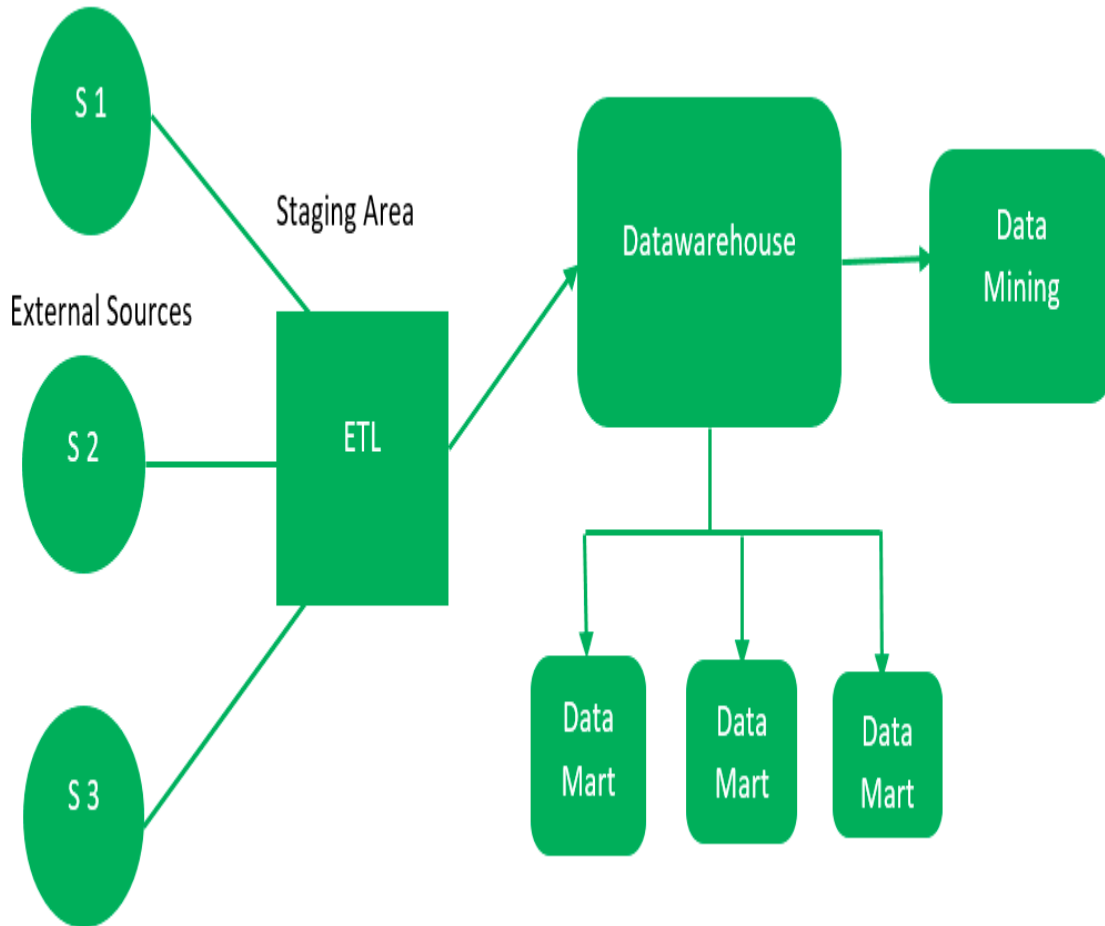
The essential components are discussed below:

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 data ware house. 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 data warehouse after transforming it into the standard format.

3. **Data-warehouse –**
   After cleansing of data, it is stored in the data warehouse as central repository. It actually stores the meta data and the actual data gets stored in the data marts. **Note** that data warehouse stores the data in its purest form in this top-down approach.
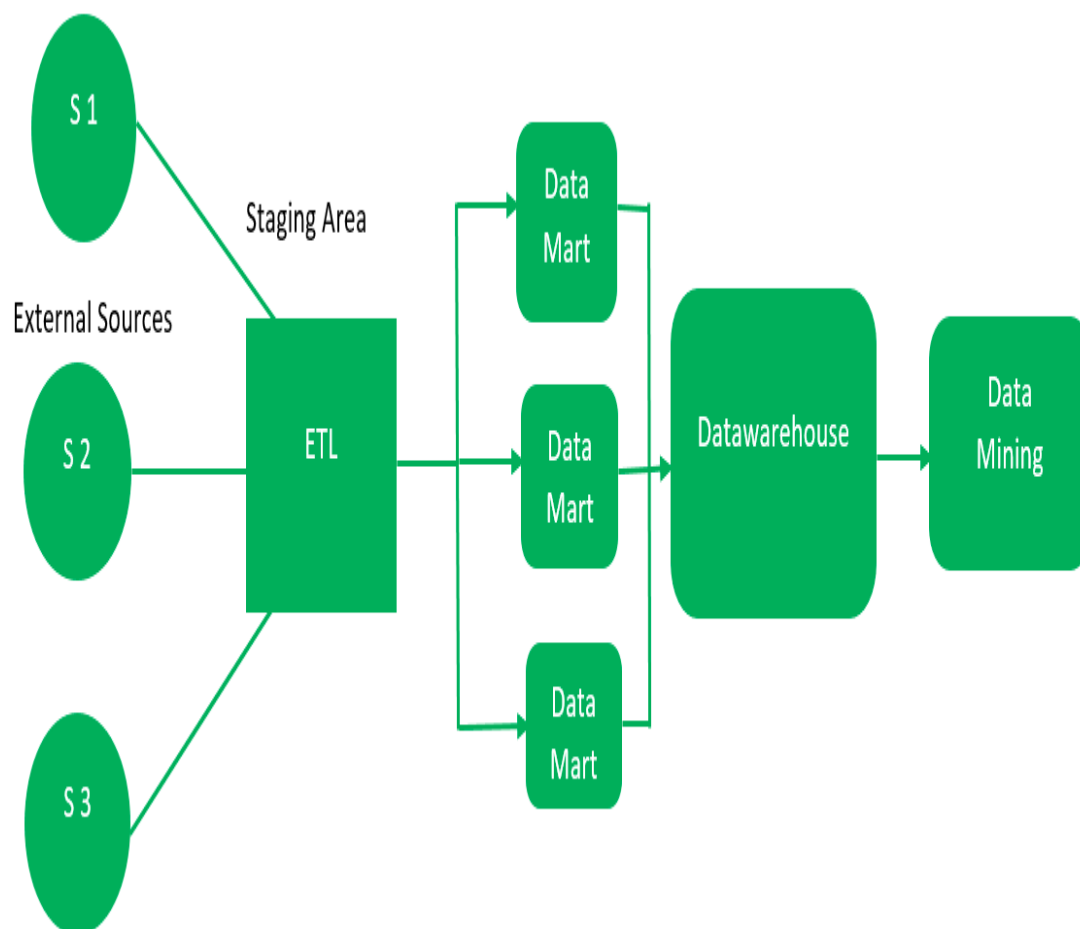
4. **Data Marts –**
   Data mart is also a part of storage component. It stores the information of a particular function of an organisation which is handled by single authority. There can be as many number of data marts in an organisation depending upon the functions. We can also say that data mart contains subset of the data stored in datawarehouse.

5. **Data Mining –**
   The practice of analysing the big data present in data warehouse 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.

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 data warehouse. The data marts are created first and provide reporting capability. It addresses a single business area.

3. These data marts are then integrated into data warehouse.

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.

16 Steps on Your Business Intelligence Roadmap

There are many benefits from implementing a BI system in your company. Once you identified the potential behind it, it is time to start planning its application. When thinking about BI we are not only talking about choosing an online data analysis tool. There are several other steps that need to be taken to ensure success

1 Go into the process with eyes wide open

When you have the right [BI solutions](#), it is easy to identify trends, pitfalls, and opportunities early on. But implementing the right tool isn't always easy. Actually, it usually isn't. We are going to be honest here, even the best software needs some initial heavy lifting to maximize its potential. If you go in with the right mindset you will be prepared to address issues like complicated data problems, change management resistance, waning sponsorship, IT reluctance, and user adoption challenges.

2. Determine stakeholder objectives

Odds are everyone at your organization could benefit from increased data access and insights. That doesn't mean they are all key stakeholders. Right off the bat, you must determine who your key stakeholders are. Then find out what they need: visible and vocal executive sponsorship is a must. Gathering and setting executive team expectations early is paramount

3. Choose a sponsor

While a business intelligence strategy should include multiple stakeholders, it is imperative to have a sponsor to spearhead the implementation. It may be tempting to place the Chief Information Officer (CIO) or Chief Technical Officer (CTO). This is usually not the best approach. It should be sponsored by an executive who has bottom-line responsibility, a broad picture of the organization's strategy and goal.

4. BI is not just a technology initiative

We are going to repeat ourselves a bit here. Because it is *that* important. To succeed, a deployment must have the support of key business areas, from the get-go. IT should be involved to ensure governance, knowledge transfer, data integrity, and the actual implementation. But every stakeholder and their respective business areas should also be involved throughout the process

5. Employ a Chief Data Officer (CDO)

Big data guru Bernard Marr [wrote about](#) *The Rise of Chief Data Officers*. In the article, he pointed to a pretty fascinating trend: "Experian has predicted that the CDO position will become a standard senior board-level role by 2020, bringing the conversation around data gathering, management, optimization, and security to the C-level." We love that data is moving permanently into the C-Suit

6. Assess the current situation

As we have already stated: usually a deployment isn't quick or easy. There is a lot of work to do on the front end. One of the biggest sections of a business intelligence roadmap should be assessing the current situation. Now that you have all the right stakeholders at the table the next step is analyzing the current software stack, and the processes and organizational structures surrounding.

7. Define a budget

Once you have defined the current situation of the business, it is time to think about a budget. Developing an accurate budget is a key step in the process of building a successful business intelligence strategy. Budgeting allows you to smartly allocate your resources to make sure you have everything you need to kick it off.

8. Think of security, privacy, and compliance

Before going all-in with data collection, cleaning, and analysis, it is important to consider the topics of security, privacy, and most importantly, compliance. Businesses deal with massive amounts of data from their users that can be sensitive and needs to be protected. Massive data breaches are a constant topic of concern and it has led to the implementation of [various legislations](#) to regulate it

9. Clean the data

10. Develop a "Data Dictionary"

With Agile development, extensive documentation has become a faux-pas. Large data dictionaries can be cumbersome and hard to keep updated. That said, for business intelligence to succeed there needs to be at least a consensus on data definitions and business calculations. The lack of agreement on definitions is a widespread problem in companies today.

11. Ensure data literacy

Paired to a well-thought data dictionary, another action you need to take to ensure your business intelligence strategy is successful is the democratization of data across the entire organization.

13. Choose the right tool/partner for your business

choose a BI software/partner. Yes, you are this far along in your business intelligence roadmap and you don't even have a tool yet. By preparing properly through steps 1-12 you

will be best suited to find the right tool and implement it successfully. During this process, you will need to choose and perform a [cloud vs on-premise comparison](). You also need to make sure to choose a solution that can start small but easily scale as your company and needs grow.

14. Rely on interactive data visualizations

Although it is not a specific step, using intuitive data visualizations still goes into the BI strategy framework as an important element to consider when choosing the right tool for your company. For decades now, data analytics has been considered a segregated task. People often think of numbers and infinite excel sheets that are almost impossible to understand.

15. Think about possible roadblocks and find solutions

clear understanding of the processes and measures you need to take for the successful implementation of your business intelligence strategy. But, as with any other business scenario, it is not without problems. In order to avoid being caught by surprise, a good practice when building a BI roadmap is to think of possible roadblocks.

16. Pursue a phased approach

Rome wasn't built in a day: neither will your BI. A successful BI strategy takes an iterative approach. Think "actionable" and take baby steps. Choose a few KPIs and build a few [business dashboards]() as examples. Gather feedback. Repeat again with new releases every few weeks. Continuously ask yourself what is working and what stakeholders are benefiting.

The key elements of a business intelligence architecture are:
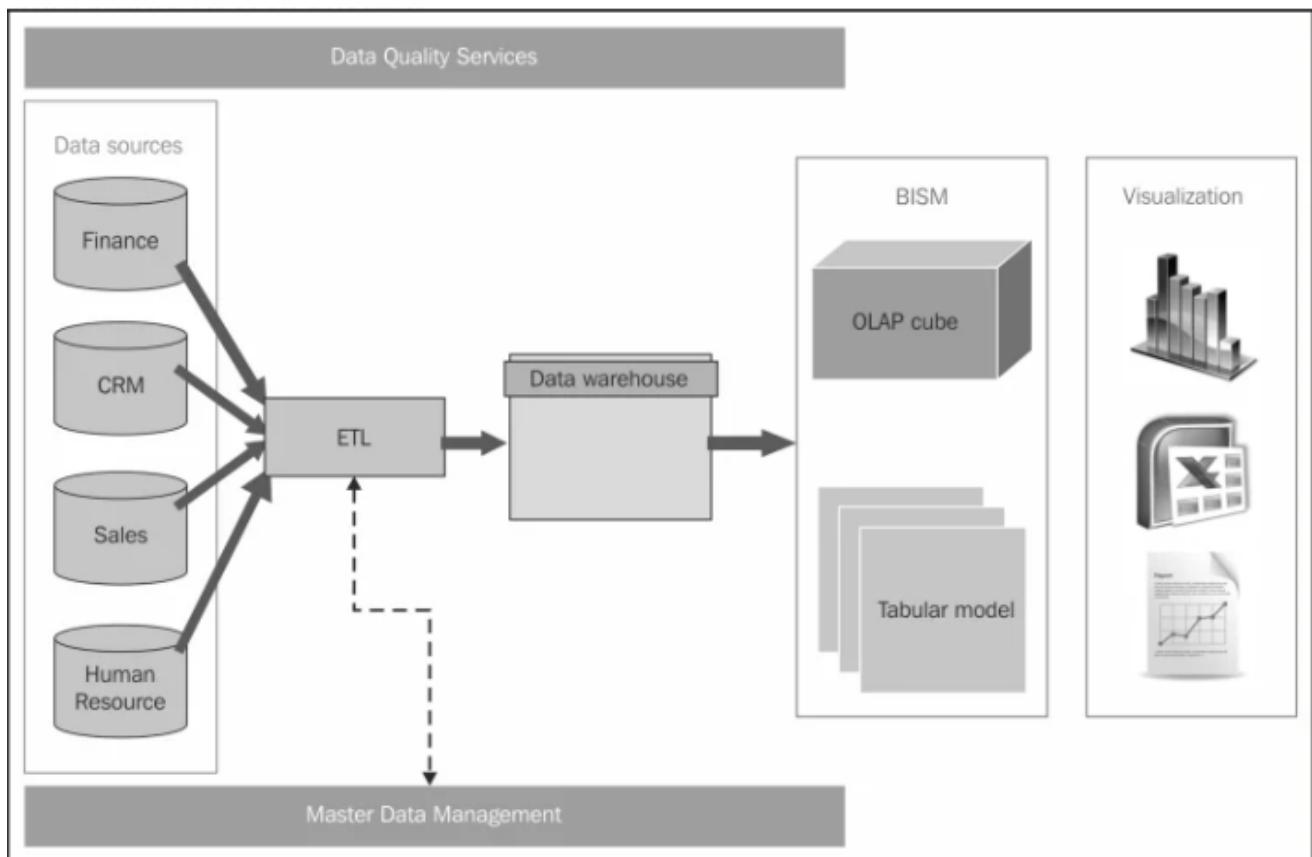
- Source systems
- ETL process
- Data modelling
- Data warehouse
- Enterprise information management (EIM)
- Appliance systems
- Tools and technologies

Source Systems – Transaction Processing Systems

This is the starting point for any BI initiative. Organization data is first created in these databases. Point to note: if you do not capture the data in the operational system, you can't analyse it.

Operational systems (OLTP) form the bulk of the data needed for the data warehousing. In addition to that, source systems may also include data from secondary sources such as market data, benchmarking data etc. Business Intelligence architecture should address all these various data sources which are of different formats and standards.

ETL Process

In an ETL process data is extracted from the operational systems and loaded into a data warehouse. ETL, which stands for Extract Transform Load, is usually done using custom solutions available in the market. IBM Websphere Data Stage, Oracle Data Integrator, Ab Initio, and Microsoft Integration Services are examples of such tools.

Data Modeling

Data modeling will help to address what exactly is needed from data sources, the format of the data, and how it will be related to other data elements. It is not feasible to extract everything from a source system as that comes with cost issues. Data modeling will help to organize the data and therefore will minimize cost of storage replication, and effort needed to build a data warehouse.

Data Warehouse

Warehouse will have data extracted from various operational systems, transformed to make the data consistent, and loaded for analysis. A data warehouse will help in achieving cross-functional analysis, summarized data, and maintaining one version of the truth across the enterprise.

Enterprise Information Management (EIM)

EMI is another BI jargon which may stump some beginners. The term usually refers to ETL tools, data modeling tools, data quality, data profiling, metadata management, and master data management.

BI Hardware

It is important to make decisions on the hardware requirements to maintain a high performance and scalable BI system. Apart from server configurations, we have data warehouse appliances to combine the server, the database, and the data storage into one system. Netezza and DATAllegro are some well-known appliances in the market.

Tools and Technologies

Another important component of business intelligence architecture is what tools and technologies to implement. It is not just the front-end UI tools, but the tools used for EIM

as well. There are cloud solutions, SaaS model, many full-fledged BI solutions (such as MSBI, Oracle BI suites, MicroStrategy and more) to choose from. BI framework should have guidelines to make decisions on what is required for the organization.
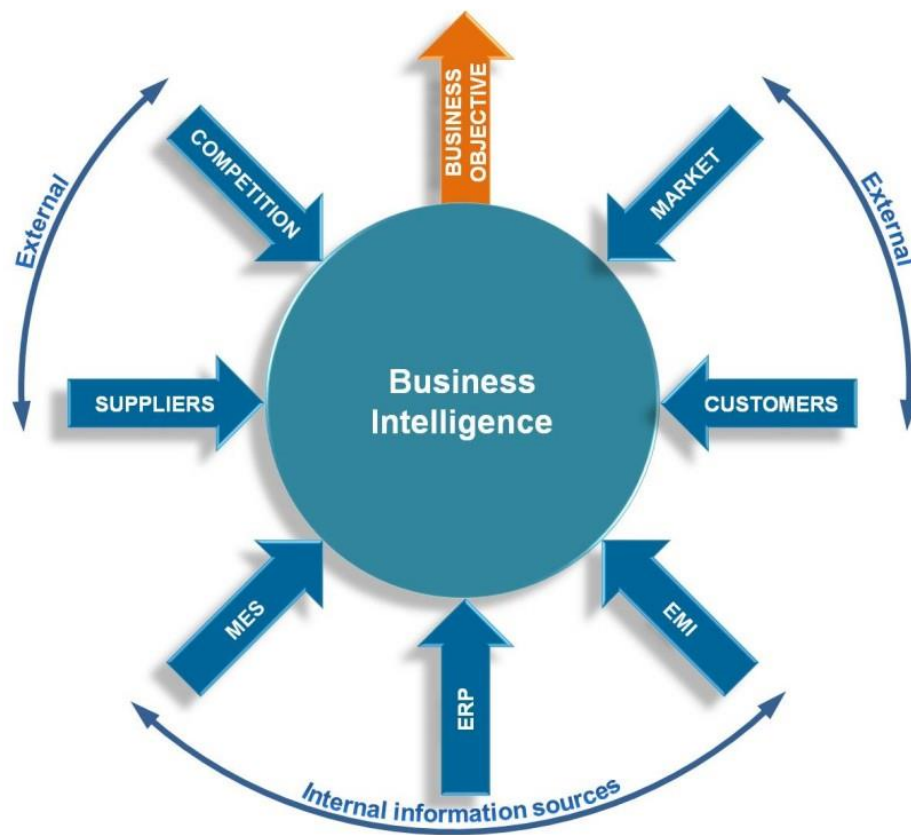
Phase 1: Setting Clear Objectives

The effective use of BI begins with an awareness that BI is fundamentally different from reporting and has strategic implications if it is used to its full potential.

At the outset of a BI project (and on a periodic basis thereafter), business leaders should define the playing field in which they mean to operate. In many organizations, the initial focus for BI is on understanding the company's existing customers, especially with respect to their buying behavior. Others may choose to zero in on sales performance and profitability by product line, division, or territory.

In most cases, BI initiatives will address multiple domains, but defining those areas of interest is a critical first step, largely because it determines how you will allocate resources, especially with respect to the second phase of the business intelligence cycle, gathering and organizing data. This is particularly important if your organization intends to enrich and extend your internal data with additional information from external sources.

Phase 2: Gathering and Organizing Data

Whatever your company's top priorities may be, it's helpful to brainstorm a list of potential questions around those topics. Ultimately, this will shed light on the kind of information you need and what you need to do to consolidate and harmonize that information.

If you're analyzing customer buying behavior, for example, you may need to aggregate information from your enterprise resource planning (ERP) system (which contains sales order transactions) and combine it with sales pipeline information from your customer relationship management (CRM) system, including sales quotes and deals that never came to fruition.

For many projects, internal data sources are likely to be sufficient. By brainstorming a list of questions in advance, you can begin to determine whether you should include data enrichment as part of your overall BI strategy. If the focus is on knowing your customer, for example, you may want to consider extending your corporate datasets with demographic details available from third-party sources.

As you consider your overall approach toward gathering and organizing the data for your BI initiative, you'll need to determine how you can best harmonize and consolidate the information and make it available to the users who will rely on your business intelligence systems for meaningful insights.

BI tools that have pre-built integration to ERP are a distinct advantage–especially if they present information in a context front-line employees can easily understand and use. Systems that require advanced database skills or custom programming will cost more to operate and ultimately generate less value because of the technical barriers that stand in the way of widespread user adoption.

Phase 3: Analyzing the Information

Analysis is where the magic happens; it's where BI distinguishes itself from business reporting in general. Business reporting is more operational in nature than BI. Operational reports often have a short-term focus, and they are used to drive the daily decisions business leaders must make. Business reports may work with real-time transactional data connected directly to the source system.

On the other hand, BI typically has a long-term focus and concerns itself with trends and patterns. BI usually involves, not real-time data, but aggregated or summarized data that may have been loaded into a data warehouse and transformed for analysis. This distinction means that the data used in BI does not necessarily have a direct connection to source systems because it doesn't need one.

Business intelligence facilitates exploration. BI makes it easy to perform ad hoc inquiries, which often prompt users to ask new questions they had never before considered.

BI makes it easy to take that next step by offering quick answers to those follow-up inquiries. It's fast, flexible, and open-ended. It helps users to explore new territory and get rapid answers to questions that challenge conventional wisdom. When leaders face "what if" scenarios, BI helps them understand their options and narrow down their choices using a data-driven approach.

This is especially valuable when BI tools are available across a broad base of users within the organization. This is referred to as "data democratization," and it can have a transformative effect because it empowers workers to better understand the forces that should impact their decisions, regardless of where they may sit within the organizational hierarchy.

Phase 4: Measuring and Monitoring KPIs

Alongside this ad hoc analysis, you'll want to leverage your BI systems to measure and monitor the key performance indicators (KPIs) that align to your organization's strategic objectives. As discussed, business intelligence differs from traditional approaches to financial reporting in its focus. Rather than focusing on short-term operational matters, BI focuses on the data at a higher-level, moving from operational thinking to a more managerial approach. Because it can aggregate high volumes of data from disparate sources, BI is naturally suited for use in tracking both financial and operational KPIs. In other words, BI has the capacity for comprehensive