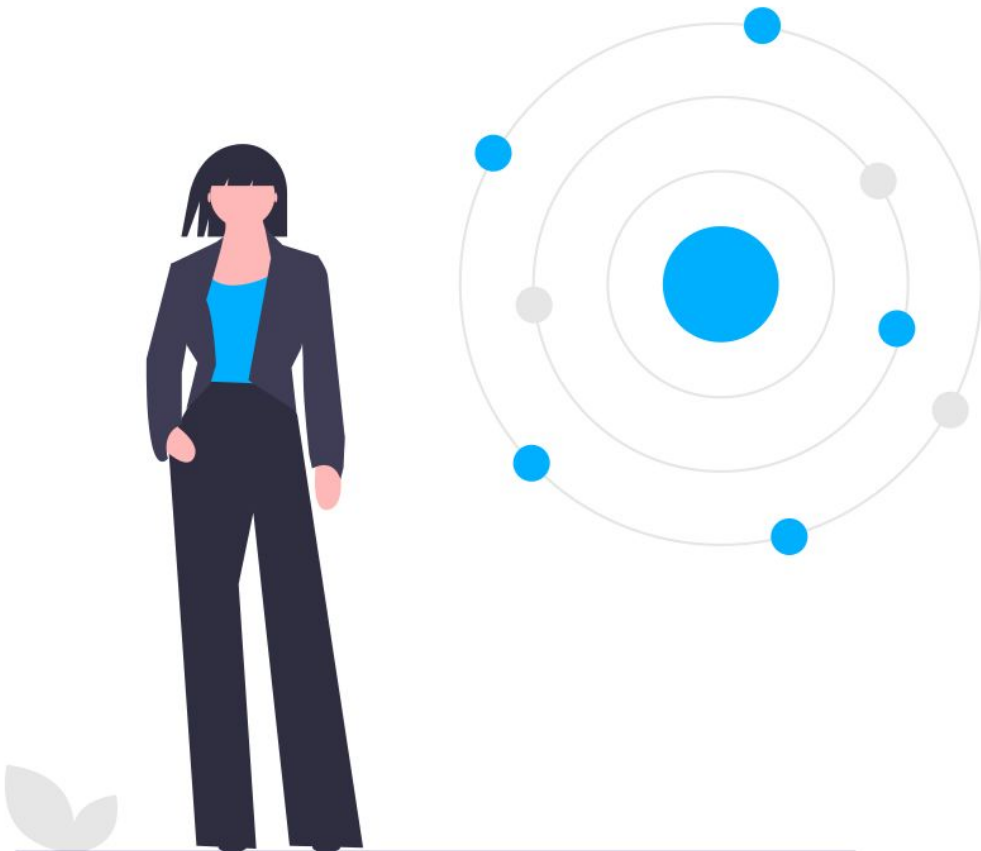


2025

Definitive Guide to Data Warehousing



Introduction

Everything You Need To Know About Data Warehousing And How It Can Help Improve Your Business!

You may have heard that data is the new gold. Does that make a data warehouse your gold reserve?
It does, in some ways.

A data warehouse is a vital element of your business intelligence. It acts as a repository that aggregates data from the multiple departments that can later be used to inform decision making. Thus, the metaphor is very real when it comes to the importance and role of a data warehouse. But things are not that simple.

For a data warehouse to serve its function one needs a well-designed architecture and a maintenance strategy. How can you maintain the particularities of the different data types while still delivering worth-telling stories? We will see that over the next ten chapters.

By the end of this guide you will know the benefits and challenges of data warehouses, their different types and how they evolved through time and have actionable insights on how to start implementing your own.

The first step is to understand the information that is critical to the organisation and identify the sources of the information.

David Foster
Partner
Applied Data Science Partners
(ADSP)



Focus Areas

This whitepaper series is a practical guide to developing your understanding of Data Warehousing.

It is organised around 10 key topics:

Chapter One:

What is data warehousing?

Chapter Two:

The Benefits of Data Warehousing

Chapter Three:

Evolution of Data Warehousing

Chapter Four:

Types of Data Warehousing

Chapter Five:

How to Maintain Your Data Warehouse

Chapter Six:

Data Warehousing For Business Intelligence & Data Support Systems

Chapter Seven:

Data Warehouse FAQs

Chapter Eight:

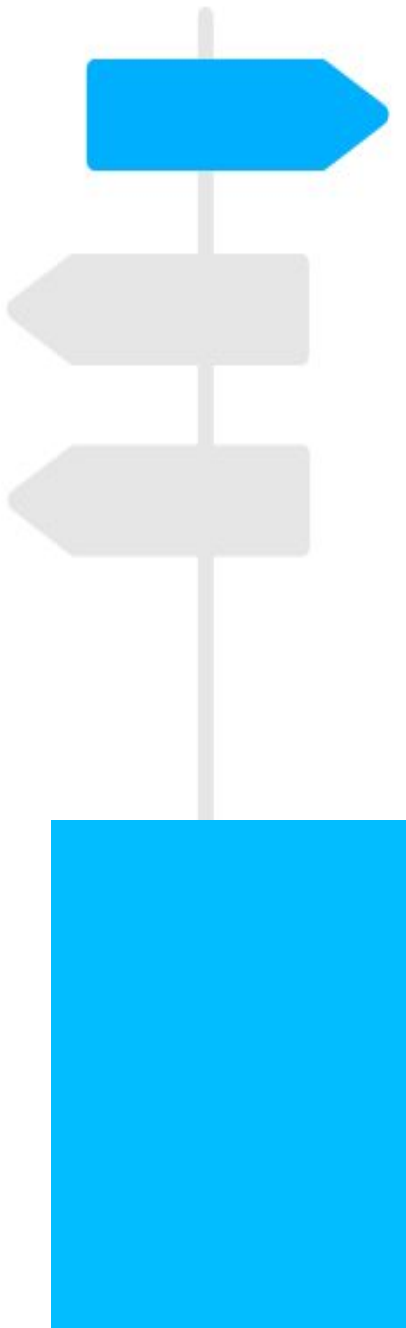
How to Implement Data Warehousing

Chapter Nine:

The Differences Between A Data Warehouse, A Database, A Data Lake & Data Mart FAQs

Chapter Ten:

Data Warehousing Tools



1 What is data warehousing?

A data warehouse is a database used for reporting and data analysis. It is a central repository of data that can be used to answer business questions. Data warehouses are used to store historical data and to support decision making.

How is it used?

Some examples of how data warehousing is used include: analysing customer behaviour, understanding product demand, and forecasting sales. Data warehouses can also be used to track changes in the business over time, such as the impact of a new marketing campaign.

Expect a diverse group of people (including business analysts, data engineers and decision makers) to interact with the data warehouse on a regular basis, attempting to extract maximum value for its own needs.

How does it work?

A typical data warehouse may include:

- A relational database. This is the backbone of the warehouse, as it is responsible for storing and managing data
- Extraction and transformation routines for querying the database
- A data science module that can derive direct statistical insights from the data
- Analytical tools for reporting and visualising these insights
- A module for advanced analytics powered by machine learning algorithms. Here, one aims at addressing complex decisions at scale

2 The Benefits of Data Warehousing

Data coming from diverse sources and different time frames that are used to inform decisions, can pose a challenge in their storage and maintenance.

Data warehousing is available because it is:

- **Subject-oriented.** A single data warehouse can tell many stories depending on who is asking
- **Integrated.** Despite its particularities, information from different sources can be combined to derive complex, cross-departmental insights
- **Non-volatile.** In contrast to a database that only holds the most recent information, a data warehouse never forgets
- **Time-variant.** Informing interventions often requires tracking change with time

Therefore a data warehouse can improve decision making by providing a consolidated view of data. The centralised storage of information comes with additional benefits in the efficiency and cost of maintaining this knowledge.

Benefits

One benefit of data warehousing is that it can help organisations make better decisions by providing them with a single, consolidated view of data. It can also help organisations save money by reducing the need for multiple data stores. Additionally, data warehousing can help organisations improve their customer service by providing them with better access to customer data.

3 Evolution of Data Warehousing

The birth of data warehousing can be traced back to the 80s, with pioneers such as Inmon introducing its definition and main principles.

That decade saw an important shift in business information systems. During that time, the main business information systems focused on optimising the efficiency of processing transactions. After the birth of the data warehouse, the ability of such systems to generate insights for decision making became more popular.

Various companies began to adopt the use of data warehousing once they realised it could help them reduce the challenge caused by storing excessive records for different business needs.

Cloud data warehousing made data warehousing accessible to everyone, regardless of size or resources. Cloud data warehousing is more flexible and scalable than traditional data warehousing, and it provides users with the ability to access their data from anywhere in the world.

With their increase in efficiency, data warehouses became capable of answering complex questions. A wide variety of applications, such as operational analytics and performance management, derives its knowledge from the broad analytics infrastructure offered by a data warehouse.

The Future of Data Warehousing

The modern era of data warehousing is seeing another revolution: machine learning and data science are vastly augmenting the capabilities of data warehouses.

This translates to two main drivers of future data warehouses:

- The complexity of insights one can derive from systems, combining such breadth of knowledge with advanced modelling and predictive algorithms, knows no limit
- The autonomous data warehouse will enable the extraction of even greater value from data while lowering costs and improving data warehouse reliability and performance

4 Types of Data Warehousing

What is the right architecture for a data warehouse?

That depends on your needs. In particular, the scale at which you will consolidate data and drive decision-making will determine which of these three types you should adopt:

Enterprise Data Warehouse (EDW)

This type provides decision support services across the enterprise. It offers a unified and versatile interface to a centralised data collection, capturing all interesting facets of enterprise functions.

Operational Data Store (ODS)

This second type of data warehousing is a complimentary element to EDW and is useful for operational reporting, controls and decision-making at the enterprise level.

The main difference between EDW and ODS is that ODS is used in real-time compared to EDW which can be reversed for tactical decision support.

Data Mart

A data mart can be seen as an ODS with a more limited scope which priorities include efficiency and completeness. Being oriented to a specific team or business line, this type of architecture can provide critical insights.

5 How To Maintain Your Data Warehouse

So you have designed, deployed and, to your great content, successfully queried your data on multiple occasions.

Are you done?

Quite the opposite. While a good design is necessary, it is not sufficient to ensure that your data warehouse will remain useful for a long time.

Maintenance is necessary for the following reasons:

- New business lines and actors will require constant refresh of permissions and may require creating new data marts
- Old metrics become outdated and new metrics emerge as your product line evolves
- The software and landscape will change

A data warehouse is solving a moving target problem and therefore needs to move itself.

6 Data Warehousing for Business Intelligence & Data Support Systems

Business Intelligence (BI) is a process for analysing data and deriving insights to help businesses make decisions. In an effective BI process, analysts and data scientists discover meaningful hypotheses and can answer them using available data.

For example, if management is asking, "how do we improve conversion rate on the website?", BI can identify a possible cause for low conversion. The cause might be lack of engagement with website content.

Within the BI system, analysts can demonstrate if engagement is really hurting conversion, and can find out which content is the root cause.

The tools and technologies that make BI possible take data stored in files, databases, data warehouses, or even on massive data lakes to run queries against that data, typically in SQL format.

Using the query results, they create reports, dashboards and visualisations to help extract insights from that data. Insights are used by executives, mid-management, and also employees in day-to-day operations for data-driven decisions.

Data warehouses applications integrate with BI tools like Tableau, Sisense, Chartio or Looker. They enable analysts using BI tools to explore the data in the data warehouse, design hypotheses, and answer them. Analysts can also leverage BI tools, and the data in the data warehouse, to create dashboards and periodic reports and keep track of key metrics.

Can you have one and not the other?

You may have business intelligence without a data warehouse (using a data lake or a database for example) but the purpose of a data warehouse is supporting BI, so no.

7 Data Warehouse FAQs

What is the difference between a data warehouse and a database?

A database is a real-time, small-scale collection of data while a data warehouse holds historical data across many business functions.

What is the difference between a data lake and a data warehouse?

A data lake is a temporary collection of disparate data while a data warehouse is a central data storage for complex decision-making.

What are the three types of data warehouses?

Enterprise Data Warehouse, operational data store and data mart.

8 How To Implement Data Warehousing

1. Determine business objectives.
2. Collect and Analyse Information.
3. Identify Core Business Processes
4. Construct a Conceptual Data Model
5. Locate Data Sources and Plan Data Transformations
6. Set tracking duration

9 The Differences Between A Data Warehouse, A Database, A Data Lake & Data Mart

Is a data warehouse the only solution to your data needs?

You may want to take a look at these different types of data storage solutions before you commit to a data warehouse.

Database

A database stores real-time information about one particular part of your business: its main job is to process the daily transactions that your company makes, e.g., recording which items have sold. Databases handle a massive volume of simple queries very quickly.

Data lake

Do you have an abundance of disparate, unstructured data that hide value but do not currently have a clear application?

Then what you need is a data lake. When organisations need low-cost storage for unformatted, unstructured data from multiple sources that they intend to use for some purpose in the future, a data lake might be the right choice.

10 Data Warehousing Tool

Which data warehousing tool you use depends on your business model. There is a plethora of cloud-based solutions you can implement, such as Amazon Redshift and integrate.io

Cloud-native data warehouses — options for moving your mission-critical data warehouse to the cloud e.g. Redshift and BigQuery.

Cloud-based ETL tools — lightweight services that help upload data or pull it directly from cloud sources, transform it, and pipe it into a data warehouse, without heavyweight planning or infrastructure e.g. Stitch and Blendo.

Cloud-based BI tools — platforms that let you connect to data warehouses, create visualizations and dashboards and share them with collaborators e.g. Tableau online and Chario.

Cloud-based data integration tools — services that help you connect to just about any application or data source, define triggers and when a specific event happens, grab the data to feed your data warehouse e.g. Zapier.

But there is one point we can't stress enough: to rip the most benefits from your platform of choice you should spend some effort on customising it to your business needs. Consider analytical dashboards to interactively mine insights and strategic dashboards for seeing the big picture.