

Data Stored Well is Data Used Well

A Prescience Decision Solutions Whitepaper

Harnessing advanced data warehousing capabilities to deliver superior business outcomes



There is no doubt that over the many years, data has evolved into an indispensable source, offering valuable insights to help make critical decisions. How important data is can be acknowledged by the fact that today it is considered equivalent to oil, soil, water and oxygen. But what is equally important is the need to store and organize data in a manner that aids seamless, accurate and efficient decision making.

Data warehousing helps in consolidating, storing and organizing data. While data warehousing has undergone significant changes in its architecture and methodology, there is an urgent need to establish a paradigm to support next-gen technologies such as big data and cloud computing along with modern analytics and reporting requirements.

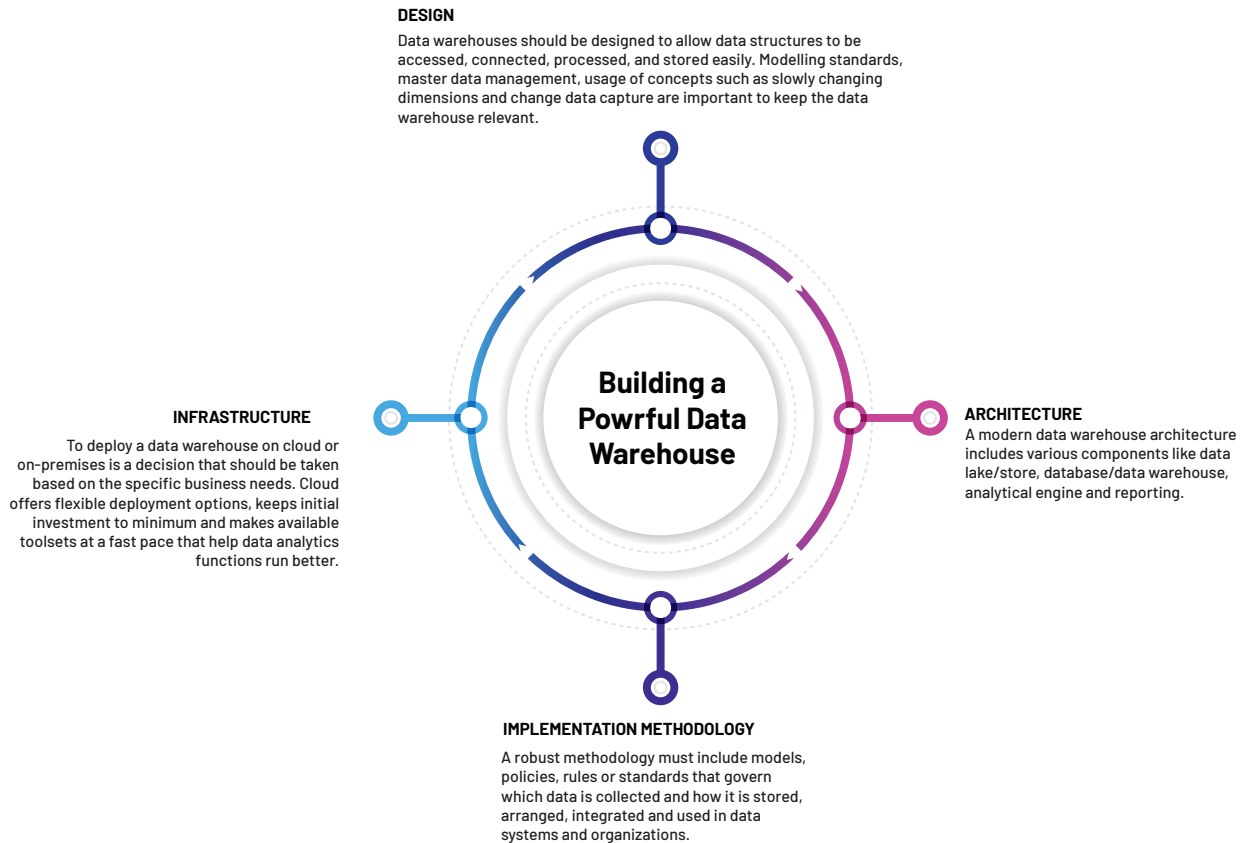
Data warehouses, when deployed effectively, are valuable in organizing data and eliminating redundancies. A well-designed warehouse can provide information in a timely manner to drive effective decision making. With a significant surge in analytical applications and prescriptive analytics, the importance of organized and clean data has definitely increased.

A modern data warehouse architecture must have the following functions to support the evolving needs of an enterprise:

- Manage and integrate both structured and unstructured data types
- Integrate support for advanced analytics processing to support new, advanced analytics use cases
- Support near-real-time or real-time access and analysis at a scale (and cost) that was not previously practical
- Move data to and from cloud services as it is with on-premises data sources and services
- Integrate transparently multiple platforms in a unified data warehouse architecture
- Support ad-hoc data / reporting requests

KEY INGREDIENTS FOR A POWERFUL DATA WAREHOUSE

A modern data warehouse must have the capabilities to store all kinds of data—structured, unstructured, semi-structured or data streaming. In addition, a data warehouse must also perform functions such as ingestion, storage, processing and reporting under one umbrella. Here are some of the key aspects that a modern data warehouse must have in today’s data-driven business landscape:



DATA WAREHOUSING PITFALLS

While data warehouse projects are among the most visible and expensive initiatives an organization can undertake, they are also among the most likely to fail. According to Gartner, more than 50 percent of data warehouses fail to make it to user acceptance. With data becoming a critical element for an enterprise’s business operations today, it is imperative that data warehousing projects are executed and implemented successfully. Some of the reasons why data warehousing projects fail are:

- **Not answering the big question** – Why does an organization need a data warehouse?
- **Using the Big Bang approach** – Delivering usable business functionality and building data warehouse incrementally.
- **Shortening testing and involving business at a much later stage for validation**
- **Neglecting maintenance** – With enterprise data and analytics requirements changing constantly, a data warehouse project has no end date.

DATA WAREHOUSING BEST PRACTICES

Every data warehousing engagement should identify and implement certain best practices for optimal technological and business returns on investment.

DESIGN

Define standards upfront

Put in place technical as well as methodology-related standards to ensure there is no confusion during the implementation stage.

Choose ELT and data lake instead of ETL

It is imperative to consider a data lake architecture along with ELT logic to ensure optimized data storage and retrieval capabilities, to manage data of multiple types easily.

Integration layer

A source-agnostic integration layer can be used to pull together information from multiple sources, allowing better business reporting and alignment with the business model.

Enable ad-hoc querying and self-service BI

A self-service BI with an analytical model to back it up makes it easy for the analytical layer to generate reports without affecting the sanctity of the underlying data model.

ARCHITECTURE

DATA MODEL

Getting a common understanding of what information is important to the business will be vital to the success of the data warehouse. Adopting recognized data warehouse architecture standards can help a long way, such as:

Star, Snowflake or Constellation Schema:

Adopting the right approach based on facts, dimensions and reporting is a key factor in meeting business requirements.

Slowly Changing Dimensions (SCD): Tracking changes to keep a historical reporting perspective and adopting the right SCD (Type 0, 1, 2, 3, 4) based on data is essential.

Normalization: Adoption of the right database design technique and the appropriate normal form is what differentiates a practical and usable warehouse from a defunct one.

Change Data Capture (CDC): CDC minimizes the resources required for ETL processes and ensures data synchronicity in an optimal manner.

MASTER DATA MANAGEMENT (MDM)

By providing a single point of reference for critical information, MDM eliminates costly redundancies that occur when organizations rely upon multiple, conflicting sources of information.

METHODOLOGY

Data Lineage

Identifying the critical data elements (CDEs), authoritative data (AD) sources and data traverses across systems (including capture of transformations) allows everyone to determine corrective actions in case of a data mismatch.

Data Governance Council

The council is responsible for the integrity and quality of data before it is ingested into the data warehouse. The members of the council should be the data team, data owners and data specialists from relevant parts of the organization who own the data.

Data Management Program

The entire methodology needs to be integrated into the Data Management Program by breaking it down into three steps:

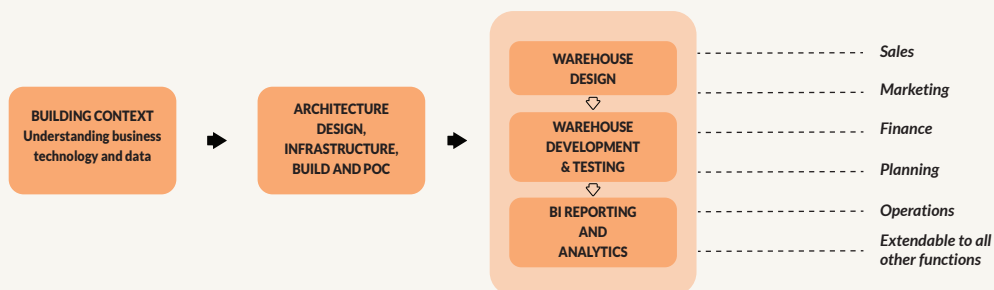
Analyze: Before creating governance policies, it is necessary for the project owner to define what data quality means for the organization, in addition to profiling and quantifying the current data landscape.

Improve: Construct the framework (for example, data governance) and run the utilities to continuously cleanse and enrich the data.

Control: Continuous monitoring and reporting of ambiguities helps in maintaining the quality of data. With greater access to high-quality data, one can finally start to monetize this information by increasing productivity, reducing waste and driving additional revenue.

IMPLEMENTATION METHODOLOGY - AGILE

One of the major reasons why data warehouse projects often fail is because of the traditional waterfall approach. On the other hand, the nature of data warehouse projects at times makes it difficult to adopt truly agile practices. Hence a modification of the approach is required as highlighted in the diagram below:



With proper planning and aligning it to a single integration layer, a data warehouse project can be broken down into smaller, faster deliverables to ensure faster and high-value returns. This also gives the teams the flexibility to change or adapt according to the dynamic business changes.

REFERENCE ARCHITECTURE OF A MODERN DATA WAREHOUSE

A robust, modern data warehouse must have the capabilities to easily consolidate all the data at any scale and provide deep, comprehensive insights through analytical dashboards, operational reports or advanced analytics for all users.

SOURCE ①



The data can be sourced as a structured data-set, semi-structured or as an unstructured one.

EXTRACT ②



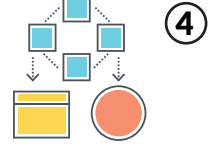
Automated data pipelines transport the data from one system to another, not necessarily transforming it. The data can also be processed in real time (or streaming) instead of batches.

STORE ③



A data lake can hold vast amounts of raw data in its native format until it is needed. Cloud object stores (AWS S3, Azure Blob, Google Cloud Storage, and more) offer high availability access for big data computing at an extremely low-cost.

TRANSFORM ④



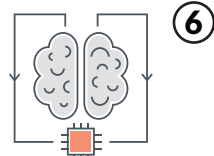
In this phase, the data is prepared, cleaned and transformed as per the need for loading into the data warehouse. An important point to be kept in mind while building a data warehouse is to integrate data from multiple heterogeneous sources that support analytical reporting, structured and/or ad hoc queries and decision making.

DATA WAREHOUSING ⑤



In-memory models and semantic layer help in addressing detailed and specific reporting needs.

ANALYTICS ⑥



By building, training and deploying artificial intelligence and machine learning models, users can derive comprehensive insights to take critical business decisions.

REPORTING ⑦



BI reports and dashboards can be created to show data insights to business user, while self-service BI can be used serve requests on-the-fly

BUSINESS BENEFITS

With a robust data quality process and data governance framework in place, data management and quality will improve over time. Some of the multiple benefits organizations can reap by putting in place robust data governance capabilities are:

- Better insights from data analytics
- Accountability for data
- Reduction in rework and costs
- Ability to track lineage and hence better business and IT agility
- Better compliance and reduced costs for compliance reporting

CASE STUDY
DERIVE DEEP INSIGHTS THROUGH EVOLVED DATA WAREHOUSING

Business challenge

The client is India's only payments company with multichannel transaction processing capabilities—web, mobile, in-store or at the time of delivery. The company sought to re-design and create an enterprise data warehouse to improve the overall system, ensure better BI experience, drive scalability, and more importantly, leverage cloud capabilities.

Solution offered

Since the aim was to drive high level of efficiencies in the data management and warehousing practices, Prescience made a detailed study of the client's existing transaction data, operational data stores, reporting infrastructure and analytics and reporting needs. Keeping the usage considerations and scope of growth in mind, the team designed the architecture with the following considerations:

- Support for structured/semi-structured/unstructured data
- Customer-centric process
- Support analytics and machine Learning – ease of availability of historical data
- Varied high-performance dashboards and reports

Business outcomes

Armed with meaningful and actionable insights through data-driven business models, the client is now on the path to achieve significant cost reductions and RoI improvements. In addition to this, the decision makers now have:

- Access to good quality and consistent data
- Holistic view of the organization's financial health, productivity and growth
- Ability to derive deep insights in the areas of customer demand trends, churn behavior and relation of churn to customer support

About Prescience Decision Solutions

Prescience is a business focused analytics firm that empowers organizations to find meaningful insights in their data. Our Business-Backward Approach helps create tangible data-driven solutions that provide users with timely inputs for astute decision making. We do this by leveraging our expertise in machine learning and advanced data science technologies, deep domain knowledge and our customers' business knowledge. Visit us at www.prescienceds.com or send us an email at info@prescienceds.com to get in touch with us. You can also follow us on LinkedIn.