

BI and Data Warehousing: Do You Need a Data Warehouse Anymore?

Business Intelligence and Data Warehousing are not synonymous anymore. The alternative: Instant BI in a data lake and automated data warehouses with ELT.

For a long time, Business Intelligence and Data Warehousing were almost synonymous. You couldn't do one without the other: for timely analysis of massive historical data, you had to organize, aggregate and summarize it in a specific format within a data warehouse.

But this dependency of BI on data warehouse infrastructure had a huge downside. Historically, data warehouses were or can be an expensive, scarce resource. They take months and millions of dollars to setup, and even when in place, they allow only very specific types of analysis. If you need to ask new questions or process new types of data, you are faced with major development efforts.

We'll define business intelligence and data warehousing in a modern context, and raise the question of the importance of data warehouses in BI.

We offer two alternatives to a traditional BI/data warehouse paradigm:

1. **Instant BI in a data lake** using an Extract-Load-Transform (ELT) strategy
2. **Automated data warehouses** that allow faster time to analysis without formal ETL

What is Business Intelligence and Analytics?

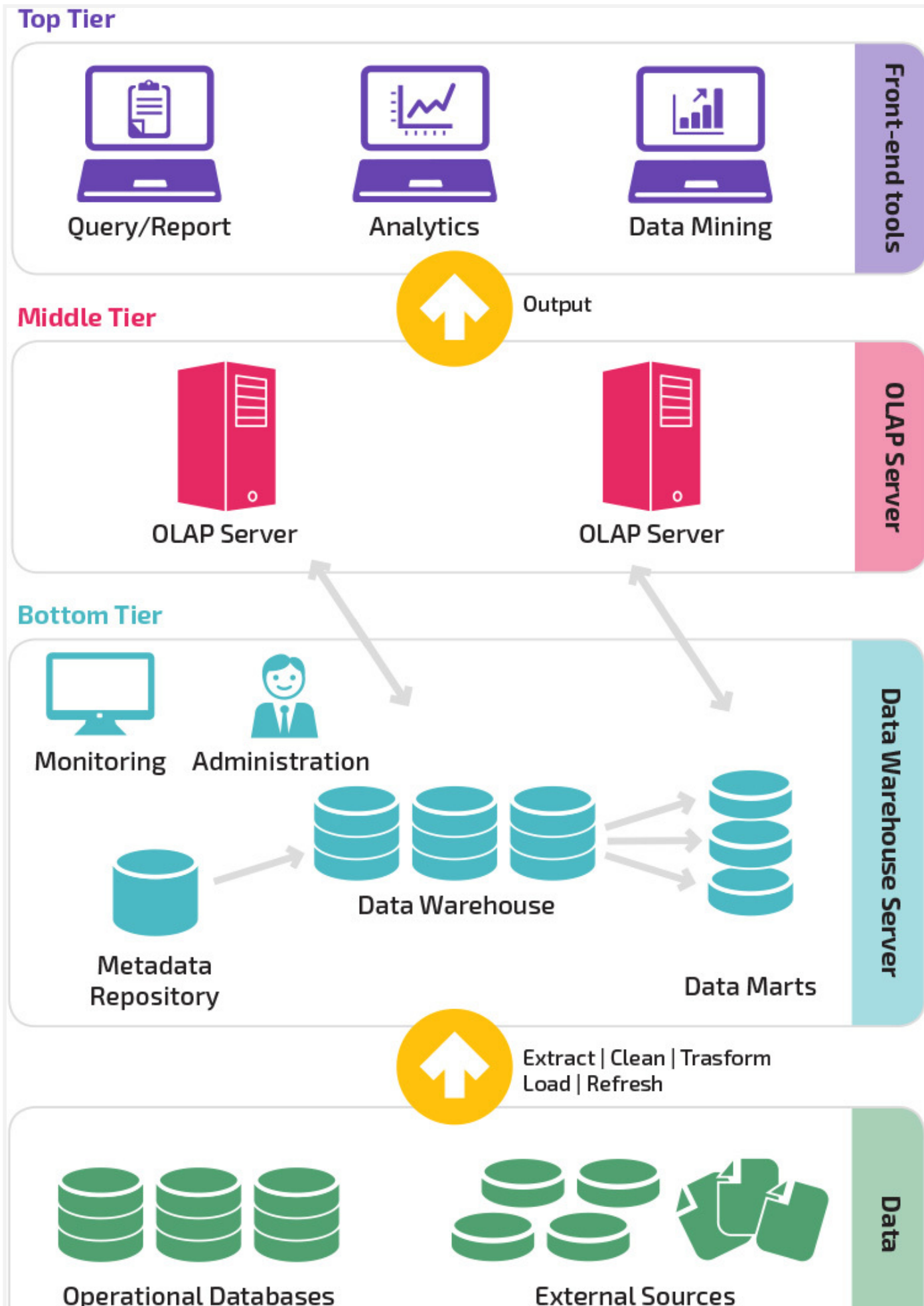
Business intelligence (BI) is a process for analyzing data and deriving insights to help businesses make decisions. In an effective BI process, analysts and data scientists discover meaningful hypotheses and can answer them using available data.

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 really is hurting conversion, and 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—and run queries against that data, typically in SQL format. Using the query results, they create reports, dashboards and visualizations 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.

What is a Data Warehouse?

A data warehouse is a relational database that aggregates structured data from across an entire organization. It pulls together data from multiple sources—much of it is typically online transaction processing (OLTP) data. The data warehouse selects, organizes and aggregates data for efficient comparison and analysis.



A data warehouse maintains strict accuracy and integrity using a process called Extract, Transform, Load (ETL), which loads data in batches, porting it into the data warehouse's desired structure.

Data warehouses provide a long-range view of data over time, focusing on data aggregation over transaction volume. The components of a data warehouse include online analytical processing (OLAP) engines to enable multi-dimensional queries against historical data.

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.

Business Intelligence and Data Warehousing: Can You Have One Without the Other?

Two decades ago most organizations used decision support applications to make data-driven decisions. These apps queried and reported directly on data in transactional databases—without a data warehouse as an intermediary. This is similar to the current trend of storing masses of unstructured data in a data lake and querying it directly.

Colin White [lists](#) five challenges experienced back in the days of decision support applications, without a data warehouse:

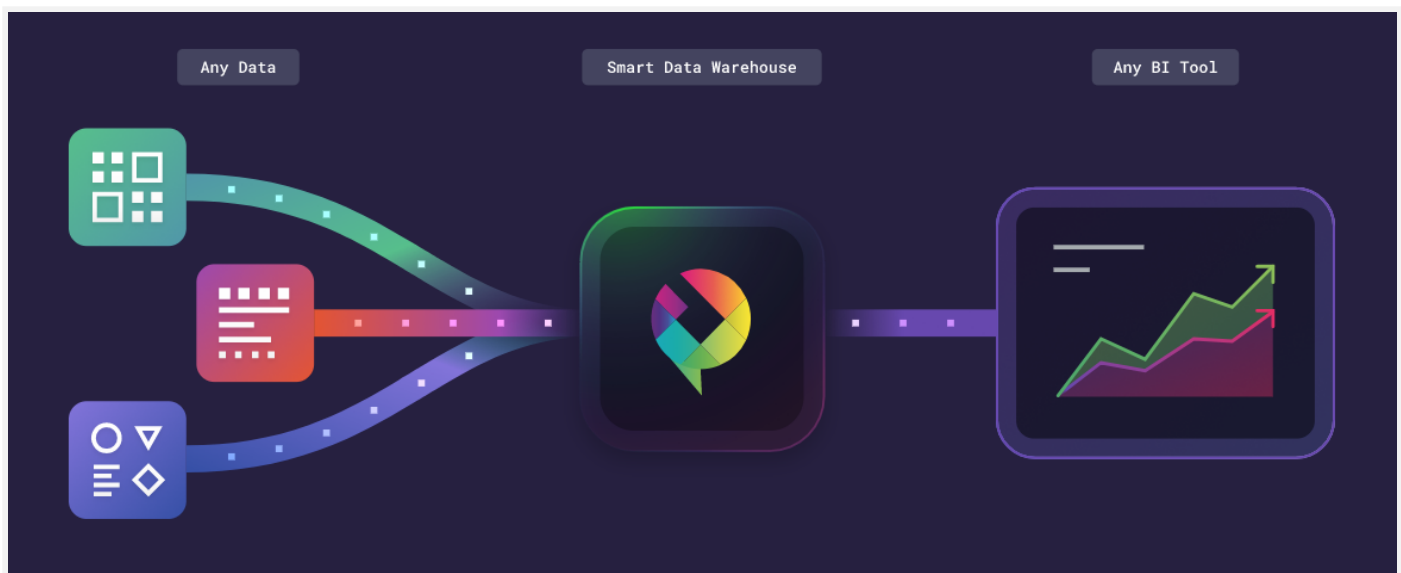
1. Data was not usually in a suitable form for reporting
2. Data often had quality issues

3. Decision support processing put a strain on transactional databases and reduced performance
4. Data was dispersed across many different systems
5. There was a lack of historical information, because transactional OLTP databases were not built for this purpose

These, among others, were the reasons almost all enterprises adopted the data warehouse model. All five of these problems still seem relevant today. So can we do without a data warehouse, while still enabling efficient BI and reporting?

BI and ETL: Running in a Data Lake without a Rigid ETL Process

With the advent of data lakes and technologies like Hadoop, many organizations are moving from a strict ETL process, in which data is prepared and loaded to a data warehouse, to a looser and more flexible process called Extract, Load, Transform (ELT).



Today ELT is mainly used in data lakes, which store masses of unstructured information, and technologies like Hadoop. Data is dumped to the data lake without much preparation or structure. Then, analysts identify relevant data, extract it from the data lake, transform it to suit their analysis, and explore them using BI tools.

Does the Data Lake Replace the Data Warehouse?

ELT is a workflow that enables BI analysis while sidestepping the data warehouse. But those same organizations that use Hadoop or similar tools in an ELT paradigm, still have a data warehouse. They use it for critical business analysis on their central business metrics—finance, CRM, ERP, and so on.

Data warehouses are still needed for the same five reasons listed above. Raw data must be prepared and transformed to enable analysis on the most critical, structured business data. If management needs to see a weekly revenue dashboard, or an in-depth analysis on revenue across all business units, data needs to be organized and validated; it can't be pieced together from a data lake.

Can such a structured analysis happen without a rigid ETL process? Or in other words, are ELT strategies relevant inside the data warehouse?

BI in an Enterprise Data Warehouse without ETL

New, [automated data warehouses](#) such as Panoply are changing the game, by allowing Extract-Load-Transform (ELT) within an enterprise data warehouse.

Panoply makes it possible to load masses of structured and unstructured data to its cloud-based data warehouse, without any ETL process at all. It uses a self-optimizing architecture with machine learning and natural language processing (NLP) to automatically prepare data for

analysis. Analysts can run queries to transform the data on the fly as needed, and work on the transformed tables in a BI tool of their choice.

Panoply solves all five problems presented above without the cost and complexity of an ETL process:

1. **Data not in suitable form for reporting** — Panoply prepares and optimizes data automatically as it is ingested to the data warehouse.
2. **Data has quality issues** — Panoply uses machine learning and NLP strategies to automatically correct many quality issues. You can fix other issues using on-the-fly transformations. Or, you can integrate with lightweight ETL tools like Stitch or Blendo, and build a cloud-based ETL pipeline in just a few clicks.
3. **Strain on transactional database performance** — not a problem because data is still being loaded to a separate data warehouse.
4. **Data dispersed across many systems** — Panoply integrates with [dozens of data sources](#), so loading data is only a matter of selecting a data source, providing credentials and selecting a destination table.
5. **Lack of historic information** — Panoply makes it possible to ingest multiple layers of historic information into the data warehouse, and easily join or aggregate the data using on-the-fly queries and transformations.

The primary benefit is **shorter time to analysis**. With an automated data warehouse, you can go from raw data to analysis in minutes or hours, instead of weeks to months.

From Monolithic Data Warehouse to Agile Data Infrastructure

Data warehouses have come a long way. The monolithic Enterprise Data Warehouse (EDW), which required a multi-million dollar project to setup, and allowed only very limited BI analysis on

specific types of structured data, is soon to be a thing of the past.

Today there are two quick, low cost ways to get from raw data to business insights:

- **Data lake with an ELT strategy** — does not allow the same critical business analysis as the EDW. But a data lake lets you do more with BI, extracting insights from enterprise data that was not previously accessible.
- **Automated data warehouse** — new tools like Panoply let you pull data into a cloud data warehouse, prepare and optimize the data automatically, and conduct transformations on the fly to organize the data for analysis. With a smart data warehouse and an integrated BI tool, you can literally go from raw data to insights in minutes.

The slow-moving ETL dinosaur is not acceptable in today's business environment. Organizations are saving money and making business decisions faster, by simplifying and streamlining process the data preparation process. For a real-life example, see [how Kimberley Clark uses Panoply](#) to gain agility and prepare data automatically for BI.

Learn More About Data Warehousing

- [Data Warehouse Architecture: Traditional vs. Cloud](#)
 - [Database vs. Data Warehouse](#)
 - [Data Mart vs. Data Warehouse](#)
 - [Amazon Redshift Architecture](#)
-



Built by Panoply