

Database vs Data Warehouse: A Comparative Review

By Drew Cardon

The important fact is that a transactional database doesn't lend itself to analytics. To effectively perform analytics, you need a data warehouse.

A question I often hear out in the field is: I already have a database, so why do I need a data warehouse for healthcare analytics? What is the difference between a database vs. a data warehouse? These questions are fair ones.

I've worked with databases for years in healthcare and in other industries, so I'm very familiar with the technical ins and outs of this topic. In this post, I'll do my best to introduce these technical concepts in a way that everyone can understand.

Before diving in to the topic, I want to quickly highlight the importance of analytics in healthcare. If you don't understand the importance of analytics, discussing the distinction between a database and a data warehouse won't be relevant to you. Here it is in a nutshell. The future of healthcare depends on our ability to use the massive amounts of data now available to drive better quality at a lower cost. If you can't perform analytics to make sense of your data, you'll have trouble improving quality and costs, and you won't succeed in the new healthcare environment.

The High-level Distinction Between Databases and Data Warehouses

What I will refer to as a "database" in this post is one designed to make transactional systems run efficiently. Typically, this type of database is an OLTP (online transaction processing) database. An electronic health record (EHR) system is a great example of an application that runs on an OLTP database. In fact, an OLTP database is typically constrained to a single application.

The important fact is that a transactional database doesn't lend itself to analytics. To effectively perform analytics, you need a data warehouse. A data warehouse is a database of a different kind: an OLAP (online analytical processing) database. A data warehouse exists as a layer on top of another database or databases (usually OLTP databases). The data warehouse takes the data from all these databases and creates a layer optimized for and dedicated to analytics.

So the short answer to the question I posed above is this: A database designed to handle transactions isn't designed to handle analytics. It isn't structured to do analytics well. A data warehouse, on the other hand, is structured to make analytics fast and easy.

In healthcare today, there has been a lot of money and time spent on transactional systems like EHRs. The industry is now ready to pull the data out of all these systems and use it to drive quality and cost improvements. And that's where a data warehouse comes into play.

Databases versus Data Warehouses: The Details

Now that you have the overall idea, I want to go into more detail about some of the main distinctions between a database and a data warehouse. Because I'm a visual person (and a database guy who likes rows and columns), I'll compare and contrast the two in table format.

	Database	Data Warehouse
Definition	Any collection of data organized for storage, accessibility, and retrieval.	A type of database that integrates copies of transaction data from disparate source systems and provisions them for analytical use.
Types	There are different types of databases, but the term usually applies to an OLTP application database, which we'll focus on throughout this table. Other types of databases include OLAP (used for data warehouses), XML, CSV files, flat text, and even Excel spreadsheets. We've actually found that many healthcare organizations use Excel spreadsheets to perform analytics (a solution that is not scalable).	A data warehouse is an OLAP database. An OLAP database layers on top of OLTPs or other databases to perform analytics. An important side note about this type of database: Not all OLAPs are created equal. They differ according to how the data is modeled. Most data warehouses employ either an enterprise or dimensional data model, but at Health Catalyst, we advocate a unique, adaptive Late-Binding™ approach. You can learn more about why the Late-Binding™ approach is so important in healthcare analytics in Late-Binding vs. Models: A Comparison of Healthcare Data Warehouse Methodologies.
Similarities	Both OLTP and OLAP systems store and manage data in the form of tables, columns, indexes, keys, views, and data types. Both use SQL to query the data.	

	Database	Data Warehouse
How used	Typically constrained to a single application: one application equals one database. An EHR is a prime example of a healthcare application that runs on an OLTP database. OLTP allows for quick real-time transactional processing. It is built for speed and to quickly record one targeted process (ex: patient admission date and time).	Accommodates data storage for any number of applications: one data warehouse equals infinite applications and infinite databases. OLAP allows for one source of truth for an organization's data. This source of truth is used to guide analysis and decision-making within an organization (ex: total patients over age 18 who have been readmitted, by department and by month). Interestingly enough, complex queries like the one just described are much more difficult to handle in an OLTP database.
Service Level Agreement (SLA)	OLTP databases must typically meet 99.99% uptime. System failure can result in chaos and lawsuits. The database is directly linked to the front end application. Data is available in real time to serve the here-and-now needs of the organization. In healthcare, this data contributes to clinicians delivering precise, timely bedside care.	With OLAP databases, SLAs are more flexible because occasional downtime for data loads is expected. The OLAP database is separated from frontend applications, which allows it to be scalable. Data is refreshed from source systems as needed (typically this refresh occurs every 24 hours). It serves historical trend analysis and business decisions.
Optimization	Optimized for performing read-write operations of single point transactions. An OLTP database should deliver sub-second response times. Performing large analytical queries on such a database is a bad practice, because it impacts the performance of the system for clinicians trying to use it for their day-to-day work. An analytical query could take several minutes to run, locking all clinicians out in the meantime.	Optimized for efficiently reading/retrieving large data sets and for aggregating data. Because it works with such large data sets, an OLAP database is heavy on CPU and disk bandwidth. A data warehouse is designed to handle large analytical queries. This eliminates the performance strain that analytics would place on a transactional system.
Data Organization	An OLTP database structure features very complex tables and joins because the data is normalized (it is structured in such a way that no data is duplicated). Making data relational in this way is what delivers storage and processing efficiencies—and allows those sub-second response times.	In an OLAP database structure, data is organized specifically to facilitate reporting and analysis, not for quick-hitting transactional needs. The data is denormalized to enhance analytical query response times and provide ease of use for business users. Fewer tables and a simpler structure result in easier reporting and analysis.

	Database	Data Warehouse
Reporting/ Analysis	<p>Because of the number of table joins, performing analytical queries is very complex. They usually require the expertise of a developer or database administrator familiar with the application. Reporting is typically limited to more static, siloed needs. You can actually get quite a bit of reporting out of today's EHRs (which run on an OLTP database), but these reports are static, one-time lists in PDF format. For example, you might generate a monthly report of heart failure readmissions or a list of all patients with a central line inserted. These reports are helpful—particularly for real-time reporting for bedside care—but they don't allow in-depth analysis.</p>	<p>With fewer table joins, analytical queries are much easier to perform. This means that semi-technical users (anyone who can write a basic SQL query) can fill their own needs. The possibilities for reporting and analysis are endless. When it comes to analyzing data, a static list is insufficient. There's an intrinsic need for aggregating, summarizing, and drilling down into the data. A data warehouse enables you to perform many types of analysis:</p> <ul style="list-style-type: none"> ➤ Descriptive (what has happened) ➤ Diagnostic (why it happened) ➤ Predictive (what will happen) ➤ Prescriptive (what to do about it) <p>This is the level of analytics required to drive real quality and cost improvement in healthcare.</p>

I hope the information I've included here has helped you understand why data warehouses are so important to the future of healthcare. Improving quality and cost requires analytics. And analytics requires a data warehouse.

An OLTP database like that used by EHRs can't handle the necessary level of analytics. My rule of thumb is this: If you get data into your EHR, you can report on it. If you get it into a data warehouse, you can analyze it.

It's that simple. 💡

About the author



Drew Cardon joined Health Catalyst in November 2011 as a data architect. Prior to this, he worked for nine years in the state tax and revenue industry as a project manager and implementation consultant with Accenture, and later with Fast Enterprises. He was involved in the installation of large information technology systems for the State Tax Commissions in Arizona, Utah, and Oklahoma. He holds a Bachelor's degree in business from Brigham Young University and an MBA from the University of Notre Dame.