

Read This Before You Make a Decision About Data Warehousing

Data warehousing for non-IT Executives

CHARTIO

Introduction: Infrastructure shouldn't drive your business intelligence decisions

Modern management requires information. Without a correct, common understanding of inputs and results, it's impossible to identify opportunities or evaluate initiatives.

But the quest for reliable information can be sidetracked by decisions about information technology. Which business intelligence tool is the best fit for my organization? How is your data stored, and how should it be? How does my business intelligence tool need for data to be stored? Do I need a data warehouse? What's a data warehouse? What kind of data warehouse do I need?

It can be confusing to know where to start. The goal of this report is a simplify the arcane problem of data warehousing for executives who know they're ready for business intelligence, but don't want to get bogged down in a giant infrastructure project.

Do you need a data warehouse?

Before continuing, it's important to consider whether you need a data warehouse at all. A data warehouse can cost a lot more than your business intelligence tool itself, require additional development resources, impose long term maintenance requirements, complicate your architecture, and delay implementation of your BI project.

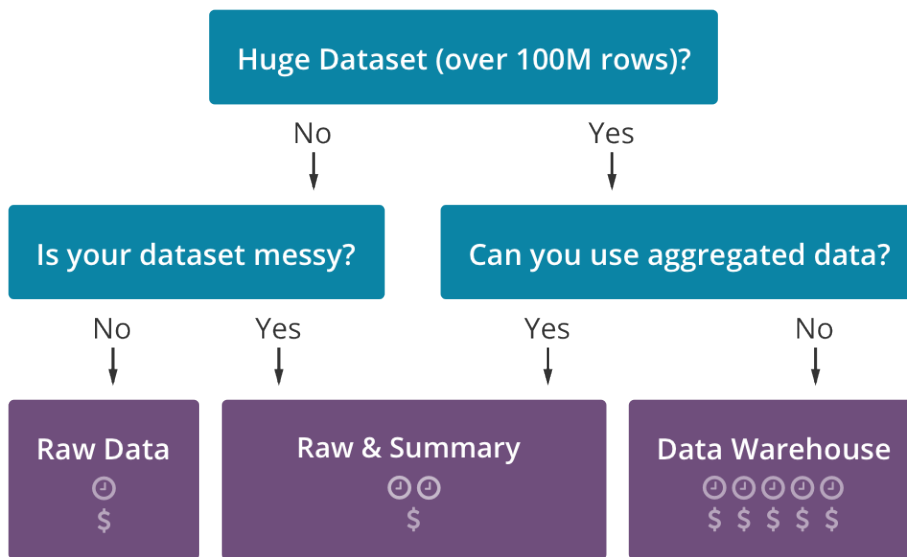
If your data is already stored in a SQL database, you may not need a data warehouse. Many of our customers either clone their operational database, or maintain a simplified version of their live data that contains only the information they need analytical and BI queries. This takes the load off their operational database without creating the need for additional hardware and software.

If your data is already stored on Amazon Redshift, your business intelligence tool should be able to access it directly.

However, there are some good reasons to build a data warehouse. Your answer to the following questions will help you understand whether you need a data warehouse, and what kind of data warehouse you need.

- **How big is your data?** If you have huge volumes of data (say, more than 100 million rows), you may need to move your data to a system that can support large scans and fast aggregations. Fortunately, today there are cloud solutions to simplify this process.
- **How messy is your data?** If your data needs to be aggregated, anonymized, or cleaned in order to be quickly and usefully processed, it may make sense to periodically preprocess the data and store it separately.
- **Do you need live data?** It's possible to create dashboards that show the up-to-the-minute status of your business by analyzing a near-real-time clone of your operational database. But if you're more concerned with trends and daily/weekly/monthly/quarterly numbers, it may be better to create summary tables of your live data.
- **Where and how is your data stored?** Chartio is able to connect quickly and easily with most SQL and many non-SQL data sources. However, if you have data in some sources that your BI tool cannot connect to, you may need to copy that data into a SQL database or data warehouse.
- **Do you need to merge databases?** If you need to merge data from multiple databases or disparate data sources (such as sales from your site and data from Google Analytics), you may need to use an ETL process (see page X) to build an analytical database. However, Chartio is able to blend up to two million rows from disparate data sources without the use of ETL or preprocessing. This is more than enough if you're aggregating your data by time, location, or product.

What Kind of Storage Do You Need?



What are your organizational requirements?

Whether you'll need a data warehouse, and what kind you need, will also be driven by the nature of your organization and how you use data.

How should your BI tool access your data?

If all your critical business data isn't already stored in a SQL database, it's time to make the move. SQL is an industry standard (ANSI/ISO) that has been optimized for data analysis over more than 30 years. Proprietary query languages promote vendor lock-in. NoSQL languages force the report writer to efficiently join and aggregate the data, which can result in poor performance. You may be using third-party or home-grown systems that don't already store business information using SQL. Building your BI system on top of SQL gives you access to a pool of skilled talent who are able to help you solve your data access problems.

If everyone on your team, and the potential users of your business intelligence system are already using SQL to pull their own reports, then you have a lot more options. Many BI tools use SQL as their primary

interface to your data, but some BI tools will not give you access to your data or their functionality using SQL. Your BI vendor should be able to help you with making your data available via SQL.

How will your team access your BI tool?

Most database analysts will be comfortable using SQL inside your BI tool. But you can't expect business users to learn SQL in order to answer their own questions. Chartio's drag-and-drop interface makes it easy for business users to answer their own questions, while giving SQL analysts the power they need to answer more difficult questions.

If you're using a home-grown system, it may be time to move to a full business intelligence tool. This will free precious development resources for projects only they can do. It will also get routine business queries out of the development queue and into the hands of the people asking the questions.

If your key reports are based on data periodically dumped into Excel, it's time to move to business intelligence. Excel is great for ad hoc analyses, but systems built on top of Excel often mutate rapidly into macro-driven multi-step processes that are time-consuming, tedious, hard to customize, error-prone, and out of date before they're delivered.

How do you control access to data?

Not everyone needs or should be allowed access to all information. There may be regulatory, financial, privacy, or security reasons for limiting access to data. Be sure that your BI tool gives you the controls you need.

A business intelligence system must be able to limit access to some kinds of data to some people. Manual and home-grown systems that are not built with access controls in mind can leak critical information.

Your data warehouse and business intelligence system must be built with access controls in mind. Your BI tool should allow you to control access to your data at the database, table, or even column level. You can set your controls for particular individuals or groups.

How will you get your data into your warehouse, i.e. do you need ETL?

In addition to determining the best way to store data for your business intelligence tool, you will need to consider how you're going to get the data into your warehouse.

The process of getting data out of your operational databases and into a data warehouse is called ETL (for extract, transform, and load). ETL is beyond the scope of this white paper, but you may encounter it as you investigate data warehousing for business intelligence.

ETL is a specialized and potentially expensive operation. However, if your data is relatively uniform and stored in accessible SQL databases, you may not need a full ETL solution. Many BI users find that they can get the job done by creating in summary tables of their data.

What are your options for data warehousing?

If you've decided that you need a data warehouse, how do you decide what to use? There are four options:

Bundled data warehouse

Some business intelligence tools require you to purchase their proprietary data warehouse. You will have to invest development and consulting resources to adapt your data to their system. You will be locked into a single supplier for data warehousing services at a time of increasing market competition, and you will not be able to use this data warehouse if you change vendors.

On-premises data warehouse

Even if your business intelligence tool doesn't require an on-premise data warehouse, there may be circumstances where it's necessary. This may include company, control, or IT-imposed requirements.

There are a couple of approaches to warehousing your data on premises. The simplest is to create a clone or a simplified copy of your database on your existing systems and connect this to your business intelligence tool.

A more complex and expensive option is to purchase a hardware and software solution purpose-built to serve as a data warehouse that resides on your site. Many BI tools require you to purchase and set up an on-premise data warehouse in order to work.

Virtual data warehouse

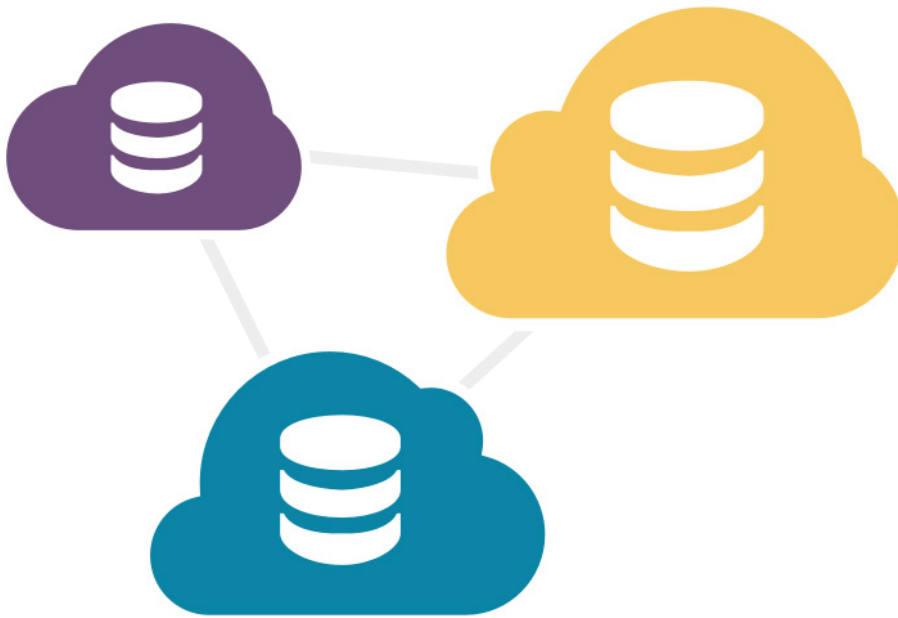
A virtual data warehouse looks like a single database to your business intelligence tool, but queries multiple underlying sources independently. Virtual data warehouses are good at integrating different data sources and simplifying access to complex data for analysts and business users. They also avoid the complexity and expense of ETL. Chartio can blend disparate data using layers, and makes it easy virtualize your database by renaming columns and tables and choosing which to show.

Cloud data warehouse

There are many good reasons why enterprises are moving their data to the cloud. Chartio is a cloud service that designed from the ground up to work with data in the cloud.

- It's quicker and easier to set up a cloud database.
- Buying cloud services is simpler than buying hardware, software, and services to set up your data warehouse.

- It's straightforward to start small and scale up as your needs increase. You never have to buy additional hardware, software, connections, or licenses.
- The expense of cloud data storage is more predictable and you don't need to buy additional consulting or services.
- The cost of cloud storage and processing is declining faster than the cost of dedicated hardware and software.



What's Agile Business Intelligence, and how can it help?

Agile Business Intelligence treats solving your BI needs as a process, and not a one-time implementation. Rather than treat the BI implementation as a step-by-step implementation of a predetermined idealized solution, Agile BI instead delivers BI tools in pieces that are manageable and usable by the business team. The system can then evolve as the organization's needs change. A good agile BI system involves collaboration between IT and business users.

Agile BI give you what you need today and prepares you to make decisions you never thought you'd face. Agile BI has a number of advantages:

- Faster time to value
- Quick implementation of new systems
- Reduced development cycles
- Faster adaptation to changes in the business or marketplace.
- Improved correction of errors
- Improved ability to adopt new technologies
- Greater flexibility

Agile BI is inherently adaptable. Cloud technology is commoditizing storage and processing power, and Agile BI can take advantage of the opportunities this creates. You know your needs will change in unanticipated ways in the next five years. Your cloud data warehouse can grow as your needs grow, or shrink as your needs become more focused.

This gives you the ability to get things done and answer your questions while you learn more about your long-term needs. You can continuously improve your systems, your preprocessing, and your reports – even while you're providing your team with quality information. As your team learns about what information is available and how they can use it, you'll be able to turn ad hoc reports into KPI's you use every day.

Combining Chartio with a cloud data warehouse makes it simple for your to adapt as your needs change. Because Chartio is a cloud application, it's continually updated to the latest versions. The beauty of Agile Business Intelligence is it adapts the way the best tools do – by breaking in as you use it every day.

How can Chartio help?

Chartio is designed to simplify access to your data, and we hope we simplified your data warehousing decision.

No BI tool connects to your data faster and more simply thanv Chartio. This will help you achieve the goal of agile business intelligence. Don't let your choice of business intelligence tool determine your data warehousing solution.

Chances are, you can get up and running with Chartio by creating a simplified copy of your operational data on your current system. If you later decide you need a data warehouse, or if your IT team believes you need a data warehouse from the get-go, we can interface quickly and easily with your warehouse either in the cloud or on your premises. This means you have maximum flexibility in your initial setup, and as your BI needs change and grow.

Data Warehouse Decision Criteria

Which BI tool do I want to use?	
How much data do I have? How many gigabytes, and how many rows?	
How messy is my data? Does it need to be aggregated by day/week/month?	
Where and how is my data stored? What databases do I need access to?	
Do I need to merge data? Can it be done in my BI tool?	
What access controls do I need?	
How do I get data into my warehouse? Do I need ETL, or something simpler?	
Does my data warehouse need to be on-premises, or can I store my data in the cloud?	

CHARTIO

445 Bryant Street | San Francisco, California 94107 | United States

+1 855 232 0320 | hello@chartio.com

Copyright ©2014 - All Rights Reserved