



Going (Cloud) Native:
**REDSHIFT, SNOWFLAKE,
AND BEYOND**

DATA: BIGGER THAN BIG

Think of the biggest number you can. Go ahead. Take a second. How big is it? Millions? Billions? Trillions? Quintillions? (Yes, that one's [real](#).) A few enlightened souls probably said “googol,” which is a 1 followed by 100 zeros or even a “googolplex” (not to be confused with Google's Silicon Valley HQ, the Googleplex), a 1 with a *googol* of zeros after it.

Wow. Those are some huge numbers.

That's what Big Data is becoming in our modern era. Data is being produced so quickly and by so many different sources that we can only make [ballpark estimates](#) of how much we're creating every day. Sensors and IoT devices crank out data constantly, and as more of our world becomes digitized, every human action creates a corresponding digital record that goes into a giant dataset and becomes grist for analytics and BI mills.

Teams all over the world are building the next generation of software, products, and services on top of these immense amounts of data. They need the right tools to store, handle, and analyze that data. One type of solution that's gained traction recently is the Cloud-Native Data Warehouse. These are warehouses like Amazon Redshift, Google BigQuery, and Snowflake.

This latest generation of data warehouses has arisen to fill a specific niche. Previous databases (and even some modern ones) were unable to handle the immense amounts of data being produced every day. These outdated systems failed at scaling, didn't offer high-enough performance, couldn't perform at scale, or weren't great at storing things like unstructured data (think images, videos, etc.).

Whatever the reason, if you're a modern company dealing with modern amounts of information, you can't afford to have a database that can't handle your data. Additionally, cloud data warehouses allow companies with rapidly-changing data needs to bring in new datasets for analysis alongside legacy information, instead of adding whole new databases. They offer a combination of scale and performance without having to maintain hefty on-premises instances or complex backend workflows and can even be set up to control costs. Combine this with the fact that a high percentage of new apps and products are being built on the cloud to begin with, and choosing a Cloud-Native Data Warehouse starts to make more and more sense.

In this ebook, we'll dig into three big players in the cloud data warehouse space (Amazon Redshift, Snowflake, and Google BigQuery) and talk about the problems that they solve and how to get the most out of them with an analytics solution.

WHY CLOUD-NATIVE DATA WAREHOUSES?

As a category, cloud-native data warehouses all offer some similar benefits. For starters, they free companies from concerns around facilities and machines: in the old days of physical servers, companies needed to worry about the server room or the server farm or at least the specific machine where their software was running or their data was stored. Building this physical infrastructure was a huge hurdle for starting or scaling a software company. Fast-forward to today, and server costs are much lower and spinning up a cloud data warehouse can be done in a few clicks. Companies pay only for what they use, processing and storing data on demand. Using the cloud can also provide increased redundancy and support for the company, as they no longer need to worry about a single server going down and crashing their entire operation. Larger cloud-service providers have multiple backup systems and can auto-scale across data centers around the world to keep everything running smoothly. It's win-win for the client company.



GREENLIGHTING AMAZON REDSHIFT

Amazon Redshift is a cloud data warehouse and one of the many AWS offerings from the Seattle tech giant.

Hosted directly on Amazon Web Services, and backed by the power and size of this mammoth company, users can scale storage and computing power quickly, easily, and to extremely high volumes. This is accomplished by activating individual nodes of different sizes as the need arises. Sometimes it pays to go with a giant corporation! Users also avoid having to shell out huge fees at the outset for storage and processing power they might not need. As they scale, they can pay to get more space or power; perfect for companies that experience a surge in growth.

Like many cloud-native options, Amazon Redshift boasts low maintenance costs, high speed, strong performance, and high availability. Connecting to live data is one of the places that Redshift shines. Companies of all kinds are turning to Redshift to improve their connections to live data and get this info into BI systems to run real-time, ad-hoc queries and deal with vital business challenges as they arise. It's also extremely efficient when it comes to performing analyses in general, clocking some of the fastest query speeds of comparable data warehouses. All good qualities to have when analyzing real-time or cached data.

Whatever you're building, being able to speak the language of your cloud-native database is vital. Redshift operates similarly to MySQL, featuring a collection of back-end tools that work with PostgreSQL, JDBC, and ODBC drivers. This means that the learning curve for the data engineers, IT teams, and developers who will be dealing with Redshift will be a lot simpler. It also connects easily with most business intelligence tools.

Amazon is serious about security. Redshift's encryption and security tools make protecting the data in your warehouse easy. VPC for network isolation plus different access control tools give you granular management capabilities, allowing you to design a security setup to suit your needs. In addition to SSL encryption for data in transit, AWS' S3 servers offer both client- and server-side encryption, giving you control over when data is viewable and accessible. If you're dealing with sensitive data, security features like these matter.



amazon
REDSHIFT

WHAT'S SPECIAL ABOUT SNOWFLAKE

Snowflake is a cloud-native, SQL data warehouse built to let users put all their data in one place for ease of access and analysis.

It's designed to handle unstructured data, can be controlled to manage resources, and also supports semi-structured data types (Variant, Object, and Array). Users can load data without worrying about the schema, converting it into a usable format that's compatible with SQL.

As a cloud-based data warehouse, it's got the flexibility and scalability that organizations depend on when choosing a cloud option. Snowflake's subscription-based model uncouples storage and computation services, allowing them to operate independently. As users build new solutions that plug into Snowflake, they only pay to store data or analyze it as needed. Additionally, the system is built with an interconnected array of cloud servers, decentralizing data and allowing individual users or groups within an org access to the specific data they need without complicated data transfers, simplifying connection and analysis.

The ability to quickly query data without affecting the underlying set and perform analyses with real-time data is a powerful feature for many cloud-native data warehouses. Since data is constantly being created by a wide array of systems, many of them native to the cloud to begin with, the ability to analyze this data in real-time is vital to modern companies. This as-needed, real-time analysis ties back into Snowflake's pricing structure by only charging for specific instances and projects without incurring higher overall costs.



THE BIG DEAL ABOUT BIGQUERY

Google BigQuery is all about scale (the “Big” in BigQuery); we’re talking billions of rows.

With an AI-powered, columnar-storage approach, users can scan their data quickly. Much like how Redshift leverages the power of Amazon’s massive infrastructure, BigQuery is built on Google Cloud Storage infrastructure and uses a SQL-like language for querying (again meaning that your SQL-trained team will have an easier time getting a handle on this data warehouse). Its AI system is actually always helping optimize your data and queries. As it detects patterns, BigQuery uses them to optimize datasets into structures better suited to the types of analytics and queries you’re performing regularly.

Another of BigQuery’s unique features is the way it dynamically distributes computing resources, which can help reduce query times and costs. Rather than a rigid structure designed around multiple clusters, you get a system that allocates computing power where it’s needed. BigQuery also offers a “serverless” build, a fully on-cloud option that offers scalability and fast queries. Its decentralized design allows users to perform queries and dig up insights even from petabytes of data.

However, having tons of data or lots of querying needs doesn’t mean your costs need to be out of control. BigQuery pricing schemes are geared towards small businesses and companies with changing analytics needs. Users can expect to be charged based on computing power and desired storage and can allocate their resource needs down to a per-second level. The platform promises 100% resource utilization, meaning you only pay for the exact computing power you’re using (vs overpaying based on fixed models). All cloud data warehouses pride themselves on

being scalable and offering costs savings, but few boast this level of granular control.

Another benefit of the cloud-native data warehousing model: real-time data access and analysis. BigQuery’s innovations in this realm are the ability to batch ingest data and its real-time ingest capacity. Batch ingesting lets you load thousands of data points into an analytics tools without incurring a computational decrease. BigQuery uses its own computing resources and does not impact real-time query abilities at all. Real-time ingest, meanwhile, can load up to 100,000 rows of data per table for instant access (this can even reach up to 1 million rows by deploying sharding). The result is incredibly fast and efficient real-time analysis.



Google
Big Query

CLOUD DATA WAREHOUSES AND ANALYTICS

By now you've seen what some of the heavy-hitter cloud-native data warehouses out there can do. How will this impact whatever you're building? And what does this have to do with your data analytics solution?

Whether you're using analytics in-house, embedding analytics into your product, or coming up with a white-labeled mobile analytics app, being tightly integrated into your data warehouse is vital.

First off, while migrating all your data (eventually) to a cloud-native data warehouse can be a great idea, it can't happen overnight. Moreover, you're almost always going to have some data that exists outside your data warehouse. This is where an analytics solution really comes in handy, giving you a convenient place to pull in all this information from different sources, analyze it, and present and share the results with your team and other stakeholders. Additionally, there's no loss in data fidelity or access as new datasets are added or your users need access to different (possibly incompatible) data sources. Basically, having a BI solution to mashup these disparate data sources eliminates data silos between your warehouse and your other systems.

Something else you no-doubt noticed about all these data warehouses is that they all have pricing structures that are based on the queries you order. If you have to query the system every time you need a piece of information, this can cut against the cost-saving factors that might have

drawn you to your cloud solution in the first place. The right analytics system will help you control when you go back to the well and incur a charge vs when you use the insights you already have from prior queries. If you have a stable well of data that doesn't change much, this can help you keep from running up charges learning what you already know. And if you're relying on mashing up real-time data with cached data, the BI system can handle this for you to, connecting to the real-time source as needed, instead of constantly paying query costs. That's huge!

Lastly, a robust analytics platform provides a semantic layer where you can interact with your data and query it, regardless of changes to the underlying data. This drastically reduces the cost of trying new things or testing query loads while concurrent users grow. Having a sandbox to experiment in without impacting the underlying data is important to builders of all stripes. This huge store of data underpins your entire operation; your product and your users depend on it. You need it to be stable and accessible, even as your engineers and developers work out new ways to build the next generation of your offering and deliver more value to your customers. That's what your analytics system offers them.

TIME TO BUILD

Now that you've learned a bit about what cloud-native data warehouses can do for you and your business and how the right business intelligence platform can help you get the most out of your selection, the ball's in your court. Whatever you're building, making a data and analytics strategy part of the foundation of your future plans will pave the way for your success and give you a strong base. Data is everywhere, customers want access to it, and building it into your products and services increases stickiness and user adoption, plus it signals to your users and competition that you understand the changing world in which we live and the role that data plays in it. If you're dealing with a huge amount of data and want to keep it all on the cloud, then a cloud-native data warehouse could be part of your plans and the right BI platform will give you the freedom to query that data your way and build powerful analytic apps on top of it.

Now you just have to get out there and make it happen. Whatever you're making, choose your tools and materials wisely, then build boldly!

[Explore Sisense + Cloud-Native Data Warehouses](#)

