snowflake

WHITE PAPER

Reinventing the Data Warehouse

When it comes to managing today's data and how it is used, current data warehousing solutions simply can't keep up. Based on assumptions and technologies from decades ago, conventional data warehouses are ill-equipped to bring together all of the data you need to analyze and to support all of the different ways in which you need to use that data.

Data Has Changed

It used to be the case that most of the data you wanted to analyze came from sources in your data center: transactional systems, enterprise resource planning (ERP) applications, customer relationship management (CRM) applications, and the like. The structure, volume, and rate of the data were all fairly predictable and well known.

Today a significant and growing share of data—application logs, web applications, mobile devices, and social media—comes from outside your data center, even outside your control. And that's without emerging new sources such as the *Internet of Things*. That data is also frequently stored in newer, more flexible data structures such as JSON and Avro. With data volume expected to increase 50-fold in the next decade, demands are increasing on both the systems themselves and on the people who manage and use them.

The harsh reality of data warehousing is that conventional solutions are simply too costly, inflexible, and complex for today's—not to mention tomorrow's—data.

The Ways Data Is Used Have Changed

At one time, it was sufficient to load updated data once a week or overnight and then generate and publish a report or dashboard every Monday morning. Not today. The value of much of today's data decays rapidly, making it a requirement to get data into the hands of analysts as quickly and easily as possible so that they can use the data to test hypotheses, create what-if scenarios, correlate trends, and project revenues.

Traditional Data Warehouses Can't Keep Up

The harsh reality of data warehousing is that conventional solutions are simply too costly, inflexible, and complex for today's—not to mention tomorrow's—data. These solutions were designed for managing predictable, slow-moving, and easily categorized data that largely came from internal enterprise applications under your control. They require customers to purchase everything they need for peak demand up front, spending hundreds of thousands of dollars (millions in some cases) just to get started. This all but guarantees that most of the technology will sit underutilized the majority of the time. As one Director of Analytics put it, "We have to buy for the 99th percentile even though we only reach that level one day per year."

> Instead of spending an hour waiting for a response, we get it in 5 minutes with Snowflake. So, we can spend more time interpreting the result or white-boarding harder questions."
>
> —James Rooney, SVP of Technology, Accordant Media

Big Data Platforms Aren't the Answer

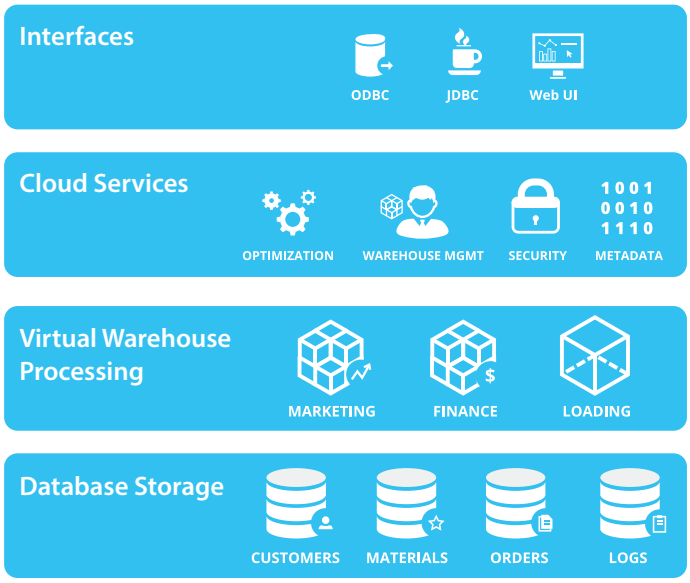
Although new solutions such as Hadoop have emerged to offer lower cost and greater flexibility, these were never designed to be data warehouse and are ill-equipped to become one. They were designed for batch processing of data, not for SQL analytics in near real-time. As a result, many customers find that once they have transformed and organized data in Hadoop, they are forced to move it to a data warehouse for fast analysis.

Equally challenging, these systems are so complex that they require specialized expertise and resources to deploy, manage, and use—operations professionals, infrastructure engineers, and data scientists. The result: access to data is restricted to a small handful of specially trained data scientists and developers charged with running analytics and deciphering their meanings, a process which often takes weeks.

Reimagining the Data Warehouse

If you were to start over, architecting the ideal data warehouse for today's and tomorrow's data needs, what would it look like? It would meet the following requirements:

* **Truly elastic.** It would be able to grow, shrink, and change in a matter of minutes to adapt to any processing demand, even going all the way back to zero when no queries are running.
* **Store and process all your data.** It would provide unlimited storage capacity at such a low cost that you never have to think about throwing out data, and it would easily accept diverse forms of data —from purely relational, structured forms (e.g. CSV files) to semi-structured data such as JSON or Avro.
* **Self-service easy.** It would make it possible for analysts to simply load data and run queries immediately without needing to worry about infrastructure, tuning, and availability.



Built from the ground up for the cloud, Snowflake's unique architecture physically separates and logically integrates compute and storage

Cloud Is the Only Way to Get There

Cloud infrastructure is the perfect platform for constructing this ideal data warehouse. Cloud infrastructure delivers near-unlimited resources, on demand, in minutes, and you pay only for what you use. That makes it possible to support virtually any scale of users and workloads without compromising performance or responsiveness.

A data warehouse built for the cloud eliminates the obstacles and pains of deploying and managing infrastructure, enabling you to focus on using your data rather than on dealing with infrastructure.

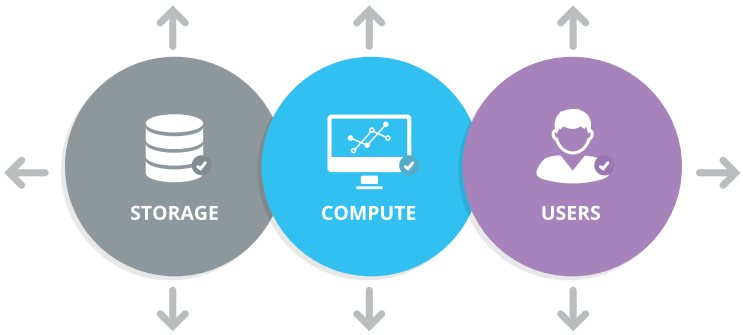
In addition, the cloud is the natural integration point for data. A rapidly increasing share of the data (as much as 80% by some estimates) that you want to analyze comes from applications and systems outside your datacenter—cloud applications like Salesforce, web applications, mobile devices, sensors, and more. Bringing that data together in the cloud is dramatically easier than building the internal infrastructure to hold all of that data —especially given that in many cases you need to experiment with and explore the data in order to determine if and where it has useful insight.

> "Snowflake is the first analytic database that really leverages the power of the cloud." —Jeff Shukis, VP Engineering and Tech Ops, VoiceBase

Snowflake: Reinventing the Data Warehouse

The Snowflake Elastic Data Warehouse is the first data warehouse designed from the ground up for the cloud and for today's needs. Delivered as a software service, it enables you and your analysts to focus on getting value out of your data while delivering the flexibility and performance to meet your ever-evolving needs.

Data warehouse as a software service. Snowflake not only handles deploying and managing infrastructure, we also automate the most resource-intensive activities associated with data warehousing. You no longer have to spend time managing and installing hardware, configuring and updating software, managing data layout, maintaining indexes, or tuning the system.



Snowflake's unique architecture enables it to flexibly support any scale of data, processing, and workloads.

Standard SQL. Snowflake is fully relational and designed for SQL processing from the start, unlike "big data" platforms that have struggled trying to bolt on incomplete SQL support. Use the skills and tools you already have, enabling any SQL analyst and any of the rich ecosystem of SQL-based tools to access data.

"Load and go" for all your business data. Easily combine all of your business data in one place because Snowflake natively supports both structured and semi-structured data in a SQL data warehouse. Snowflake understands and automatically optimizes how semi-structured data is stored and queried, allowing you to correlate structured and semi-structured data in a single system with optimal performance.

Multidimensional elasticity. Snowflake's unique architecture enables it to flexibly support any scale of data, processing, and users. You can independently scale storage, compute, and users up and down without data movement or disruption to deliver exactly the scale you need, exactly when you need it. Snowflake's technology makes it possible to run any number of workloads—loading, querying, ad hoc analysis, and more—at the same time without performance degradation.

A New Standard in Data Warehousing

The Snowflake Elastic Data Warehouse brings a new standard in innovation to data warehousing, innovation that is sorely needed. It offers a highly efficient, cost-effective, and easily manageable data warehouse service designed for today's business needs.

Snowflake is a truly elastic data warehouse, one that can scale up and down on-demand to support any scale of data, processing, and users. That's why Snowflake customers report up to ten times better performance at 90% lower cost than their existing solutions.

To learn more about Snowflake, visit us on the web at at www.snowflake.net or contact us at customer@snowflake.net to schedule a product demonstration by a Snowflake specialist.

About Snowflake

Snowflake Computing, the cloud data warehousing company, was founded in 2012 by a team of experts who hold over 120 patents in the fields of database architecture, data warehouses, query optimization and parallelization. Snowflake was created with the vision to reinvent the data warehouse, bringing together all users, all data and all workloads in a single repository. The company is backed by leading investors including Redpoint Ventures, Sutter Hill Ventures and Wing Ventures. Snowflake is headquartered in Silicon Valley and can be found online at www.snowflake.net.

WP_RDW_1_0_10/2014