



Shifting from Hadoop to the Snowflake Data Cloud

573 of the 2022 Forbes Global 2000 (G2K) use Snowflake Data Cloud to power their businesses.

- Forrester

While Databricks serves as a comprehensive analytics workspace, Snowflake excels in offering scalable and easy-to-use cloud-based data warehousing for structured and unstructured data. Both platforms are equally powerful and designed to address the unique challenges of big data processing and analytics.

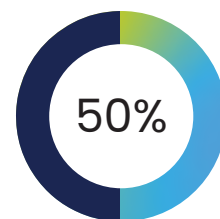
Organizations transitioning their existing Hadoop environments, including the Cloudera, Hortonworks, and MapR Hadoop distributions, to Snowflake can benefit from the latter's capabilities in streamlining analytics, leveraging automation, enhancing scalability, and enabling a more user-friendly approach to managing and deriving insights from data. Hadoop to Snowflake migration involves Schema Migration and Data Migration.

Why do organizations migrate to Snowflake?

Organizations migrating to Snowflake can expect an ROI of

612%

&



reduction in the time-to-market of their products over 3 years.

Source: Forrester

Hadoop Vs Snowflake



Infrastructure Management and Licensing Costs

Hadoop's substantial data center management expenses, coupled with storage and compute costs, and the cost of maintaining skilled resources for its complex maintenance and upgrades can be quite intense. Full-time technical administrators are necessary for managing clusters, upgrades, and security.

Snowflake operates as a managed cloud service and eliminates on-premises infrastructure costs with its pay-as-you-go model. This approach substantially reduces the overall infrastructure management expenses, minimizing the need for dedicated maintenance teams.

Performance and Query Efficiency

Hadoop's traditional batch processing and disk-based storage model often lead to performance challenges, particularly for real-time analytics. High-latency issues can arise, impacting query efficiency.

Snowflake optimizes analytical workloads, providing high-performance results for ad-hoc SQL-based queries and reporting. Its Massively Parallel Processing (MPP) architecture allows scalable and concurrent processing, ensuring efficient querying without performance constraints.

Scalability and Innovation

With companies increasingly connecting data and AI wherever they can, Hadoop faces challenges in meeting the modern expectations of scalability and innovation.

For enterprises seeking a self-service cloud data warehouse with automation and BI prowess, Snowflake becomes a perfect fit. It enables data teams to explore broader use cases, including ML/data science, data marketplace, and data collaboration.

In response to Hadoop's various limitations, Snowflake stands out for its user-friendly interface, impressive performance, seamless functionality, and a track record of uninterrupted operation over the years. However, companies also express their apprehensions about cost management, ecosystem intricacies, governance issues, and the overall ease of management in the Snowflake cloud-based data warehouse.

Migrating From Hadoop to Snowflake

A Case in Point

How a Healthcare Company Saves Cost and Optimizes Operations with Snowflake Data Cloud



Business Challenge

A leading healthcare staffing solutions provider needed timely access to unified data and insights. However, their existing Hadoop-based data platform created data silos and presented data pipeline complexities along with administration and scalability challenges. They faced difficulties in connecting directly to production databases, leading to resource contention and delays in reporting.

Solution

The company switched to Snowflake's robust data platform from the Hadoop environment. Snowflake enabled the creation of a single source of truth. The integration of Microsoft Azure Data Factory with Snowflake alleviated the data pipeline complexities and Microsoft Power BI streamlined data visualization for effective reporting and analytics.

Value Gained

With Snowflake's multi-cluster shared data architecture increasing per-second pricing transparency, the company was able to manage costs far more efficiently. Besides, the near-zero maintenance of the new solution freed up technical resources for other high-value tasks, helping the company further save millions of dollars while improving operational efficiency and reducing complexity. The annual cost savings resulting from this Hadoop to Snowflake transition were around \$2.2 million.

Mapping Hadoop Technologies to Snowflake Components

Hadoop Component	Snowflake Component
Hadoop Distributed File System (HDFS)	Snowflake Object Store
MapReduce	Snowflake's Query Execution Engine
Apache Hive	Snowflake's Virtual Warehouses (for SQL querying)
Apache Pig	Snowflake's Data Loading and Transformation Tools
Apache Spark	Snowflake's Query Processing Engine (for certain workloads)
Apache HBase	Snowflake Variant and Array Data Types
Apache Kafka	Snowpipe for Real-time Data Ingestion
Apache Flink	Snowflake Streaming Data Features
Apache Sqoop	Snowflake Data Loading Tools
Apache Oozie	Snowflake Task Scheduler and Snowflake Stored Procedures



About KANINI

KANINI is a digital transformation enabler, providing cutting-edge software services and solutions that help enterprises drive innovation and business growth. We create impeccable customer experiences through thoughtfully designed digital solutions that help improve our customer's efficiency, scale, and revenues.

We specialize in Cloud Modernization, Data Analytics & AI, Product Engineering, and ServiceNow Consultation and Implementation—all delivered through flexible engagement models.

We focus on empowering Banking and Financial Services, Healthcare, and Manufacturing, among other industries to harness the power of cloud technologies and solutions by implementing agile development practices and a global delivery framework. Find more about our Data Analytics & AI solutions and Consulting services here:

<https://kanini.com/data-analytics-consulting/>

