CLOUDERA DATA WAREHOUSE

Unlock the Power of Your Data

A high-performing, secure, and flexible data warehouse

At a Glance

High-Performance & Intelligent Analytics: Advanced SQL engines for high-speed querying, intelligent automations, AI-assisted SQL, and optimized workload performance.

Secure & Unified Data Platform: End-to-end security, fine-grained access control, governance, and compliance ensure seamless, secure data access across multi-cloud, private cloud, and on-premises environments.

Flexible & Cost-Effective Scalability: Cloud-native, open-source engines with cost-efficient scaling for all data types deliver workload flexibility, full control over your data, and no vendor lock-in.

A Comprehensive Data Platform

At the core of Cloudera's anywhere cloud platform for data and AI is the Cloudera Open Data Lakehouse, which integrates data ingestion, processing, analysis, and storage across hybrid and multi-cloud environments.

Within this platform, Cloudera Data Warehouse provides robust data management and analytics for both structured and unstructured data, and Cloudera Shared Data Experience (SDX) enforces consistent security, governance, and control across all data and analytics workflows.

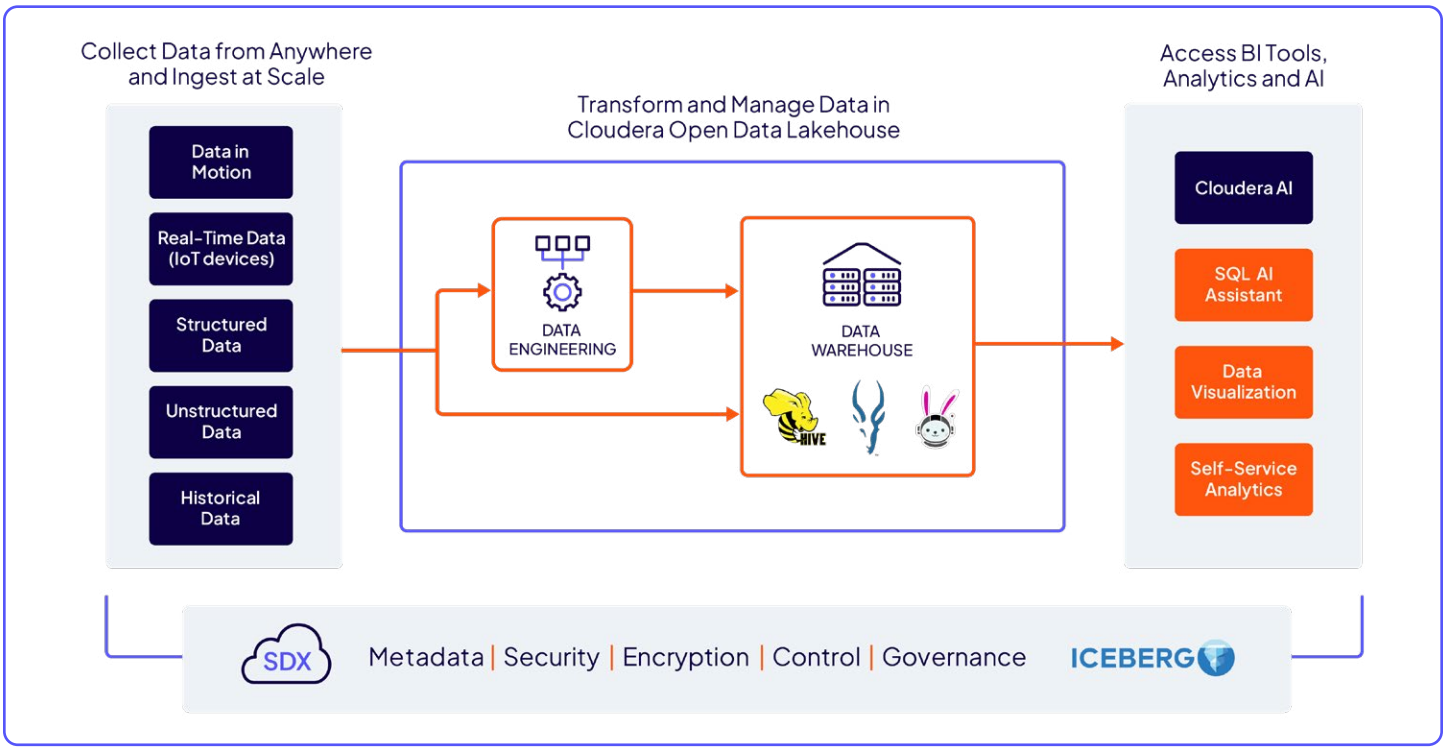

FIGURE 1: Cloudera Data Warehouse is a key offering in the Cloudera Open Data Lakehouse architecture. Collect and Ingest data at scale, transform and manage that data, and unlock critical business insights.

Superior Resource Management

Cloudera Data Warehouse provides intelligent automation and optimization features designed for efficiency at scale. With auto-provisioning and Workload Aware Auto-Scaling (WAAS), resources are dynamically allocated based on mixed workload demands, ensuring consistent performance without manual intervention. Self-service workload management empowers users to manage and prioritize mixed workloads independently, enhancing agility and responsiveness. Built-in cloud optimization ensures cost-effective resource usage across hybrid and multi-cloud environments. Integration with Cloudera Observability provides deep visibility into resource utilization and cost control, enabling informed decision-making and budget adherence. These capabilities result in enhanced financial governance and lower total cost of ownership.

Accelerated Insights and Innovation

Cloudera Data Warehouse enhances decision-making by rapidly processing transactions, events, and other data streams. Its cloud-native PaaS architecture delivers high-performance, low-cost SQL-compute services with a focus on speed, security, flexibility, and control. The integration of Cloudera's AI Assistant allows users to query data using natural language, while Cloudera Data Visualization enables real-time insight generation and interactive analytics. This combination of automation, intelligence, and usability ensures that enterprises can trust and scale their data operations globally and cost-effectively.

High-Performance Analytics for Every Use Case

By leveraging industry-leading SQL engines such as Trino, Apache Impala, and Apache Hive, Cloudera Data Warehouse provides the agility needed for both real-time interactive queries and large-scale batch processing. Support for modern open table formats like Apache Iceberg ensures high-performance analytics on massive datasets, time travel, and concurrency, maintaining consistency and governance across hybrid and multi-cloud environments. Integration with dbt Labs streamlines data transformation workflows to enable faster, more reliable analytics delivery. When integrated with Cloudera Data Engineering and Cloudera AI, organizations can seamlessly execute multi-functional use cases, from business intelligence (BI) to AI-driven insights, with zero ETL and zero data copy.

Data Analytics & Federated Queries	Trino executes federated queries across multiple, disparate data sources, enabling users to combine and analyze data from various systems in a single SQL query. The manageability that Cloudera provides Trino enables data analysts and data scientists to work with diverse data sources without the need for data migration or complex integration processes.
Decision Support & Business Intelligence	Apache Impala is optimized for interactive, low-latency analytics and is best positioned for decision support and BI use cases. With its fast query execution, built-in optimizations, and scalable architecture, Impala is a must-have for data analysts and BI professionals that require fast, interactive analytics and real-time insights on large datasets.
ETL Pipelines & Batch Processing	Apache Hive is best suited for ETL pipelines and batch processing, where large volumes of structured data need to be transformed and prepared for downstream analysis and reporting. Its strength lies in executing complex queries over massive datasets, making it ideal for scheduled, high-volume jobs.

See Cloudera Data Warehouse in action by requesting a demo:
https://www.cloudera.com/products/cloudera-data-platform/cdp-demos.html

CLOUDERA

Cloudera, Inc. | 5470 Great America Pkwy, Santa Clara, CA 95054 USA | cloudera.com

Cloudera is the only true hybrid platform for data, analytics, and AI. With 100x more data under management than other cloud-only vendors, Cloudera empowers global enterprises to transform data of all types, on any public or private cloud, into valuable, trusted insights. Our open data lakehouse delivers scalable and secure data management with portable cloud-native analytics, enabling customers to bring GenAI models to their data while maintaining privacy and ensuring responsible, reliable AI deployments. The world's largest brands in financial services, insurance, media, manufacturing, and government rely on Cloudera to be able to use their data to solve the impossible—today and in the future.

To learn more, visit Cloudera.com and follow us on LinkedIn and X.