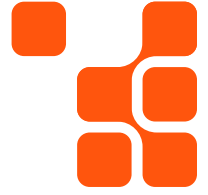


CLUDERA DATA ENGINEERING



Deliver AI and Advanced Analytics with Simplified, Secure, and Cost-Effective Data Pipelines



Your Data Strategy, Engineered for Innovation

As AI and advanced analytics reshape industries, data is the fuel for innovation. It is capable of driving growth, reducing risks, and enhancing operational efficiency. With data volumes skyrocketing and formats diversifying, enterprises continue to struggle with breaking down silos and realizing business value due to the increasing complexity of data.

To fully harness the power of data, organizations rely on data engineering teams to build and maintain data pipelines. However, data teams need more than a simple processing engine or pipeline tool to manage complex data at scale. Cloudera Data Engineering addresses the key challenges that modern enterprises face by providing an end-to-end, practitioner-centric solution for building secure, simplified, and cost-effective data pipelines—whether in the cloud, in data centers, or across hybrid environments—ready to power the next breakthrough.

Business Values

Improved

- Data quality for your AI and analytics needs
- Metadata and lineage via the open data lakehouse
- Productivity powered by DataOps

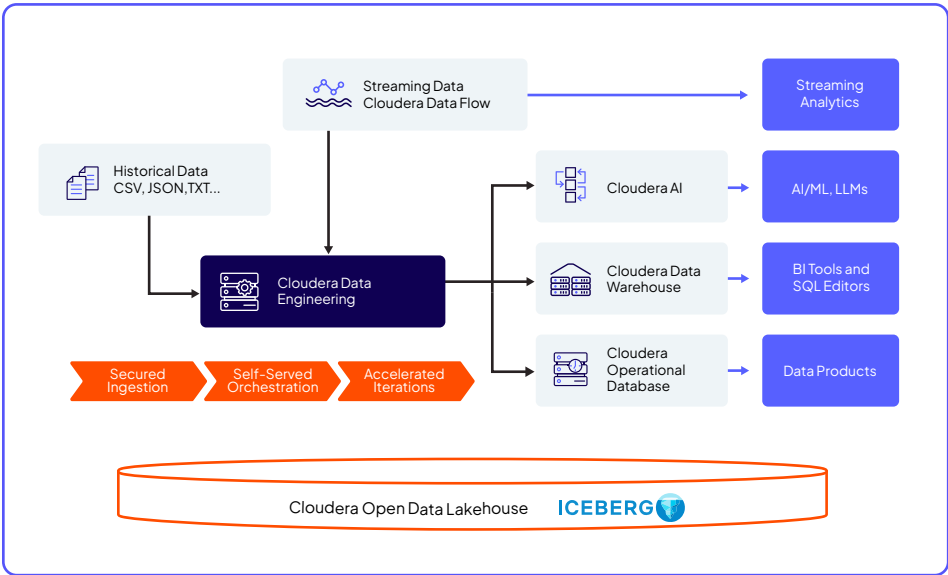
Reduced

- Security risk
- Total Cost of Ownership
- Time to development

Key Challenges of Data Engineering Teams	What We Offer
Integrating Diverse Data Sources	Secure Ingestion, Data Sharing, and Governance <ul style="list-style-type: none"> • Apache Spark is a unified processing framework for both batch and streaming. • Apache Iceberg powers cross-platform data sharing via the REST catalog, with built-in schema evolution and time-travel for governed access. • The open data lakehouse delivers unified data management across hybrid and multi-cloud environments.
Coordinating Data Pipelines at Scale	Simplified Workflow Orchestration <ul style="list-style-type: none"> • Apache Airflow powers workflow orchestration and provides API integrations for external data tools. • Granular access controls and automatic lineage for end-to-end data provenance. • Full lifecycle management from development to deployment, as well as operational monitoring.

Key Challenges of Data Engineering Teams	What We Offer
Managing Cost Efficiency Across Data Workflows	Accelerated Time to Value with DataOps Tooling <ul style="list-style-type: none"> • Self-service orchestration, management, and scheduling. • Interactive sessions for quick iterations. • External IDE Connectivity for remote access to Spark. • Workload-level observability and financial governance. • Visual troubleshooting and performance tuning.

Unlocking Full Potential from Your Data Lakehouse



“Modern data pipelines transcend data center boundaries and are capable of working in dynamic data environments where data can flow seamlessly across multiple data platforms, both on-premises and in the cloud.”

Gartner®, Build Better Data Platforms With Data Engineering Patterns and Practices, September 2024.

GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

Cloudera Data Engineering is central to unlocking the true potential of the open data lakehouse. Beyond building and orchestrating enterprise-grade data pipelines, data teams can tap into Data Services, from Cloudera Data Flow to Cloudera AI, to drive multi-functional analytic use cases from streaming analytics to AI and ML.

With an integrated suite of platform services, Cloudera Data Engineering supports quality data management with Apache Iceberg, the leading open table format purpose-built for managing exabyte-scale data lakes and delivering high-performance queries. It unifies data across hybrid and multi-cloud environments within the data lakehouse, providing consistent and reliable management no matter where the data resides. From transforming raw information into actionable insights to creating data assets for future use cases, Cloudera Data Engineering provides the agility required to build a data-driven organization.

Visit us on our website and learn more: [Cloudera.com/products/data-engineering.html](https://cloudera.com/products/data-engineering.html)



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](https://cloudera.com) and follow us on [LinkedIn](#) and [X](#).