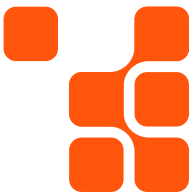


CLUDERA DATA ENGINEERING



# 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. Cludera 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, ready to power the next breakthrough.

### Business Values

#### Improved

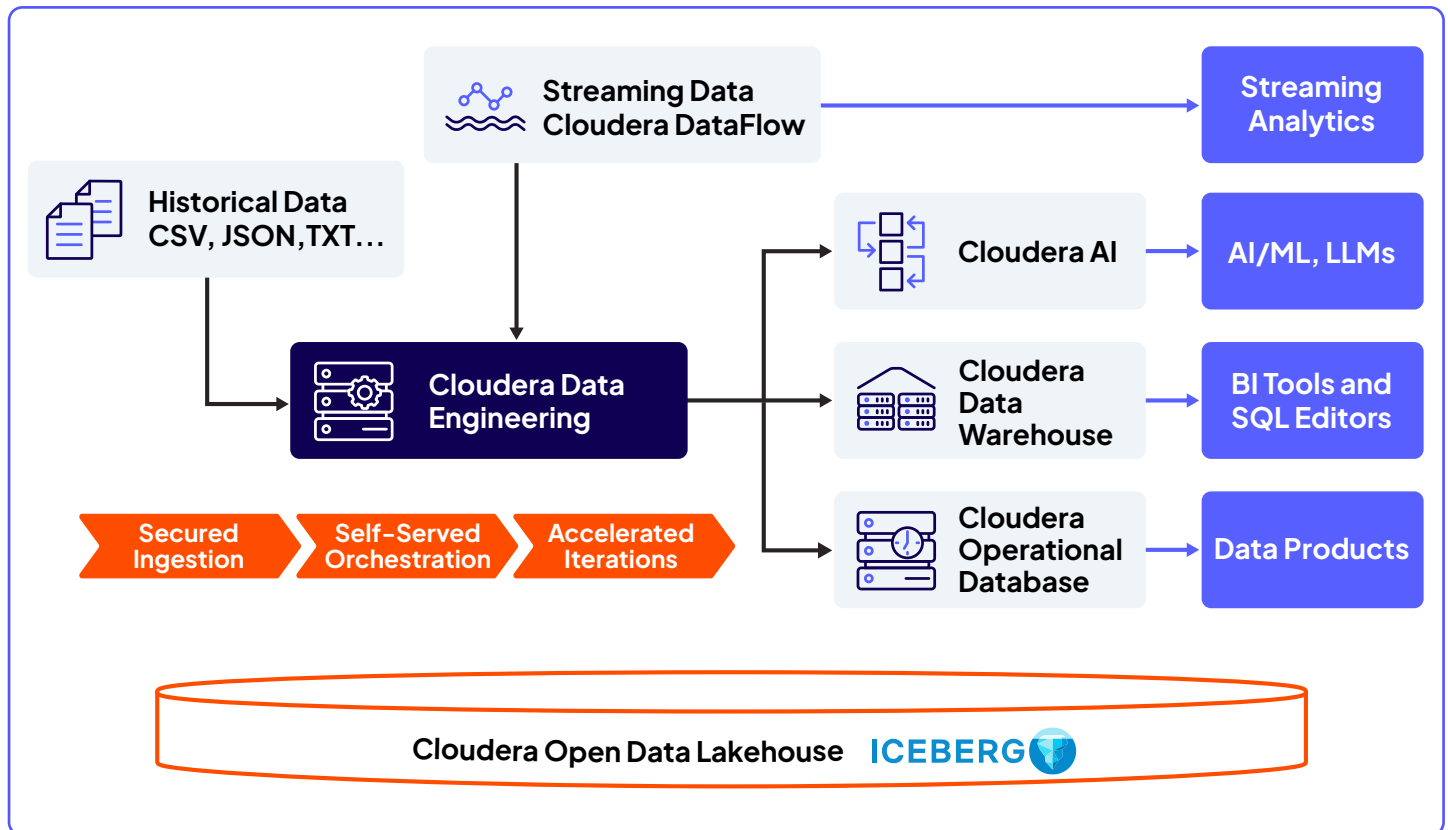
- Data quality for your AI and analytics needs
- Metadata management supported by Cludera’s open data lakehouse
- Productivity with workflow orchestration

#### Reduced

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

KEY CHALLENGES OF DATA ENGINEERING TEAMS	WHAT WE OFFER
Integrating Diverse Data Sources	<div>Secure Ingestion and Enterprise-Grade Data Governance</div> <ul style="list-style-type: none"><li>• Apache Spark is a unified processing framework for both batch and streaming.</li><li>• Apache Iceberg provides enhanced data preparation, including branching, tagging, and time-travel capabilities.</li><li>• The open data lakehouse delivers unified data management across hybrid and multi-cloud environments.</li></ul>
Coordinating Data Pipelines at Scale	<div>Simplified Workflow Orchestration</div> <ul style="list-style-type: none"><li>• Apache Airflow powers workflow orchestration and provides API integrations for external data tools.</li><li>• Granular access controls and automatic lineage for end-to-end data provenance.</li><li>• Full lifecycle management from development to deployment, as well as operational monitoring.</li></ul>
Managing Cost Efficiency Across Data Workflows	<div>Accelerated Time to Value</div> <ul style="list-style-type: none"><li>• Self-service orchestration, management, and scheduling</li><li>• Interactive sessions for quick iterations</li><li>• Workload-level financial governance</li><li>• Operational observability with visual troubleshooting and performance tuning</li></ul>

## Unlocking Full Potential from Your Data Lakehouse



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 Machine Learning, 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 about [Cloudera Data Engineering](#).



**Leveraging the open data lakehouse architecture, Cloudera Data Engineering delivers enterprise-grade data integration while empowering practitioners with secure, democratized access for diverse use cases.**

DIPTO CHAKRAVARTY  
Chief Product Officer at Cloudera

# CLUDERA

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](#).