

Cloud Data Fusion

Fully managed, cloud-native data integration at any scale.

New customers get \$300 in free credits to spend on Data Fusion. All customers get the first 120 hours of pipeline development free per month, per account, not charged against your credits.

- Visual point-and-click interface enabling code-free deployment of ETL/ELT data pipelines
- Broad library of 150+ preconfigured [connectors and transformations](#), at no additional cost
- Natively integrated best-in-class Google Cloud services
- End-to-end data lineage for root cause and impact analysis
- Built with an open source core ([CDAP](#)) for pipeline portability

BENEFITS

Avoid technical bottlenecks and lift productivity

Data Fusion's intuitive drag-and-drop interface, pre-built connectors, and self-service model of code-free data integration remove technical expertise-based bottlenecks and accelerate time to insight.

Lower total cost of pipeline ownership

A serverless approach leveraging the scalability and reliability of Google services like Dataproc means Data Fusion offers the best of data integration capabilities with a lower total cost of ownership.

Build with a data governance foundation

With built-in features like end-to-end data lineage, integration metadata, and cloud-native security and data protection services, Data Fusion assists teams with root cause or impact analysis and compliance.

Key features

Open core, delivering hybrid and multi-cloud integration

Data Fusion is built using open source project CDAP, and this open core ensures data pipeline portability for users. CDAP's broad integration with on-premises and public cloud platforms gives Cloud Data Fusion users the ability to break down silos and deliver insights that were previously inaccessible.

Integrated with Google's industry-leading big data tools

Data Fusion's integration with Google Cloud simplifies data security and ensures data is immediately available for analysis. Whether you're curating a data lake with [Cloud Storage](#) and [Dataproc](#), moving data into [BigQuery](#) for data warehousing, or transforming data to land it in a relational store like [Cloud Spanner](#), Cloud Data Fusion's integration makes development and iteration fast and easy.

Data integration through collaboration and standardization

Cloud Data Fusion offers [pre-built transformations](#) for both batch and real-time processing. It provides the ability to create an internal library of custom connections and transformations that can be validated, shared, and reused across teams. It lays the foundation of collaborative data engineering and improves productivity. That means less waiting for ETL developers and data engineers and, importantly, less sweating about code quality.

All features

Code-free self-service	Remove bottlenecks by enabling nontechnical users through a code-free graphical interface that delivers point-and-click data integration.
Collaborative data engineering	Cloud Data Fusion offers the ability to create an internal library of custom connections and transformations that can be validated, shared, and reused across an organization.
Google Cloud-native	Fully managed Google Cloud-native architecture unlocks the scalability, reliability, security, and privacy features of Google Cloud.
Real-time data integration	Replicate transactional and operational databases such as SQL Server, Oracle and MySQL directly into BigQuery with just a few clicks using Data Fusion's replication feature. Integration with Datastream allows you to deliver change streams into BigQuery for continuous analytics. Use feasibility assessment for faster development

	iterations and performance/health monitoring for observability.
Batch integration	Design, run and operate high-volumes of data pipelines periodically with support for popular data sources including file systems and object stores, relational and NoSQL databases, SaaS systems, and mainframes.
Enterprise-grade security	Integration with Cloud Identity and Access Management (IAM) , Private IP , VPC-SC and CMEK provides enterprise security and alleviates risks by ensuring compliance and data protection.
Integration metadata and lineage	Search integrated datasets by technical and business metadata. Track lineage for all integrated datasets at the dataset and field level.
Seamless operations	REST APIs, time-based schedules, pipeline state-based triggers, logs, metrics, and monitoring dashboards make it easy to operate in mission-critical environments.
Comprehensive integration toolkit	Built-in connectors to a variety of modern and legacy systems, code-free transformations, conditionals and pre/post processing, alerting and notifications, and error processing provide a comprehensive data integration experience.
Hybrid enablement	Open source provides the flexibility and portability required to build standardized data integration solutions across hybrid and multi-cloud environments.

What is Cloud Data Fusion?

Cloud Data Fusion is a fully managed, cloud-native, enterprise data integration service for quickly building and managing data pipelines.

The Cloud Data Fusion web UI lets you to build scalable data integration solutions to clean, prepare, blend, transfer, and transform data, without having to manage the infrastructure.

Cloud Data Fusion is powered by the open source project [CDAP](#).