# Simplifying Big Data in a Multi-Cloud World with Streamlined Workflows

Introduction

Big data projects often entail moving data between multiple cloud and legacy on-premise environments. A typical scenario involves moving data from a cloud-based source to a cloud-based normalization application, to an on-premise system for consolidation with other data, and then through various cloud and on-premise applications that analyze the data. Processing and analysis turn the disparate data into business insights delivered through dashboards, reports, and data warehouses—often using cloud-based apps.

The workflows that take data from ingestion to delivery are highly complex and have numerous dependencies along the way. Speed, reliability, and scalability are crucial. So, although data scientists and engineers may do things manually during proof of concept, manual processes don't scale.

THE BIG DATA PROCESS

Processing

Ingestion

Analysis

Delivery
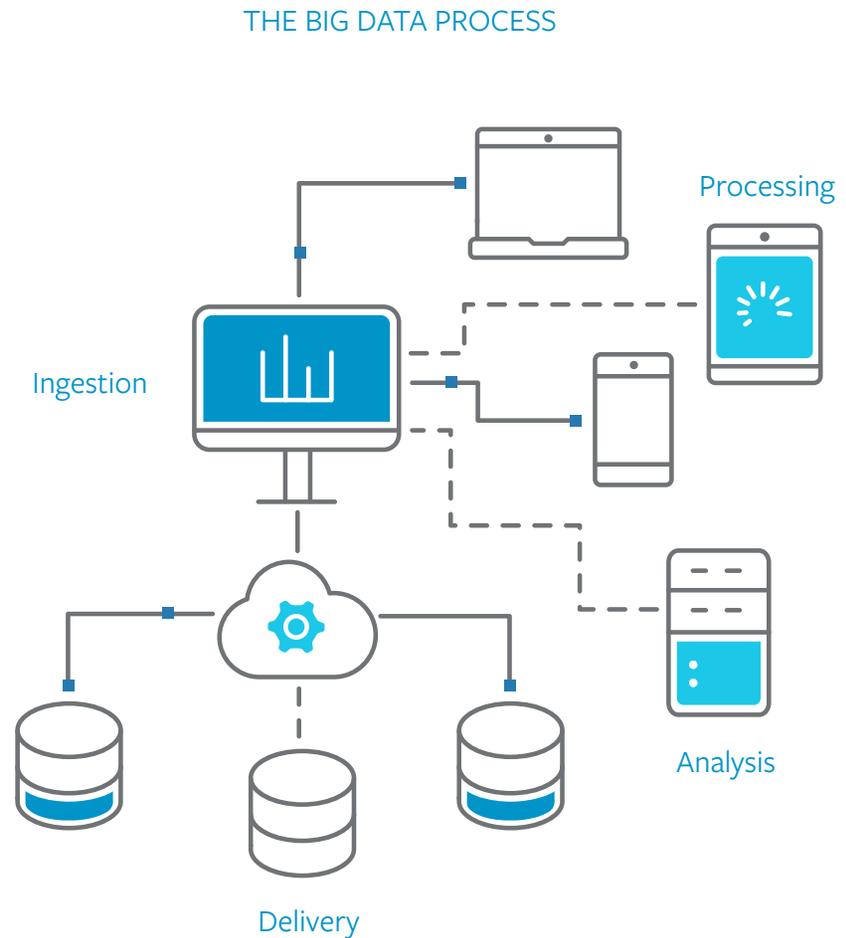
Automation and orchestration simplify Big Data in the cloud

To deal with the complexity, you need industrial-strength workload automation and business application orchestration capabilities. Many tools can orchestrate big data workflows. Some of them— such as Oozie or tools bundled with public cloud services such as Amazon Web Services (AWS) and Microsoft Azure—are "free."

But most of those tools are platform-specific and limited in functionality. So, you have to cobble together multiple tools to orchestrate complex workflows across multi-cloud and traditional environments. That's when free becomes expensive. Not only are tool integrations costly, they can also drag down end-to-end process reliability and your productivity.

Big data drives important business decisions. You need a reliable, fail-safe way to automate and orchestrate every step of big data processing across all involved environments.

Automation and orchestration:

- Accelerate big data processing

- Free up time for engineers and data scientists to focus on strategic efforts

- Speed the delivery of business insights to the consumers of data

Orchestrating the big data lifecycle

Control-M speeds the implementation of big data projects by replacing manual scripting with automated workflow management and data integration. Control-M also provides visibility into workflows with an end-to-end picture of data pipelines at every stage. Having visibility into service level agreements (SLAs) for service delivery SLAs enables you to resolve critical issues before deadlines are missed.

Control-M in Action

A leading manufacturer of commercial trucks, buses, defense vehicles, and engines, leverages big data technologies to improve its products and create innovative offerings and new revenue streams. The company captures millions of data points every day and transforms them into business insights.

1

They partner with a company that collects data from its vehicles and stores it on public cloud.

2

Control-M pulls the data from the cloud and puts it into Hadoop where it can be tapped with queries built by the engineers.

3

The big data team uses Tableau to process and display the information on dashboards.

Control-M orchestrates this complex process, freeing up engineers who previously spent all their time pulling data from the cloud, aggregating it, and putting it into spreadsheets. Plus, design engineers now have real-time access to data on dashboards.

Delivering the goods

Getting insights into the hands of the right people—whether they are employees within your business or customers who rely on actionable information to run their businesses—is a critical step in the big data lifecycle. Delivery often occurs via the cloud in the form of dashboards, traditional reports, and SQL queries into data warehouses.

Control-M in Action

A large provider of consumer data draws on more than 100,000 sources, processing more data in an hour than many businesses process in a month. The company transforms this data into meaningful reports that help millions of consumers make informed buying decisions.

- Control-M runs 130,000 processes daily across 350 nodes in five data centers, managing data ingestion and movement, traditional workflows, and SLA compliance.

- Much of the processing and analysis happens on premise but most customer-facing front-end systems for delivery run in public cloud.
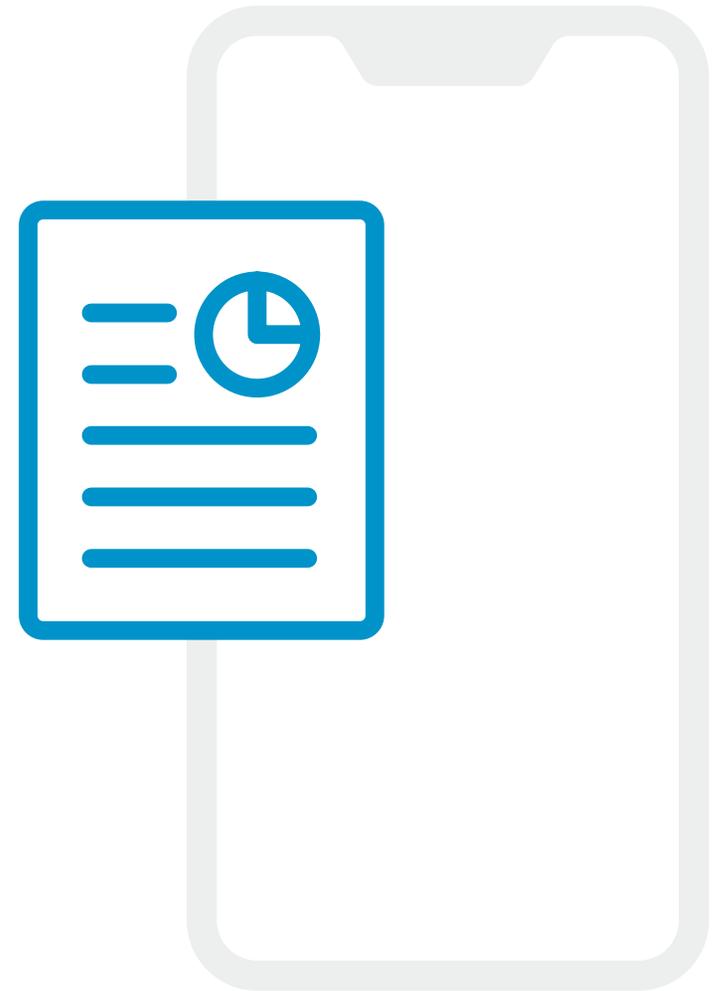
Control-M for Big Data

Control-M overcomes big data complexity by automating and orchestrating complex application workflows that span multiple applications across multi-cloud and on-premise environments. Native support for Hadoop, traditional platforms and applications, and file transfers, further simplifies Hadoop workflow processing.

For more information

Learn more about Control-M for Big Data **here**.

Explore the advantages of Control-M on the Cloud **here**.

See Control-M in action **here**.

506972