

HOW CONTROL-M HELPS AT EVERY STAGE OF THE BIG DATA JOURNEY



Getting started with Big Data is like preparing for a cross country drive. You're not going to get it done in a day. You have to consider different routes – in the Big Data world, this means '*What platform will you use?*'; '*How will you set up the IT architecture?*' etc. At various points in the journey, you'll need to find bridges (from existing data sources and IT infrastructure to the Big Data environment) and there will be tolls to pay (planning and development time, and possibly new tools you may need to buy). You have to pack carefully – you don't want to bring a different suitcase for every day, just like you don't want to be burdened by a different set of software solutions and infrastructure for every stage of the journey.

BMC can help. We can't replace the drive with a direct flight where you arrive at your Big Data destination a few hours after beginning the journey. But we can help you navigate, and even do a lot of the driving for you. We do that by: providing a bridge for getting your data from your existing systems to the Big Data environment, giving you tools to simplify Big Data workflow management, making your workflow development, scheduling and execution consistent and compatible between your existing and Big Data infrastructures, and saving time by automating at every stage of the journey.

In this blog, we invite you to take a road trip with us - starting by loading the car with data for the drive.

Ingesting data

One of the first forks in the road on a Big Data journey is deciding how to ingest the source data that will fuel the Big Data program. The raw data that ultimately becomes Big Data insights is often in forms that existing systems do not support – such as social media streams, Internet of Things (IoT) input, output from machine learning, customer service call recordings, plus more traditional structured data from ERP and other enterprise systems. There are open source tools for working with these data sources (for example, in the Hadoop world there is Sqoop for ingesting data from RDBMS sources and Flume for streaming data) and file transfer for traditional data sources. Working with multiple, single-purpose tools is akin to driving cross-country on two-lane roads instead of the highway – it's slow and presents a lot of problems with wrong turns.

Control-M helps you execute data ingestion without slowing down. Control-M automates file transfer for reliable, automated execution across existing and Big Data environments, both on-premise and in the cloud. It also supports Sqoop and the ETL functionality embedded in many leading Big Data and business intelligence solutions, including Cognos, Informatica, Oracle Business Intelligence, SAP Business Objects, and SQL Server SSIS – plus the Cloudera, Hortonworks, and MapR Hadoop distributions and the IBM Big Insights Distribution. With Control-M, you only have to go down one road to meet all your ETL needs for Big Data and other environments.

The big turn: turning data into value

Once data is ingested it needs to be transformed and made valuable. That is done when the data is processed, which in turn is done by the workflows that are developed. Here Control-M can guide you. You don't have to sort through and select new toolsets for Big Data, or slow down your development while you learn to use them. Control-M simplifies and automates Big Data development and execution in several ways:

- Control-M automates many steps in the development, testing, scheduling, promotion and execution processes.
- It lets developers and operations staff work in their familiar environments. [Control-M Automation API](#) is a set of programmatic interfaces (both APIs and CLIs) that let developers and [DevOps](#) engineers use Control-M in the agile application release process. Now Big Data jobs can be developed as code, by embedding workflow automation in the application while it is being developed. The Jobs-as-Code approach makes the development environment identical to the production environment and thereby saves time by preventing many common failures and routine delays that occur when workflows are tested and promoted to production.
- Meanwhile, operations can schedule and execute Big Data jobs just like any other enterprise workflows – with no separate solutions or new scripting required. If you were thinking of going down the Oozie road, [don't take it](#).

You've arrived, we'll unpack

The final destination is the delivery of new insight to business users. This often requires delivering your Big Data output to data visualization and business intelligence applications. Control-M drives these processes by automating data transfers and workload execution – applying predictive analytics to prevent job failures; automatically retrying jobs that were interrupted; and presenting user dashboards and self-service capabilities. Business users get the insights they need, the

operations staff gets to be proactive because many Big Data processing tasks are automated, and the development team gets to focus on delivering new services instead of debugging earlier ones. Creating insight from Big Data is a journey. We can help you by providing automated navigation at every turn.