CONTROL-M PROVIDES QUALITY CONTROL FOR BMC'S CUSTOMER ZERO DATA WAREHOUSE

At BMC we learned the hard way that our enterprise data warehouse was much better at managing data volume than it was at managing data quality. For example, our data volumes are growing insanely fast and the load process may take hours. We can't afford not to know where we are in the load process, especially if it fails and we have to recover. In today's analytical world, you can't tell business users their jobs now will be ready in 12 hours instead of six because something broke earlier and we didn't know about it. Unfortunately, that's what used to happen. We also got burned many times when reports were wrong because the data was only partially refreshed.

It became clear we needed to build more quality control and automation into our workflows, especially for extract, transfer and load (ETL) operations involving our enterprise data warehouse. Fortunately, we also learned that Control-M could help us with workflow management and quality control in several ways.

Life Without Automation

Let me set the scene. BMC's enterprise data landscape requires the ingestion of many different data sources. This includes numerous SaaS applications like SalesForce.com and Eloqua that use web services, on premise applications and databases like Oracle® CRM, flat files, unstructured data and other external sources that require custom parsing. Getting these sources into the warehouse for loading, processing and integration requires many different tools and processes. Managing the numerous tools and orchestrating the complex load process is where we felt the most pain.

In the early days, we leveraged custom scripting across the different tool sets to attempt to manage

the hybrid data landscape. This became impossible to manage as the environment grew, and new technologies where added which greatly increased the complexity. Custom scripting is not scalable or manageable across many different tools.

Our data volumes began to scale significantly as we got more involved in analytics, business intelligence, big data, cloud and hybrid infrastructure, and we began having more data quality problems. We were also using more sources of data, and had to find ways to validate the data before it went into our enterprise data warehouse.

Two problems surfaced at this stage of our digital development. First, we hit the scalability limitations of using multiple toolsets and scripting. Second, the resulting consequences went from being an IT/Operations problem to becoming a business problem. Data quality issues could cause people not to be paid on time. Bad data might cause jobs to fail, or might not be apparent until jobs and reports were complete, which would necessitate the data being updated and the entire job having to be rerun. Situations like that threatened our ability to complete quarterly closing on time. Many of the problems we experienced are preventable. For example, we can validate data by making sure all the required rows and columns are in place, and that values are within normal ranges. But that takes time and effort – time we didn't have.

Integrate. Automate. Orchestrate. – Life After Control-M

We knew improving the quality of data going into our data warehouse, and improving the orchestration of data warehouse-dependent jobs with systems throughout the enterprise, were what we needed to satisfy our volume and speed pressure. When we dug deeper, we determined it was critical to have an abstraction layer that can sit above applications to manage ETL, workflows and custom scripts.

Control-M lets us do that. We can develop workflows that manage the load process across the disparate tools just like we do for any other job – we don't need separate tools or scripts. And because everything is integrated, we can get complete visibility on job status. If an Informatica job is delayed because it can't start until another workflow completes (for example a file transfer or ETL transaction) we'll be proactively alerted. We no longer have to manually check five or six systems to get different status reports so we can figure out when a job will complete. We no longer have to tell anyone that the job that was supposed to be ready today isn't even going to start for another 12 hours. Plus, with Control-M Self Service, business users can monitor their own jobs, on their cellphones.

We can also do some data validation. Before, we might have been able to do a rudimentary quality check, for example to check whether a needed file was there or not. Now we can go further and know if there is anything in the file, if the file is in the right format, if it loaded correctly and more.

Sometimes bad data still happens, but now it's OK. Instead of 12 hours to learn about the problem and recover, now we can identify the problem, get an answer and move on in five minutes.

Click here to learn more about how Control-M integrates with Informatica and other business applications to manage ETL operations and much more.