Cisco IT Automates Workloads for Big Data Analytics Environments

Cisco Tidal Enterprise Scheduler eliminates time-consuming script management, integrates with enterprise applications, and includes a mobile app.

EXECUTIVE SUMMARY

CHALLENGE
- Keep up with growing demand for internal big data analytics
- Help developers become productive more quickly
- Avoid spending time developing scripts to connect to various data sources and applications

SOLUTION
- Replaced previous scheduling tool with Cisco Tidal Enterprise Scheduler
- Used built-in adapters to connect to various data sources and applications

RESULTS
- Eliminated steep learning curve for previous scheduler by providing intuitive browser interface
- Accelerated deployment of services from months to hours in some cases
- Placed heartbeat-monitoring process in production in just 30 minutes
- Improved work-life balance for Cisco IT staff by providing a mobile app for monitoring and managing processes

NEXT STEPS
- Use Cisco Tidal Enterprise Scheduler with JAWS Predictive Analytics to assign priority to workloads on SLA critical path

Challenge

Cisco business units are adopting big data analytics to unlock the business intelligence in vast amounts of data with the goals of increasing sales, improving the customer experience, enhancing product quality, and more. As one example, Cisco IT developed a big data analytics solution that processes 1.5 billion customer records daily to identify partner sales opportunities. Anticipated incremental revenue from this solution, which is in production, is approximately US$40 million in FY13 alone.

To keep up with burgeoning demand for big data analytics, Cisco IT needed an easy-to-use workload automation solution. Depending on the business need, the tool would have to orchestrate processes involving Hadoop, Informatica, Teradata, SAP HANA, SAP BusinessObjects, and other custom and enterprise applications and data warehouses.

"The third-party scheduler that we used initially was difficult to learn," says Sudharshan Seerapu, a Cisco IT engineer specializing in workload automation for big data services. "One problem was a complex command-line interface (CLI), which took months to master even for experienced data warehousing software engineers."

Another drawback of the original scheduling tool was the complexity of connecting with the various components in the big data processing environment. "We had to write shell scripts to connect to each application," Seerapu says. "Development and debugging took months for each application, postponing the business value."

Finally, the previous scheduling software required one physical or virtual server for every application, which meant that the operational burden scaled linearly. At one point, Cisco IT had to manage more than 200 separate scheduling servers.

Solution

Cisco IT found a simpler, more scalable solution to automate workloads for big data applications by adopting Cisco® Tidal Enterprise Scheduler (TES), an end-to-end workload automation solution with built-in adapters for

Hadoop Sqoop and Hive tools, as well as for leading enterprise resource planning (ERP), databases, data warehouses, data integration, and business intelligence applications (Figure 1). "We use Cisco Tidal Enterprise Scheduler to manage big data workloads that move data in and out of application data sets and to execute big data jobs within Hadoop," Seerapu says. Figure 1 shows TES integration points and the basic architecture.

In addition to using TES to schedule time-based batch jobs, Cisco IT uses it to automate workloads by taking actions based on events. For example, an event such as the creation of a new customer record might trigger the action of moving a set of records into a data warehouse. Other examples of events that can trigger an action include a self-service request for a report, a change to the contents of an FTP folder, or email actions.

Figure 1. Cisco Tidal Enterprise Scheduler Architecture
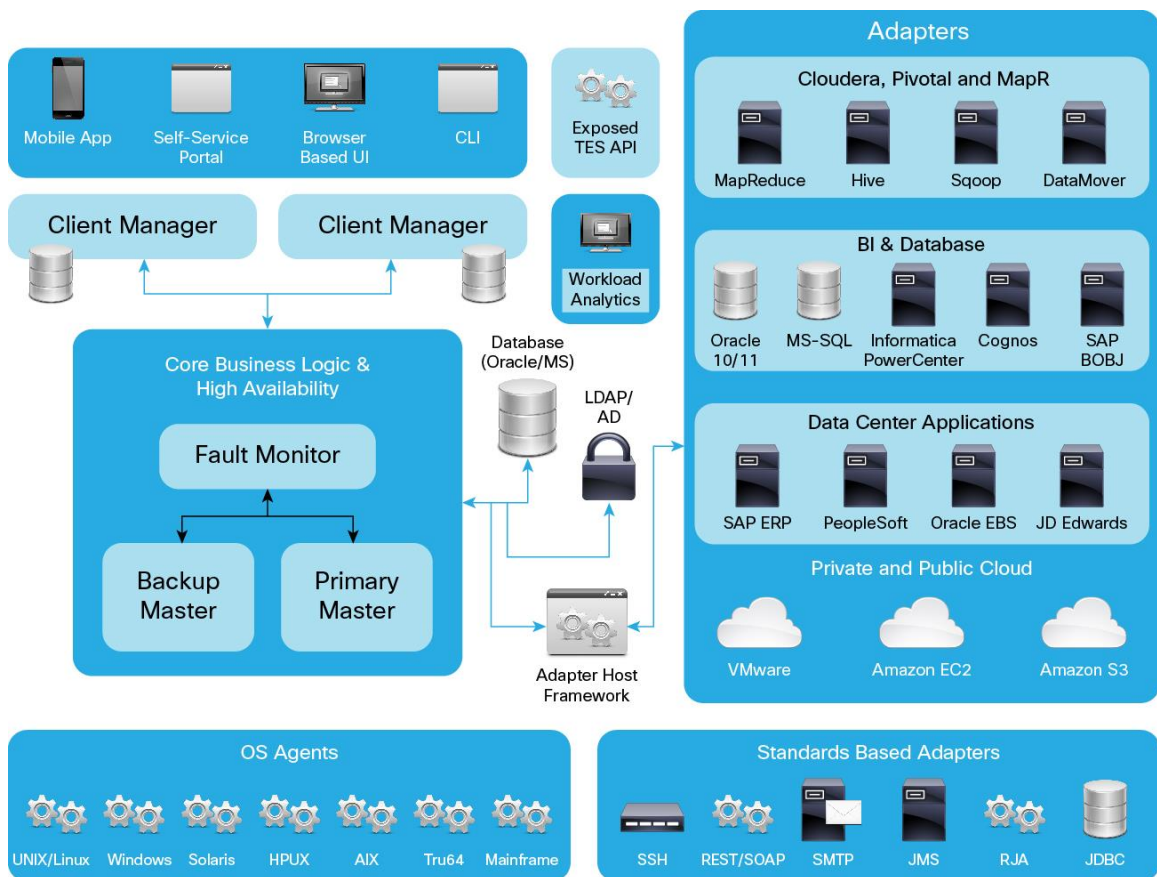
TES provides important operational advantages in Cisco IT's big data environment that the previous tool lacked:

- Simplifies the IT infrastructure because its single, centralized server can initiate all jobs and events.
- Scales to meet growing demand for big data analytics from Cisco users because it can manage many thousands of jobs daily.
- Maximizes the data sources that Cisco can mine for business intelligence because it has adapters for nearly all data sources in the enterprise.
- Includes a mobile app, which Cisco IT engineers appreciate because it enables them to receive alerts, check logs, and remediate errors from anywhere rather than having to drive to the office at night or on weekends (Figure 2).
- Provides a self-service portal for internal users to submit jobs, increasing user satisfaction while offloading Cisco IT administrators.

Figure 2. Mobile App Enables Cisco IT Engineers to Manage Workloads from Anywhere

Use Cases

Cisco IT has identified more than 10 big data analytics use cases for TES. The following use cases are in production or in the proof-of-concept stage.

> "We estimate that using Cisco Tidal Enterprise Scheduler for end-to-end data management reduced development time by 90 percent compared to using traditional scripting tools."
> — Sudharshan Seerapu, Cisco IT Engineer

Use Case 1: Coordinate Database, Data Warehouse, and Hadoop Process for Sales-Related Data Mining

The first production application to use TES is the 360 Data Foundation, which Cisco IT developed to identify Cisco Services sales opportunities for partners. The application uses Informatica to transform data from various data sources, Hadoop to store and process the data, and Teradata to store the data for analysis (Figure 3). Using built-in adapters, TES coordinates all workloads within this processing environment. "We estimate that using Cisco Tidal Enterprise Scheduler for end-to-end data management reduced development time by 90 percent compared to using traditional scripting tools," says Seerapu.

Figure 3. Cisco Tidal Enterprise Scheduler Orchestrates Workloads for Big Data Analytics Application That Identifies Revenue Opportunities
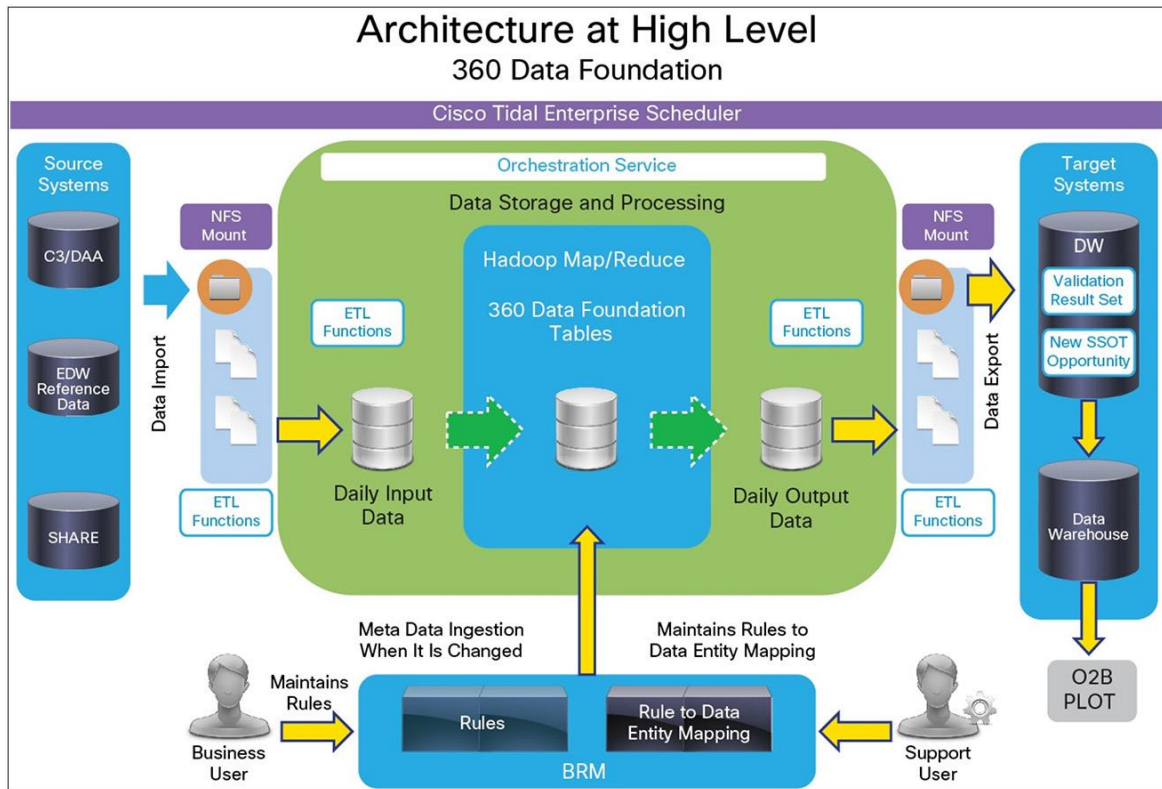
Use Case 2: Monitor Big Data Analytics Application Performance

Cisco IT developed an executive dashboard, which operates on the Cisco Unified Computing System™ (UCS®), to display sales metrics based on SAP HANA. Executives eagerly anticipated this big data analytics tool, but the day before it was scheduled to go live, Cisco IT discovered a major reliability issue: although Cisco IT's internal Enterprise Management (EMAN) application monitored the Cisco UCS host, the application itself was not monitored. Cisco IT quickly resolved the problem by using the built-in Web Services adapters in TES to poll for application status every 5 to 10 minutes. "If an abnormal string appears, Tidal Enterprise Scheduler can execute an event-based process to patch the application," Seerapu says. Cisco IT placed the new workload into production in just 30 minutes, enabling the highly visible launch to take place as planned.

Use Case 3: Monitor Hive Process

Cisco IT also uses TES to monitor the health of the 360 Data Foundation, the platform that identifies sales opportunities. In this case, Cisco IT developers used TES to create a Hive heartbeat process. If an abnormal string appears, TES sends an email or pager alert to a designated administrator. "We estimate that writing scripts for this process would have taken one month," says Seerapu. "With Cisco Tidal Enterprise Scheduler, we were able to place it in full production in just one day."

Results

Using TES is helping Cisco IT scale to meet growing demand for big data analytics by minimizing training time, accelerating deployment, and reducing management burden.

Minimized Training Time

"The biggest advantage of Cisco Tidal Enterprise Scheduler compared to our previous tool is the ease of use, which accelerates deployment and delivers value to our business units more quickly," Seerapu says. For example, developers no longer need to learn to use Sqoop to write a script to import or export data sets. "Tidal Enterprise Scheduler prompts for the information it needs, such as the table, the format, and the user names of administrators who can run the job," says Seerapu.

> "The biggest advantage of Cisco Tidal Enterprise Scheduler compared to our previous tool is the ease of use, which accelerates deployment and delivers value to our business units more quickly. Tidal Enterprise Scheduler prompts for the information it needs, such as the table, the format, and the user names of administrators who can run the job."
> **— Sudharshan Seerapu, Cisco IT Engineer**

Accelerated Deployment of Big Data Applications that Improve Sales, Service, and Quality

The previous scheduling tool required a dedicated physical or virtual server for each instance, increasing costs and delaying the availability of applications to improve sales, service, or quality. "Before, we had to deploy infrastructure for each big data application," Seerapu says. "Now, with Cisco Tidal Enterprise Scheduler, all applications access the same centralized server, which reduces costs."

Improved Work-Life Balance by Providing a Mobile App

Many business intelligence and analytics applications operate 24 hours a day. To handle a high volume of these jobs - many with service-level agreements (SLAs) - Cisco IT needs the tools to receive and respond to alerts from anywhere. "Using the Cisco Tidal Enterprise Scheduler client on our iPhones and iPads, we can receive alerts, log in to learn more about jobs, and respond by holding or restarting jobs," says Seerapu.

Next Steps

Cisco IT has standardized on TES and will use it for all future workload automation.

The team is evaluating numerous other use cases. One is increasing the resources available to big data workloads using Terma Labs' JAWS, a workload analytics tool that performs critical-path analysis and can alert TES about workload constraints. For example, suppose an operation discovers that 2 of the 10 Informatica jobs running are critical. The plan is that Cisco IT will use TES to view this information and then increase the priority of these jobs in real time. "Today, we can't handle critical-path SLAs," Seerapu says. "Gaining this ability will improve productivity, lower operational costs, and enable IT to develop more services in less time."

For More Information

To read additional Cisco IT case studies on a variety of business solutions, visit Cisco on Cisco: Inside Cisco IT.

For more information on Cisco Common Platform Architecture (CPA) for big data analytics, visit:

- Introducing Cisco UCS Common Platform Architecture (CPA) for Big Data
- Big Data
- Cisco UCS Common Platform Architecture Version 2 for Big Data

Note

This publication describes how Cisco has benefited from the deployment of its own products. Many factors may have contributed to the results and benefits described; Cisco does not guarantee comparable results elsewhere.

CISCO PROVIDES THIS PUBLICATION AS IS WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Some jurisdictions do not allow disclaimer of express or implied warranties, therefore this disclaimer may not apply to you.