

Automic[®] Workload Automation

Benefits of Automating Data Warehouses and Big Data

Executive Summary

This white paper explains how Automic[®] Workload Automation can integrate all the software used in data warehouse systems into enterprise process automation, ensuring the timely delivery of trusted BI reports to your business users.

Challenge

In today's economic climate, reliable and up-to-date business intelligence (BI) is crucial for decision makers to deliver world-class business performance and stay on top of their competition. Trusted and fast data warehouses are key for delivering business-critical reports to business users at the right time, anywhere.

The amount of data that needs to be collected across a company's business units, applications, and external sources grows exponentially. Additionally, more and more applications are hosted as a service in the cloud and within big data instances such as Hadoop.

Ensuring that information is available at the right time to make strategic, tactical, and operational decisions can be a challenge. Typical data warehouse operations deal with extremely large amounts of data. Missing one step in the process, or executing a step at the wrong time, can result in a significant amount of wasted processing time—or, in the worst case scenario, bad data.

Opportunity

Often, the data needed for the data warehouse is extracted from a variety of systems and applications to a central repository. It is not unusual for a corporation's applications to reside on different platforms, such as mainframes, UNIX, AS/400, Microsoft Windows, and big data in the cloud. These applications typically store data in a mix of different databases and/or non-relational data formats. For example, in application A, gender may be stored as *Male*, in application B as *M* and in application C as *01*. A data warehouse converts the data to a uniform format and provides the ability to view, compare and correlate information at an enterprise level.

Benefits

A data warehouse provides content for information displayed in BI reports or portal pages. Additionally, data warehouses can be the backend to custom-developed enterprise information systems (EIS).

The Process

There are five major steps in a data warehouse process. The first three involve the building and subsequent loading of the data warehouse, commonly referred to as extraction, transformation and loading (ETL). The last two steps consist of creating data marts and BI reporting.

1. Extracting data from applications.

In this step, the data is extracted from the application's data structure and most often placed in a temporary data structure, or occasionally left in flat files.

2. Transforming, or staging, the data (getting it in the proper format).

In the transformation step, the different data formats are reconciled. For example, all data formats for gender are transformed in a uniform way to either *M* or *F*.

3. Loading the transformed data into the data warehouse.

4. Creating data marts from the data warehouse.

5. BI/reporting.

In this step, the data is turned into valuable information and distributed to the right person at the right time.

The Problem

Running a data warehouse requires the coordination of many operations across many applications, databases, and systems. In a large company, this might be as many as 20 discrete operations that must be performed in the correct sequence, at the correct time, under the correct conditions. Typical data warehouse operations deal with extremely large amounts of data. It is not unusual for a large organization to load terabytes of data nightly; data warehouses can grow to sizes larger than one petabyte. Missing one step in the process, or executing a step at the wrong time, can result in a significant amount of wasted processing time or, in the worst case scenario, bad data.

The Product

Automatic Workload Automation can automate data warehousing processes end-to-end.

- Each step in the process becomes part of end-to-end process flows.
- Process flows can be run on a set schedule, and individual steps in the process can be executed based on a wide range of criteria, such as the number of transactions or the number of records in the data warehouse.
- Process steps are application-aware and ensure that required application tasks are run prior to data extractions.
- Data extracts, often long processes, can be checked periodically to ensure that the data being extracted is in the correct format.
- Notification of nightly failures can be automated.
- Complex recovery sequences can be automated.
- Reports based on big data tools such as Hadoop can be run and distributed automatically as part of the process flows.
- Workloads on the systems can be leveled using work queues and dynamic load balancing.
- The whole data warehouse environment is automated, including all applications, scripts and tools such as:
 - Enterprise applications such as ERP, CRM, and Financials.
 - Databases such as Oracle or Microsoft SQL Server, and big data platforms like Hadoop.
 - ETL tools, and OLAP tools.
 - BI reporting tools.

Benefits

There are many benefits of using Automic Workload Automation to automate, accelerate, and control your data warehousing processes:

- Faster data integration cycles
 - Removes any latency between different vendor tools and applications
 - Balances the load placed on the servers by the enterprise applications, ETL tools and reporting tools to achieve optimum throughput without overloading your systems
 - Provides for the proper execution sequence, resulting in fewer errors and recoveries
- Immediate report update and distribution
 - Immediate update of reports when ETL has finished
 - Immediate distribution of updated reports to business users and portals
- Faster implementation of new data warehousing processes
 - Script-less, standardized automation of ETL and BI processes
 - Templates for quick implementation
 - Reduce development and maintenance efforts
- Facilitates compliance of your data warehousing process
 - Every task is visible
 - Every execution is tracked
- Monitoring and control
 - Full visibility of the entire process
 - Notifications in case of error

Background

A data warehouse is a very large database containing data from many sources in an enterprise. The data warehouse is at the center of an entire data collection and distribution process. This process extracts the data from enterprise sources such as ERP, CRM, and financial systems. These sources often reside on different platforms, such as UNIX, Windows, mainframe, and AS400. These can be hosted on premises or in the cloud.

If doing traditional ETL:

- Transforms and cleanses the data so that data coming from the different sources is standardized or normalized. In the transformation step, the different data formats are reconciled. For example, all data formats for gender are transformed to standardized format—either *M* or *F*.
- Places the normalized data in a staging area.
- Loads the data into the actual data warehouse database and operational data store. The data warehouse is usually a relational database running on a UNIX platform.

If using big data:

- Aggregates raw data inside platforms like Hadoop to make it available as a data lake for further analysis.
- Loads the data through OLAP tools into small databases called data marts for fast data access by specific groups within the enterprise. The data is usually stored in data cubes. A data cube presents data based on two or more attributes. For example, it might summarize sales by region (two attributes) or sales by region and by product (three attributes). Products such as Cognos PowerPlay create optimized cubes of data for distribution. Data marts are often distributed to remote sites for faster access. Reports are usually generated from the data marts rather than the data warehouse.
- Turns the data into valuable information using various reporting and BI tools.

Benefits to the Data Warehouse Team

Building a data warehouse generally requires a team of experts. The different roles and the benefits the Automatic Workload Automation product brings to each role are described below.

NOTE: One person on the team may function in more than one role.

Project Manager

This person will oversee the progress and be responsible for the success of the data-warehousing project. Automatic Workload Automation benefits the project manager during the implementation and maintenance phase of the project. A script-less approach reduces implementation times and provides an agile framework for quicker promotion of changes to the production environment.

Database Analyst (DBA)

This person is responsible for keeping the database running smoothly. Additional tasks are to plan and execute a backup/recovery plan, as well as performance tuning. Automatic Workload Automation benefits DBAs during the maintenance phase of the project. The DBA can automate routine tasks, get notifications of failed processes, automate recovery, and proactively check the health of database structures prior to loading.

Technical Architect

This person is responsible for developing and implementing the overall technical architecture of the data warehouse, from the backend hardware/software to the client desktop configuration. Automatic Workload Automation simplifies the technical architecture by providing a single point of control for the whole data warehouse process.

ETL and OLAP Developers

These people are responsible for planning, developing and deploying the extraction, transformation, and loading routine for the data warehouse. ETL tools provide their own scheduling system. The ETL developers may be resistant to moving control of ETL processes to external automation technology. However, the major benefit that Automatic Workload Automation provides is to offload production support to the operations team: monitoring, notifications of failures, and automation of recovery procedures.

BI Developer and Data Scientist

These people are responsible for developing reports, dashboards, and algorithms from which data is transformed into valuable information for decision-making. BI and big data tools require extensive scripting, which creates complexity. Automatic Workload Automation provides a reusable object-oriented framework combined with out-of-the box actions and templates, and drag-and-drop workflow designer. Data scientists and non-technical users can quickly build complex data flows, reducing development effort and time.

For more information, please visit ca.com/automation.

Broadcom, the pulse logo, Connecting everything, CA Technologies, the CA technologies logo, and Automic are among the trademarks of Broadcom and/or its affiliates in the United States, certain other countries, and/or the EU.

Copyright © 2019 Broadcom. All Rights Reserved.

The term “Broadcom” refers to Broadcom Inc. and/or its subsidiaries. For more information, please visit www.broadcom.com.

Broadcom reserves the right to make changes without further notice to any products or data herein to improve reliability, function, or design. Information furnished by Broadcom is believed to be accurate and reliable. However, Broadcom does not assume any liability arising out of the application or use of this information, nor the application or use of any product or circuit described herein, neither does it convey any license under its patent rights nor the rights of others.