workload automation for the cloud

Prabhakar Gopalan
Sr Principal Product Marketing Manager
CA Service Automation

we can
## Table of Contents

### Executive Summary

**SECTION 1** 04

- **Workload automation today — challenges** 04
- First things first 04
- Evolution from job scheduling 04

**Workload automation for the cloud: evaluation criteria** 04

- Workload attributes 05
- Cloud adoption maturity 07
- Business adoption alignment 08

**SECTION 2** 09

- **Characteristics of a dynamic workload automation solution** 09
- Cross platform visibility and control 09
- On agent-less scheduling and job definitions 09

**Governance and security** 10

**Cloud services — service oriented architecture (ITWAB) and cloud bursting** 11

- Service oriented architecture for workload automation broker 11

**SECTION 3** 13

**How can CA Technologies help you?** 13

**CA Workload Automation** 14

For the dynamic datacenter 14

**SECTION 4** 15

**About the author** 15

**SECTION 5** 15

**About CA Technologies** 15
executive summary

As cloud computing has become a strategic initiative for large enterprises, the new method of delivering and consuming IT services has forced its users to rethink domains like job scheduling that have long been considered mature and stable.

This is presenting both opportunities and challenges to datacenter administrators and CIOs. For datacenter administrators, opportunities exist to consolidate operations in order to optimize management and be more responsive to ever increasing business needs. Cloud computing offers CIOs and line of business owners the opportunity to showcase IT as a strategic differentiator, not only to help reduce capital and operational expenditures, but also to gain competitive advantage by becoming more agile and dynamic.

To fully benefit from cloud computing, these decision makers need to understand the role of workload automation and the challenges they face in taking their existing and future workloads to operate in a cloud. They also need to evaluate vendor strategies that will enable them to successfully gain all the benefits cloud computing has to offer. An abrupt change from traditional operational models to cloud computing is risky and unsound.

This paper discusses workload automation and job scheduling as it pertains to cloud computing and helps enterprise IT leaders understand the workload automation landscape by raising questions they need to carefully consider for long-term success of their datacenter workload automation strategy.
Workload automation today — challenges

First things first
What exactly is a workload? Definitions abound on what a workload is. In this paper we define and use workload as an abstraction of a process or set of processes that can be componentized, individually operated upon, and produce a determinate result, with the abstraction being above the network, hardware, and operating system layers. An Oracle Stored Procedure that is run nightly to update warehouse inventory, for instance, can be called a workload. This stored procedure is most likely a part of a series of components in a programmatic workflow that implements a business process, like a customer order shipping J2EE application that processes orders placed on an e-commerce website in real time. “Workload” refers to both batch jobs and programs executed in real time (application workloads) and adapted to the modern data center.

Evolution from job scheduling
A job scheduler is a tool that allows management and scheduling of jobs or workloads using a calendaring system. The need for job schedulers arose when built-in features in operating systems (e.g. cron in Distributed Unix environments & Job Control Language in the mainframe) needed to address jobs spanning multiple instances of the OS.

Workload automation is the evolution of job scheduling with advanced workload management capabilities for the dynamic data center. The aspects of scheduling workloads here include automatically resolving complex dependencies on various platforms and application tiers and triggering workloads based on IT and business events.

Workload automation for the cloud: evaluation criteria
Cloud computing is the next stage in the datacenter service delivery model. It provides a pay-as-you-go model that is designed to reduce significant capital expenses and operating expenses for consumers of IT services and offers the flexibility to expand or shrink capacity dynamically. IT managers and workload automation software administrators can reduce costs and improve business processes in running datacenters, by embracing the cloud and moving workloads to the cloud — be it a private cloud, a public cloud or a hybrid model of those.

When moving workloads to a cloud computing model, it is important to pay attention to three key considerations:

1. The attributes of workloads that can be placed in the cloud (both private and public clouds);
2. A clear understanding of where your organization is in its cloud adoption maturity curve, and the technology underpinnings necessary to orchestrate, automate and manage workloads in the cloud; and
3. A cost-benefit analysis for moving workloads to the cloud and its alignment with business objectives. Without these considerations, any abrupt move to cloud adoption for workload automation is full of risks. Let us discuss these sections at a little more depth.
Workload attributes
The first question that immediately comes up when talking about moving workloads to the cloud is whether that workload is a good candidate for the cloud. How do you evaluate this? What workload attributes need to be evaluated?

a) Types of workloads
We discuss two major categories of workloads: batch workloads and online transaction processing workloads, which cover the majority of business process automation use cases.

Batch workloads process large amounts of data, anywhere from hundreds to millions of records. For example, the monthly cell phone billing statements of a service provider with several million customers depend on batch workloads that process large amounts of customer cell phone usage records to generate the billing statements. These are processed in scheduled batch windows programmatically executing a sequence of business processes including capturing and processing exceptions along the way. This is all automated without any manual intervention.

Online transaction processing workloads (OLTP) process data in real time with end-user interaction and computing response provided almost instantly, for instance ATM transactions. Web workloads are a subset of OLTP workloads with an architecture that includes an application server, a web server, applications, a database and supporting infrastructure components such as load balancers. However, the essential characteristics of web workloads remain the same as OLTP workloads — web workloads also require immediate end-user response and transactional.

Batch workloads with predictable demand that run infrequently can free up expensive and dedicated hardware by using the cloud. OLTP workloads that need elastic scaling to meet infrequent spikes in demand can use the cloud to provide a theoretically unlimited but fully managed resource pool.

Figure 1

Benefits of workloads in the cloud

- Batch workloads in the Cloud
  - Free up dedicated hardware and computing power (reduce CapEx)
  - Reduce dedicated maintenance and support expenses (reduce OpEx)
  - Improve governance by increasing resource-based availability (AvP)

- OLTP workloads in the Cloud
  - Elastic resource scaling for spikes and spikes in demand (improve Resilience)
  - Free up front and back end for occasional spikes (increased capacity)
  - Help meet SLA driven workloads by providing on-demand compute capability
b) Size of data transfer for operating the workload
A workload hosted in a cloud that requires 20 GB of data such as database requests/transfers or movement of an enterprise database application, or a reporting engine on a dedicated 100mbps link with no latency, takes approximately one hour to complete a round transfer. Analysis of total costs to business for data transfer versus total benefits of moving workload to the cloud, with due SLA considerations, would give a clear picture of what type of workloads are best suited for operating in the cloud. This would vary from organization to organization, depending on their network infrastructure and workload architecture.

c) Compute capacity required to process the workload
If the compute capacity requirements of the workload are CPU-intensive and provided by specialized hardware and software combinations not available in cloud operating models, then dedicated computing environments may still be the best option for these workloads. Typical examples include database workloads that are processed in a mainframe operating environment. These workloads may be tied down to even the specific processor types within the mainframe (e.g. a DB2 workload in a mainframe that is executed in total control block mode).

d) Workload architecture
Application architecture is another important consideration. If the workload has many application dependencies and moving the workload to the cloud requires moving those applications to the cloud, then a move to the cloud would be on a much larger scale than moving only the workload. Often this evaluation would cross organizational and business processes and have significant impact on Service Level Agreements (SLA) and Key Performance Indicators (KPI) of IT services. Organizations that have already implemented a Service Oriented Architecture (SOA) for their workloads are best suited for the cloud delivery model, as their application dependencies are already abstracted to Application Programming Interfaces (API) and service connectors between applications and workloads.

e) Networking needs
Networking needs become particularly important when workload size and transfer speeds to and from the cloud have an impact on workload schedule windows. Some of these concerns can be mitigated if the workloads and dependencies are also hosted on the cloud and workloads are turned on and off in the cloud without needing to transfer huge chunks of data in and out of the cloud. Bandwidth impacts the time to process workloads (and hence the SLAs) and the cost of data transfers — fixed one-time infrastructure costs and variable carrying costs for metered pipes. Public cloud providers charge not only for computing and storage, but also for data transfer size. Apart from bandwidth requirements, availability, failover and security are other major networking needs. Measuring, monitoring, and managing these networking needs also add to additional hardware (for private clouds), software, and management costs.

f) Frequency of workload processing
Workloads such as batch jobs — which require processing only once a month or other predetermined frequencies, and in present-day scenarios, consume dedicated infrastructure — are excellent candidates for moving to the cloud (if all other workload attributes are also deemed good for the move). This is where the cost benefits in operating in the cloud come into play. For instance, a 4GB workload that is processed once every 24 hours, needing one hour of computing power, with a data
transfer in/out of 4GB, costs less than $50 a month on Amazon’s EC2 public cloud. The costs of running similar workloads operated, managed, and maintained on dedicated hardware can run as high as several hundreds, if not thousands, of dollars in traditional non-cloud hosting environments.

9) Security
Security is a priority, not only for public or hybrid clouds, but for private clouds as well. At the technology level, security needs include securing ports, firewalls, enabling appropriate access, authorization and authentication to the workload infrastructure. The governance needs include audit, control, and compliance. (Also discussed in detail in the governance and security section of this paper.)

Cloud adoption maturity

Cloud adoption maturity within your IT organization is an important consideration for workload automation in the cloud. Virtualization, service automation, and service management are the technology underpinnings for enabling an IT organization to move along the cloud adoption maturity curve as observed in the above figure, CA Technologies cloud adoption maturity path. Most IT organizations are in the server consolidation phase.

Workloads that use virtual machine resources are logical candidates for the cloud. Most datacenters are yet to realize full server consolidation. As datacenters continue their server virtualization projects, opportunities for workloads to exploit the virtual machine resources will become more commonplace. At that point, subsequent technology enablers in the cloud, such as automated server provisioning and configuration management, will become even more critical to IT organizations.

The dynamic datacenter vision can be delivered in an approach that starts with the necessary technology to consolidate servers and manage virtualized resources first, to automate their provisioning in the cloud through service automation next, and to manage the SLAs of the workloads through service management technologies in conjunction with those. Workloads would have to have
Workload automation cannot be deployed in the cloud without this foundation.

**Business process alignment**
Business process alignment adds a third dimension to the challenges in transforming workload automation to the cloud. While workload attributes and cloud adoption maturity path focused on the technology aspects of workload, business process alignment is essential to meeting business objectives and executing business strategy. How workloads behave in the cloud, how workload processing will respond to business events, how exceptions in workload processing will be captured, reported and acted upon in cloud operating environments, all need to rationalized just as they are in traditional computing environments.
Most IT organizations have established automated process automation tools (or even manual processes) that are tightly integrated with their workload automation solutions (see Figure 3: a typical process automation view integrated with workload automation). Moving workloads to the cloud should help ensure process dependencies are also carried forward in the cloud scenario. This imposes new cloud operating requirements, not just on the workload automation tools, but also on the process automation tools. The process automation tools need to be cloud-aware and the corresponding adapters and interfaces in the cloud to be operational and aligned with business processes.

Characteristics of a dynamic workload automation solution

In this section, we discuss the three areas that can help decision makers in IT organizations evaluate vendor products in choosing a workload automation solution that is ready for both traditional IT operating environments and the journey into the cloud.

Cross platform visibility and control

A primary function of a good workload automation solution is to provide visibility into enterprise-wide workloads, regardless of where the workload or the workload automation solution is physically located. However, workloads themselves are not operated along platform lines of separation. They have cross-platform dependencies for computing needs and for application dependencies. For instance, the workload automation solution could be on a mainframe but the workloads could be running on distributed platforms, or vice-versa. Most vendors have separate solutions for each platform, making it difficult for IT operations to understand workload dependencies across platforms or virtual servers.

For a dynamic workload automation solution, it becomes even more complex when workloads are run in the cloud, another virtual resource. This makes it important for the workload automation solution to be able to offer full flexibility in its ability to operate agents across platforms, virtual resources and the cloud, and visibility into all of these workloads from a single place. To cite an example, CA Workload Automation solution’s CA Workload Command Center displays visibility into workloads in mainframe, distributed and Amazon EC2 cloud — all in a single pane. This gives workload administrators visibility into enterprise-wide workload infrastructure².

The second aspect of cross-platform workload management, beyond visibility as discussed above, is control. Workload administrators need the ability to apply job definitions that abstract out the platform differences sufficiently in order to avoid recreating multiple job definitions for each platform. This saves time, not only for adding new job definitions, but also on maintenance and service, and helps IT operations be more responsive to business needs.

On agent-less scheduling and job definitions

Technologies such as agent-less scheduling have both advantages in overcoming deployment and time to market challenges, but also some inadequacies in addressing enterprise workload processing needs for scalability on server side, choice in use of workload specific adapters (not all workload types can be automated effectively by remote interfaces), and security requirements for the cloud. For instance, SSH
could be a non-starter in many enterprise deployments. The value of an agent-based or agent-less monitoring solution is no longer a purely technical decision, but must factor in other requirements. Agent-based solutions are far more resilient, reliable and preferred for running mission critical applications. These requirements can be easily extended to workload automation solutions too, which process mission critical applications. In the end, a solution that has the flexibility to offer both agent-based and agent-less technologies addresses a larger set of requirements.

**Governance and security**

Just as in traditional IT delivery models, cloud computing imposes a separate set of governance and security requirements for reliable and secure delivery of IT. The Computer Emergency Response Team (CERT) at Carnegie Mellon University observes, “Information security governance means viewing adequate security as a non-negotiable requirement of being in business.” While traditional IT governance has established frameworks for managing and controlling resources, policies and risks in accordance with objectives of the business, security is now becoming an important constituent of that framework. With the adoption of cloud delivery models, security is at the forefront of IT governance for three reasons — data, control, and access. Enterprise data is no longer residing in traditional IT infrastructure; control of flow of data developer, host and end user has become blurry; and securing access to the data from both internal and external threats is extremely challenging for IT. It is in this context that we need to look at how a workload automation solution can help IT enable and support deployments in all three models of the cloud — private, public, and hybrid.

As you evaluate your current workload automation solution for supporting workloads in the cloud, key questions to ask should be:

- How is the security design in the solution architected?
- Does the solution supports object-level AAA (authentication, authorization, and accounting)?
- Does the solution supports and implements protocols (such as LDAP) for integration with directory servers for user AAA?
- How does it secure the connection of the agents to these servers?
- Does it have adequate security at each client and server level technology (in workload scheduler, engine, and agent)?
- Does it support VPN integration within or outside the solution?
- Does use of an agent-less solution poses risks to opening ports requiring governance exceptions?
- Are those risks are adequately addressed?
- Is the workload going to be deployed in an infrastructure as a service (IaaS) or platform as a service (PaaS) model?

The data and application security requirements are different for each of these.

These questions are not limited to the workload automation solution, but relate back to application design itself — applications already in use and design for new applications. Ensuring applications that will run in a traditional infrastructure, with the option to run in a private cloud in the near term, a
public cloud in the long term, and meet security designs for all of these models, is not simply an architectural design choice, but also a potential cost saving business benefit. The costs of retooling workload implementations can become big hurdles to surmount if the design practice does not adequately cover these scenarios. Particularly in this regard a workload automation solution that can be delivered as a service and implemented with the foundation of SOA design principles becomes an automatic choice.

Cloud services — service oriented architecture (ITWAB) and cloud bursting
For a dynamic workload automation solution, apart from the capabilities discussed earlier which are all extensions to existing capabilities, two entirely new sets of capabilities are required — IT workload automation broker, using SOA, and a cloud bursting service. Both these capabilities not only take advantage of the cloud as a resource pool, but also shift IT operations into providing workload automation itself as a service so that business units within an enterprise can order workload automation from the service catalogue. Add metering, billing, and reporting facilities (which are built-in features of public clouds), and you have true end-to-end click-and-pay service automation capabilities.

Service oriented architecture for workload automation broker
For workload automation solutions to scale to the cloud, they need to evolve beyond simply serving as cross-platform job scheduling engines and begin functioning as service brokers. It is not sufficient to optimize workloads to run within individual physical or virtual resource pools. The workload automation solution needs to be embedded in the business process so that when service requests are made, based on policies and service level agreements, resources can be automatically provisioned and de-provisioned by combining all underlying IT infrastructure and optimizing that for cost and performance.

This end-to-end workload automation is possible only if the workload automation solution is built-in on open standards to provide interoperability and integration with legacy jobs and jobs for the cloud. A workload automation solution that supports web services, for example, can provide the right set of integration points for different workloads to talk to the same broker (workload automation engine) to mediate and schedule jobs based on events, calendaring, business process, resource optimization needs, resource pool availability, etc — thus creating a true workload broker/service/service requestor pattern for workload automation.

Cloud bursting of workload automation
The second new capability that is required for a workload automation solution to take advantage of the cloud computing model is cloud bursting. Cloud bursting is the ability to provision and de-provision resources in the cloud automatically, based on needs from a traditional computing environment. If the traditional computing environment is unable to handle a certain load, cloud bursting quickly helps in obtaining new computing resources and — once the work is complete — contract back to initial state, providing a way to handle sudden increasing computing needs.

To illustrate cloud bursting through an example, consider a scenario where a financial institution is running month-end client reports for fund performance. These reports are typically completed through batch jobs run on dedicated hardware, possibly in a combination of mainframe (database) and
distributed (reporting engine) resources, and a multi-tier custom financial application (say, SAP). If the financial institution switches to leveraging the cloud to offload some of these batch jobs to save on dedicated hardware costs, the workload automation solution needs to be able to burst those batch jobs into the cloud. For such an operation, the workload automation solution needs to be resource-aware for physical and virtual pools of hardware and software, have an understanding of workload characteristics to determine which ones to place in the cloud, apply policy on those workloads, connect to the cloud, ‘wake-up’ or provision (for new) workloads in the cloud, execute, report and decommission/’retire’ workloads after job completion. This seemingly straight forward process requires a number of supporting technology requirements beginning with a service oriented architecture of the workload automation broker. The figure below illustrates CA Service Automation’s cloud bursting of 10 jobs that get instantiated in Amazon’s EC2 public cloud and automatically de-provisioned on completing the work.
How can CA Technologies help you?

CA Technologies takes a pragmatic approach in helping IT organizations move along the path to the dynamic datacenter. Instead of offering ad hoc point products focused on narrow technology initiatives, we offer industry leading solutions to help IT in their journey from virtualization to cloud. We recognize enterprises have significant investments and business risks associated with their workloads and one size doesn’t fit all.

CA Technologies offers a five-step process to embrace the cloud:

1. Pairing virtualization with management to drive immediate benefits;
2. Provisioning and automating workloads with heterogeneous cloud adapted tools;
3. Integrating lifecycle management tools across traditional IT silos like asset management, configuration management, capacity and performance management;
4. Integrating management across server, network, storage, physical and virtual domains, and;
5. Combining virtualization management, service automation and service management offerings into a single integrated solution to deliver the dynamic datacenter.

In this holistic view, workload automation is delivered as a composite service for the cloud, as are other automation services (e.g. server provisioning, network provisioning).
CA Workload Automation — a scalable, secure, cloud-ready solution

For the dynamic datacenter

CA Workload Automation provides complete end-to-end views and management of enterprise workloads, whether they are running on distributed systems, on the mainframe or in the cloud. CA Workload Automation is designed to enable IT organizations to reduce costs in scheduling and managing workloads and help derive significant business benefits from offloading workload processing to cloud computing environments on-demand or through cloud bursts. CA Workload Automation’s cloud-aware workload automation engine is designed to intelligently respond to business needs and events by managing and processing workloads in physical and virtual resource pools, regardless of underlying platforms, optimizing infrastructure, saving costs and improving agility in operations. Native integration to IT process management (run book automation) and IT service management provides a simplified, centralized interface to all workload management needs, enforces policy directives to workload processing, and aligns IT operations with business process to help meet service level objectives.

CA Workload Automation is an industry leading solution for solving complex multiplatform scheduling challenges. With more than 50 workload specific agent types, CA Workload Automation provides end-to-end automation solutions for workloads on all major operating systems, common enterprise business applications such as Oracle E-Business Suite, SAP and PeopleSoft; agents for custom applications; and object types such as POJOs, EJBs, JMS and web services. CA Workload Automation’s event-driven scheduling, business event sensors, critical path analysis and forecasting, and resource optimization features provide advanced workload automation capabilities necessary for building the dynamic datacenter.

For more details on CA Workload Automation, please visit: http://www.ca.com/us/workload-automation.aspx
About the author

Prabhakar Gopalan leads product marketing strategy for workload automation solutions at CA Technologies. Before joining CA Technologies, Prabhakar managed product and services strategy for systems and application management technologies in product management, product marketing, IT Architect and Consultant roles at Red Hat Inc., Novell Inc., and IBM Corporation. Prabhakar has a graduate degree in MIS from Texas A&M University, College Station, and an MBA from The University of Texas at Austin. He can be reached via email at prabhakar.goplan@ca.com and you can follow him @pgopalan on Twitter.

Acknowledgements

Many thanks go to the CA Technologies team that came together for content, review and feedback in writing this paper. Without the kind assistance of Jim Anderson, Ryan Shopp, Charles Ealy, Andi Mann, and Joseph Offenberg, this effort would not have been possible.