

## A Survey of Cloud Workflow

Huang Hua <sup>1,a</sup>, Zhang Yi-Lai <sup>2,b</sup>, Zhang Min <sup>3,c</sup>

<sup>1,2,3</sup> Jingdezhen Ceramic Institute, Jingdezhen, Jiangxi, 333001, China

<sup>a</sup>jdz\_hh@qq.com, <sup>b</sup>tyzyl@qq.com, <sup>c</sup>191148074@qq.com

**Keywords:** Cloud Computing; Workflow; Cloud Workflow; Business Process

**Abstract:** Traditionally, enterprises have approached IT modernization and back office process enhancement in silos, by introducing multiple vendors with different technologies, platforms and operating systems, which limits process integrity and leads to disseminated responsibilities. In contrast, the Cloud revolution has made it much easier for enterprises to approach business transformation in a holistic manner, leading to a quantum leap in their business and IT practices and performance. In other words, Cloud-based workflow (which is shorted for cloud workflow) technology is now regarded as a pretty good solution: Cloud workflow helps organizations in process harmonization, optimal organization design and change management offering benefits beyond cost reduction, it enables a complex application instance to be abstractly defined, flexibly configured and auto-operated. This paper introduces shortly the technology of both cloud computing and workflow and then presents the concepts and features of cloud workflow. Meanwhile, the following part discusses the application scenario and the application cases of cloud workflow. Lastly a possible trend of workflow in the future is proposed.

### Introduction

With the development of the Internet in the past few years, WEB2.0, middleware, virtualization, grid computing and other technologies were merged into a great power, named the power of cloud computing. Almost all well-known IT companies publicize and market their cloud plans, which makes the application field of cloud computing have expanded largely within a very short time from enterprise-level cloud services (such as IBM's Blue Cloud) to small client anti-virus software (e.g., Rising's "cloud Antivirus"), this means the arrival of the cloud era.

However, along with the development of cloud computing, a series of technical problems also arose. One of the most prominent problems is how to reduce running costs and maximize revenues with the premise of maintaining or even improving the quality of service (QoS). In this regard, we believe that a viable solution is cloud-based workflow (referred to as the "cloud workflow"). Seen from the user's point of cloud computing, cloud workflow provides abstract definition of complex applications, flexible configuration and automated operation, and improves the quality of service; From the perspective of the providers of the cloud computing services, cloud workflow offers automatic scheduling of tasks, optimization and management of resources, and then saves the operating costs of cloud computing. As a optimization solution of cloud computing, cloud workflow is gaining a lot of attention and study.

In this paper, we make a survey of cloud workflow in detail. First we introduces two major foundations of cloud workflow-----cloud computing and workflow technology, on this basis workflow applications in the cloud computing are elaborated, and then describe the concept and technical characteristics of cloud workflow, and give several application cases of cloud workflow. At the end of this paper, we forecast the technology development prospects of cloud workflow.

### Cloud computing and workflow technology

#### Overview of Cloud Computing

Cloud Computing is the development of distributed processing, parallel processing and Grid Computing, or commercial implementations of these concepts of computer science.

The basic principles of cloud computing is that computing is distributed in a large number of distributed computers, rather than the local computer or a remote server and the running of enterprise data centers is more like the Internet, which allows companies to be able to switch resources to the applications needed and access to a computer and storage systems on demand. This is a revolutionary initiative, which is similar to that the old single generator mode is shifted into a plant centralized power supply mode. It means that computing power can also be as a kind of commodity for circulation and using, like gas, water and electricity, being accessed easily and low-cost. The biggest difference is that it is transmitted through the Internet. The cloud computing blueprint has been pretty clear: in the future, having only a laptop or a mobile phone, we can achieve everything we need by network services, even including supercomputing. From this point of view, the end users are the real owners of the cloud computing. In summary, cloud computing has the following characteristics:

1) **Service-Oriented Architecture (SOA)**. What the cloud computing provides are services, the specific mechanism of which is transparent to users.

2) **Component Based, Extensible and Scalable**. The system instances of cloud computing are the integration of functional components, which support extending functions to be seamlessly compatible with the new functional components. The running scale of computing resources can be scalable according to computing needs, which means that the size and performance of calculating can be scaled to scale-down and scale-up.

3) **Multi-tenancy**. Multi-tenant characteristic of the cloud computing is not only limited to the IaaS layer, including the PaaS layer (such as an application server, Java virtual machine, etc.), and the SaaS application layer (such as database, business logic, workflow and user interface) also needs multi-tenant characteristic. Only in this way, each tenant or application can fully enjoy all of the universal service.

4) **Service On-demand**. The cloud is a huge resource pool modeled by pay-as-you-go style, which is accounted like using water, electricity and gas.

5) **High Availability**. Cloud computing system has strong fault tolerance and debugging performance, automatically detecting the failed node, some failure computing nodes do not affect the service performance of the entire cloud computing system.

6) **Low-cost**. The formation of the commercial computer groups providing computing power in the cloud computing system cost less than the deployment costs of a supercomputer with equivalent performance.

## Overview of Workflow Technology

### Theoretical basis of workflow

According to the definition of WFMC, workflow is a calculation model of the business process and is an abstract, summary and description of the business rules and its steps. A workflow is an ordered sequence of activities events, designed to achieve a defined business objective. We can regard the entire business process as a river, what flows through which is the workflow.

Workflow Management System (WFMS) is a system for fully defining, implementing and managing the workflow, the software is derived to run in an orderly manner by workflow logic represented by computer. Its main functions are to define, implement and manage the workflow and to support exchanging information between the group members and tasks. In fact, the workflow management system is a software system, but it also differs from the enterprise's business management system, providing a support environment for running business systems. In this system, workflow engine is the core and is an integral part of a software system, which is tailored for adapting to different transmission of information in different business environments according to the priorities, tasks and conditions of the roles. In a way, workflow engine holds the right to the allocation of system resources. A typical model of the workflow engine is shown in Figure 1.

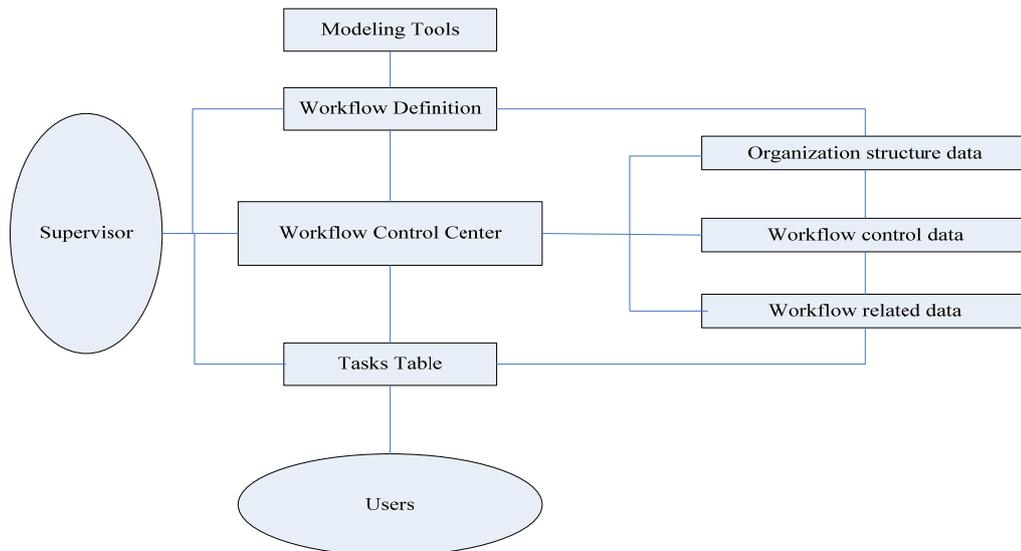


Figure.1 Workflow Engine Model

**Development of Workflow Technology**

Many famous workflow commercial software vendors have publicized new products, among of which there are some well-known enterprise-class workflow products, e.g., BEA's WLI (Web Logic Integration Studio) and Aqua Logic BPM Suite, IBM 's Web Sphere MQ Workflow and Web Sphere Process Server, especially the workflow system based web service, achieving cross-organizational and cross-platform business interworking via the Internet, has become a new trend of workflow products, such as "Ultimus" business Process Management products, etc. . But there are still many business difficulties, such as the integration between the application of internal processes and applications and the integration between the internal business processes and enterprise application.

In addition, a theory of service-oriented workflow is emerging, which is the integration between traditional workflow technologies and SOA, with role attribute and human task. The BPEL4People draft was released in August 2007, but there is still a distance from WFMC. Meanwhile, the other research areas of Workflow is also emerging, such as flexible work flow, distributed workflow, context workflow, scientific workflow, cloud workflow and so on, respectively corresponding to supporting the flexible, configurable, real-time and multi-tenant feature of workflow.

**Applications of Workflow in Cloud Computing**

Workflow technology provides a solution for cloud computing, which can be applied in two different hierarchies: above-the-cloud and in-the-cloud.

**Above-the-cloud applications of workflow**

To define, model, re-design, automatic execute application projects in cloud computing by using the workflow's idea is a feasible solution. Firstly, workflow can be well implanted into the cloud computing project. The flow of cloud computing is a workflow with different and serialized transactions, which includes the traditional workflow affairs and cloud-based transaction. Pulling workflow into cloud computing makes the processes of cloud computing to be abstracted and modeled and distinguish cloud transactions with traditional workflow transactions. Secondly, a workflow management system (WMS) can reduce the cost for operating transaction processes, allow users to model a specific application instance and support the dynamic allocation of computing resources and the automatic execution of transaction flows. In this case, the related processes of cloud computing become a part of the whole workflow model, the cloud computing will be performed at a higher level than the "cloud" for users. Workflow management system can be integrated with cloud computing platforms or agent (Broker), monitoring and managing the running process of cloud computing.

Cloud workflow management systems above the cloud have many advantages, such as modeling and integrating of cloud computing processes, selecting the carrier cloud and flexible scheduling of service processes. In addition, the process' scheduling and monitoring mechanism of workflow

management system can also be used for real-time monitoring and performance optimization of the running process of cloud computing.

### **In-the-cloud applications of workflow**

Cloud computing systems follow a service-oriented framework, with features of high availability and real-time. In actual operation, however, due to network congestion, node failure and node communication loss and other factors, the system's performance is often affected. Workflow's in-the-cloud applications offer a solution: The process of cloud computing system's responding to a service request is the same as the sequence of many transactions. These transactions are implemented in the cloud, which are transparent to users. The cloud workflow management system parse the process of requesting and responding in the cloud into a number of definable cloud transactions, which are modeled and redesigned by users using workflow modeling tools, so the operation process of these transactions in the cloud are controlled. In fact the above process is that the above-the-cloud cloud computing transactions are resolved and modeled as a sub-workflow model running in the cloud. The cloud workflow management system is being deployed in the cloud, which not only can achieve more effective control to in-the-cloud processes, but also bring the underlying commercial computers of cloud computing into the workflow management system, help to strengthen the self-configuration, scalability, fault tolerance of cloud computing system, optimize node communication, and reduce operating costs.

Regardless of above the cloud or in the cloud, cloud workflow can reduce operating costs of cloud computing system and improve the quality of services. As to users, cloud workflow provides abstract definition, flexible configuration and automated operation of complex applications; to the providers of cloud service, cloud workflow can provide automatic scheduling of tasks and optimization and management of resources. Research and application of cloud workflow is promoted and developed by users and providers.

### **Concept and Technical Feature of Cloud Workflow**

#### **Concept of Cloud Workflow**

Cloud workflow is a workflow system based on cloud computing, a new mode of application of workflow management system in the cloud environment, and is a solution for optimizing the system performance and running costs of cloud computing, which integrates workflow with cloud computing system, reduces costs of cloud computing and improves the quality of cloud services. The cloud workflow generally has the following characteristics:

**1) Transparency.** In run-time, provide a mechanism of non-user-visible task scheduling, load-balancing, self-configuration.

**2) Scalability.** Cloud workflow management system can achieve the self-configuration of underlying computing resources by increasing or decreasing the running nodes according to the operating conditions and control the cost of computing with the premise of guaranteed performance. It means that the size of the services' underlying computing can be real-time scalable based on computing demands.

**3) Multi-tenant Architecture.** Cloud workflow has the multi-tenant characteristics of cloud computing, a number of different tenants can simultaneously design, deploy and run their own workflows, and achieve three kinds of isolations: data isolation, performance isolation and execution isolation.

**4) Real-time Monitoring.** Provide a mechanism of load balancing, fault controlling and node scale controlling by detecting the running status of transaction processes in the cloud environment and being combined with the network status and node performance.

#### **Technical features of cloud workflow**

Before cloud workflow, business workflow, grid workflow has been matured and applied in practice. The inheritance and continuity of workflow development decides that cloud workflow cannot be pulled out of the development path of the workflow. Cloud workflow has many similar properties and characteristics with other workflows, such as business workflow and grid workflow; it is insufficient to get a full understanding of the technical characteristics of cloud workflow simply by the above isolated concept. So the technology characteristics of cloud workflow will be elaborated

from three perspectives of workflow design, workflow execution and scheduling as well as the host computing environment.

### **Workflow Design**

1) **Model Topology:** A workflow model is assembled by a number of independent and interrelated transactions, control flow of which reflects the running process of workflow. If a transaction is abstracted as a point and a control flow as a directed side, a topology of the directed graph will be formed. The topology of cloud workflow model can be divided into DAG (Directed Acyclic Graph) and Non-DAG (Non-Directed Acyclic Graph), that there is a directed acyclic graph with non-directed acyclic graph.

2) **Model type:** Cloud workflow model is abstract model, which means that no transactions are mapped to specific computing resources in the definition phase.

3) **Focus on describing model:** According to the different focuses of control flow and data flow in the model definition, it can be divided into: focusing on controlling, focusing on data and both. Cloud workflow model focuses on both sides.

4) **Model generation:** Cloud workflow has two generating modes: user-defining and automatically generating. The mode of user-defining generation requires users to manually create a workflow model, which is achieved by directly using workflow language (such as XML-based Workflow Language), or by means of a visual workflow modeling tool; The model of automatically generation can automatically generate the workflow model matched by system based on semantic analysis of users' needs.

### **Execution and scheduling of workflow**

1) **Architecture of scheduling system:** The architecture of the cloud workflow scheduling system can be divided into centralized type and distributed type. Centralized scheduling system is controlled by a central scheduler node for mapping and running workflow transactions; Distributed scheduling system is consisted of many nodes with the same level, which can communicate with each other.

2) **Response mechanism:** The mode of responding to transactions in cloud workflow engine is asynchronous response.

3) **Instance generation:** The instances generation is a mechanism of transforming an abstract model of workflow into a meaningful workflow instance (in particular a transformation template), which may be determined at run time or be predefined statically. Instance generation of cloud workflow have two dynamic and static ways.

4) **Decision Making:** When a transaction is mapped to a specific computing node in a workflow management system, the data to support making decision are from two sources: the running status of the current transaction, as well as the overall running status of the workflow. The decision making of cloud workflow is divided into "Local" and "Global" types.

5) **Middleware:** Cloud workflow's middleware is a service-oriented one.

### **Host computing environment**

1) **Composition and Size:** Ultra-large-scale commercial computers and a cluster of servers.

2) **Scalability:** Computing resources can be expanded or reduced by self-configuring with response to changes in operating performance, and supporting for interactive computing.

3) **Virtualization Degree:** Cloud workflow has a higher degree of virtualization. Virtualization means that the specific computing resources is abstracted and encapsulated to make the process to be transparent for users, API or Service is only exposed to users.

4) **Programming Model:** In terms of data management, different workflow systems adapt different programming model, the Map-Reduce programming model is used cloud workflow.

### **Application Cases of Cloud Workflow**

#### **1) Megha Workflow Management System (MWMS).**

MWMS is a running platform of scientific applications projects based distributed computing resources, which is developed by Professor Rajkumar Buyya and his team from Melbourne University, Australia. MWMS is practical application on the Amazon Elastic Cloud, it is proved that the scalability of MWMS is ensured and the execution time of this system is reduced by dynamic

scheduling of tasks, when thousands of transactions are executed in a limited number of cloud computing resources. MWMS based xWFL, is an Above-the-Cloud cloud workflow management system. MWMS provides a users' portal, integrating workflow modeling tools, job submission interface, running and resource monitoring interface. Users can create and edit models online, submit tasks and monitor running state. In MWMS, users can perform cloud computing above the cloud: using cloud computing services of workflow model, modeling running process and controlling the real-time operating conditions.

## 2) IBM LotusLive

In April 2009, IBM publicized LotusLive, as a solution for online business collaboration and social networking services. LotusLive is an IBM-hosted online services portfolio, which can provide scalable and secure solutions of e-mail, Web conferencing and collaboration. LotusLive is realized by SaaS model. LotusLive website, users can rent many collaboration services deployed in the cloud via LotusLive website. LotusLive is a typical cloud-based SaaS model: the underlying cloud is used for deploying different services and providing computing support, users can get the services they need through the portal. Workflow is an important part of composing services that can be hired. Services provided by LotusLive are on-premise solutions, which includes the pre-setting of workflow: a service as a set of pre-setting transaction process. Service requests are sent from users, these services will run in the cloud as a workflow. To classify from the workflow architecture-level, workflows integrated in IBM LotusLive belong to in-the-cloud workflows. It is worth mentioned that the "Click to Cloud" makes workflow models in IBM LotusLive to have excellent scalability, which allows seamless bonding between the online LotusLive services and the existing services in user services host, it means that the workflows of LotusLive support scalable process model.

### Forecast of Cloud Workflow

In summary, cloud workflow is an only way in the development process of cloud computing. That is because that running cost and quality of computing services are the most important factors of the development of cloud computing industry, and cloud workflow undoubtedly has its unique advantages in reducing the operational costs of cloud computing and improving the service quality of cloud computing. Cloud workflow provides users abstract definition, flexible configuration and automated operation of complex applications for users in cloud computing, as well as automatic scheduling of tasks and optimization and management of resources for cloud computing. Today, more and more IT vendors publicize their own "cloud plans", cloud workflow is bound to get more and more powerful impetus of development, thus to promote the development of cloud computing industry.

## Acknowledgment

This work was supported by National 863 Program (NO: 2012BAH25F02)

## References

- [1] IBM. Google and IBM announced university initiative to address internet-scale computing challenge [EB/OL] <http://www-03.ibm.com/press/us/en/pressrelease/22414.wss>, 2007 October.
- [2] Google and IBM join in 'cloud computing' research [EB/OL] [http://www.nytimes.com/2007/10/08/technology/08cl-oud/html?\\_r=1&%20r=1&ei=5088&en=92a8c77c354521ba&ex=1349582400&oref=slogin&partner=rssnyt&emc=rss&pagewanted=print](http://www.nytimes.com/2007/10/08/technology/08cl-oud/html?_r=1&%20r=1&ei=5088&en=92a8c77c354521ba&ex=1349582400&oref=slogin&partner=rssnyt&emc=rss&pagewanted=print), 2007 October.
- [3] Mladen A Vouk. Cloud computing issues research and implementations [J]. Journal of Computing and Information Technology CIT, 2008, 16(4): 235-246.
- [4] Deng Qian-ni, Chen Quan. Cloud computing and key technology [J]. Development & Application of High Performance Computing, 2009, 26(1): 2-6.

- [5] Chen Kang,Zheng Wei-min.Cloud computing: system instances and current research [J] . Journal of Software,2009,20 ( 5 ) :1337-1348.
- [6] Greg Boss,Padma Malladi,Dennis Quan,et al. Cloud computing,IBM White paper [EB /OL] . www. Ibm.com /developerworks /websphere /zones/hipods/,2010.
- [7] Rajkumar Buyya,Suraj Pandey,Christian Vecchiola. Cloudbus toolkit for market-oriented cloud computing [C] . Proceeding of the 1st International Conference on Cloud Computing ( CloudCom 2009,Springer, Germany ) , Beijing, China, December 1-4,2009.
- [8] Suraj Pandey,Dileban Karunamoorthy,Kapil Kumar Gupta,et al. Megha workflow management system for application workflows [C] .IEEE Science & Engineering Graduate Research Expo 2009,Melbourne,Australia,2009.
- [9] IBM. Online collaboration: lotusLive [EB /OL] . <https://www.lotuslive.com/en/>,August,2010.