



A survey and taxonomy on workload scheduling and resource provisioning in hybrid clouds

Bo Wang¹ · Changhai Wang¹ · Ying Song^{2,3} · Jie Cao¹ · Xiao Cui¹ · Ling Zhang¹

Received: 25 February 2019 / Revised: 1 November 2019 / Accepted: 6 January 2020 / Published online: 5 February 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Hybrid cloud is a cost-efficient way to address the problem of insufficient resources for meeting the peak demand of its users for a service provider, which elastically scales up or down the cloud capability based on demand by combining local infrastructures and one or more public clouds. While, the combination introduces new challenges that must necessarily be addressed before adoption. To address these new challenges for improving the resource efficiency in hybrid clouds, much work tried to solve the decision problem of workload scheduling, resource provisioning, or both, where workload scheduling answers how to efficiently map workloads to available resources, and resource provisioning addresses how to optimally provision resources based on demand. In this article, we proposed a comprehensive taxonomy of workload scheduling and resource provisioning in hybrid cloud environments to investigate and classify 146 related research articles. Based on the investigation, we summarized the challenges which have not been addressed by these researches, and discussed future directions and trends in the area.

Keywords Cloud computing · Hybrid cloud · Resource provisioning · Workload scheduling

1 Introduction

Cloud computing has received increasing attention in both research and business for about one decade due as it provides a great deal of benefits, e.g., elastic resource provisioning, pay-for-use, economies of scale, high reliability, dynamic customization, etc. [175]. But although resources appear to be infinite to users, a cloud has limited resources in the real world. Thus a cloud should have enough resources for satisfying the peak demand of its users'

requests to satisfy all Quality of Service (QoS) requirements of users for its reputation [45].

When there are insufficient resources for a local cloud meeting the peak demand of its users, three methods can be exploited. The first one is to reject some unimportant requests, e.g., cheap requests, to make room for important requests whose rejections cost much more [106]. While this method would reduce the cloud provider's reputation [45], and may further result in the loss of potential users. Second, the cloud provider increases infrastructures enough for the peak demand. While, in real production environments, the peak resource demand of users is usually much more than the average one, but transient [14, 99], in a cloud, which leads to lots of idle resources most of the time if using the second method. Besides, most of the small to medium enterprises have insufficient capital for infrastructure investments. The third method, hybrid cloud (a.k.a. cloud bursting), is a cost-efficient way to address the problem, which elastically scales up or down the cloud capability based on demand by combining local infrastructures (clusters, grids, or private clouds) and one or more public clouds. Surveyed by the European Network and Information Security Agency (ENISA), most of the

✉ Bo Wang
wangb@zzuli.edu.cn

Changhai Wang
wangb@zzuli.edu.cn

¹ Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou 450002, China

² Beijing Advanced Innovation Center for Materials Genome Engineering, Beijing Information Science and Technology University, Beijing 100101, China

³ Beijing Key Laboratory of Internet Culture and Digital Dissemination, Beijing Information Science and Technology University, Beijing 100101, China

small to medium enterprises prefer a mixture of cloud computing models (public cloud, private cloud) [26]. Nowadays, both commercial and open-source virtualization tools support basic cloud bursting functionalities, e.g., VMware [182], Open Nebula [150], OpenStack [151]. As a new computing paradigm, hybrid cloud computing plays a crucial role not only in developing cloud computing, but also in integration of cloud computing and Internet of Things (IoTs) [25] as others, e.g., edge computing [165], fog computing [51], mobile cloud computing [62].

For providing services, cloud providers must solve the problems of provisioning the optimal number of resources based on demand, i.e. resource provisioning, and mapping users' workloads to available resources efficiently, i.e. workload scheduling [130]. Workload scheduling is to decide the order (priority) of workloads to be executed on available resources in various scheduling units, e.g., virtual machines (VMs), tasks, user requests, while resource provisioning judges which and what amount of resources should be allocated to the scheduled workloads. For optimally providing services, the cloud provider should schedule workloads based on the characteristics of available resources, e.g., heterogeneity [55, 89], reliability [18], and provision resources considering the features of workloads to be run, such as various QoS [72], interdependences between service components [100], interferences between workloads [195], etc. Thus, workload scheduling and resource provisioning are intimately related with each other, and both of which are essential to a cloud management.

However, for a service provider, the hybrid cloud resource management not only has all of the challenges for provisioning services both on a private cloud, e.g., server consolidation [183], and on public clouds, e.g., elasticity management [10], but also introduces new ones, such as the heterogeneity between clouds in terms of various resources [131, 187], the decision of which services and/or which part of a service to be outsourced to the public cloud [100, 145], the performance overhead caused by the network connection between clouds with much lower bandwidth [73, 131], and so on.

In this paper, we surveyed the articles about workload scheduling and resource provisioning for hybrid clouds in recent 10 years. We first presented a comprehensive taxonomy of workload scheduling and resource provisioning in hybrid cloud environments to investigate these 146 related research works in detail. Then, we discussed the challenges which have not been addressed, and suggested several promising directions for future research, based on the detailed investigation for these related papers. To our best knowledge, no work has extensively and thoroughly reviewed workload scheduling and resource provisioning

in hybrid clouds. We believe our review work is helpful to academia and industry concerning hybrid clouds.

The rest of this paper was organized as follows. Section 2 presented the hybrid cloud architecture, which is helpful to understand the remainder of the paper, the survey works related to the resource provisioning and the workload scheduling in hybrid clouds, and the method of collecting related works reviewed. Section 3 introduced in detail the comprehensive taxonomy of workload scheduling and resource provisioning on hybrid cloud management and investigated related works in depth. Section 4 summarized the challenge and the opportunity for future work. And finally, Sect. 5 concluded the paper.

2 Background

In this section, we first provided a simple overview of hybrid cloud architecture, which is helpful to understand the remainder of the paper. Then, we presented the previous work surveying hybrid cloud resource management, and the search method for the related literature reviewed in this paper.

2.1 Hybrid cloud architecture

In a hybrid cloud environment, as shown in Fig. 1, users request and pay for services with various QoS requirements from the service provider with local or public resources including computing, storage, network resources, and so on. The service provider provisions its local resources on which it schedules user requests to process, and rents resources from public clouds when local resources are not enough to satisfy any QoS requirement.

The local resources can be provisioned as either physical resources, e.g., clusters and grids, or virtualized resources which are managed with virtualization tools, e.g., Xen [16], KVM [103], and Docker [64]. Virtualization introduces many benefits, e.g., better isolation and manageability of resources, while with significant performance overheads [120, 185]. The public computing resources are usually

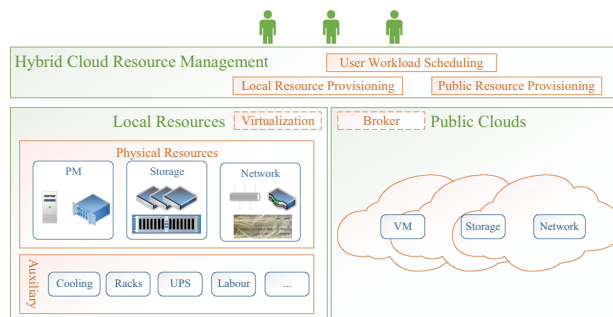


Fig. 1 The hybrid cloud architecture

provisioned in the form of VM, e.g., Amazon EC2 [2], Alibaba Cloud [1], etc. The local and public resources are different in various characteristics, which will be explained in detail in Sect. 3.3, and the resource provisioning policies should be designed with their own peculiarities, respectively. In a hybrid cloud, the public resources may be provisioned by more than one public clouds to avoid vendor lock-in. Then a broker of multi-cloud service composition may be introduced to save the service provider the complex issue of choosing the best fit public cloud [12]. While the use of broker may lead to the service provider losing some optimization opportunities, e.g., the communication cost between public clouds, as the execution of workloads on the public clouds are transparent to the provider.

In this article, we focused on the hybrid cloud resource management including workload scheduling and resource provisioning on hybrid clouds.

2.2 Related survey work

Although there have been plenty of researches surveying resource management on a cloud, only a few works concerned the hybrid cloud.

Bittencourt et al. [22] compared the performance of seven scheduling algorithms on hybrid clouds, where only two work was specially designed for hybrid clouds. And they assessed the impact of communication links on schedules, concluding that the increase of bandwidth reduced the costs and the makespan, and that HCOC [21, 23], a scheduler in hybrid clouds, outperformed MDP [196] designed for utility computing.

Fadel and Fayoumi [68] surveyed 19 works published over 5 years ago, tackling the issue of cloud bursting whose challenges were choosing the best workload to burst and choosing the best resource to provision.

Chopra and Singh [42] investigated only six task scheduling methods for hybrid clouds, considering only three aspects: their optimization criteria, multi-core processor awareness and the number of workflows supported. Their criteria involved only the cost for renting public computational resources and workflows' finish time.

Manikandan and Suguna [134] reviewed 10 papers about resource provisioning for clouds, where there was only one work [217] focusing on hybrid clouds.

Bhosale and Bandari [19] reviewed three works [27, 173, 191] of Aneka cloud platform developed by CLOUDS Laboratory in the University of Melbourne, which provisioned resources for multiple non-interactive workloads involving a large number of files in hybrid clouds.

de Assunção et al. [53] investigated whether a local cluster can benefit from using clouds to improve its

requests' performance by evaluating seven scheduling strategies designed based on conservation [143], aggressive [121], and selective backfilling [172], in terms of various performance metrics, e.g., average weighted response, job slowdown, number of deadline violations, number of jobs rejected and the cost for using clouds.

In this article, we reviewed 146 research works studying on the workload scheduling or/and resource provisioning for various service deliveries, e.g., scientific computing, web services, infrastructures (i.e., VMs), in hybrid clouds. We presented a comprehensive taxonomy to categorize these related works for helping us to summarize the challenges which have not been addressed and propose promising directions for future research. We hope that our work is helpful for both research and business in hybrid cloud management.

2.3 Literature search

The literatures we reviewed include the followings:

- (1) the relevant papers obtained by querying the Engineering Village Compendex database [67] and the Web of Science Core Collection [190] with the searching conditions (in the form of the query statement in Engineering Village Compendex database), (“hybrid cloud” OR “hybrid clouds” OR “cloud bursting”) AND (“task scheduling” OR “job scheduling” OR “application scheduling” OR “request scheduling” OR “service scheduling” OR “task management” OR “job management” OR “request management” OR “application management” OR “service management” OR “service migration” OR “resource scheduling” OR “resource provision” OR “resource provisioning” OR “resource management”), to cover all high quality research papers;
- (2) the relevant references of the papers obtained in (1), (2) and (3) (a recursive procedure);
- (3) and the relevant literatures citing the papers obtained in (1), (2) and (3), which were achieved by Google Scholar [77] (a recursive procedure).

3 Taxonomy

This section presented the detailed taxonomy for workload scheduling and resource provisioning in hybrid cloud environments. We classify related works in six ways according to the properties of the hybrid cloud optimization problem they solved, as shown in Fig. 2. This classification can help us to review related works in detail and summarize them for leading out challenges and opportunities of

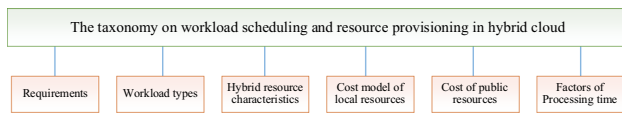


Fig. 2 The taxonomy of workload scheduling and resource provisioning in hybrid cloud environments

optimizing the use of resources in hybrid clouds. The taxonomy is detailed as followings.

- (1) **Requirement.** When exploiting cloud resources, users have various QoS requirements for their workloads, e.g., response time, security, etc. In reviewed works, these requirements are treated as objective(s) of optimizing one or more QoS metrics, or constraints of restricting some metric values, illustrated in Sect. 3.1 in details.
- (2) **Workload types.** There are various types of workload to be executed on clouds. Distinct types of workload have requirements of various characteristics of resource, and thus should be managed by different methods to achieve the best result [37]. The workload type is a useful differentiating factor for related literatures to understand the applicability of a scheduling/provisioning method to a type of workload. Workload types concerned are detailed in Sect. 3.2.
- (3) **Resource characteristics.** Private and public resources have differences in various aspects, e.g., cost–performance ratio, security, reliability, etc., which will be described in Sect. 3.3, leading to different types and different amounts of hybrid resources required by various workloads with diverse QoS requirements. In general, cloud resources are heterogeneous because of continuously infrastructure updates during operation in a cloud, which is one of basic challenges complicating scheduling in clouds [56, 57]. The mix of private and public clouds make various resources more heterogeneous, which brings much more challenges. While, existing related works simplified the resource optimization problem by ignoring more or less heterogeneity or diversity of hybrid cloud resources, such as ignoring the heterogeneity between private resources and public resources, illustrated in Sect. 3.3.
- (4) **Private resource cost model.** The total costs of cloud operations are mainly composed of investment costs for buying infrastructures and operational costs which consist of the electricity costs for power, software copyright costs, hardware/software maintenance costs, and so on [171, 184]. In general, investment costs are consider as “sunk costs”, and operational costs are evaluated by consumed power

due to its very largest part of operational costs [71]. When concerning the cost of executing workloads on hybrid clouds, one should consider costs of both private and public resources used. The ways how to deal with the private resource costs by related works are shown in Sect. 3.4.

- (5) **Public resource costs.** When renting resources from public clouds, service providers, which is also public cloud users, are charged based on the amount of rented resources and the rent time. In existing related literatures, there are three types of resources, computing (VMs), storage, and network bandwidth, concerned to be charged by the public cloud. The cost model for each type of resources for each work is presented in Sect. 3.5.
- (6) **Factors of workload processing time.** The performance of workloads, e.g., finish time, response time, is one of the most concern in clouds. While there are a number of factors affecting the performance, and one work cannot address all of these factors. Therefore, each work has its own concern on some factors to solve the scheduling or provisioning problem in a hybrid cloud environment with simplification which was believed reasonable. Performance factors concerned by related works are shown in Sect. 3.6.

3.1 Requirements

Service providers have various requirements in different hybrid cloud environments concerning various QoS when managing hybrid resources. These requirements can be handled as optimization objectives or constraints in a hybrid cloud resource management problem. The requirements concerned by current works are as follows.¹ Table 1 summarizes requirements concerned by each work.

3.1.1 Profit or cost

For a service provider, the profit is of the most concern. The profit is the difference between the revenue and the cost (Fig. 3),

$$Profit = Revenue - Cost. \quad (1)$$

A user pays for its required services according to service level agreement (SLA) contracts with service providers.

¹ The objectives and constraints concerned by reviewed papers respectively correspond to the second and third columns in tables respectively summarizing related works for each type of workloads in “Appendix” which reviewed in detail each work by the categories classified by the workload type they focused on: the job with independent tasks, the workflow job, the (One-tier) web service as a whole, the (Multi-tier) web service with multiple components, etc.

Table 1 Classifying literatures based on requirements

Requirement	Literatures
Cost/profit (76.0%) ^a	[3, 4, 6–9, 11, 13, 15, 18, 20, 21, 23, 24, 30–32, 34–36, 38–41, 43, 44, 46, 47, 50, 59–61, 63, 65, 66, 70, 74, 76, 78–81, 83–87, 92–94, 98, 100, 101, 107, 108, 110, 113–119, 122–128, 131, 133, 136, 137, 139, 140, 147–149, 153, 154, 156–160, 162–164, 166, 167, 169, 177, 181, 186, 187, 189, 193, 194, 197, 198, 201, 205, 208–212, 215–217]
Performance (76.7%)	[4, 6–9, 13, 15, 18, 20, 20, 21, 23, 28, 29, 31, 31–33, 36–41, 43, 44, 46–49, 54, 58–61, 63, 65, 66, 70, 80, 81, 83–85, 90, 91, 95–98, 100–102, 108–110, 113–117, 119, 122, 124, 126, 128, 131, 133, 136–139, 141, 142, 145, 147–149, 152–154, 156–160, 166, 167, 174, 176, 178, 181, 186, 187, 189, 193, 194, 197, 198, 205, 208, 209, 209, 210, 210–217]
Resource amount (13.0%)	[3, 28, 29, 33–35, 40, 41, 74, 119, 131, 138, 162, 163, 170, 176, 178, 213, 214]
Utilization (2.1%)	[24, 33, 213]
Reliability availability (4.1%)	[13, 18, 37, 90, 91, 126]
Security (14.4%)	[11, 30, 50, 76, 80, 92, 94, 104, 111, 126, 132, 141, 149, 159, 160, 164, 166, 167, 177, 199, 201]
SLA (9.6%)	[30, 50, 94, 107, 118, 137, 158, 159, 164, 177, 202–204, 216]
Others	[80, 125, 192] (Balance, 2.1%) [86, 87](Contention, 1.4%) [213](RentTime) [119](QLength, 2.1%)
Multi-objective (21.2%)	[18, 32, 33, 44, 46, 47, 50, 65, 66, 70, 83–85, 100, 101, 110, 113–117, 122, 126, 159, 164, 174, 189, 193, 205, 213, 216]

^aThe percentage below a feature is the proportion of literatures concerning the feature among all related literatures (146)

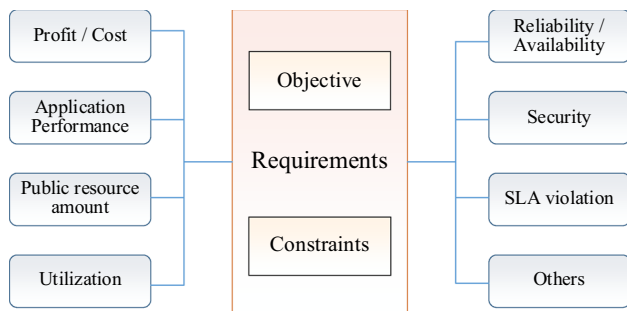


Fig. 3 The taxonomy by concerned requirements

The revenue of a service provider is the accumulated payment from all of its users for their service requests. Thus, when user requests and the payment of each request are known, the revenue of the service provider is a constant (*C*), and then, the profit maximization is equal to the cost minimization for a service provider.

$$Profit = C - Cost \Rightarrow \max Profit = C - \min Cost. \quad (2)$$

In a hybrid cloud environment, the cost of a service provider includes the costs for operating the local resources and renting the public resources, which would be illustrated in Sects. 3.4 and 3.5, respectively, as well as the penalty cost due to SLA violations,

$$Cost = Cost_{pri} + Cost_{pub} + Cost_{penalty}. \quad (3)$$

The service provider must pay the penalty when it is in breach of any SLA contract with users. Meanwhile, some

potential users may be lost, and thus the revenue and the profit may be reduced, if the service provider has violated some SLA contracts which reduces its reputation.

In the most of related works, the profit or the cost was concerned as the/an optimization objective (Profit²—profit maximization, Cost—cost minimization) or a constraint (Budget—the upper limit to the cost).

From Table 1, we can see that the profit/cost is one of the most concerns for providing services in hybrid clouds as it is one of the most important factors considered by cloud providers for increasing their incomes.

3.1.2 Application performance

A user always wants to get the result of its requests as soon as possible. Thus, the turnaround time, the time between submitting a request instance and receiving the result, usually is concerned as a QoS metric. For a batch job, the turnaround time is generally expressed as the finish time, or the makespan which is the period elapsed from the submission to the completion, while for a web service, it is expressed as the response time influenced by many factors, e.g., the communication delay between the user and the private cloud, the queuing delay, the processing delay, the communication delay between the private and public clouds.

For batch tasks/jobs, the finish time and the makespan are equivalent when the turnaround time is considered as

² We use the Courier font to represent possible values for the properties of related works in tables in “Appendix”.

the optimization objective (*Finish/Makespan*—finish time/makespan minimization). The finish time can be also considered as a constraint metric, *Deadline*—the time before which the job must be finished, which depends on various factors, e.g., the started time (*Start*), the execution time, the transfer time of input data (*Transfer*), etc.

For web services, the response time was concerned by some works as an optimization objective or a constraint metric (*Response*). While, most of works focusing on optimization of hybrid cloud resource management for web services only concerned a part of the response time, e.g., the queuing delay (*Queuing*), the processing delay (*Delay*), the communication delay between the private and public clouds (*Communication*). The part of the response time was also considered as an optimization objective or a constraint for each related work.

As shown in Table 1, the performance (turnaround time) is one of the most concerns in hybrid clouds as it is one of the most important factors considered by cloud users paying for qualified services delivered by clouds. The fail of satisfying performance requirements would increase the cost of cloud providers due to a violation of contracts, and hurt their reputation, leading to a lose of some potential users.

3.1.3 Public resource amount

The amount of public resources was considered by several works in two ways: some works concerned that the number of VMs may be limited in a public cloud, in practice, such as a user can run up to 20 instances at a time with a default plan in Amazon EC2; the public resources should be sufficient for satisfying users' requirements as the private resources are fixed.

In the first case, the public resource amount was considered as a constraint metric expressing the limitation of rented public resources in VM number (*VM Number*) [33, 40, 41, 74, 119, 138, 162, 162, 213, 214] or in amount of each resource type (*Amount Limit*) [131].

In the second case, the public resource amount was concerned as a metric of the objective or a constraint. When concerned as an objective metric, the amount was minimized with satisfying the performance requirements of provided services (*Amount Minimum*) [3, 28, 29, 34, 35, 176, 178], which indirectly reduces the cost of renting public resources. When the amount was considered as a constraint metric, a lower bound was set, representing the minimum requirements of users (*Amount Required*) [170] to indirectly guarantee the application performance.

3.1.4 Resource utilization

To improve the efficiency of resources, the resource utilization is one metric should be concerned. In general, the resource utilization and efficiency are positively correlated. Thus, there is a few work tried to maximize the resource utilization (*Utilization*) [33, 213]. While, a high utilization may reduce the reliability of infrastructures [17], and thus, some work restricted the resource utilization to an upper bound (*Utilization Limit*) [24].

3.1.5 Security

Security is a essential factor for whether users, especially enterprise users, exploit the public cloud because of the security and privacy issues of internal data or codes. Thus, several works have concerned the security as a constraints (*Security*) to restrict the location for processing workloads requiring high security. There usually are multiple levels for security requirements, and most of the related works considered simple two-level security model: the high security level requiring the workload must be processed in the private cloud, and the low security level representing the workload can be processed in both the private and public clouds.

3.1.6 Reliability/availability

Due to the increasing functionality and complexity of hybrid cloud computing, failures are inevitable, which may degrade the performance of processing workloads. For example, a task can be finished with its requirement without any failure, while its requirement may be violated if there is a failure which interrupts the task's execution. Thus, a few works improved the reliability or the availability of processing workloads, which had a great influence on the performance (*Reliability*, Arantes2017, Choi2015, Ben-Yehuda2012, Liu2015 and *Availability*) [90, 91].

Usually, the reliability and the availability can be both improved by redundant executions of workloads. While, the redundant execution requires more resources, and thus costs more. Therefore, there is a tradeoff between the reliability (or availability) and the cost.

3.1.7 SLA violation

There will be SLA violations when there is no resource with enough power for satisfying a workload, or when the service provider puts the profit first and there are some requests whose rejections cost less than their acceptations. Thus, there are some works concerned SLA violation, e.g., request rejections/failures, as an objective or a constraint metric (*SLA*) to minimize the number of SLA violations

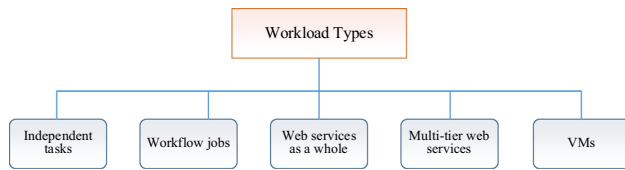


Fig. 4 The taxonomy by workload types

[30, 50, 137, 159, 164, 216] or to restrict the number to an upper bound [94, 107, 118, 158, 177, 202–204].

3.1.8 Others

A very few work respectively concerned the load balance between the private and public clouds (Balance) [80, 125, 192], which may improve the turnaround time by finishing workloads in clouds close to one another, the length of request queues (QLength) which was constrained within a bound for optimizing the queue time [119], or the resource contentions between services (Contention) [86, 87], e.g., the supports of library and operating system versions for each service, the conflicts of communication ports for services, etc.

If all rented (homogeneous) VMs have a same price, minimizing the cost and minimizing the total rent time of public VMs are equivalent. With this assumption, a very few work concerned the optimization of the total rent time instead of the cost for public VMs (RentTime) [213] (Fig. 4).

3.1.9 Single- or multi-objective

Cloud computing provides services for users usually with multiple QoS requirements. There are two ways to concern multiple metrics presented in the previous paragraphs simultaneously. The first one is to consider one metric as the optimization objective while others as constraints, formulating the hybrid cloud resource management as a single objective optimization problem. With another way, some works considered multiple metrics as objectives and modelled the hybrid cloud resource management as a multi-objective problem.

For a single objective optimization problem, there is always an optimization solutions not inferior to any other solution, while there usually is no such solution for a multi-objective problem. Therefore, all Pareto-efficient solutions should be solved to provide candidate solutions for service providers, where a Pareto-efficient solution outperforms any other solution in at least one objective, i.e., an objective cannot be improved without sacrificing other objectives.

Most of works optimizing multiple objectives focused on the tradeoff between cost and performance, two of the

most important factors concerned, where the performance is improved by increasing allocated resources, generally, leading to more costs for a workload. While there is few research concerning other factors, e.g., security, reliability, availability, when considering the tradeoff between/among multiple objectives.

3.2 Workloads

There are mainly two types of workloads focused by related works: batch jobs and long-running web services, explained as follows. There are also a few works focusing on the infrastructure (VM) service delivery using hybrid clouds regardless of workloads.

3.2.1 Batch jobs

Batch jobs, e.g., recommendation computing, financial analysis, weather forecast, generally require complex calculations, which take from a few seconds to a few days to complete, and thus mostly are not sensitive to short-term performance fluctuations. A batch job is usually divided into multiple tasks which are respectively dispatched to available resources within a duration so that all tasks of the job are completed as soon as possible or within its deadline.

Batch jobs can be classified into two categories, jobs with independent tasks and workflow jobs. There are plenty of jobs consisted of a number of trivially parallel tasks (called Jobs with Independent Tasks), such as parallel image rendering, data analysis. For these jobs, any two tasks do not depend on each other, and thus, all tasks can be executed in parallel. Such jobs are a kind of very common application in the parallel and distributed systems [75, 88]. Thus, a number of researchers focused on completing jobs with independent tasks on hybrid clouds.

While, a number of scientific computing jobs consist of multiple tasks with logic or data dependency relationships (called Workflow Jobs), i.e., a task can be started only when all tasks it depends on are finished. A workflow job is usually abstracted into a directed acyclic graph (DAG) where nodes are tasks and edges represent the dependences between corresponding tasks. More difficult than focusing on jobs with independent tasks, the work scheduling workflow jobs on hybrid clouds should take task dependences into account for maximizing the degree of task parallelism to improve the performance of workflow jobs.

3.2.2 Web services

Web services are long-running services handling short-lived latency-sensitive requests, where each request takes only a few milliseconds to a few hundred milliseconds.

Table 2 Classifying literatures based on workload types

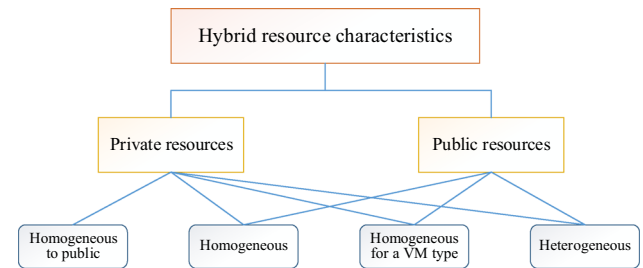
Type	Literatures
Independent batch tasks (49.3%)	[4, 6, 13, 15, 18, 20, 28, 29, 31, 32, 37–39, 43, 44, 46, 47, 54, 58–61, 63, 69, 70, 83–85, 93, 98, 102, 108–110, 113–117, 122, 123, 126, 133, 137, 139, 140, 142, 145, 153, 154, 162, 163, 169, 170, 174, 176, 178, 181, 186, 187, 189, 192–194, 197, 198, 208–211, 215, 216]
Workflow (21.2%)	[7–9, 21, 23, 33, 36, 40, 41, 48, 49, 65, 66, 74, 81, 95–97, 104, 124, 131, 136, 138, 152, 158, 160, 166, 167, 199, 213, 214]
One-tier web service (11.6%)	[24, 86, 87, 100, 101, 118, 119, 125, 128, 147–149, 156, 157, 202–204]
Multi-tier web service (6.1%)	[3, 34, 35, 78–80, 94, 107, 212]
VM (11.6%)	[11, 30, 50, 76, 90–92, 111, 127, 132, 141, 159, 164, 177, 201, 205, 217]

Such services are used for end-user-facing products such as web search, online video, business transaction, and for internal infrastructure services (e.g., distributed databases). Web services should return results to their requests as soon as possible as the request respond time has significant influence on the service providers' profit [82].

A web service usually consists of several tiers (or components). For example, the 3-tier web application architecture consisting of presentation, application and data tiers, has been widely used. For a web service, tiers have different requirements of resources due to their different functionalities, which motivates the use of virtualization for consolidating the instances of web service tiers to benefit from the complementarity of tiers' resource requirements.

While, a few works considered a web service instance as a whole for simplification (One-tier Web Service). Such works studied on methods vertically (reconfiguring the resources allocated to service instances) and/or horizontally (tuning the number of service instances) scaling web services to improve the cost of the service provider with guaranteed performance requirement.

For the works focusing on Multi-tier Web Services, the vertical and horizontal scale of instances should be concerned for each tier. Meanwhile, these works should improve the performance of the connection between instances of different tiers by reducing the communication distance using instance deployment/migration, which has significant influence on the response time of requests. Such things increase much more complexity for providing web services.

**Fig. 5** The taxonomy by hybrid resource characteristics

3.2.3 Literature review

From Table 2, we can see that there are about half of literatures focusing on independent tasks. This is mainly because its suitability of being outsourced to public clouds as there is no data transfer between tasks, which makes the performance of tasks not degraded by the scarce network resource between two clouds as (almost) no communication between tasks. There are also 21.2% researches optimizing the execution of workflow applications on hybrid clouds as the problems of resource heterogeneities, poor network resources between clouds, etc. can be addressed by carefully designing methods of workload scheduling to guarantee performance requirements (Fig. 5).

Web services have stringent requirements in performance due to their interacting with users, and are tunable for each component in instance number [188]. Thus it is more challenging when providing them in hybrid clouds. Therefore, there are relatively fewer studies focusing on providing web services on hybrid clouds, being only about a quarter of studies focusing on batch jobs in amount.

There are only about 11.6% literatures focusing on infrastructure (VM) service delivery regardless of workload features in the perspective of an IaaS provider, as it is more useful and more efficient when executing applications on clouds concerning the application characteristics while it is more simple than providing web services. When using these research result, a service provider must decide the amounts of required private resources and rented public resources, according to its load and service characteristics, which is a challenging question.

3.3 Hybrid resource characteristics

Resources of a hybrid cloud are composed of Local Resources and resources rented from public clouds (Public Resources). There are many different characters between local and public resources, as shown in Table 3.

In general, local resources have better performance and less costs than public resources, and thus almost all of

Table 3 The comparison between local resources and public resources

	Local resources	Public resources
Performance	Better	Worse
Performance/price	Higher	Lower
Resource amount	Limited	(Seem) unlimited Some limited
Homogeneity	Mostly heterogeneous	Homogeneous
Security	Higher	Lower
Reliability	Lower	Higher
Elasticity	Almost no	Better

related works used local resources whenever possible as their high performance cost ratio. While, the amount of local resources is limited, thus, the public resources are rented when the workloads are so high that local resources cannot satisfy all QoS requirements. Even though a public cloud provisions “unlimited” resources, there are some restrictions for public cloud users in the form of resource amount (VM instance number).

Usually, local resources are heterogeneous (*Heter*) as servers are gradually provisioned and replaced over the operation of the private cloud (or clusters, grids) [55]. While, plenty of related works (about 65.5% as shown in Table 4) considered local resources as homogeneous for simplicity. There are two degrees of homogeneity considered by existing related works: (i) all of local resources, physical machines (PMs), VMs or CPU cores, are homogeneous (*Homo*); (ii) similar to the public cloud, local resources are homogeneous for VMs with one type (*HomoType*). For simplicity, 88.2% related works, as shown in Table 4, treated local resources the same as public resources for seamlessly combining the private and public clouds (*Seamless*). In general, a public cloud provisions homogeneous VM instances with one type (*HomoType*), while, some related works regarded all public resources as homogeneous for simplicity (*Homo*) or as heterogeneous for generalization (*Heter*).³

Compared to the public cloud, the local cloud provided more security/private services as it only serves internal users. While local resources are often regarded as less reliable as it is too expensive to maintain resources with high reliability, for example, traditional and desktop grids have yearly resource availability averages of 70% or less [105]. Public clouds, in contrast, have SLAs that guarantee resource availability averages of over 99%. Thus, some related works tried to migrate services to public clouds for

Table 4 Classifying literatures based on resource characteristics

Character	Literatures
Alike - (24.1%) ^a	[3, 31, 40, 41, 44, 46, 47, 63, 80, 93, 104, 111, 113–118, 122, 123, 127, 136, 140, 141, 145, 149, 156, 157, 169, 174, 181, 192, 194, 212, 214]
Alike HT (24.8%)	[4, 6, 15, 30, 33, 38, 39, 43, 50, 59–61, 74, 81, 83–85, 110, 132, 133, 137, 152–154, 158, 159, 162, 163, 166, 167, 201, 208–211, 217]
Alike Homo (28.9%)	[11, 13, 20, 28, 29, 32, 34, 35, 37, 54, 58, 65, 66, 76, 78, 79, 86, 87, 90–92, 95–97, 100, 101, 119, 125, 128, 138, 142, 147, 148, 176–178, 193, 197, 198, 205, 213, 215]
Alike Heter (10.4%)	[7–9, 21, 23, 36, 70, 98, 102, 107, 124, 126, 131, 160, 189, 216]
Heter HT (1.4%)	[186, 187]
Heter Homo (1.4%)	[18, 109]
Homo Heter (0.7%)	[24]
Homo HT (0.7%)	[108]
- (7.6%)	[48, 49, 69, 94, 139, 164, 170, 199, 202–204]

^aIn the resource characteristics column, the first and second rows represent the characteristics of local and public resources, respectively. ‘-’ represents that the work did not consider the resource characteristic

increasing reliability, which could prove prohibitively expensive [18].

3.4 Cost model of local resources

For service providers, the investment costs of local infrastructures, which cannot be reduced by runtime resource managements, are usually considered as “sunk costs”. Thus, the vast majority of works in resource managements concerned the reduction of the operational costs consisted of the electricity costs for power, hardware/software maintenance costs, and so on [171, 184]. As these electricity costs make up the largest part of operational costs, and above 90% of energy is consumed for computing ($Energy_{Com}$), networking ($Energy_{Net}$) and cooling ($Energy_{Coo}$) [71], the electricity costs for powering PMs, network and cooling equipments can be regarded approximately as the operating cost of local resources (private cost for short),

$$Cost_{Pri} \approx price_e \cdot (Energy_{Com} + Energy_{Net} + Energy_{Coo}) + C, \quad (4)$$

where $price_e$ is the unit price of electricity, C is a constant representing relatively fixed private costs including software copyright costs, electricity costs for powering auxiliary equipment, salaries of staffs, etc. (Fig. 6).

³ The Resource Characteristics of the local and public clouds regarded by reviewed papers respectively correspond to the fourth and fifth columns in tables in “Appendix”.

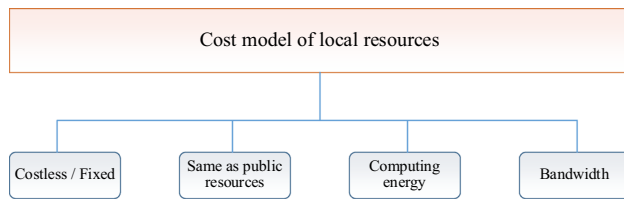


Fig. 6 The taxonomy by local resource cost model

Table 5 Classifying literatures based on concerned local resource costs

costs	Literatures
Costless/fixed (67.1%)	[3, 7–9, 13, 15, 20, 20, 21, 23, 28–37, 40, 41, 43, 44, 46–50, 54, 58, 63, 69, 70, 76, 78–81, 90, 91, 93–98, 102, 104, 109, 111, 113–117, 119, 122, 124–127, 131, 132, 136–142, 145, 147, 149, 152, 156–160, 170, 174, 176–178, 186, 187, 192, 193, 202–204, 211–216]
Same (24.7%)	[4, 6, 18, 38, 39, 59–61, 65, 66, 74, 83–87, 100, 101, 107, 110, 133, 148, 153, 154, 162, 163, 166, 167, 169, 181, 189, 201, 208–210, 217]
Energy (7.5%)	[11, 24, 92, 108, 123, 128, 164, 194, 197, 198, 205]
Bandwidth (0.7%)	[118]

For a decade, reducing the electricity costs in a private cloud, a cluster, or a grid has been studied by many works [179, 200], which are worth borrowing for optimizing costs in hybrid cloud environments. While, most of existing works focusing on hybrid cloud resource managements, about 67.1% as shown in Table 5, did not consider the costs of local resources, or considered them costless (–) or a Fixed value, considering that operational costs of local resources were much less than the rent cost of public resources. Most of works concerning the costs of using local resources, about 75.1% ($= 24.7\% / (1 - 67.1\%)$) as shown in Table 5, regarded the cost model the same as that of public resources (Same), without considering their difference. Only 7.5% related works concerned the computing Energy cost for the local resources with simple energy consumption model, e.g., the linear relationship between consumed energy of the private cloud and the number of tasks/requests [194, 197, 198], the linear relationship between energy consumed by a data center and the number of active PMs/VMs [128]. There was also a very few related works assumed that the service provider pay for used network bandwidth (Net) in the local cloud [118].⁴

⁴ The concerned cost of local resources (Local Cost) in related works corresponds to the sixth column in the tables in “Appendix”.

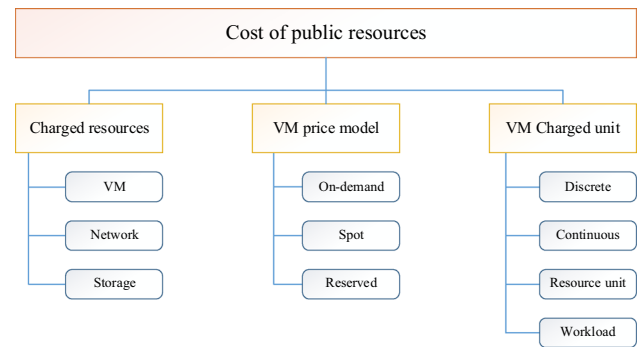


Fig. 7 The taxonomy by public resource costs

3.5 Cost of public resources

Service providers should pay for various resources, e.g., computing (C), network (N), and storage resources (S), rented from public clouds. Usually, public resources are charged on a basis of unit time ($Time_{unit}$), for example, Amazon EC2 charges its VMs hourly. The resource with the lease time less than one unit is charged for one time unit, for example, if one rents a VM with \$0.2/h price for 2.2 h, it needs to pay \$0.6 ($\$0.2/h \times \lceil 2.2 \rceil$ h) instead of \$0.44 (Fig. 7).

In general, the computing resources provisioned by public clouds are in the forms of VMs. Thus, the cost for renting a VM ($Cost_{VM}$) is

$$Cost_{VM} = Price_{VM} \cdot \left\lceil \frac{Time_{VM}}{Time_{unit}} \right\rceil, \quad (5)$$

where $Time_{VM}$ is the lease time of the VM.

There are usually three price model for rent VMs, On-demand, Spot and Reserved. With on-demand model, the public cloud provisions VMs as soon as its users pay and withdraws them when their rents are due. The price of on-demand VMs is usually stable during a relatively long period. Spot VMs are bid by multiple users. The public cloud provisions spot VMs for a user only when it bids higher than the spot price which varies depending on market supply and demand, and will withdraw them either when their rents are due or when the spot price increases to higher than the user bid. For most of time, spot VMs are cheaper than on-demand VMs, providing cost saving opportunities with proper bid strategy, while they are more expensive sometimes. The reserved VMs are rented and paid by users for a long time, e.g., weeks, months. Reserved VMs are cheaper than on-demand VMs in unit price, saving cost for service providers often having relatively high workload. Thus, the combination of these three kind of VMs help service providers to optimize their costs satisfying various requirements, while almost all related

Table 6 Classifying literatures based on public VM price model

price model	Literatures
On-demand (70.5%)	[3, 4, 6–9, 11, 13, 15, 18, 20, 21, 23, 24, 30–36, 38–41, 43, 44, 46, 47, 50, 59–61, 63, 65, 66, 70, 74, 76, 78, 79, 81, 83–85, 92, 93, 95, 98, 100, 101, 107, 108, 110, 113–117, 119, 122, 124–128, 131, 133, 136, 137, 139, 140, 147, 148, 152–154, 158–160, 162–164, 166, 167, 169, 177, 181, 186, 187, 189, 193, 197, 198, 201, 205, 208–212, 215–217]
Spot (3.2%)	[54, 58, 119, 139, 194]
Reserved (0.7%)	[119]

Table 7 Classifying literatures based on charged unit of public VMs

Price model	Literatures
Discrete (32.2%)	[4, 15, 20, 31, 38, 39, 43, 59–61, 108, 133, 139, 140, 153, 154, 162, 163, 181, 186, 187][24, 30, 74, 78, 79, 81, 83–85, 98, 119, 124, 125, 128, 131, 147, 148, 158, 164, 166, 167, 194, 197, 198, 201, 205, 217]
Continuous (13.0%)	[6, 11, 13, 18, 21, 23, 50, 63, 65, 66, 70, 107, 110, 159, 208–211, 215]
Resource (9.6%)	[3, 34, 35, 86, 87, 93, 100, 101, 127, 137, 177, 189, 212, 216]
Workload (11.6%)	[32, 33, 44, 46, 47, 113–117, 122, 123, 126, 136, 160, 169, 193]

works, as shown in Table 6, only concerned on demand VMs. Only 3.2% and 0.7% related works respectively exploited spot VMs and reserved VMs in hybrid cloud environments.

Even though, in real world, most of public clouds charge their provisioned VMs in the discrete form of time (Discrete), e.g., Hourly, plenty of related works (67.8% as shown in Table 7) simplified the VM price model for their hybrid environments. 13% related works considered that the public VMs were charged by seconds or continuous time (Continuous). Many other related works (21.2%) used the price model of VMs with the bill unit being Workload (task or request) or Resource unit (e.g., VM) regardless of the length of the rent time. Very few work considered that there is a constant saved cost for using a public VM [212], optimizing cost benefits by migrating some workloads to a public cloud. Several works even did not concerned the cost for renting the public resources (-), which optimized the amount of rented public resources.⁵

⁵ In the tables, the public resources concerned to be charged (charged), the Price Model and the Charge Unit for public VMs correspond to the seventh, eighth, and ninth columns, respectively, in “Appendix”.

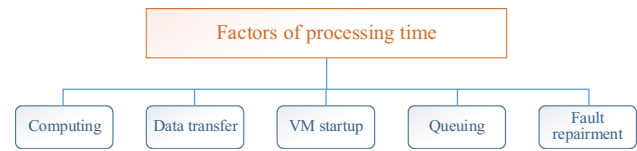


Fig. 8 The taxonomy by factors of processing time

For network resources, public clouds charge their users on the basis of network bandwidths (BW) and time. The time for using public network resources is proportional to the amount of transferred data ($Data_{Transfer}/BW$). Thus the cost for renting network resources in a public cloud

$$\begin{aligned}
 Cost_{Net} &= Price_{BW} \cdot BW \cdot \left[\frac{Data_{Transfer}}{BW} \right] \\
 &\approx Price_{BW} \cdot Data_{Transfer},
 \end{aligned}
 \tag{6}$$

where $Price_{BW}$ is the price of network resources in bandwidth unit and time unit. The network resources are paid for only the uplink and downlink data transfers in public clouds whose internal bandwidths are free to use. Thus, the data transfers between two clouds are charged. The broker is not recommended for service providers as the data transmission between two public clouds is transparent to them in both performance and cost, losing some opportunities for optimization in workload scheduling or resource provisioning (Fig. 8).

Public clouds charge the storage resources according to the amount of stored data ($Data_{STO}$) and its during ($Time_{STO}$),

$$Cost_{STO} = Price_{STO} \cdot Data_{STO} \cdot \left[\frac{Time_{STO}}{Time_{unit}} \right],
 \tag{7}$$

where $Price_{STO}$ is the cost for a unit of data, e.g., KiloBytes, MegaBytes, GigaBytes, per time unit.

Then a user should pay for rented computing, network and storage resources in a public cloud

$$Cost_{public} = \sum_{VM} Cost_{VM} + Cost_{Net} + Cost_{STO}.
 \tag{8}$$

Although public clouds charge their computing, network and storage resources, in general, there only about 6.2% related works, as shown in Table 8, concerning the costs of these three types of resources, and more than half of related researches only considered the cost of rented VMs for the sake of simplicity.

In hybrid cloud environments, service providers should minimizing their total cost. The outsourcing of workload may reduce the private cost while increase the public cost by increasing the rented public resources. Thus, a service provider should concern the tradeoff between the private and public costs to achieve the optimal overall cost.

Table 8 Classifying literatures based on public resources charged

Resource	Literatures
VM (53.5%)	[4, 7–9, 13, 15, 18, 24, 28–32, 36, 38–41, 44, 46, 47, 50, 54, 58, 63, 65, 66, 70, 74, 76, 83–87, 92, 93, 98, 108, 110, 113–117, 119, 122, 123, 125, 127, 128, 131, 137–140, 152–154, 158, 159, 162, 163, 176, 178, 181, 186, 187, 193, 194, 197, 198, 201, 205, 208–210, 217]
VM and network (20.5%)	[3, 6, 11, 21, 23, 34, 35, 43, 59–61, 80, 81, 94, 100, 101, 124, 133, 147, 148, 160, 164, 166, 167, 169, 177, 211, 212, 215]
VM and storage (0.7%)	[149]
VM and network and storage (6.2%)	[20, 33, 78, 79, 107, 126, 136, 189, 216]
Network and storage (2.1%)	[118, 156, 157]

3.6 Factors of processing time

In a cloud, the turnaround time of a workload (a task or a request) depends on various factors, e.g., the computing time (C), the transfer time for input data (T), the startup time of provisioned resources (S), the queue time (Q), the recover time when there is a failure (F), etc.

The computing time of a workload is decided by its computing load, e.g., the number of instructions to be executed, and the computing power of the resource (PM or VM) it is assigned to. The VM performance can be affected by the heterogeneity in underlying hardware [89], such as, VMs with same type (resource configuration) provided by heterogeneous architectures, e.g., POWER and X86, can have different performance. The evaluation of the computing time can be conducted by exploiting either the linear model of the load and the resource capacity or other complex models captured by some data analysis tools, e.g., machine learning [55].

The data transfer time is influenced by the bandwidth of the transmission link and the data amount. When outsourcing workloads to public clouds to improve the performance, service providers should concern the heterogeneity and the dynamics of available network resources [73]: (i) the bandwidth between two clouds is much less than that within a cloud; (ii) the bandwidths in public clouds may fluctuate much as the shared public resources by various users. While, less than half of related researches, as shown in Table 9, concerned the delay of data transfer and no related work considered the fluctuation of bandwidths, to our best of knowledge.

The startup time/delay, concerned by only 8.2% related literatures, is the difference between starting a VM and scheduling the workload on it, which is also known as

Table 9 Classifying literatures based on factors of workload processing time

Factor	Literatures
Transfer (41.8%)	[6–9, 20, 21, 23, 28, 29, 33, 36, 40, 41, 43, 44, 46, 47, 65, 66, 70, 74, 78–81, 94–97, 100–102, 104, 109, 113–117, 122, 124, 126, 131, 133, 136, 142, 145, 147, 156, 157, 160, 166, 167, 174, 176, 178, 189, 211–213, 216]
Startup (8.2%)	[15, 28, 29, 78, 79, 138, 139, 176, 178, 192, 213]
Queuing (2.7%)	[119, 147, 156, 157]
Recovering (0.7%)	[6]
Process time as a whole (5.5%)	[86, 87, 128, 148, 149, 202–204]

bootstrapping time, service initiation time or VM provisioning delay, made up by the time loading VM image, starting the operating system, installing software, configuring network and so on. Thus the available network is a factor affecting the startup time by influencing the image loading time. The types of services and VMs as well as cloud service providers are also important factors influencing the startup time [135]. The startup time ranges from seconds to dozens of minutes [135, 161], which may have a significant impact on the performance of applications, especially for latency-sensitive web services.

The queue time quantifies how long it takes to start executing a task/request from its arrival, which is an important factor for the performance of workloads, e.g., the finish time of batch jobs, the response time of web services. The queue time fluctuates strongly in time, and thus using the average value as the evaluation/prediction one, which is used by most of related works concerning the queue time for web services, may lead to many SLA violations. Using a percentile of the queue time, e.g. 90th, 95th, or a complex tools, e.g., stochastic process analysis [5], may be more suitable to evaluations.

As the increased scale and complex of clouds, the failures of processing workloads are inevitable, which also influence the workload performance. Service providers should apply some recovery approaches to handle these failures, which take some time and consume some resources.

All of these above factors contribute to the turnaround time of workloads, while all of existing related works, to the best of our knowledge, either concerned only a portion of them or used the average one as the turnaround time for homogeneous workloads (P), with simplifications.⁶

⁶ In “Appendix”, the last column of tables showed the factors of processing time concerned by related works.

4 Challenges and directions

In this section, we resumed the main issues related to hybrid cloud resource management still requiring research efforts and put forward some advice for future research direction.

4.1 Potential of using distributed public clouds

To avoid vendor lock-in and to improve the cost, service providers rent resources from multiple public clouds instead of only one public cloud, as there is no public cloud always having best cost-performance ratio due to the commercial competition. While the usage of multiple public clouds increases the complexity as it introduces several resource heterogeneities. The service provider should be careful to dispatch workload among clouds as the introduction of several public clouds may degrade the performance due to the low network performance between every two public clouds, which has been concerned by few works. Especially in the era of big data, there are plenty of data analysis applications whose performance is largely limited by the network resources. Thus, there is a tradeoff between benefits from more diversity of used public clouds and the overall workload performance.

4.2 Cost evaluation

It is necessary to establish a cost model for hybrid clouds to provide the cost optimization objective function and an evaluation method of resource management strategies for service providers. While modelling the cost is difficult as there are different costs for different resources or different resource amounts and as the price models of private and public resources is very different. In the private cloud, the cost of resources is influenced by many factors [112], the utilizations of computing and network infrastructures, supply of cooling, etc., each of which has various challenges need to be addressed [52]. In a public cloud, the service provider pays for their rented resources according to the resource amounts and the rent times. While, the prices of public resources, especially spot VM instances, vary with time [129], which should be modelled as a time series model for example. Intuitively, there is a positive correlation between the total cost and the allocated workload size in a cloud, thus, there is a tradeoff between the costs of the private and public clouds, which has not concerned by related works as most of related works ignored the cost of operating the private cloud.

4.3 Performance evaluation

In general, the performance requirements of users/workloads are defined as QoS, e.g., the finish time of batch jobs and the response time of web services. While, almost all related works used resource amount to express the requirements of workloads/users which simplifies the hybrid cloud resource management problem. Thus, existing works have to be employed with the relationship between QoS values and resource amounts in hybrid clouds, scarcely studied by researches. Therefore, it is necessary to establish the model mapping QoS requirements to various resources to address the problems, *how many resources and which hybrid resources should be provided to satisfy the QoS requirements?*

4.4 The VM provisioning delay

On a cloud, starting a VM instance needs seconds or minutes [161]. Ignoring the time consumed by VM provisioning may violate QoS requirements, e.g., deadline constraints of batch workloads, response time requirements of web services, which is what have done by almost all of related works. Thus the evaluation and the concern of the VM provisioning delay, the duration between the request and the running of a VM instance, is essential to the service provisioning in clouds. There are many variables for estimating VM provisioning delays should to be considered [135, 146], e.g., the virtualization technology, the instance type, the VM image loading, the software installing, the network configuration, the time of the day, the data center location, etc. The heterogeneous between private and public clouds also should be considered, due to the infrastructure heterogeneous and the available information for the service provider. Therefore, the evaluation of provisioning delays is still a challenging open problem need to be solved.

4.5 Workload prediction

To eliminate the negative impact of the VM provisioning delay on the workload performance, the prediction of workload sizes is necessary for provisioning VMs in advance. Thus, the time of predicting workload sizes have to be no longer than the VM provisioning delay. While few forecasting models were fast enough under these highly dynamic hybrid cloud circumstances [206], which may result in a prediction delay and provision insufficiency to deal with traffic bursts.

4.6 Reliability

The diversity, frequency, and number of failures are all increased with the hardware and software complexities in cloud platforms [105, 168]. Failures may increase the penalty cost of service providers by decreasing the workload performance, and thus violating SLA. There are many service providers having lost substantial revenue because of failures [168]. While failures are hard to diagnose/forecast or hard to repair due to the high dynamics when clouds are operated, the complex relationships among failures, the different characters of various resource/workload reliabilities (e.g., hardware vs. software, data vs. process) [144, 168], etc. Existing related works concerning reliability in hybrid clouds simplified the reliability analysis by assuming the reliability of running a workload in a resource was known. Reliability models which are applicable to real comprehensive hybrid cloud environments need to be researched to avoid more penalty costs for service providers.

4.7 Security

Users, especially enterprise users, have requirements of their data security and privacy. Security issues are one of the most factors enterprises move their data on to a public cloud [155]. Existing related works concerning security or privacy considered two levels of data security, private and public. The workload with private level cannot be outsourced to the public cloud, i.e. having location constraints, while the workload with public level can be processed in both the public and private clouds. While, the data protection technologies, e.g., encryption algorithms, data integrity auditing, access control, etc., provide opportunities for outsourcing some private workload or data to public clouds, to our best knowledge, which have employed by no research work related to hybrid cloud resource management, to overcome the problem of lacking private resources for the private workload or data, or to reduce the overheads of moving some executing workloads from the private cloud to the public cloud to make room for some new private workloads. However, there are resource overheads consumed by data protection technologies, and thus there is a tradeoff between the overheads of using protection technologies and of moving executing workloads for idling some private resources.

4.8 Optimization for hybrid workloads

In many cases, there are complementarities of resource requirements for different types of workloads in time or/and amount, e.g., compute-intensive batch jobs vs.

network-intensive web services. In production environments, a lot of service providers provide hybrid services. For example, Google clusters concurrently run long-running services handling short-lived latency-sensitive requests and batch jobs that take from a few seconds to a few days to complete [180, 207]. Thus, the resource efficiency can be improved by consolidating heterogeneous workloads with different characteristics, i.e., concurrently executing hybrid workloads, in hybrid clouds. While, existing related works did not focus on executing hybrid workloads in hybrid clouds, which is one of the most promising directions to improve the profit of service providers.

5 Conclusion

In this paper, we presented a taxonomy to classify the research works of resource provisioning and workload scheduling in hybrid clouds according to various factors considered, the optimization objective, the constraints, the workload type, the heterogeneities of hybrid resources, the cost model of local and public resources, and the concerned factors of turnaround time, and investigated the current research status based on the taxonomy. Then, we presented several issues as well as research directions about hybrid cloud management still requiring research efforts, the potential of using multiple public clouds, the cost model of hybrid resources, the performance evaluation in hybrid clouds, the concern of VM provisioning delay, the reliability guarantee, the security guarantee, and optimization for hybrid workloads in hybrid clouds. We believe our survey work is helpful for industrial circles and academic interested in hybrid clouds.

Acknowledgements The research was supported in part by the Key Scientific Research Projects of Henan Higher School (Grant No. 19A520043), the Key Science and Technology Program of Henan Province (Grant Nos. 192102210291, 172102210540), the National Natural Science Foundation of China (Grant Nos. 61872043, 61975187), Qin Xin Talents Cultivation Program, Beijing Information Science and Technology University (No. QXTCP B201904), the Fund of the Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (Grant No. ICDDXN004), the Foundation Training Program for Young Key Teachers of Zhengzhou University of Light Industry, and the Research Fund for the Doctoral Program of Zhengzhou University of Light Industry.

Appendix

See Tables 10, 11, 12, 13, 14, and 15.

Table 10 Workload scheduling and resource provisioning for jobs with independent tasks in hybrid cloud with one optimization objective

Literature	Requirements		Resource characteristics		Local cost	Cost of public resources			TT
	Optimization	Constraints	Local	Public		Charged	Price model	Charge unit	
Ditarso et al. [63]	Cost	Deadline	Alike	–	Fixed	C	On-demand	Continuous	C
Mattess et al. [139]	Cost	Deadline	–	–	–	C	Spot, on-demand	Hourly	S C
Zinnen and Engel [215]	Cost	Deadline	Alike	Homo	Fixed	C N	On-demand	Continuous	C
den Bossche et al. [59–61]	Cost	Deadline	Alike	HT	Same	C N	On-demand	Hourly	C
Bicer et al. [20]	Cost	Deadline	Alike	Homo	–	C N S	On-demand	Discrete	C T
Malawski et al. [133]	Cost	Deadline	Alike	HT	Same	C N	On-demand	Hourly	C T
Chu & Simmhan [43]	Cost	Deadline	Alike	HT	–	C N	Spot, on-demand	Hourly	C T
Wang et al. [186, 187]	Cost	Deadline	Heter	HT	–	C	On-demand	Hourly	C
Ahene et al. [6]	Cost	Deadline	Alike	HT	Same	C N	On-demand	Continuous	C T F
Balagoni and Rao [15]	Cost	Deadline	Alike	HT	–	C	On-demand	Hourly	C S
Peláez et al. [153, 154]	Cost	Deadline	Alike	HT	Same	C	On-demand	Hourly	C
Zhang et al. [211]	Cost	Deadline	Alike	HT	–	C N	On-demand	Continuous	C T
Lee & Lian [108]	Cost	Deadline	Homo	HT	Fixed + energy	C	On-demand	Hourly	C
Abdi et al. [4]	Cost	Deadline	Alike	HT	Same	C	On-demand	Hourly	C
Zhang and Sun [208]	Profit	Deadline	Alike	HT	Same	C	On-demand	Continuous	–
Caron and de Assunção [31]	Cost	Deadline, budget	Alike	–	–	C	On-demand	Hourly	C
Vilutis et al. [181]	Cost	Start	Alike	–	Same	C	On-demand	Hourly	C
Ruiz-Alvarez et al. [162, 163]	Cost	VM number	Alike	HT	Same	C	On-demand	Hourly	C
Arantes et al. [13]	Cost	Queuing, reliability	Alike	Homo	–	C	On-demand	Continuous	C
Jiang et al. [93]	Cost	–	Alike	–	–	C	On-demand	Resource	C
Mattess et al. [140]	Cost	–	Alike	–	–	C	On-demand	Hourly	C
Shifrin et al. [169]	Cost	–	Alike	–	Same	C N	On-demand	Workload	C
Choi and Kim [38, 39]	Cost/finish	Deadline	Alike	HT	Same	C	On-demand	Hourly	C
Yuan et al. [197, 198]	Cost/profit	Deadline	Alike	Homo	Energy	C	On-demand	Discrete	C
Xie et al. [194]	Profit	Queuing	Alike	–	Energy	C	Spot	Discrete	–
Fan et al. [69]	–	–	–	–	–	–	–	–	–
Lee and Zomay [109]	–	Deadline	Heter	Homo	–	–	–	–	C T
Siddiqui et al. [170]	–	Amount required	–	–	–	–	–	–	–
Lilienthal [123]	Cost	–	Alike	–	Fixed+ operate	C	–	Workload	–
Kim et al. [102]	Finish	–	Alike	Heter	–	–	–	–	T
Bicer et al. [20]	Finish	Budget	Alike	Homo	–	C N S	On-demand	Discrete	C T
Nahir et al. [145]	Finish	–	Alike	–	–	–	–	–	C T
Kang et al. [98]	Finish/cost	Deadline	Alike	Heter	–	C	On-demand	Hourly	C
Morla et al. [142]	Finish	–	Alike	Homo	–	–	–	–	C T
Choi et al. [37]	Finish	Reliability	Alike	Homo	–	–	–	–	C
Delamare et al. [54, 58]	Finish	–	Homo	Homo	–	C	Spot	Hourly	C
Zhang et al. [209, 210]	Makespan	Budget	Alike	HT	Same	C	On-demand	Continuous	C
Buyya et al. [28, 29, 176, 178]	Amount Minimum	Deadline	Alike	Homo	–	C	–	–	S C T
Marcu et al. [137]	SLA	Deadline, budget	Alike	HT	–	C	On-demand	Resource	C
Wu et al. [192]	Balance	–	Alike	–	–	–	–	–	S C

Table 11 Workload scheduling and resource provisioning for jobs with independent tasks in hybrid cloud with multiple optimization objectives

Literature	Requirements		Resource characteristics		Local cost		Cost of public resources		TT
	Optimization	Constraints	Local	Public	Charged	Price model	Charge unit		
								Local	
Ben-Yehuda et al. [18]	Cost, makespan	Reliability	Heter	Homo	C	On-demand	Continuous	C	
Wang et al. 2013 [189]	Cost, makespan	Deadline, budget	Alike	Heter	C N S	On-demand	Resource	C T	
Liu et al. 2015 [126]	Cost, makespan	Deadline, budget, security, reliability	Alike	Heter	C N S	On-demand	Workload	C T	
HoseinyFarahbady et al. [83–85]	Cost/+ makespan	Deadline	Alike	HT	C	On-demand	Hourly	C	
Champati and Liang [32]	Cost + makespan	–	Alike	Homo	C	On-demand	Workload	–	
Wu et al. [193]	Cost + finish	–	Alike	Homo	C	On-demand	Workload	C	
Leena et al. [110]	Cost, finish	–	Alike	HT	C	On-demand	Continuous	C	
Fan et al. [70]	Cost, finish	Deadline, budget	Alike	Heter	C	On-demand	Continuous	C T	
Taheri et al. [174]	Finish, transfer	–	Alike	–	–	–	–	C T	
Zuo et al. [216]	Finish, cost, SLA	Deadline, budget	Alike	Heter	C N S	On-demand	Resource	C T	
Li et al. [44, 46, 47, 113–115, 116, 117, 122]	Finish + profit	Deadline, budget, energy	Alike	–	C	On-demand	Workload	C T	

Table 12 Workload scheduling and resource provisioning for workflow jobs in hybrid cloud

Literature	Requirements		Resource characteristics		Local cost		Cost of public resources		TT
	Optimization	Constraints	Local	Public	Charged	Price model	Charge unit		
								Local	
Bittencourt et al. [21, 23]	Cost	Deadline	Alike	Heter	C N	On-demand	Continuous	C T	
Hoenisch et al. [81]	Cost	Deadline	Alike	HT	C N	On-demand	Discrete	C T	
Lin et al. [124]	Cost	Deadline	Alike	Heter	C N	On-demand	Hourly	C T	
Mao et al. [136]	Cost	Deadline	Alike	–	C N S	On-demand	Workload	C T	
Rahman et al. [160]	Cost	Deadline, budget, security	Alike	Heter	C N	On-demand	Workload	C T	
Sharif et al. [166, 167]	Cost	Deadline, security	Alike	HT	C N	On-demand	Hourly	C T	
Chopra and Singh [40, 41]	Cost	Deadline, VM number	Alike	–	C	On-demand	–	C T	
Genez et al. [74]	Cost	VM number	Alike	HT	C	On-demand	Discrete	C T	
Qiu et al. [158]	Cost	Deadline, SLA	Alike	HT	C	On-demand	Discrete	C	
Maheshwari et al. [131]	Cost	Amount limit	Alike	Heter	C	On-demand	Hourly	C T	
Ahn et al. [7–9, 36]	Cost/finish	Deadline	Alike	Heter	C	On-demand	–	C T	
Ko et al. [104]	–	Security	Alike	–	–	–	–	C T	
Kailasam et al. [95]	Finish	Deadline	Alike	Homo	–	On-demand	–	C T	
Kailasam et al. [96, 97]	Makespan	–	Alike	Homo	–	–	–	C T	

Table 12 (continued)

Literature	Requirements	Resource characteristics				Local cost			Cost of public resources			TT
		Constraints		Public		Charged	Price model	Charge unit	Charged	Price model	Charge unit	
		Local	Public	Local	Public							
Pasdar et al. [152]	Makespan	Deadline	HT	HT	0	C	On-demand	-	-	-	C	
Mattess et al. [138]	VM number	Deadline	Alike	Homo	-	C	-	-	-	-	S C	
Zhu et al. [214]	VM number	Deadline	Alike	-	-	-	-	-	-	-	C	
Clemente-Castelló et al. [48, 49]	Communication	-	-	-	-	-	-	-	-	-	-	
Duan et al. [65, 66]	Cost, makespan	-	Alike	Homo	Same	C	On-demand	Continuous	-	-	C T	
Zhu et al. [213]	VM number, RentTime, utilization	Deadline	Alike	Homo	-	-	-	-	-	-	S C T	
Chang et al. [33]	VM number, utilization	Deadline	Alike	HT	-	C N S	On-demand	Workload	-	-	C T	
Yuan et al. [199]	-	Security	-	-	-	-	-	-	-	-	-	

Table 13 Workload scheduling and resource provisioning for the one-tier web services in hybrid cloud

Literature	Requirements	Resource characteristics				Local cost			Cost of public resources			TT
		Constraints		Public		Charged	Price model	Charge unit	Charged	Price model	Charge unit	
		Local	Public	Local	Public							
Li et al. [118]	Cost	SLA	Alike	-	Net	N S	-	-	-	-	-	
Björkqvist et al. [24]	Cost	Utilization limit	Homo	Heter	Fixed + energy	C	On-demand	Hourly	Hourly	Hourly	C	
Li et al. [119]	Cost	Response, VM number, QLength	Alike	Homo	Fixed	C	On-demand, spot, reserved	Discrete	Discrete	Discrete	Q C	
Qiu et al. [156, 157]	Cost	Response	Alike	-	-	N S	-	-	-	-	Q T	
Ogawa et al. [148]	Cost	Response	Homo	Homo	Same	C N	On-demand	Hourly	Hourly	Hourly	P	
Hwang [86, 87]	Cost	Contention	Alike	Homo	Same	C	-	Resource	Resource	Resource	P	
Liu et al. [125]	Cost	Balance	Alike	Homo	-	C	On-demand	Discrete	Discrete	Discrete	-	
Lu et al. [128]	Profit	Response	Alike	Homo	Energy	C	On-demand	Discrete	Discrete	Discrete	P	
Zhang et al. [202–204]	-	SLA	-	-	-	-	-	-	-	-	P	
Oktay et al. [149]	Response	Budget, security	Alike	-	-	C S	-	-	-	-	P	
Niu et al. [147]	Response	Budget	Alike	Homo	-	C N	On-demand	Discrete	Discrete	Discrete	Q C T	
Kaviani et al. [100, 101]	Cost, response	-	Alike	Homo	Same	C N	On-demand	Resource	Resource	Resource	C T	

Table 14 Workload scheduling and resource provisioning for multi-tier web services in hybrid cloud

Literature	Requirements		Resource characteristics		Local cost	Cost of public resources		TT
	Optimization	Constraints	Local	Public		Charged	Price model	
Hajjat et al. [80]	Profit	Security, response, balance	Alike	–	–	C N	–	C T
Guo et al. [78, 79]	Cost	–	Alike	Homo	–	C N S	On-demand	S T
Charrada et al. [34, 35]	Cost	Amount minimum	Alike	Homo	–	C N	On-demand	Resource
Juan-Verdejo and Baars [94]	Profit	Security, SLA	–	–	–	C N	–	C T
Abbes et al. [3]	Cost	Amount minimum	Alike	–	–	C N	On-demand	Resource
Zhou et al. [212]	Cost	Response	Alike	–	–	C N	On-demand	Resource
Labba et al. [107]	Cost	SLA	Alike	Heter	Same	C N S	On-demand	Continuous

Table 15 Workload scheduling and resource provisioning for the infrastructure service delivery in hybrid cloud

Literature	Requirements		Resource characteristics		Local cost	Cost of public resources		TT
	Optimization	Constraints	Local	Public		Charged	Price model	
Grewal et al. 2013 [76]	Cost	Security	Alike	Homo	–	C	On-demand	–
Altmann et al. [11]	Cost	–	Alike	Homo	Fixed+ operate	C N	On-demand	Continuous
Unuvar et al. [177]	Cost	Budget, security, SLA	Alike	Homo	–	C N	On-demand	Resource
Jha & Gupta [92]	Consumed power	Security	Alike	Homo	Same	C	–	Energy
Cao et al. [30]	Cost	Security, SLA	Alike	HT	–	C	On-demand	Discrete
Zhang et al. [201, 217]	Profit	Deadline	Alike	HT	Same	C	On-demand	Hourly
Liu et al. [127]	Profit	–	Alike	–	–	C	On-demand	Resource
Makris et al. 2011 [132]	–	Security	Alike	HT	–	–	–	–
Javadi et al. [90, 91]	–	Deadline, Availability	Alike	Homo	–	–	–	C
Leitner et al. [111]	–	Security	Alike	–	–	–	–	–
Mechtri et al. 2017 [141]	–	Latency security	Alike	–	–	–	–	–
D’Agostino et al. [50, 159]	SLA, profit	Security	Alike	HT	–	C	On-demand	Continuous
Zhang et al. [205]	Cost, response	–	Homo	Homo	Energy	C	On-demand	Discrete
Saber et al. [164]	SLA, cost	Security	–	–	Energy	C N	On-demand	Discrete

References

- Alibaba Cloud: an integrated suite of cloud products, services and solutions. <https://www.alibabacloud.com/> (2018). Accessed 18 Jan 2020
- Amazon Elastic Compute Cloud (Amazon EC2). <http://aws.amazon.com/ec2/> (2018). Accessed 18 Jan 2020
- Abbes, W., Kechaou, Z., Alimi, A.M.: A new placement optimization approach in hybrid cloud based on genetic algorithm. In: 2016 IEEE 13th International Conference on e-Business Engineering (ICEBE), pp. 226–231 (2016). <https://doi.org/10.1109/ICEBE.2016.046>
- Abdi, S., PourKarimi, L., Ahmadi, M., Zargari, F.: Cost minimization for deadline-constrained bag-of-tasks applications in federated hybrid clouds. *Future Gener. Comput. Syst.* **71**, 113–128 (2017). <https://doi.org/10.1016/j.future.2017.01.036>
- Adam, O., Lee, Y.C., Zomaya, A.Y.: Stochastic resource provisioning for containerized multi-tier web services in clouds. *IEEE Trans. Parallel Distrib. Syst.* **28**(7), 2060–2073 (2017). <https://doi.org/10.1109/TPDS.2016.2639009>
- Ahene, E., Acheampong, K.N., Xu, H.: Fault-tolerant resource provisioning with deadline-driven optimization in hybrid clouds. *Int. J. Adv. Comput. Sci. Appl.* **7**(12), 379–389 (2016)
- Ahn, Y., Choi, J., Jeong, S., Kim, Y.: Auto-scaling method in hybrid cloud for scientific applications. In: The 16th Asia-Pacific Network Operations and Management Symposium, pp. 1–4 (2014). <https://doi.org/10.1109/APNOMS.2014.6996527>
- Ahn, Y., Kim, Y.: Auto-scaling of virtual resources for scientific workflows on hybrid clouds. In: Proceedings of the 5th ACM Workshop on Scientific Cloud Computing, ScienceCloud '14, pp. 47–52. ACM, New York (2014). <https://doi.org/10.1145/2608029.2608036>
- Ahn, Y., Kim, Y.: VM auto-scaling for workflows in hybrid cloud computing. In: 2014 International Conference on Cloud and Autonomic Computing, pp. 237–240 (2014). <https://doi.org/10.1109/ICCAC.2014.34>
- Al-Dhuraibi, Y., Paraiso, F., Djarallah, N., Merle, P.: Elasticity in cloud computing: state of the art and research challenges. *IEEE Trans. Serv. Comput.* **PP**(99), 1–1 (2017). <https://doi.org/10.1109/TSC.2017.2711009>
- Altmann, J., Kashef, M.M.: Cost model based service placement in federated hybrid clouds. *Future Gener. Comput. Syst.* **41**, 79–90 (2014). <https://doi.org/10.1016/j.future.2014.08.014>
- Amato, A., Venticinque, S.: Multiobjective optimization for brokering of multicloud service composition. *ACM Trans. Internet Technol.* **16**(2), 13:1–13:20 (2016)
- Arantes, L., Friedman, R., Marin, O., Sens, P.: Probabilistic Byzantine tolerance scheduling in hybrid cloud environments. In: Proceedings of the 18th International Conference on Distributed Computing and Networking, ICDCN '17, pp. 2:1–2:10. ACM, New York (2017). <https://doi.org/10.1145/3007748.3007770>
- Arlitt, M., Jin, T.: A workload characterization study of the 1998 World Cup Web site. *IEEE Netw.* **14**(3), 30–37 (2000). <https://doi.org/10.1109/65.844498>
- Balagoni, Y., Rao, R.R.: A cost-effective SLA-aware scheduling for hybrid cloud environment. In: 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), pp. 1–7 (2016). <https://doi.org/10.1109/ICIC.2016.7919621>
- Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I., Warfield, A.: Xen and the art of virtualization. In: Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, SOSP '03, pp. 164–177. ACM, New York (2003)
- Beloglazov, A., Abawajy, J., Buyya, R.: Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future Gener. Comput. Syst.* **28**(5), 755–768 (2012). Special Section: Energy efficiency in large-scale distributed systems
- Ben-Yehuda, O.A., Schuster, A., Sharov, A., Silberstein, M., Iosup, A.: ExPERT: Pareto-efficient task replication on grids and a cloud. In: 2012 IEEE 26th International Parallel and Distributed Processing Symposium, pp. 167–178 (2012)
- Bhosale, A.S., Bandari, S.D.: Survey on resource provisioning of hybrid cloud with Aneka. *Int. Adv. Res. J. Sci. Eng. Technol. Spec. Issue* **4**(11), 38–41 (2017)
- Bicer, T., Chiu, D., Agrawal, G.: Time and cost sensitive data-intensive computing on hybrid clouds. In: 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012), pp. 636–643 (2012). <https://doi.org/10.1109/CCGrid.2012.95>
- Bittencourt, L.F., Madeira, E.R.M.: HCOC: a cost optimization algorithm for workflow scheduling in hybrid clouds. *J. Internet Serv. Appl.* **2**(3), 207–227 (2011). <https://doi.org/10.1007/s13174-011-0032-0>
- Bittencourt, L.F., Madeira, E.R.M., Fonseca, N.L.S.D.: Scheduling in hybrid clouds. *IEEE Commun. Mag.* **50**(9), 42–47 (2012)
- Bittencourt, L.F., Senna, C.R., Madeira, E.R.M.: Scheduling service workflows for cost optimization in hybrid clouds. In: 2010 International Conference on Network and Service Management, pp. 394–397 (2010). <https://doi.org/10.1109/CNSM.2010.5691241>
- Björkqvist, M., Chen, L.Y., Binder, W.: Cost-driven service provisioning in hybrid clouds. In: 2012 Fifth IEEE International Conference on Service-Oriented Computing and Applications (SOCA), pp. 1–8 (2012). <https://doi.org/10.1109/SOCA.2012.6449447>
- Botta, A., de Donato, W., Persico, V., Pescapé, A.: Integration of integration of cloud computing and Internet of Things: a survey. *Future Gener. Comput. Syst.* **56**, 684–700 (2016). <https://doi.org/10.1016/j.future.2015.09.021>
- Boutaba, R., da Fonseca, N.L.: Cloud architectures, networks, services, and management. In: Cloud Services, Networking, and Management, pp. 1–22. Wiley, Hoboken (2015)
- Buyya, R., Barreto, D.: Multi-cloud resource provisioning with Aneka: a unified and integrated utilisation of Microsoft Azure and Amazon EC2 instances. In: 2015 International Conference on Computing and Network Communications (CoCoNet), pp. 216–229 (2015). <https://doi.org/10.1109/CoCoNet.2015.7411190>
- Calheiros, R., Buyya, R.: Cost-effective provisioning and scheduling of deadline-constrained applications in hybrid clouds. In: X. Wang, I. Cruz, A. Delis, G. Huang (eds) Web Information Systems Engineering—WISE 2012. Lecture Notes in Computer Science, vol. 7651, pp. 171–184. Springer, Berlin (2012)
- Calheiros, R.N., Vecchiola, C., Karunamoorthy, D., Buyya, R.: The Aneka Platform and QoS-driven resource provisioning for elastic applications on hybrid clouds. *Future Gener. Comput. Syst.* **28**(6), 861–870 (2012)
- Cao, Y., Lu, L., Yu, J., Qian, S., Zhu, Y., Li, M., Cao, J., Wang, Z., Li, J., Xue, G.: Online cost-aware service requests scheduling in hybrid clouds for cloud bursting. In: Bouguettaya, A., Gao, Y., Klimenko, A., Chen, L., Zhang, X., Dzerzhinskiy, F., Jia, W., Klimenko, S.V., Li, Q. (eds.) Web Information Systems Engineering—WISE 2017, pp. 259–274. Springer, Cham (2017)

31. Caron, E., de Assunção, M.D.: Multi-criteria malleable task management for hybrid-cloud platforms. In: 2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech), pp. 326–333 (2016). <https://doi.org/10.1109/CloudTech.2016.7847717>
32. Champati, J.P., Liang, B.: One-restart algorithm for scheduling and offloading in a hybrid cloud. In: 2015 IEEE 23rd International Symposium on Quality of Service (IWQoS), pp. 31–40 (2015). <https://doi.org/10.1109/IWQoS.2015.7404699>
33. Chang, Y.S., Fan, C.T., Sheu, R.K., Jhu, S.R., Yuan, S.M.: An agent-based workflow scheduling mechanism with deadline constraint on hybrid cloud environment. *Int. J. Commun. Syst.* **31**(1), e3401 (2018). <https://doi.org/10.1002/dac.3401>
34. Charrada, F.B., Tata, S.: An efficient algorithm for the bursting of service-based applications in hybrid clouds. *IEEE Trans. Serv. Comput.* **9**(3), 357–367 (2016). <https://doi.org/10.1109/TSC.2015.2396076>
35. Charrada, F.B., Tebourski, N., Tata, S., Moalla, S.: Approximate placement of service-based applications in hybrid clouds. In: 2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 161–166 (2012). <https://doi.org/10.1109/WETICE.2012.76>
36. Choi, J., Ahn, Y., Kim, S., Kim, Y., Choi, J.: VM auto-scaling methods for high throughput computing on hybrid infrastructure. *Clust. Comput.* **18**(3), 1063–1073 (2015). <https://doi.org/10.1007/s10586-015-0462-8>
37. Choi, J., Kim, S., Adufu, T., Hwang, S., Kim, Y.: A job dispatch optimization method on cluster and cloud for large-scale high-throughput computing service. In: 2015 International Conference on Cloud and Autonomic Computing, pp. 283–290 (2015). <https://doi.org/10.1109/ICAC.2015.42>
38. Choi, J., Kim, Y.: An adaptive resource provisioning method using job history learning technique in hybrid infrastructure. In: 2016 IEEE 1st International Workshops on Foundations and Applications of Self* Systems (FAS*W), pp. 72–77 (2016). <https://doi.org/10.1109/FAS-W.2016.27>
39. Choi, J., Kim, Y.: Adaptive resource provisioning method using application-aware machine learning based on job history in heterogeneous infrastructures. *Clust. Comput.* **20**(4), 3537–3549 (2017). <https://doi.org/10.1007/s10586-017-1148-1>
40. Chopra, N., Singh, S.: Deadline and cost based workflow scheduling in hybrid cloud. In: 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 840–846 (2013). <https://doi.org/10.1109/ICACCI.2013.6637285>
41. Chopra, N., Singh, S.: HEFT based workflow scheduling algorithm for cost optimization within deadline in hybrid clouds. In: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1–6 (2013). <https://doi.org/10.1109/ICCCNT.2013.6726627>
42. Chopra, N., Singh, S.: Survey on scheduling in hybrid clouds. In: Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1–6 (2014)
43. Chu, H.Y., Simmhan, Y.: Cost-efficient and resilient job life-cycle management on hybrid clouds. In: 2014 IEEE 28th International Parallel and Distributed Processing Symposium, pp. 327–336 (2014)
44. Chunlin, L., Jianhang, T., Youlong, L.: Distributed QoS-aware scheduling optimization for resource-intensive mobile application in hybrid cloud. *Clust. Comput.* (2017). <https://doi.org/10.1007/s10586-017-1171-2>
45. Chunlin, L., Layuan, L.: A distributed multiple dimensional QoS constrained resource scheduling optimization policy in computational grid. *J. Comput. Syst. Sci.* **72**(4), 706–726 (2006). <https://doi.org/10.1016/j.jcss.2006.01.003>
46. Chunlin, L., LaYuan, L.: Optimal scheduling across public and private clouds in complex hybrid cloud environment. *Inf. Syst. Front.* **19**(1), 1–12 (2017). <https://doi.org/10.1007/s10796-015-9581-2>
47. Chunlin, L., Min, Z., Youlong, L.: Elastic resource provisioning in hybrid mobile cloud for computationally intensive mobile applications. *J. Supercomput.* **73**(9), 3683–3714 (2017). <https://doi.org/10.1007/s11227-017-1965-2>
48. Clemente-Castelló, F.J., Nicolae, B., Mayo, R., Fernandez, J.C.: Performance model of MapReduce iterative applications for hybrid cloud bursting. *IEEE Trans. Parallel Distrib. Syst.* **PP**(99), 1–1 (2018). <https://doi.org/10.1109/TPDS.2018.2802932>
49. Clemente-Castelló, F.J., Nicolae, B., Rafique, M.M., Mayo, R., Fernández, J.C.: Evaluation of data locality strategies for hybrid cloud bursting of iterative MapReduce. In: Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid '17, pp. 181–185. IEEE Press, Piscataway (2017). <https://doi.org/10.1109/CCGRID.2017.96>
50. D'Agostino, D., Galizia, A., Clematis, A., Mangini, M., Porro, L., Quarati, A.: A QoS-aware broker for hybrid clouds. *Computing* **95**(1), 89–109 (2013). <https://doi.org/10.1007/s00607-012-0254-4>
51. Dastjerdi, A.V., Buyya, R.: Fog computing: helping the Internet of Things realize its potential. *Computer* **49**(8), 112–116 (2016). <https://doi.org/10.1109/MC.2016.245>
52. Dayarathna, M., Wen, Y., Fan, R.: Data center energy consumption modeling: a survey. *IEEE Commun. Surv. Tutor.* **18**(1), 732–794 (2016). <https://doi.org/10.1109/COMST.2015.2481183>
53. de Assunção, M.D., di Costanzo, A., Buyya, R.: A cost-benefit analysis of using cloud computing to extend the capacity of clusters. *Clust. Comput.* **13**(3), 335–347 (2010)
54. Delamare, S., Fedak, G., Kondo, D., Lodygensky, O.: SpeQuloS: a QoS service for hybrid and elastic computing infrastructures. *Clust. Comput.* **17**(1), 79–100 (2014). <https://doi.org/10.1007/s10586-013-0283-6>
55. Delimitrou, C., Kozyrakis, C.: Paragon: QoS-aware scheduling for heterogeneous datacenters. In: Proceedings of the Eighteenth International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS '13, pp. 77–88. ACM, New York (2013). <https://doi.org/10.1145/2451116.2451125>
56. Delimitrou, C., Kozyrakis, C.: QoS-aware scheduling in heterogeneous datacenters with Paragon. *ACM Trans. Comput. Syst.* **31**(4), 12:1–12:34 (2013). <https://doi.org/10.1145/2556583>
57. Delimitrou, C., Kozyrakis, C.: Quasar: resource-efficient and QoS-aware cluster management. In: Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS '14, pp. 127–144. ACM, New York (2014). <https://doi.org/10.1145/2541940.2541941>
58. Delamare, S., Fedak, G., Kondo, D., Lodygensky, O.: SpeQuloS: A QoS service for BoT applications using best effort distributed computing infrastructures. In: Proceedings of the 21st International Symposium on High-Performance Parallel and Distributed Computing, HPDC '12, pp. 173–186. ACM, New York (2012). <https://doi.org/10.1145/2287076.2287106>
59. den Bossche, R.V., Vanmechelen, K., Broeckhove, J.: Cost-optimal scheduling in hybrid IaaS clouds for deadline constrained workloads. In: 2010 IEEE 3rd International Conference on Cloud Computing, pp. 228–235 (2010). <https://doi.org/10.1109/CLOUD.2010.58>
60. den Bossche, R.V., Vanmechelen, K., Broeckhove, J.: Cost-efficient scheduling heuristics for deadline constrained workloads on hybrid clouds. In: 2011 IEEE Third International Conference

- on Cloud Computing Technology and Science, pp. 320–327 (2011). <https://doi.org/10.1109/CloudCom.2011.50>
61. den Bossche, R.V., Vanmechelen, K., Broeckhove, J.: Online cost-efficient scheduling of deadline-constrained workloads on hybrid clouds. *Future Gener. Comput. Syst.* pp. 973–985. (2013). <https://doi.org/10.1016/j.future.2012.12.012>. Special Section: Utility and Cloud Computing
 62. Dinh, H.T., Lee, C., Niyato, D., Wang, P.: A survey of mobile cloud computing: architecture, applications, and approaches. *Wirel. Commun. Mob. Comput.* **13**(18), 1587–1611 (2013). <https://doi.org/10.1002/wcm.1203>
 63. Ditarso, P., Figueiredo, F., Maia, D., Brasileiro, F., Coelho, A.: On the planning of a hybrid IT infrastructure. In: NOMS 2008—2008 IEEE Network Operations and Management Symposium, pp. 496–503 (2008). <https://doi.org/10.1109/NOMS.2008.4575173>
 64. Docker: build, manage and secure your apps anywhere. <https://www.docker.com/> (2018). Accessed 18 Jan 2020
 65. Duan, R., Prodan, R.: Cooperative scheduling of bag-of-tasks workflows on hybrid clouds. In: 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, pp. 439–446 (2014). <https://doi.org/10.1109/CloudCom.2014.58>
 66. Duan, R., Prodan, R., Li, X.: Multi-objective game theoretic scheduling of bag-of-tasks workflows on hybrid clouds. *IEEE Trans. Cloud Comput.* **2**(1), 29–42 (2014). <https://doi.org/10.1109/TCC.2014.2303077>
 67. Engineering Village. <https://www.engineeringvillage.com/> (2018). Accessed 18 Jan 2020
 68. Fadel, A.S., Fayoumi, A.G.: Cloud resource provisioning and bursting approaches. In: 2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp. 59–64 (2013)
 69. Fan, C.T., Wang, W.J., Chang, Y.S.: Agent-based service migration framework in hybrid cloud. In: 2011 IEEE International Conference on High Performance Computing and Communications, pp. 887–892 (2011). <https://doi.org/10.1109/HPCC.2011.127>
 70. Fan, Y., Liang, Q., Chen, Y., Yan, X., Hu, C., Yao, H., Liu, C., Zeng, D.: Executing time and cost-aware task scheduling in hybrid cloud using a modified DE algorithm. In: Li, K., Li, J., Liu, Y., Castiglione, A. (eds.) *Computational Intelligence and Intelligent Systems*, pp. 74–83. Springer, Singapore (2016)
 71. Fang, S., Kanagavelu, R., Lee, B.S., Foh, C.H., Aung, K.M.M.: Power-efficient virtual machine placement and migration in data centers. In: 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, pp. 1408–1413 (2013)
 72. Faniyi, F., Bahsoon, R.: A systematic review of service level management in the cloud. *ACM Comput. Surv.* **48**(3), 43:1–43:27 (2015). <https://doi.org/10.1145/2843890>
 73. Genez, T.A.L., Bittencourt, L., Fonseca, N., Madeira, E.: Estimation of the available bandwidth in inter-cloud links for task scheduling in hybrid clouds. *IEEE Trans. Cloud Comput.* **PP**(99), 1–1 (2015). <https://doi.org/10.1109/TCC.2015.2469650>
 74. Genez, T.A.L., Bittencourt, L.F., Madeira, E.R.M.: On the performance-cost tradeoff for workflow scheduling in hybrid clouds. In: Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing, UCC '13, pp. 411–416. IEEE Computer Society, Washington, DC (2013). <https://doi.org/10.1109/UCC.2013.82>
 75. Goder, A., Spiridonov, A., Wang, Y.: Bistro: scheduling data-parallel jobs against live production systems. In: Proceedings of the 2015 USENIX Conference on Usenix Annual Technical Conference, USENIX ATC '15, pp. 459–471. USENIX Association, Berkeley (2015)
 76. Grewal, R.K., Pateriya, P.K.: Chap. 5. In: *A Rule-Based Approach for Effective Resource Provisioning in Hybrid Cloud Environment*, pp. 41–57. Springer, Berlin (2013)
 77. Google Scholar. <https://scholar.google.com/> (2018). Accessed 18 Jan 2020
 78. Guo, T., Sharma, U., Shenoy, P., Wood, T., Sahu, S.: Cost-aware cloud bursting for enterprise applications. *ACM Trans. Internet Technol.* **13**(3), 10:1–10:24 (2014)
 79. Guo, T., Sharma, U., Wood, T., Sahu, S., Shenoy, P.: Seagull: intelligent cloud bursting for enterprise applications. In: Proceedings of the 2012 USENIX Conference on Annual Technical Conference, USENIX ATC '12, pp. 33–33. USENIX Association, Berkeley (2012)
 80. Hajjat, M., Sun, X., Sung, Y.W.E., Maltz, D., Rao, S., Sripanidkulchai, K., Tawarmalani, M.: Cloudward bound: planning for beneficial migration of enterprise applications to the cloud. In: Proceedings of the ACM SIGCOMM 2010 Conference, SIGCOMM '10, pp. 243–254. ACM, New York (2010). <https://doi.org/10.1145/1851182.1851212>
 81. Hoenisch, P., Hochreiner, C., Schuller, D., Schulte, S., Mendling, J., Dustdar, S.: Cost-efficient scheduling of elastic processes in hybrid clouds. In: 2015 IEEE 8th International Conference on Cloud Computing, pp. 17–24 (2015). <https://doi.org/10.1109/CLOUD.2015.13>
 82. Hoff, T.: Latency is everywhere and it costs you sales—how to crush it. <http://highscalability.com/blog/2009/7/25/latency-is-everywhere-and-it-costs-you-sales-how-to-crush-it.html> (2009)
 83. HoseinyFarahabady, M., Lee, Y., Zomaya, A.: Randomized approximation scheme for resource allocation in hybrid-cloud environment. *J. Supercomput.* **69**(2), 576–592 (2014)
 84. HoseinyFarahabady, M., Lee, Y.C., Zomaya, A.: Pareto-optimal cloud bursting. *IEEE Trans. Parallel Distrib. Syst.* **25**(10), 2670–2682 (2014)
 85. HoseinyFarahabady, M., Samani, H., Leslie, L., Lee, Y.C., Zomaya, A.: Handling uncertainty: Pareto-efficient BoT scheduling on hybrid clouds. In: 2013 42nd International Conference on Parallel Processing (ICPP), pp. 419–428 (2013)
 86. Hwang, J.: Computing resource transformation, consolidation and decomposition in hybrid clouds. In: 2015 11th International Conference on Network and Service Management (CNSM), pp. 144–152 (2015). <https://doi.org/10.1109/CNSM.2015.7367350>
 87. Hwang, J.: Toward beneficial transformation of enterprise workloads to hybrid clouds. *IEEE Trans. Netw. Serv. Manag.* **13**(2), 295–307 (2016). <https://doi.org/10.1109/TNSM.2016.2541120>
 88. Iosup, A., Epema, D.: Grid computing workloads. *IEEE Internet Comput.* **15**(2), 19–26 (2011)
 89. Jackson, K.R., Ramakrishnan, L., Muriki, K., Canon, S., Cholia, S., Shalf, J., Wasserman, H.J., Wright, N.J.: Performance analysis of high performance computing applications on the Amazon Web Services Cloud. In: 2010 IEEE Second International Conference on Cloud Computing Technology and Science, pp. 159–168 (2010)
 90. Javadi, B., Abawajy, J., Buyya, R.: Failure-aware resource provisioning for hybrid Cloud infrastructure. *J. Parallel Distrib. Comput.* **72**(10), 1318–1331 (2012). <https://doi.org/10.1016/j.jpdc.2012.06.012>
 91. Javadi, B., Abawajy, J., Sinnott, R.O.: Hybrid Cloud resource provisioning policy in the presence of resource failures. In: 4th IEEE International Conference on Cloud Computing Technology and Science Proceedings, pp. 10–17 (2012). <https://doi.org/10.1109/CloudCom.2012.6427521>
 92. Jha, R.S., Gupta, P.: Power aware resource allocation policy for hybrid cloud. In: 2015 Third International Conference on Image

- Information Processing (ICIP), pp. 336–341 (2015). <https://doi.org/10.1109/ICIP.2015.7414791>
93. Jiang, W.Z., Sheng, Z.Q.: A new task scheduling algorithm in hybrid cloud environment. In: 2012 International Conference on Cloud and Service Computing (CSC), pp. 45–49 (2012)
 94. Juan-Verdejo, A., Baars, H.: Decision support for partially moving applications to the cloud: the example of business intelligence. In: Proceedings of the 2013 International Workshop on Hot Topics in Cloud Services, HotTopiCS '13, pp. 35–42. ACM, New York (2013). <https://doi.org/10.1145/2462307.2462316>
 95. Kailasam, S., Dhawalia, P., Balaji, S.J., Iyer, G., Dharanipragada, J.: Extending MapReduce across Clouds with BStream. *IEEE Trans. Cloud Comput.* **2**(3), 362–376 (2014). <https://doi.org/10.1109/TCC.2014.2316810>
 96. Kailasam, S., Gnanasambandam, N., Dharanipragada, J., Sharma, N.: Optimizing service level agreements for autonomic cloud bursting schedulers. In: 2010 39th International Conference on Parallel Processing Workshops, pp. 285–294 (2010). <https://doi.org/10.1109/ICPPW.2010.54>
 97. Kailasam, S., Gnanasambandam, N., Dharanipragada, J., Sharma, N.: Optimizing ordered throughput using autonomic cloud bursting schedulers. *IEEE Trans. Softw. Eng.* **39**(11), 1564–1581 (2013). <https://doi.org/10.1109/TSE.2013.26>
 98. Kang, H., Koh, J., Kim, Y., Hahm, J.: A SLA driven VM auto-scaling method in hybrid cloud environment. In: 2013 15th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp. 1–6 (2013)
 99. Kang, X., Zhang, H., Jiang, G., Chen, H., Meng, X., Yoshihira, K.: Measurement, modeling, and analysis of internet video sharing site workload: a case study. In: IEEE International Conference on Web Services, 2008. ICWS '08, pp. 278–285 (2008)
 100. Kaviani, N., Wohlstadter, E., Lea, R.: MANTICORE: a framework for partitioning software services for hybrid cloud. In: 4th IEEE International Conference on Cloud Computing Technology and Science Proceedings, pp. 333–340 (2012). <https://doi.org/10.1109/CloudCom.2012.6427541>
 101. Kaviani, N., Wohlstadter, E., Lea, R.: Partitioning of web applications for hybrid cloud deployment. *J. Internet Serv. Appl.* **5**(1), 14 (2014). <https://doi.org/10.1186/s13174-014-0014-0>
 102. Kim, S., Won, J., Han, H., Eom, H., Yeom, H.Y.: Improving Hadoop performance in intercloud environments. *SIGMETRICS Perform. Eval. Rev.* **39**(3), 107–109 (2011). <https://doi.org/10.1145/2160803.2160873>
 103. Kivity, A., Kamay, Y., Laor, D., Lublin, U., Liguori, A.: KVM: the Linux virtual machine monitor. In: Proceedings of the Linux Symposium, pp. 225–230 (2007)
 104. Ko, S.Y., Jeon, K., Morales, R.: The HybrEx model for confidentiality and privacy in cloud computing. In: Proceedings of the 3rd USENIX Conference on Hot Topics in Cloud Computing, HotCloud'11, p. 8. USENIX Association, Berkeley (2011)
 105. Kondo, D., Javadi, B., Iosup, A., Epema, D.: The failure trace archive: enabling comparative analysis of failures in diverse distributed systems. In: 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 398–407 (2010). <https://doi.org/10.1109/CCGRID.2010.71>
 106. Koutsandria, G., Skevakis, E., Sayegh, A.A., Koutsakis, P.: Can everybody be happy in the cloud? Delay, profit and energy-efficient scheduling for cloud services. *J. Parallel Distrib. Comput.* **96**(Supplement C), 202–217 (2016). <https://doi.org/10.1016/j.jpdc.2016.05.013>
 107. Labba, C., Saoud, N.B.B., Dugdale, J.: A predictive approach for the efficient distribution of agent-based systems on a hybrid-cloud. *Future Gener. Comput. Syst.* **86**, 750–764 (2018). <https://doi.org/10.1016/j.future.2017.10.053>
 108. Lee, Y.C., Lian, B.: Cloud bursting scheduler for cost efficiency. In: 2017 IEEE 10th International Conference on Cloud Computing (CLOUD), pp. 774–777 (2017). <https://doi.org/10.1109/CLOUD.2017.112>
 109. Lee, Y.C., Zomaya, A.Y.: Rescheduling for reliable job completion with the support of clouds. *Future Gener. Comput. Syst.* **26**(8), 1192–1199 (2010). <https://doi.org/10.1016/j.future.2010.02.010>
 110. Leena, V.A., Ajeena Beegom, A.S., Rajasree, M.S.: Genetic algorithm based bi-objective task scheduling in hybrid cloud platform. *Int. J. Comput. Theory Eng.* **8**(1), 7–13 (2016). <https://doi.org/10.7763/IJCTE.2016.V8.1012>
 111. Leitner, P., Rostyslav, Z., Gambi, A., Dustdar, S.: A framework and Middleware for application-level cloud bursting on top of infrastructure-as-a-service clouds. In: Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing, UCC '13, pp. 163–170. IEEE Computer Society, Washington, DC (2013). <https://doi.org/10.1109/UCC.2013.39>
 112. Lent, R.: Evaluating the cooling and computing energy demand of a datacentre with optimal server provisioning. *Future Gener. Comput. Syst.* **57**, 1–12 (2016). <https://doi.org/10.1016/j.future.2015.10.008>
 113. Li, C., Li, L.: Efficient market strategy based optimal scheduling in hybrid cloud environments. *Wirel. Pers. Commun.* **83**(1), 581–602 (2015). <https://doi.org/10.1007/s11277-015-2410-6>
 114. Li, C., Li, L.: Hierarchical scheduling optimization scheme in hybrid cloud computing environments. *J. Circuits Syst. Comput.* **24**(08) (2015). <https://doi.org/10.1142/S021812661550111X>
 115. Li, C., Li, L.: Hybrid cloud scheduling method for cloud bursting. *Fundam. Inf.* **138**(4), 435–455 (2015). <https://doi.org/10.3233/FI-2015-1220>
 116. Li, C., Yan, X., Li, L.: Agents collaboration-based service provisioning strategy for large enterprise business in hybrid cloud. *Trans. Emerg. Telecommun. Technol.* **28**(3) (2017). <https://doi.org/10.1002/ett.2965>
 117. Li, C., Zhang, J., Chen, Y., Li, L.: Efficient QoS aware two-layer service allocation in hybrid mobile cloud. *Autom. Softw. Eng.* **25**(3), 569–593 (2018). <https://doi.org/10.1007/s10515-018-0233-x>
 118. Li, H., Zhong, L., Liu, J., Li, B., Xu, K.: Cost-effective partial migration of VoD services to content clouds. In: 2011 IEEE 4th International Conference on Cloud Computing, pp. 203–210 (2011). <https://doi.org/10.1109/CLOUD.2011.41>
 119. Li, S., Zhou, Y., Jiao, L., Yan, X., Wang, X., Lyu, M.T.: Towards operational cost minimization in hybrid clouds for dynamic resource provisioning with delay-aware optimization. *IEEE Trans. Serv. Comput.* **8**(3), 398–409 (2015)
 120. Li, Z., Kihl, M., Lu, Q., Andersson, J.A.: Performance overhead comparison between hypervisor and container based virtualization. In: 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), pp. 955–962 (2017). <https://doi.org/10.1109/AINA.2017.79>
 121. Lifka, D.A.: The ANL/IBM SP scheduling system. In: Job Scheduling Strategies for Parallel Processing: IPPS '95 Workshop Santa Barbara, CA, USA, April 25, 1995 Proceedings, pp. 295–303. Springer, Berlin (1995)
 122. Lijun, X., Chunlin, L.: Dynamic service provisioning and selection for satisfying cloud applications and cloud providers in hybrid cloud. *Int. J. Coop. Inf. Syst.* **26**(04), 1750005 (2017). <https://doi.org/10.1142/S0218843017500058>
 123. Lilienthal, M.: A decision support model for cloud bursting. *Bus. Inf. Syst. Eng.* **5**(2), 71–81 (2013). <https://doi.org/10.1007/s12599-013-0257-5>
 124. Lin, B., Guo, W., Lin, X.: Online optimization scheduling for scientific workflows with deadline constraint on hybrid clouds.

- Concurr. Comput. Pract. Exp. **28**(11), 3079–3095 (2016). <https://doi.org/10.1002/cpe.3582>
125. Liu, F., Luo, B., Niu, Y.: Cost-effective service provisioning for hybrid cloud applications. *Mob. Netw. Appl.* **22**(2), 153–160 (2017). <https://doi.org/10.1007/s11036-016-0738-0>
 126. Liu, Y., Li, C., Yang, Z., Chen, Y., Xu, L.: Research on cost-optimal algorithm of multi-QoS constraints for task scheduling in hybrid-cloud. *J. Softw. Eng.* 33–49 (2015)
 127. Liu, Z., Li, C., Wu, W., Jia, R.: A hierarchical approach for resource allocation in hybrid cloud environments. *Wirel. Netw.* (2016). <https://doi.org/10.1007/s11276-016-1416-7>
 128. Lu, P., Sun, Q., Wu, K., Zhu, Z.: Distributed online hybrid cloud management for profit-driven multimedia cloud computing. *IEEE Trans. Multimed.* **17**(8), 1297–1308 (2015)
 129. Luong, N.C., Wang, P., Niyato, D., Wen, Y., Han, Z.: Resource management in cloud networking using economic analysis and pricing models: a survey. *IEEE Commun. Surv. Tutor.* **19**(2), 954–1001 (2017). <https://doi.org/10.1109/COMST.2017.2647981>
 130. Madni, S.H.H., Latiff, M.S.A., Coulibaly, Y., Abdulhamid, S.M.: Resource scheduling for infrastructure as a service (IaaS) in cloud computing: challenges and opportunities. *J. Netw. Comput. Appl.* **68**(Supplement C), 173–200 (2016)
 131. Maheshwari, K., Jung, E.S., Meng, J., Morozov, V., Vishwanath, V., Kettimuthu, R.: Workflow performance improvement using model-based scheduling over multiple clusters and clouds. *Future Gener. Comput. Syst.* **54**, 206–218 (2016). <https://doi.org/10.1016/j.future.2015.03.017>
 132. Makris, P., Skoutas, D.N., Rizomiliotis, P., Skianis, C.: A user-oriented, customizable infrastructure sharing approach for hybrid cloud computing environments. In: 2011 IEEE Third International Conference on Cloud Computing Technology and Science, pp. 432–439 (2011). <https://doi.org/10.1109/CloudCom.2011.64>
 133. Malawski, M., Figiela, K., Nabrzyski, J.: Cost minimization for computational applications on hybrid cloud infrastructures. *Future Gener. Comput. Syst.* **29**(7), 1786–1794 (2013). <https://doi.org/10.1016/j.future.2013.01.004>
 134. Manikandan, M., Suguna, M.: A survey on temporal task scheduling for profit maximization in hybrid clouds. *Int. J. Innov. Adv. Comput. Sci.* **6**(1), 20–24 (2017)
 135. Mao, M., Humphrey, M.: A performance study on the VM startup time in the cloud. In: 2012 IEEE Fifth International Conference on Cloud Computing, pp. 423–430 (2012). <https://doi.org/10.1109/CLOUD.2012.103>
 136. Mao, X., Li, C., Yan, W., Du, S.: Optimal scheduling algorithm of MapReduce tasks based on QoS in the hybrid cloud. In: 2016 17th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), pp. 119–124 (2016). <https://doi.org/10.1109/PDCAT.2016.038>
 137. Marcu, O.C., Negru, C., Pop, F.: Dynamic scheduling in real time with budget constraints in hybrid clouds. In: Altmann, J., Silaghi, G.C., Rana, O.F. (eds.) *Economics of Grids, Clouds, Systems, and Services*, pp. 18–31. Springer, Cham (2016)
 138. Mattess, M., Calheiros, R.N., Buyya, R.: Scaling MapReduce applications across hybrid clouds to meet soft deadlines. In: 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), pp. 629–636 (2013). <https://doi.org/10.1109/AINA.2013.51>
 139. Mattess, M., Vecchiola, C., Buyya, R.: Managing peak loads by leasing cloud infrastructure services from a spot market. In: 2010 IEEE 12th International Conference on High Performance Computing and Communications (HPCC), pp. 180–188 (2010). <https://doi.org/10.1109/HPCC.2010.77>
 140. Mattess, M., Vecchiola, C., Garg, S., Buyya, R.: *Cloud Computing: Methodology, Systems, and Applications*. CRC Press, Boca Raton (2011)
 141. Mechtri, M., Hadji, M., Zeglache, D.: Exact and heuristic resource mapping algorithms for distributed and hybrid clouds. *IEEE Trans. Cloud Comput.* **5**(4), 681–696 (2017). <https://doi.org/10.1109/TCC.2015.2427192>
 142. Morla, R., Gonçalves, P., Barbosa, J.: A scheduler for cloud bursting of map-intensive traffic analysis jobs. In: *Proceedings of the Second International Workshop on Sustainable Ultrascale Computing Systems (NESUS 2015)*, pp. 11–21 (2015)
 143. Mu’alem, A.W., Feitelson, D.G.: Utilization, predictability, workloads, and user runtime estimates in scheduling the IBM SP2 with backfilling. *IEEE Trans. Parallel Distrib. Syst.* **12**(6), 529–543 (2001). <https://doi.org/10.1109/71.932708>
 144. Nachiappan, R., Javadi, B., Calheiros, R.N., Matawie, K.M.: Cloud storage reliability for big data applications: a state of the art survey. *J. Netw. Comput. Appl.* **97**, 35–47 (2017). <https://doi.org/10.1016/j.jnca.2017.08.011>
 145. Nahir, A., Orda, A., Raz, D.: Workload factoring with the cloud: a game-theoretic perspective. In: 2012 Proceedings IEEE INFOCOM, pp. 2566–2570 (2012). <https://doi.org/10.1109/INFOCOM.2012.6195654>
 146. Nguyen, T.L., Lebre, A.: Virtual machine boot time model. In: 2017 25th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), pp. 430–437 (2017). <https://doi.org/10.1109/PDP.2017.58>
 147. Niu, Y., Luo, B., Liu, F., Liu, J., Li, B.: When hybrid cloud meets flash crowd: towards cost-effective service provisioning. In: 2015 IEEE Conference on Computer Communications (INFOCOM), pp. 1044–1052 (2015). <https://doi.org/10.1109/INFOCOM.2015.7218477>
 148. Ogawa, Y., Hasegawa, G., Murata, M.: Cloud bursting approach based on predicting requests for business-critical web systems. In: 2017 International Conference on Computing, Networking and Communications (ICNC), pp. 437–441 (2017). <https://doi.org/10.1109/ICNC.2017.7876168>
 149. Oktay, K.Y., Khadilkar, V., Hore, B., Kantarcioglu, M., Mehrotra, S., Thuraisingham, B.: Risk-aware workload distribution in hybrid clouds. In: 2012 IEEE Fifth International Conference on Cloud Computing, pp. 229–236 (2012). <https://doi.org/10.1109/CLOUD.2012.128>
 150. Open Nebula: the open source toolkit for data center virtualization. <http://www.opennebula.org> (2018). Accessed 18 Jan 2020
 151. Openstack: open source software for creating private and public clouds. <http://www.openstack.org> (2018). Accessed 18 Jan 2020
 152. Padsar, A., Almi’ani, K., Lee, Y.C.: Data-aware scheduling of scientific workflows in hybrid clouds. In: Shi, Y., Fu, H., Tian, Y., Krzhizhanovskaya, V.V., Lees, M.H., Dongarra, J., Sloot, P.M.A. (eds.) *Computational Science—ICCS 2018*, pp. 708–714. Springer, Cham (2018)
 153. Peláez, V., Campos, A., García, D.F., Entrialgo, J.: Autonomic scheduling of deadline-constrained bag of tasks in hybrid clouds. In: 2016 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), pp. 1–8 (2016). <https://doi.org/10.1109/SPECTS.2016.7570526>
 154. Peláez, V., Campos, A., García, D.F., Entrialgo, J.: Online scheduling of deadline-constrained bag-of-task workloads on hybrid clouds. *Concurr. Comput. Pract. Exp.* e4639 (2018)
 155. Popović, K., Ž. Hocenski: cloud computing security issues and challenges. In: *The 33rd International Convention MIPRO*, pp. 344–349 (2010)
 156. Qiu, X., Li, H., Wu, C., Li, Z., Lau, F.C.M.: Dynamic scaling of VoD services into hybrid clouds with cost minimization and

- QoS guarantee. In: 2012 19th International Packet Video Workshop (PV), pp. 137–142 (2012). <https://doi.org/10.1109/PV.2012.6229726>
157. Qiu, X., Li, H., Wu, C., Li, Z., Lau, F.C.M.: Cost-minimizing dynamic migration of content distribution services into hybrid clouds. *IEEE Trans. Parallel Distrib. Syst.* **26**(12), 3330–3345 (2015). <https://doi.org/10.1109/TPDS.2014.2371831>
 158. Qiu, X., Yeow, W.L., Wu, C., Lau, F.C.M.: Cost-minimizing preemptive scheduling of mapreduce workloads on hybrid clouds. In: 2013 IEEE/ACM 21st International Symposium on Quality of Service (IWQoS), pp. 1–6 (2013). <https://doi.org/10.1109/IWQoS.2013.6550284>
 159. Quarati, A., Danovaro, E., Galizia, A., Clematis, A., D’Agostino, D., Parodi, A.: Scheduling strategies for enabling meteorological simulation on hybrid clouds. *J. Comput. Appl. Math.* **273**, 438–451 (2015). <https://doi.org/10.1016/j.cam.2014.05.001>
 160. Rahman, M., Li, X., Palit, H.: Hybrid heuristic for scheduling data analytics workflow applications in hybrid cloud environment. In: 2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and PhD Forum, pp. 966–974 (2011). <https://doi.org/10.1109/IPDPS.2011.243>
 161. Razavi, K., Kolk, G.V.D., Kielmann, T.: Prebaked μ VMs: scalable, instant VM startup for IaaS clouds. In: 2015 IEEE 35th International Conference on Distributed Computing Systems, pp. 245–255 (2015). <https://doi.org/10.1109/ICDCS.2015.33>
 162. Ruiz-Alvarez, A., Humphrey, M.: Toward optimal resource provisioning for cloud MapReduce and hybrid cloud applications. In: 2014 IEEE/ACM International Symposium on Big Data Computing, pp. 74–82 (2014). <https://doi.org/10.1109/BDC.2014.14>
 163. Ruiz-Alvarez, A., Kim, I.K., Humphrey, M.: Toward optimal resource provisioning for cloud mapreduce and hybrid cloud applications. In: 2015 IEEE 8th International Conference on Cloud Computing, pp. 669–677 (2015). <https://doi.org/10.1109/CLOUD.2015.94>
 164. Saber, T., Thorburn, J., Murphy, L., Ventresque, A.: VM reassignment in hybrid clouds for large decentralised companies: a multi-objective challenge. *Future Gener. Comput. Syst.* **79**, 751–764 (2018). <https://doi.org/10.1016/j.future.2017.06.015>
 165. Satyanarayanan, M., Simoens, P., Xiao, Y., Pillai, P., Chen, Z., Ha, K., Hu, W., Amos, B.: Edge analytics in the Internet of Things. *IEEE Pervasive Comput.* **14**(2), 24–31 (2015). <https://doi.org/10.1109/MPRV.2015.32>
 166. Sharif, S., Taheri, J., Zomaya, A.Y., Nepal, S.: MPHc: preserving privacy for workflow execution in hybrid clouds. In: 2013 International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 272–280 (2013). <https://doi.org/10.1109/PDCAT.2013.49>
 167. Sharif, S., Taheri, J., Zomaya, A.Y., Nepal, S.: Online multiple workflow scheduling under privacy and deadline in hybrid cloud environment. In: 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, pp. 455–462 (2014). <https://doi.org/10.1109/CloudCom.2014.128>
 168. Sharma, Y., Javadi, B., Si, W., Sun, D.: Reliability and energy efficiency in cloud computing systems: survey and taxonomy. *J. Netw. Comput. Appl.* **74**, 66–85 (2016). <https://doi.org/10.1016/j.jnca.2016.08.010>
 169. Shifrin, M., Atar, R., Cidon, I.: Optimal scheduling in the hybrid-cloud. In: 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), pp. 51–59 (2013)
 170. Siddiqui, U., Tahir, G.A., Rehman, A.U., Ali, Z., Rasool, R.U., Bloodsworth, P.: Elastic JADE: dynamically scalable multi agents using cloud resources. In: 2012 Second International Conference on Cloud and Green Computing, pp. 167–172 (2012). <https://doi.org/10.1109/CGC.2012.60>
 171. Speitkamp, B., Bichler, M.: A mathematical programming approach for server consolidation problems in virtualized data centers. *IEEE Trans. Serv. Comput.* **3**(4), 266–278 (2010). <https://doi.org/10.1109/TSC.2010.25>
 172. Srinivasan, S., Kettimuthu, R., Subramani, V., Sadayappan, P.: Selective reservation strategies for backfill job scheduling. In: Job Scheduling Strategies for Parallel Processing: 8th International Workshop, JSSPP 2002 Edinburgh, Scotland, UK, July 24, 2002 Revised Papers, pp. 55–71. Springer, Berlin (2002)
 173. Sukumar, K., Vecchiola, C., Buyya, R.: The structure of the new IT frontier: Aneka platform for elastic cloud computing applications. *Strateg. Facil. Mag.* **25**(6), 599–616 (2010)
 174. Taheri, J., Zomaya, A.Y., Siegel, H.J., Tari, Z.: Pareto frontier for job execution and data transfer time in hybrid clouds. *Future Gener. Comput. Syst.* **37**, 321–334 (2014)
 175. Tian, W., Zhao, Y.: *Optimized Cloud Resource Management and Scheduling: Theories and Practices*, 1st edn. Morgan Kaufmann Publishers Inc., San Francisco (2014)
 176. Toosi, A.N., Sinnott, R.O., Buyya, R.: Resource provisioning for data-intensive applications with deadline constraints on hybrid clouds using Aneka. *Future Gener. Comput. Syst.* **79**, 765–775 (2018). <https://doi.org/10.1016/j.future.2017.05.042>
 177. Unuvar, M., Steinder, M., Tantawi, A.N.: Hybrid cloud placement algorithm. In: 2014 IEEE 22nd International Symposium on Modelling, Analysis Simulation of Computer and Telecommunication Systems, pp. 197–206 (2014). <https://doi.org/10.1109/MASCOTS.2014.33>
 178. Vecchiola, C., Calheiros, R.N., Karunamoorthy, D., Buyya, R.: Deadline-driven provisioning of resources for scientific applications in hybrid clouds with Aneka. *Future Gener. Comput. Syst.* **28**(1), 58–65 (2012)
 179. Verma, A., Ahuja, P., Neogi, A.: pMapper: power and migration cost aware application placement in virtualized systems. In: Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware, Middleware ’08, pp. 243–264. Springer, New York (2008)
 180. Verma, A., Pedrosa, L., Korupolu, M., Oppenheimer, D., Tune, E., Wilkes, J.: Large-scale cluster management at Google with Borg. In: Proceedings of the Tenth European Conference on Computer Systems, EuroSys ’15, pp. 18:1–18:17. ACM, New York (2015). <https://doi.org/10.1145/2741948.2741964>
 181. Vilutis, G., Daugirdas, L., Kavaliūnas, R., Štutienė, K., Vaidelys, M.: Model of load balancing and scheduling in Cloud computing. In: Proceedings of the ITI 2012 34th International Conference on Information Technology Interfaces, pp. 117–122 (2012). <https://doi.org/10.2498/iti.2012.0460>
 182. VMware: public and hybrid cloud computing. <http://www.vmware.com/> (2018). Accessed 18 Jan 2020
 183. Vogels, W.: Beyond server consolidation. *Queue* **6**(1), 20–26 (2008). <https://doi.org/10.1145/1348583.1348590>
 184. Wang, B., Song, Y., Cui, X., Cao, J.: Mathematical programming for server consolidation in cloud data centers. In: 2017 4th International Conference on Systems and Informatics (ICSAI), pp. 678–683 (2017). <https://doi.org/10.1109/ICSAI.2017.8248374>
 185. Wang, B., Song, Y., Cui, X., Cao, J.: Performance comparison between hypervisor- and container-based virtualizations for cloud users. In: 2017 4th International Conference on Systems and Informatics (ICSAI), pp. 684–689 (2017). <https://doi.org/10.1109/ICSAI.2017.8248375>
 186. Wang, B., Song, Y., Sun, Y., Liu, J.: Managing deadline-constrained bag-of-tasks jobs on hybrid clouds. In: Proceedings of the 24th High Performance Computing Symposium, HPC ’16, pp. 22:1–22:8. Society for Computer Simulation International, San Diego (2016). <https://doi.org/10.22360/SpringSim.2016.HPC.039>

187. Wang, B., Song, Y., Sun, Y., Liu, J.: Managing deadline-constrained bag-of-tasks jobs on hybrid clouds with closest deadline first scheduling. *KSIIT Trans. Internet Inf. Syst.* **10**(7), 2952–2971 (2016). <https://doi.org/10.3837/tiis.2016.07.005>
188. Wang, B., Song, Y., Sun, Y., Liu, J.: Analysis model for server consolidation of virtualized heterogeneous data centers providing internet services. *Clust. Comput.* **22**(3), 911–928 (2019). <https://doi.org/10.1007/s10586-018-2880-x>
189. Wang, W.J., Chang, Y.S., Lo, W.T., Lee, Y.K.: Adaptive scheduling for parallel tasks with QoS satisfaction for hybrid cloud environments. *J. Supercomput.* **66**(2), 783–811 (2013)
190. Web of Science. <http://apps.webofknowledge.com/> (2018). Accessed 18 Jan 2020
191. Wei, Y., Sukumar, K., Vecchiola, C., Karunamoorthy, D., Buyya, R.: Aneka Cloud Application Platform and Its Integration with Windows Azure. *CoRR arXiv:abs/1103.2590* (2011)
192. Wu, H., Ren, S., Garzoglio, G., Timm, S., Bernabeu, G., Kimy, H., Chadwick, K., Jang, H., Noh, S.Y.: Automatic cloud bursting under FermiCloud. In: 2013 International Conference on Parallel and Distributed Systems (ICPADS), pp. 681–686 (2013)
193. Wu, X., Gu, Y., Li, G.: Game analysis of workload factoring with the hybrid cloud. In: 2013 First International Symposium on Computing and Networking, pp. 263–269 (2013). <https://doi.org/10.1109/CANDAR.2013.46>
194. Xie, H., Song, X., Bi, J., Yuan, H.: VCG auction based idle instance bidding to increase IaaS provider's profit in hybrid clouds. In: Mohamed Ali, M.S., Wahid, H., Mohd Subha, N.A., Sahlan, S., Md. Yunus, M.A., Wahap, A.R. (eds.) *Modeling, Design and Simulation of Systems*, pp. 359–368. Springer, Singapore (2017)
195. Xu, F., Liu, F., Jin, H.: Heterogeneity and interference-aware virtual machine provisioning for predictable performance in the cloud. *IEEE Trans. Comput.* **65**(8), 2470–2483 (2016)
196. Yu, J., Buyya, R., Tham, C.K.: Cost-based scheduling of scientific workflow applications on utility grids. In: First International Conference on e-Science and Grid Computing (e-Science'05), pp. 140–147 (2005)
197. Yuan, H., Bi, J., Tan, W., Li, B.H.: Temporal task scheduling with constrained service delay for profit maximization in hybrid clouds. *IEEE Trans. Autom. Sci. Eng.* **14**(1), 337–348 (2017). <https://doi.org/10.1109/TASE.2016.2526781>
198. Yuan, H., Bi, J., Tan, W., Zhou, M., Li, B.H., Li, J.: TTSA: an effective scheduling approach for delay bounded tasks in hybrid clouds. *IEEE Trans. Cybern.* **47**(11), 3658–3668 (2017)
199. Yuan, X., Weng, J., Wang, C., Ren, K.: Secure integrated circuit design via hybrid cloud. *IEEE Trans. Parallel Distrib. Syst.* **29**(8), 1851–1864 (2018). <https://doi.org/10.1109/TPDS.2018.2807844>
200. Zakarya, M., Gillam, L.: Energy efficient computing, clusters, grids and clouds: a taxonomy and survey. *Sustain. Comput. Inform. Syst.* **14**, 13–33 (2017). <https://doi.org/10.1016/j.suscom.2017.03.002>
201. Zhang, G., Zuo, X.: Deadline constrained task scheduling based on standard-PSO in a hybrid cloud. In: Tan, Y., Shi, Y., Mo, H. (eds.) *Advances in Swarm Intelligence. ICSI 2013*, pp. 200–209. Springer, Berlin (2013)
202. Zhang, H., Jiang, G., Yoshihira, K., Chen, H.: Proactive workload management in hbrid cloud computing. *IEEE Trans. Netw. Serv. Manag.* **11**(1), 90–100 (2014). <https://doi.org/10.1109/TNSM.2013.122313.130448>
203. Zhang, H., Jiang, G., Yoshihira, K., Chen, H., Saxena, A.: Intelligent workload factoring for a hybrid cloud computing model. In: 2009 Congress on Services—I, pp. 701–708 (2009). <https://doi.org/10.1109/SERVICES-I.2009.26>
204. Zhang, H., Jiang, G., Yoshihira, K., Chen, H., Saxena, A.: Resilient workload manager: taming bursty workload of scaling internet applications. In: Proceedings of the 6th International Conference Industry Session on Autonomic Computing and Communications Industry Session, ICAC-INDST '09, pp. 19–28. ACM, New York (2009). <https://doi.org/10.1145/1555312.1555318>
205. Zhang, P., Lin, C., Li, W., Ma, X.: Long-term multi-objective task scheduling with Diff-Serv in hybrid clouds. In: Bouguet-taya, A., Gao, Y., Klimenko, A., Chen, L., Zhang, X., Dzerzhinskiy, F., Jia, W., Klimenko, S.V., Li, Q. (eds.) *Web Information Systems Engineering—WISE 2017*, pp. 243–258. Springer, Cham (2017)
206. Zhang, Q., Chen, H., Shen, Y., Ma, S., Lu, H.: Optimization of virtual resource management for cloud applications to cope with traffic burst. *Future Gener. Comput. Syst.* **58**, 42–55 (2016). <https://doi.org/10.1016/j.future.2015.12.011>
207. Zhang, X., Tune, E., Haggmann, R., Jnagal, R., Gokhale, V., Wilkes, J.: CPI2: CPU performance isolation for shared compute clusters. In: Proceedings of the 8th ACM European Conference on Computer Systems, EuroSys '13, pp. 379–391. ACM, New York (2013). <https://doi.org/10.1145/2465351.2465388>
208. Zhang, Y., Sun, J.: Novel efficient particle swarm optimization algorithms for solving QoS-demanded bag-of-tasks scheduling problems with profit maximization on hybrid clouds. *Concurr. Comput. Pract. Exp.* **29**(21), e4249:1–e4249:19 (2017). <https://doi.org/10.1002/cpe.4249>
209. Zhang, Y., Sun, J., Wu, Z.: An heuristic for bag-of-tasks scheduling problems with resource demands and budget constraints to minimize makespan on hybrid clouds. In: 2017 Fifth International Conference on Advanced Cloud and Big Data (CBD), pp. 39–44 (2017). <https://doi.org/10.1109/CBD.2017.15>
210. Zhang, Y., Sun, J., Wu, Z., Xie, S., Xu, R.: Scheduling parallel intrusion detecting applications on hybrid clouds. *Secur. Commun. Netw.* **2018**, 2863793:1–2863793:12 (2018). <https://doi.org/10.1155/2018/2863793>
211. Zhang, Y., Sun, J., Zhu, J.: An effective heuristic for due-date-constrained bag-of-tasks scheduling problem for total cost minimization on hybrid clouds. In: 2016 International Conference on Progress in Informatics and Computing (PIC), pp. 479–486 (2016). <https://doi.org/10.1109/PIC.2016.7949548>
212. Zhou, B., Zhang, F., Wu, J., Liu, Z.: Cost reduction in hybrid clouds for enterprise computing. In: 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 270–274 (2017). <https://doi.org/10.1109/ICDCSW.2017.13>
213. Zhu, J., Li, X., Ruiz, R., Xu, X.: Scheduling stochastic multi-stage jobs to elastic hybrid cloud resources. *IEEE Trans. Parallel Distrib. Syst.* **29**(6), 1401–1415 (2018). <https://doi.org/10.1109/TPDS.2018.2793254>
214. Zhu, J., Li, X., Ruiz, R., Xu, X., Zhang, Y.: Scheduling stochastic multi-stage jobs on elastic computing services in hybrid clouds. In: 2016 IEEE International Conference on Web Services (ICWS), pp. 678–681 (2016). <https://doi.org/10.1109/ICWS.2016.94>
215. Zinnen, A., Engel, T.: Deadline constrained scheduling in hybrid clouds with Gaussian processes. In: 2011 International Conference on High Performance Computing Simulation, pp. 294–300 (2011). <https://doi.org/10.1109/HPCSim.2011.5999837>
216. Zuo, L., Shu, L., Dong, S., Chen, Y., Yan, L.: A multi-objective hybrid cloud resource scheduling method based on deadline and cost constraints. *IEEE Access* **5**, 22067–22080 (2017). <https://doi.org/10.1109/ACCESS.2016.2633288>

217. Zuo, X., Zhang, G., Tan, W.: Self-adaptive learning PSO-based deadline constrained task scheduling for hybrid IaaS cloud. *IEEE Trans. Autom. Sci. Eng.* **11**(2), 564–573 (2014). <https://doi.org/10.1109/TASE.2013.2272758>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Bo Wang is currently a Lecturer at Software Engineering College, Zhengzhou University of Light Industry. He received the PhD Degree in Computer Science from Xi'an Jiaotong University (XJTU), Xian, China, in 2017, and the BS Degree in Computer Science from Northeast Forest University (NEFU), Harbin, China, in 2010. He was a Guest Student in State Key Laboratory of Computer Architecture, Institute of Computing Technology (ICT),

Chinese Academy of Sciences (CAS) during 2012 to 2016. His research interests include distributed systems, cloud computing, edge computing, resource management and big data computing platform. He has published more than 10 publications in these areas recent years, and served in various academic journals and conferences.



Changhai Wang was born in Liaocheng, Shandong, China, in 1987. He received the MS Degrees in Computer Science and technology from Jiangsu University in 2012 and the PhD Degree from Nankai University in 2016. He has been with the Zhengzhou University of Light Industry since 2017. In the last 5 years, he has published nine papers related to activity recognition. His research interests include activity recognition, wearable computing, and

mobile computing. He received the Best Paper Award at the 10th IEEE Asia-Pacific Services Computing Conference and the Excellent Paper Award at the Journal of Communication in 2016.



Ying Song received the PhD Degree in Computer Engineering from the Institute of Computing Technology (ICT), Chinese Academy of Sciences. She is an Associate Professor at the Computer School, Beijing Information Science and Technology University. Her work thus far has covered topics such as performance modeling, resource management, cloud computing, and big data computing platform. Her main research interests include com-

puter architecture, parallel and distributed computing, and

virtualization technology. She has authored or coauthored more than 30 publications in these areas since 2007, and she served in various academic conferences.



Jie Cao was born in China in 1980. He received the PhD Degree in Computer Software and Theory from Tongji University, Shanghai, in 2015. He received his MS Degree in Computer Application Technology from Shanghai Maritime University, Shanghai, in 2010. She received his BS Degree in Mathematics and Applied Mathematics from Anyang Normal University, Henan, China, in 2004. He is now a Lecturer of Software College at

Zhengzhou University of Light Industry. He is a Member of CCF. His major research contains parallel computing, Grid and Cloud computing, and reliability.



Xiao Cui is currently an Associate Professor in the Software Engineering College, Zhengzhou University of Light Industry. He received his BE Degree from the School of Physical Sciences, University of Science and Technology of China (USTC) in 1997 and obtained his PhD Degree from the Key Lab of Wide Band-Gap Semiconductor Materials and Devices, School of Micro-Electronics Xidian University (XDU) in 2014. Later, he

worked as an Associate Professor in the Software Engineering College, Zhengzhou University of Light Industry (2014–2018). His research interests include biosensor and machine learning.



Ling Zhang received the PhD from the School of Computer Science and Engineering, Beijing University of Posts and Telecommunications in 2014. She is a Teacher of Software Engineering College of Zhengzhou University of Light Industry, the mainly researches are network security, artificial intelligence and cloud computing.