Journal of Network and Computer Applications 102 (2018) 17–37

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Analytics in/for cloud-an interdependence: A review

Kumar Ravi[a,b], Yogesh Khandelwal[a], Boora Shiva Krishna[a], Vadlamani Ravi[a,*]

[a] Center of Excellence in Analytics, Institute for Development and Research in Banking Technology, Castle Hills Road No. 1, Masab Tank, Hyderabad 500057, India

[b] School of Computer & Information Sciences, University of Hyderabad, Hyderabad 500046, India

ARTICLE INFO

Keywords:
Analytics in cloud
Analytics for cloud
Cloud resource demand prediction
Cloud resource usage optimization
Cloud provisioning

ABSTRACT

Cloud computing has brought a paradigmatic shift in providing data storage as well as computing resources. With the ever-increasing demand for cloud computing, the number of cloud providers is also increasing evidently, which poses challenges as well as opportunities for consumers and providers. From a consumer point of view, efficient selection of cloud resources at a minimum cost is a big challenge. On the other hand, a provider has to meet consumers' requirements with sufficient profit in the fiercely competitive market. The relationship between cloud computing is truly symbiotic in the sense that cloud computing makes the practice of analytics more pervasive while analytics makes cloud computing more efficient and optimal in a lot of ways. In addressing these issues, analytics plays an important role. In this paper, we reviewed some important research articles, which focus on cloud computing from the viewpoint of analytics. Analytics and cloud computing are found to be quite interdependent. From analytics perspective, cloud computing makes available high-end computing resources even to an individual customer at an affordable price. We call this thread "Analytics in Cloud". From the point of view of cloud computing, efficient management, allocation, and demand prediction can be performed using analytics. We call this thread "Analytics for Cloud". This review paper is mainly based on these two threads of thought process. In this regard, we reviewed eighty-eight research articles published during 2003–2017 related to the formidable duo of cloud computing and analytics.

1. Introduction

Data-driven analytics and decision support, while harnessing the cloud computing and big data, opens up unprecedented opportunities for revolutionizing different industries that use machine learning, data mining, and optimization. Data glut pervasive in many disciplines manifests in both structured and unstructured forms. Analysis of available data helps in decision-making and non-obvious knowledge generation. The advancement in data storage, data processing, and data mining technologies help in making data-driven decisions for any organization (Hashem et al., 2015).

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources e.g. networks, servers, storage, applications, and services. These computing services can be easily provisioned and released with minimal management effort or service provider interaction. The cloud service models are mainly available in three forms viz. Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS). Deployment of these three models is possible in four forms namely *private cloud, community cloud, public cloud,* and *hybrid cloud*. A service model of different forms should have five essential characteristics viz. *on-demand self-service, broad network access, resource pooling, rapid elasticity, measured service,* and *resource usage* (Mahmud et al., 2016).

Cloud computing has become a powerful architecture to perform large-scale and complex computing. Some of the advantages of cloud computing can be listed as *virtualized resources, parallel processing, security,* and *scalable data storage*. Cloud computing helps an organization in minimization of operational cost by restricting organization's owned software and hardware. The requirements of computing resources are fulfilled by ad-hoc procurement from service providers, which further helps in efficient management of users' dynamic demand. The dynamic allocation of resources is performed using virtualization. Virtualization is the technology behind the implementation of cloud computing. Virtualization can be defined as a process of sharing resources and separation of underlying hardware so that computer resources can be utilized efficiently and in a scalable manner (Mahmud et al., 2016).

* Corresponding author.
E-mail addresses: ankitaravi.ravi00@gmail.com (K. Ravi), ykhandelwal1612@gmail.com (Y. Khandelwal), shivakrishnaboora@gmail.com (B.S. Krishna), padmarav@gmail.com (V. Ravi).

https://doi.org/10.1016/j.jnca.2017.11.006
Received 2 August 2017; Received in revised form 15 October 2017; Accepted 14 November 2017
Available online 16 November 2017
1084-8045/

As the demand for cloud computing is increasing day-by-day, the cloud providers must scale the technologies in order to survive in the fierce competition. The overarching theme in 2016 was hybrid cloud adoption (Weins, 2016). It is growing exponentially as the cloud providers and clients are leveraging cloud computing at a fast pace. In order to meet clients' varying and dynamic requirements, the provider needs to predict the demand and cater the clients' needs accordingly. For prediction purpose, analytics is proved quite useful. Furthermore, the latest organizational demands like social media analytics and big data analytics are also computing intensive operations, which can be immensely benefitted from cloud computing too.

On an elaboration of future demand area of cloud computing, Forbes[1] published an article entitled "5 Cloud Service Provider Predictions For 2016: Analytics, CRM, & CPQ (Configure Price Quote) Accelerate Sales" (Columbus, 2016). It reported that there would be much greater sales of cloud services in the area of analytics, Customer Relationship Management (CRM), health care, etc. in the forthcoming years. Out of them, CRM is quite necessary and useful in Banking, Financial Services, and Insurance (BFSI) sectors. These different kinds of services will be dependent on historical data and hence can be utilized for forecasting different requirements. The report emphasized on five points: (i) there is a great demand for cloud-based analytics, CRM and CPQ in future, (ii) cloud integration expertise turns into a margin multiplier for service providers in 2016, (iii) cloud-based analytics will move beyond dashboards to cognitive intelligence, which can drive recommender systems this year, (iv) service providers with expertise tailoring analytics, CRM, and CPQ apps to unique business requirements will grow so fast this year, and (v) the marketing and selling analytics adoption are already accelerating in 2016 driven by widespread cloud adoption, hence CMO (Chief Marketing Officers) need to quantify marketing performance and analyse sales cycles for better insights.

As the number of customers and competitors started increasing, the data related to the customers accessing the cloud services and sessions is exploding. In order to handle such huge and ever-increasing data, the cloud providers can employ Big Data Analytics (BDA). However, there exist many complexities with respect to tools and infrastructure available for dealing with Big Data. Lack of suitable human resources further increases the difficulty to employ BDA.

Cloud is about lowering the cost of ownership for enterprises by delivering value to them. Cloud brings this by glorifying "As-a-Service" model (Narayanan, 2013). This "As-a-Service" model is very much needed for their abstraction of complexity and problems in scalability and elasticity of self-service application. In order to handle Big Data, Hadoop[2] addresses the challenges of scaling horizontally like large-scale distributed processing. The Hadoop provides the large-scale distributed computing without letting an end-user to worry about scalability, node failure, and fault tolerance. "As-A-Service" (AAS) model hinders the complexity in building a scalable elastic self-service application. For many organizations, a variety of technologies is implemented either on-premise or conventional cloud services, which may not be effective in terms of cost and response time. Rather than implementing complex solutions on-premise, organizations can use the services provided by different cloud providers. Cloud services can be provided using different price models, which will be helpful to adapt scalability, abstraction from various issues, and the dynamic needs. Therefore, the simplification that is provided by Cloud and Hadoop for Big data is the main reason behind migration towards Big data and Cloud.

"Analytics as a Service" is a buzz word nowadays, which refers to providing on-demand Analytics services to any organization. The simplification can be enabled for companies through analysing their earlier problems faced and customising their needs. The historical data of the company can be analysed by employing various data mining tasks. In order to manage huge computing resources, analytics has been employed in several studies. Cloud computing is needed to handle ever-growing big data at enterprise-level. Therefore, we realize that analytics and cloud computing are interdependent, which needs to be dealt together in many cases.

In this regard, we present a first-of-its-kind survey paper, which focuses broadly on two major interrelated topics viz. "Analytics in Cloud" and "Analytics for Cloud". Here, the former one represents "analytics that can be placed in the cloud environment to derive the well-known advantages of cloud computing paradigm", whereas the latter one represents "analytics that are performed for improving the quality of cloud services". Hence, the goal of this study is to review the work done so far when analytics and cloud are merged together for improving the efficiency of either one in a symbiotic manner. We covered major aspects of analytics and cloud computing as well as the relationship between them. Further, some research challenges are also discussed focusing on 'analytics in cloud' and 'analytics for cloud' perspectives.

The rest of the paper is organized as follows: Section 2 discusses motivation and review methodology for this survey paper. We presented earlier reviews related to cloud and analytics used together in Section 3. Section 4 reviews articles regarding "analytics in cloud". In Section 5, a reviews studies related to different aspects of "analytics for cloud" is presented. Section 6 presents discussion along with some open issues. Section 7 concludes the paper with future directions.

2. Motivation and review methodology

In order to gain insights from historical data, data analytics is quite useful for better decision making. In the last few years, the organizations from various fields are employing data mining, machine learning, and cloud computing, etc. for analysing and storing their data, which is in turn used for predictive analytics purpose. Storing data on cloud enables researchers and practitioners for using data in a cost-efficient and timely manner in order to get better and informed decision making. In this paper, we demonstrated how cloud computing and analytics are interdependent on each other. In this regard, we reviewed 88 papers regarding two threads of thought viz. *Analytics in Cloud* and *Analytics for Cloud* as depicted in Fig. 1. The former one is addressing how analytics can be performed in a cloud environment to leverage the availability of computing and storage resources. The latter one describes that how analytics can be employed for efficient management of computing and storing resources for the cloud environment.

In this study, we covered papers related to cloud computing, data analytics, integration of cloud computing and data analytics. We included only those papers which addressed the problems related to

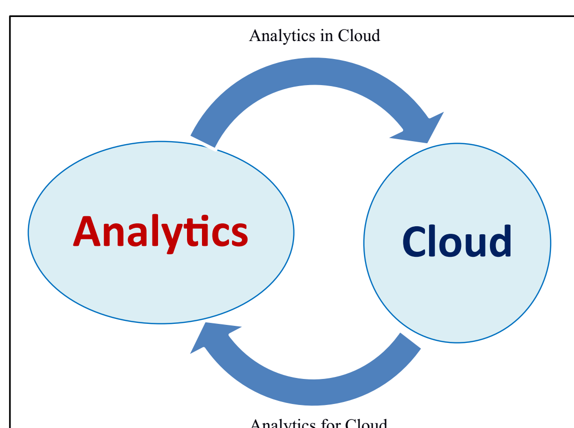

Fig. 1. Symbiotic relation between analytics and cloud.

[1] www.forbes.com.
[2] hadoop.apache.org.

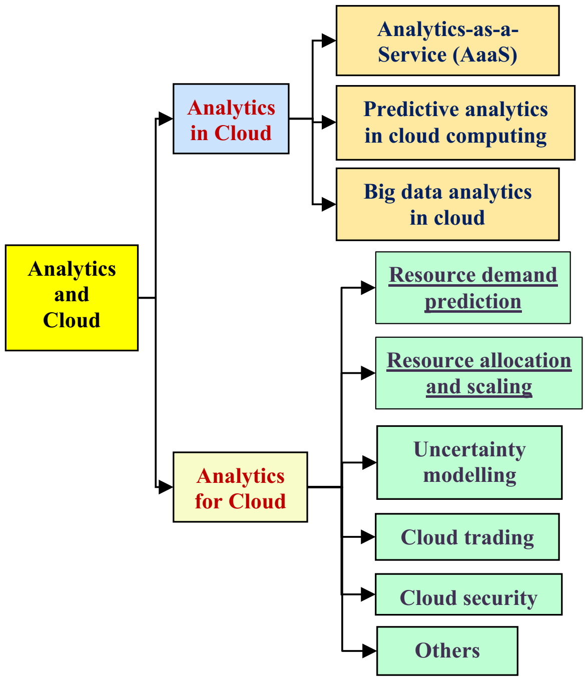

Fig. 2. Theme of the review depicting interdependence between Cloud and Analytics.

analytics in cloud computing and analytics for cloud computing. In this regard, we reviewed 88 most relevant research articles with respect to above-mentioned two threads of thought. For *analytics in cloud* part, we reviewed only those articles, which clearly mentioned that their experiments were carried out in a cloud environment. Regarding *analytics for cloud* part, we identified some key motivators for our review like functional problems in cloud computing, quality-of-service, cost optimization, management of resources, security, trading, etc. On the basis of the identified key motivators, we grouped the reviewed articles under six categories: *(i) resource demand prediction, (ii) resource allocation and scaling, (iii) Cloud trading, (iv) Cloud security, (v) uncertainty modelling,* and *(vi) others*. In our opinion, this categorization succinctly captures all the aspects of cloud computing operations, where analytics plays a quintessential role. In order to collect research articles related to both aspects, we searched relevant papers in some reputed journals, book chapters, and conferences. We used different sources like Science Direct, Google Scholar, IEEE Explore, and ACM digital library for our search requirements. The theme of the current review is depicted in Fig. 2.

3. Earlier reviews

Masdari et al. (2017) reviewed various forms of PSO to be used for optimizing different functions and resources of cloud computing. Reviewed forms of PSO include multi-objective PSO, learning PSO, jumping PSO, bi-objective PSO, and hybrid PSO. The functions and resources of cloud include *minimizing task execution time, minimizing task transferring time, minimizing task execution cost, minimizing makespan, meeting the service-level agreement (SLA) and guaranteeing user-level Quality of Service (QoS) in task execution, increasing elasticity, raising the availability of resources, load balancing,* and *decreasing energy consumption*. Ghomi et al. (2017) surveyed a list of load-balancing algorithms in cloud computing. They considered different aspects of load balancing to organize one hundred and four articles. The major aspects considered for the survey are: *load balancing schedulers in Hadoop, MapReduce optimization in Hadoop, natural-phenomena-based load balancing, agent-based load balancing, general load balancing, application-oriented load balancing, network-aware task scheduling and load balancing,* and *workflow specific scheduling*. Amiri and Mohammad-Khanli (2017) presented a detailed literature survey on prediction models used for application-oriented resources provisioning. They listed application-oriented attributes such as performance (throughput and response time), SLA parameters, and workload (the number of requests, resource demand, and resource utilization). They considered some of the functions and measures of resource provisioning comprising *the number of requests/jobs, resources utilization, execution time of jobs, SLA parameters, future demand of resources, the number of users, the number of requests of data objects, performance prediction, power consumption, the number of virtual machine (VM),* and *the number of PMs* to organize the whole review. Yang et al. (2017) presented a taxonomy about cloud computing and big data. They further reviewed some existing literature regarding *on-demand resource provision, scheduling, scalability, data locality, cloud computing for social media and other streamed data, quality of service, cloud computing benchmark and adoption, diversity and interoperability, hybrid computing infrastructure* and *virtual resource bursting*. Masdari et al. (2017) did not consider other architectures of soft computing and machine learning methods to be used for cloud and analytics. Ghomi et al. (2017) confined their study to load balancing. Amiri and Mohammad-Khanli (2017) considered to review a list of studies related to resource provisioning only. Finally, Yang et al. (2017) was mainly concerned with preparing big data taxonomy. Moreover, they considered very few literature related to abovementioned categories.

Therefore, we observe that the present survey is distinct in that it is an overarching and broad one encompassing all the operational, business and scientific aspects of their symbiotic relationship.

4. Analytics in cloud

In this section, we focused on “How data analytics can be performed using cloud computing?”, “What are the technologies available?”, “How big data analytics can be implemented in the cloud?”, and “What challenges companies are facing to exploit cloud for analytics purpose?”. In order to make this paper self-contained, we presented a brief introduction to cloud computing and analytics in Section 3.1 and Section 3.2 respectively. The remaining three sections are about *Analytics-as-a-Service* (AaaS), *predictive analytics* in *cloud computing*, and *big data analytics in cloud computing* as depicted in Fig. 2.

4.1. Overview of cloud computing

Cloud computing is a type of Internet-based computing paradigm that provides shared computer processing, resources, and data, which can be utilized by a number of customers on different computing devices on demand basis. It is a model for enabling ubiquitous and on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications, services, etc.) (Mell and Grance, 2011; Neto, 2011). Following are some of the characteristics of cloud computing:

- Agility: Cloud computing increases user's flexibility for growing, re-provisioning, and expanding infrastructure resources.
- Cost: Reduces the cost of infrastructure and services as well as some other costs.
- Multi-tenancy: Cloud Computing enables resource sharing to a large pool of users.
- Scalability: Provide ability to user to scale up or down
- Productivity: Increases productivity

Fig. 3 depicts a deployment model, describing attributes, organization of infrastructure, resources and delivery models of cloud computing. Different types of service models include IaaS, PaaS, SaaS, etc., whereas service-oriented architecture advocates “everything as a service” (with acronyms EaaS or XaaS or simply AAS) (Weins, 2016).

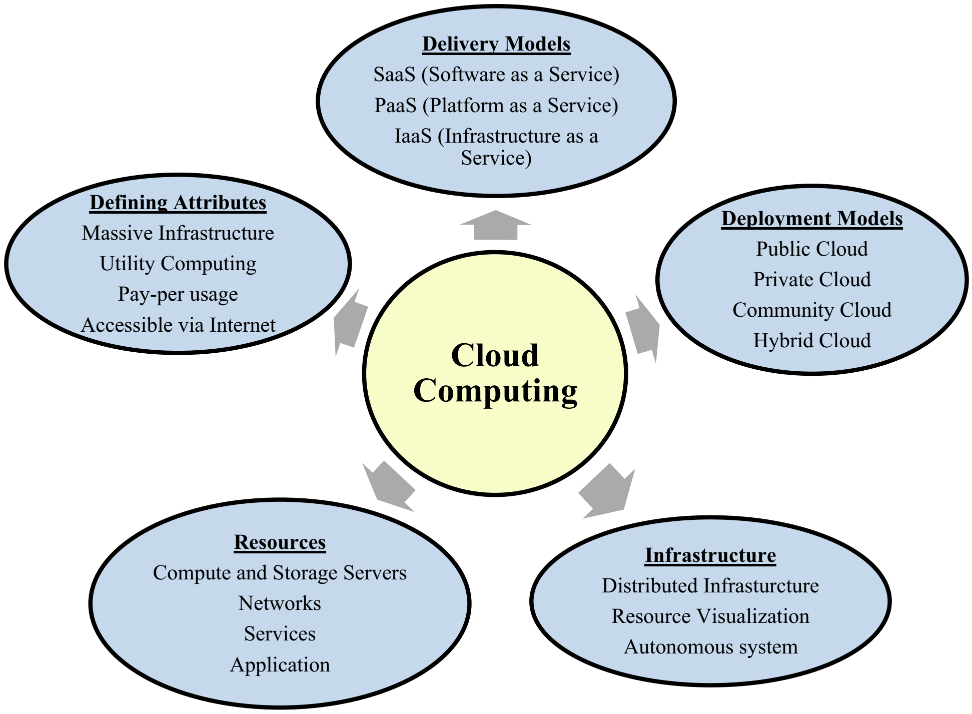

Fig. 3. Cloud Computing paradigm (Kurdi et al., 2010).

Cloud computing models are of three types: public, private, community or hybrid cloud.

4.2. Overview of analytics

Data analytics or simply analytics is the process of inspecting, cleaning, processing and modelling data with the aim of gaining useful patterns, insights, and conclusions that support decision-making ("Data analysis," n.d.). Data analytics tasks can be divided into following three categories:

4.2.1. Descriptive analytics

Descriptive analytics is the phenomenon through different measures. The primary purpose of descriptive analytics is to answer the questions like "what happened earlier?" (Banerjee et al., 2013). It uses exploratory data analytics tools on existing data like visualization tools for depicting "what is happening?". Some people use the word diagnostic analytics to focus on questions such as "why something happened?" (Banerjee et al., 2013). However, it is also a part of descriptive analytics. It uses subjects like descriptive statistics, online analytical processing (OLAP), etc.

4.2.2. Predictive analytics

Predictive analytics answers some questions like "What is likely to happen?". Predictive analytics uses the concepts from data mining subsuming predictive statistics and machine learning. Some examples of the predictive analytics include: predicting the propensity of the customer to buy a product, predicting customer churn, default prediction, fraud detection and so on (Banerjee et al., 2013).

4.2.3. Prescriptive analytics

Prescriptive analytics goes beyond describing, explaining and predicting. It mainly focuses on "what should be done about it?" or "what course of action can be taken?" to optimize the whole business process and achieve the business objectives efficiently (Banerjee et al., 2013). It uses all kinds of optimization techniques, visualizations, etc.

4.3. Analytics-as-a-service (AaaS)

Almost every organization uses analytics or data based applications for business planning, problem-solving or decision support system in the form of dashboards or visuals. According to Forrester Research, business intelligence and analytics will trend as self-service, pervasive, social, scalable, cloud-based, and real-time (Kobielus, 2011). Analytics-as-a-service (AaaS) sometimes also referred to as agile analytics, which means converting utility computing into a service-based model for analytics purpose. One of the advantages of AaaS is that it is not restricted to a single database or software, rather AaaS based platform can share its utility for an enterprise which is focused on virtualization of analytical services (Demirkan and Delen, 2013). Analytics across the enterprise is emerging very rapidly to solve complex problems. The new and improved database architectures are required to process the vast amount of structured and unstructured data in the shortest time possible with higher accuracy. Companies like Amazon, Microsoft, eBay, Opera Solutions, etc. are already facilitating "analytics-as-a-service" model. For example, eBay employees access a virtual slice of the main data warehouse server, where they can store and analyse their datasets. The eBay's virtual private data marts have been quite successful. Hundreds of them have been created with 50–100 in operation at one time. They eliminated the company's need for new physical data marts that cost an estimated $1 million apiece and require the full-time attention of several skilled employees to provision (Winter, 2008).

AaaS in the cloud has the economies of scale and scope by providing many virtual analytical applications with better scalability and higher cost savings. Cloud supported analytics solution should be able to analyse real-time events and help to analyse different types of data. Service oriented analytical solution should not only be restricted to data or text mining problems, but it should be able to address the following type of problems too:

- Large scale optimization
- Highly complex multi-criteria decision-making
- Distributed simulation model

Cloud analytics is an emerging alternative solution for large-scale data analysis. Data oriented cloud systems include storage and computing in a distributed and virtualized environment. These solutions also face some challenges, such as security, service level, data governance, etc. Even though prolific research has been reported in this area, but these issues need to be addressed at a finer level. As a result, there is ample opportunity to bring analytical, computational, and conceptual modelling into the context of service science, service orientation, and cloud intelligence (Demirkan and Delen, 2013). Analytics plays a crucial role in any organization. Therefore, it is imperative for the companies to have analytics-savvy culture. Companies can realize the benefits of analytics services on the cloud without investing a huge amount in setting up infrastructure. In the subsequent sections, we presented how predictive analytics can be realized in cloud computing.

4.4. Predictive analytics in cloud

Predictive analytics encompasses a variety of statistical techniques from predictive modelling, machine learning, that constitutes data mining which analyzes current and historical data to make predictions about future or otherwise unknown events ("Predictive analytics," n.d.). Predictive analytics in BFSI area is used to answer some of the following questions:

- What is the likelihood of customer churn?
- How likely is a claim fraudulent?
- What will maximize customer's profitability?
- Demand prediction of product?

The value proposition of predictive analytics using the cloud is based on following points:

4.4.1. Scalability

Compute and data resources can be scaled up as per requirement. Cloud-based resources make it easy and cost-effective (James, n.d.).

4.4.2. Deployment agility

Deploying predictive analytics project is a complex and challenging process for many organizations. Cloud as a deployment platform reduces cost and time to deploy predictive analytics and also increases agility (James, n.d.).

4.4.3. Pervasiveness

Always available nature of cloud-based systems enhances the ubiquity of predictive analytics in an organization; Predictive analytics addresses more problems than the descriptive analytics. Now-a-days, more and more data is available in the cloud so cloud-based predictive analytics solutions increase efficiency with lower latency (James, n.d.).

Above points support following three primary use cases of predictive analytics in the cloud:

1) **Pre-Packaged cloud-based solution**: A complete decision-making solution where predictive analytics is embedded in the solution framework, which is used for decision-making. Such solutions are usually known as Decision-as-a-Service model. For example, a multi-channel cross-sell solution wants to decide which product they should offer to the customers as a cross-sell product in different channels. They solved it using predictive analytics based models by predicting the likelihood of customers to buy a product. The kind of predictive analytics they employed was based on rules, which helped them take the decisions (James, n.d.).
2) **Cloud-based predictive modelling**: Development of predictive models is switching over to the cloud swiftly, which helps in taking advantages of scalable processing as well as fast access to data. As most of the organizations already using SaaS, the sufficient amount of data is already available on the cloud, which they are using or managing e.g. web analytics data, credit bureau data, CRM data, transactional data, etc. Apart from the traditional data sources, social media data and various other types of data is already available on the cloud which can be used by the analyst for faster analysis and a 360-degree view of the customer (James, n.d.).
3) **Cloud-based deployment of predictive analytics**: Deploying predictive models on the cloud and make it pervasive in an organization e.g. a European bank adopted a cloud-deployed model to predict the Probability of Default (PD) for residential mortgages during the origination process. This on-premise developed model has been deployed in a real-time engine on a private cloud. This allowed the model to be included in an origination process that orchestrated different services inside and outside the bank (James, n.d.).

4.5. Big data analytics in cloud

Big data refers to massive data sets that are large (Volume); more diverse i.e. structured, unstructured and semi-structured (Variety); and arriving faster (Velocity) than you and your organization deal before. This data deluge is usually generated from weblogs, social networking sites, IoT devices, RFID, etc. These data are heterogeneous in terms of formats like text, audio, video, etc. Insights generated from this data can prove to be valuable. Big data analytics is a combination of advanced technologies which use sophisticated data mining, text mining and web mining methods to explore the data and to discover interrelationships and patterns (Assunção et al., 2015). Analytics and big data are gaining popularity across various industries. However, they are facing a lot of difficulties to put all things into practice because of its complex nature and time-consuming endeavor. An organization which is willing to adopt analytics usually acquires expensive software licenses, deploy massive infrastructure, and pay the huge amount to consulting companies. Such efforts are expensive and lack flexibility (Assunção et al., 2015). Cloud computing, on the other hand, is becoming a reality for businesses. It provides flexibility to the organization regarding the pay-as-you-go model, availability of resources, and cost reduction. Clouds distinguish between technology and implementation, hence come with infrastructure, platform and software as a service. Cloud saves organizations from maintaining large infrastructure, which they are unlikely to use most of the time. At first, we will see exemplary steps performed for big data analytics and then relationships between cloud computing and big data.

Fig. 4 explains the typical processes performed during big data analytics. The first block depicts data sources like Data Mart's, streams and data warehouses, which are used for model building. Data can be of any one format of structured, unstructured or semi-structured. The second block depicts that a large volume of data in different format requires pre-processing, cleansing, transformation and filtering for the modelling purpose. The preprocessed data is used for training purpose and estimation of parameters. Different machine learning, data mining, and statistical techniques are used for modelling purpose. Once the model is estimated, the model needs to be validated. Normally model validation is performed on original input data, which can be carried out using different techniques.

Cloud-based big data analytics has a number of core benefits as follows:

- After wide acceptance of cloud computing in different industries, a significant percentage of data resides in the cloud, which can be used for big data analytics in cloud computing.
- Scalability of the cloud allows a larger volume of data to be handled efficiently and used for modelling.
- With cloud computing, it is easy to access, integrate and manage different types of new datasets.

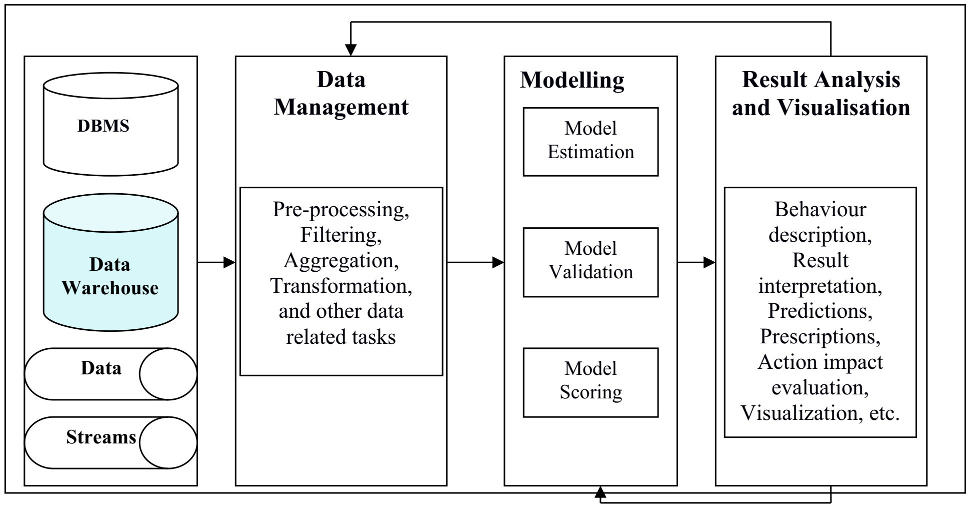

Fig. 4. Overview of the analytics workflow for Big Data (Assunção et al., 2015).

Table 1
Various service providers for big data analytics.

	Google	Microsoft	Amazon	Cloudera[a]
Big Data Storage	Google cloud services[b]	Azure[c]	S3[d]	
MapReduce	AppEngine	Hadoop on Azure	Elastic MapReduce	MapReduce YARN
Big Data Analytics	BigQuery	Hadoop on Azure	Elastic MapReduce	Elastic MapReduce (Hadoop)
Relational Database	Cloud SQL	SQL Azure	MySQL or Oracle	MySQL, Oracle, PostgreSQL
NoSQL Database	AppEngineDatastore	Table Storage	DynamoDB	Apache Accumulo
Streaming Processing	Search API	StreamInsight	Nothing prepackaged	Apache Spark
Machine Learning	Prediction API	Hadoop + Mahout	Hadoop + Mahout	Hadoop + Oryx
Data Import	Network	Network	Network	Network
Data Sources	A few sample dataset	Windows Azure Marketplace	Public Datasets	Public Datasets
Availability	Some services in private beta	Some services in private beta	Public Production	Industries

[a] https://www.cloudera.com/.
[b] https://cloud.google.com/.
[c] https://azure.microsoft.com/.
[d] https://aws.amazon.com/s3/.

Despite the Cloud-based big data analytics promises several benefits, it raises some concerns regarding:

- Data management, integration, and processing.
- Model development and evaluation in cloud
- Data visualization
- Security

Big data analytics and cloud computing are interrelated. Big data analytics provides the solution for analysing massive datasets. Big data provides users the ability to use commodity computing for processing distributed query across multiple data sets promptly. On the other hand, cloud computing provides the underlying engine, scalability, and elasticity to handle massive data sets on the cloud. Big data works on a distributed storage technology rather than on the local storage. Cloud computing provides facilities for computation and processing of big data as a service model. Table 1 compares several big data cloud platforms (Talia, 2013).

5. Analytics for cloud

Analytics is very much needed to solve several operational and business problems faced by service providers while offering cloud services. By resorting to analytics, they can offer cloud services in the most objective, optimal and scientific manner without compromising the QoS attributes/requirements. In order to efficiently predict, manage, optimize, and handle probable uncertainties in the cloud, we need to employ various kinds of analytical techniques at different stages of cloud computing. Based on this idea, we divided the process of deployment of analytics for the cloud with respect to six aspects as depicted in Fig. 2. We believe that the categorization under this stream, is succinct, optimal and complete in all respects. For each of the reviewed paper, we mainly considered four aspects viz. problem addressed, the techniques employed to address the problem, the dataset or experimental environment, and the results reported.

5.1. Resource demand prediction

According to the requirements of a client, the demand for resources will vary a lot. In order to manage resources efficiently, we need to predict the resource requirements of the client based on the type of services and the number of parameters. Some of the studies employed agent-based approach in order to tackle dynamic demand for resources (Dutreilh et al., 2011).

In order to predict the resource demand in advance, Wood et al. (2008) considered the profiles of different kinds of overheads of virtualization to employ a regression-based model to predict application resource requirements. The proposed model predicts the resource requirements of any application for any platform. They had used Xen VM monitor to measure the effectiveness of the model proposed. This developed a fully automated model building approach, which was able to characterize the various virtualization problems of different kinds of platforms. This model has median prediction error less than 5% for both the RUBiS (Cecchet et al., 2002) and TPC-W benchmarks (TPC, 2003). Jiang et al. (2011) presented an online temporal data mining system called ASAP to model and predict the cloud VM demand. ASAP extracts high-level provisioning request stream and notify the provisioning system to prepare VM in advance. They introduced an asymmetric and heterogeneous measurement “Cloud Prediction Cost” in the context of cloud service to evaluate the quality of the prediction results. ASAP is a three-module system that leverages two-level ensemble algorithms to predict the multi-type VM demands based on the basis of historical usage patterns. At the first level, different regression models like moving average, auto regression, MLP, SVM, and gene expression programming were employed. At the second-level correlation based time-series were considered. The proposed model was evaluated on the real historical data obtained from IBM's current cloud service platform. The dataset contained tens of thousands of VM requests with more than 100 different types collected during 13th Feb 2011 to 2nd June 2011. The results indicated that ASAP can effectively reduce the customer waiting time for VM while not causing much idle resource. Then, Mark et al. (2011) proposed an evolutionary optimal VM placement algorithm with a demand forecaster. As we know that, there are two types of resource provisioning methods, namely *reservation plan* and *on-demand plan*. Every user tries to optimize their cost i.e. increase revenue or decrease cost that incurs in operations. Accordingly, a user determines the number of cloud providers for different purposes. The users can opt for multiple providers for different purposes especially when some providers may not provide sufficient amount of resources at some point in time. In order to meet IT needs, the users must optimize the allocation of resources by splitting their demands to various chunks to multiple providers and should decide the best combination that incurs the smallest cost. In order to solve this problem, they introduced an Evolutionary Optimal Virtual Machine Placement (EOVMP) algorithm. The EOVMP algorithm helps in getting near optimal decisions of VM placement. The objective of EOVMP is to minimize the cost that incurs by predicting the demand of users. Further, EOVMP optimizes the VM placement by utilizing users' usage history. The algorithm yielded similar results to that of stochastic integer programming (Mark et al., 2011).

Dutreilh et al. (2011) proposed a Reinforcement Learning (RL) based approach, which is employed in an agent-based system to determine the dynamic demand. RL was based on a proposed Q-function, which is easy to learn from experience. They found that an appropriate initialization of resources at initial stages and proper convergence speedup throughout the learning phases could efficiently allocate the resources to the clients. They proposed an autonomic framework, which combines reinforcement learning and allocation strategy together. They have simulated Olio (Sobel et al., 2008) and the VirRL (Dutreilh et al., 2010) decision agent.

The problems of database applications hosting include optimizing database performance, isolating database users, pricing the services, etc. In this regard, Mozafari et al. (2013b) proposed a study called DBSeer which addresses these challenges. This DBSeer tool is an initial step towards better resource prediction for database users. The tool generates models for online transaction processing databases, which is dependent on cache model (Tran et al., 2008) and a simple iterative algorithm (Mozafari et al., 2013a). The solutions provided by the tool are good at tackling above-mentioned issues (Mozafari et al., 2013b).

Verma et al. (2013) proposed a two-level prediction framework. At the first level, they employed seven classification algorithms viz. logistic regression, probabilistic neural network, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), REPTree, Structural Learning Algorithm on Vague Environment (SLAVE), and C4.5 to predict the increase in demand for resources. At the second level, they employed regression models like exponential moving average, trend seasonality model, polynomial regression, and ARX/ARMAX for short-term and long-term demand prediction. They experimented with a dataset, which was developed by simulating a set of 1000 realistic service tenants. For classification purpose, REPTree yielded the best sensitivity of 94.6%, whereas trend seasonality model yielded the least mean squared error of 15.75%. Zhong et al. (2017) proposed an effective hybrid of Wavelet Transform (WT) and SVM and called as Wavelet Support Vector Machine (WSVM) for load forecasting. The model combines WT's advantage of analysing the cycle and frequency of input signals with SVM's characteristic of nonlinear regression analysis. They experimented with Google cluster dataset and compared the performance with baseline methods like ARIMA and ANFIS. This dataset contains the usage information of nearly 12,000 machines for 29 days in the Google cluster dataset. Furthermore, the Google cluster dataset contained 672,003 jobs and 25,462,157 tasks of scheduling events, resource requirements, and records usage. Liu et al. (2017) proposed an adaptive method for workload pattern classification. This approach firstly categorized the workloads into different classes like CPU usage, memory usage, local disk usage, etc. which were given as inputs to some prediction models according to workload features. Furthermore, the workload classification problem is transformed into a task assignment one by establishing a mixed 0–1 integer programming model, which is made available as an online solution. Models were evaluated using a real-world load trace, which was sampled from the Google Data Center (Google, 2015). The Google cluster traces are publicly available, which contained traces of approximately 12,500 machines over 29 days for nearly 650,000 jobs (Reiss et al., 2011).

5.2. Resource allocation and scaling

With the growing adoption of virtualized computing resources and datacenters, the allocation and sizing of various computing resources such as CPU, memory and I/O bandwidths are very important. Cloud computing provides the cost-saving benefit using dynamic scaling of resources, which can be accomplished by using VM. But, the setup time for virtual resources under cloud framework takes a lot of time. In order to predict the estimated time needed for a particular setup, we can employ machine learning algorithms as well as resource optimization algorithms to reduce the setup time (Amiri and Mohammad-Khanli, 2017).

In the case of fluctuating demand, cloud provisioning can be used to allocate resources to a client application in dynamic mode. But, dynamic provisioning suffers due to lack of instantaneous demand of infrastructure. Moreover, the start-up time of the resources is another issue to be faced during dynamic provisioning. The start-up time will affect the QoS of the cloud clients. Consequently, the companies may churn out from a cloud provider. In order to mitigate this problem, the cloud providers use predictive approach to determine future platform usages. Using the predictive approach, cloud providers will be able to predict scaling requirements as well as the direction of scaling by compensating the resource start-up time. So, the historical resource usage patterns can help in automatic infrastructure resource scaling decisions for cloud clients (Jiang et al., 2013b). The cloud provider can detect the usage patterns of the client by mining the historical resource usage records. The future usage pattern can also depend on the domains of different cloud clients. The cloud clients working in the same domain will have a close similarity of resource usage. So, exploiting the past usage data with respect to application domains of different clients can help in better prediction.

As the demand for cloud computing further increases, *scaling* need to be employed carefully to utilize resources efficiently. To allocate required resources to the requesting client, we need to optimize the available resources in order to meet every client's requirements. As the requirements exceed the available resources, investments in new infrastructure will be needed. New investments can be made using a *two-period model*, which may involve a third party. The third party acts as a bridge between clients and cloud providers. The third party uses two mechanisms that help the provider in future capacity planning. First, he extracts probabilities from clients through an options market. This is to determine the clients' requirements for the future, which can be used to schedule workloads. Then the third party uses the previous market demand to predict whether the cost can be reduced by reserving for a longer period. This procedure is proved to be profitable for clients as well as a provider even in fluctuating market conditions (Rogers and Cliff, 2012). Second, cloud providers must be able to interoperate in order to meet the clients' dynamic expectations, which will also help consumers to switch over between different providers easily. There are two problems viz. *interpolation* and *lock-in,* which prevent the consumers from switching over from one provider to others. Here, the interpolation is the process of exchanging of resources, which enables the use of centralized compute-resource. The lock-in is the period, during which a client cannot switch over to other provider (Rogers and Cliff, 2012).

Keeping above-mentioned issues in mind, we reviewed a number of papers under this section. The reviewed literature are grouped under two category viz. resource allocation based study (Banditwattanawonga et al., 2016; Bennani and Menasce, 2005; Calheiros et al., 2015; Caron et al., 2010; Díaz et al., 2017; Grande et al., 2011; Guiyi et al., 2010; Imam et al., 2011; Kundu et al., 2012; Pandey et al., 2010; Rahman et al., 2011; Rankothge et al., 2017; Tsai et al., 2013; Zhang et al., 2017; Zhao et al., 2015, 2017) and scaling based study (Ashraf et al., 2016; Chou et al., 2016; da Rosa Righi et al., 2017; Gong et al., 2010; Jiang et al., 2013a, 2013b; Mao and Humphrey, 2011; Roy et al., 2011; Shahin, 2016).

In a cloud environment, if resource allocation policy is effective and efficient then it saves considerable cost for the service provider. Cloud vendor needs to create multiple new VMs in order to serve a continuous load increase. However, a substantial amount of data is required for the creation and setup of a VM, which directly affects the QoS. In this regard, Bennani and Menasce (2005) considered solving a datacenter resource allocation problem. They employed predictive multi-class queuing network model (Menasce et al., 2004) to predict the response time and throughput for CPU and disks. They experimented using three Applications Environments (AE) and each containing twenty-five servers. Among three AE, two were in online and one in a batch environment. They reported that an analytic performance model can be employed to design a controller, which can switch the server on-demand from one AE to other efficiently. Guiyi et al. (2010) proposed a game theoretic method for fair resource allocation of cloud computing resources. They proposed a two steps solution in order to solve the resource allocation problem. In most of the resource allocation problem, scheduling algorithms assume that tasks are independent of each other. However, the majority of cloud-based complex applications consist of multiple subtasks and require communications among tasks. In this regard, they presented a game-theoretic method for scheduling of cloud-based computing services with collaborative QoS requirements. A cost is incurred at each service that depends on the amount of computation. Each computing task is associated with multiple dependent and homogenous subtasks, which are sensitive to the execution time. They used binary integer programming method to obtain initial independent optimization. Based on initial results, an evolutionary mechanism is designed to achieve the final optimal and fair solution. They also reported that Nash equilibria always exist if the resource allocation game has feasible solutions.

Based on resource usage similarity, Caron et al. (2010) proposed to predict the future usage of cloud resources. They followed an approach similar to a string matching problem called Knuth-Morris-Pratt (KMP) algorithm (Caron et al., 2010). The algorithm has many parameters to tune like the maximum number of matches, the length of the predicted sequence, acceptable errors, etc. Imam et al. (2011) proposed prediction model based allocation of resources which can improve the quality of service of the cloud platform. More specifically, time delay neural network and regression methods were employed to predict the future workload in the grid or cloud platform. They experimented with the dataset of workload trace from the Large Hadron Collider Computing Grid (LCG). LCG is the grid established to analyse the 15 petabytes of data annually produced by the Large Hadron Collider in CERN (European Organization for Nuclear Research) ("Worldwide Large Hadron Collider Grid," n.d.). The time delay neural network performed well in predicting the grid workload. The user applications involve high data transfer between resources. In this regard, Pandey et al. (2010) minimized the provisioning cost and data transfer cost using Particle Swarm Optimization (PSO) based heuristic for scheduling workflow applications. PSO was employed to optimize computational cost and data transmission cost for scheduling of applications. The results yielded by PSO were as good as three times saving the costs compared to the *best resource selection* algorithm. Further, the proposed approach performed well in the distribution of workloads for resources allocated (Pandey et al., 2010).

Grande et al. (2011) proposed several efficient time series model to enhance prediction precision for HLA (High-level Architecture) simulation systems. High-Level Architecture (HLA) is a framework for the design and management of distributed simulations through a set of management services. These variants are extended from Holt's model in order to address issues related to a dynamic load balancing system (Grande et al., 2011). A set of migration decision-making techniques is also proposed to promote a modular construction, which was independent of any prediction model. Calheiros et al. (2015) proposed Auto-Regressive Integrated Moving Average (ARIMA) based model for predicting cloud workload for SaaS providers. For experimental purpose, they setup a simulated environment of datacenters, which contained 1000 hosts. Each of hosts contained 8 cores and 16 GB of RAM. The proposed workload analyser is evaluated in this environment. They reported different performance measures like root mean squared deviation, normalized root mean square deviation, mean absolute deviation, Mean Absolute Percentage Error (MAPE) is also reported. The best MAPE reported was 0.09, which indicates that there is an efficient utilization of resource and minimal impact on the QoS.

Meta-heuristics based approaches for scheduling of resources lack the ability to consider workflow level optimization and user quality of service constraints. Rahman et al. (2011) proposed an Adaptive Hybrid Heuristic (AHH) for user constrained data analytics. This model integrates the dynamic nature of heuristic based approaches and workflow-level optimization. The proposed model was evaluated against the existing model using a comprehensive case study. The model is able to handle the dynamic nature of cloud environment and is able to satisfy users' budget in a given time stamp (Rahman et al., 2011).

Kundu et al. (2012) modelled the relationship between the resource allocation to a virtualized application using a limited amount of training data. They employed MLP and SVM to model the performance of a VM-hosted application as a function of the resources allocated to the VM (Kundu et al., 2012). They experimented with five virtualized applications from RUBiS (Cecchet et al., 2002) and Filebench suite of benchmark (McDougall et al., 2011). They reported that median and 90th percentile predictions errors were within 4.36% and 29.17% respectively. These techniques outperformed regression-based approaches. They also developed per-application performance models, which can be used for VM resource provisioning (Kundu et al., 2012). Tsai et al. (2013) proposed Improved Differential Evolution Algorithm (IDEA) to optimize scheduling tasks and resource allocation based on cost and time models. The IDEA is a hybrid of Differential Evolution Algorithm (DEA) and Taguchi method. The DEA is powerful at finding a global optimal solution on a macro-space level by using less number of control parameters. On the other hand, Taguchi method provides effective statistical experimental design processes for evaluating and implementing any enhancements for processes and products. As the DEA is powerful at global exploration on macro-space, the Taguchi method is powerful at exploiting better individuals on micro-space with systematic reasoning ability (Tsai et al., 2013). The cost model involves the rent cost for receiving and processing of subtasks, whereas the time model includes receiving, processing, and waiting time. The IDEA outperformed the DEA and NSGA-II in searching the better Pareto-optimal values (Tsai et al., 2013).

To support AaaS platform for big data, Zhao et al. (2015) suggested effective scheduling of cloud resources for BDAAs (Big Data Analytics Applications) and provisioning of BDAAs as consumable services. They developed admission control and resource scheduling algorithm that admits queries based on QoS requirements to ensure SLA. They proposed an architecture for AaaS platform and formulated the resource scheduling problem using Mixed Integer Linear Programming (ILP). The proposed architecture is evaluated on the simulated environment of the datacenter. The simulated environment consists of 500 physical nodes, and each node consists of 50 CPU cores and 100 GB memory, 10 TB storage, and 10 GB/s network bandwidth.

In order to implement greener Big Data analytics in a cloud environment, we need to address several problems like network bandwidth, increase in external private-cloud data-out charge imposed, long-delayed cloud service responsiveness, data transfer bottlenecks, data transfer costs, scalability, and SaaS SLA responsiveness. The Client-Side Cloud Caching (CSCC) is one of the solutions to scalability, economy and responsiveness requirements of cloud computing services. Amazon Cloud Front offers a caching service through the content delivery network. CSCC retrieves the data from nearby consumer premises (Cache hit) or downloads the updated copies from the cloud (Cache miss). Some of the characteristics of CSCC include multi-tenancy support, Cache-as-a-Service (CaaS) models, etc. Here, the multi-tenancy support provides sharing of hardware among different enterprises. CasS model provides RAM Multi-tenancy Isolated database and SSD Multi-tenancy Isolated database. CSCC can be provided in a multi-provider cloud environment (Petcu, 2013; Wright et al., 2011) or interconnected as a hybrid cloud (Mell and Grance, 2011). In this regard, Banditwattanawonga et al. (2016) proposed enterprise-level client-side shared cloud caching towards multi-provider clouds. They presented an intelligent cloud cache replacement policy, i-Cloud (named so for its intended application domain) as depicted in Fig. 5. It is based on the non-uniform cost model. They experimented on two cloud service providers namely Google storage network and Amazon web service (AWS). They employed MLP in order to predict the future price. They considered a 31-day BO trace from a user community in Boulder from 16th August to 15th September 2012, and a 31-day NY trace was collected from the other user community in New York from 16th July to 15th August 2012. They reported byte-hit ratio, cost-saving ratio, delay-saving ratio, and hit rate as performance metrics. The proposed approach outperformed least recently used, GDSF, and LFU-DA.

Díaz et al. (2017) proposed LLOOVIA (Load Level based Optimization for VIrtual machine Allocation), which optimally allocates the VMs required for a service. Moreover, it ensures minimization of cost and guaranteeing the required level of performance. LLOOVIA is an allocation strategy, which minimizes the cost, type, number, price schema and a number of the VMs required in a multi-cloud environment. They evaluated the proposed technique on the real workloads of Wikipedia (English), which is publicly available (Baaren and Erik-Jan, 2009). Zhao et al. (2017) reported that SLA violations along with energy consumption could decrease the datacenters cost efficiency. They proposed two online VM placement methods viz. First-Fit and Harmonic algorithm in the non-migration environment and another two namely Least-Reliable-First (LRF) and Decreased-Density-Greedy (DDG) in migration environment. Without migration, First-Fit yielded more revenue by 1.7% than Harmonic one. In migration environment, DDG generated more revenue by 1.23% than LRF. First-Fit, LRF, and DDG outperformed the Stack and Spread algorithms of Openstack on real traces. They reported consistent results with synthetic traces too. Zhang et al. (2017) proposed an online primal-dual optimization framework and a randomized reduction algorithm for VM allocation. The proposed solution was able to maximize the social welfare with server costs and profit for the provider. They experimented with exploiting Google cluster-usage data (Reiss et al., 2011) and outperformed existing offline solution on similar frameworks. Rankothge et al. (2017) presented a comprehensive analysis of GA and GP based resource allocation algorithms for Virtualized Network Functions (VNFs). The employed algorithm is able to allocate resources for initial requests and is able to scale according to future requirements. The employed algorithms outperformed Integer Linear Programming resource allocation technique in terms of time efficiency, where they used two different fitness functions viz. *VNFs provisioning* and *scaling out/in*. They experimented with a Resource Manager comprising policy requests of the size of 400 VNFs, where 50% of the network function center servers need to be used to allocate resources for policy requests.

For dynamic scaling of resources, Gong et al. (2010) presented a novel approach PRedictive Elastic reSource Scaling (PRESS) for cloud systems. PRESS extracts the fine-grained dynamic patterns in an application resource demand and adjusts their resource allocation automatically. They employed fast Fourier transform to determine the variation of resource-usage and discrete-time Markov chain for short-term prediction of dynamic application requirements. The proposed algorithm is implemented on the Xen, which was tested on RUBiS and an application load trace from Google (Reiss et al., 2011).

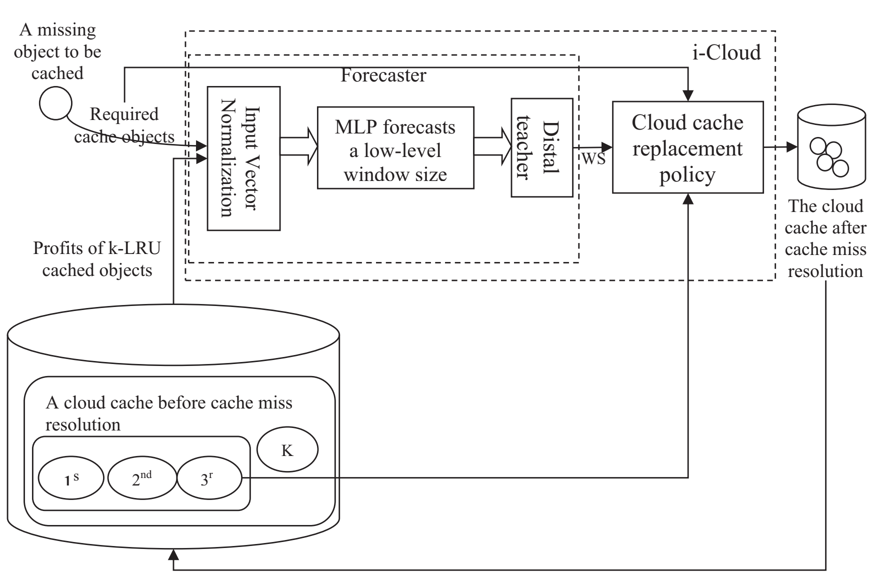

Fig. 5. The conceptual framework of i-Cloud cache replacement policy (Banditwattanawonga et al., 2016).

The results obtained indicated that resource usage predictions can be predicted more accurately. Further, the resource allocation achieved better service provider profitability compared to other approaches across the range of workloads. Shahin (2016) employed Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) for a dynamic threshold based auto-scaling algorithm that predicts the computing resources. They performed auto-scaling of virtual resources based on predicted values of future demand. Evaluation is performed using CloudSim simulator and NASA Log and reported MAPE, a number of running VMs vs. Time, average response time in milliseconds, and a number of completed requests.

Large-scale cloud resource users expect QoS guarantees in the SLA between the customer and the service providers. Load forecasting helps the vendor in keeping operational cost low and meet SLA. In this regard, Roy et al. (2011) reported overcoming the lack of effective techniques for forecasting and optimal resource allocation. They listed out challenges involved in auto-scaling in the cloud. They developed a predictive model for workload forecasting which is further employed for resource auto-scaling. They employed Second Order ARMA (Auto-Regressive Moving Average) for forecasting the workload. Finally, the empirical results were provided, which demonstrates that resources can be allocated and de-allocated by using the proposed algorithm. They reported that the allocation process can satisfy application QoS while keeping operational cost low. Jiang et al. (2013a) proposed a novel cloud resource auto-scaling scheme at the VM-level for web application providers. The proposed model predicts the number of web requests and discovers optimal cloud resource demand with cost latency trade-off. The proposed approach is evaluated on Amazon cloud platform using three real-world web-log datasets namely AOL,[3] Sogou,[4] and UTS (University of Technology, Sydney) library.[5] They predicted the number of requests using machine learning techniques and time series analysis. They were further able to predict an optimal number of VM by using queuing theory and multi-objective optimization. The proposed scheme is able to achieve the resource auto-scaling along with an optimal cost-latency trade-off and low SLA violations.

Cloud practitioners generally pursue schedule based and rule based mechanism in order to automate the matching between the computing requirement and available computing resources. But, most of the "auto-scaling" mechanism only supports simple resource utilization indicators and do not consider budget concern and performance requirements. Mao and Humphrey (2011) considered some additional factors of the auto-scaling problem to have a general application model. Further, they considered individual (non-uniform) job deadlines. The proposed approach is evaluated on three different types of workloads and patterns. The goal is accomplished by dynamically allocating or de-allocating VMs using scheduling of the most cost-efficient instances. The results indicated that the proposed approach can save cost from 9.8% to 40.4% as compared to the existing approach approaches. Furthermore, Jiang et al. (2013b) presented two different problems related to service quality viz. capacity planning for cloud vendor and instant VM provisioning. Traditionally, capacity planning was straight forward, i.e. in the case of increasing demand the service vendor simply upgrade the datacenters by scaling up the infrastructure. However, capacity planning becomes a difficult job in an elastic environment where the demand is highly fluctuating. Both underestimation and overestimation of the cloud capacity can lead to a huge revenue loss. The fluctuating demand of cloud resources needs provisioning and de-provisioning of VM very frequently. The time delay in provisioning and de-provisioning is not acceptable for a task, which needs scaling out during computation. In order to address this problem, they indicated that a practical, effective and achievable solution is to predict the demand and prepare the VMs in advance. They proposed an ensemble of five prediction techniques viz. moving average, auto regression, MLP, SVM, and gene expression programming. The prediction obtained using five techniques were combined using a weighted linear strategy. They conducted a series of simulation experiments based on the trace data of IBM Smart Cloud Enterprise (IBM, 2011). The trace data consists of VM request for more than 4 months (March 2011 to July 2011). The dataset contains requests for more than 100 types of VMs. Each record contains the value of 21 features such as customer id, VM type, request start time, request end time, etc.

Chou et al. (2016) proposed a Dynamic Power Saving Resource Allocation (DPRA) mechanism based on Particle Swarm Optimization (PSO). DPRA considers energy consumption of physical machine and VM as well as energy efficiency ratio of air conditioning systems. DPRA utilizes least square regression method for forecasting PM (Physical machine) resource utilization for allocating VM and eliminating VM migrations. The proposed DPRA is compared using simulation on three existing allocation schemes, i.e. best-fit algorithm with random VM selection (BFrand), a best-fit algorithm with minimum CPU (BFminU), and Best-Fit with minimum memory (BFminM). DPRA outperformed the traditional schemes in terms of energy consumption. It has been shown that the DPRA energy consumption is decreased by 17.85%, 15.03%, 18.04%, and 11.36% in total energy consumption while comparing with BFrand, BFminU, BFminM, and the PSO in (Dashti and Rahmani, 2016). Ashraf et al. (2016) developed a cost-efficient VM provisioning and admission control approach for multi-tier applications based on predictions. They proposed an automatic deployment and scaling of multiple web applications on a given IaaS. The resource requirements of a web application vary over time, which further depends on the type of task need to be performed. Multiple simultaneous users were traditionally handled by deploying 3-tiered architecture. However, if the number of users increases beyond the limit, an application will demand more resources. On the other hand, the less number of users and more resources will cause a lot of unused resources and involve an opportunity cost. The proposed approach for VM provisioning and admission control approach multi-tier web application is based on prediction and cost-efficient provisioning. The proposed approach provides automatic deployment and scaling of multiple simultaneous third party web applications on a given IaaS cloud in a shared hosting environment. The approach is based on OSGi component model to share VM resource among deployed applications. The proposed approach comprises three sub-approaches: (i) a reactive VM provisioning approach called ARVUE (Aho et al., 2011), (ii) a hybrid reactive-proactive VM provisioning approach called Cost-efficient Resource Allocation for Multiple web application with Proactive scaling (CRAMP), and (iii) a session based adaptive admission control approach called adaptive admission controlled for virtualized application servers. CRAMP is a two-step prediction algorithm comprising exponential moving average and linear regression model. The algorithms were applied on both synthetic and real datasets. The comparison of results among various approaches indicated that VM provisioning experiment with ARVUE and CRAMP provide good performance in terms of average response time, CPU load average, and memory utilization. In terms of responsiveness, CRAMP outperformed ARVUE. In another study, da Rosa Righi et al. (2017) developed a hybrid elasticity service for a master-slave parallel application named Helpar, which provides resource provisioning on parallel applications. The proposed approach is able to overcome drawbacks related to proactive and reactive approaches. They proposed Live Thresholding (LT) technique for controlling the elasticity. The proposed algorithm provided a lightweight plug and play service for PaaS and evaluated on OpenNebula and numerical integration application.

5.3. Uncertainty modelling

Despite the wide acceptance of cloud computing, it suffers from an

[3] http://www.infochimps.com/datasets/aol-search-data.
[4] http://www.sogou.com/labs/dl/q-e.html.
[5] https://library.sydney.edu.au/.

uncertainty issue. If an organization wants to utilize storage feature and store its data, it has no or little control over the data. So, there is a challenge or uncertainty that the data stored on the cloud may be prone to hacking or some other security breaches. A survey conducted by InformationWeek Analytics[6], the clients' concern about security was at the top for not using services of the cloud. They reported that even the Cloud provider giants like Google are also facing the same issue of security concern. So, we can say that the clients are very concerned about their information or data that is deployed in the cloud. There's another uncertainty that the applications or infrastructure that is provided by the Cloud provider may not be the latest or the output of the services may not be up to the mark (Farah, 2013). Similarly, some of the major uncertainties are (Farah, 2013):

- Information missing
- Trusting the available information
- Inconsistency in the available information
- No relevant information
- Interpretable information

In order to avoid unexpected uncertainty, clients should prefer cloud providers, who are accountable. The important features like transparency and control are very much needed in the assessment of providers. They may include SLAs, quality control procedures, dependence on vendors, etc. (Farah, 2013). The uncertainties that a client may face based on the services provided by Cloud provider (Farah, 2013):

- **Governance**: The client may be uncertain towards governance of the infrastructure if there are security breaches.
- **Lock-in**: The uncertainty of lock-in means that the client becomes dependent on single provider preventing the client to change the provider. This may affect the quality of results or information that is provided by the Cloud provider.
- **Isolation failure**: A client may face uncertainty if there exists a separation of storage, memory, and routing among various clients. The client may not be aware of separation failures.
- **Compliance risks**: The uncertainty may also occur if the Cloud provider will not allow a client to audit. Further, if the provider gives any compliance client about the requirements needed.
- **Data protection**: clients may face uncertainty if they cannot get any information about data handling processes or if the provider doesn't provide.
- **Insecure or incomplete data deletion**: If any client wants to delete their data from the cloud, the provider may not delete the complete data or not interested in deleting.

Usually, cost/benefit analysis is performed on the infrastructure or software that is included in the IT portfolio. So, another kind of uncertainty is a function of cost/benefit analysis (Farah, 2013). The cost for utilizing cloud also includes uncertainties like data security, down time, data link connection performance, communication, dependability of the customer service, the cost of posting the application on the cloud computing, etc. The down-time depends on the duration of service repairing. The cost of posting the application depends on the rate and time to deploy an application on the cloud. For example, a client has to pay according to the number of days, during which an application is online. Because the service may not be needed on the weekend as well as some service may not be needed at the night. Estimating the cost may also depend upon the previous clients' experience with a particular provider or various cloud providers. The better you estimate, the better you will decide to place an application on the cloud (Farah, 2013).

The historical performances of the cloud provider can be considered while calculating the fraction of the revenue that the client wants to save. Sensitivity analysis can be employed to assess the robustness of the decision for a varying (or interval) values of the fraction. This approach provides a framework for dealing with the assessment of the fraction of the savings. So, such systematic approach could help in reducing the uncertainty about posting applications on the cloud. The fractions of the savings can be calculated by employing Analytic Hierarchy Processes (AHP) or decision trees with monetary values associated with possible outcomes. In case the assessment of cloud is not credible, clients can use qualitative methods like rankings rendered by a panel of experts. In some cases, both quantitative and qualitative methods can also be used (Farah, 2013). An organization wants to save money by placing an application in the cloud, which depends on some variables. The variables include human resource (IT staff) reduction, the cost of buying or upgrading the software, the maintenance cost, the requirements of the servers, software requirements, etc. Some variables are used to assess the cost for uncertainty like the security issues, maintenance cost, availability of customer care center, and the cost incurred for the customization (if the customization of the service is needed) (Farah, 2013).

Due to elastic nature of resource demand, large swings are very common. Someone cannot model the resource demand accurately based on a number of session request because it shows very large variability and poor auto-correlation. In order to address such demand uncertainty, Saripalli et al., (2011) proposed a load prediction algorithm for cloud platforms, using a two-step approach of load prediction, using cubic spline interpolation and hotspot detection algorithm for sudden spikes. Two step approaches consist of load tracking followed by the load prediction. Load tracking was a representative view of load trend from raw data in the form of load tracker. It is an apriori approach to filter out noises from a raw sequence of uncorrelated resource measures sampled over time.

Infrastructure-as-a-Service (IaaS) cloud infrastructure allows users to lease resources in a pay-as-you-go-fashion. The complex nature of IaaS makes it highly prone to the performance anomalies due to many reasons such as software bug, resource contentions, hardware failures, etc. On one hand, it is very hard to detect such kind of anomalies manually on tens of thousands of VMs. On the other hand, the delayed anomaly detection can cause long service-level objective violation time, which is often related to a financial penalty. In this regard, Dean et al. (2012) presented Unsupervised Behavioural Learning (UBL) for IaaS cloud m-computing infrastructures. The proposed techniques employed Self-Organizing Maps (SOM) to capture emergent system behaviour and predict unknown anomalies. The prototype of UBL is implemented on top of the XEN platform. They performed extensive experiments on a real-world distributed systems running inside a production cloud. UBL yielded up to a sensitivity of 98% and 1.7% false positive rate and raises advance alarms within up to 47s lead time. Due to its lightweight nature, it is suitable for large-scale cloud computing infrastructures. Minarolli et al. (2017) considered to address uncertainty of long-term predictions using non-parametric density estimation method. They were able to detect overload, underload, not-overload, and probabilistic overload. They employed Weka (Hall et al., 2009) based Gaussian process for regression for long-term time series prediction. Variable Kernel Method with Gaussian Smoothing is used for kernel density estimation.

5.4. Cloud trading

As clouds are becoming mature and popular, marketplaces will be developed to fulfil clients' needs. For the better exploitation of marketplace, the clients will be facilitated with the composition of different kinds of services across multiple providers. Here, the marketplace is an environment that offers different kinds of services at the single place. The marketplace acts as an intermediate for searching, choosing and

[6] http://www.informationweek.com/cloud/software-as-a-service/time-to-think-about-cloud-computing/d/d-id/1073198.

selling or buying of services. The marketplaces must enable clients to use the whole lifecycle of services and strengthen the offerings provided by various cloud providers. So, a one-stop marketplace is needed to solve the above-mentioned problems. A one-stop marketplace provides the different business case optimization. For multi-cloud services, we need an environment, which allows an integration of business and technical aspects without any interruption into the service offering procedure (Menychtas et al., 2014).

In future, the internet should concede the communication between layers and players principle. According to this principle, the various providers must be able to exchange information about their choices and priorities. This communication in a marketplace not only restricted to combining, collecting and evaluating but also for taking decisions in a multi-provider environment. This marketplace mechanism must allow the decision-making by the ways that decipher high-level business and service requirements. Such mechanism provides requirement proposals and service composition recommendations. The marketplaces must be able to go beyond SLA-based services selection or services discovery based on analysis of performance. These marketplaces can use techniques by analysing the dependencies between the service providers for respective products. These are useful for pricing models and business resolution revenue sharing. Furthermore, these marketplaces will be helpful in proper delivery of the service as well as compensate a proper payment of the services for a customer. The marketplace analytics can analyse the income that can be helpful for cloud providers (Menychtas et al., 2014).

Menychtas et al. (2014) proposed a marketplace, which is a pool of cloud marketplace solutions. Such marketplace tackles several above-mentioned problems. It provides a simulation of various kinds of business cases by optimizing the product offerings, which use different kinds of parameters in the wide range (Menychtas et al., 2014). They proposed an advanced compact price models by combining a new resolution approach that integrates the business intelligence, which includes *find* and *choose* procedures. Such approach provides various kinds of pricing alternatives as well as services customization. These pricing models will have a big impact on the large share value of the market revenue. A provider may suffer while offering services if one does not have experience of past offerings. Another possible reason of suffering may be a lack of knowledge of the particular market, which may increase the profit of an organization. So, these marketplaces help them a lot with the help of historical data again. The 4CaaSt marketplace is an analytical tool that helps in taking proper decisions while defining offerings or optimization of existing offers. The tool provides various options for visualization, simulation, and optimization (Menychtas et al., 2014).

Cloud computing is one of the important requirements for many companies, especially for start-ups. By adopting cloud computing, they can avoid the cost of procuring traditional infrastructure resources, which also takes several months to setup. However, a cloud-hosting provider offers this near-infinite resource on demand using a different kinds of pricing models. One such example is pay-as-you-go model. These models are dynamic in nature i.e. it can be allotted to the application provider as per the requirement. Pay-as-you-go kind of model and dynamic resource provisioning feature of the cloud reduces the challenges of over- and under-provisioning, which happens in the case of static provisioning of resources. As compared to traditional month-long procurement of resources, instantiation of new VM on demand is relatively less time-consuming. In this regard, Sadeka et al. (2012) devised a smart way for dynamic provisioning of the resources, which is effective in terms of both cost and performance. They proposed a prediction framework based on statistical models, which are able to speculate future surge in the resource requirement. They employed machine learning techniques on the basis of a sliding window technique. They experimented with TPC-W (TPC, 2003) benchmark dataset in the Amazon EC2 cloud. They reported the effectiveness of the prediction framework using MAPE, PRED, etc.

Generally, cloud providers offer two provisioning plans for computing resources namely on-demand plans and reservation plans (Foster et al., 2008). The cost of utilization of computing resources provisioned by reservation plan is generally cheaper than that of the on-demand plan. However, a reservation of the resources is very difficult to achieve due to uncertainty in consumers future demand and providers resource prices. To address this problem, Chaisiri et al. (2012) proposed OCRP (Optimal Cloud Resource Provisioning) algorithm, which is applicable to provision computing resources under short-term as well as a long-term plan. OCRP algorithm was proposed by formulating a stochastic programming model, which considered the uncertainty of price and demand. In addition to the stochastic approach, they also employed deterministic equivalent formulation, sample-average approximation, and Benders decomposition. The performance of OCRP algorithm was evaluated using numerical studies and simulations. They reported that the algorithm could optimally adjust the trade-off between reservation of resources and allocation of on-demand resources. In cloud computing, the resources were allotted to consumers as per the requirements.

The broker mechanism is controlled through SLA. According to SLAs, there is a provision for compensation if the targets are not achieved. In order to maintain the relation between clients and providers, SLA can be adopted. According to SLA, the providers will have to compensate the clients according to predefined terms and conditions. In order to prevent from compensation and loss of reputation, the provider should go for adequate investment in new technology and optimal utilization of existing capacity. With the help of clients' resource reservation, their future requirements can be predicted but it may suffer due to two reasons viz. *variable nature of IT usage* and *the trustworthiness of clients' information*. Due to former one, it is difficult to fix pricing of the service too. Furthermore, *competitiveness* and *profitability* need to be considered for the pricing model. The issue of *variable IT usage* can be addressed by employing analytics, whereas *trustworthiness* can be ensured by employing a reservation model. The reservation model will lead to a truthful reservation on the clients' side. In addition to that, a third party (Broker) can be involved for receiving and resolving resource requests. The mechanism of reservation can be made profitable in a multiple-user, heterogeneous and variable-demand market to both clients and providers by making available information about future demands in advance (Rogers and Cliff, 2012). The broker will be beneficial only when the resources are abundantly available and there is no surplus or deficit of the resources purchased by the broker. In order to address these problems, Rogers and Cliff (2012) proposed an agent-based approach involving a two-step process for switching over providers. In the first step, the user submits the probability of demand to the broker, which may not be the real probability, but an approximated probability. The broker must reserve the resources that to be required in the future. In the second step, the broker allocates the resources to the clients. For reserved instances, the broker has to pay some reduced cost, whereas the clients need to pay to the broker according to their requirements. In the case of agent-based simulation, there is a less probability that the cloud provider will have to keep resources idle or less profitable (Rogers and Cliff, 2012). They experimented in a simulated environment and reported that historical data can help in increasing the profit of a broker by 28%.

Various cloud service providers made available various forms of cloud services. The selection of cloud service is a quite cumbersome task and need to carry out at the large scale. In order to fulfil the user requirements, we need to optimize the process of selection of services. Here, user requirements can be either consumer- or application-oriented. Due to the availability of multi-cloud environments for different applications, the client will use multiple cloud providers, instead of running on a single provider solely. As the complexity in cloud market is increasing, the broker is interacting with the cloud providers on behalf of cloud consumers to provide decision support for consumers. The Cloud Resource Management Problem (CRMP) in this

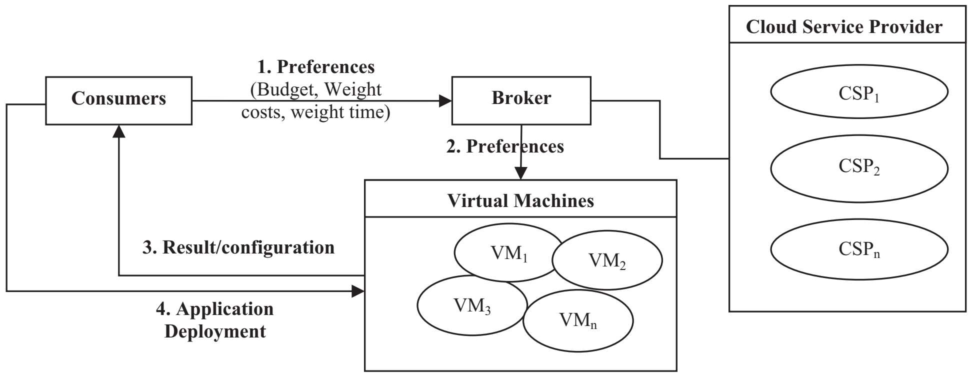

Fig. 6. Broker Architecture (Heilig et al., 2016).

multi-cloud environment (MCE) is reducing the cost and execution time of consumer applications. Some of the benefits of MCE are: the utilization of services offered by different cloud providers (Petcu, 2013), system maintenance at planned downtime (Bardhan and Milojicic, 2012), load balancing (Ghomi et al., 2017; Genez et al., 2013), confidentiality protection (Ermakova and Fabian, 2013), and risk mitigation (AlZain et al., 2012; Ermakova and Fabian, 2013) The clients need real-time and high-quality solutions to economically automate cloud resource management in High-Performance Computing (HPC) environment. HPC is not only needed for public or private industries, but also for research and development. HPC is quite useful for time- and resource-intensive applications like eScience (Hoffa et al., 2008), data mining, and big data applications (Grossman and Gu, 2008). With the evolution of Big Data technologies, business managers are quite dependent on analytics to increase the return on investment. Some of the applications of analytics involve analysing online shopping behaviour in real-time, flight arrival time prediction (Tejasviram et al., 2015), fraud analytics (Kamaruddin and Ravi, 2016), click stream analytics (Chen et al., 2012), web-log mining, sentiment analysis (Ravi and Ravi, 2015), etc. Handling of Big Data needs a large amount of hardware and software services. In order to speed up the execution of Big Data applications, hardware and software service can be scaled up, which in turn need optimization of cost.

Management and utilization of the third-party infrastructure according to user's requirement are called IaaS. IaaS helps in reduction of time and cost. In order to reduce the time, on the one hand, the infrastructure can be scaled horizontally by using additional VMs (VMs). On the other hand, vertical scaling can be accomplished by allocating additional computing devices to the VMs. Increasing the number of VMs or computing devices can be referred as the number and size of VMs respectively. The execution cost and time for a cloud project depend on the type of VMs as well as the constellations of the VMs. Some other constraints like legal and compliance regulations of the client will also have a great impact on the cost. In most of the cases, the clients opt to the multi-cloud environment in order reduce the risk of being dependent on a single provider. Although IaaS provides a lot of opportunities to accomplish one's task in less time and cost, it poses a major challenge in the form of decision making from wide market availability. Decision making is needed in terms of available resources with respect to specific requirements. They need to maximize the value in terms of cost and service quality. Decision-making becomes so complex due to the availability of various resources and services provided by different cloud providers that are offering different features at different prices. Decision-making needs to consider one additional challenge like a variation of the requirements as well as prices with respect to time. This kind of variations can be managed slightly by allowing vertical and horizontal scaling. But, decision making regarding vertical/horizontal scaling again needs to be supported by cloud service provider. Generally, the threshold-based approach is adopted in current cloud service provider. In this approach, when the load is increased from the given threshold then load balancing can be adopted. In addition to that, we need to address three challenges like (i) re-deployment needs to consider individual features and prices of different VM types, (ii) optimization of resources according to the requirement of individual customers, and (iii) time-to-time comparison of prices of VM types and its features by different cloud providers. In all such circumstances, decision-making becomes so difficult due to having several conflicting goals. Therefore, in order to take the appropriate decision at the varying requirement, a brokerage mechanism is a suitable approach for CRMP.

Along this line, Coutinho et al. (2013) proposed the CRMP, which is considered as to maximize the value of cost and performance using Integer Programming (IP) model and referred as CC-IP fed. The CC-IP was extended to be implemented on the multi-cloud environment and called as Greedy Randomized Adaptive Search Procedure-CC-fed (GRASP-CC-fed) (Coutinho et al., 2015). The mathematical model considers the time and budget limits of consumers, different application requirements in terms of resource demands, and resource capabilities of available VM types provided by different cloud providers. Furthermore, internal and external communication costs are considered. Using CC-IP-fed or GraspCC-fed, however, is computationally expensive for multiple problem instances. Van den Bossche et al. (2010) considered maximization of internal computing resources usage and minimization of cost using an optimization problem formulation for hybrid clouds. Coutinho et al. (2013) formulated the CRMP as a cloud computing optimization problem in the multi-cloud environments, where they considered consumers and cloud brokers. In order to formulate an optimization problem for the CRMP, they needed to choose an appropriate combination of cloud resources from different cloud providers that satisfy application requirements, budget and time limitation, the costs of communication between different cloud providers, and the costs of communication between different applications on the same cloud environment, etc. Heilig et al. (2016) proposed Biased Random-Key Genetic Algorithm (BRKGA) to solve cloud resource management problem on a cloud brokerage mechanism using IaaS as depicted in Fig. 6. BRKGA works in the multi-cloud environments in order to reduce the monetary cost and the execution time of consumer applications. Furthermore, it helps in economically automating IaaS and corresponding deployment processes.

For online auction of IaaS, Shi et al. (2017) proposed two models viz. Social Welfare Maximizing Online Auction (SWMOA) and Provider Revenue Maximization Online Auction (PRMOA) for dynamic virtual cluster provisioning in geo-distributed clouds. They experimented with google cluster data (Reiss et al., 2011). The mechanisms are truthful, individually rational, time efficient, and guarantee a $O(2 \log \mu + 1)$ competitive ratio in social welfare and a $O(\log \mu)$ competitive ratio in provider revenue in expectation, respectively. Here, μ is the problem size.

5.5. Cloud security

Due to online availability of data, security becomes a very big challenge. There are several aspects of cloud computing, which needs to be taken care in security aspect. Some of the important security issues regarding cloud computing are reviewed in Khan (2016). They listed out issues and countermeasures related to four components of cloud viz. network, VM, storage, and applications. Network based attacks include port scanning, botnets, and spoofing. VM based components involve cross VM side channel attacks, VM creation attacks, VM migration and rollback attacks, and VM scheduler based attacks. Storage based attacks comprises data scavenging and data deduplication. Applications based attacks consists of malware injection and steganography attacks, shared architectures, and web services & protocol based attacks. Khan (2016) did not review log analytics, hence we mainly focused on cloud security using log analytics.

Every organization maintains a log file to record all activities related to all events occurred in the respective organization. The log analytics can be proved quite useful while dealing with the security problem. For an instance, one can analyse network log files for finding the usage of bandwidth of the network in a particular time span. Another application could be that an application developer can use the logs of applications for bug fixing in the code. One can also exploit the log files to identify malicious behaviour, to analyse the performance of a service, to optimize the resource usage, avoiding security threats, etc. Cloud Log Forensics (CLF) investigates the suspicious behaviours or attacks on the system by analysing logs of the system. Forensic investigators are also dependent on log files, which are maintained by cloud providers (Suleman et al., 2016). In the scarce of storage, an organization can opt for a Log-as-a-Service (LaaS), which is a cloud logging service. There are various cloud providers, who offer LaaS system like Papertrail, Logentries, Splunkstorm, Loggly, etc. IBM Smart Cloud analytics is a framework for analysing the logs. It uses IBM cloud infrastructure for analysing the operational data. It has been integrated with different kinds of sources. The IBM Smart Cloud analytics is helpful in detecting, analysing and solving operational data related issues. It is also helpful in the root-cause analysis as well as in the reduction of processing time. It implements the filtering, quick search, and visualization of data using a single interface. There are significant log services that perform accurate and quick analysis by integrating various logs. The IBM Smart Cloud analytics has many features like repair mean time reduction, dynamic warning message generation, service availability improvement, domain related issues categorization, results visualization, etc. Such features make the cloud provider as one of the leading log-as-a-service providers (Suleman et al., 2016).

Suleman et al. (2016) presented a case study of Banca Intesa in Serbia. They leveraged the system of log analysis for security mechanisms. The Banca Intesa bank has around 2 million Customers. It processes around 11 million events on a daily basis, which are generated by the network system, security systems, and the databases. Being a financial organization, the bank must be secured from various attacks that can exploit the client records. So, the bank subscribed a centralized log management service, which is helping them to analyse network activities and user related information. The bank wanted to investigate log information using root-cause analysis for suspicious activities and threats (Suleman et al., 2016). They used HP ArcSight logger for searching the logs for any activities, which might affect their bank's infrastructure. The HP ArcSight logger provides a wide-range of log forensic services to the bank. It helps in analysing different kinds of log files like real-time events log files, alert notifications, log information correlations, threat analysis, etc. The log forensics helped Banca Intesa Bank to access some crucial information like kind of data accessed, data accessed by, kind of actions performed on data, etc. The security investigators identified the perpetrators of attacks before they could damage the infrastructure of the bank. A strong security is provided by HP ArcSight logger for the bank, which helps in analysing the logs and finding the root cause of the attacks in the real-time mode (Suleman et al., 2016). In another study, Dayarathna et al. (2017) presented a knowledge graph based novel solution recommender system based log streams, where the knowledge graph is prepared using GraphX (Zaharia et al., 2011). They employed WSO2 (Inc, 2017a, 2017b) API manager to collect logs that are stored and indexed using NoSQL.

5.6. Others

Under this section, we reviewed six articles that employed analytics for optimization or prediction of related attributes of cloud computing. The reviewed six articles can be categorized as factors affecting cloud adoption (Sharma et al., 2016;), QoS value prediction (Ye et al., 2011; Ma et al., 2017), hard disk drive failure prediction (Vishwanath and Nagappan, 2010), location optimization for datacenter (Zhang et al., 2017) and ROI prediction (Wildstrom et al., 2008).

Although cloud computing is being a recent phenomenon, the research on the issues like adaption of cloud computing by both individual and organization is in the nascent stage. Most of the research in this area is performed from the organization's perspective. A very few research is carried out from an individual's perspective. Sharma et al. (2016) developed a hybrid model to predict the motivators influencing the cloud adaption of the cloud computing services by the IT professionals in the developing countries. Data collection is performed by survey questionnaire and is collected during March 2015 to May 2015. The analysis is performed on 101 useful responses. They proposed a new hybrid model by extending the Technology Adaption Model (TAM) with three external constructs self-efficacy, trust, and job opportunity. Collected data is analysed using Multiple Linear Regression (MLR) and neural networks. The main objective of the research was to develop and test empirical model for understanding the influence of various factors like determinants, job opportunity, self-efficacy, and trust on the willingness to adapt cloud computing services by the IT professionals in the developing countries. The neural networks outperformed MLR in terms of RMSE value.

There are two ways in which cloud services can be categorized viz. application services and utility computing services. Applications available on the internet are application-oriented services, e.g. rail ticket booking, hotel booking, etc. Utility computing services include software or virtualized hardware resources e.g. CPU, computing, and storage services. Compositions in the application are similar to the web service composition in Service Oriented Computing (SOC). Similarly, the composition at utility level is scheduling in grid and task matching. An extensive QoS model is proposed to calculate the QoS values of services in the cloud computing (Ye et al., 2011). A genetic algorithm based approach is proposed to compose services in the cloud computing. The proposed approach was compared with the exhaustive search algorithms and the random selection algorithms. In another study, Ma et al. (2017) proposed a multivalued collaborative approach to predict the unknown QoS value via time series analysis for potential users. Time series data measured in one time-slot and two time-slots for one user from consumers are transformed into cloud models (Deyi et al., 1995), which outperformed some existing models. Time series and multivalued both features of QoS are considered for this analysis. To detect exactly the neighboring users for potential users, a new vector comparison method combining the orientation similarity and dimension similarity to improve the accuracy of similarity calculation between cloud models and employed the Fuzzy Analytic Hierarchy Process (FAHP) method to determine the objective weights of periods according to application requirements of potential users. Experiments are performed on WS-DREAM dataset, which contains QoS data regarding real services ("WS-DREAM, 2016). The dataset contains response time and throughput of 4532 services provided by 142 users in 64 timeslots based on PlanetLab (Zhang et al., 2011). A detailed survey on web service composition can be found in (Alamri et al., 2006).

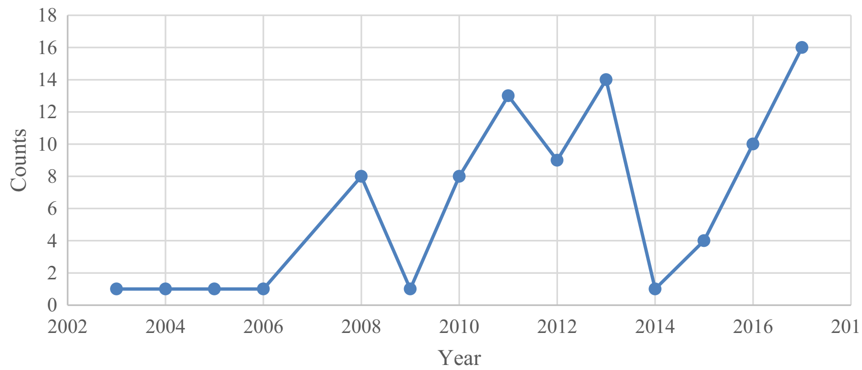

Fig. 7. A number of articles considered from 2003 to 2017.

A hard disk is an important hardware that is to be replaced because of its least reliability and high dominance (Vishwanath and Nagappan, 2010). In this study, the author found that 8% of all servers can lead to at least one failure in a year and will be more for hardware having multiple hard disks. The approximate cost for a datacenter is over a million dollars. In this study, they explored if the failures could be anticipated by using variables like environmental variables (datacenter name, manufacturer, time zone), design variables (location, a number of disks, slots the server has, physical memory) and operation metrics (OS version, last updated time, etc). They used more than 50 variables in order to build a CHi-squared Automatic Interaction Detector (CHAID) model. They observed two variables are very important namely datacenter name and manufacturer name as the CHAID reported them as statistically important variables (Vishwanath and Nagappan, 2010).

The VMs placement is an important factor, which needs to be considered for reducing the latency or communication gap as it affects QoS. In cloud computing, the datacenters play an important role, they need a very large space and distributed over various geographical regions. A cloud provider tries to minimize the distance between datacenters, so that the communication latency is decreased and bandwidth is minimized to reduce the cost. In order to optimize the inter-distance of datacenters, Zhang et al. (2017) proposed an efficient method called CBMinDia (ClusteringBasedMinDiameter), which is suitable for large scale problems. In the proposed cluster-based model, the time complexity is reduced drastically to $O(\mathit{log}n)$ from $O(n^2)$ (Zhang et al., 2017).

Wildstrom et al. (2008) proposed a Cognitive Agent for Resource Value Estimation (CARVE) for prediction return-on-investment on physical memory in a partitioned system running a multi-partition, multi-process distributed benchmark. M5′ tree of Weka tool (Hall et al., 2009) was employed in CARVE for prediction purpose. They reported that CARVE uses a low-level of statistics for making an informed decision. The evaluation is performed on TPC-W benchmark (TPC, 2003).

6. Discussion

In this survey, we considered eighty-eight articles published during 2003–2017 as depicted in Fig. 7. Fig. 8 presents the distribution of surveyed articles considered from different publishers. Maximum number of articles appeared in IEEE journals and conferences. Table 2 presents some of the prominent journals, in which reviewed articles appeared. Most of the articles appeared in the journal *Future Generation Computer Systems*. The distribution of articles with respect to different analytical techniques is presented in Fig. 9, which indicates that MLP and linear regression are the most employed techniques. Fig. 10 depicts the number of articles considered for review under different sub-topic of analytics for cloud. In the last few decades, the growth in cloud computing technology and adoption of VM-hosted applications is becoming increasingly important. Our survey indicates that analytics for cloud computing is a very fertile area for research and development. More specifically, demand prediction of cloud resources by cloud service provider is very important. In this study, we reviewed almost all the aspects of analytics and its association with cloud computing. Some of the observations as well as findings of this survey are as follows:

- With respect to dynamic demand, train once and predict anytime approach will not work in the real life. In order to overcome this issue, we need to develop a strategy, which will automatically upgrade the prediction model periodically. Dynamic demand can be predicted very well by the use of spiking neural networks, online algorithms such as general regression neural network etc. Another solution is that we need to develop a short-term predictor that predicts the future demand in terms of hours, whereas a long-term

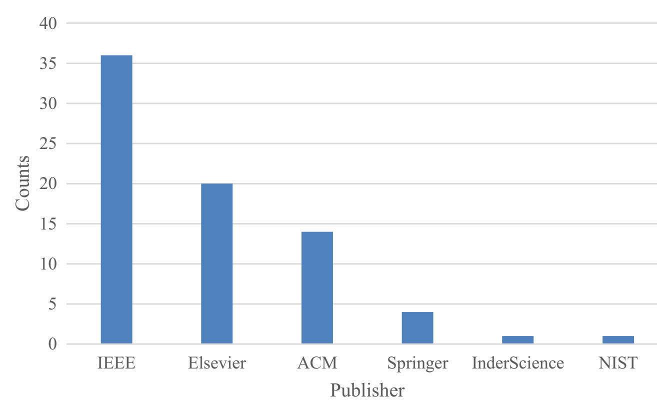

Fig. 8. A number of articles considered from different publishers.

Table 2
Number of papers reviewed from various journals.

Sl. No.	Journal Name	#Articles
1	Future Generation Computer Systems	9
2	Journal of Network and Computer Applications	4
3	ACM Transactions on Storage	2
4	IEEE Transactions on Service Computing	2
5	Information Systems	1
6	Vikalpa: The Journal for Decision Makers	1
7	Decision Support Systems	1
8	Journal of Parallel and Distributed Computing	1
9	Journal of Cloud Computing	1
10	Journal of Supercomputing	1
11	Computers & Industrial Engineering	1
12	Journal of Management Policy and Practice	1
13	ACM Computing Surveys	1
14	Computers in Human Behaviour	1
15	International Journal of Web and Grid Services	1
16	Computers & Operations Research	1
17	IEEE Transactions on Cloud Computing	1

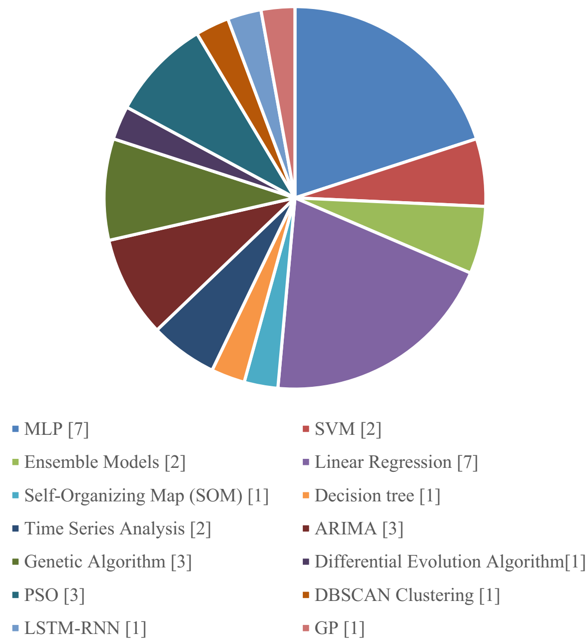

Fig. 9. Popular intelligent techniques applied in a number of articles.

predictor will predict the resource demand for the next week or month.

- Prediction of resource demand for long-term period will induce more uncertainty. In such cases non-parametric and probabilistic approaches can be proved quite useful as given in Minarolli et al. (2017).
- Cloud provisioning can be used to allocate resources to a client application in dynamic mode. However, dynamic provisioning suffers a lot due to lack of instantaneous demand of infrastructure. Moreover, the start-up time of the resources is another issue to be faced during dynamic provisioning. The start-up time will affect the quality of services of the cloud clients. Consequently, the companies may churn out from a cloud provider. In order to reduce the start-up time, an appropriate optimization technique can be employed, whereas churner can be predicted using analytics.
- According to our study, we observed that the number of study carried out for scaling of resources i.e. dynamic allocation are less than that of the static resource allocation. However, we need to focus more on scaling of resources to improve the adaptability in cloud environment.
- The problem of latency in VMs can also be addressed using binary classification task. The classification task will predict that whether the resource demand will increase in future or not. This kind of prediction framework can provision the resources in advance more accurately with the help of probabilistic neural network. This can be applied to multi-tenant service clouds also. Data mining techniques can extract high-level characteristics from historical demand, which can help in the allocation of resources in advance to the requesting application.
- Auto-scaling mechanism holds the promise of assuring QoS simultaneously maintaining the efficient use of the resources and keeping operational cost low. Despite the advantages of auto-scaling, the realization of its full potential is difficult due to various challenges like precisely estimating resource usage when the client workload pattern has significant variability.
- The prediction of resources by pattern matching will enable cloud providers to predict resources by using the past usage patterns in an auto-scaling approach, which indicates that effective cloud resource provisioning can be performed using pattern matching and predictive modelling approach.

- From this review, we can observe that the various kinds of cloud related issues can be addressed using a prediction model for the capacity planning of cloud resources, instant VM provisioning, etc. For prediction purpose, various forecasting techniques have been employed like linear regression, SVM, neural networks, exponential moving average, trend seasonality model, polynomial regression, etc. We observed that neural networks outperformed linear regression (Imam et al., 2011; Sharma et al., 2016).
- Comparing machine learning to optimization-based methods, most of the studies employed machine learning methods for resource demand prediction purpose.
- Fig. 7 helps us infer that the trend in research work with respect to analytics for cloud computing increased in the last five years. With the emerging demand of cloud in the forthcoming years, we may need more robust efficient approaches for prediction of cloud resources & demand respectively.
- From Table 3, it can be noticed that RUBiS and TPC-W benchmark were widely considered for the simulation and evaluation purpose.
- In order to improve upon the cloud service quality, Jiang et al. (2013b) proposed an ensemble of five prediction techniques viz. moving average, auto regression, MLP, SVM, and gene expression programming. Some other studies like (Imam et al., 2011; Roy et al., 2011), and (Guiyi et al., 2010) employed existing data mining techniques.
- Fig. 10 indicates that majority of the studies are carried out on resource allocation and scaling. The rest of area needs significant attention to improve upon existing methodologies.
- From customer's view point, cloud deployment using multi-providers is a cost effective approach, which needs effective analytical methods in order to utilize resources for varying needs of a customer.
- From broker's view point, analytical methods are quite helpful in trading cloud services from different service providers. Because there exists a fierce competition among service providers.
- A unique log based recommender solution is developed in Dayarathna et al. (2017). They exploited NoSQL and Apache Spark based GraphX for recording and developing recommender system.

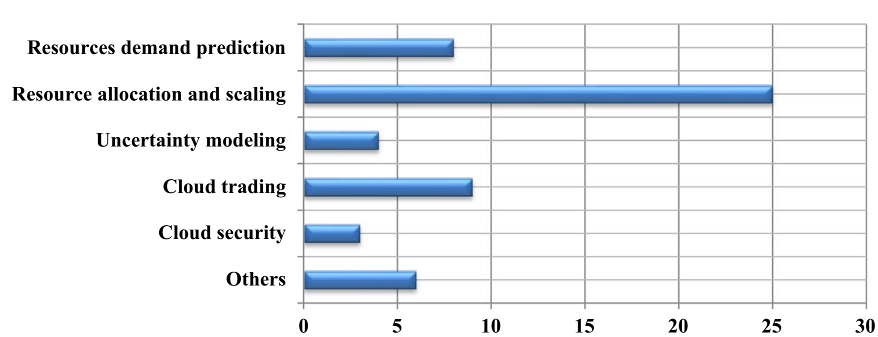

Fig. 10. A number of articles considered under the different category of Analytics for Cloud.

Table 3
Datasets analysed in different articles.

Articles	Datasets experimented with
(Imam et al., 2011)	Data From Large Hadron Collider Computing Grid (LCG) ("Worldwide Large Hadron Collider Grid," n.d.)
(Gong et al., 2010)	Google cluster data (Reiss et al., 2011)
(Zhong et al., 2017)	Google cluster data (Reiss et al., 2011)
(Kundu et al., 2012), (Gong et al., 2010)	RUBiS (Cecchet et al., 2002)
(Kundu et al., 2012)	Filebench (McDougall et al., 2011)
(Jiang et al., 2013b)	Trace log of IBM Smart Cloud Enterprise (IBM, 2011)
(Roy et al., 2011)	1998 Soccor world cup data (Roy et al., 2011)
(Jiang et al., 2013a)	Search log dataset of AOL, Sogou, and UTS library (Jiang et al., 2013)
(Liu et al., 2017)	Google cluster data (Reiss et al., 2011)
(Rogers and Cliff, 2012)	UK market data (Genez et al., 2013)
(Banditwattanawonga et al., 2016)	31-day BO trace was gathered from a user community in Boulder from 16th August to 15th September 2012, and a 31-day NY trace was collected from the other user community in New York from 16th July to 15th August 2012 (Banditwattanawonga et al., 2016)
(Saripalli et al., 2011)	Class Extract input is implemented to generate the input files from the database (Saripalli et al., 2011)
(Sadeka et al., 2012)	Simulation Data from TPC-W (TPC, 2003)
(Vishwanath and Nagappan, 2010)	Logs data of hardware repair incidents (Vishwanath and Nagappan, 2010)
(Zhang et al., 2017)	Google cluster data (Reiss et al., 2011)
(Minarolli et al., 2017)	Synthetic data using CloudSim simulator and PlanetLab workloads

7. Conclusion and future directions

In this paper, we presented a comprehensive review of eighty-eight papers appeared during the year 2003–2017 in the area of analytics deployed in the cloud & analytics used for solving operational and business problems in the cloud. This review brings forth the symbiotic relationship between cloud computing and analytics.. We categorized the studies into two important categories viz. *Analytics in Cloud* and *Analytics for Cloud*. *Analytics in cloud* becomes evident due to ever increase in the amount of data to be handled by an organization. In this regard, we reviewed the key stages of analytics workflows, different types of analytics offerings in cloud computing, predictive modelling in the cloud, and big data analytics in the cloud. *Analytics for cloud* is even more crucial than *analytics in cloud*. In order to efficiently manage cloud resources, predictive analytics and optimization techniques are found to be quite important for different aspects of cloud computing. We found some of the important aspects as a prediction of demand at individual's and provider's end, reservation of resources to effectively manage the dynamic demand, optimization of a pool of resources in broker environment, agent-based approach to handle dynamic demand, etc. We listed out approaches available to tackle issues with respect to different aspects of analytics for cloud. This survey can help a budding researcher to explore possible application areas of analytics with respect to cloud. This study also elaborates that prediction of demand of cloud computing resources is one of the most fertile research areas. Some of the future research directions include:

- There is a lot of scope for employing evolutionary techniques to optimize the usage of cloud resources in the multi-provider environment.
- Optimization of resources and price is very much needed for management and sale of cloud resources in the broker's perspectives too. There is no sufficient research available to predict the seasonal demand for cloud resources.
- Deep learning is becoming increasing popular in many domains. Similarly, GPU-based (CUDA) programming is also gaining a lot of traction among many researchers and practitioners. The new trend is to run powerful but time-consuming deep learning algorithms in GPU environment. Therefore, GPU based deep learning has immense potential to be placed in the cloud for the variety of reasons already discussed in the paper. This will have important ramifications in fields like banking and finance, bioinformatics, and a number of engineering disciplines.
- For optimization purpose, some of the applied techniques in the literature are PSO, IDEA, etc. Different forms of PSO to be used for task and workflow scheduling are reviewed in (Masdari et al., 2017). From the myriad of evolutionary techniques available in the literature (Krishna and Ravi, 2016), a very few have been tried for optimization purpose. In future, some powerful techniques can be exploited in order to better optimize the resources.
- With an ever increase in demand of cloud resources and market players, a multi-cloud solution becomes an evident approach. But, there is a need to develop efficient approaches to manage demand for consumers. Some of the proposed solutions in this regard are BRKGA, i-Cloud, OCRP, etc. The platform i-Cloud addressed the issue of network infrastructure optimization for enterprises. It reduces public cloud data-out expenses and optimizes network and various caching performances.
- To develop more robust approaches, a real-life dataset should be considered for experiments rather than synthetic datasets so, that the utility of work can be realized. Furthermore, there is an immense scope to develop some benchmark datasets in the field of multi-tenant service clouds.
- The provisioning of resources can be made efficient in multi-tenant service clouds using data mining techniques. In this environment, an agent-based simulation approach can also be employed, which can help in utilizing the infrastructure and can invest in expanding the infrastructure.
- Web application provider has the potential to perform scaling of VM resources in cloud computing environment. It is one of the advantages of the on-demand cloud environment. True elasticity and cost-effectiveness in the pay-per-use have not yet been achieved.
- Different kinds of uncertainty often occurs with respect to cloud usage and allocation. Uncertainty can be handled using unsupervised learning methods and outlier detection methods (Menasce et al., 2004).
- As clouds are becoming ubiquitous, we need a one-stop marketplace, which incorporates business intelligence and is useful for trading cloud service where the user can search and select the best-suited product. There is also security involved in the system of cloud services. A solution for identifying the malicious behaviour is by analysing the logs. There is some state of the art cloud log forensic services discussed here in the paper.
- There is a need for systems that predict the hardware failures. An organization may incur a huge loss due to hardware failure. Hence, there is a need for an automated solution to estimate the resource requirements, which may need due to shifting the native systems to a virtualized one.
- Chaos, neural networks and evolutionary computing based soft computing hybrids (Pradeepkumar and Ravi, 2017; Ravi et al., 2017) can be employed to predict the time series of demand and scaling requirements of resources more accurately.

- Dynamic resource scaling system is not able to scale proportionally with sudden traffic surge due to a special offer or market campaigns, which sometimes turnout to be catastrophic for the application performance. Pro-active prediction-based resource scaling is required for such kind of situations.
- For allocating VMs, the placement of datacenters spawning a large geographical area is an important problem. In order to optimize the inter-datacenter distance, a clustering mechanism can be employed. Further, energy optimization of such datacenters is a critical issue and can be solved effectively by a host of optimization algorithms.
- Logs are generated at a very fast pace, voluminous amount, and in unstructured format. Hence, big data analytics can be employed for analysis of logs, which will be used for security, troubleshooting, etc.
- In the reviewed literature, we observed that the studied related to security using log analytics are very limited. Hence, there are a lot of scope in this direction to exploit logs to overcome the security breach. As given in Dayarathna et al. (2017), developing log analytics based recommender solution is another research area, where researchers can contribute.
- We need to develop suitable methods for system reliability optimization, which can address issues related to cost, hardware failure, response time, etc.
- There is a symbiotic relation between analytics and cloud. The petabytes of data generated at Large Hadron Collider at CERN (the European Centre for Nuclear Research) can be placed on cloud in order to avoid any accident (Binning, n.d.). This strategy can be replicated in all those disciplines where such glut of all forms of data is prevalent.
- Agent-based automation is helpful to save time significantly. Hence, agent-based resource prediction need to be developed in future.
- There is an immense need to work on dynamic virtual cluster provisioning in future. This provisioning has two key challenges regarding computation and communication resources: (1) optimal placement of virtual clusters and inter-VM traffic routing, which is an NP-hard problem; (2) there is very much need of an efficient and adaptive pricing mechanism.
- Electronic word of mouth, in the form of customer reviews, about the quality of cloud providers are available online. Sentiment analysis performed on these reviews together with fuzzy multi criteria decision making new customers in selecting the most suitable cloud provider (Timmaraju et al., 2017). A lot of scope exists in this area, where aspect-oriented sentiment analysis can be performed.

References

Aho, T., Ashraf, A., Englund, M., Katajamäki, J., Koskinen, J., Lautamäki, J., Nieminen, A., Porres, I., Turunen, I., 2011. Designing IDE as a service. Commun. Cloud Softw. 1, 1–10.

Alamri, A., Eid, M., Saddik, A., et al., 2006. Classification of the state-of-the-art dynamic web services composition techniques. Int. J. Web Grid Serv. 2, 148–166.

AlZain, M., Pardede, E., Soh, B., Thom, J., 2012. Cloud computing security: From single to multi-clouds, In: Proceedings of the 45th Hawaii International Conference on System Science. IEEE, pp. 5490–5499. doi:10.1109/HICSS.2012.153.

Amiri, M., Mohammad-Khanli, L., 2017. Survey on prediction models of applications for resources provisioning in cloud. J. Netw. Comput. Appl. 82, 93–113. http://dx.doi.org/10.1016/j.jnca.2017.01.016.

Ashraf, A., Byholm, B., Porres, I., 2016. Prediction-based VM provisioning and admission control for multi-tier web applications. J. Cloud Comput. 5, 15. http://dx.doi.org/10.1186/s13677-016-0065-9.

Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A.S., Buyya, R., 2015. Big Data computing and clouds: trends and future directions. J. Parallel Distrib. Comput. 79, 3–15.

Baaren, V., Erik-Jan,, 2009. Wikibench: A Distributed, Wikipedia Based Web Application Benchmark. VU University, Amsterdam.

Banditwattanawonga, T., Masdisornchote, M., Uthayopas, P., 2016. Multi-provider cloud computing network infrastructure optimization. Future Gener. Comput. Syst. 55, 116–128.

Banerjee, A., Bandyopadhyay, T., Acharya, P., 2013. Data analytics: hyped up aspirations or true potential? Vikalpa 38, 1–12.

Bardhan, S., Milojicic, D., 2012. A mechanism to measure quality-of-service in a federated cloud environment, In: Proceedings of Workshop on Cloud Services, Federation, and the 8th Open Cirrus Summit. ACM, pp. 19–24. doi:10.1145/2378975.2378981.

Bennani, M.N., Menasce, D.A., 2005. Resource allocation for autonomic data centers using analytic performance models, In: Proceedings of the Second International Conference on Autonomic Computing. pp. 229–240.

Binning, D., n.d. ⟨http://www.computerweekly.com/feature/CERNs-LHC-pioneers-quantum-leap-in-cloud-computing⟩ [WWW Document].

Calheiros, R.N., Masoumi, E., Ranjan, R., Buyya, R., 2015. Workload prediction using ARIMA model and its impact on cloud applications' QoS. IEEE Trans. Cloud Comput. 3, 449–458.

Caron, E., Desprez, F., Muresan, A., 2010. Forecasting for grid and cloud computing on-demand resources based on pattern matching, In: Proceedings of International Conference on Cloud Computing Technology and Science. IEEE, pp. 456–463. doi:10.1109/CloudCom.2010.65.

Cecchet, E., Marguerite, J., Zwaenepoel, Willy, 2002. Performance and Scalability of EJB Applications, In: Proceedings of the OOPSLA'02. Seattle, WA, USA.

Chaisiri, S., Lee, B.-S., Niyato, D., Wei, G., Vasilakos, A.V., Zheng, Y., Xiong, N., Tsai, J.-T., Fang, J.-C., Chou, J.-H., Sharma, S.K., Al-Badi, A.H., Govindaluri, S.M., Al-Kharusi, M.H., Jiang, Y., Perng, C.S., Li, T., Chang, R.N., Islam, S., Keung, J., Lee, K., Liu, A., Chang, V., Ramachandran, M., Yao, Y., Kuo, Y.H., Li, C.-S., Chaisiri, S., Lee, B.-S., Niyato, D., Ashraf, A., Byholm, B., Porres, I., 2012. Optimization of resource provisioning cost in cloud computing. IEEE Trans. Serv. Comput. 5, 164–177. http://dx.doi.org/10.1016/j.ijinfomgt.2015.09.008.

Chen, H., Chiang, R.H., Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. MIS Q. 36, 1165–1188.

Chou, L.-D., Chen, H.-F., Tseng, F.-H., Chao, H.-C., Chang, Y.-J., 2016. DPRA: dynamic power-saving resource allocation for cloud data center using particle swarm optimization. IEEE Syst. J., 1–12. http://dx.doi.org/10.1109/JSYST.2016.2596299.

Columbus, L., 2016. 5 Cloud Service Provider Predictions For 2016: Analytics, CRM & CPQ Accelerate Sales [WWW Document]. Forbes. URL ⟨http://www.forbes.com/sites/louiscolumbus/2016/01/30/5-cloud-service-provider-predictions-for-2016-analytics-crm-cpq-accelerate-sales/⟩.

Coutinho, R.deC., Drummond, L.M., Frota, Y., 2013. Optimization of a cloud resource management problem from a consumer perspective, In: Proceedings of European Conference on Parallel Processing. pp. 218–227.

Coutinho, R., de, C., Drummond, L.M., Frota, Y., de Oliveira, D., 2015. Optimizing virtual machine allocation for parallel scientific workflows in federated clouds. Future Gener. Comput. Syst. 46, 51–68. http://dx.doi.org/10.1016/j.future.2014.10.009.

da Rosa Righi, R., Rodrigues, V.F., Rostirolla, G., André da Costa, C., Roloff, E., Navaux, P.O.A., 2017. A lightweight plug-and-play elasticity service for self-organizing resource provisioning on parallel applications. Future Gener. Comput. Syst., (doi:10.1016/j.future.2017.02.023).

Dashti, S.E., Rahmani, A.M., 2016. Dynamic VMs placement for energy efficiency by PSO in cloud computing. J. Exp. Theor. Artif. Intell. 28, 97–112.

Data analysis [WWW Document], n.d. URL ⟨https://en.wikipedia.org/wiki/Data_analysis⟩.

Dayarathna, M., Akmeemana, P., Perera, S., Jayasinghe, M., 2017. Solution recommender for system failure recovery via log event pattern matching on a knowledge graph. In: Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems 331–334.

Dean, D.J., Nguyen, H., Gu, X., 2012. UBL: Unsupervised behavior learning for predicting performance anomalies in virtualized cloud systems, In: Proceedings of the 9th International Conference on Autonomic Computing. ACM, pp. 191–200. doi:10.1145/2371536.2371572.

Demirkan, H., Delen, D., 2013. Leveraging the capabilities of service-oriented decision support systems: putting analytics and big data in cloud. Decis. Support Syst. 55, 412–421.

Deyi, L., Haijun, M., Xuemei, S., 1995. Membership clouds and membership cloud generators. J. Comput. Res. Dev. 32, 15–20.

Díaz, J.L., Entrialgo, J., García, M., García, J., García, D.F., 2017. Optimal allocation of virtual machines in multi-cloud environments with reserved and on-demand pricing. Future Gener. Comput. Syst. 71, 129–144.

Dutreilh, X., Kirgizov, S., Melekhova, O., Malenfant, J., Rivierre, N., Truck, I., 2011. Using reinforcement learning for autonomic resource allocation in clouds: towards a fully automated workflow, In: Proceedings of the Seventh International Conference on Autonomic and Autonomous Systems. Venice, Itlay, pp. 67–74.

Dutreilh, X., Moreau, A., Malenfant, J., Rivierre, N., Truck, I., 2010. From data center resource allocation to control theory and back, In: Proceedings of the 3rd International Conference on Cloud Computing (CLOUD). IEEE, pp. 410–417.

Ermakova, T., Fabian, B., 2013. Secret sharing for health data in multi-provider clouds, In: Proceedings of the 15th Conference on Business Informatics. IEEE, pp. 93–100.

Farah, B.N., 2013. A model for managing uncertainty on the cloud. J. Manag. Policy Pract. 14, 19–24.

Genez, T., Bittencourt, L., Madeira, E., 2013. Using time discretization to schedule scientific workflows in multiple cloud providers, In: Proceedings of the Sixth International Conference on Cloud Computing. IEEE, pp. 123–130.

Ghomi, E.J., Rahmani, A.M., Qader, N.N., 2017. Load-balancing algorithms in cloud computing: a survey. J. Netw. Comput. Appl. 88, 50–71.

Gong, Z., Gu, X., Wilkes, J., 2010. PRESS: Predictive elastic resource scaling for cloud systems, In: Proceedings of International Conference on Network and Service Management. IEEE, pp. 9–16. doi:10.1109/CNSM.2010.5691343.

Grande, A., And, Boukerche, R.E., A, 2011. Predictive dynamic load balancing for large-scale hla-based simulations, In: Proceedings of the 15th International Symposium on Distributed Simulation and Real Time Applications, IEEE/ACM, IEEE Computer Society, pp. 4–11.

Grossman, R., Gu, Y., 2008. Data mining using high performance data clouds: Experimental studies using sector and sphere, In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, Las Vegas, NV, USA, pp. 920–927. doi:10.1145/1401890.1402000.
Guiyi, W., Zheng, Athanasios V., Vasilakos, Y., Xiong, N., Wei, G., Vasilakos, A.V., Zheng, Y., Xiong, N., 2010. A game-theoretic method of fair resource allocation for cloud computing services. J. Supercomput. 54, 252–269. http://dx.doi.org/10.1007/s11227-009-0318-1.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explor. Newsl. 11, 10–18.
Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U., 2015. The rise of "big data" on cloud computing: review and open research issues. Inform. Syst. 47, 98–115.
Heilig, L., Lalla-Ruiz, E., Voß, S., 2016. A cloud brokerage approach for solving the resource management problem in multi-cloud environments. Comput. Ind. Eng. 95, 16–26. http://dx.doi.org/10.1016/j.cie.2016.02.015.
Hoffa, C., Mehta, G., Freeman, T., Deelman, E., Keahey, K., Berriman, B., Good, J., 2008. On the use of cloud computing for scientific workflows, In: Proceedings of the 4th International Conference on eScience (eScience). IEEE, Indianapolis, IN, USA, pp. 640–645. doi:10.1109/eScience.2008.167.
Foster, I., Zhao, Lu, S, Y., 2008. Cloud computing and grid computing 360-degree compared, In: Proceedings Grid Computing Environments Workshop (GCE '08).
IBM, 2011. IBM Smart Cloud Enterprise (SCE) trace data [WWW Document].
Imam, M.T., Miskhat, S.F., Rahman, R.M., Amin, M.A., 2011. Neural network and regression based processor load prediction for efficient scaling of Grid and Cloud resources, In: Proceedings of International Conference on Computer and Information Technology. IEEE, pp. 333–338. doi:10.1109/ICCITechn.2011.6164809.
Inc., W., 2017a. Analyzing logs with the log analyzer.
Inc., W., 2017b. Smart Analytics.
James, T., n.d. Predictive analytics in the cloud - decision management solutions [WWW Document]. URL ⟨http://www.decisionmanagementsolutions.com/predictive-analytics-solutions-in-the-cloud/⟩ (accessed 3 September 2017).
Jiang, J., Lu, J., Zhang, G., Long, G., 2013a. Optimal cloud resource auto-scaling for web applications, In: Proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing. IEEE, pp. 58–65.
Jiang, Y., Perng, C.-S., Li, T., Chang, R.N., 2013b. Cloud analytics for capacity planning and instant VM provisioning. IEEE Trans. Netw. Serv. Manag. 10, 312–325. http://dx.doi.org/10.1109/TNSM.2013.051913.120278.
Jiang, Y., Perng, C.S., Li, T., Chang, R., 2011. ASAP: A self-adaptive prediction system for instant cloud resource demand provisioning, In: Proceedings of IEEE International Conference on Data Mining. IEEE, pp. 1104–1109. doi:10.1109/ICDM.2011.25.
Kamaruddin, S., Ravi, V., 2016. Credit card fraud detection using big data analytics: Use of PSOAANN based one-class classification, In: Proceedings of the International Conference on Informatics and Analytics. ACM press, Puducherry, pp. 1–8.
Khan, M.A., 2016. A survey of security issues for cloud computing. J. Netw. Comput. Appl. 71, 11–29. http://dx.doi.org/10.1016/j.jnca.2016.05.010.
Kobielus, J., 2011. Predictions and plans for business analytics in 2011 [WWW Document]. URL http://blogs.forrester.com/james_kobielus/11-01-06-predictions_and_plans_for_business_analytics_in_2011.
Krishna, G.J., Ravi, V., 2016. Evolutionary computing applied to customer relationship management: a survey. Eng. Appl. Artif. Intell. 56, 30–59.
Kundu, S., Rangaswami, R., Gulati, A., Zhao, M., Dutta, K., 2012. Modeling virtualized applications using machine learning techniques. ACM SIGPLAN Not., 3. http://dx.doi.org/10.1145/2365864.2151028.
Kurdi, H., Li, M., Al-Raweshidy, H.S., 2010. Taxonomy of Grid Systems. In: Antonopoulos, Nick, Exarchakos, George, Li, Maozhen, Liotta, Antonio (Eds.), Handbook of Research on P2P and Grid Systems for Service-oriented Computing: Models, Methodologies and Applications. IGI Global.
Liu, C., Liu, C., Shang, Y., Chen, S., Cheng, B., Chen, J., 2017. An adaptive prediction approach based on workload pattern discrimination in the cloud. J. Netw. Comput. Appl. 80, 35–44. http://dx.doi.org/10.1016/j.jnca.2016.12.017.
Ma, H., Zhu, H., Hu, Z., Tang, W., Dong, P., 2017. Multi-valued collaborative QoS prediction for cloud service via time series analysis. Future Gener. Comput. Syst. 68, 275–288.
Mahmud, S., Iqbal, R., Doctor, F., 2016. Cloud enabled data analytics and visualization framework for health-shocks prediction. Future Gener. Comput. Syst. 65, 169–181. http://dx.doi.org/10.1016/j.future.2015.10.014.
Mao, M., Humphrey, M., 2011. Auto-scaling to minimize cost and meet application deadlines in cloud workflows, In: Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis (SC). IEEE, pp. 1–12. doi:10.1145/2063384.2063449.
Mark, C.C.T., Niyato, D., Chen-Khong, T., 2011. Evolutionary optimal virtual machine placement and demand forecaster for cloud computing, In: Proceedings of International Conference on Advanced Information Networking and Applications. IEEE, pp. 348–355. doi:10.1109/AINA.2011.50.
Masdari, M., Salehi, F., Jalali, M., Bidaki, M., 2017. A survey of PSO-based scheduling algorithms in cloud computing. J. Netw. Syst. Manag. 25, 122–158.
McDougall, R., Nadgir, N., Debnath, S., Shepler, S., Wilson, A., Kay, D., Mauro, J., Schmidt, M., Debnath, S., Kustarz, E., Rockwood, B., 2011. Filebench: a Framew. Simula. Appl. file Syst..
Mell, P., Grance, T., 2011. The NIST definition of cloud computing.
Menasce, D.A., Almeida, V.A., Dowdy, L.W., Dowdy, L., 2004. Performance by design: computer capacity planning by example. Prentice Hall Professional.
Menychtas, A., Vogel, J., Giessmann, A., Gatzioura, A., Gomez, S.G., Moulos, V., Junker, F., Müller, M., Kyriazis, D., Stanoevska-Slabeva, K., Varvarigou, T., 2014. 4caast marketplace: an advanced business environment for trading cloud services. Future Gener. Comput. Syst. 41, 104–120.
Minarolli, D., Mazrekaj, A., Freisleben, B., 2017. Tackling uncertainty in long-term predictions for host overload and underload detection in cloud computing. J. Cloud Comput. 6, 4. http://dx.doi.org/10.1186/s13677-017-0074-3.
Mozafari, B., Curino, C., Jindal, A., Madden, S., 2013a. Performance and resource modeling in highly-concurrent OLTP workloads, In: Proceedings of the ACM Sigmod International Conference on Management of Data. ACM, pp. 301–312.
Mozafari, B., Curino, C., Madden, S., 2013b. DBSeer: Resource and performance prediction for building a next generation database cloud, In: Proceedings of Conference on Innovative Data Systems Research (CIDR).
Neto, P., 2011. Demystifying cloud computing, In: Proceeding of Doctoral Symposium on Informatics Engineering.
Pandey, S., Wu, L., Guru, S.M., Buyya, R., 2010. A particle swarm optimization based heuristic for scheduling workflow applications in cloud computing environments, In: Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications. IEEE, Perth, WA, USA, pp. 400–407.
Petcu, D., 2013. Multi-cloud: Expectations and current approaches, In: Proceedings of International Workshop on Multi-Cloud Applications and Federated Clouds. ACM, New York, NY, USA, pp. 1–6.
Pradeepkumar, D., Ravi, V., 2017. Forecasting financial time series volatility using Particle Swarm Optimization trained Quantile Regression Neural Network. Appl. Soft Comput. 58, 35–52.
Predictive analytics [WWW Document], n.d. URL ⟨https://en.wikipedia.org/wiki/Predictive_analytics⟩.
Rahman, M., Li, X., Palit, H., 2011. Hybrid heuristic for scheduling data analytics workflow applications in hybrid cloud environment, In: Proceedings of Parallel and Distributed Processing Workshops and PhdForum (IPDPSW). IEEE, pp. 966–974.
Rankothge, W., Le, F., Russo, A., Lobo, J., 2017. Optimizing resource allocation for virtualized network functions in a cloud center using genetic algorithms. IEEE Trans. Netw. Serv. Manag. 14, 343–356.
Ravi, K., Ravi, V., 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl.-Based Syst. 89, 14–46.
Ravi, V., Pradeepkumar, D., Deb, K., 2017. Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms. Swarm Evolut. Comput. 35, 136–149.
Reiss, C., Wilkes, J., Hellerstein, J.L., 2011. Google Cluster-Usage Traces: Format+ Schema. Google Inc. White Paper, Mountain View, CA, USA.
Rogers, O., Cliff, D., 2012. Forecasting Demand for Cloud Computing Resources-An Agent-based Simulation of a Two Tiered Approach. ICAART, 106–112.
Roy, N., Dubey, A., Gokhale, A., 2011. Efficient autoscaling in the cloud using predictive models for workload forecasting. In: Proceedings of International Conference on Cloud Computing (CLOUD) 500–507.
Sadeka, I., Keunga, J., Lee, K., Liu, A., 2012. Empirical prediction models for adaptive resource provisioning in the cloud. Future Gener. Comput. Syst. 28, 155–162.
Saripalli, P., Kiran, G., Shankar, R.R., Narware, H., Bindal, N., 2011. Load prediction and hot spot detection models for autonomic cloud computing, In: Proceedings of International Conference on Utility and Cloud Computing. IEEE, pp. 397–402. doi:10.1109/UCC.2011.66.
Shahin, A.A., 2016. Automatic cloud resource scaling algorithm based on long short-term memory recurrent neural network. Int. J. Adv. Comput. Sci. Appl. 7, 279–285.
Sharma, S.K., Al-Badi, A.H., Govindaluri, S.M., Al-Kharusi, M.H., 2016. Predicting motivators of cloud computing adoption: a developing country perspective. Comput. Human. Behav. 62, 61–69. http://dx.doi.org/10.1016/j.chb.2016.03.073.
Shi, W., Wu, C., Li, Z., 2017. An online auction mechanism for dynamic virtual cluster provisioning in geo-distributed clouds. IEEE Trans. Parallel Distrib. Syst. 28, 677–688.
Sobel, W., Subramanyam, S., Sucharitakul, A., Nguyen, J., Wong, H., Klepchukov, A., Patil, S., Fox, A., Patterson, D., 2008. Cloudstone: Multi-platform, multi-language benchmark and measurement tools for web 2.0, in: Proceedings of CCA.
Sudheesh Narayanan, 2013. Cloud and Big Data: Is there a Relation between the two [WWW Document]. URL ⟨http://www.infosysblogs.com/cloud/2013/05/cloud_and_big_data_is_there_a_html⟩.
Suleman, K., Gani, A., Wahab, A.W.A., Bagiwa, M.A., Shiraz, M., Khan, S.U., Buyya, R., Zomaya, A.Y., 2016. Cloud log forensics: foundations, state of the art, and future directions. ACM Comput. Surv. (CSUR) 49, 7. http://dx.doi.org/10.1145/2906149.
Talia, D., 2013. Clouds for scalable big data analytics. Computer 46, 98–101.
Tejasviram, V., Solanki, H., Ravi, V., Kamaruddin, S., 2015. Auto associative extreme learning machine based non-linear principal component regression for big data applications, In: Proceedings of International Conference on Digital Information Management (ICDIM). IEEE, South Korea, pp. 223–228. doi:10.1109/ICDIM.2015.7381854.
Timmaraju, S., Ravi, V., Gangadharan, G.R., 2017. Ranking of cloud services using opinion mining and multi-attribute decision making. In: Trivedi, S.K., Dey, S., Kumar, A., Panda, T.K. (Eds.), Handbook of Research on Advanced Data Mining Techniques and Applications for Business Intelligence. IGI Global, USA, 379–396.
TPC, 2003. TPC-W benchmark. Transaction Processing Performance Council (TPC).
Tran, D.N., Huynh, P.C., Tay, Y.C., Tung, A.K., 2008. A new approach to dynamic self-tuning of database buffers. ACM Trans. Storage (TOS) 4, 3.
Tsai, J.-T., Fang, J.-C., Chou, J.-H., 2013. Optimized task scheduling and resource allocation on cloud computing environment using improved differential evolution algorithm. Comput. Oper. Res. 40, 3045–3055. http://dx.doi.org/10.1016/j.cor.2013.06.012.
Van den Bossche, R., Vanmechelen, K., Broeckhove, J., 2010. Cost-optimal scheduling in hybrid iaas clouds for deadline constrained workloads, In: Proceedings of the 3rd International Conference on Cloud Computing. IEEE, pp. 228–235. doi:10.1109/

CLOUD.2010.58.
Verma, M., Gangadharan, G.R., Ravi, V., Narendra, N., 2013. Resource demand prediction in multi-tenant service clouds, In: Proceedings of IEEE International Conference on Cloud Computing in Emerging Markets (CCEM). , pp. 1–8.
Vishwanath, K.V., Nagappan, N., 2010. Characterizing cloud computing hardware reliability, In: Proceedings of the 1st ACM Symposium on Cloud Computing. ACM, pp. 193–204.
Weins, K., 2016. Cloud Computing Trends: 2016 State of the Cloud Survey [WWW Document]. URL ⟨http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2016-state-cloud-survey⟩.
Wildstrom, J., Stone, P., Witchel, E., 2008. CARVE: A cognitive agent for resource value estimation, In: Proceedings of International Conference on Automonic Computing. IEEE, pp. 182–191.
Winter, R., 2008. Ebay turns to analytics as a service. Intell. Enterp., 5.
Wood, T., Cherkasova, L., Ozonat, K., Shenoy, P., 2008. Predicting application resource requirements in virtual environments.
Worldwide Large Hadron Collider Grid [WWW Document], n.d. URL ⟨http://wlcg.web.cern.ch/⟩.
Wright, P., Harmer, T., Hawkins, J., Sun, Y.L., 2011. A commodity-focused multi-cloud marketplace exemplar application, In: IEEE International Conference on Cloud Computing. IEEE, pp. 590–597.
WS-DREAM, 2016. [WWW Document]. URL ⟨http://wsdream.com/dataset.html⟩.
Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F., 2017. Big Data and cloud computing: innovation opportunities and challenges. Int. J. Digit. Earth 10, 13–53. http://dx.doi.org/10.1080/17538947.2016.1239771.
Ye, Z., Zhou, X., Bouguettaya, A., 2011. Genetic algorithm based QoS-aware service compositions in cloud computing, In: International Conference on Database Systems for Advanced Applications. Springer, HongKong, pp. 321–334. doi:10.1007/978-3-642-20152-3_24.
Zaharia, M., Xin, R.S., Wendell, P., Das, T., Armbrust, M., Dave, A., Al, X.M., et al., 2011. Apache spark: a unified engine for big data processing. Commun. ACM 59, 56–65.
Zhang, J., Wang, X., Huang, H., Chen, S., 2017. Clustering based VMs placement in distributed cloud computing. Future Gener. Comput. Syst. 66, 1–10. http://dx.doi.org/10.1016/j.future.2016.06.018.
Zhang, X., Huang, Z., Wu, C., Li, Z., Lau, F., 2017. Online auctions in IaaS clouds: welfare and profit maximization with server costs. IEEE/ACM Trans. Netw. (TON) 25, 1034–1047.
Zhang, Y., Zheng, Z., Lyu, M.R., 2011. WSPred: A time-aware personalized QoS prediction framework for Web services, In: Proceedings of the 22nd International Symposium on In Software Reliability Engineering (ISSRE). IEEE, pp. 210–219.
Zhao, L., Lu, L., Jin, Z., Yu, C., 2017. Online VM placement for increasing cloud provider's revenue. IEEE Trans. Serv. Comput. 10, 273–285.
Zhao, Y., Calheiros, R.N., Gange, G., Ramamohanarao, K., Buyya, R., 2015. SLA-based resource scheduling for big data analytics as a service in cloud computing environments, In: Proceedings of the 44th International Conference on Parallel Processing (ICPP). IEEE, pp. 510–519.
Zhong, W., Zhuang, Y., Sun, J., Gu, J., 2017. The cloud computing load forecasting algorithm based on wavelet support vector machine, In: Proceedings of the Australasian Computer Science Week Multiconference ACSW '17. ACM Press, New York, New York, USA, pp. 1–5. doi:10.1145/3014812.3014852.

Kumar Ravi is a Ph. D. Scholar at Institute for Development and Research in Banking Technology, Hyderabad and University of Hyderabad. He did his M. Tech. (CSE) from Pondicherry University, Pondicherry in 2013. He published 5 journals and 5 conference papers. He has 218 citations for his publications. One of his articles is listed among the top 25 hottest articles in Knowledge-Based Systems. His research area of interests are machine learning, social media analytics, natural language processing, sentiment analysis, formal concept analysis, ontological engineering, and big data analytics.

Yogesh Khandelwal is currently working as a data scientist with Tata Technologies, Mumbai. Earlier, he was associated with IDRBT, Hyderabad and was working as a research associate under Prof. V.Ravi. He completed M.Tech in future studies and planning and B.E. in computer engineering from Devi Ahilya Vishwavidyalaya, Indore (M.P). His major area of interest lies in CRM analytics, Text analytics and Business data analytics.

Boora Shiva Krishna is working as a Research Associate in IDRBT under Prof. Vadlamani Ravi in Center of Excellence in Analytics since March 2015. He has been working on different kinds of banking related business problems like Loan Defaulter Prediction, Customer Churn Prediction, Market Basket Analysis, RFM Analysis, Customer Segmentation, Social Media Analytics, and Big Data Analytics. He was involved in training programmes conducted for various public and private sector banks in India. He participated in FIRE 2015 task named Mixed Script Information Retrieval conducted by Microsoft. He also co-authored in "Language Identification in Mixed Script Social Media Text, FIRE-2017, IISC Bangalore". His research interests include Text mining and Big Data analytics. He did his B. Tech. (CSE) from Jyothishmathi Institute of Technology and Science, Karimnagar.

Dr. Vadlamani Ravi is the Professor at the Institute for Development and Research in Banking Technology, Hyderabad since June 2014. He obtained his Ph.D. in the area of Soft Computing from Osmania University, Hyderabad and RWTH Aachen, Germany (2001); MS (Science and Technology) from BITS, Pilani (1991) and M.Sc. (Statistics & Operations Research) from IIT, Bombay (1987). At IDRBT, he spearheads the Center of Excellence in Analytics, first-of-its-kind in India and evangelizes Analytical CRM and Non-CRM related analytics in a big way by conducting customized training programmes for bankers on OCRM & ACRM; Data Warehousing, Data and Text Mining, Big Data Analytics, Fraud Analytics, Risk Analytics, Social Media Analytics, Credit Recovery Analytics, Business Analytics and conducting POC for banks etc. He has 193 papers to his credit with the break-up of 82 papers in refereed International Journals, 6 papers in refereed National Journals, 83 papers in refereed International Conferences and 3 papers in refereed National Conferences and 19 invited book chapters. His papers appeared in Applied Soft Computing, Soft Computing, Asia-Pacific Journal of Operational Research, Decision Support Systems, European Journal of Operational Research, Expert Systems with Applications, Engineering Application of Artificial Intelligence, Fuzzy Sets and Systems, IEEE Transactions on Fuzzy Systems, IEEE Transactions on Reliability, Information Sciences, Journal of Systems and Software, Knowledge Based Systems, Neurocomputing,IJUFKS, IJCIA, IJAEC, IJDMMM, IJIDS, IJDATS, IJISSS, IJECRM, IJISSC, IJCIR, IJCISIM, IJBIC, JIPS, Computers and Chemical Engineering, Canadian Geotechnical Journal, Biochemical Engineering Journal, Computers in Biology and Medicine, Applied Biochemistry and Biotechnology, Bioinformation, Journal of Services Research etc. He also edited a Book entitled "Advances in Banking Technology and Management: Impacts of ICT and CRM" (http://www.igi-global.com/reference/ details.asp?id=6995), published by IGI Global, USA, 2007. Some of his research papers are listed in Top 25 Hottest Articles by Elsevier and World Scientific. He has an H-index of 31 and more than 4121 citations for his papers (http://scholar.google.co.in/). His profile was among the Top 10% Most Viewed Profiles in LiknkedIn in 2012. He is recognized as a Ph.D. supervisor at Department of Computer and Information Sciences, University of Hyderabad and Department of Computer Sciences, Berhampur University, Orissa. He is an invited member in Marquis Who's Who in the World, USA in 2009, 2015. He is also an invited member in 2000 Outstanding Intellectuals of the 21st Century 2009/2010- published by International Biographical Center, Cambridge, England, He is an Invited Member of "Top 100 Educators in 2009" published by International Biographical Center, Cambridge, England. So far, 3 Ph.D. students graduated under his supervision, 1 submitted the thesis, and 4 more are currently working towards Ph.D. So far, he advised more than 50 M.Tech./MCA/M.Sc projects and 20 Summer Interns from various IITs. He currently supervises 3 M.Tech students. He is on the IT Advisory Committee of Canara Bank for their DWH and CRM project; IT Advisor for Indian Bank for their DWH and CRM project and Principal Consultant for Bank of India for their CRM project; Expert Committee Member for IRDA for their Business Analytics and Fraud Analytics projects. He is a referee for 40 International Journals of repute. Moreover, he is a member of the Editorial Review Board for the "International Journal of Information Systems in Service Sector" published by IGI Global, USA; "International Journal of Data Analysis Techniques and Strategies" published by Inderscience, Switzerland; International Journal of Information and Decision Sciences (IJIDS), Inderscience, Switzerland; International Journal of Strategic Decision Sciences (IJSDS), IGI Global, USA and International Journal of Information Technology Project Management (IJITPM), IGI Global, USA. International Journal of Data Science (IJDS), Inderscience, Switzerland; Editorial Board Member for Book Series in Banking, Inderscience Switzerland, He is on the PC for some International Conferences and chaired many sessions in International Conferences in India and abroad. His research interests include Fuzzy Computing, Neuro Computing, Soft Computing, Evolutionary Computing, Data Mining, Text Mining, Web Mining, Big Data, Privacy Preserving Data Mining, Global/Multi-Criteria/Combinatorial Optimization, Bankruptcy Prediction, Risk Measurement, Customer Relationship Management (CRM), Fraud Analytics, Sentiment Analysis, Social Media Analytics, Big Data Analytics, Churn Prediction in Banks and firms and Asset Liability Management through Optimization. In a career spanning 27.5 years, Dr. Ravi worked in several cross-disciplinary areas such as Financial Engineering, Software Engineering, Reliability Engineering, Chemical Engineering, Environmental Engineering, Chemistry, Medical Entomology, Bioinformatics and Geotechnical Engineering. At IDRBT, he held various administrative positions such as Coordinator, IDRBT-Industry Relations (2005-06), M. Tech (IT) Coordinator, (2006–2009), Convener, IDRBT Working Group on CRM (2010-11); Ph. D. Coordinator (2014–2016). As the convener, IDRBT Working Group on CRM, he co-authored a Handbook on Holistic CRM and Analytics (http://www.idrbt.ac.in/PDFs/Holistic%20CRM-%20Booklet_Nov2011.pdf), where a new framework for CRM, best practices and new organization structures apart from HR issues for Indian banking industry are suggested. He has 28.5 years of research and 15 years of teaching experience. He designed and developed a number of courses in Singapore and India at M. Tech level in Soft Computing, Data Warehousing and Data Mining, Fuzzy Computing, Neuro Computing, Quantitative Methods in Finance, Soft Computing in Finance etc. Further, he designed and developed a number of short courses for Executive Development Programmes (EDPs) in the form of 2-week long CRM for executives, Data Mining, Big Data and its relevance to Banking, Fraud Analytics etc. He conducted ACRM proof of the concept (POC) for 14 banks on their real data. He established excellent research collaborations with University of Hong Kong, University of Ghent, Belgium, IISc., Bangalore and IIT Kanpur. He co-ordinated Two International EDPs in University of Ghent, Belgium on ACRM to banking executives jointly with Prof Dr Dirk Van den Poel, University of Ghent, Belgium in 2011 and 2012. As part of academic outreach, he is a Guest Speaker in IIM Kolkata's PGP programme and an invited Resource Person in various National Workshops and Faculty development programmes on Soft Computing, Data Mining, Big Data funded by AICTE and organized by SRM University, JNTU, UoH and some Engineering Colleges in India. He is an invited Panel Member/Chair in several National forums like IBA on Big Data Analytics in Banking in India. He is a member of task force set up to suggest changes to curriculum by NSDC and administered by NASSCOM to develop Data Scientists. Further, he contributed to the Roadmap for Big Data in India developed by DST. He is an External examiner for Ph.D. in Auckland University of Technology, New Zealand and Christ University, India. Further, he is an External Expert to review Research Project Proposals on Analytics and ACRM submitted by the Belgian Academics to Belgian Government for funding. He was the Associate Professor at IDRBT between February 2010-June 2014. Prior to joining IDRBT as Assistant Professor in April 2005, he worked as a Faculty at the

Institute of Systems Science (ISS), National University of Singapore (April 2002 - March 2005). At ISS, he was involved in teaching M. Tech. (Knowledge Engineering) and research in Fuzzy Systems, Neural Networks, Soft Computing Systems and Data Mining & Machine Learning. Further, he consulted for Seagate Technologies, Singapore and Knowledge Dynamics Pte. Ltd., Singapore, on data mining projects. Earlier, he worked as Assistant Director (Scientist E1) from 1996 −2002 and Scientist C from 1993 to 1996 respectively at the Indian Institute of Chemical Technology (IICT), Hyderabad. He was deputed to RWTH Aachen (Aachen University of Technology) Germany under the DAAD Long Term Fellowship to carry out advanced research during 1997–1999. He earlier worked as Scientist B and Scientist C at the Central Building Research Institute, Roorkee (1988–1993) and was listed as an expert in Soft Computing by TIFAC, Government of India.