Usage Analytics: Research Directions to Discover Insights from Cloud-based Applications

Duc-Tien Dang-Nguyen, Manoj Kesavulu, and Markus Helfert
Lero – Irish Software Research Organization
School of Computing, Dublin City University, Dublin, Ireland
{*duc-tien.dang-nguyen, markus.helfert*}*@dcu.ie; manoj.kesavulu@lero.ie*

Keywords: Usage Analytics, Cloud-based Applications, User Satisfaction, Resource Provisioning

Abstract: Usage in the software field deals with knowledge about how end-users use the application and how the application reacts to the users' actions. In a complex and heterogeneous cloud computing environment, the process of extracting and analysing usage data is difficult since the usage data is spread across various front-end interfaces and back-end underlying infrastructural components of the cloud that host the application. In this paper we propose **Usage Analytics**, a set of potential research directions that could help tackle various challenges in the cloud domain. We provide an overview of usage analytics in the cloud environment and propose how to discover insights using these analytics solutions. We give some discussions about challenges in discovering insights from the usage data as well as provide vision of how usage data will bring benefits to the cloud environment.

1 Introduction

Majority of the in-house applications nowadays are moving to cloud-based environment, with the market size grows to \$246.8B in 2017[1], increasing more than 10 times in just half a decade. Cloud computing provides the users with the possibility of using different devices to use (access) the cloud-based services seamlessly (Mell and Grance, 2011). It, however, brings more challenges to application developers and architects to understand how the applications work and how their customer satisfy with the provided services. To have an example, let us recall the scenario of the cloud-based picture. Cloud providers deliver the same service to different customers, who share physical and/or virtual resources transparently, this concept is referred as Multi-tenancy (Kabbedijk et al., 2015). A cloud-based application lets customers share the same hardware resources, by offering them one shared application and database instance, while allowing them to configure the application to fit their needs as if it runs on a dedicated environment, these resources are shared transparently, meanwhile guaranteeing substantial cost savings (Bezemer et al., 2010). End-users or consumers interact with the cloud applications through various interfaces (those being web browsers, mobile applications, and command-line interfaces). The Infrastructure as a Service (IaaS) model offers computer - physical or virtual machines - and other resources, such as raw block storage, file or object storage, virtual local area networks (VLANs), IP addresses, and firewalls. In the Software as a Service (SaaS) model, applications deployed on cloud infrastructure and are provided to end-users as services over the internet. The more resources are shared by multiple different users, the more resources are managed by cloud providers. Extraction of usage data of the features provided by the cloud applications could help software developers and architects to make an informed decision for the development/improvement of functionalities of the system according to end-user usage patterns (Pachidi et al., 2014).

Analytical solutions refer to the use of various analysis techniques and methods such as data mining, machine learning, reasoning, and other methods to extract useful knowledge and insights from large data set. For example, a company can use analysis techniques to understand customers' behaviour and predict how they are engaged or which customers are least likely to quit. These insights can be discovered via customers' profiles, memberships they subscribe to, or their generated content (comments, clicks, and other interactions). Developers can understand if

[1] According to Forbes, http://bit.ly/forbescloudapps2017

some functions do not work properly via the usage data generated by the users. User interests can be modelled by extracting browsing behaviour when accessing web application (Gasparetti, 2016). Such analytical solutions are considered as increasingly critical tools for modern enterprise to get an informational advantage, and have evolved from a matter of choice to a fundamental requirement in the present competitive business environments. Applying these solutions, thus, is a key to discover insights from the applications' usage.

An intuitive solution is to survey users on how customers use these applications through well-designed studies (interviews or surveys). Unfortunately, this approach has different limitations such as cost to conduct the studies, inability to include large population and users may not be willing to or able to self-identify and so on. These issues can be addressed by using data analytics on the usage data, namely **usage analytics**, which aims to obtain insightful and actionable information for data-driven tasks, around applications and services. With the improving of the data mining tools, these usage data can be gathered from online services by collecting all traces of user activity to produce clickstreams, sequences of timestamped events generated by user actions. For web-based services, these might include detailed HTTP requests. For mobile applications, clickstreams can include everything from button clicks, to finger swipes and text or voice input (Wang et al., 2016). Insightful information is information that conveys meaningful and useful understanding or knowledge towards providing the target service or user satisfaction to that service (Zhang et al., 2011). Typically insightful information is not easily achievable by directly investigating the raw data without aid of analytical solutions. Developing a usage analytics project typically goes through iterations of four phases: *task definition, data preparation, analytic-technology development*, and *deployment and feedback gathering*. Task definition is to define the target task to be assisted by software analytics. Data preparation is to collect data to be analysed. Analytic-technology development is to develop problem formulation, algorithms, and systems to explore, understand, and discover insights from the data. Deployment and feedback gathering involves two typical scenarios. One is that, as researchers, we have obtained some insightful information from the data and we would like to ask domain experts to review and verify. The other is that we ask domain experts to use the analytic tools that we have developed to obtain insights by themselves. Most of the times it is the second scenario that we want to enable.

In this paper, we will show that when applying usage analytics in practice, we should fuse a broad spectrum of domain knowledge and expertise, from management, machine learning, data processing to information visualization. By using usage analytics, we are aiming at proposing powerful tools to address the following challenges (deeper discussions can be seen in (Kesavulu et al., 2018)):

1. **Resource provisioning**: based on the usage data, predict the resources that may be allocated to an application.
2. **Problem diagnosis**: analyzing the usage data, which are the logs in this case, to understand how to localize the node that is the source of performance problems.
3. **Understanding user satisfaction**: instead of surveying and asking feedback, how users satisfy with an application and be revealed via their usage data.
4. **Discovering user behaviour patterns**: Every user has their own pattern when using an application or a service. Understanding these patterns could help to improve the service or discover the trends in advance. These patterns, can be discovered from the usage data.

Consequently, the aims of this paper are:

- To provide an overview of what is usage analytics in the cloud environment and propose a usage data classification;
- To inspire and motivate researchers to use their know-how in this new emerging and important area;
- To propose how future usage data in the cloud applications should be extracted, and from that insights can be discovered via appropriate analytics techniques;
- To discuss the challenges in discovering insights from the usage data and provide vision of how usage data will bring benefits to the cloud environment.

The remainder of the paper is structured as follows: in Section 2, we discuss the problems and challenges that exist in understanding what does *usage data* mean, challenges in extracting information from usage data and building an usage data analytics framework and discuss the general requirements for the usage analytics framework. We then describe the potential analytics methods that aim to provide solutions in Section 3. In Section 4, we discuss the potential of this research and the corresponding challenges involved. Finally, in Section 5, we provide conclusions drawn.

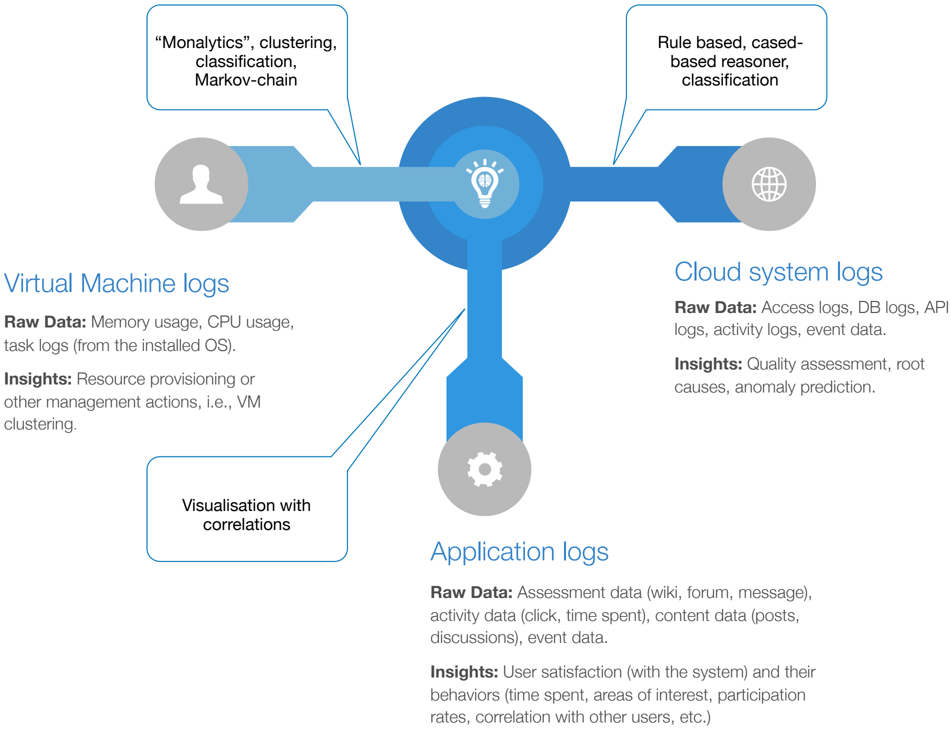

Figure 1: Three main sources of usage data in cloud-based environment.

2 Problems, Challenges and Requirements

In this section, we address a couple of basic questions in this topic: "what are the usage data?", "where can we get them?", "what are the challenges?" and "what are the requirements in analysing these data?".

2.1 What are the usage data?

Usage data, as its name suggested, are data generated when users or customers are using the applications and services. They can be extracted from any stage when services and/or applications are being provided. In this study, we propose to group these data into three groups, coming from three main sources, as follows: the system logs from the cloud services from the back-end of the cloud system, the application logs, and the logs from the virtual machines (VMs). Fig. 1 shows a summary of the three main sources of usage data and the answers for the above questions.

System logs contain a wealth of information to help manage systems. Most systems print out logs during their executions to record system runtime actions and states that can directly reflect system runtime behaviours. System developers and architects usually use these logs to track a system to detect and diagnose system anomalies.

The second type of this data is the user-level usage data generated as a result of user interaction with a cloud-based application. Some examples of usage data are application logs, for example the assessment data (wiki, forum, message), the activity data (clicks, time spent), server logs, and so on. They can be extracted by the applications themselves or via Web cookies (from web browser). Such data in the cloud is spread across various interfaces such as Web browser, mobile applications and command line interfaces on the front-end and server and database on the back-end.

The last type of usage data is the VM logs, typically generated from the VMs running the applications or services. This type of logs contains the usage of the CPU, memories, as well as running tasks, time of starting and stopping and others.

2.2 Challenges in Extracting Information from Usage Data

As mentioned above, usage data can be extracted at any stage, and the major challenge is they can be in any form and format, which brings many challenges to the analysis. The main questions for usage data extractor is what usage data should be extracted and how to map the raw usage data with the right applications or services. Considering the multi-tenant architecture of the cloud, different applications share the same physical and virtual resources. This raises challenge as in how to separate and extract the logs that represent each application from the instance (VM) cohosting the applications.

Another important challenge is handling with different contextual information. A system usually has a lot of branches, and thus the systems behaviors may be quite different under different input data or environmental conditions. Knowing the execution behavior under different inputs or configurations can greatly help system operators to understand system behaviors. However, there may be a large number of different combinations of inputs or parameters under different system behaviors. Such complexity poses difficulties for analyzing contextual information related to the state of interest.

2.3 Challenges in Building Usage Analytics Framework

Despite the advantages provided by analytical solutions, the solution implementation is usually very costly, which hinders enterprises, especially the SMBs (Small and Medium Business), to start such projects (Sun et al., 2012). Normally, storing huge volumes of data requires a large storage system as well as buying (and training how to use) expensive analytics software. This will come also with a large number of clusters and powerful machines to run data analytics algorithms.

Another challenge that related to cost-effective is limited to only a small number of large companies or enterprises can effort to run usage analytics framework. They normally have to pay a lot to maintain the complex software and hardware only for occasional usages, for example when a financial quarter is over or some unusual events happen (Sun et al., 2012).

The framework also have to be constantly updated since the cloud-based environment is also moving very quickly. The built framework should have the ability to predict and adapt with the changes of technologies and able to process new coming types of usage data or new types of services and applications.

Last but not least, usage analytics framework requires high skill analysts to run the analytical solutions since it requires constant tuning, validating, and updating according to the changing business context, as well as the manner of services and applications.

2.4 Usage Analytics Requirements

Usage analytics requires technologies to efficiently process and discover insights from large and diverse usage data. Summing up from several previous studies in software analytics and cloud computing((Aceto et al., 2012), (Jehangiri et al., 2013)), we point here the general requirements for a usage analytics framework:

1. Scalability: The system should be scalable, i.e., it should be able to handle a large number of usage data extractors. This requirement is very important in the cloud environments due to a large number of services and structures of cloud systems that may grow elastically.
2. Heterogeneous data: The system should consider a heterogeneous group of metrics. It should allow to deal with usage data at different levels: service level (application), virtual IT-infrastructure level (e.g., VM logs), and fine-grained physical IT-infrastructure level (e.g., the cloud systems logs).
3. Relationship: There is a hierarchical relationship between applications, VMs and the physical machines. These relationships can be changed dynamically, thus, the analytics system has to cope also this aspect.
4. Meaningful: The extracted usage data must be meaningful from a variety of sources, i.e., the system should be able to filter out non-relevant information, e.g., noise data. Furthermore, data extractors should easily be extended, for example, by adding more plugins.
5. Abstraction: The usage analytics need to exhaustively aggregate runtime data from different sources and consolidate information at a high level of abstraction.
6. Identification of Influential Metrics: The system should be able to identify the metrics or parameters that strongly influence the decision making, which will help in decreasing the time and complexity in analysis.

3 Addressing the Challenges

In this section, we describe the potential analytics methods that aim to provide useful tools to solve the four major problems (summarised in Table 1) mentioned in the introduction: resource provisioning, problems diagnosis, understanding user satisfaction, and discovering user behaviour patterns. Some evaluation methods will be also introduced.

Table 1: The potential solutions for usage analytics problems.

Problem			Potential Solution
Problem Type	Problem Area	Problem Example	
Resource Provision	Cloud Infrastructure Level	VM Selection	Prediction methods
		In-time decision making	Monalytics
		VM Capacity Management	Markov-chain
Problem diagnosis	Application & Cloud Infrastructure Level	System Execution Behavior	Contextual Analysis
Understanding User Satisfaction	Application level	Feedback analysis	Data Visualisation
	User level	User feedback	User Engagement
Discovering User Behavior Patterns	User level	User patterns	Descriptive statistic, Graph-based decision making

3.1 Resource Provisioning

Let us start with the usage data from the left-most branch from the diagram in Figure 1. A typical problem on cloud-based environment is the network resource management, for example the acceptable Virtual Machine (VM) configuration to minimize the resource consumed by certain services deployed on these VMs. A common problem experienced in data centers and utility clouds is the lack of knowledge about the mappings of the services being run by or offered to external users to the sets of virtual machines (VMs) that implement them (Wang et al., 2011). It can be done by exploiting analytics methods, for example by predictive analysis on the usage data from the systems logs from each VM, to predict the suitable configuration for future VM deployments.

For an in-time decision making, Wang et al. in (Wang et al., 2011) proposed a system integrating monitoring with analytics, termed Monalytics, which can capture, aggregate, and incrementally analyze data on-demand and in real-time, thus increasing accuracy and reducing human intervention in the analysis process. It was done by applying a clustering algorithm and a top-k flow analysis (Kumar et al., 2004) on the data gathered from the CPU usage data on each VM, identifying the VMs that are responsible for the majority of the traffic flow in the group. This provides information on critical VM combinations to include in the same group to achieve maximum cost benefit during VM migration.

Usage data on VM can also be exploited to predict VM states that can be used as the inputs of the existing networking capacity management techniques. For example, in (Sun, 2016), the authors proposed a method named Smart Predictive Capacity Management (SPCM) that is designed to assist cloud networking deployment in estimating the acceptable network capacity for a specific configuration of interdependent VMs by predicting individual VM states. It is done by applying Markov chain techniques to address the data analytics for potential states in heterogeneous cloud computing environment. This work could help enterprises to optimise the VM configurations to attain significant performance improvement.

3.2 Problem Diagnosis

With the increasing scale and complexity of the cloud-based applications, it has become more and more difficult for system operators to understand the behaviors of system for tasks such as system problem diagnosis. For example, system operators need to understand system execution behaviors to identify symptoms and root-causes of anomalous nature of the system. System behaviors include a series of actions executed by the system and the corresponding changes in the system states. Although operators usually investigate a system starting from a specific state of interest, e.g., a hang state or failure state, it is critical to identify the series of states the system traversed to reach the current unstable state.

The study in (Fu et al., 2013) proposes a new approach for the contextual analysis of system logs to better understand a systems behaviors. In particular, they used execution patterns extracted from the system logs that ultimately reflect the runtime behavior of the application, and propose an algorithm to mine execution patterns from the system logs. Based on the execution patterns, their approach further learns the essential contextual factors by modelling the relationships among execution patterns that are responsible for the execution of specific branch of the system.

In this study, we also propose a contextual analysis method, inspired by study in (Fu et al., 2013), by analysing the application logs to better understand the correlation between users behaviors and the corresponding system. In particular, we propose to use the Formal Concept Analysis (Ganter and Wille, 1997) to mine execution patterns from the usage data from all sources: application, hosting VM(s), and underlying cloud logs. The execution patterns in this context can be considered as reflections of user's interactions with the application. The mining and learning results can help system operators understand both the behaviors of the customers as well as the execution logic of their services and applications.

3.3 Understanding User Satisfaction

Visualisation is a typical way to help the system analysts understand how user satisfy with their services. These information can be grasped more easily and quickly when presented through comprehensible information visualizations, for example by the means of charts, graphs, from the basic interaction usage data (as shown in the bottom branch of the schema in Figure 1). This problem has been studied for decades in learning and multiple ways of visualizing data increase the perceived value of different feedback types (Dyckhoff et al., 2012).

A typical way to access the cloud-based applications by the end-user is through a web-browser. Data analytics techniques, e.g., web-mining, in this way, can be employed to obtain interaction insights. In (Bucklin and Sismeiro, 2009), the authors provided an overview of *Clickstream* data, defined as the electronic record of a user's activity, represents the traces of an end-user takes while accessing the cloud application. Analyse such kind of information can discover how a user satisfy with the provided services based on the interaction of the users (obtained via the number of clicks, time spent, and other usage data).

Usage analytics can also reveal the engagement level of customers during the development and evaluation process of a software analytic project. It is well recognized that engaging customers is a challenging task especially in the context of software engineering tools. Customers always tend to keep their existing way of carrying out a task or the way of using a service. Furthermore, it is usually lacking of investing time to understand the pros and cons of the proposed tools due to tight development schedule. Thus, understanding customer engagement has significant impact on the development of the applications and services. By visualisation also predictive analysis, analytics tool can provide providing effective visualization and manipulation of analysis results as well as predict the engagement level which helps the evaluation of the applications and services.

3.4 Discovering User Behavior Patterns

In order to understand user behavior, descriptive statistics, e.g., mean, total, standard variation, most frequent value, etc., are typically used to obtain meaningful insights such as the basic behaviors of the users. These information can be also used to classify the user based on the correlation and demographic similarities among them. In order to understand the patterns from user behavior, we propose to exploit all of the usage data from multiple layers of the cloud environment, usage data of a cloud-based application is spread across front-end interfaces (web-browser, smart phone app/client and command-line interface) and the back-end (server instance and database instance) in a cloud environment (Kesavulu et al., 2017) and formulate as the transition states of a graph. This type of graph can be used to mine execution patterns and to model relationships among different user behavior patterns. This kind of approach can also be used to discover some problems under some specific context. To discover these contextual factors, we propose to use the decision trees to learn the conditions, which allows us to determine any possible connections between the contexts and change in behavior of the user.

It is worth noting that these usage data potentially can be exploited in situation-emotional analytics (Märtin et al., 2017), which aims at recognizing the emotions and changes of software situations in order to improve the quality and user experience levels. These emotional information are now extracted via external biometric recording devices, e.g., recording devices that record the eye and gaze-tracking signal. We firmly believe that, usage information at the application levels, will be very useful for this type of learning and potentially can replace eye and gaze-tracking information.

3.5 Evaluations and Validations

Evaluation itself is a challenge in usage analytics. For example, if an analytical solution provides some information about the user satisfaction, it is non-trivial to evaluate if that information shows the real "picture" of how user satisfaction can be achieved. Traditionally, we need to run some surveys and/or interviews with the actual users to evaluate results from the analytical solutions. Another way is to ask some domain experts to evaluate the proposed solutions. This method, however, is also costly and very subjective. In this study, we propose a novel way for the evaluation by using another type usage information, gathering from the snapshots of the users' device, the screen-shots of the interfaces used to access the application. By collecting these information (in the testing phase), we can provide users with the ability to recall and re-access their previous computer usage and the content they engage with.

4 Realising the Potentials

We firmly believe that usage analytics will very soon be a phenomenon in anyone working on cloud-based environment, and will positively impact on everyone who uses the technology. In this section, we point out the potential applications as well as possible research challenges and projects on usage analytics.

4.1 Potential Applications

In our vision, usage analytics opens up a new paradigm of opportunities, namely:

- **Enhancing productivity without running never-ending surveys.** Most of companies run surveys and collecting feedback from customers to know how their services being used. This is costly and time-consuming. With usage analytics, these information can be obtained almost in real-time, and even better, more reliable. Given a set of usage data from different users over time (historical usage data), how they use the applications, what could make errors, how much resources should be allocated, and so on can be provided.
- **A greater knowledge of the system.** Usage analytics can discover hidden user behavior patterns, providing information which would go unnoticed. They can identify trends and patterns from their customers as well as their own systems, allowing better services with less expensive resources.
- **Improving the services and the architecture behind them.** Traditionally, software and cloud-based applications are upgraded as an increasing of their versions, which requires a lot of time for collecting customers' feedback and system diagnosis results. By using usage analytics, this process can be simplified and the services potentially can be constantly update, seamlessly and reliably.

4.2 Possible Research Challenges and Projects

Usage analytics comes also with challenges and opportunities for researchers. It is important that the research community helps to address the challenges in this emerging and important field. We cannot easily apply our existing analytics methods on this type of data and hope for success. Therefore we need specific approaches addressing the specific challenges. As a first way-point for researchers we are proposing different research topics and research questions.

- **How can we identify and extract important information from the usage data?** Deciding what should be extracted from the data that is important is a nontrivial task. Going beyond standard analysis like predicting some satisfaction level will be important, forcing researchers to think creatively and go beyond simple analysis. Many research questions arises, such as, how to combine information from different raw usage data, or how to efficiently process the data. Also an important part here is to explore how context and situations can be taken into account to improve the quality of the analysis.

- **How can we present the results to the user, i.e., the company?** Reporting the results to the users is one of the most important parts of the analysis of this data. Nevertheless, this is not trivial since the amount of data and information that can be extracted is huge. It will be important to research novel interfaces that allow users easily getting the root causes of the errors or understanding the engagement of their customers. Generating summaries and automatic reports will be another topic that is important for this data since there will be a need from the user side for such summaries with respect to, for example, weekly report from the whole systems.

- **How can information and usage data be processed efficiently?** Systems that have to process a huge amount of data in a complex way have to be efficient to make them useful to the users. This comes with challenges for researchers in terms of how to parallelize and process data efficiently in a reasonable amount of time, how to combine different research fields, from software analytics to machine learning, together.

The potential for usage analytics is enormous. We do acknowledge that there are challenges to be overcome, such as finding the right analytics techniques, synchronization, data extraction, and the development of a new generation of analytics tools on usage data, but we believe that these will be overcome and that we are on the cusp of a positive turning point for cloud-based applications community.

5 Conclusions

We presented the challenges in getting insights from cloud based applications. We pointed out that analytical solutions on usage data, namely usage analytics, can help to overcome these challenges. A complete picture of how to apply usage analytics to get insights from the cloud-based applications and services is shown and discussed. Some potential applications were also addressed. Our future work aims (i) at designing and developing methods/techniques to collect, extract and/or aggregate the usage-data from Applications, VMs hosting the application and the cloud system hosting the VMs; (ii) to develop an experiment to evaluate the usage data extraction and analysis methods.

ACKNOWLEDGEMENTS

This work was supported with the financial support of the Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie).

REFERENCES

Aceto, G., Botta, A., de Donato, W., and Pescap, A. (2012). Cloud monitoring: Definitions, issues and future directions. In *IEEE International Conference on Cloud Networking (CLOUDNET)*, pages 63–67.

Bezemer, C.-P., Zaidman, A., Platzbeecker, B., Hurkmans, T., and Hart, A. . (2010). Enabling multi-tenancy: An industrial experience report. In *IEEE International Conference on Software Maintenance*, pages 1–8.

Bucklin, R. E. and Sismeiro, C. (2009). Click here for internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive Marketing*, 23(1):35–48.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., and Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Educational Technology and Society*, 15(3):58–76.

Fu, Q., Lou, J.-G., Lin, Q., Ding, R., Zhang, D., and Xie, T. (2013). Contextual analysis of program logs for understanding system behaviors. *Proceedings of the 10th Working Conference on Mining Software Repositories*, pages 397–400.

Ganter, B. and Wille, R. (1997). *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition.

Gasparetti, F. (2016). Modeling user interests from web browsing activities. *Data Mining and Knowledge Discovery*, 31(2):1–46.

Jehangiri, A. I., Yaqub, E., and Yahyapour, R. (2013). Practical aspects for effective monitoring of slas in cloud computing and virtual platforms. In *International Conference on Cloud Computing and Services Science*, pages 447–454.

Kabbedijk, J., Bezemer, C.-P., Jansen, S., and Zaidman, A. (2015). Defining multi-tenancy: A systematic mapping study on the academic and the industrial perspective. *Journal of Systems and Software*, 100:139–148.

Kesavulu, M., Dang-Nguyen, D.-T., Helfert, M., and Bezbradica, M. (2018). An Overview of User-level Usage Monitoring in Cloud Environment. In *The UK Academy for Information Systems (UKAIS)*.

Kesavulu, M., Helfert, M., and Bezbradica, M. (2017). A Usage-based Data Extraction Framework for Cloud-Based Application - An Human-Computer Interaction approach. In *International Conference on Computer-Human Interaction Research and Applications (CHIRA)*, Madeira, Portugal.

Kumar, A., Sung, M., Xu, J. J., and Wang, J. (2004). Data streaming algorithms for efficient and accurate estimation of flow size distribution. *SIGMETRICS Perform. Eval. Rev.*, 32(1):177–188.

Märtin, C., Herdin, C., and Engel, J. (2017). Model-based User-Interface Adaptation by Exploiting Situations, Emotions and Software Patterns. *International Conference on Computer-Human Interaction Research and Applications*.

Mell, P. and Grance, T. (2011). The NIST definition of cloud computing. *NIST Special Publication*, 145:7.

Pachidi, S., Spruit, M., and Van De Weerd, I. (2014). Understanding users' behavior with software operation data mining. *Computers in Human Behavior*, 30(January):583–594.

Sun, X. (2016). Virtual Machine Optimizations Using Markov Chain Data Analytics in Heterogeneous Cloud Computing. In *International Conference on Smart Cloud*, pages 248–253.

Sun, X., Gao, B., Fan, L., and An, W. (2012). A cost-effective approach to delivering analytics as a service. In *International Conference on Web Services*, pages 512–519.

Wang, C., Schwan, K., Talwar, V., Eisenhauer, G., Hu, L., and Wolf, M. (2011). A flexible architecture integrating monitoring and analytics for managing large-scale data centers. *ACM International Conference on Autonomic Computing*, page 141.

Wang, G., Zhang, X., Tang, S., Zheng, H., and Zhao, B. Y. (2016). Unsupervised Clickstream Clustering for User Behavior Analysis. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 225–236.

Zhang, D., Dang, Y., Lou, J.-G., Han, S., Zhang, H., and Xie, T. (2011). Software analytics as a learning case in practice. *Proceedings of the International Workshop on Machine Learning Technologies in Software Engineering - MALETS '11*, pages 55–58.