Minimizing and Managing Clouds failures

PatriciaTakako Endo, Universidade de Pernambuco

Guto Leoni Santos, Daniel Rosendo, Demis Moacir Gomes, André Moreira, Judith Kelner, Djamel Sadok, Universidade Federal de Pernambuco

Glauco Estácio Gonçalves, Universidade Federal Rural de Pernambuco

Mozhgan Mahloo, Ericsson Research

I. INTRODUCTION

Many enterprises have migrated their applications to the cloud due to guaranteed flexibility, scalability, and lower investment. However, availability of cloud services is still a big concern for many enterprises. To consider a service highly available (99.999%), service downtime should be less than 5.25 minutes per year, [1]. Therefore, guaranteeing such stringent availability level is a main challenge that cloud providers are facing nowadays. Lack of standard solutions to handle the failures as well as high dependency of service availability on the data center management strategies used to deal with the full-stack failures' occurrences (extending from physical infrastructure to software levels) contribute to the challenge.

Non-planned data center interruptions are expensive for both cloud providers and users, and require special attention. For instance, cloud service providers, like Amazon and Microsoft Azure, have an estimated cost of $336,000 per hour in case of a failure, while a disruption in a credit card authorization service incurs loss of around $2,600,000 per hour [2]. The amount of such financial losses can differ a lot for businesses.

In addition to the direct financial losses, these failures result in business interruptions, decreased productivity, and damaged business reputation. It is therefore important to understand the reasons behind cloud service failures, as the first step towards offering highly available cloud services. What are the main causes that lead to service unavailability? How can cloud providers safeguard against failures? These are some of the important questions that every successful cloud provider should be able to address.

II. WHAT CAUSES CLOUD APPLICATION OUTAGES?

Figure 1 shows where the main failures can originate in a data center. Though **planned outages** can occur to facilitate preventive maintenance or tests, **non-planned failures** are the main concerns for data center operators and cloud users. Non-planned failures can occur due to infrastructure failures (e.g., UPS system failures are 25% of outages causes, while CRAC failures are responsible for 11% [10]), planning mistakes, software failures, and, mistakes by humans (that represent 22% of outages [10]), including external attacks and misconfiguration.

With regard to **infrastructure failures**, a first problem stems from **equipment failures**, including power, cooling and IT subsystems. Moreover, sometimes, low cost equipment is used in an attempt to reduce data center spending, which might lead to increased failure rate of cooling and power systems (such as power distribution unit (PDU) or circuit breaker), and consequently triggering IT equipment shutdown. Recently, a failure in the British Airways (BA) data center resulted in 600 flights canceled affecting about 75,000 passengers with a cost of $112 million. According to BA's chief executive, Alex Cruz[1], the failure occurred due to a power surge at the data center that resulted in a failure on their complex IT system.

[1] https://www.theguardian.com/business/2017/may/31/ba-it-shutdown-caused-by-uncontrolled-return-of-power-after-outage

Problem can arise from **planning mistakes** and increased data center usage. This often leads to issues, such as performance bottlenecks on servers and network components, overloading of the UPS system, and excessive heat to be dissipated by the computer room air conditioning (CRAC). Last year, the Google Cloud data center struggled under heavy demand from the phenomenon Pokémon Go game. There are reports of server problems with server connection and log-in (access) issues.

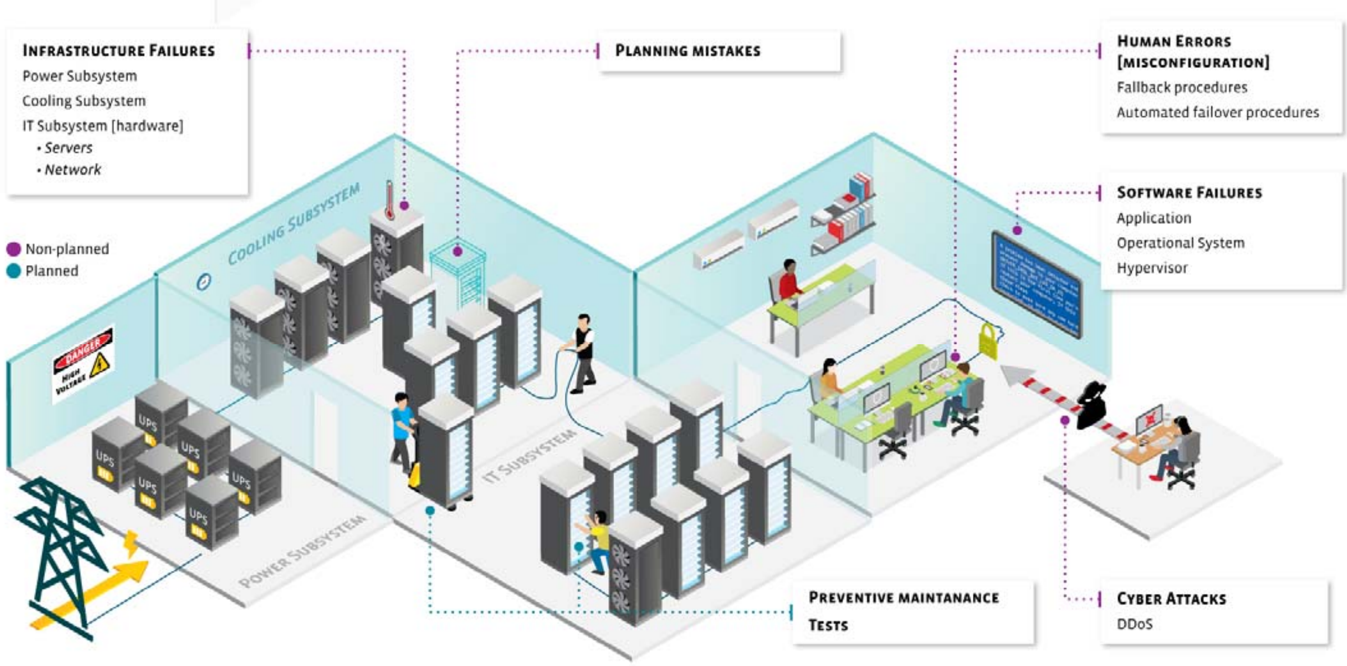


Figure 1. Main data center point of failures, divided in non-planned and planned outages.

At a logical level, data center operators have to deal with **software failures**, that may occur in applications, operational systems or hypervisors. Some of those failures can come from Heisenbugs [3]. These are software bugs whose roots are undefined. A case of software failure occurred in 2014 and led to Facebook services being down during about 31 minutes[2]. The service interruption took place while an update process of their software system was carried out.

Failures caused by **human intervention** can be divided into two categories: failures from external attacks and failures caused by internal personal, such as misconfigurations or typos. Recently, the Amazon Cloud suffered a downtime due a simple typo affecting a large Internet site for about 5 hours[3]. The company stated that it is making some changes in the tool employees use to avoid a repeat of this type of human error.

These recent and other real-world cloud failures illustrate that non-planned data center interruptions can bring huge financial losses. Further, cloud failures can result in business disruption, lost revenue, and diminished end user productivity in addition to damaged reputation. Hence, cloud service providers must address cloud failures.

III. How to Minimize Cloud Failures’ Impact?

To build and offer highly available cloud, cloud providers must implement mechanisms such as **monitoring**, geo-**distributed storage**, and **disaster recovery (DR).**

[2] https://www.theguardian.com/technology/2014/jun/19/facebook-unavailable-longest-outage-four-years

[3] https://aws.amazon.com/message/41926/

Monitoring

Monitoring is a basic but crucial mechanism related to checking the applications and (physical and virtual) resources on cloud infrastructure. Most of current monitoring mechanisms are not only a passive agent collecting and logging information; they go beyond, trigging failure alerts, suggesting diagnoses and routines to prevent possible failures.

Geo-Distributed Storage

The geo-distributed storage provides many advantages, such as scalability and access locality. It spreads the data backup (replicas)over multiple data centers around the world, increasing protection and security against disasters and failures in the active main data center. Some cloud providers offer distributed storage solutions, and most of them rely on different geographic zones with synchronization mechanisms to allow an efficient failover procedure.

Disaster Recovery (DR)

Despite all the high availability mechanisms put in place, Cloud providers ought to have a DR plan to be executed in case of any disaster occurrence in order to minimize the downtime of their data centers. DR plans rely on many other mechanisms to recover all data center components from a failure. For instance, a distributed storage with replicas and an automatic failover procedure is essential to recover the data to another entity and get the service up again as soon as possible. Some cloud providers offer DR plans at infrastructure, storage, or application levels, as well as a failover process at physical and virtual machine levels.

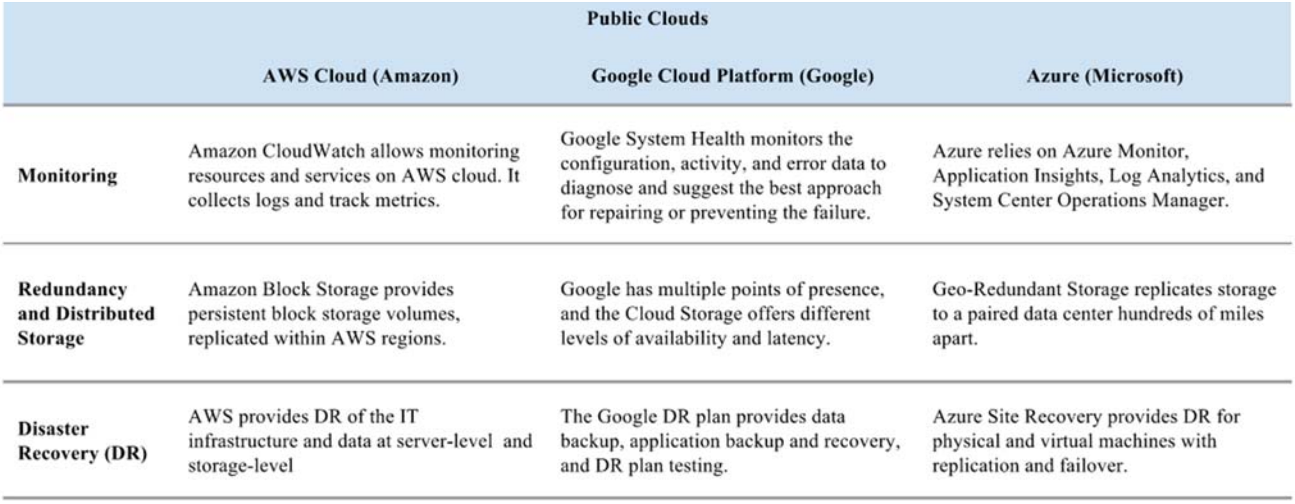
IV. What Cloud Service Providers are Doing

Public Cloud Providers

Table I presents three examples of public cloud providers and their respective mechanisms for monitoring, redundancy and geo-distributed storage, and DR.

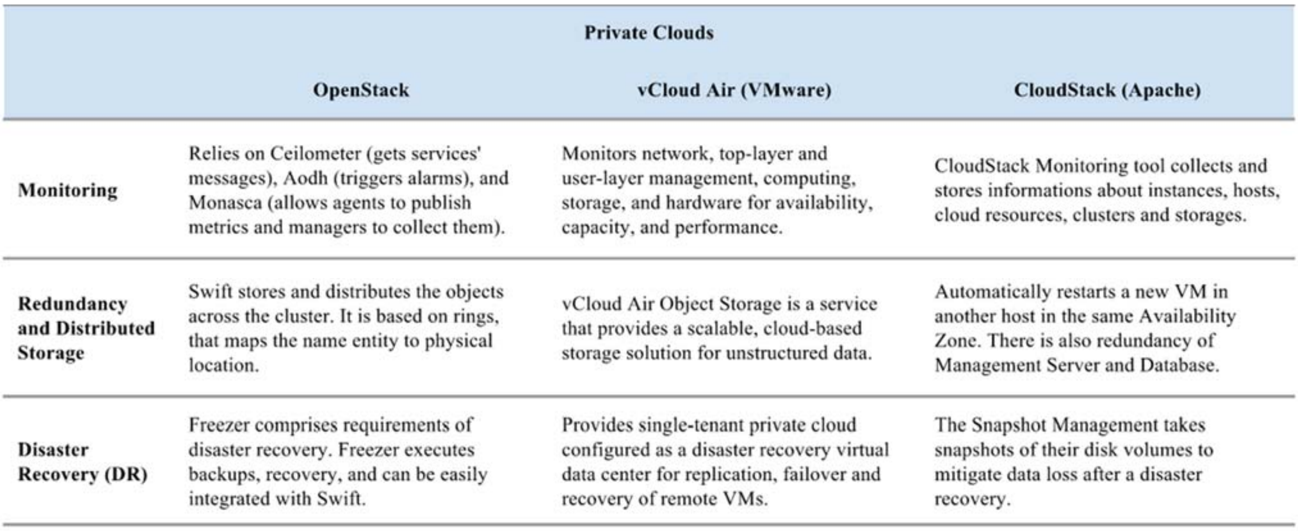
AWS Cloud (https://aws.amazon.com) extends across 16 geographic regions around the world and offers services such as computing (EC2 and Elastic Beanstalk), storage (S3 and DynamoDB), and content delivery

Table I. Public Cloud Mechanisms



	Public Clouds		
	AWS Cloud (Amazon)	Google Cloud Platform (Google)	Azure (Microsoft)
Monitoring	Amazon CloudWatch allows monitoring resources and services on AWS cloud. It collects logs and track metrics.	Google System Health monitors the configuration, activity, and error data to diagnose and suggest the best approach for repairing or preventing the failure.	Azure relies on Azure Monitor, Application Insights, Log Analytics, and System Center Operations Manager.
Redundancy and Distributed Storage	Amazon Block Storage provides persistent block storage volumes, replicated within AWS regions.	Google has multiple points of presence, and the Cloud Storage offers different levels of availability and latency.	Geo-Redundant Storage replicates storage to a paired data center hundreds of miles apart.
Disaster Recovery (DR)	AWS provides DR of the IT infrastructure and data at server-level and storage-level	The Google DR plan provides data backup, application backup and recovery, and DR plan testing.	Azure Site Recovery provides DR for physical and virtual machines with replication and failover.

Table II. Private Cloud Mechanisms



	Private Clouds		
	OpenStack	vCloud Air (VMware)	CloudStack (Apache)
Monitoring	Relies on Ceilometer (gets services' messages), Aodh (triggers alarms), and Monasca (allows agents to publish metrics and managers to collect them).	Monitors network, top-layer and user-layer management, computing, storage, and hardware for availability, capacity, and performance.	CloudStack Monitoring tool collects and stores informations about instances, hosts, cloud resources, clusters and storages.
Redundancy and Distributed Storage	Swift stores and distributes the objects across the cluster. It is based on rings, that maps the name entity to physical location.	vCloud Air Object Storage is a service that provides a scalable, cloud-based storage solution for unstructured data.	Automatically restarts a new VM in another host in the same Availability Zone. There is also redundancy of Management Server and Database.
Disaster Recovery (DR)	Freezer comprises requirements of disaster recovery. Freezer executes backups, recovery, and can be easily integrated with Swift.	Provides single-tenant private cloud configured as a disaster recovery virtual data center for replication, failover and recovery of remote VMs.	The Snapshot Management takes snapshots of their disk volumes to mitigate data loss after a disaster recovery.

(CloudFront) to its more than 1 million customers. In case of failure, it provides a hot standby solution that enables a rapid failover.

Google Cloud Platform (https://cloud.google.com/) offers management tools (Monitoring, Debugger, Shell, Billing) and provides cloud services, such as computing (Compute Engine, App Engine, and Cloud Functions), data storage (Cloud Datastore and Storage), data analytics (Cloud Bigtable and BigQuery) and machine learning (Video Intelligence, Cloud Vision, Cloud Speech, and Natural Language).

Microsoft Azure (https://azure.microsoft.com) is available in 34 regions around the world and offers management tools (Resource Manager, Site Recovery, Scheduler, Log Analytics, among others) and computing (Virtual Machines and Cloud Services), mobile (Azure App), storage (Azure Storage and Cosmos DB), data management (Azure Search and Backup), and machine learning (Azure ML) services.

Private Cloud Providers

Table II presents solutions for monitoring, redundancy and geo-distributed storage, and DR of three private clouds: OpenStack, vCloud Air and CloudStack.

OpenStack (https://www.openstack.org/) offers a set of software tools to manage computer, storage, and network resources, working with enterprise and open source technologies. They provide management at high level (dashboard) and low-level (API).

vCloud Air (https://www.vmware.com/cloud-services/infrastructure.html) is a VMware solution to manage the cloud. Its DR plan is a recovery-as-a-service in which users can test different failover scenarios combining different failback and migration procedures.

CloudStack (https://cloudstack.apache.org/)offers a solution that covers the entire stack, including computer orchestration and network of virtual machines management with high availability and scalability.

V. FURTHER STEPS

A comprehensive modeling of the data center infrastructure and its assets represents the right step towards understanding potential points of failure and subsequently preventing service outages. Such models also make it possible to estimate the end-to-end availability of the offered services, and can also answer many related performance questions, such as where a cloud provider should dedicate more effort and money to minimize the impact of such failures.

To obtain a comprehensive model, one needs to cover both the physical systems, as well as the software and applications hosted in the data center. In [4] and [5], we have proposed a set of stochastic models to represent a data center infrastructure and cloud applications, in addition to other works (such as [7], [8], and [9]). On the basis of these models, we can perform sensitivity analysis to find the components with high impact factors on the total data center availability, and detect performance bottlenecks, in order to design an effective failure recovery policy and implement the necessary protection mechanisms [6]. Thought initially, these models are separated, we plan to integrate them in order to improve the understanding of overall data center behavior.

Please note that providing such comprehensive models is not an easy task due to the complex cloud infrastructure and variety of solutions and implementations. Beyond that, care must be taken to reduce models´ state-space complexity and ensure that they remain tractable.

Other important aspects are the DR plan testing and maintenance. The DR plan project is not complete until the cloud provider tests it to make sure that the plan is effective and will attend all requirements when faced with a disastrous situation. Multiple points of data restoration must be designed and tested regularly in order to verify the need of adjustments and also to train the staff to handle emerging situations when necessary. Testing and maintenance should always check the backup and replica services integrity, as well as the networking components.

ACKNOWLEDGEMENTS

This work was supported by the RLAM Innovation Center, Ericsson Telecomunicações S.A., Brazil.

REFERENCES

[1] Toeroe M, Tam F (2012) Service Availability: Principles and Practice. John Wiley & Sons. http://www.wiley.com/WileyCDA/WileyTitle/productCd1119954088.html

[2] Gagnaire, M., Diaz, F., Coti, C., Cerin, C., Shiozaki, K., Xu, Y., Delort, P., Smets, J.P., Le Lous, J., Lubiarz, S. and Leclerc, P., 2014. Downtime statistics of current cloud solutions. International Working Group on Cloud Computing Resiliency, Tech. Report. Available at http://iwgcr.org/wp-content/uploads/2014/03/downtime-statistics-current-7.pdf

[3] Grottke, Michael, and Kishor S. Trivedi. "A classification of software faults." Journal of Reliability Engineering Association of Japan 27.7 (2005): 425-438.

[4] Gomes, D., Endo, P., Gonçalves, G., Rosendo, D., Santos, G., Kelner, J., Sadok, D., and Malhoo, M. Evaluating the Cooling Subsystem Availability on a Cloud Data Center. In IEEE Symposium on Computers and Communications (ISCC), 2017.

[5] Santos, G., Endo, P., Gonçalves G., Rosendo, D., Gomes, D., Kelner, J., Sadok, D., and Malhoo, M. Analyzing the IT Subsystem Failure Impact on Availability of Cloud Services. In IEEE Symposium on Computers and Communications (ISCC), 2017.

[6] Junior, R.M., Guimaraes, A., Camboim, K., Maciel, P. and Trivedi, K., 2011, April. Sensitivity analysis of availability of redundancy in computer networks. In *CTRQ 2011, The Fourth International Conference on Communication Theory, Reliability, and Quality of Service* (pp. 115-121). Available at: https://www.researchgate.net/profile/Rubens_Matos/publication/258628553_Sensitivity_analysis_of_availability_of_redundancy_in_computer_networks/links/0c960528b97097e385000000.pdf

[7] Callou, G., Maciel, P., Tutsch, D. and Araújo, J., 2012, June. Models for dependability and sustainability analysis of data center cooling architectures. In *Dependable Systems and Networks Workshops (DSN-W), 2012 IEEE/IFIP 42nd International Conference on* (pp. 1-6). IEEE. Available at: http://ieeexplore.ieee.org/abstract/document/6264697/

[8] Govindan, S., Wang, D., Chen, L., Sivasubramaniam, A. and Urgaonkar, B., 2011, October. Towards realizing a low cost and highly available datacenter power infrastructure. In *Proceedings of the 4th Workshop on Power-Aware Computing and Systems* (p. 7). ACM. Available at: http://dl.acm.org/citation.cfm?id=2039259

[9] Addabbo, T., Fort, A., Mugnaini, M., Vignoli, V., Simoni, E. and Mancini, M., 2016. Availability and reliability modeling of multicore controlled UPS for datacenter applications. *Reliability Engineering & System Safety*, *149*, pp.56-62. Available at: http://www.sciencedirect.com/science/article/pii/S0951832015003622

[10] Cost of Data Center Outages. Data Center Performance Benchmark Series. January 2016. Available at: https://www.vertivco.com/globalassets/documents/reports/2016-cost-of-data-center-outages-11-11_51190_1.pdf

PATRICIA TAKAKO ENDO is a professor in the Universidade de Pernambuco – Campus Caruaru and a researcher at Grupo de Pesquisa em Redes de Computadores e Telecomunicações. Contact her at patricia.endo@upe.br.

GUTO LEONI is a Master student at Universidade Federal de Pernambuco and a researcher at Grupo de Pesquisa em Redes de Computadores e Telecomunicações. Contact him at guto.leoni@gprt.ufpe.br

DANIEL ROSENDO is a Master researcher at Grupo de Pesquisa em Redes de Computadores e Telecomunicações. Contact him at daniel.rosendo@gprt.ufpe.br

DEMIS MOACIR GOMES is a Master student at Universidade Federal de Pernambuco and a researcher at Grupo de Pesquisa em Redes de Computadores e Telecomunicações. Contact him at demis.gomes@gprt.ufpe.br

ANDRÉ MOREIRA is a PhD researcher at Grupo de Pesquisa em Redes de Computadores e Telecomunicações. Contact him at andre@gprt.ufpe.br

JUDITH KELNER is an associate professor at Universidade Federal de Pernambuco. Contact her at jk@gprt.ufpe.br

DJAMEL SADOK is an associate professor at Universidade Federal de Pernambuco. Contact him at jamel@gprt.ufpe.br

GLAUCO ESTÁCIO GONÇALVES is a professor at Universidade Federal Rural de Pernambuco and a researcher at Grupo de Pesquisa em Redes de Computadores e Telecomunicações. Contact him at glauco@gprt.ufpe.br

MOZGHAN MAHLOO works as the researcher in cloud technology department of Ericsson.