Michael Crowder | Brian Kolovich | Gerardo Garza

Master of Science in Data Science Candidates at Southern Methodist University, Dallas, TX

[[1]](#footnote-1)

Virtualization Comparison with Popular Public Cloud Providers

*Statement of the Problem*—With more public cloud providers offering more user-friendly interfaces that are helping to democratize scale in software as a service businesses we hope to compare and contrast the major public cloud virtualization capabilities.

Cloud computing has reduced hardware cost and energy costs. A downside of virtualization is that the user will not get the same performance as they would on a physical system. We hope to evaluate virtual machines from some public cloud providers as well as their user interfaces to see the pros and cons of the major public cloud providers. This paper will help those new to cloud computer understand tradeoffs between the major cloud providers and which fits their organization’s needs the best.

We have not developed testing methods yet, but will as we deepen our research.

# INTRODUCTION

Each day 2.5 quintillion bytes of data are produced[[2]](#footnote-2). This data is produced when you walk, get in your car, adjust a thermostat, etc. This data known as Big Data comes in the form of structured and unstructured data that is defined by the three V’s. Big Data comes in high *volumes* from a *variety* of sources. Streams of data come in at a high *velocity* and must be cleansed and understood in a timely manner. Formats of data come in many types of *variety*, from structured like traditional tables to unstructured documents like email, video and audio.[[3]](#footnote-3) For example each day UPS delivers 20 million packages through a fleet of over 119 thousand ground vehicles and almost 2 thousand daily flights throughout the world. Their customers request 142.7 million package tracking request each day. Almost all of their vehicles contain real time telematics data and their customers are requesting and pushing data to them throughout the day. This is the very definition of big data.[[4]](#footnote-4)

The evolution of 5G cellular, Internet of Things (IOT) and data analytics is going disrupt the information and Communications Technology (ICT) ecosystem. Mobile Cloud Sensing, Big Data, and 5G Networks will enable more communication in areas such as public safety and healthcare that will need analytics in real-time. 5G networks will make it possible to sense and analyze at the edge. New analytical models will be needed to trigger decisions at the edge as well.

With the lead-up to changes with greater volumes, velocity and variety of data public cloud providers are providing solutions to help manage, analyze and store this data. Industries such as government, environmental, financial and agriculture are filled with users that are non-traditional computing professionals. The cost of deploying and maintaining large-scale dedicated server clusters can be prohibitively high. There is also a gap in technical skills required to run these large server clusters and use them effectivity [1]. Public Cloud providers have emerged in recent years that allow users to rent computing and storage in a pay-as-you-go manner.

## Public Cloud Overview and Public Cloud Providers in The United States

Public Cloud is a computing service offered by third-party providers over the public internet making it available to anyone who wants to use public cloud for usage of CPU cycles, storage or bandwidth that is consumed and paid for with on-demand services.[[5]](#footnote-5)

Cloud computing enables enterprises to obtain as much storage and computation resources as needed by only paying for the resources they require. Before cloud computing an enterprise would have to purchase hardware, talent and other resources to scale their business as needed.

The below table demonstrates some of the service models that public cloud providers distribute to their customers.

|  |  |
| --- | --- |
| Service | Description |
| Infrastructure as a Service (IaaS) | Online services that provide underlying network infrastructure like physical computing resources |
| Platform as a Service (PaaS) | The ability to deploy consumer-created or acquired applications using programming languages, libraries, services and tools supported by the provider. |
| Software as a Service (SaaS) | Applications that are accessible from client devices through either a thin client like interface such as. A web browser, or a program interface. |

The major cloud providers in the United States (U.S.) are Amazon Web Services, Microsoft Azure Google Cloud Platform and IBM Cloud.

References

1. Wang, D. & Liu, J. Optimizing Big Data Processing Performance in the Public Cloud: Opportunities and Approaches, IEEE Network, September 2015

1. [↑](#footnote-ref-1)
2. Patil, D. A Code of Ethics for Data Scientists, Medium, https://medium.com/@dpatil/2-5-quintillion-bytes-of-data-are-created-every-day-c908c951f3bf [↑](#footnote-ref-2)
3. SAS, Big Data History and Current Considerations, https://www.sas.com/en\_us/insights/big-data/what-is-big-data.html [↑](#footnote-ref-3)
4. UPS, 2018 UPS Fact Sheet, https://pressroom.ups.com/assets/pdf/pressroom/fact%20sheet/UPS\_Fact\_Sheet.pdf [↑](#footnote-ref-4)
5. Microsoft Azure, What is public cloud? Microsoft Azure Overview, https://azure.microsoft.com/en-us/overview/what-is-a-public-cloud/ [↑](#footnote-ref-5)