[1]. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M. A view of cloud computing. Communications of the ACM. 2010 Apr 1;53(4):50-8.

The problem being addressed?

The study conducted by Armbrust et al. aimed to define cloud computing and eliminate confusion by providing clear terminology. They distinguished between cloud and conventional computing using simple diagrams and briefly explained the classes of utility computing and the economics of cloud computing, emphasizing its importance when considering any platform. Additionally, the authors identified and compared the top technical and non-technical challenges and opportunities in Cloud Computing, providing a basis for further improvement.

Summary of Proposed Work?

The authors defined cloud computing as the delivery of system software, applications, and hardware over the Internet, for providing services to users. They also subclassified cloud services into three models, PAAS (Platform as a Service), SAAS (Software as a Service), and IAAS (Infrastructure as a Service). Furthermore, the authors mentioned that computation, storage, and network communication are the three fundamental classes of utility computing needed for building applications. They highlighted some products like EC2 (Compute), S3 (Storage), and App engine from various public cloud vendors such as AWS (Amazon Web Services), GCP (Google Cloud Platform), and Azure. They also presented the concept of cloud economics with necessary use cases, discussing unexpected spikes from users and the provisioning of resources by managing cost associativity. Later, the authors succinctly explained the difference between cloud and conventional computing, citing features such as infinite compute resources on demand, elimination of upfront commitment by cloud users, and the ability to pay for use on demand. This simplifies operations and improves utilization using resource virtualization. Private data centers or conventional computing

environments do not support the above principles; instead, public cloud handles the problem of under-provisioning and over-provisioning. To support these principles, the authors provided a use case from Animoto, a company that started offering its services on Facebook and experienced a spike in users. This led them to launch compute from 50 to 3500 servers in three days, and later reduced the number of servers when the demand subsided. Therefore, this problem correlates with identifying under-provisioning, scaling up based on demand, and also connects with elasticity and pay-as-you-use, addressing the operational requirement.

Furthermore, the authors identified ten obstacles and opportunities in the field of cloud computing, where the first three are around adoption, the next five affect growth, and the last two are about policy. Some of the obstacles and opportunities mentioned are regarding business continuity and availability, where everyone expects 24/7 availability. The authors suggested that having a multi-cloud strategy is the solution. Data lock-in is another issue when using the cloud because every vendor has its own formats and standards. To solve this problem, the authors suggested standardizing APIs across platforms to enable rapid adoption in hybrid cloud environments. Other obstacles include data confidentiality, transfer bottlenecks, and performance unpredictability, which can be resolved by maintaining strong firewalls, encrypting data, increasing the bandwidth of disks, and improving VM (Virtual Machine) support. Additionally, scalable storage, dynamic scaling, and bugs in large-scale distributed systems can be rectified by inventing scalable storage and creating an auto scaler technique that uses machine learning for dynamic scaling, and by developing a debugger specifically for distributed VMs. Lastly, for reputation fate sharing and software licensing, the authors advocated having reputation-guarding services and pay-for-use licenses.