Ahmedabad University

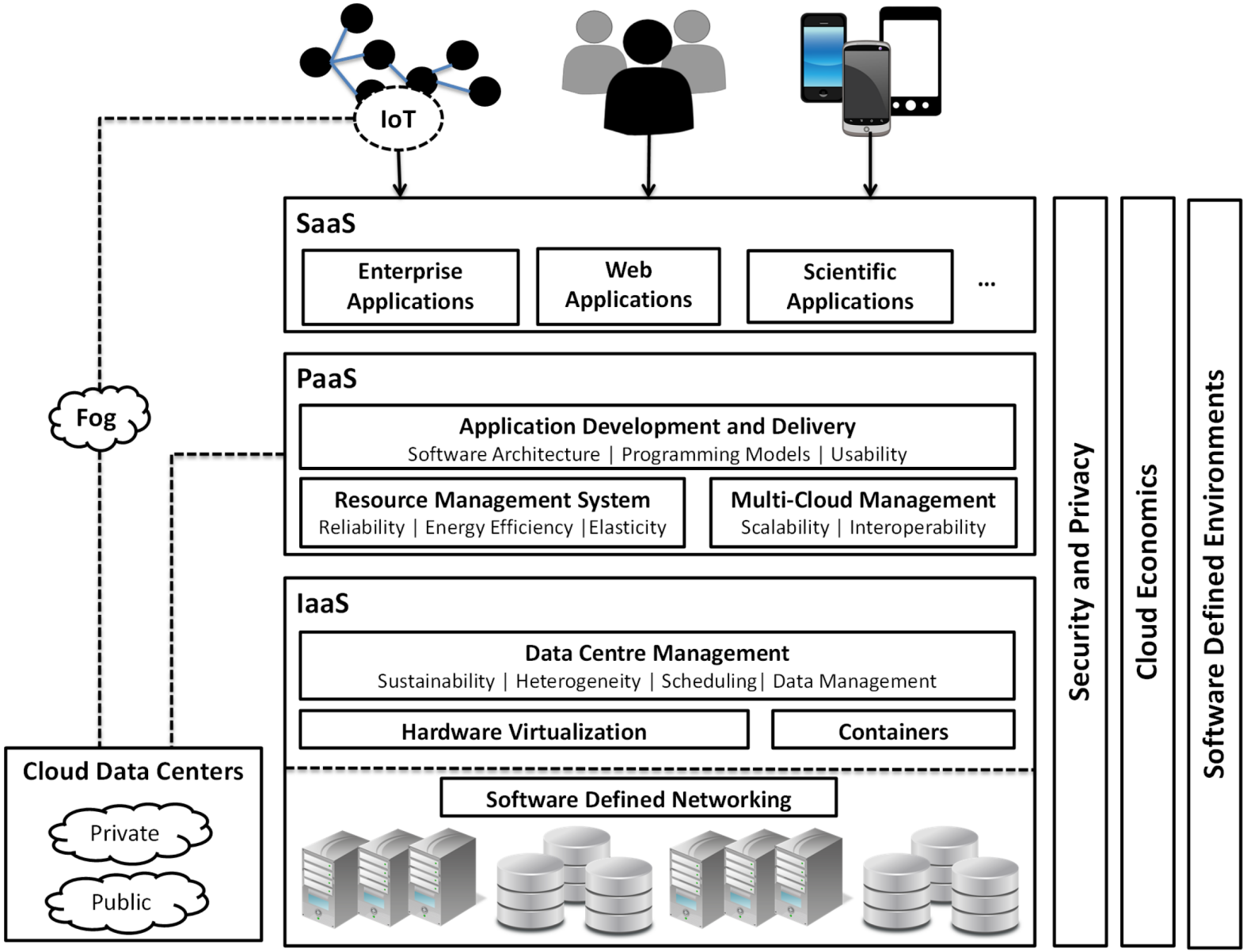
Research Advances in Cloud Computing

Sanjay Chaudhary
Interim Dean and Professor,
School of Engineering and Applied Science,
Dean of Students, Ahmedabad University

Motivation

Ahmedabad University

- IDC's (International Data Corporation) World- wide Semiannual Public Cloud Services Spending Guide reported that Cloud services were expected to grow from $70 billion in 2015 to more than $203 billion in 2020
- an annual growth rate almost seven times the rate of overall IT spending growth.
- This extensive usage of Cloud computing in various emerging domains is posing several new challenges and is forcing us to rethink the research strategies and re-evaluate the models that were developed to address issues such as
  - scalability, resource management, reliability, and security for the realisation of next-generation Cloud computing environments



Components of the Cloud computing paradigm

Fog Computing*

Ahmedabad University

- Provides a distributed infrastructure at the edge of the network, resulting in low-latency access and faster response to application requests.
- With this new level of computing capacity, new forms of resource allocation and management can be developed to take advantage of the fog infrastructure.

- ***Source:** 'Mobility-Aware Application Scheduling in Fog Computing', Luiz F. Bittencourt, Javier Diaz-Montes, Rajkumar Buyya, Omer F. Rana, and Manish Parashar, IEEE Cloud Computing Published By The IEEE Computer Society

Fog Computing (cont)

Ahmedabad University

- With Internet-of-Things (IoT), countless devices scattered and connected to the Internet, producing and consuming data requires scalable resource management at unprecedented levels.
- The data dynamism and heterogeneity resulting from this expected explosive expansion of connected devices, commonly referred in a broad sense as Big Data, also requires new processing models and infrastructures to support its main dimensions: data volume, variety, velocity, veracity, value...

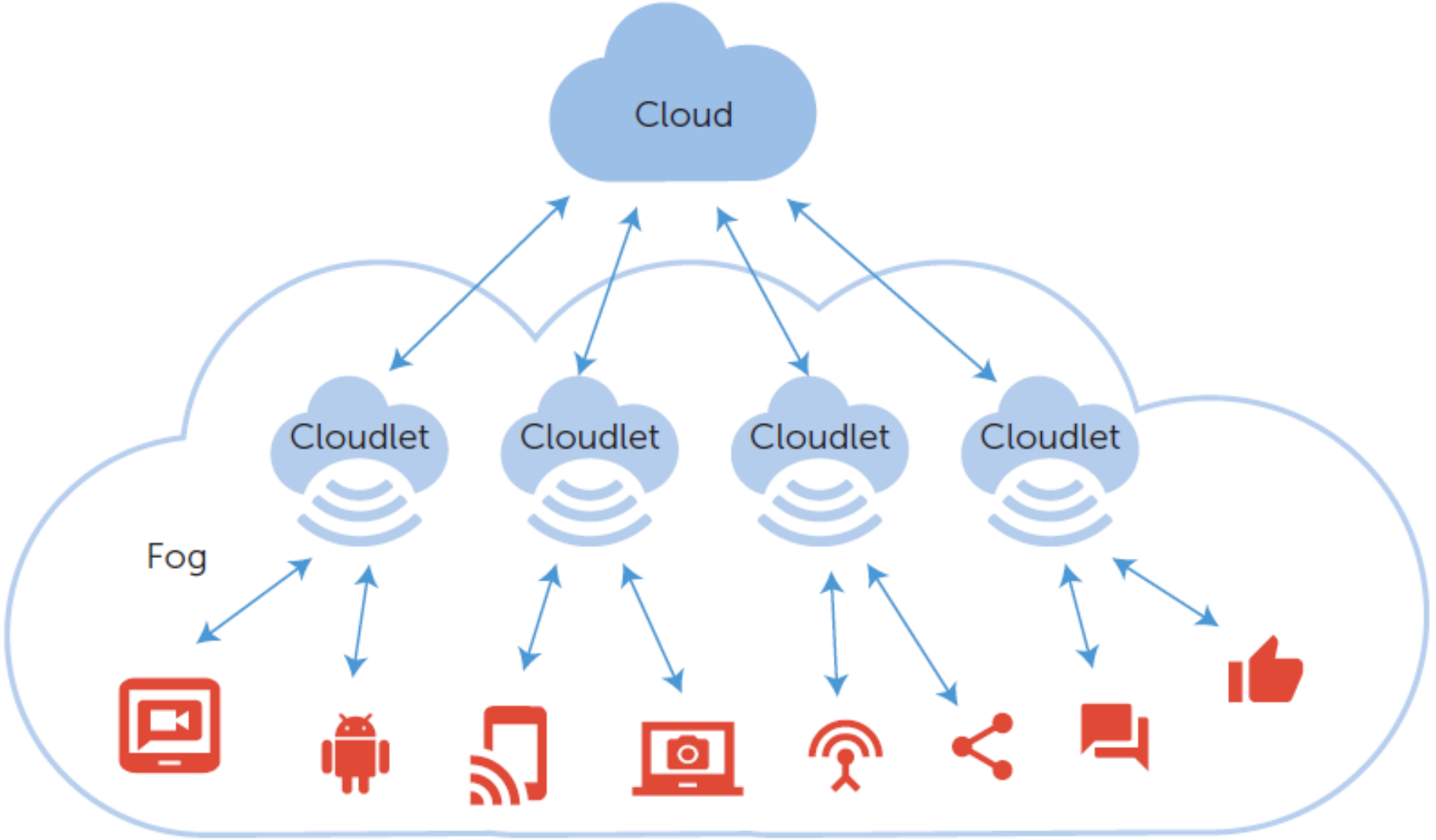


FIGURE 1. Fog computing: cloud, cloudlets and edge devices/ applications ecosystem.

Cloudlets

Ahmedabad University

- With the introduction of computing capacity at the edge of the network, these access points can be extended to also provide computing and storage services referred as the cloudlets.
- Previous Figure illustrates the cloudlets concept within the hierarchical infrastructure of the fog.
- This fog computing architecture presents a hierarchical, bi-directional computing infrastructure: edge devices communicate with cloudlets and cloudlets communicate with clouds.
- Cloudlets can also communicate with each other to perform data and process management in order to support application requirements, and to exchange fog control/management data (such as user device and application state).

Applications to benefit

Ahmedabad University

- Applications with low latency requirements, such as
  - pedestrian and traffic security, surveillance, applications for vision, hearing, or mobility impaired users, online gaming, augmented reality, etc. can benefit from lower latencies because of a single hop connection to a cloudlet.
- Raw data collected by many devices often does not need to be transferred to the cloud for long term storage:
  - data can be processed, filtered, or aggregated to extract knowledge and produce reduced data sets, which in turn are to be stored; or it can be processed and utilized right-away to other edge devices in the so-called sensor/actuator loop.
- In both cases, the fog computing paradigm can reduce network traffic from the edge to data centers.
- Cloudlets can provide reduced latencies and help in avoiding/reducing traffic congestion in the network core.

Challenges

Ahmedabad University

- In fog computing, sensors and other devices pervasively present at the edge of the fog generate data and consume data that have to be processed using the cloudlets and the clouds.
- More complex and sophisticated resource management and scheduling mechanisms are needed.
- This raises new challenges to be overcome, e.g.,
  - dynamically deciding what, when, and where (device/fog/cloud) to carry out processing of requests to meet their quality of service requirements.

Challenges

Ahmedabad University

- Furthermore, with smart and wearable devices, such mechanisms must incorporate mobility of data sources and sinks in the fog.
- Traditional resource management and scheduling models for distributed systems do not consider mobility and timeliness of data production and consumption in the resource management and allocation process.
- Fog computing scheduling must bring user's location to the resource allocation policies to uphold the benefits of fog computing proximity to the user.

Challenges (cont)

Ahmedabad University

- It aims at moving decision making operations as close to the data sources as possible by leveraging resources on the edge such as
  - mobile base stations, gateways, network switches and routers, thus reducing response time and network latencies.
- Additionally, as a way of fulfilling increasingly complex requirements
  - the composition of multiple services and
  - as a way of achieving reliability and improving sustainability, services spanning across multiple geographically distributed CDCs have also become more widespread.

Microservices

Ahmedabad University

- By 2013, the first major design pattern for cloud native applications began to emerge.
- It was clear that to achieve scale and reliability, it was essential to decompose applications into very basic components, which we now refer to as microservices.
- Microservice paradigm design rules dictate that each microservice must be managed, replicated, scaled, upgraded, and deployed independently of other microservices.

Microservices

Ahmedabad University

- Each microservice must have a single function and operate in a bounded context; that is, it has limited responsibility and limited dependence on other services.
- All microservices should be designed for constant failure and recovery and therefore they must be as stateless as possible.
- Representational State Transfer web service calls, remote procedure call (RPC) mechanisms such as
  - Google's Swift, and
  - the Advanced Message Queuing Protocol.

Application Scenario

Ahmedabad University

- Suppose you have a mechanism to continuously mine on-line data sources, looking for articles with scientific content.
- You want to build a cloud-native application that will use machine learning to characterize and classify the topic of these articles and store this information in a database.
- Three types of microservices.
  - M1–M6 are the data monitors that skim the relevant RSS feeds and news sites looking for relevant sounding articles.
  - A1–A5 are analysis services and
  - D1–D3 are database managing services.

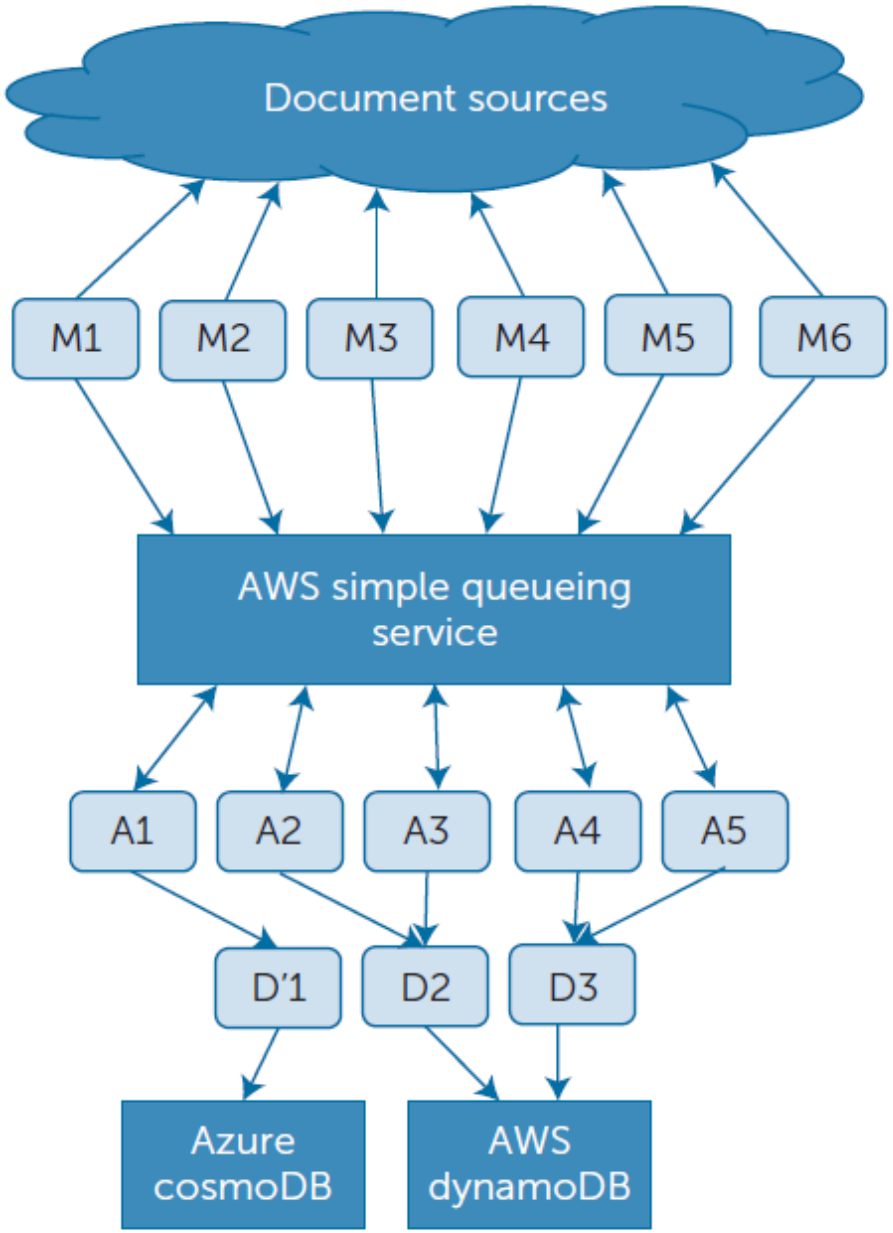


FIGURE 1. Simple microservice cloud-native application.

Serverless Computing

Ahmedabad University

- The major disadvantage of the microservice model as illustrated in the previous example is that we still need to provision a cluster of compute resources to run the services, then manage and scale these compute resources
- Serverless computing is a style of cloud computing where you write code and define the events that should cause the code to execute and leave it to the cloud to take care of the rest.
- AWS's approach to this is called Lambda. For Azure it is called Functions and for Google it is called Cloud Functions.
- The concept is very simple.
- In the case of AWS Lambda, examples of the types of events that can trigger the execution of the function are the arrival of a new object in the S3 storage system, or the arrival of a record in the Amazon Kinesis data streaming service.

Serverless Computing (cont)

Ahmedabad University

- “fully managed” services because the service manages all of the infrastructure resourcing, management, and scaling, along with the workflow needed to carry out your computation.
- There is no need for the user to allocate resources.
- For example, Azure CosmosDB allows a user to add their own functions and procedures to their databases.
- These functions are executed by triggers or by user queries.

Book: '**Research Advances in Cloud Computing**'

Editors: Sanjay Chaudhary, Gaurav Somani, and Rajkumar Buyya

Springer Nature, ISBN 978-981-10-5025-1, 2017

Ahmedabad University

Preface

- This book on “Research Advances in Cloud Computing” discusses various new trends, designs, implementations, outcomes, and directions in the various areas of cloud computing.
- This book has been organized into three sections:
  1. Programming model, infrastructure, and runtime
  2. Resource Management
  3. Security

Research challenges [1]

Ahmedabad University

- Automated service provisioning
- Virtual machine migration
- Server consolidation
- Energy management
- Traffic management and analysis
- Data security
- Software frameworks
- Storage technologies and data management
- Novel cloud architectures

Challenges, Open Issues and Future Research Direction [2]

Ahmedabad University

- Scalability and Elasticity
- Resource Management and Scheduling
- Reliability
- Sustainability
- Heterogeneity
- Interconnected Clouds
- Empowering Resource-Constrained Devices
- Security and Privacy

Challenges, Open Issues and Future Research Direction (cont) [2]

Ahmedabad University

- Economics of Cloud Computing
- Application Development and Delivery
- Data Management
- Networking
- Usability

Discussion [2]

Ahmedabad University

- IaaS
  - Heterogeneous hardware such as CPUs and accelerators (e.g., GPUs and TPUs) and special purpose Clouds for specific applications (e.g., HPC and deep learning).
  - The future generation Clouds should also be ready to embrace the non-traditional architectures, such as neuromorphic, quantum computing, adiabatic, nanocomputing, and so on.
  - Emerging trends such as containerisation, SDN and Fog/Edge computing
- PaaS
- SaaS

Discussion [2]

Ahmedabad University

- PaaS
  - Resource management and scheduling.
  - The need for programming abstractions, models, languages and systems supporting scalable elastic computing and seamless use of heterogeneous resources are proposed leading to energy-efficiency, minimised application engineering cost, better portability and guaranteed level of reliability and performance.
  - It is also foreseeable that the ongoing interest for ML, deep learning, and AI applications will help in dealing with the complexity, heterogeneity, scale and load balancing applications developed through PaaS.
  - Serverless computing is an emerging trend in PaaS, which is a promising area to be explored with signifcant practical and economic impact.
  - Interesting future directions are proposed such as function-level QoS mannagement and economics for serverless computing.
  - In addition, future research directions for data management and analytics are also discussed in detail along with security, leading to interesting applications with platform support such as edge analytics for real-time stream data processing, from the IoT and smart cities domains.

Discussion [2]

Ahmedabad University

- SaaS
  - SaaS should mainly see advances from the application development and delivery, and usabil- ity of Cloud services.
  - Translucent programming models, languages, and APIs will be needed to enable tackling the complexity of application development while permitting control of application delivery to future-generation Clouds.
  - A variety of agile delivery tools and Cloud standards (e.g., TOSCA) are increasingly being adopted during Cloud application development.
  - The future research should focus at how to continuously monitor and iteratively evolve the design and quality of Cloud applications.

Discussion [2]

Ahmedabad University

- SaaS
  - extend DevOps methods and de ne novel programming abstractions to include within existing software development and delivery methodologies, a support for IoT, Edge computing, Big Data, and serverless computing.
  - Focus should also be at developing effective Cloud design patterns and development of formalisms to describe the workloads and workflows that the application processes, and their requirements in terms of performance, reliability, and security are strongly encouraged.
  - It is also interesting to see that even though the technologies have matured, certain domains such as mobile Cloud, still have adaptability issues.
  - Novel incentive mechanisms are required for mobile Cloud adaptability as well as for designing Fog architectures.

Ahmedabad University

References

1. ‘Research Advances in Cloud Computing’, Editors: Sanjay Chaudhary, Gaurav Somani, and Rajkumar Buyya, Springer Nature, ISBN 978-981-10-5025-1, 2017
2. ‘A Manifesto for Future Generation Cloud Computing: Research Directions for the Next Decade’, Rajkumar Buyya et. al., ACM Computing Surveys, Vol. 51, No. 5, Article 105. Publication date: November 2018

Ahmedabad University

Thank you for your kind attention