THE 3RD SUSTAINABILITY AND RESILIENCE CONFERENCE: CLIMATE CHANGE
November 15-16, 2021, Online

Facilitating Role of Cloud Computing in Driving Big Data Emergence

Teah Yi Fan [1] & Dr. Muhammad Ehsan Rana [2]

School of Computing

Asia Pacific University of Innovation & Technology, Malaysia

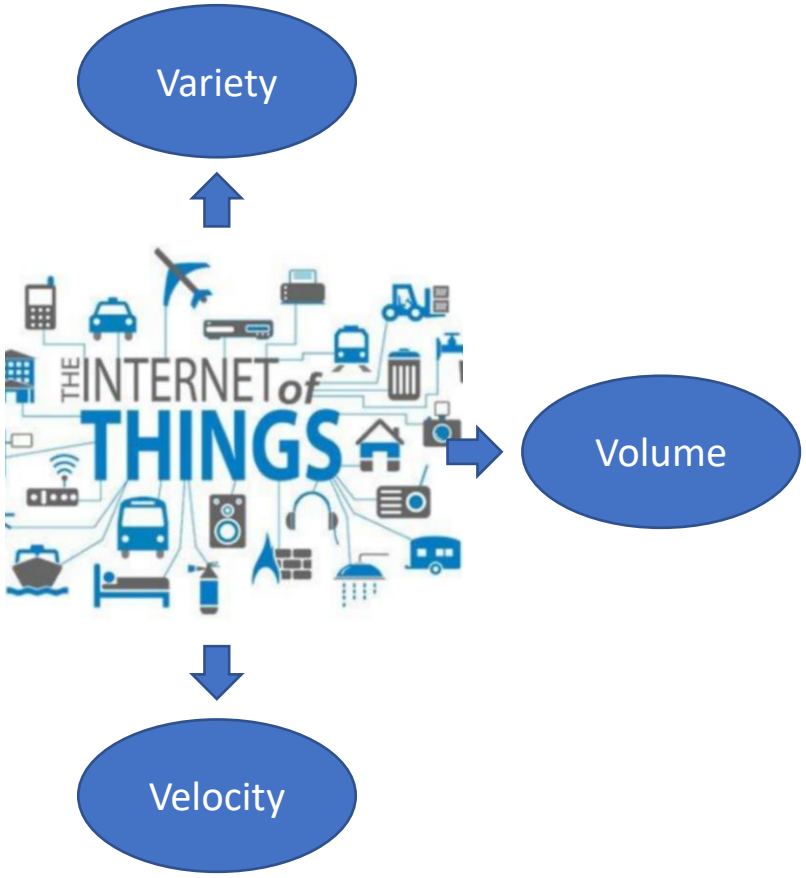
[1] tp061817@mail.apu.edu.my

[2] muhd_ehsanrana@apu.edu.my

Agenda:

1. **Introduction**
   - Big Data Emergence
   - Concepts of Cloud Computing
2. **Cloud Deployment Models**
3. **Cloud Service Models**
4. **Advantages, Risk and Challenges of Driving Big Data in Cloud Environment**
5. **Conclusions and Future Research**

Introduction:
Big Data Emergence



Business Insights

Query, Analysis
Reports

Descriptive, Predictive,
Prescriptive Analytics

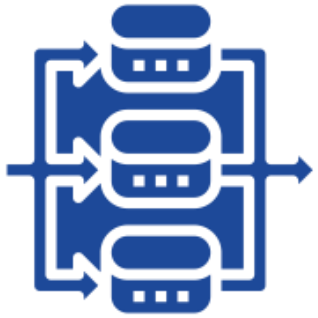
Storage **Processing** **Analytics**

Big Data Technologies: Parallel storage and Processing Framework
Human Genome Project : personalised treatment and therapy
IoT-based Smart Manufacturing : predictive maintenance
IoT-based Smart Cities : energy consumption and environmental preservation

Introduction: Virtualisation via Cloud Computing

Cloud Deployment Model

Cloud Service Model



Business Insights

Query, Analysis
Reports

Descriptive, Predictive,
Prescriptive Analytics

Storage

Processing

Analytics

Cloud Computing Technologies: Virtualisation
On demand self service
Measured services
Broad network access
Rapid elasticity
Resource pooling

Cloud Deployment & Service Models

Where is the cloud and what is in the cloud?

Review of Cloud Deployment Models

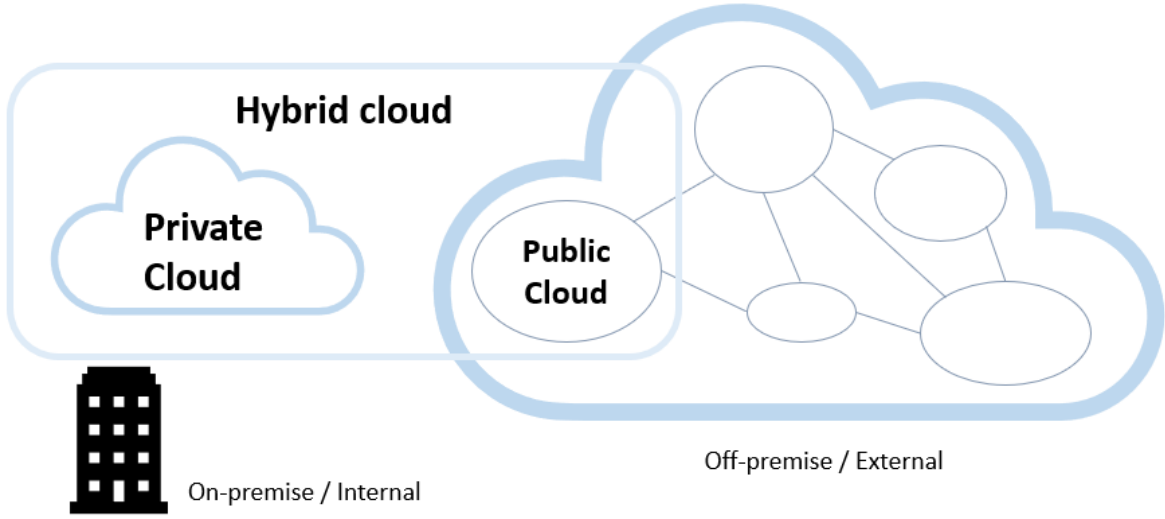
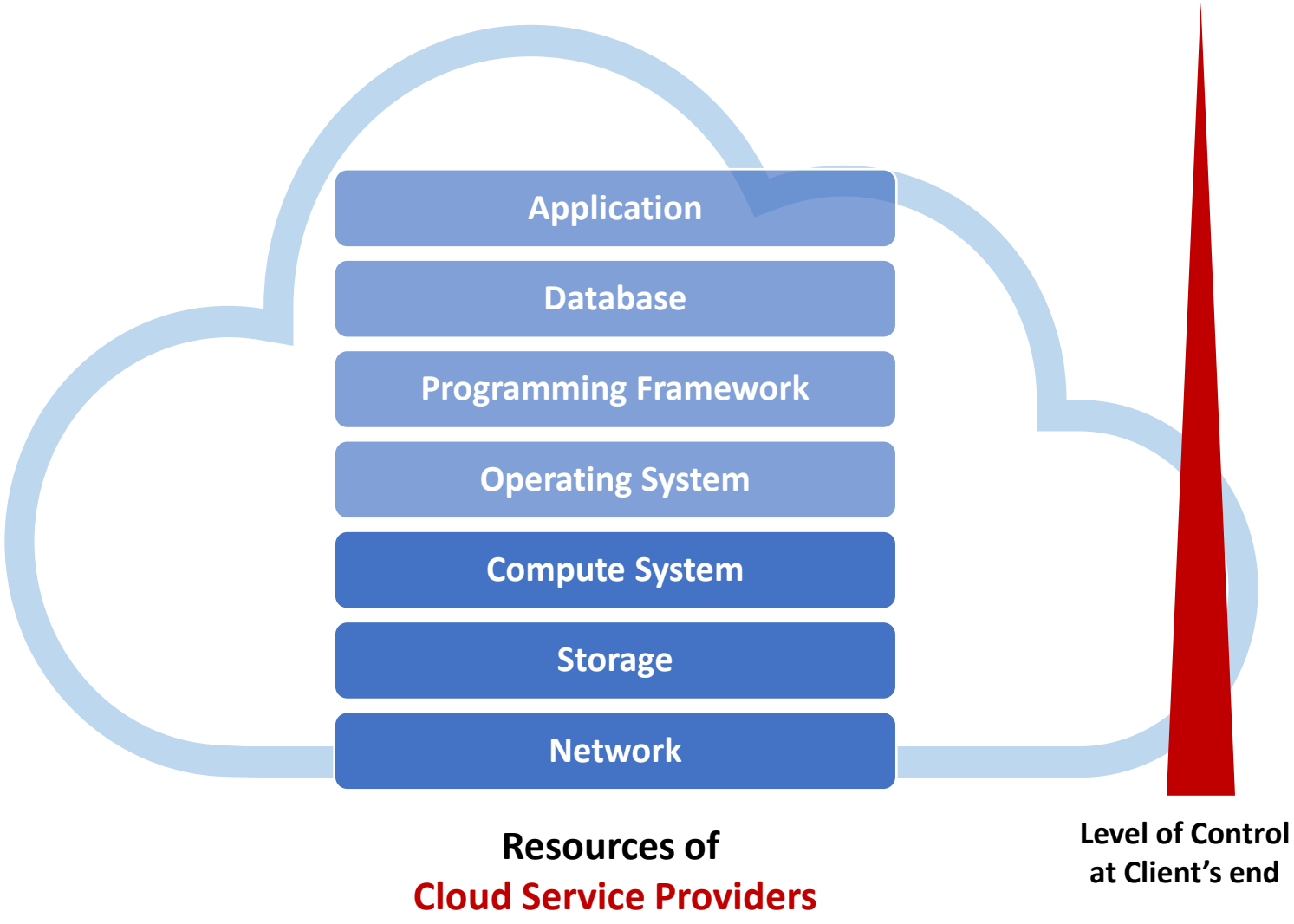


TABLE I. COMPARISONS OF REQUIREMENT FOR ADOPTION OF CLOUD DEPLOYMENT MODELS

	Public	**Private**	**Community**	**Hybrid**
Infrastructural and Technological setup	Minimal	Requires specialised cloud proficiency	Requires specialised cloud proficiency	Requires specialised cloud proficiency
Access and Control to Data	Low	High	High	High
Scalability and flexibility	High	High	Fixed capacity	High
Cost of adoption	The most cost effective model	Cost-intensive; the most expensive model	Cost is shared among community members	Depends on model
Requirements on In-house On-premise	No	Depends on on-premise or off-premise	Depends on on-premise or off-premise	Depends on the model combinantion

Cloud Service Models

Cloud Clients
Web browser, mobile app, thin client, terminal emulator



SaaS
Software-as-a-Service
Consume
Cloud Service Providers: ALL
Cloud Service Client: None (Configurations only)

PaaS
Platform-as-a-Service
Build
Cloud Service Providers: Compute System, Storage, Network, OS, Programming framework, database
Cloud Service Client: Application

IaaS
Infrastructure-as-a-Service
Host
Cloud Service Providers: Compute System, Storage, Network
Cloud Service Client: OS, Programming framework, database, application

Cloud Service Models that Facilitates Big Data Adoption

Database-as-a-service
DBaaS :

Storage

Big Data-as-a-service
BDaaS :

Processing

Analytics-as-a-service
AaaS :

Analytics

TABLE II. SUMMARY OF BIG DATA-RELATED CLOUD SERVICE MODELS

Big Data Application	Cloud Computing Service Model	Cloud Deployment Model	Cloud Service Provider	Platform	Framework/Technology/Infrastructure
DBaaS Relational Storage	SaaS	Public, Private, or Hybrid	Amazon	Amazon RDS Amazon RedShift	MySQL, Oracle Database, SQL Server, PostgreSQL
			Google	Google Cloud SQL	MySQL, PostgreSQL, and SQL Server.
			Microsoft	Azure SQL Database	MySQL
			Cloudera	CDP Operational Database	MySQL, Oracle Database, PostgreSQL
DBaaS Non-relational Storage			Amazon	Amazon DynamoDB	Dynamo DB
			Google	Google Cloud Firestore Google Cloud BigTable	Mongo DB HBase
			Microsoft	Azure CosmosDB	CosmosDB
			Cloudera	Cloudera Accumulo	Apache Accumulo
BDaaS Big Data Processing Framework	PaaS		Amazon	Amazon EMR	Hadoop HDFS and MapReduce, Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi, and Presto.
			Google	Google Cloud Dataproc Google Cloud Dataflow	Apache Hadoop, Apache Spark
			Microsoft	Microsoft Azure HDInsight	Apache Hadoop, Spark, Hive, Kafka etc.
			Cloudera	Cloudera Distribution of Hadoop (CDH) Cloudera Data Hub	Hadoop HDFS and Map Reduce, YARN, Apache Spark
AaaS Batch Analytics	SaaS		Amazon	Amazon Athena	ETL, interactive query
			Google	Google Cloud BigQuery Google Data Studio	SQL query, machine learning, geospatial analysis interactive query
	PaaS & SaaS		Microsoft	Azure Synapse Analytics Azure Machine Learning Azure Analysis Services	interactive query, machine learning
	SaaS		Cloudera	DXC Analytics Platform	interactive query
AaaS Stream Analytics	SaaS		Amazon	Amazon Kinesis	Apache Flink
			Google	Google DataFlow	In-house proprietory
			Microsoft	Azure Stream analytics	In-house proprietory
			Cloudera	Cloudera DataFlow	Apache Flink

Advantages, Risk and Challenges of Driving Big Data in Cloud Environment

Advantages

- *High scalability* ensures ***business agility***
- *High availability* ensures ***business continuity*** – 24/7/365

Risk and Challenges

- ***Security***
  - intrusion of virtual layer and physical trespassing at data center
  - **Multitenancy** amplifies the velocity of attack
- ***Privacy***
  - Loss and Breaches of Sensitive Data
  - Data Ownership
- ***Data governance***
  - Monetisation of client's data by the Cloud Service Providers

Conclusion and Future Research

Conclusion

- Investment into **Big Data storage, processing and analytics infrastructures** is the bottleneck of Big Data adoption
- **Cloud computing** solves the issues by providing the resources through **virtualisation** in an *economically feasible* way.
- The **advantages outweighs** the risk and challenges especially in the *small and medium scale enterprises.*

Future Research

- The availability of cloud services are overwhelmed.
  i. **Standardisation of service quality** among and across the cloud providers.
  ii. **Open standards for interoperability and security** may warrant further development

Acknowledgement

The work described in this paper was supported by the Asia Pacific University of Technology & Innovation (APU) Malaysia.

The authors of this paper are also thankful to the University of Bahrain that provided the opportunity to present this paper free of cost under IEEE platform.

IEEE
Advancing Technology for Humanity
BAHRAIN IEEE SECTION

UNIVERSITY OF BAHRAIN

A·P·U
ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION