



# Cloud Data Management: A Report

**Shubham Kapoor**  
**University of Helsinki**



# Outline

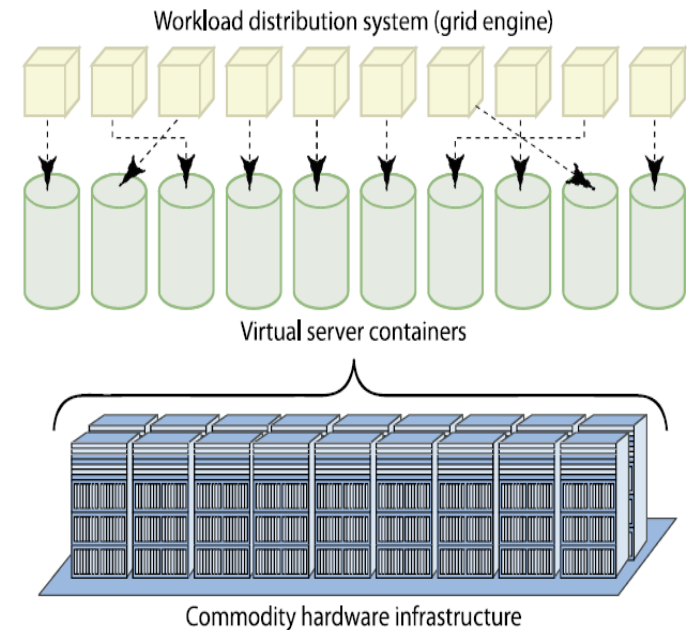
- Background
- Cloud Characteristics
- Cloud Data Management
  - Motivation
  - Wishlist
  - Challenges
- Cloud Data Management Applications
  - Transactional Data Management
  - Analytical Data Management
- Conclusion
- Future Directions





# Cloud Computing

- **Cloud Computing:** Delivery of computing services—servers, storage, databases, networking, software, analytics and more—over the Internet.
- “A model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” : working definition by **US-NSIT**



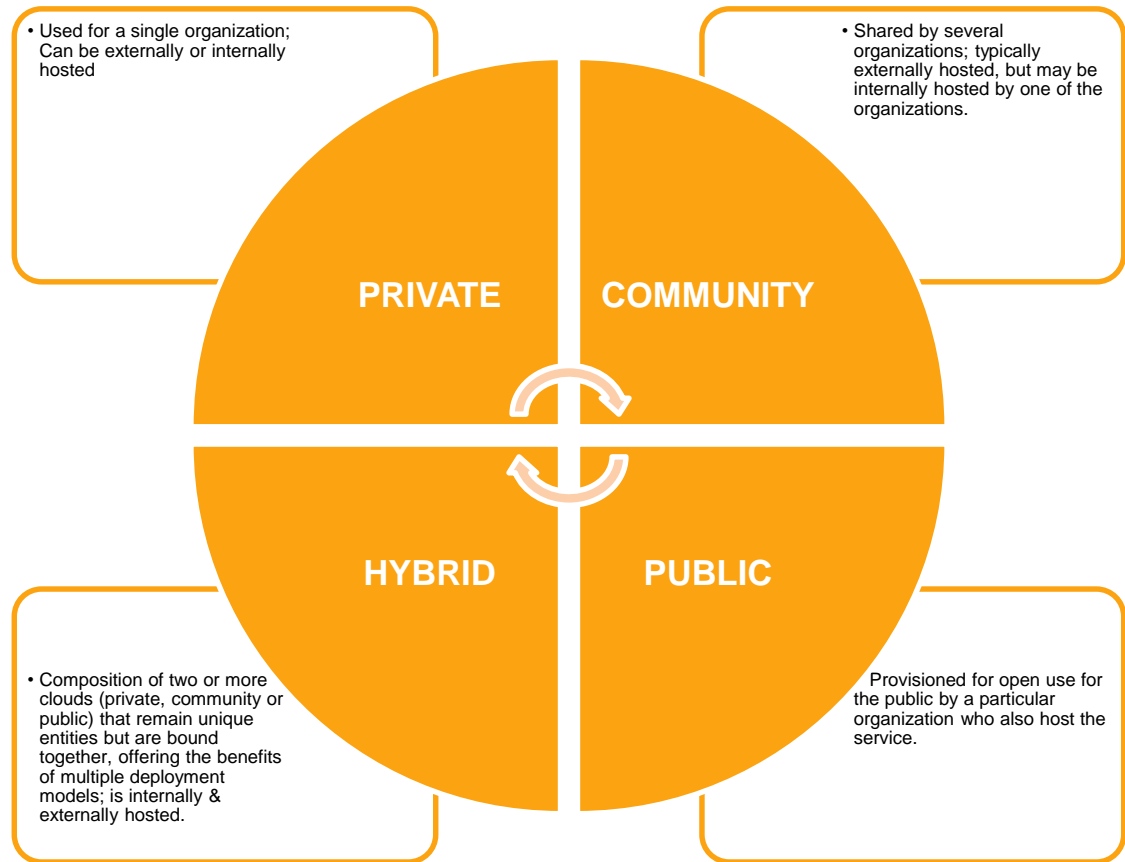
Virtualization in cloud Computing [2]



# Cloud Deployment Models

## ➤ Cloud Computing has 4 deployment models

- Private Cloud :
- Community Cloud
- Public Cloud
- Hybrid Cloud

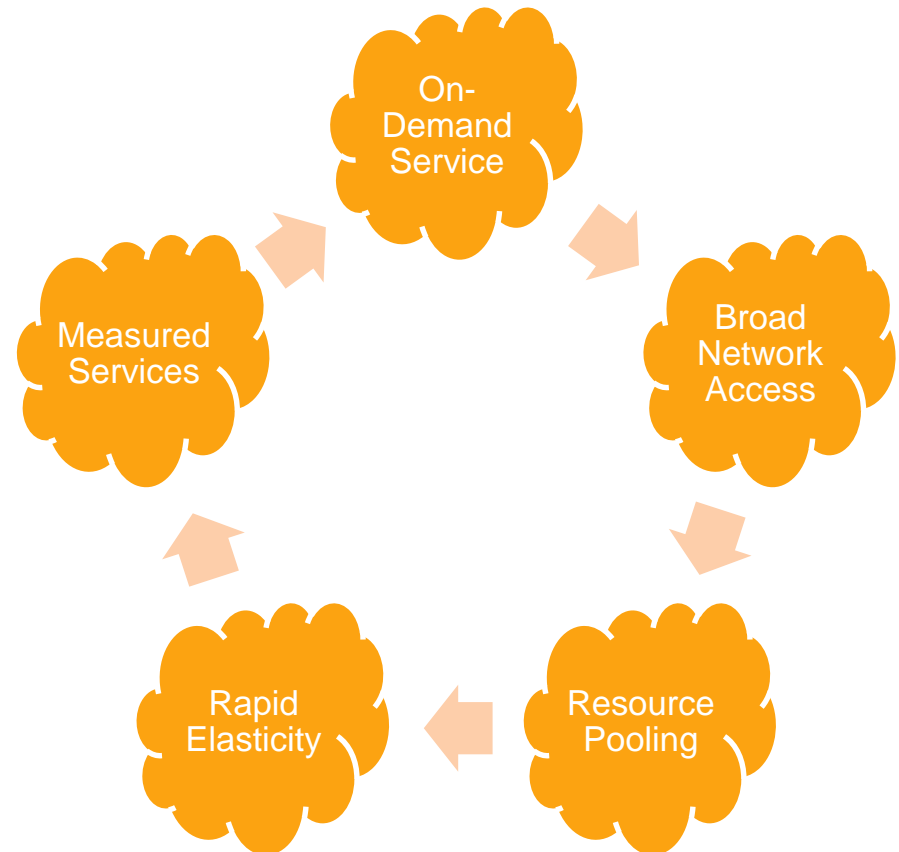




# Cloud Computing Characteristics

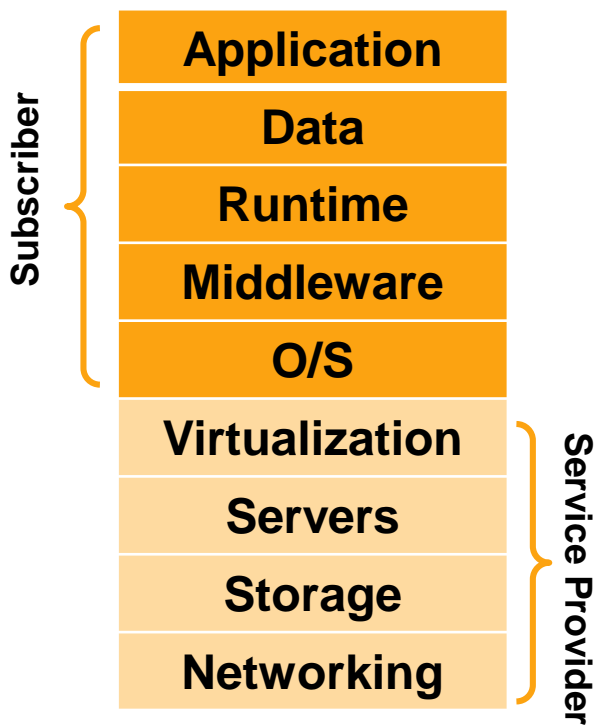
## ➤ Cloud Computing has 5 essential characteristics

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured Service

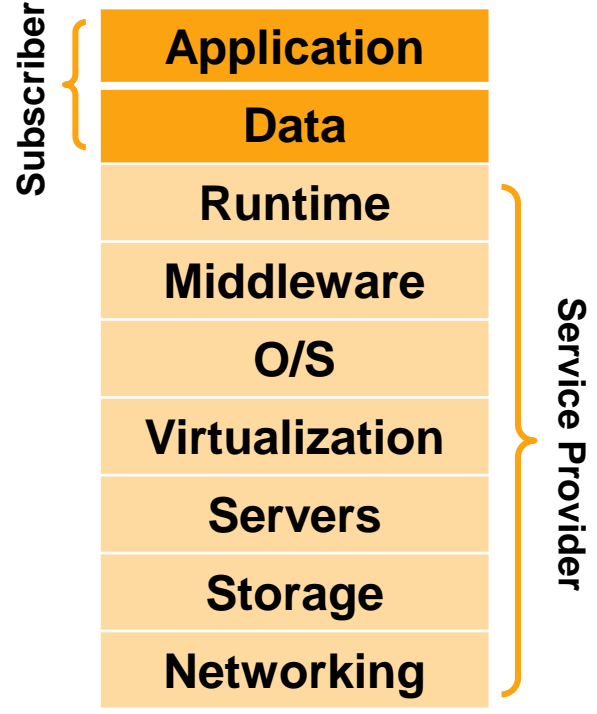




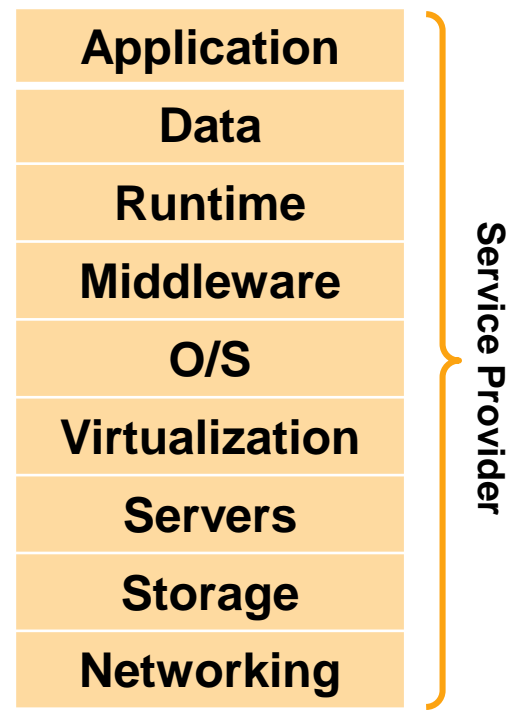
# Cloud Computing Service Models



Infrastructure as a Service (IaaS)  
Amazon EC2, Amazon S3,  
Rackspace....



Platform as a Service (PaaS)  
Google AppEngine, Microsoft Azure  
Amazon RDs....



Software as a Service (SaaS)  
Salesforce.com, Google Apps,  
Microsoft Apps...

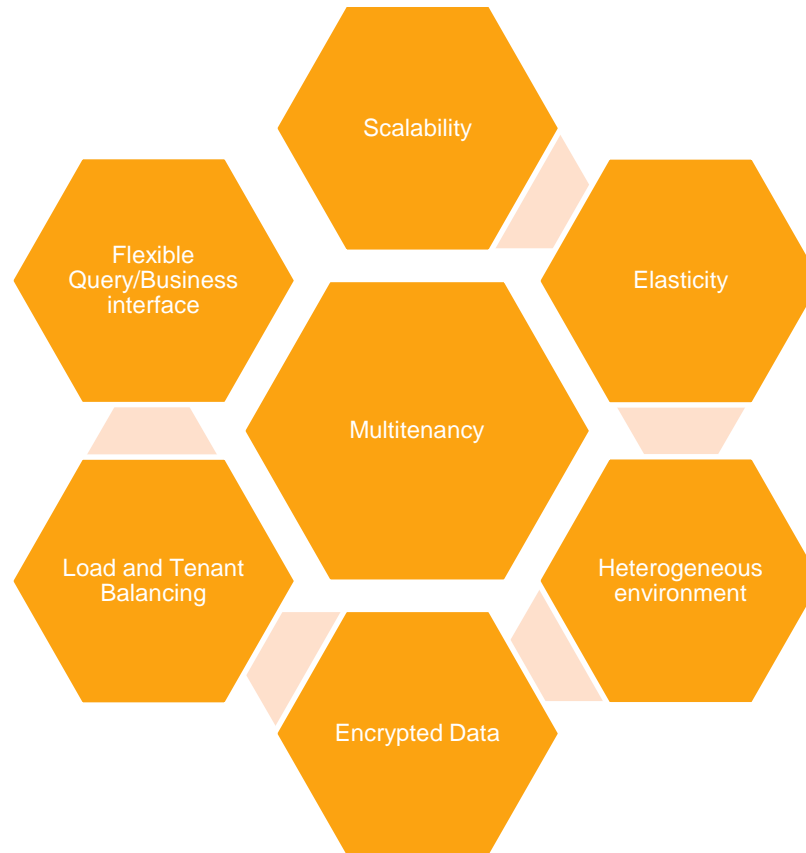


# Cloud Data Management: Motivation

- **Data Management Applications** : Potential Candidate for deployment in cloud
- **On-premises Enterprise Database System:**
  - Large Upfront Cost (Both Hardware & Software),
  - Maintenance of Hardware,
  - May be prohibitive for many companies (Start-Ups & Mid-Sized companies)
- **Cloud based Data Management System :**
  - Pay as You go computing model
  - No worry for hardware
  - Lucrative platform for many companies.
- Cloud Computing reminiscent of Application Service provider and Database as a Service (DaaS) paradigm



# Cloud Data Management Applications : Wishlist







# Cloud Data Management Applications: Challenges





# Cloud Data Management Applications : Types

## Transactional Data Management Applications



DB2 on Cloud  
IBM

SYBASE®

## Analytical data management Applications

TERADATA®





# Transactional Data Management Applications

- Used for back banking, airline reservation, online e-commerce, and supply chain management applications.
- Rely heavily on the ACID guarantees
  - ACID (Atomicity, Consistency, Isolation and Durability)
  - Hard to maintain all ACID guarantees together.
  - Consistency usually compromised over Availability while replicating data over geographical regions.
  - Example: Amazon's SimpleDB and PNUTS relax consistency by using eventual consistency model, Google BigTable relax Atomicity



# Transactional Data Management Applications

- Shared-nothing architecture : Not trivial to use.
  - Shared Nothing Architecture : A set of independent machines accomplish a task with minimal resource overlap
  - Recommended for cloud applications
  - None of Oracle, IBM DB2, Microsoft SQL Server, and Sybase support it fully
- Enormous risks in storing transactional data on an untrusted host
  - Transactional databases typically contain complete set of operational data
  - Operational data may contain includes sensitive information like customer data or credit card numbers
- Thus Not Ideal Applications for Cloud deployment.



# Analytical data management Applications

- Used for business planning, problem solving and decision support
- Scale of operation is generally larger than the transactional data management applications.
- Shared-Nothing Architecture good match for application hosting
  - increasing amount of data involved in the data analysis workload:  
driver for use of this architecture



# Analytical data management Applications

- Strict ACID guarantees are not needed.
  - Consistency can be relaxed.
  - Ok for applications to perform the analysis on a recent snapshot of the data (rather than on up-to-the-second most recent data)
- Sensitive data can be left out of analysis or anonymized
- Thus Suitable for deployment in cloud



# Analytical Data Management Applications: Cloud Deployment

- Two solutions are available for deployment of Analytical data management Applications in Cloud
- Using MapReduce-like software
  - Example Microsoft's Dryad/SCOPE stack, open source Hadoop etc.
- Shared-Nothing Parallel Databases
  - Example Teradata, Netezza, IBM DB2, Greenplum

Wishlist	Map Reduce Like Software	Shared-Nothing Parallel Databases
<b>Fault Tolerance</b>	Yes	Yes
<b>Hetrogenous Enviornment</b>	Yes	No
<b>Encrypted Data</b>	No	No
<b>Business Interface</b>	No	Yes
<b>Efficiency</b>	Yes	Yes



# Conclusion

- Transactional Data Management Applications not ideally suitable for cloud deployment.
- Analytical Data Management Applications suitable for cloud deployment.
  - But available deployment models not ideally satisfy cloud DBMS wishlist.
- Need for Hybrid solutions which fulfill all of cloud data management application wishlist.
- Need for hybrid solution that combines
  - Fault Tolerance
  - Heterogeneous Cluster,
  - Integrated use of Map-Reduce,
  - Performance
  - Tool pluggability using shared nothing parallel database systems.





# Future Directions

- Eye on Industry movement towards Hybrid solutions
  - The H-Store project : Claims for strict ACID guarantees but still in vision state
  - Pig at Yahoo and SCOPE project at Microsoft
    - Integrating declarative query constructs into Map-Reduce like software.
    - Allows data independence, code reusability, and automatic query optimization.
  - Greenplum and Aster Data
    - Ability to write MapReduce functions
- Interesting topics for future research
  - Effectively Combining advantages of Map-Reduce software with traditional data management applications.
  - Creating Performance enhancing data structure for new hybrid environments
  - How to balance tradeoff between fault tolerance and performance.



# References

1. Abadi, D. J., Data management in the cloud: Limitations and opportunities. IEEE Data Eng. Bull., 32,1(2009), pages 3–12.
2. Agrawal, D., El Abbadi, A., Antony, S. and Das, S., Data management challenges in cloud computing infrastructures. International Workshop on Databases in Networked Information Systems. Springer, 2010, pages 1–10
3. S. Mohammad, S. Breß and E. Schallehn , "Cloud Data Management: A Short Overview and Comparison of Current Approaches" , Proc. 24<sup>th</sup> GI- Workshop Foundations of Databases (Grundlagen von Datenbanken)



# THANK-YOU !