Cloud Workload Characterization and Monitoring

I Gde Dharma Nugraha

Agenda

Motivation

Background on Workload Characterization

Top-Level Cloud Workload Categorization

Cloud Workload Categories

Low-Level or Hardware Metrics of Computer Utilization

Cloud Management Requirements

Motivation

- Evolution of Cloud services which has different workload challenges.
- Each player in the Cloud has different business and technical needs.
- Therefore, each player has a different viewpoint regarding their Cloud Services' workload.

Background on Workload Characterization

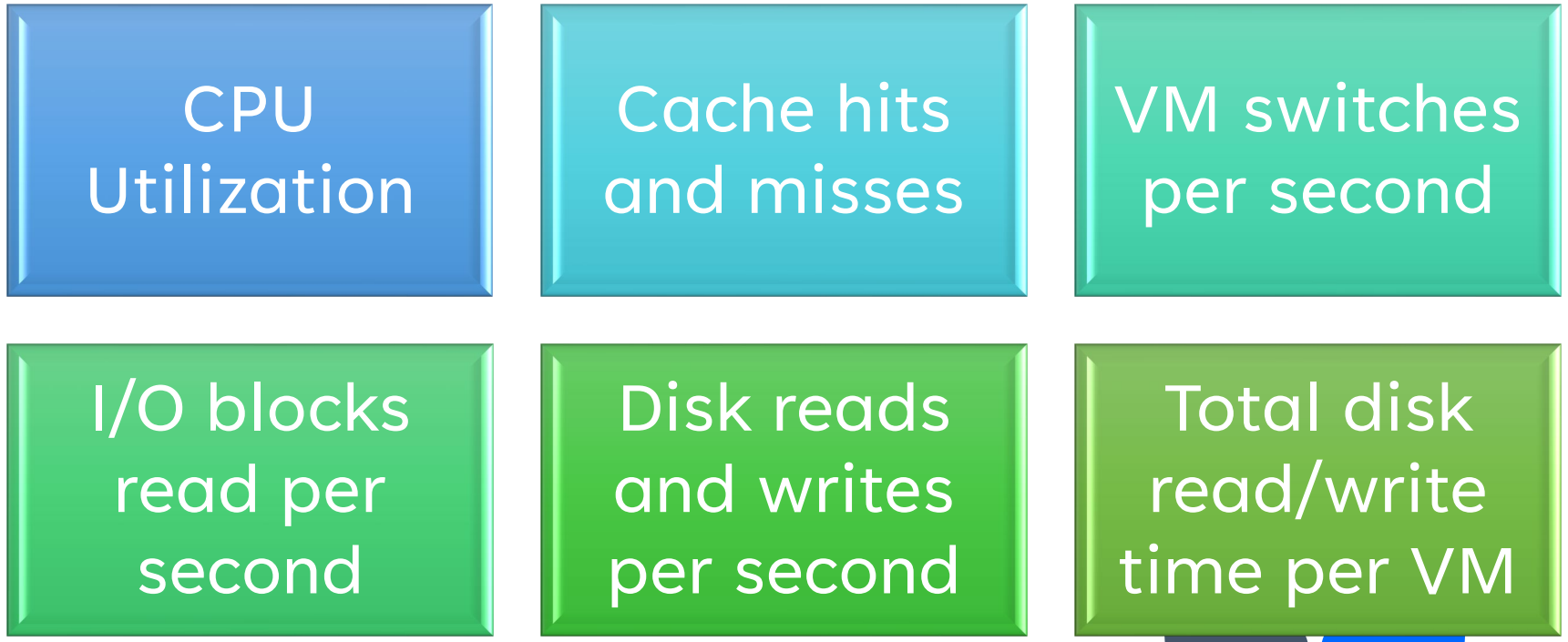
- Characterization of computer workloads has been extensively studied with many practical applications.
- Several existing studies for workload characterization have used targeted benchmarks for resource utilization.
- Jackson et al. evaluate the performance of HPC applications on Amazon's EC2 to understand the trade-offs in migrating HPC workloads to Public Cloud.
- Xie and Loh studied the dynamic memory behavior of workloads in a shared cache environment.

Background on Workload Characterization

- Xie and Loh grouped workloads into four classes:
  1. Applications that do not make much use of the cache (turtle)
  2. Applications that are not perturbed by other applications (sheep)
  3. Applications that are sensitive and will perform better when they do not have to share the cache (rabbit)
  4. Applications that do not benefit from occupying the cache and negatively impact other applications (devil)

Background on Workload Characterization

- Koh et al. [20] studied hidden contention for resources in a virtualized environment.

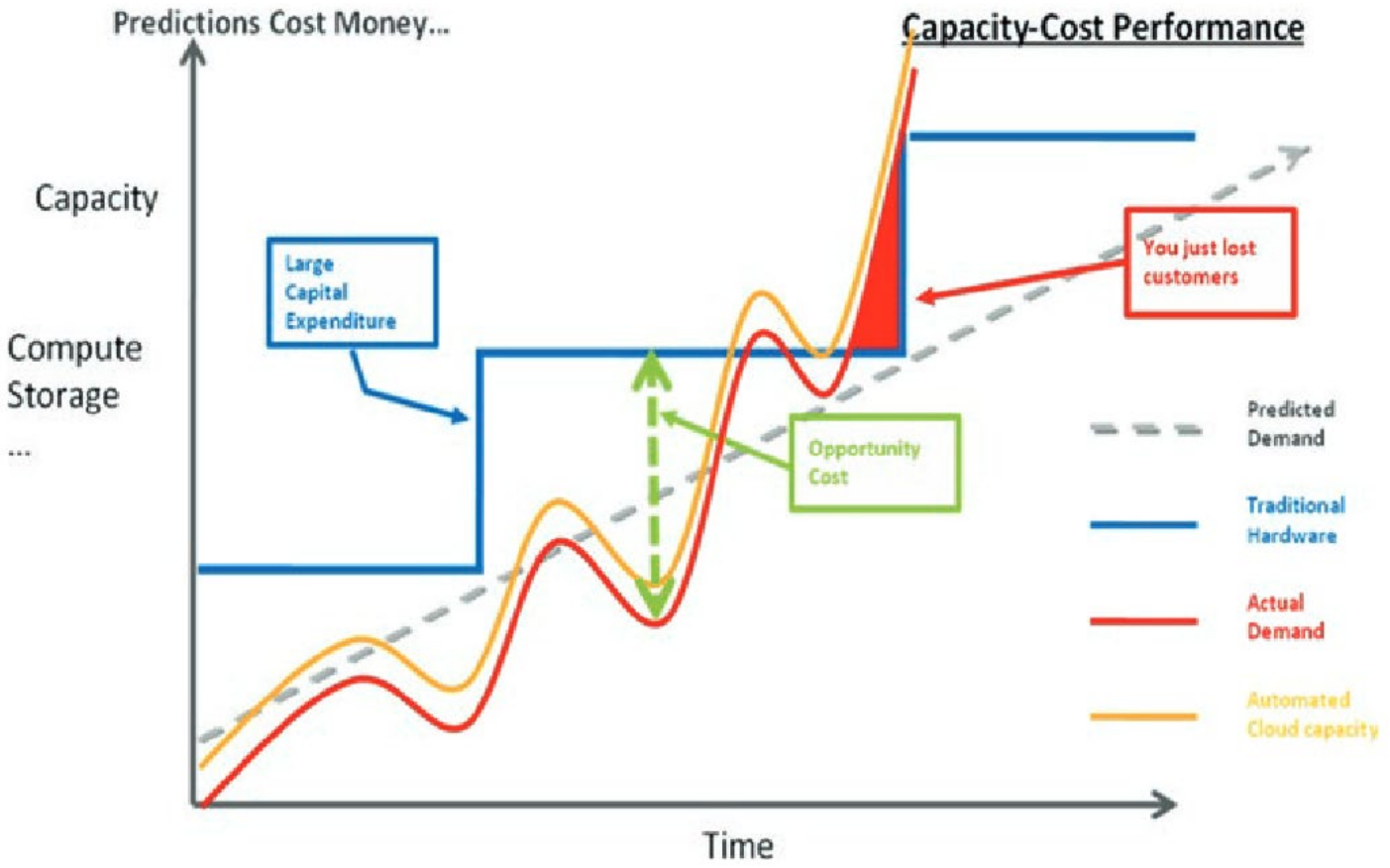


Top-Level Cloud Workload Categorization

- Cloud Computing users have a variety of needs and wants.
- Cloud Computing platform (or services) suppliers have resource choices for allocation, planning, purchasing, and pricing.
- Workload categories can be split in two ways:
  - Static architecture
  - Dynamic behavior
- Also:
  - Interactive
  - Batch-mode jobs

Top-Level Cloud Workload Categorization

- Considerations:



Cloud Workload Categories

Big Streaming Data	Big Database Calculation	Big Database Access	Big Data Storage	Many Tiny Tasks (ants)
Tightly coupled calculation intensive HPC	Highly Interactive Single Person Tasks	Highly interactive multi-person jobs	Single compute intensive jobs	Private local tasks
Slow communication	Real-time local tasks	Real-time geographically dispersed tasks,	Access Control	Voice or Video over IP

Computing Resources

Persistent Storage

Compute power/Computational Capability

Network bandwidth

Broadcast transmission receivers

Data buses within a server

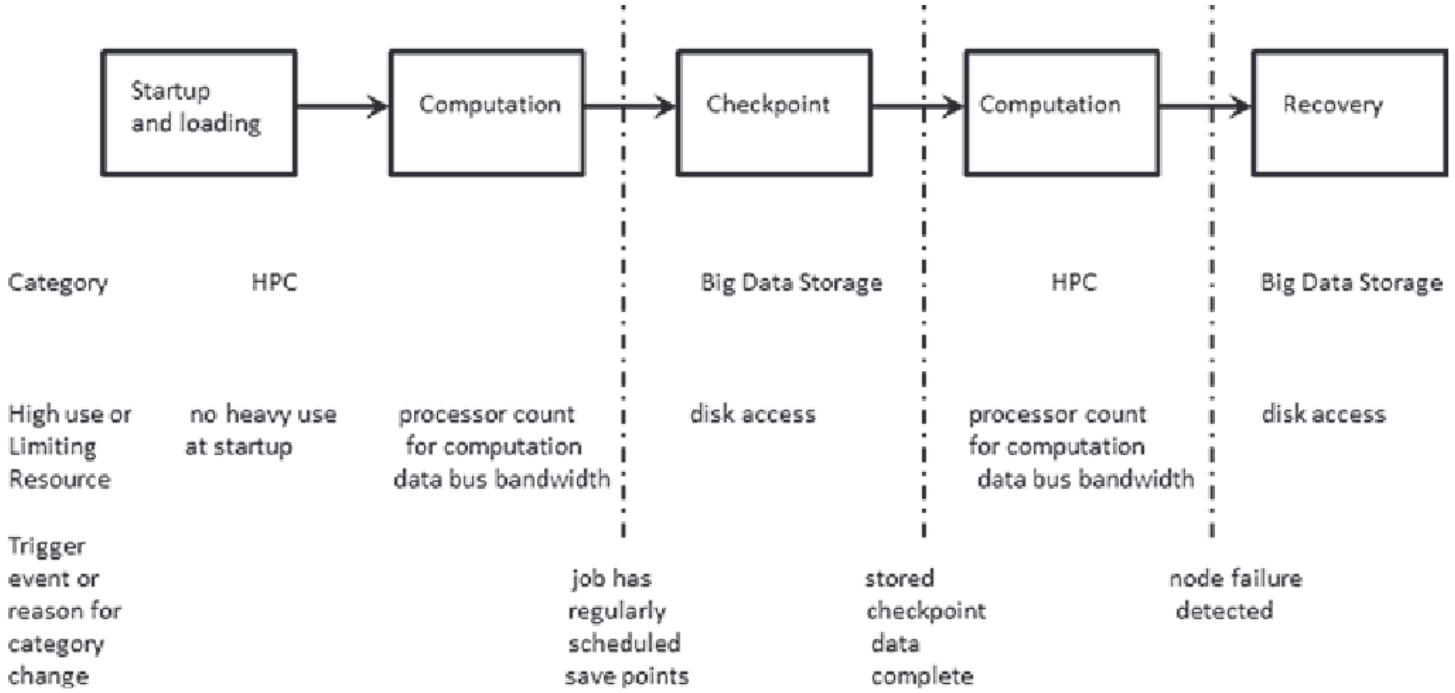
- USB
- Network Types
- Cache memory
- Software capability
- Main Memory

Temporal Variability of Workloads

- There are two different cases in which the workload category would change.
  - when a job's next step or phase is a different category than the current one.
  - when the job is incorrectly categorized.

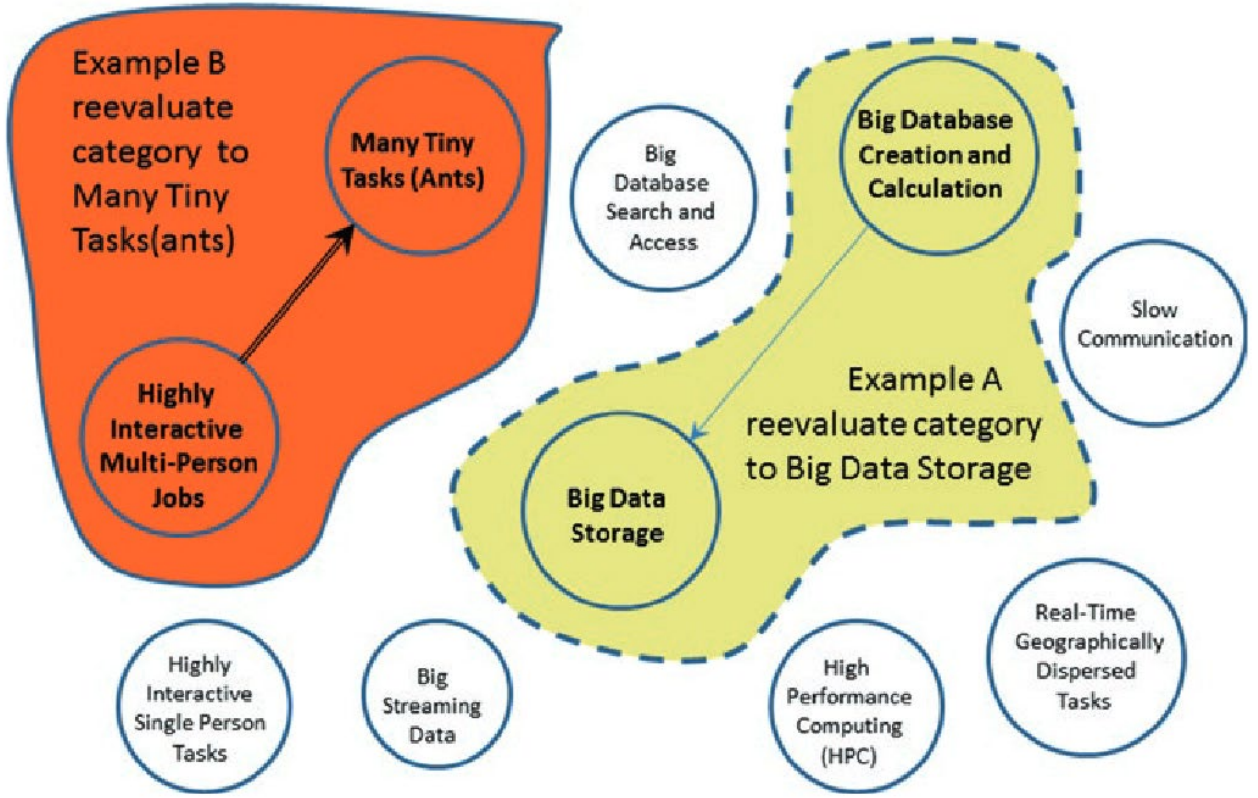
Temporal Variability of Workloads

Example of the changing Cloud Workload categories for HPC Job



Temporal Variability of Workloads

Example of miscategorized workloads



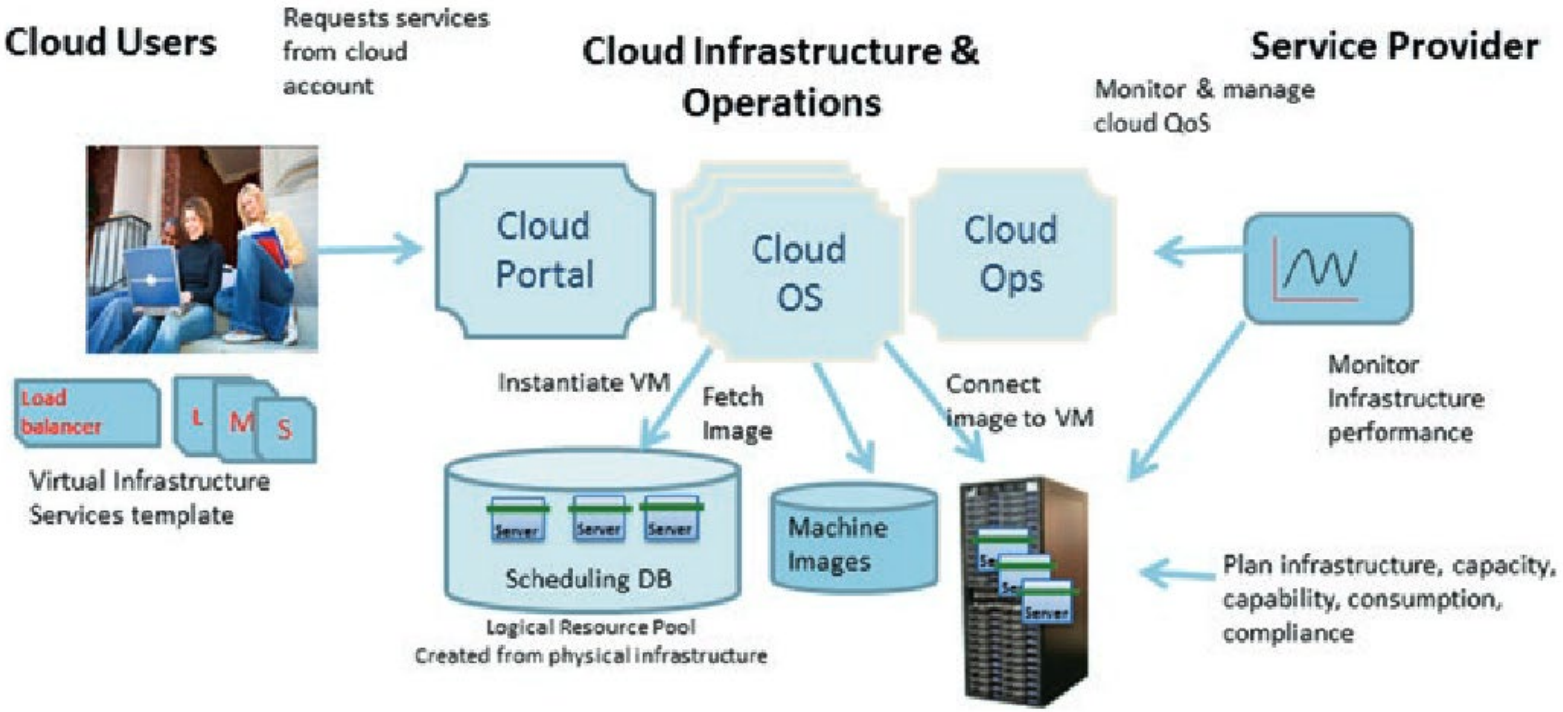
Low-Level or Hardware Metrics of Computer Utilization

Intel's VTune Amplifier XE documentation provides a fair amount of details about performance counters:

- Instruction per Cycle (IPC)
- LLC misses
- L1 data cache misses
- Cycles for which Instruction Queue is Full
- L1 Instruction Cache Misses
- Cycles during which reservation stations are full
- Number of lines fetched from memory

Benefits to Cloud Service Providers

SLA and QoS in IaaS Cloud. Cloud OS can benefit from fine-grained platform resource monitoring and controls to assure predictable performance and efficient and secure operations



Cloud Management Requirements

- Cloud management tools can be of two types: in-band (IB) or out-of-band (OOB).
  - In-band refers to an agent that typically runs in an OS or VM, collects data, and reports for monitoring purposes. However, it may interfere with other processes running in that VM, slowing it down or creating additional resource contentions.
  - OOB refers to monitoring tools that typically use a baseboard management controller with a processor and memory system to observe the main server's health metrics.

Monitoring Metrics

Metrics name	Description	Units
CPUUtilization	The percentage of allocated compute units	*Percent*
DiskReadOps	Completed read operations from all disks available to the VM	*Count*
DiskWriteOps	Completed write operations to all disks available to the VM	*Count*
DiskReadBytes	Bytes read from all disks available to the VM	*Bytes*
DiskWriteBytes	Bytes written to all disks available to the VM	*Bytes*
NetworkIn	The number of bytes received on all network interfaces by the VM	*Bytes*
NetworkOut	The number of bytes sent out on all network interfaces by the VM	*Bytes*

Monitoring Metrics

Frequency

Monitored resources	Frequency
VM instance (basic)	Every 5 minutes
VM instance (detail)	Every 1 minute
Storage volumes	Every 5 minutes
Load balancers	Every 5 minutes
DB instance	Every 1 minute
SQS queues	Every 5 minutes
Network queues	Every 5 minutes

Some Example of Monitoring Tools

Amazon Cloud Watch

Nagios

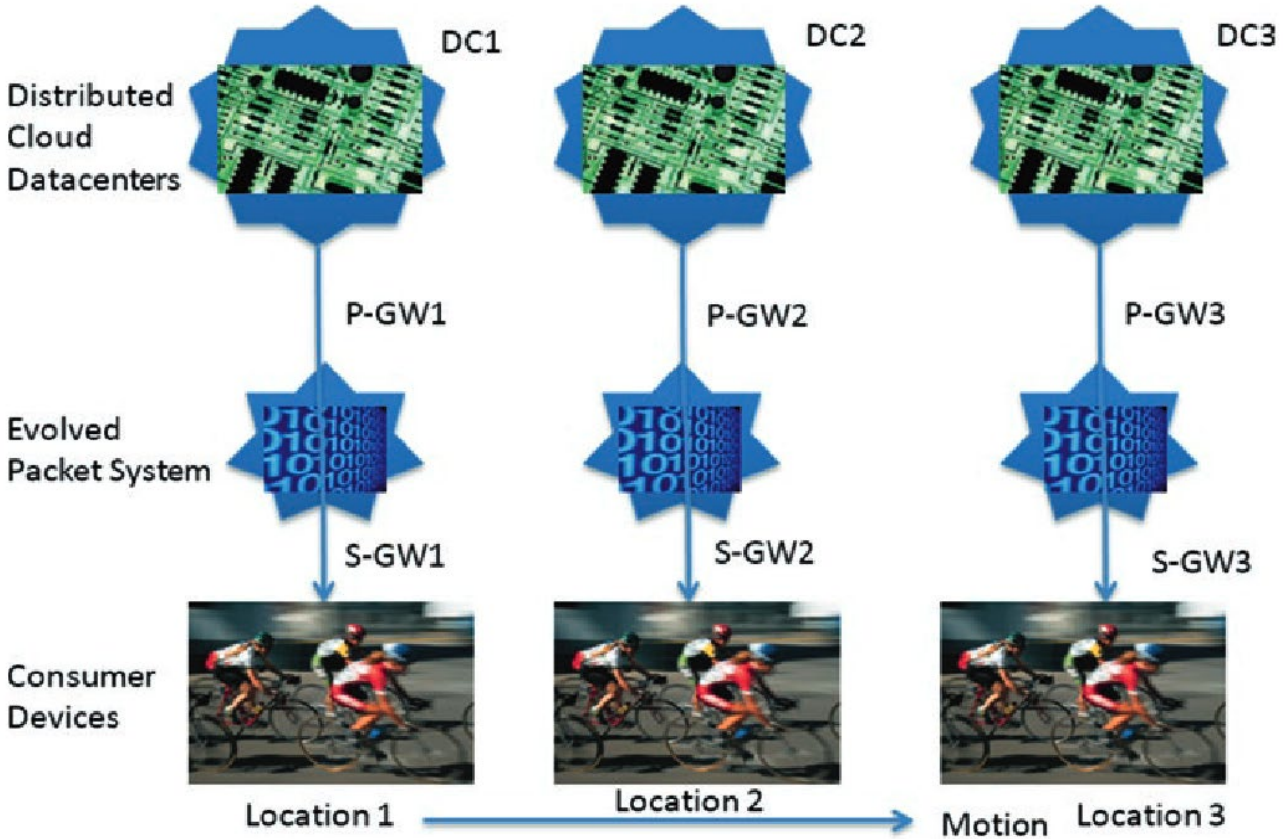
New relic

Follow-ME Cloud

- Major consumers of Public Cloud services are the numerous mobile user devices, with their needs of live video calls and always-connected social networks, with minimal latency requirements.
- However, these users are not stationary. A need to provide them with continuous IP (Internet Protocol)-based services while optimizing support from the nearest data center is nontrivial.
- The ability to smoothly migrate a mobile user from one data center to another in response to the physical movement of a user's equipment without any disruption in the service, is called Follow-ME Cloud (FMC)

Follow-ME Cloud

- Migration of a user's with changes in location




thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
спасибо
obrigado
dziękuję
bedankt
merci
grazie
arigatō
takk
dakujem
мерси
sukriya
kop khun krap
terima kasih
go raibh maith agat
mochchakkeram
tapadh leat
감사합니다
ευχαριστώ
ありがとう
hvala
paldies
рахмат
Баярлалаа
хвала
obrigado