

# Chapter 4

## Cloud Compute

## 04.01 Compute Introduction

CH04: Cloud Compute

## Cloud Compute Defined

- Computer calculations in the cloud
  - Operating systems
  - Services
  - Applications
  - Functions

### Cloud Compute Benefits

- Dynamic performance improvement
- Use only when required
- Test new hardware capabilities
- Evaluate new software
- Implement specialized processing

## Cloud Compute Challenges

- Latency of results
  - Traffic has to cross the Internet and back
- Learning new methodologies
- Understanding cloud architectures
  - Cloud structure and networking
- Understanding service provider options

## 04.02 Processing Capabilities

CH04: Cloud Compute

## Central Processing Units (CPUs)

- The core compute engine
- Cloud providers offer varying technologies
- Plan for proposed resources
  - What is required at peak
- Implement available resources
  - What is required at all times
  - Use elasticity to achieve peak requirements

## CPU Technologies

- Hyperthreading
  - Multiple threads of concurrent operation
  - Results in multiple virtual CPUs
  - For example, a 4-core hyperthreaded CPU = 8 virtual CPUs



## CPU Technologies

- VT-x
  - Virtualization technology in the CPU
  - Intel's solution
  - AMD implements AMD-V

## Overcommitment Ratios

- Utilize real resources for virtual machines well
- Scenario:
  - 2 CPUs
  - Each is quad core
  - Each is hyperthreaded
  - Total of 16 virtual CPUs
  - Run 4 virtual machines, each with 8 CPUs
  - Result is a 2:1 overcommitment ratio
    - 2 virtual processors for each of the CPUs (including hyperthreading)
- Overcommitment is the primary factor in private clouds
- Cloud service providers hide this from you and perform it themselves

## 04.03 Memory Requirements

CH04: Cloud Compute

Memory Requirement Factors

- Operating system
- Services
- Applications
- Processes

## Memory Ballooning

- A feature of virtualization platforms
  - Unused, allocated memory for one guest can be used by another
  - Allows for overcommitment of memory
- Mostly used in private clouds from a configuration perspective
  - Service providers may use it, but you won't configure it
- Bursting
  - The action of ballooning

## 04.04 Performance Considerations (Lab)

CH04: Cloud Compute

## Performance Factors

- CPU
- Memory
- Disks
- Network

## DEMO

- Optimizing CPU options in AWS

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-optimize-cpu.html>

- Storage Optimized Instances in AWS

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/storage-optimized-instances.html>

- Load testing in GCP

<https://cloud.google.com/community/tutorials/load-testing-iot-using-gcp-and-locust>



## 04.05 Cost Considerations

CH04: Cloud Compute

## Hands-On

- Azure Pricing Calculator  
<https://azure.microsoft.com/en-us/pricing/calculator>
- AWS Pricing Calculator  
<https://calculator.aws/#/>
- GCP Pricing Calculator  
<https://cloud.google.com/products/calculator/>

## 04.06 Energy Savings

CH04: Cloud Compute

### Public/Community Cloud Energy Savings

- Shared resources = energy savings
  - Only use what you need on a shared system

### Traditional Private Deployments

- Departmental servers
- Multiple data centers
- Localized servers in distributed companies
- High-powered desktops in many cases

### Private Cloud Deployment Energy Savings

- Virtualization changed everything
- Private cloud is basically automated virtualization
  - With some extra bells and whistles
- The new deployment that saves energy
  - Multiple virtual servers on a single box
  - Services accessed across the Internet in the private cloud
  - Virtual desktops with high computing power
    - Possible shared among multiple resources

## 04.07 Dedicated vs. Shared Compute

CH04: Cloud Compute

## DEMO

- AWS Dedicated Hosts and Instances

- <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/dedicated-hosts-overview.html>

- Azure Dedicated Host

- <https://azure.microsoft.com/en-us/services/virtual-machines/dedicated-host/>



## 04.08 High Availability and Disaster Recovery for Compute CH04: Cloud Compute

### HA/DR Effect for Compute

- High availability and disaster recover (HA/DR)
- Server or service must be there
  - Virtual servers demand the full virtual server be available
  - Serverless compute requires the function to be available

## HA/DR Effect for Compute

- Availability functions for compute
  - Clustering
    - Multiple instances with a primary and failover
  - Load balancing
    - Multiple instances with workload rotating between them
  - Serverless limits
    - Imposed by service provider

## DEMO

- AWS Lambda Limits

- <https://docs.aws.amazon.com/lambda/latest/dg/limits.html>

## 04.09 Monitoring (Lab)

CH04: Cloud Compute

## Monitoring Terminology

- Target object
  - Baselines
  - Anomalies
- Alerts
- Events
  - System can log
  - Event collection for analysis

## Event Correlation

- Event timestamps are used to correlate
- Ex:
  - Event A happened at 10:17:32 and Event B happened at 10:17:33
  - Event B and A are related
  - Maybe Event B was caused by Event A
- Correlation benefits
  - Determination of cause
  - Locating attack points
  - Identifying errant code

## Hands-On

- Monitoring in AWS



## 04.10 Forecasting

### CH04: Cloud Compute

## Forecasting Required Resources

- Forecasting is looking into the future to determine needs
  - Look at today to predict tomorrow
- Baseline
  - Standard normal performance today
  - Current average utilization
  - Recent peak utilization
  - How often?

### Forecasting Required Resources

- Upsize/increase or downsize/decrease resources to meet future demands
  - CPU
  - Memory
  - Storage

## 04.11 Policies

CH04: Cloud Compute

## Policies and Monitoring

- Monitoring may reveal sensitive data
  - Can monitor down to the process level
  - Identifying processes can give you insight into points of attack

## Policies and Monitoring

- Policies should be in place
- Policies in support of event collection
  - What can be monitored?
  - When should it be monitored?
  - What can be correlated?
- Policies to communicate alerts appropriately
  - How should it be reported?
  - Who should be notified?