Cloud Cost and Pricing

Leila Sharifi
l.sharifi@urmia.ac.ir

Outline

- Breaking Down the Cost of Cloud Computing
    - Main Components of Cloud Costs
    - Calculating Cloud Costs vs. Traditional Infrastructure
- 5 Cloud Cost Models
- 6 Cloud Cost Management Strategies
- Cloud Cost Optimization Tools
    - First Party Tools
    - Third Party Tools
- Optimizing Cloud Costs with Spot by NetApp

Main Components of Cloud Costs

- Compute
- Storage
- Network

Calculating Cloud Costs vs. Traditional Infrastructure

- **Capital costs**—server software, licensing, and hardware, as well as network infrastructure, storage environments, and backup systems.

- **Operational costs**—including support for server and network infrastructure, as well as storage warranty, data center facilities, existing system administration labor costs, and IT staff training and turnover.

- **Indirect business costs**—including unplanned and planned downtime.

Migration to Cloud Cost

- **Direct costs**—include hardware, software, management, maintenance, and staff, as well as any physical facility. Direct costs are usually straightforward and easy to estimate.

- **Indirect costs**—include loss of productivity which might be a result of several factors like server downtime, loss of customer trust, and reputational damage. Indirect costs are typically more complex to predict and estimate.

Pricing Models

- Pay as you go
- Prepaid/Fixed Subscriptions
- Reserved Instances
- Reserved instances: allow companies to commit to cloud resources for a long period of time, typically 1 or 3 years.
    - Cloud providers typically offer discounts of 50-75% compared to pay-as-you-go rates for reserved instances with the same capabilities.
    - Reserved instances are suitable for steady state loads and long running systems.

- AWS Savings Plan

- Similar to **reserved instances**, Savings Plans are a flexible pricing model that allows organizations to enjoy lower than on-demand pricing, in exchange for a one-year or three-year specific usage commitment. The commitment is expressed in terms of **spend per hour** on Amazon services.

AWS offers three types of Savings Plans

- Compute Savings Plans - apply to all usage of Amazon compute services usage, including EC2, AWS Lambda and Fargate.

- EC2 Savings Plans - applies only to usage of Amazon EC2 instances.

- SageMaker Savings Plans - applies only to SageMaker usage.

Savings plan offer three payment methods

- No upfront - does not require an upfront payment, bills customers according to actual usage each month. This grants the minimal savings plan discount.

- Partial upfront payment - with this option, more than half of your contract is prepaid and the rest is billed monthly, which grants an additional discount.

- Full upfront payment - the full commitment is paid upfront, which grants the deepest discount.

Spot Instances

- Spot instances are usually the lowest-cost computing option, offering discounts of up to 90% compared to pay-as-you-go rates. Spot instances are used by cloud providers to sell off spare capacity. The discount comes with a catch—spot instances can be interrupted at very short notice.

- Ordinarily, spot instances can only be used for workloads that are stateless, fault tolerant, or processes that can be stopped and restarted.

Cloud Cost Management Strategies

- **Budget control**—as a first step, organizations must establish budgets for cloud services and ensure teams are aware of them, and cannot exceed the budget for their specific project.

- **Right sizing**—ensuring compute instances, storage volumes and other services are provisioned at the level actually required by the business. It is very common for cloud resources to be provisioned and not fully utilized.

- **Autoscaling**—scale resources up and down dynamically according to application demand, ensuring you only pay for extra cloud resources during peak usage.

- **Scheduling**—many cloud services are not required 24/7, and can be scheduled to shut down when not needed. For example, services used by a US-based team can be shut down outside US business hours.

- **Detecting unused resources**—compute instances, storage volumes, load balancers, snapshots, and many other resources can easily be created and forgotten. Organizations must be able to scan their cloud deployment for unused resources and delete them to conserve costs.

- **Smartly applying discounts**—discounted pricing models like reserved instances and spot instances can significantly drive down cloud costs, but they must be used appropriately. Tools like Cloud Analyzer from Spot by NetApp can help you understand which of your applications and workloads is best suited for discounted price models.

Cloud Cost Optimization Tools

- First Party Tools:
    - Limited functionality—most first-party tools are limited in their ability to identify wasted costs and maximize savings using multiple cost models.
    - Limited to one vendor—most first-party tools only work with one cloud provider, and are not suitable for organizations that require multi-cloud cost management.
    - Conflict of interest—a cloud provider is, in the end, interested in maximizing profits.
- Third Party Tools