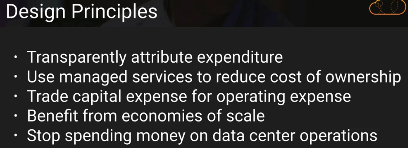
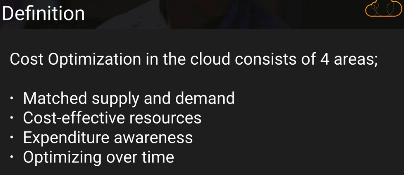
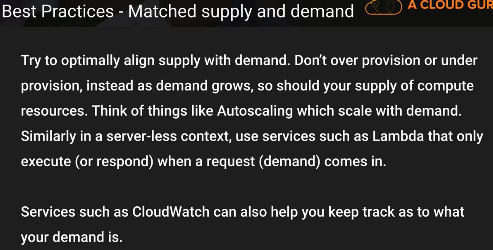


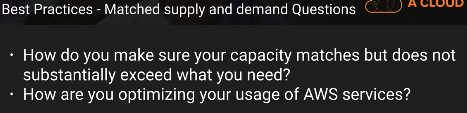
Transparently attribute expenditure

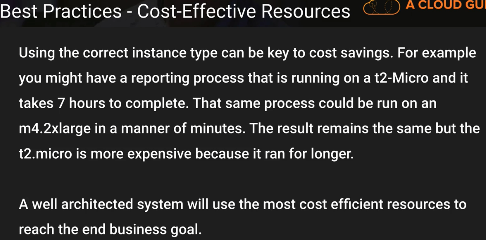


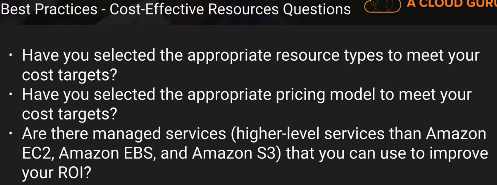


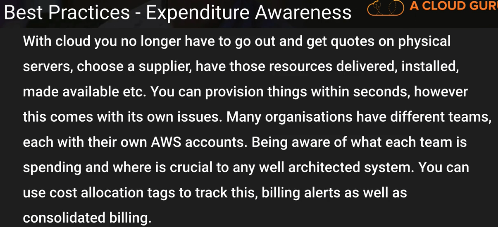
Example : Autoscaling..

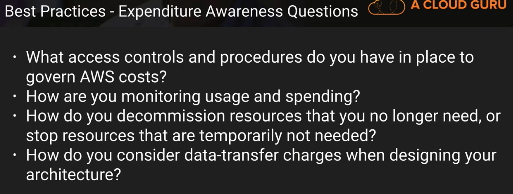


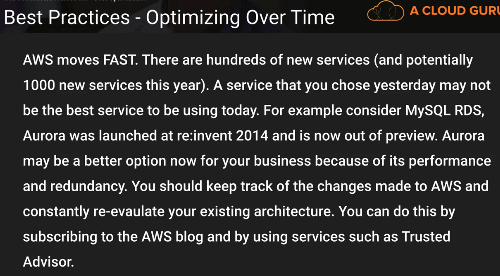






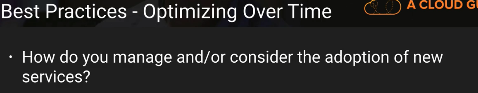


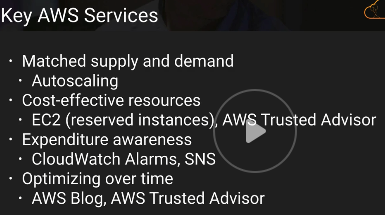




**Trust advisor**

<https://aws.amazon.com/premiumsupport/trustedadvisor/best-practices/>





Thanks & Regards,

Shantaram Vernekar

A cost-optimized system will fully utilize all resources, achieve an outcome at the lowest possible price point, and meet your functional requirements.

Design Principles

Adopt a consumption model

Pay only for the computing resources you consume, and increase or decrease usage depending on business requirements

Measure overall efficiency

Measure the business output of the system and the costs associated with delivering it

Stop spending money on data center operations – AWS takes care of it

Analyze and attribute expenditure

The cloud makes it easier to accurately identify the usage and cost of systems, which then allows transparent attribution of IT costs to individual business owners.

Use managed services to reduce cost of ownership

Definition

Cost optimization in the cloud is composed of four areas:  Cost-effective resources  Matching supply with demand  Expenditure awareness  Optimizing over time

Cost-effective resources

In AWS there are a number of different approaches:  Appropriate provisioning  Right sizing  Purchasing options: On Demand Instances, Spot Instances, and Reserved Instances  Geographic selection  Managed services

Appropriate provisioning

when you use managed services the vendor provisions and manages the underlying resources. Typically managed services have attributes that you can set to ensure sufficient capacity to meet your needs. You need to set and monitor these attributes so that your excess capacity is kept to a minimum and performance is maximized for end users.

Key Service – CloudWatch

Right sizing

Right sizing is using the lowest cost resource that still meets the technical specifications of a specific workload

Monitor resources and alarms to provide the data for right sizing. This monitoring can also provide triggers for the next right-sizing cycle

Key Services

CloudWatch and Amazon CloudWatch Logs,

Others - AWS Trusted Advisor

Purchasing options:

On Demand Instances, - you pay a flat hourly rate, and you have no long-term commitments

Spot Instances - Spot Instances are ideal for use cases such as batch processing, scientific research, image or video processing, financial analysis, and testing

Reserved Instances

two types of Amazon EC2 Reserved Instances:  Standard  Convertible

Key Service : **Cost Explorer**

Geographic selection

When you architect your solutions, a best practice is to seek to place computing resources closer to users to provide lower latency and strong data sovereignty

For global audiences you might need to use multiple locations to meet these needs. In addition, you want to select the geographic location that minimizes your costs

Key Services – Region

Ohers – Route 53 and CloudFront

Managed services

AWS managed services remove the burden of undifferentiated heavy lifting required to maintain a service

Use managed services to reduce the cost of managing infrastructure, but you should also consider the time savings that will allow your team to focus on value-adding features.

Matching Supply and Demand

However, the economic benefits of just-in-time supply needs to be balanced against the need to provision to account for resource failures, high availability, and provision time.

In AWS, you can use a number of different approaches to match supply with demand. The following sections describe how to use these approaches:  Demand-based  Buffer-based  Time-based

**Demand Based**  -

Achieved using auto scaling and elastic load balancing

Key AWS Services The key AWS service that supports a demand-based approach is Auto Scaling, which allows you to add or remove resources to match demand without overspending.

Others : CloudWatch and Elasctice load balancing

**Buffer Based**

A buffer-based approach to matching supply and demand uses a queue to accept messages (units of work) from producers

A buffer is a mechanism to ensure that applications can communicate with each other when they are running at different rates over time

The key AWS services that support a buffer-based approach are Amazon SQS and Amazon Kinesis, which make it simple and cost-effective to decouple the components of a cloud application

Others :  Ec2 spot instances and Lambda

**Time Based**

A time-based approach ensures that resources are available at the specific time they are required, and can be provided without any delays due to start-up procedures and system or consistency checks.

Key service : Autoscaling

Others : CloudFormation

Expenditure awareness

Accurate cost attribution allows you to understand how profitable business units and products are, and allows you to make more informed decisions about where to allocate resources within your business

Consider taking a multi-faceted approach to becoming aware of your expenditures. Your team needs to gather data, analyze, and then report. These are the key factors to consider:

 Stakeholders  - Financial, Business Unit owners, Tech Leads, Third Parties

 Visibility and controls - Before taking action, a detailed level of visibility into your AWS environment will enable you to identify opportunities for saving.

 Cost attribution - You can use cost attribution to drive cost management by assigning costs to parts of your organization, such as departments, business units, products, or internal teams.

 Tagging - Tags allow you to overlay business and organizational information onto your billing and usage data. This helps you categorize and track your costs by meaningful, relevant business information. You can apply tags that represent business categories (such as cost centers, application names, projects, or owners) to organize your costs across multiple services and teams

 Entity lifecycle tracking – AWS config, Clould Trail

Key AWS Services The key AWS service for expenditure awareness is AWS Billing and Cost Management.

Cost Explorer: Use this tool to visualize and analyze your costs.

Tagging: Use tags to overlay business and organizational information onto your billing and usage data.

Amazon CloudWatch alerts: Create billing alerts that notify you when usage of your services exceeds financial thresholds you define.

Optimizing over time

In AWS, you can use a number of different approaches to optimize over time. The following sections describe how to use these approaches:  Measure, monitor, and improve  Staying ever green (move to the newest services, features, and instance types)

**Measure, Monitor, and Improve**

Minimize AWS resources/system utilization gap

There are four key ways to set this up:

 Establish a cost optimization function – Team to -> responsibility and accountability of cost optimization for the entire organization

Establish goals and metrics - Establish goals and metrics that your organization can use to measure its progress.

 Gather insight and perform analysis

 Report and validate

**Staying Ever Green**

As AWS releases new services and features, it is a best practice to review your existing architectural decisions to ensure that they remain cost effective and stay ever green.

As your requirements change, be aggressive in decommissioning resources, components, and workloads that you no longer require.

Trusted Advisor, AWS blog,Whats new