Migrations

Introduction

A CLOUD GURU

Exam Blueprint

A CLOUD GURU

Domain 1: Design for Organizational Complexity

1.1. Determine cross-account authentication and access strategy for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

1.2. Determine how to design networks for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

1.3. Determine how to design a multi-account AWS environment for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

Domain 2: Design for New Solutions

2.1. Determine security requirements and controls when designing and implementing a solution.

2.2. Determine a solution design and implementation strategy to meet reliability requirements.

2.3. Determine a solution design to ensure business continuity.

2.4. Determine a solution design to meet performance objectives.

2.5. Determine a deployment strategy to meet business requirements when designing and implementing a solution.

Exam Blueprint

A CLOUD GURU

Domain 3: Migration Planning

3.1. Select existing workloads and processes for potential migration to the cloud.

3.2. Select migration tools and/or services for new and migrated solutions based on detailed AWS knowledge.

3.3. Determine a new cloud architecture for an existing solution.

3.4. Determine a strategy for migrating existing on-premises workloads to the cloud.

Domain 4: Cost Control

4.1. Select a cost-effective pricing model for a solution.

4.2. Determine which controls to design and implement that will ensure cost optimization.

4.3. Identify opportunities to reduce cost in an existing solution.

Domain 5: Continuous Improvement for Existing Solutions

5.1. Troubleshoot solution architectures.

5.2. Determine a strategy to improve an existing solution for operational excellence.

5.3. Determine a strategy to improve the reliability of an existing solution.

5.4. Determine a strategy to improve the performance of an existing solution.

5.5. Determine a strategy to improve the security of an existing solution.

5.6. Determine how to improve the deployment of an existing solution.

Cloud Migration and Hybrid Architecture

Migration Strategies

Migration Strategies

A CLOUD GURU

Migration Strategy	Description	Example
Re-Host	“Lift and Shift”; simply move assets with no change.	Move on-prem MySQL database to EC2 Instance.
Re-Platform	“Lift and Reshape”; Move assets but change underlying platform.	Migrate on-prem MySQL database to RDS MySQL.
Re-Purchase	“Drop and Shop”; Abandon existing and purchase new.	Migrate legacy on-prem CRM system to salesforce.com
Rearchitect	Redesign application in a cloud-native manner.	Create a server-less version of legacy application.
Retire	Get rid of applications which are not needed.	End-of-life the label printing app because no-one uses it anymore.
Retain	“Do nothing option”; Decide to reevaluate at a future date.	“Good news Servers, you’ll live to see another day.”

Migration Strategies

A CLOUD GURU

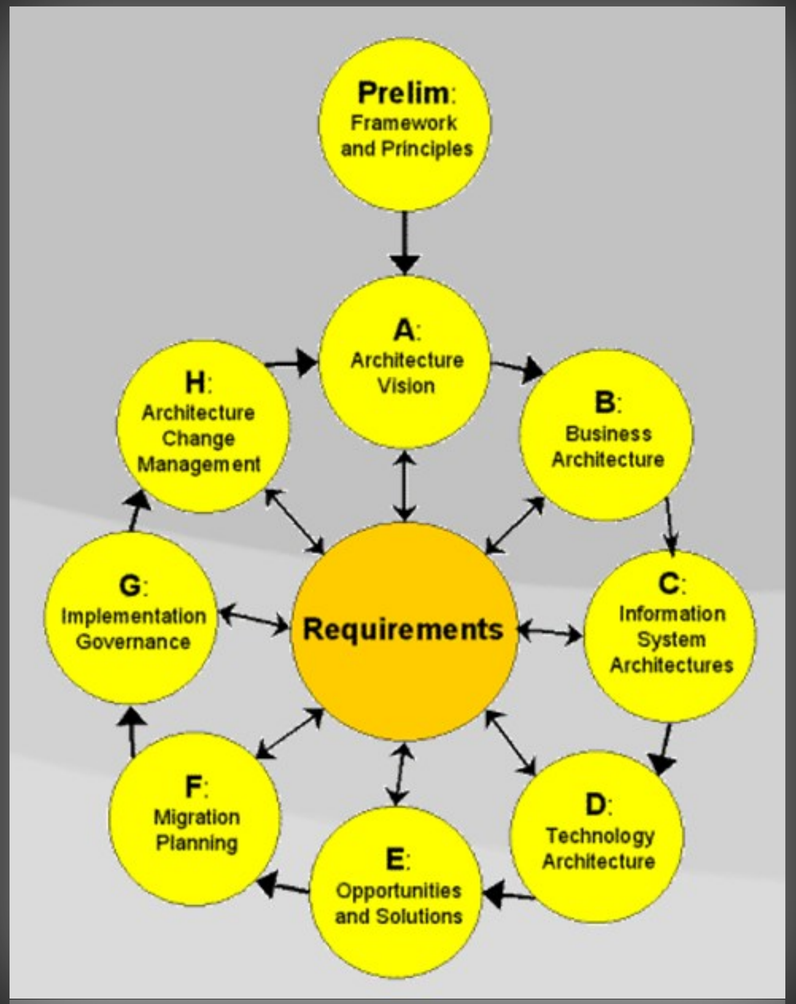
Migration Strategy	Effort (Time and Cost)	Opportunity to Optimize
Re-Host	★★	★
Re-Platform	★★★★	★★★
Re-Purchase	★★★	★
Rearchitect	★★★★★	★★★★★
Retire		
Retain	★	

Cloud Migration and Hybrid Architecture

Cloud Adoption Framework

TOGAF

- The Open Group Architectural Framework
- Approach for designing, planning, implementing and governing enterprise IT architectures
- Started development in 1995
- De facto standard in EA practice
- Favored enterprise architecture framework for most Fortune 500 companies



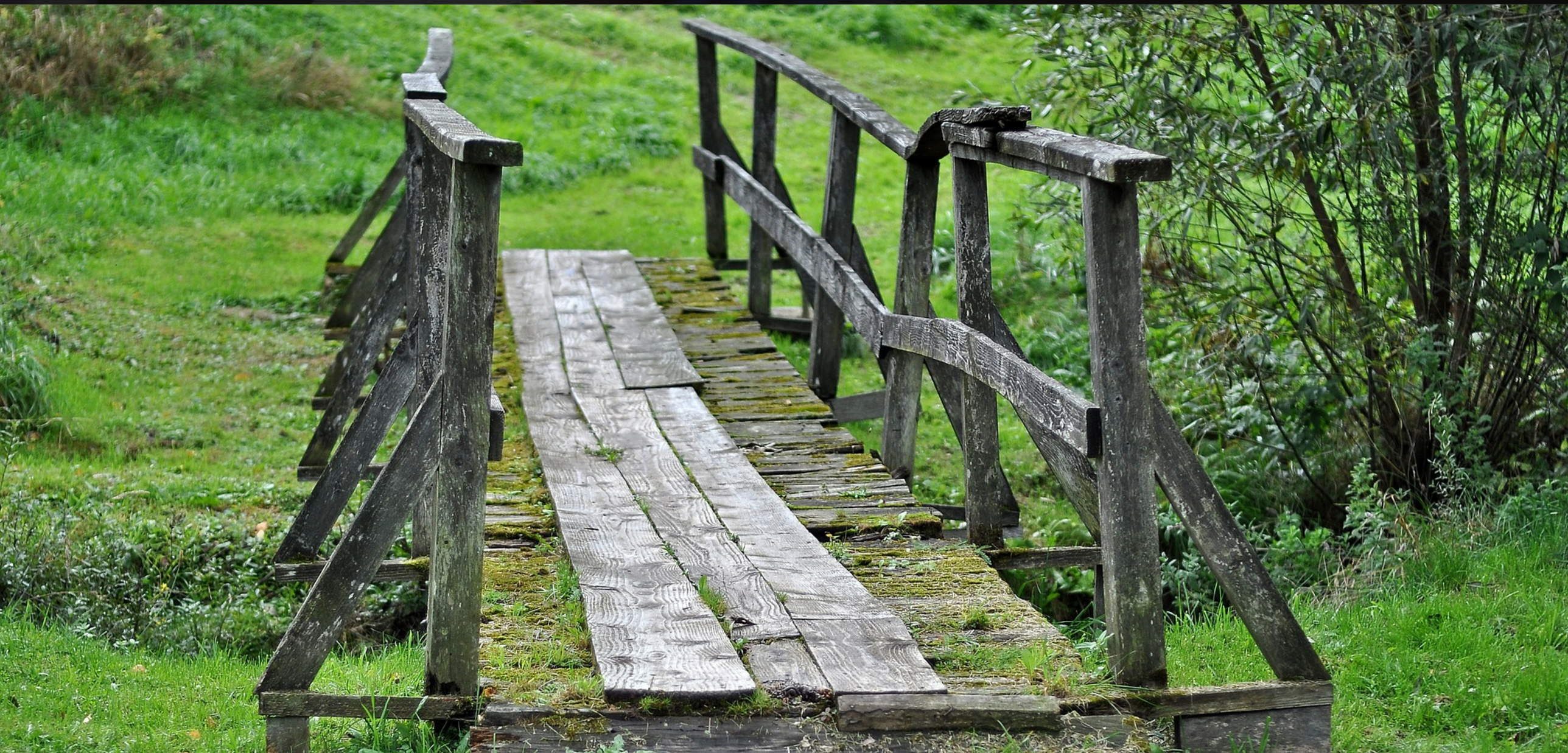
TOGAF

A CLOUD GURU

- Enterprise Architecture is not TOGAF…
- But TOGAF often fills a vacuum.
- Often misunderstood and victim of unreasonable expectations
- Not all practitioners are great practitioners.
- TOGAF is not a cookbook.
- People take it too literally.
- “Architects gonna architect”

Boundary Spanners

A CLOUD GURU



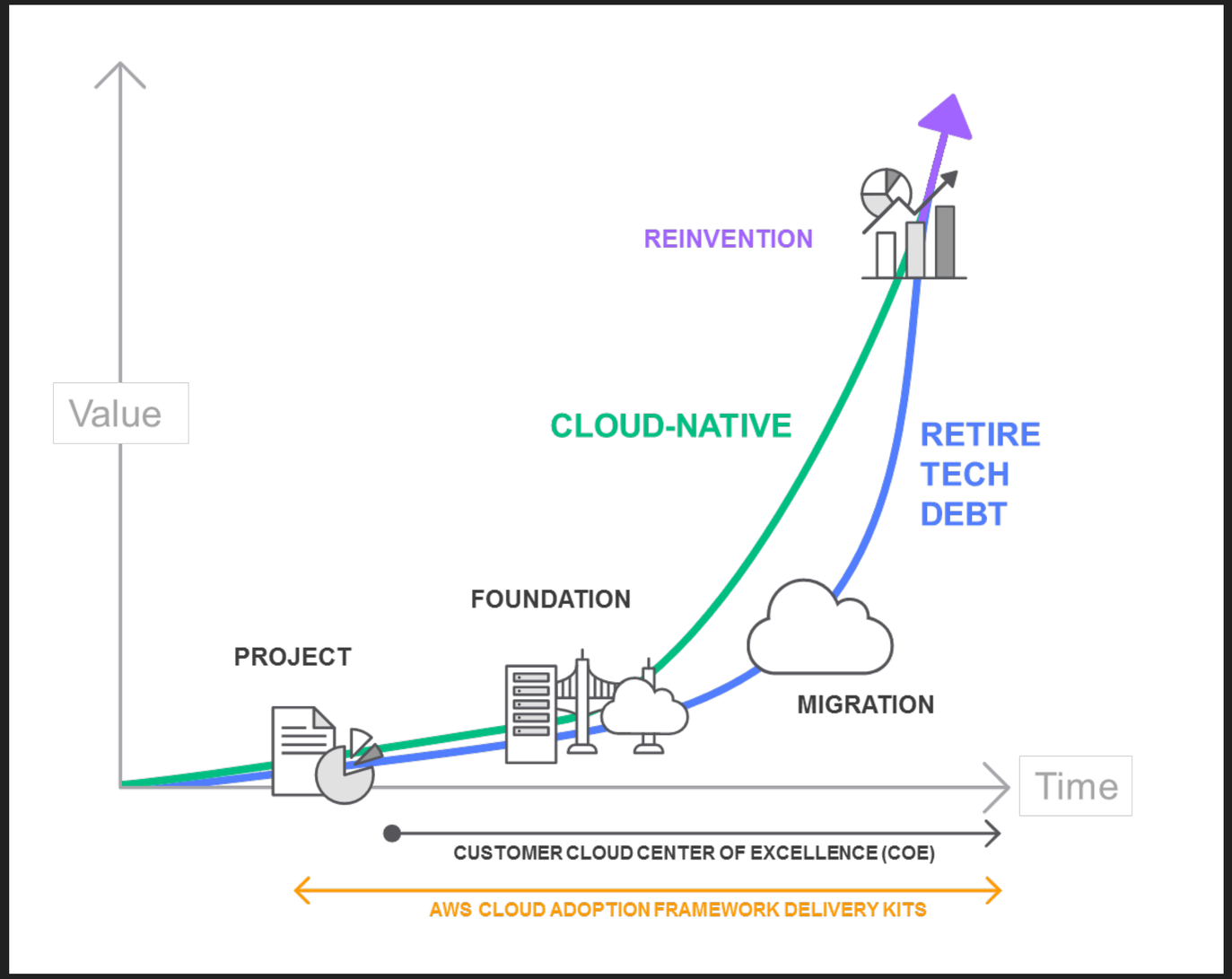
What is a Framework?

A framework…

- IS some information to help you get your mind around a problem.
- IS open for localization and interpretation.
- IS something you should adapt to your organizational culture.
- IS NOT a perfect step-by-step recipe to success.
- IS NOT something to hide behind with big words.

Typical Cloud Adoption Phases

A CLOUD GURU



PROJECT
In the project phase, you are running projects to get familiar and experience benefits from the cloud.

FOUNDATION
After experiencing the benefits of cloud, you then build the foundation to scale your cloud adoption. This includes creating a landing zone (a pre-configured, secure, multi-account AWS environment), Cloud Center of Excellence (CCoE), operations model, as well as assuring security and compliance readiness.

MIGRATION
In this stage, you migrate existing applications including mission-critical applications or entire data centers to the cloud as you scale your adoption across a growing portion of your IT portfolio.

REINVENTION
Now that your operations are in the cloud, you can focus on reinvention by taking advantage of the flexibility and capabilities of AWS to transform your business by speeding time to market and increasing the attention on innovation.

Cloud Adoption Framework

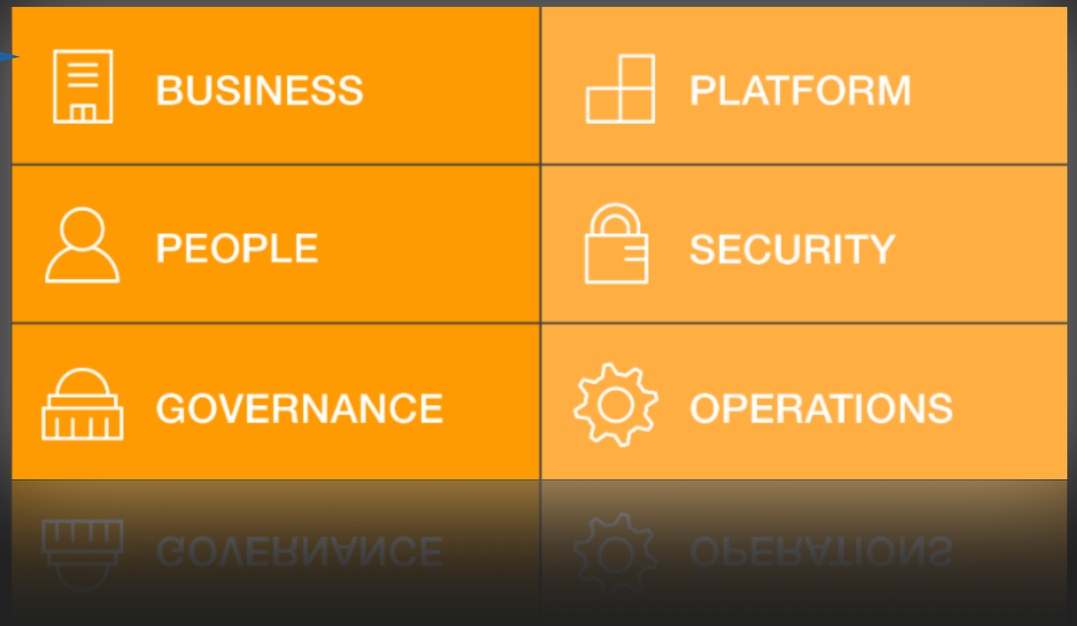
A CLOUD GURU

- There's more to cloud adoption than technology.
- To fully unlock the potential benefits of a cloud migration, a holistic approach must be considered.

Cloud Adoption Framework

A CLOUD GURU

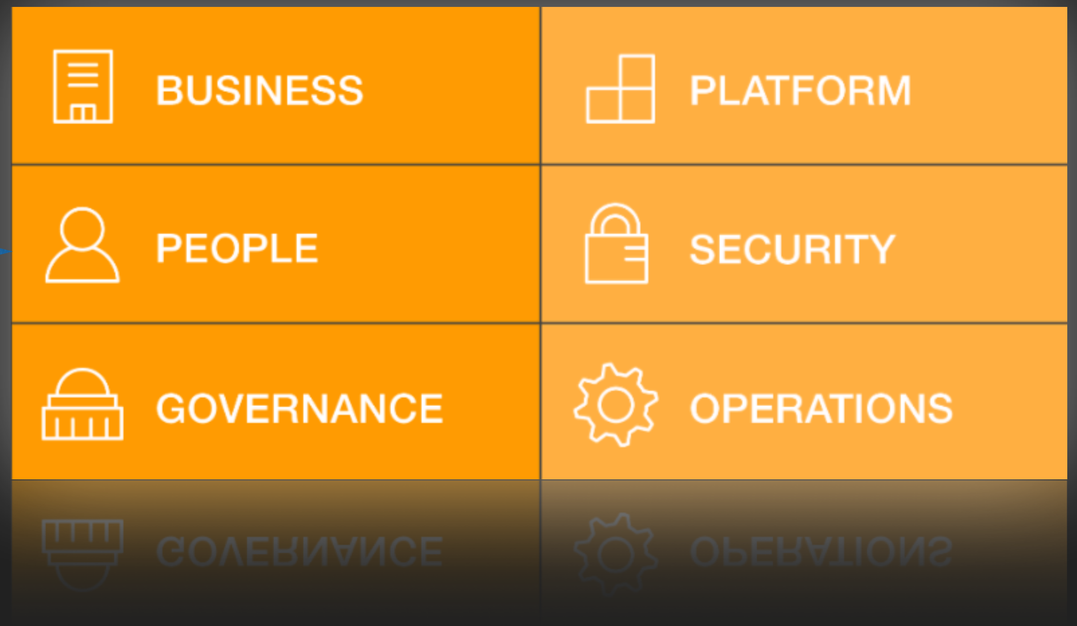
- Creation of a strong business case for cloud adoption
- Business goals are congruent with cloud objectives
- Ability to measure benefits (TCO, ROI)



Cloud Adoption Framework

A CLOUD GURU

- Evaluate organizational roles and structures, new skills and process needs and identify gaps
- Incentives and Career Management aligned with evolving roles
- Training options appropriate for learning styles



Cloud Adoption Framework

A CLOUD GURU

- Portfolio Management geared for determining cloud eligibility and priority
- Program and Project management more agile projects
- Align KPI's with newly enabled business capabilities

BUSINESS	PLATFORM
PEOPLE	SECURITY
GOVERNANCE	OPERATIONS

Cloud Adoption Framework

A CLOUD GURU

BUSINESS	PLATFORM
PEOPLE	SECURITY
GOVERNANCE	OPERATIONS

- Resource provisioning can happen with standardization.
- Architecture patterns adjusted to leverage cloud-native
- New application development skills and processes enable more agility.

Cloud Adoption Framework

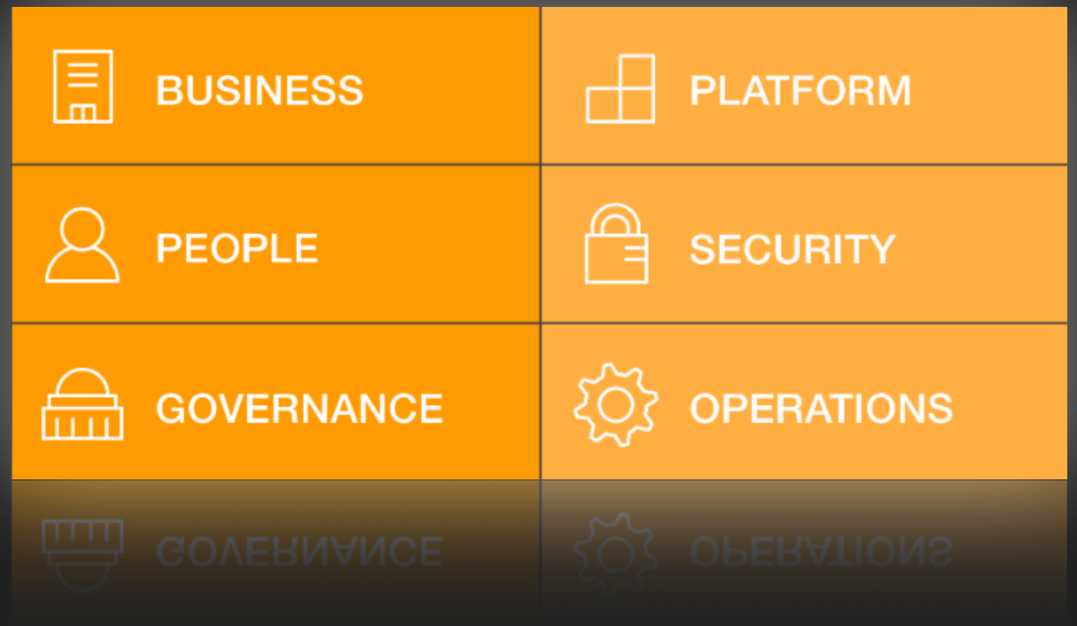
A CLOUD GURU

BUSINESS	PLATFORM
PEOPLE	SECURITY
GOVERNANCE	OPERATIONS

- Identity and Access Management modes change
- Logging and audit capabilities will evolve.
- Shared Responsibility Model removes some and adds some facets.

Cloud Adoption Framework

A CLOUD GURU



- Service monitoring has potential to be highly automated.
- Performance management can scale as needed.
- Business continuity and disaster recovery takes on new methods in the cloud.

Cloud Migration and Hybrid Architecture

Hybrid Architectures

Hybrid Architectures

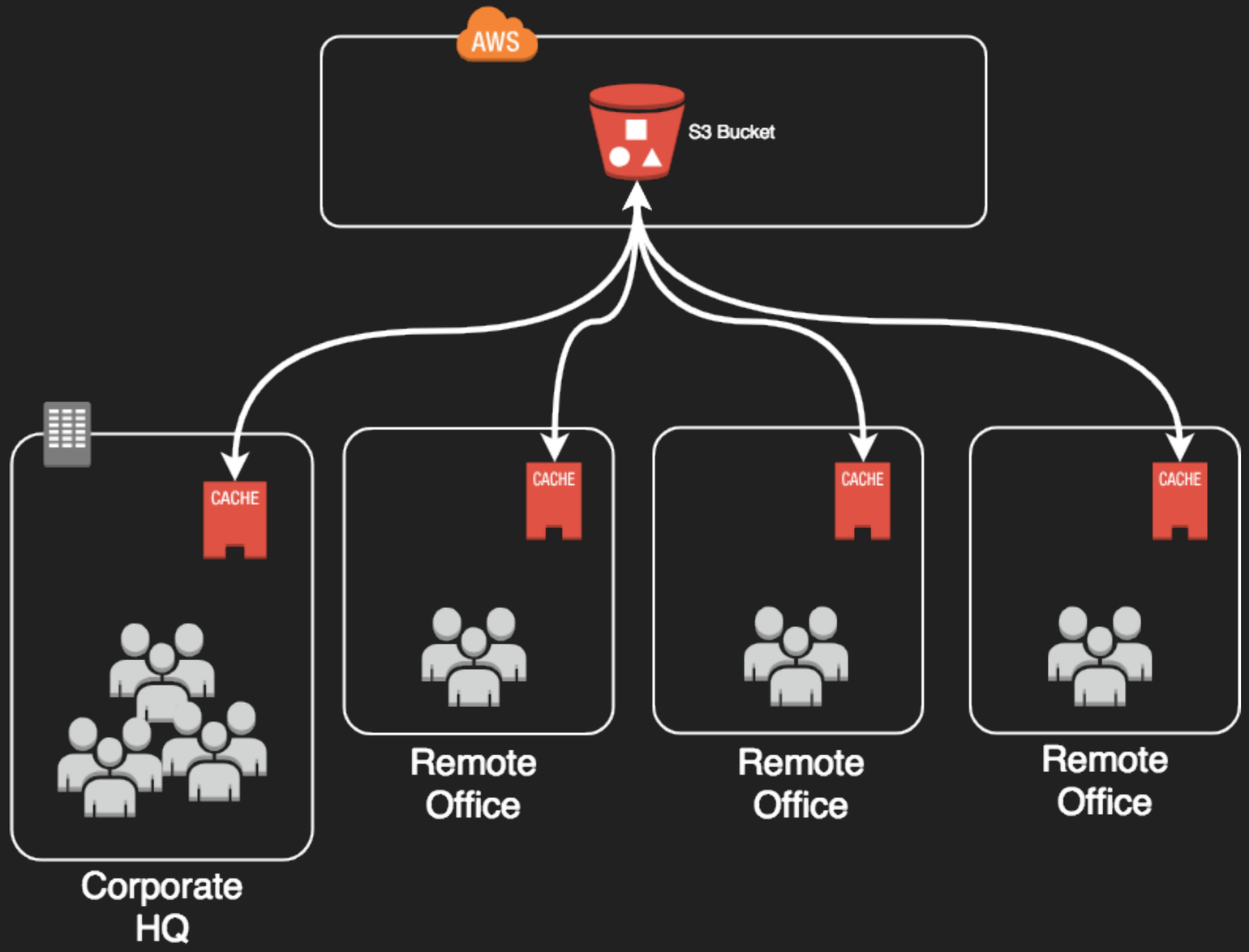
A CLOUD GURU

- Hybrid Architectures make use of cloud resources along with on-premises resources.
- Very common first step as a pilot for cloud migrations.
- Infrastructure can augment or simply be extensions of on-prem platforms — VMWare, for example.
- Ideally, integrations are loosely coupled — meaning each end can exist without extensive knowledge of the other side.

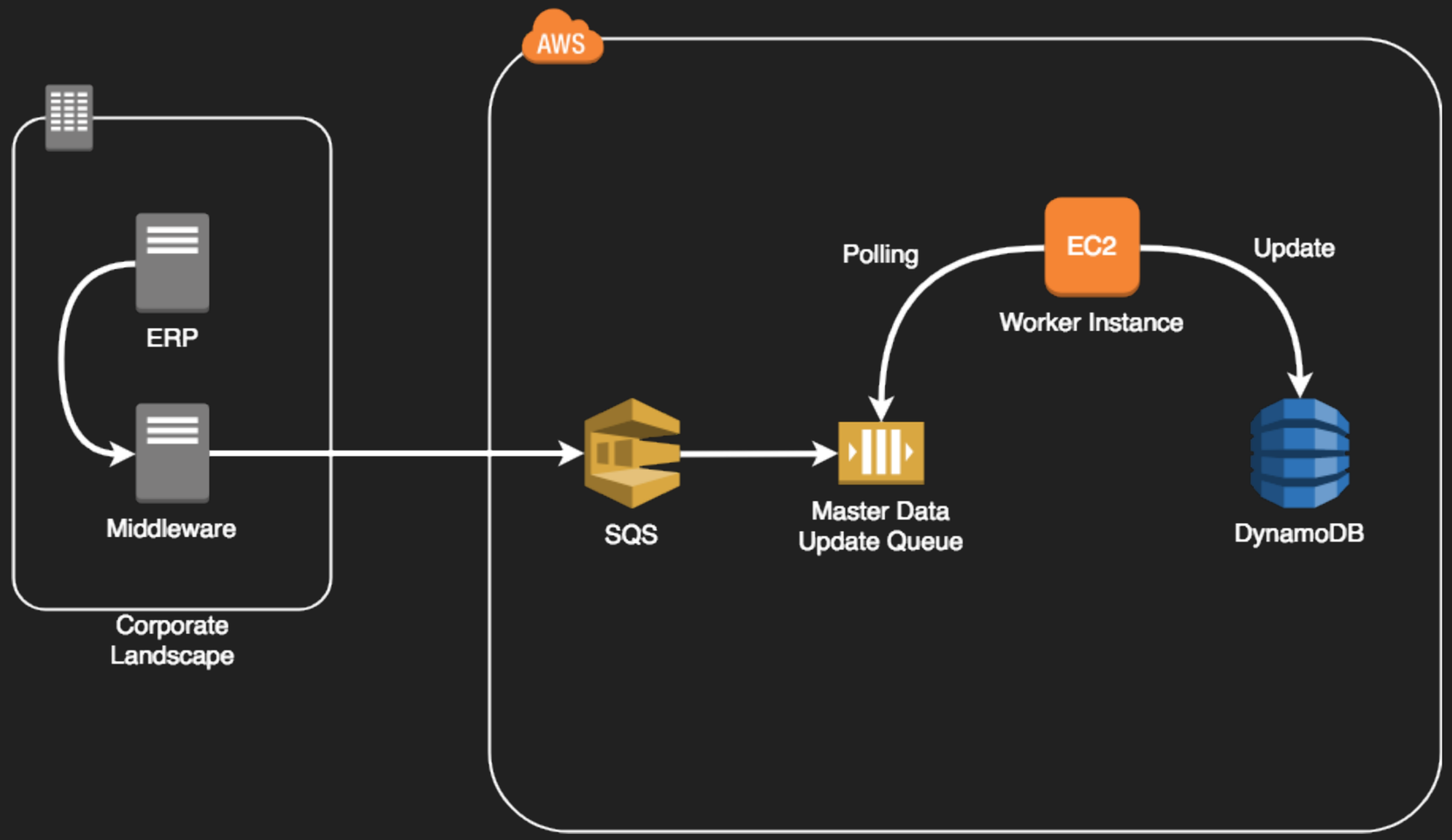
Hybrid Architectures

A CLOUD GURU

- Storage Gateway creates a bridge between on-prem and AWS.
- Seamless to end-users
- Common first step into the cloud due to low risk and appealing economics



Hybrid Architectures

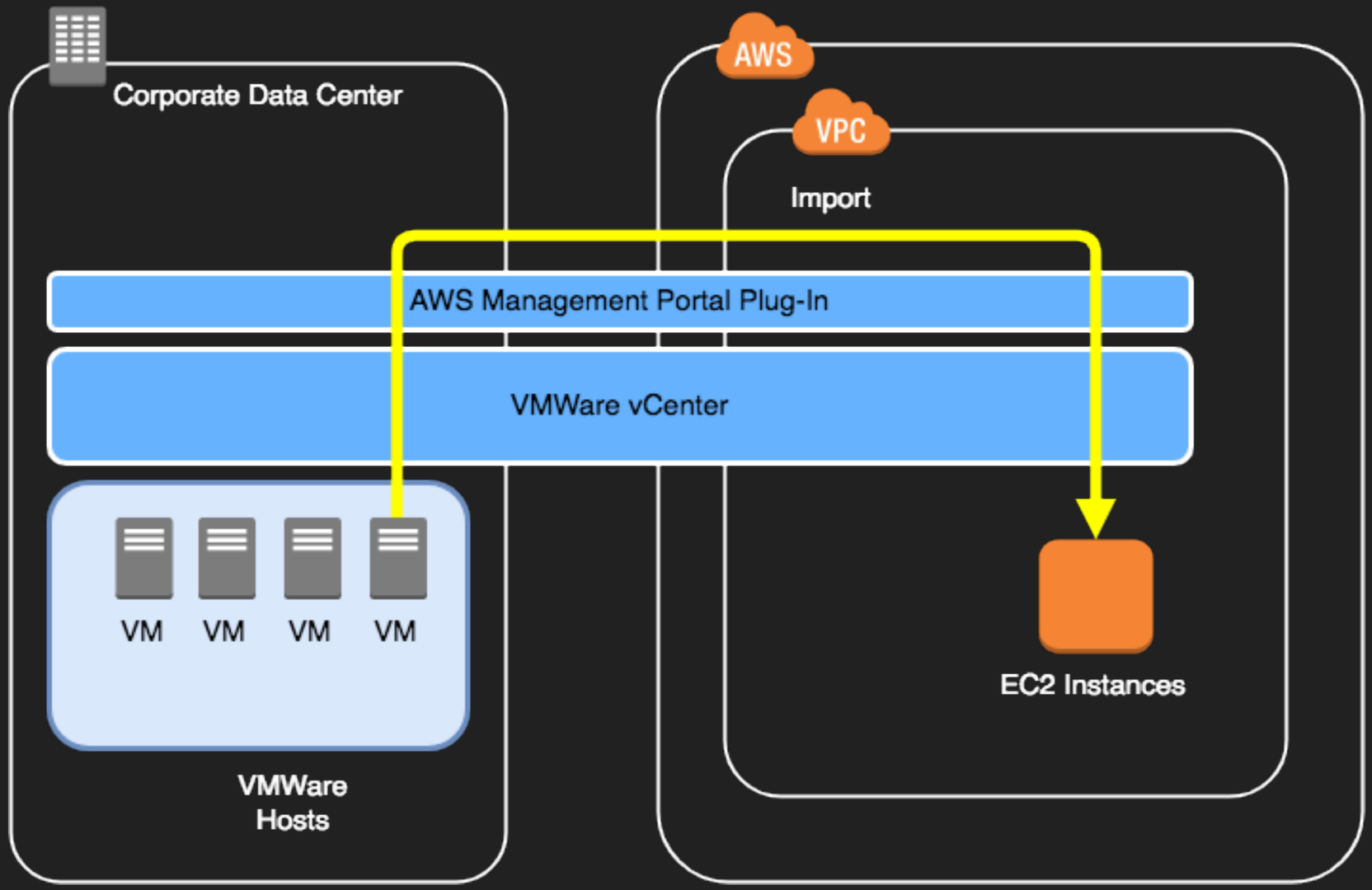


- Middleware often a great way to leverage cloud services.
- Loosely-coupled, canonical-based.

Hybrid Architectures

A CLOUD GURU

- VMware vCenter Plug-in allows transparent migration of VMs to and from AWS.
- VMware Cloud furthers this concept with more public cloud-native features.



Cloud Migration and Hybrid Architecture

Migration Tools

Migration Tools

A CLOUD GURU

- Storage Migration
- Server Migration Service
- Database Migration Service
- Application Discovery Service
- AWS Migration Hub

Storage Migration

A CLOUD GURU

AWS Storage Gateway

AWS Snowball

Server Migration Service

A CLOUD GURU

- Automates migration of on-premises VMware vSphere or Microsoft Hyper-V/SCVMM virtual machines to AWS.
- Replicates VMs to AWS, syncing volumes and creating periodic AMIs.
- Minimizes cutover downtime by syncing VMs incrementally.
- Supports Windows and Linux VMs only (just like AWS.)
- The Server Migration Connector is downloaded as a virtual appliance into your on-prem vSphere or Hyper-V setup.

AWS
Server Migration
Service

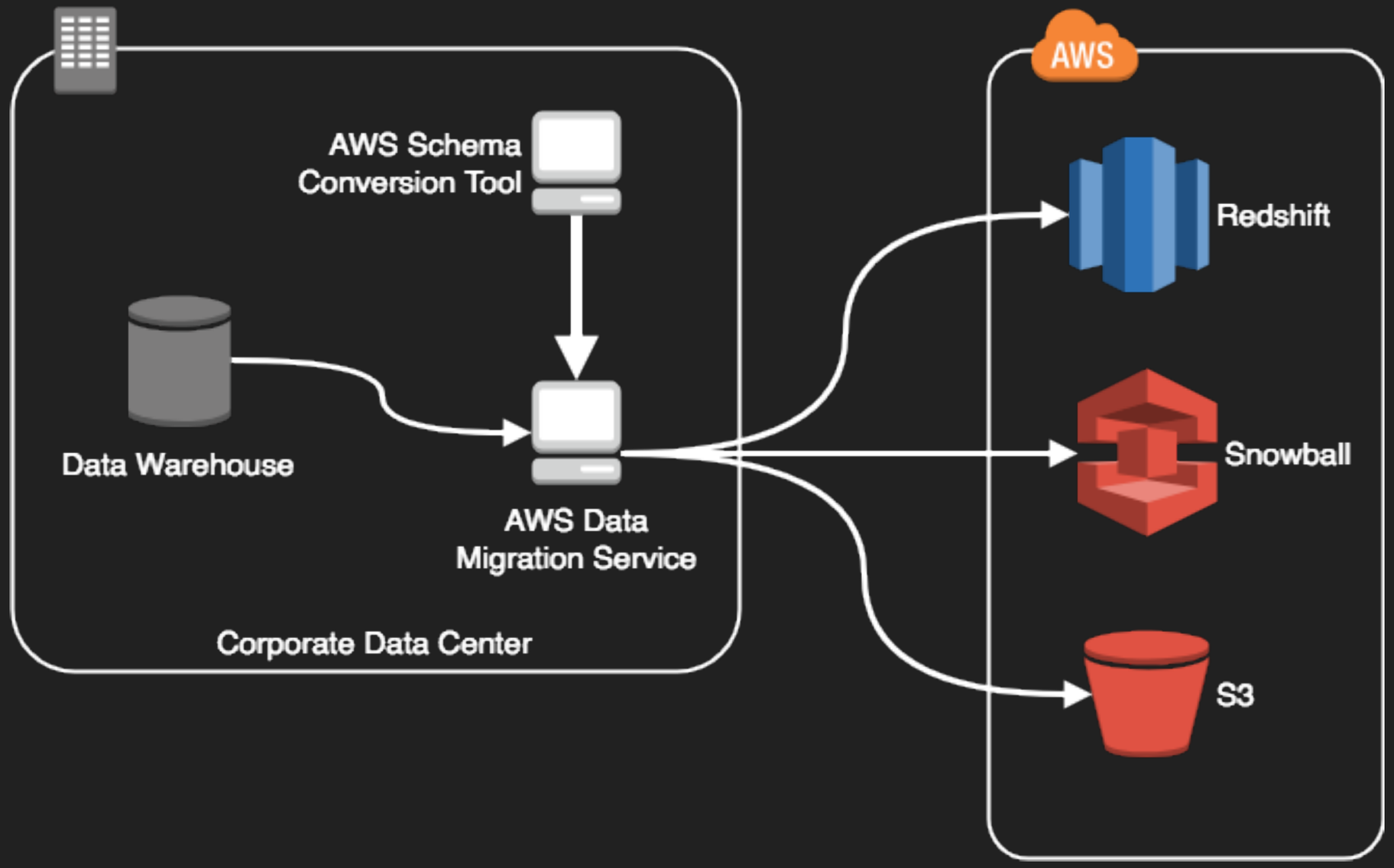
Database Migration Service

A CLOUD GURU

- Data Migration Service (DMS) along with the Schema Conversion Tool (SCT) helps customers migrate databases to AWS RDS or EC2-based databases.
- Schema Conversion Tool can copy database schemas for homogenous migrations (same database) and convert schemas for heterogeneous migrations (different database.)
- DMS is used for smaller, simpler conversions and also supports MongoDB and DynamoDB.
- SCT used for larger, more complex datasets like data warehouses.
- DMS has replication function for on-prem to AWS or to Snowball or S3.

AWS
Database Migration
Service

Database Migration Service



Application Discovery Service

A CLOUD GURU

- Gathers information about on-premises data centers to help in cloud migration planning.
- Often customers don't know the full inventory or status of all their data center assets, so this tool helps with that inventory.
- Collects config, usage and behavior data from your servers to help in estimating TCO of running on AWS.
- Can run as agent-less (VMware environment) or agent-based (non-VMware environment).
- Only supports those OSes that AWS supports (Linux and Windows.)

AWS
Application Discovery
Service

AWS Migration Hub

A CLOUD GURU

aws Services Resource Groups scott @ pletcher-acg-master Oregon Support

Migration Hub

- Dashboard
- Discover
  - Servers
  - Applications
  - Data Collectors
  - Tools
- Migrate
  - Applications
  - Updates
  - Tools
- Help & Support
- Feedback

AWS Migration Hub

Migration Hub simplifies and accelerates discovery and migration from your data centers to the AWS Cloud. Learn more

Discover
- Deploy AWS discovery tools (Optional)

Migrate
- Connect migration tools
- Migrate using connected tools
- Group servers as applications

Track
- Track status of migrations

View example

What would you like to do?

Get started with discovery OR Get started migrating

Feedback English (US) © 2008 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Cloud Migration and Hybrid Architecture

CIDR Reservations

A CLOUD GURU

- Ensure your IP addresses will not overlap between VPC and on-prem.
- VPCs support IPv4 netmasks range from /16 to /28.

  /16 = 255.255.0.0 = 65,536 addresses

  /28 = 255.255.255.240 = 16 addresses

- Remember: 5 IPs are reserved in every VPC subnet that will take up addresses.

  /28 = 255.255.255.240

  = 16 addresses

  = 14 “usable” minus 3 reserved addresses in a VPC = 11 available in VPC

Network Migrations

A CLOUD GURU

- Most organizations start with a VPN connection to AWS.
- As usage grows, they might choose Direct Connect but keep the VPN as a backup.
- Transition from VPN to Direct Connect can be relatively seamless using BGP.
- Once Direct Connect is set-up, configure both VPN and Direct Connect within the same BGP prefix.
- From the AWS side, the Direct Connect path is always preferred…
- …but you need to be sure the Direct Connect path is the preferred route from your network to AWS and not VPN (through BGP weighting or static routes.)

Data Storage

Amazon Snow Family

Great Barrier Island

A CLOUD GURU

Copyright 2018 Google Maps

Wreck of the SS Wairarapa

A CLOUD GURU

Pigeon Post of Great Barrier Island

Amazon Snow Family

A CLOUD GURU

- Evolution of AWS Import/Export process
- Move massive amounts of data to and from AWS
- Data transfer as fast or as slow as you’re willing to pay a common carrier (UPS, DHL, FedEx, etc.)
- Encrypted at rest
- And also in transit!

AWS Snow Family

Amazon Snow Family

A CLOUD GURU

AWS Import/Export	Ship an external hard drive to AWS. Someone at AWS plugs it in and copies your data to S3.
AWS Snowball	Ruggedized NAS in a box AWS ships to you. You copy over up to 80TB of your data and ship it back to AWS. They copy the data over to S3.
AWS Snowball Edge	Same as Snowball, but with onboard Lambda and clustering.
AWS Snowmobile	A literal shipping container full of storage (up to 100PB) and a truck to transport it.

Amazon Snow Family

A CLOUD GURU

Amazon Snow Family

Amazon Snow Family

A CLOUD GURU

Cloud Migration and Hybrid Architecture

Wrap-Up

Exam Blueprint

A CLOUD GURU

8	**Domain 8.0: Cloud Migration and Hybrid Architecture**
8.1	Plan and execute for applications migrations
8.2	Demonstrate ability to design hybrid cloud architectures

For Further Study

A CLOUD GURU

AWS Whitepapers

- AWS Migration Whitepaper
- An Overview of the AWS Cloud Adoption Framework (version 2)
- Migrating Applications to AWS: Guide and Best Practices
- AWS Cloud Transformation Maturity Model

2017 re:Invent Videos

- How To Assess Your Organization's Readiness to Migrate at Scale (ENT211)
- Migrating Databases and Data Warehouses to the Cloud (DAT317)
- Deep Dive: Using Hybrid Storage with AWS Storage Gateway to Solve On-Premises Storage Problems (STG309)

Migrations

Exam Tips

Exam Tips

A CLOUD GURU

Migration Strategies

- Understand the different strategies that companies might undertake when deciding if the cloud is right for them.
- Understand the typical trade-offs and relative benefits for each strategy.

Cloud Adoption Framework

- Know what a “framework” is and the realistic expectations that should accompany it.
- Understand the high-level components of the Cloud Adoption Framework.
- Most importantly, know that cloud adoption is only partially a technology effort.

Exam Tips

A CLOUD GURU

Hybrid Architectures

- Be able to speak to some typical hybrid architectures that leverage both on-prem and cloud assets.
- Know that VMware has some nifty tools for abstracting workloads across on-prem and cloud, such as the Import plug-in.

Migrations

- Understand the different services and tools available for migrating servers, storage and databases.
- Tool usage specifics won’t likely be on the exam, except maybe Storage Gateway.

Exam Tips

A CLOUD GURU

Network Migration and Cutover:

- Know various hybrid networking architectures (See Networking chapter).
- Understand that smooth transitions from and to VPN and Direct Connect can be done using BGP routing; abrupt route changes risk downtime.

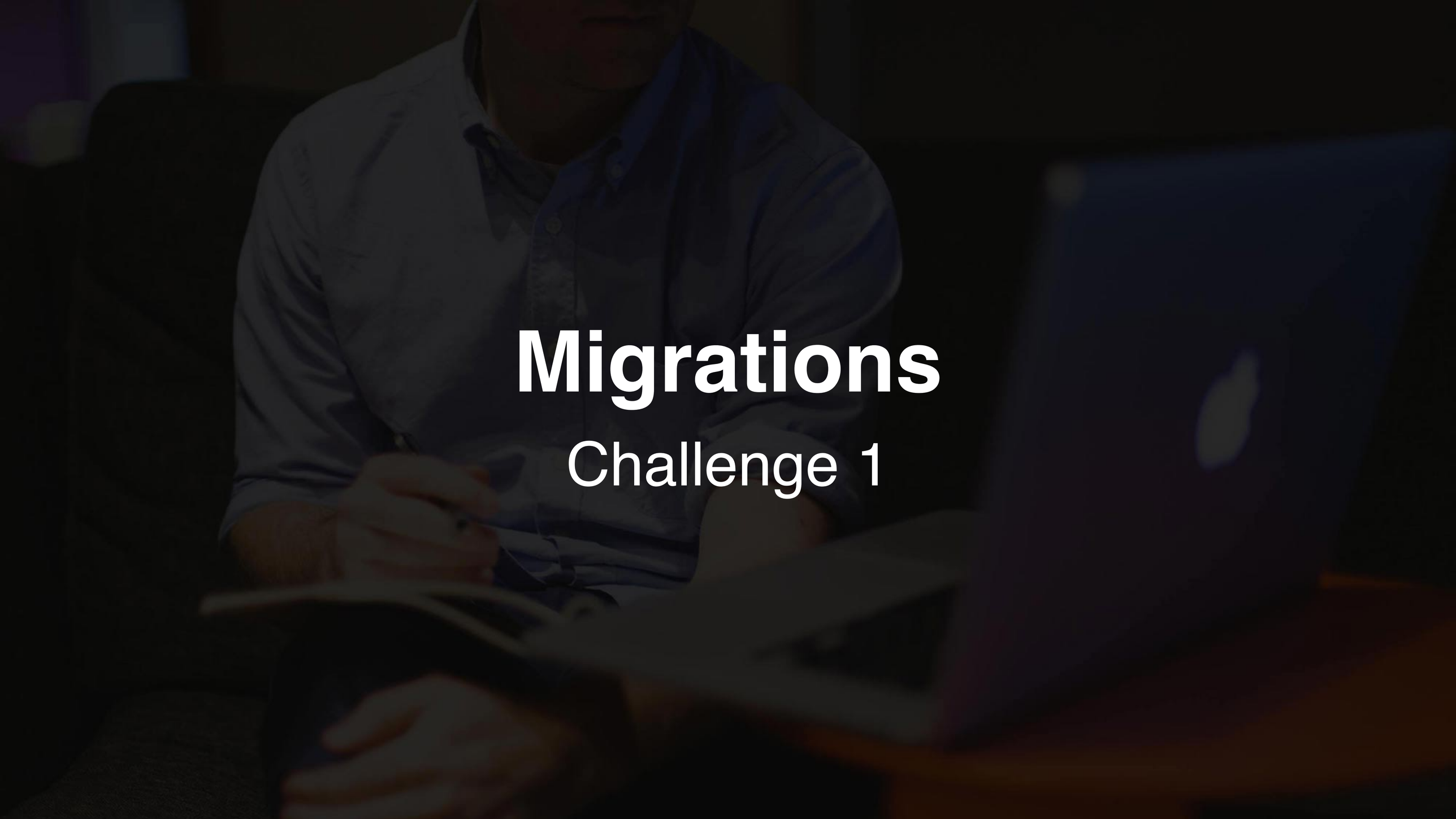
Migrations

Pro Tips

Pro Tips

A CLOUD GURU

- Technology is often a minor part of a cloud migration project.
- Project management discipline is a must — don’t underestimate this.
- Adapt the Cloud Adoption Framework for your own organization’s culture.
- Leverage the CAF to get buy-in by acknowledging the enterprise nature of cloud migrations.
- Be a Boundary Spanner!



Migrations

Challenge 1

Challenge 1

A CLOUD GURU

You have been hired by a company to help them design a comprehensive cloud migration strategy. You choose the leverage the Cloud Adoption Framework as a template to create an agenda for a kickoff meeting.

Which of the following points are part of the CAF?

A. With the shift toward cloud, some roles and skills transition from commodity to highly valued in the marketplace so the organization should consider its compensation structure.

B. Because customers may be impacted with the transition to the cloud, the organization needs to consider the appropriate way to communicate this transition to external stakeholders.

C. Transitioning to the cloud can change the risk profile of certain parts and functions of the organization so a risk management aspect should be included.

D. The organization needs a way to evaluate and measure creation of benefits, if any, to see if the cloud migration lives up to its business case.

E. The organization should consider how their existing software licensing can or cannot be used in a future cloud deployment.

F. A cloud migration often creates an opportunity to move from CapEx to OpEx, meaning the organization should probably consider a different accounting structure.

Challenge 1

A CLOUD GURU

You have been hired by a company to help them design a comprehensive cloud migration strategy. You choose the leverage the Cloud Adoption Framework as a template to create an agenda for a kickoff meeting.

Which of the following points are part of the CAF?

A. With the shift toward cloud, some roles and skills transition from commodity to highly valued in the marketplace so the organization should consider its compensation structure.

B. Because customers may be impacted with the transition to the cloud, the organization needs to consider the appropriate way to communicate this transition to external stakeholders.

C. Transitioning to the cloud can change the risk profile of certain parts and functions of the organization so a risk management aspect should be included.

D. The organization needs a way to evaluate and measure creation of benefits, if any, to see if the cloud migration lives up to its business case.

E. The organization should consider how their existing software licensing can or cannot be used in a future cloud deployment.

F. A cloud migration often creates an opportunity to move from CapEx to OpEx, meaning the organization should probably consider a different accounting structure.

Challenge 1

A CLOUD GURU

You have been hired by a company to help them design a comprehensive cloud migration strategy. You choose the leverage the Cloud Adoption Framework as a template to create an agenda for a kickoff meeting.

Which of the following points are part of the CAF?

A. With the shift toward cloud, some roles and skills transition from commodity to highly valued in the marketplace so the organization should consider its compensation structure.

B. Because customers may be impacted with the transition to the cloud, the organization needs to consider the appropriate way to communicate this transition to external stakeholders.

C. Transitioning to the cloud can change the risk profile of certain parts and functions of the organization so a risk management aspect should be included.

D. The organization needs a way to evaluate and measure creation of benefits, if any, to see if the cloud migration lives up to its business case.

E. The organization should consider how their existing software licensing can or cannot be used in a future cloud deployment.

F. A cloud migration often creates an opportunity to move from CapEx to OpEx, meaning the organization should probably consider a different accounting structure.

Challenge 1

A CLOUD GURU

You have been hired by a company to help them design a comprehensive cloud migration strategy. You choose the leverage the Cloud Adoption Framework as a template to create an agenda for a kickoff meeting.

Which of the following points are part of the CAF?

A. With the shift toward cloud, some roles and skills transition from commodity to highly valued in the marketplace so the organization should consider its compensation structure.

B. Because customers may be impacted with the transition to the cloud, the organization needs to consider the appropriate way to communicate this transition to external stakeholders.

C. Transitioning to the cloud can change the risk profile of certain parts and functions of the organization so a risk management aspect should be included.

D. The organization needs a way to evaluate and measure creation of benefits, if any, to see if the cloud migration lives up to its business case.

E. The organization should consider how their existing software licensing can or cannot be used in a future cloud deployment.

F. A cloud migration often creates an opportunity to move from CapEx to OpEx, meaning the organization should probably consider a different accounting structure.

Challenge 1

A CLOUD GURU

You have been hired by a company to help them design a comprehensive cloud migration strategy. You choose the leverage the Cloud Adoption Framework as a template to create an agenda for a kickoff meeting.

Which of the following points are part of the CAF?

A. With the shift toward cloud, some roles and skills transition from commodity to highly valued in the marketplace so the organization should consider its compensation structure.

B. Because customers may be impacted with the transition to the cloud, the organization needs to consider the appropriate way to communicate this transition to external stakeholders.

C. Transitioning to the cloud can change the risk profile of certain parts and functions of the organization so a risk management aspect should be included.

D. The organization needs a way to evaluate and measure creation of benefits, if any, to see if the cloud migration lives up to its business case.

E. The organization should consider how their existing software licensing can or cannot be used in a future cloud deployment.

F. A cloud migration often creates an opportunity to move from CapEx to OpEx, meaning the organization should probably consider a different accounting structure.

Challenge 1

A CLOUD GURU

You have been hired by a company to help them design a comprehensive cloud migration strategy. You choose the leverage the Cloud Adoption Framework as a template to create an agenda for a kickoff meeting.

Which of the following points are part of the CAF?

A. With the shift toward cloud, some roles and skills transition from commodity to highly valued in the marketplace so the organization should consider its compensation structure.

B. Because customers may be impacted with the transition to the cloud, the organization needs to consider the appropriate way to communicate this transition to external stakeholders.

C. Transitioning to the cloud can change the risk profile of certain parts and functions of the organization so a risk management aspect should be included.

D. The organization needs a way to evaluate and measure creation of benefits, if any, to see if the cloud migration lives up to its business case.

E. The organization should consider how their existing software licensing can or cannot be used in a future cloud deployment.

F. A cloud migration often creates an opportunity to move from CapEx to OpEx, meaning the organization should probably consider a different accounting structure.

Migrations

Challenge 2

Challenge 2

A company has chosen to adopt a cloud-first IT architecture strategy and they are looking for advice on some applications in their portfolio.

They currently host on-prem a large Sybase database which holds a customer database and a smaller MongoDB database used for tracking serial numbers of manufactured products. Both databases are currently deployed as highly available across dual data centers and on VMware virtualized systems.

They are looking for improvements in long-term cost and manageability without compromising availability. Modifications to their existing applications which use the databases may be in order but the extent of changes required is not yet known.

How would you recommend they proceed with migrating these systems to AWS?

A. Take a wait-and-see approach as their current deployment is likely reliable and cost effective enough for the short term.

B. Use SCT to do a trial conversion of both the Sybase and MongoDB databases into an instance of Amazon Aurora. Evaluate the results of that exercise and decide the next step.

C. Use the VMware Import process to migrate the VM's into AWS as EC2 instances. Deploy across at least two AZs to ensure HA. Automate management activities with OpsWorks.

D. Use DMS to run a trial conversion of the MongoDB database to DynamoDB and evaluate the results. Use the VMware Import process to migrate the Sybase database into AWS as EC2 instances and deploy across two AZs.

E. Restore a recent backup of the Sybase database to RDS. Restore a recent backup of MongoDB to DynamoDB. Evaluate the results of that exercise and decide the next step.

Challenge 2

A company has chosen to adopt a cloud-first IT architecture strategy and they are looking for advice on some applications in their portfolio.

They currently host on-prem a large Sybase database which holds a customer database and a smaller MongoDB database used for tracking serial numbers of manufactured products. Both databases are currently deployed as highly available across dual data centers and on VMware virtualized systems.

They are looking for improvements in long-term cost and manageability without compromising availability. Modifications to their existing applications which use the databases may be in order but the extent of changes required is not yet known.

How would you recommend they proceed with migrating these systems to AWS?

A. Take a wait-and-see approach as their current deployment is likely reliable and cost effective enough for the short term.

B. Use SCT to do a trial conversion of both the Sybase and MongoDB databases into an instance of Amazon Aurora. Evaluate the results of that exercise and decide the next step.

C. Use the VMware Import process to migrate the VM's into AWS as EC2 instances. Deploy across at least two AZs to ensure HA. Automate management activities with OpsWorks.

D. Use DMS to run a trial conversion of the MongoDB database to DynamoDB and evaluate the results. Use the VMware Import process to migrate the Sybase database into AWS as EC2 instances and deploy across two AZs.

E. Restore a recent backup of the Sybase database to RDS. Restore a recent backup of MongoDB to DynamoDB. Evaluate the results of that exercise and decide the next step.

Challenge 2

A company has chosen to adopt a cloud-first IT architecture strategy and they are looking for advice on some applications in their portfolio.

They currently host on-prem a large Sybase database which holds a customer database and a smaller MongoDB database used for tracking serial numbers of manufactured products. Both databases are currently deployed as highly available across dual data centers and on VMware virtualized systems.

They are looking for improvements in long-term cost and manageability without compromising availability. Modifications to their existing applications which use the databases may be in order but the extent of changes required is not yet known.

How would you recommend they proceed with migrating these systems to AWS?

A. Take a wait-and-see approach as their current deployment is likely reliable and cost effective enough for the short term.

B. Use SCT to do a trial conversion of both the Sybase and MongoDB databases into an instance of Amazon Aurora. Evaluate the results of that exercise and decide the next step.

C. Use the VMware Import process to migrate the VM's into AWS as EC2 instances. Deploy across at least two AZs to ensure HA. Automate management activities with OpsWorks.

D. Use DMS to run a trial conversion of the MongoDB database to DynamoDB and evaluate the results. Use the VMware Import process to migrate the Sybase database into AWS as EC2 instances and deploy across two AZs.

E. Restore a recent backup of the Sybase database to RDS. Restore a recent backup of MongoDB to DynamoDB. Evaluate the results of that exercise and decide the next step.

Architecting to Scale

Introduction

A CLOUD GURU

Exam Blueprint

A CLOUD GURU

Domain 1: Design for Organizational Complexity

1.1. Determine cross-account authentication and access strategy for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

1.2. Determine how to design networks for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

1.3. Determine how to design a multi-account AWS environment for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

Domain 2: Design for New Solutions

2.1. Determine security requirements and controls when designing and implementing a solution.

2.2. Determine a solution design and implementation strategy to meet reliability requirements.

2.3. Determine a solution design to ensure business continuity.

2.4. Determine a solution design to meet performance objectives.

2.5. Determine a deployment strategy to meet business requirements when designing and implementing a solution.

Exam Blueprint

A CLOUD GURU

Domain 3: Migration Planning

3.1. Select existing workloads and processes for potential migration to the cloud.

3.2. Select migration tools and/or services for new and migrated solutions based on detailed AWS knowledge.

3.3. Determine a new cloud architecture for an existing solution.

3.4. Determine a strategy for migrating existing on-premises workloads to the cloud.

Domain 4: Cost Control

4.1. Select a cost-effective pricing model for a solution.

4.2. Determine which controls to design and implement that will ensure cost optimization.

4.3. Identify opportunities to reduce cost in an existing solution.

Domain 5: Continuous Improvement for Existing Solutions

5.1. Troubleshoot solution architectures.

5.2. Determine a strategy to improve an existing solution for operational excellence.

5.3. Determine a strategy to improve the reliability of an existing solution.

5.4. Determine a strategy to improve the performance of an existing solution.

5.5. Determine a strategy to improve the security of an existing solution.

5.6. Determine how to improve the deployment of an existing solution.

Architecting to Scale

Concepts

Architectural Patterns

A CLOUD GURU

Loosely Coupled Architecture

A CLOUD GURU

"Components can stand independently and require little or no knowledge of the inner workings of the other components."

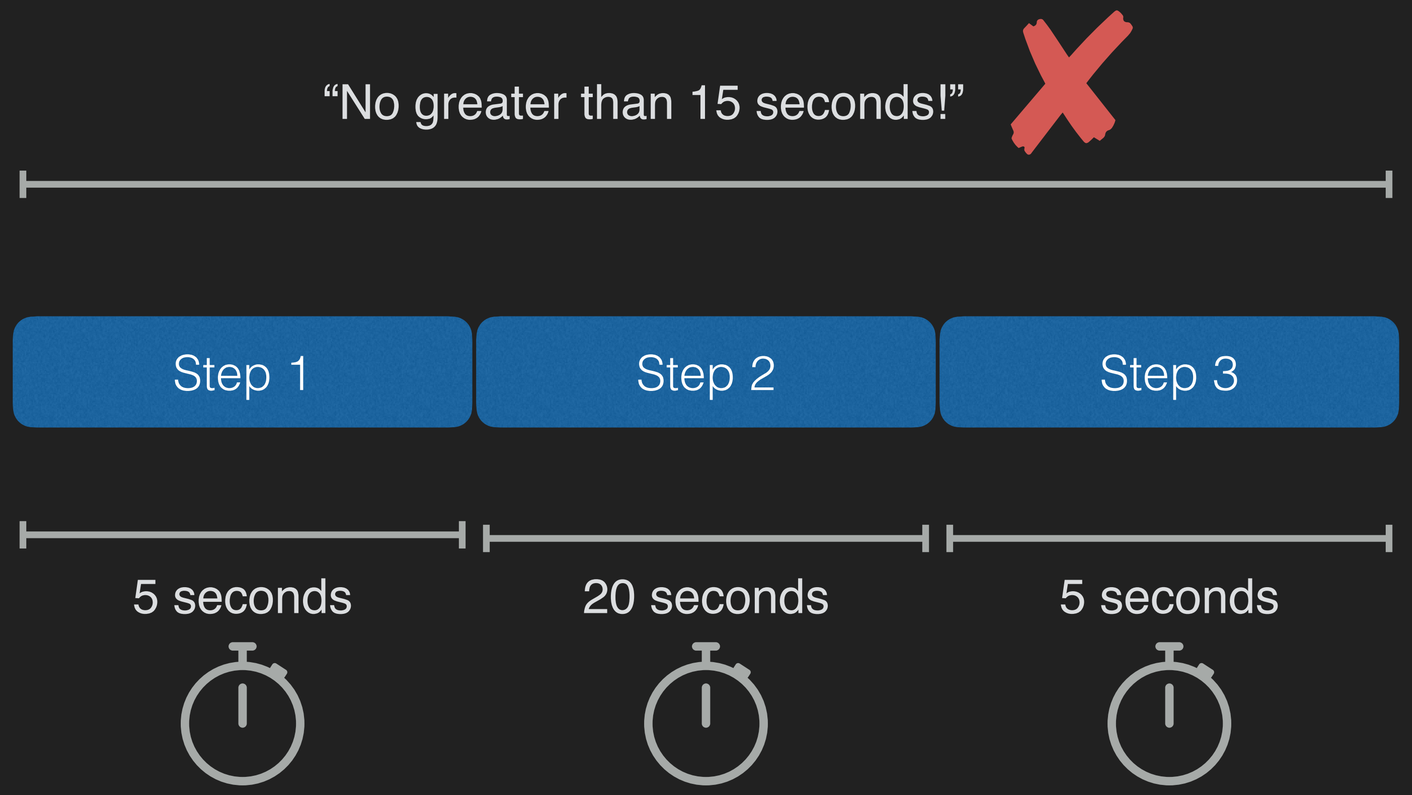
Loosely Coupled Architecture

A CLOUD GURU

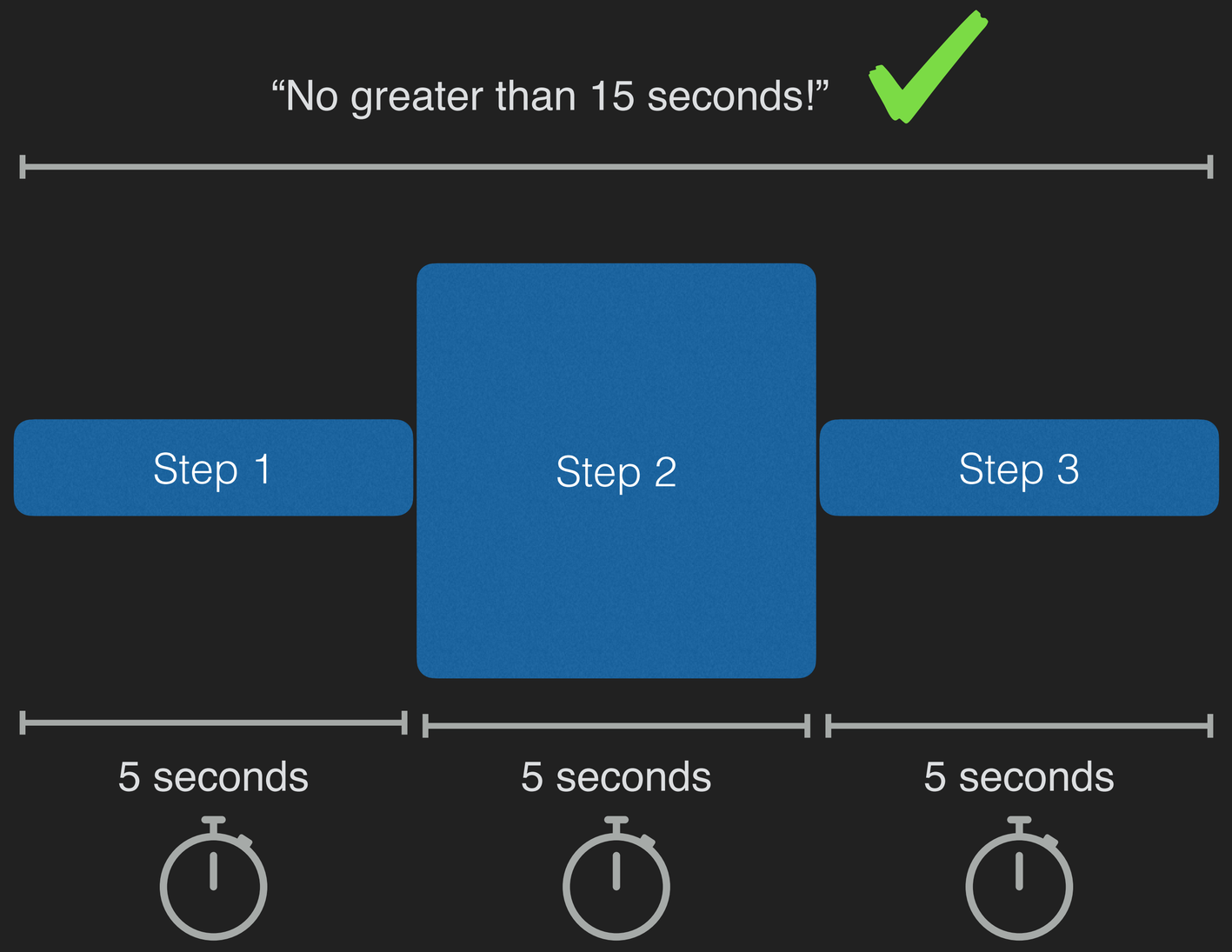
- Layers of abstraction
- Permits more flexibility
- Interchangeable components
- More atomic functional units
- Can scale components independently

Let’s Look At A Process…

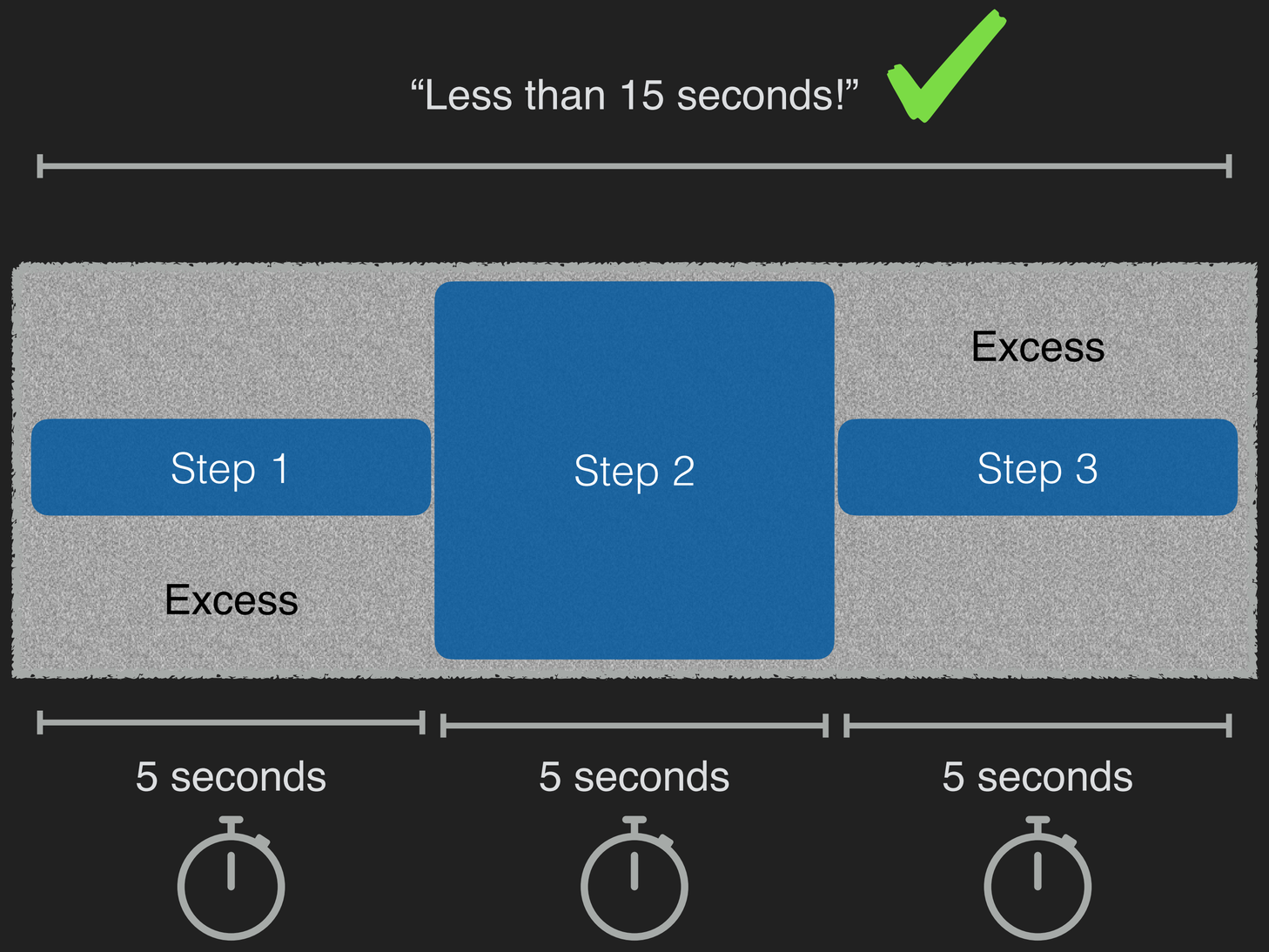
A CLOUD GURU



Accommodating A Service Level

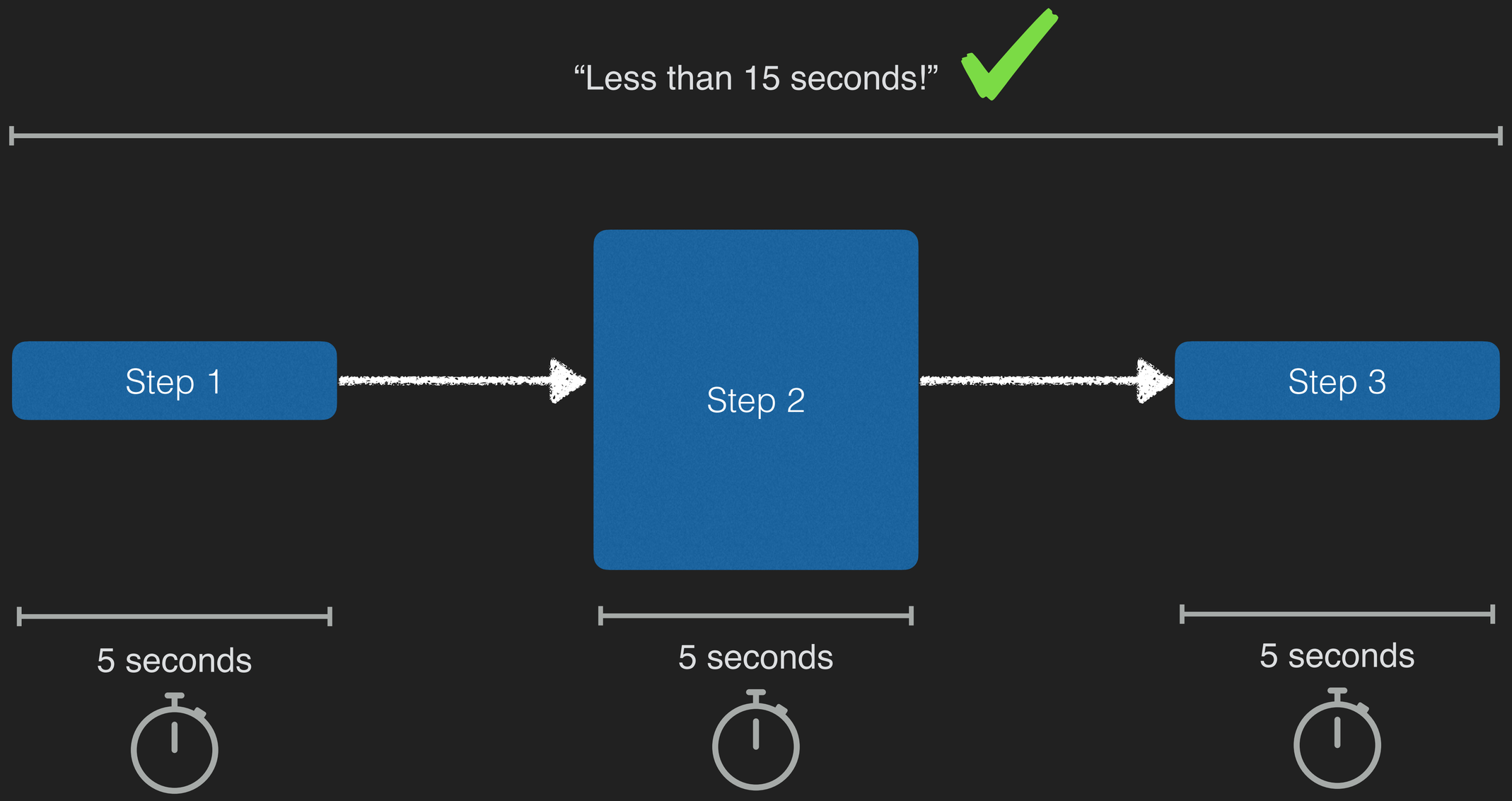


Tightly Coupled



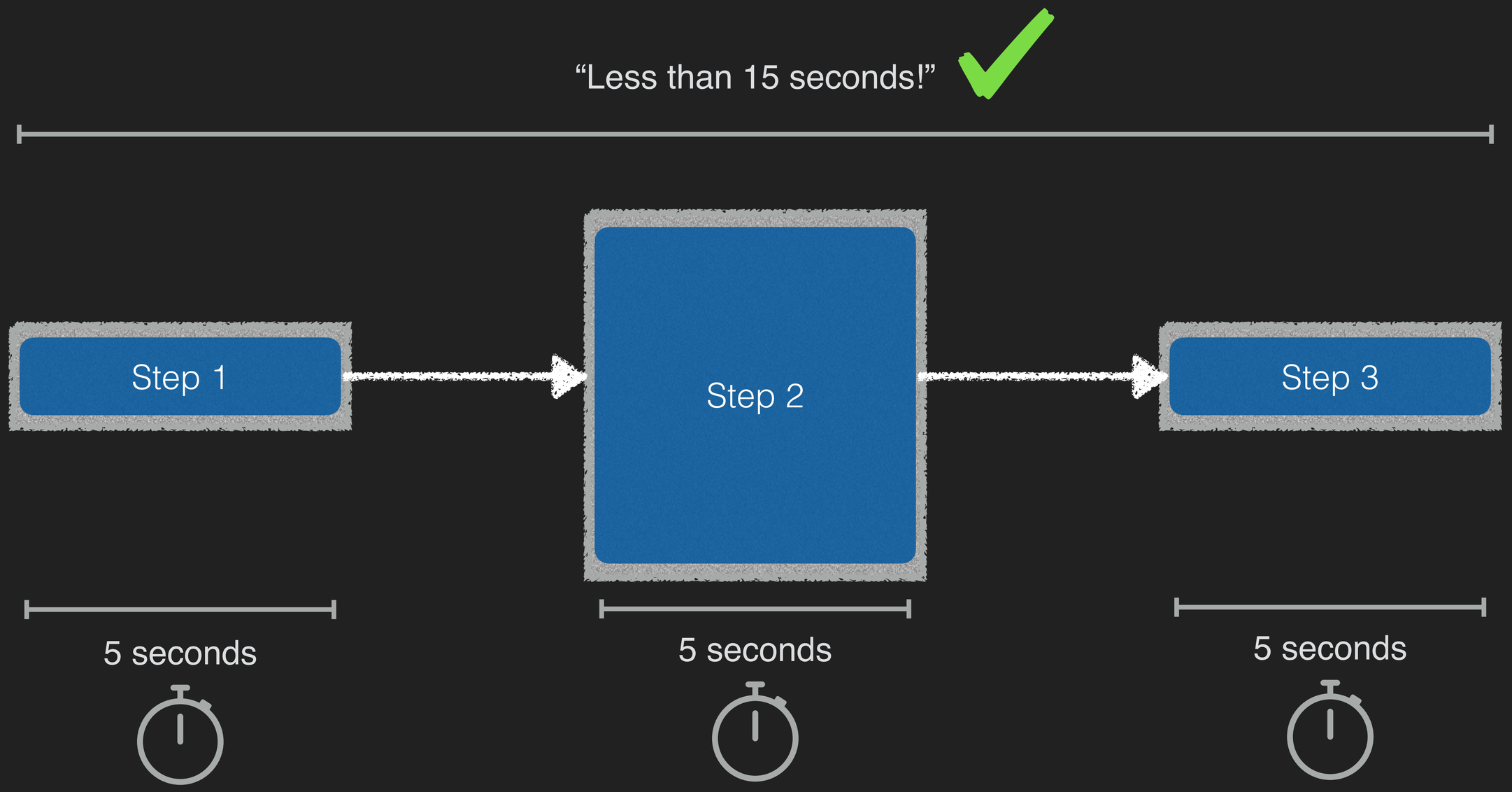
Loosely Coupled

A CLOUD GURU

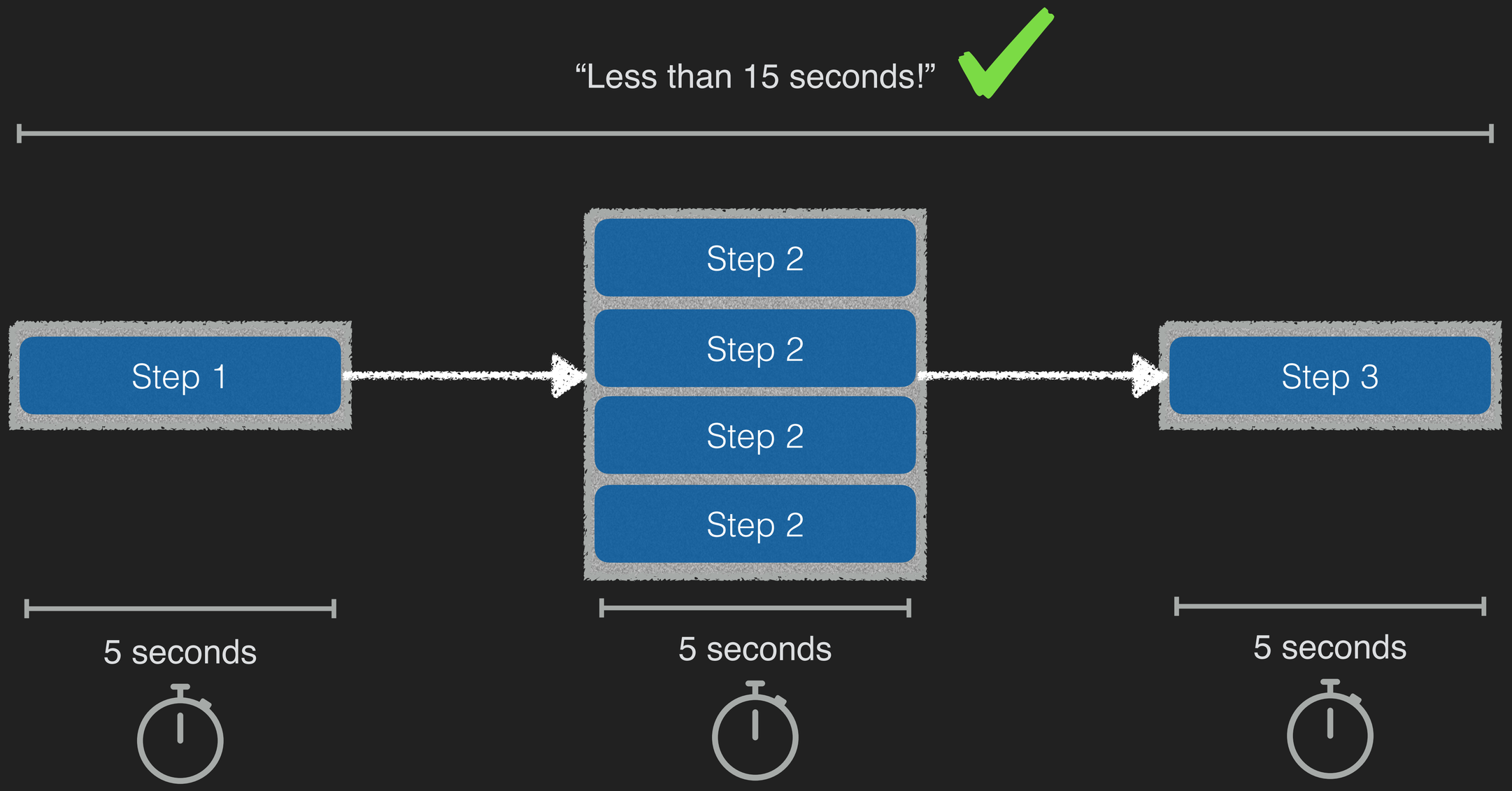


Scale Vertically

A CLOUD GURU



Scale Horizontally



Horizontal vs. Vertical Scaling

A CLOUD GURU

Horizontal Scaling	Vertical Scaling
Add more instances as demand increases	Add more CPU and/or RAM to existing instance as demand increases
No downtime required to scale up or down	Requires restart to scale up or down
Automatic using Auto-scaling Groups	Would require scripting to automate
(Theoretically) Unlimited	Limited by instance sizes

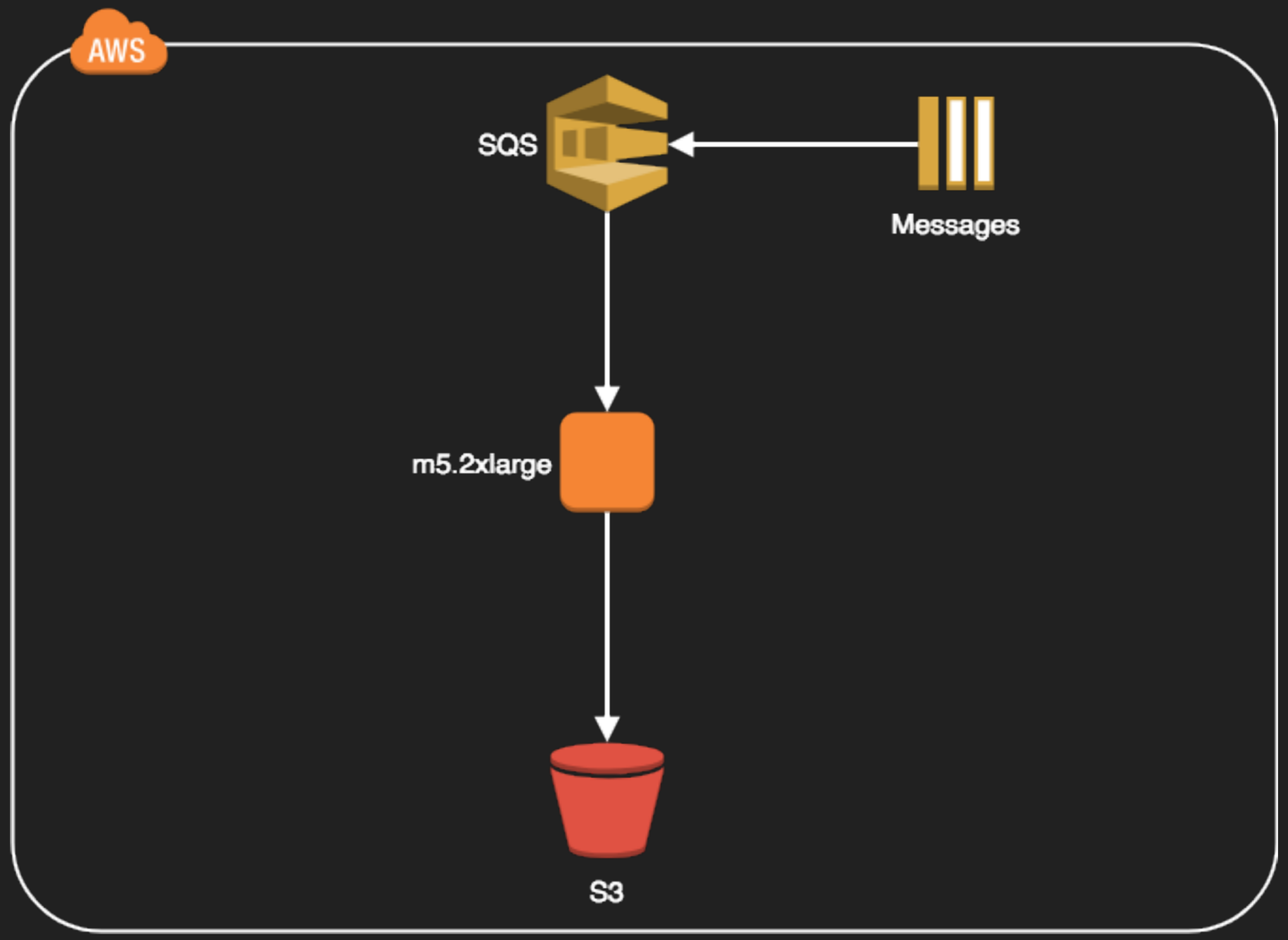
Scale-Out Scale-In Scale-Up Scale-Down

A CLOUD GURU

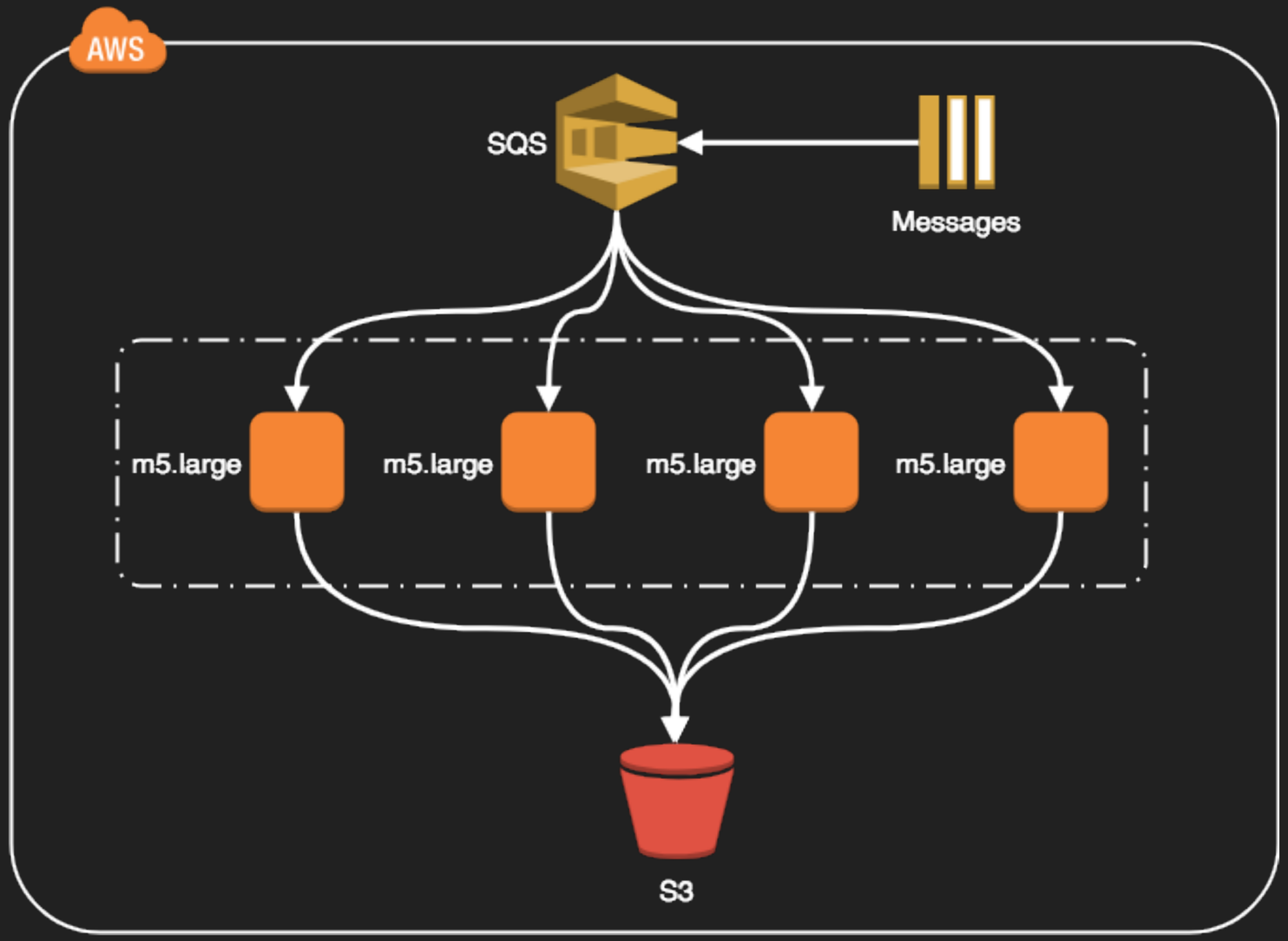
Scale Out	
Scale Up	
Scale In	
Scale Down	

Sample Message Processing Architecture

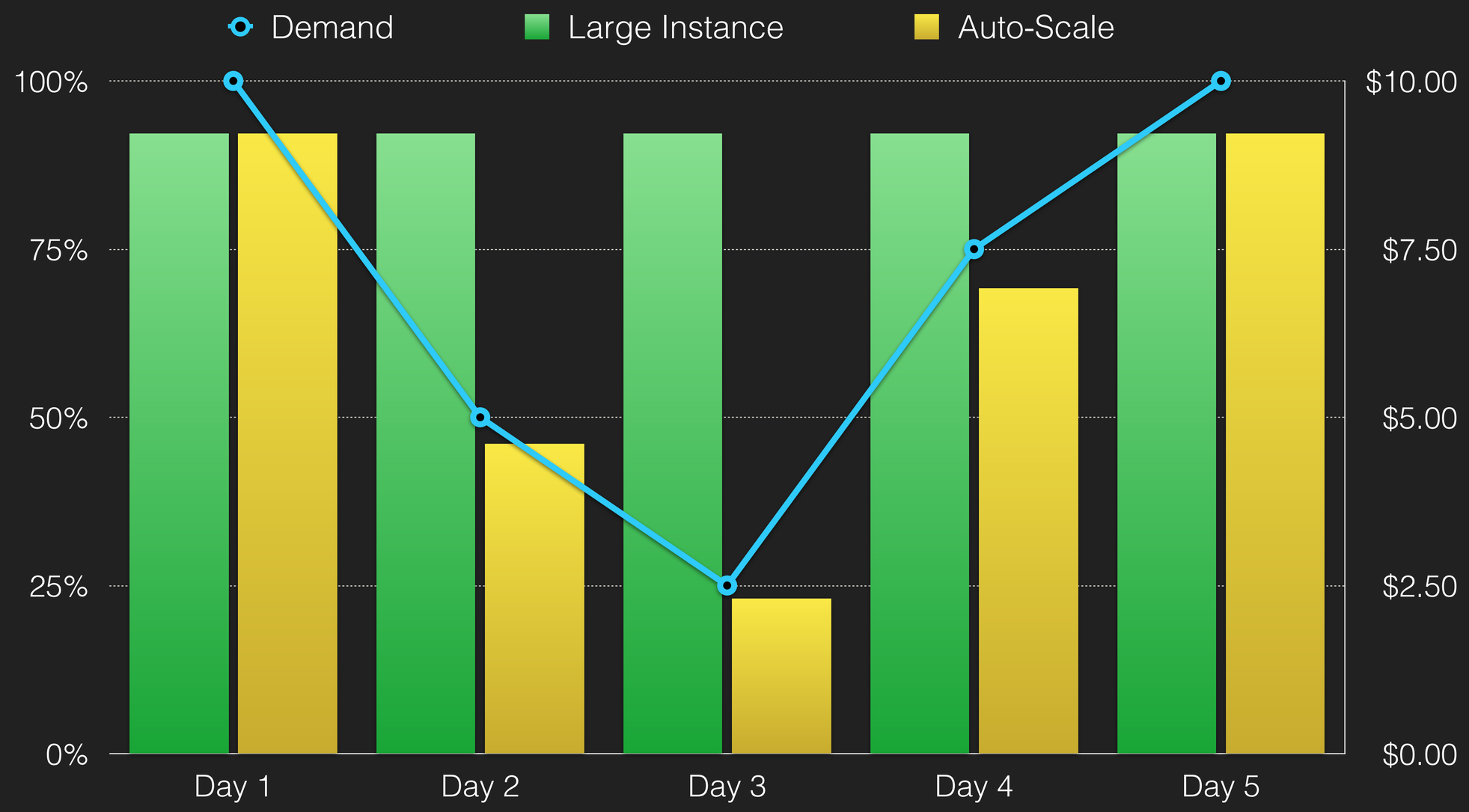
A CLOUD GURU



Scale-Out Architecture



Scaling to Match Demand



Scaling Economics

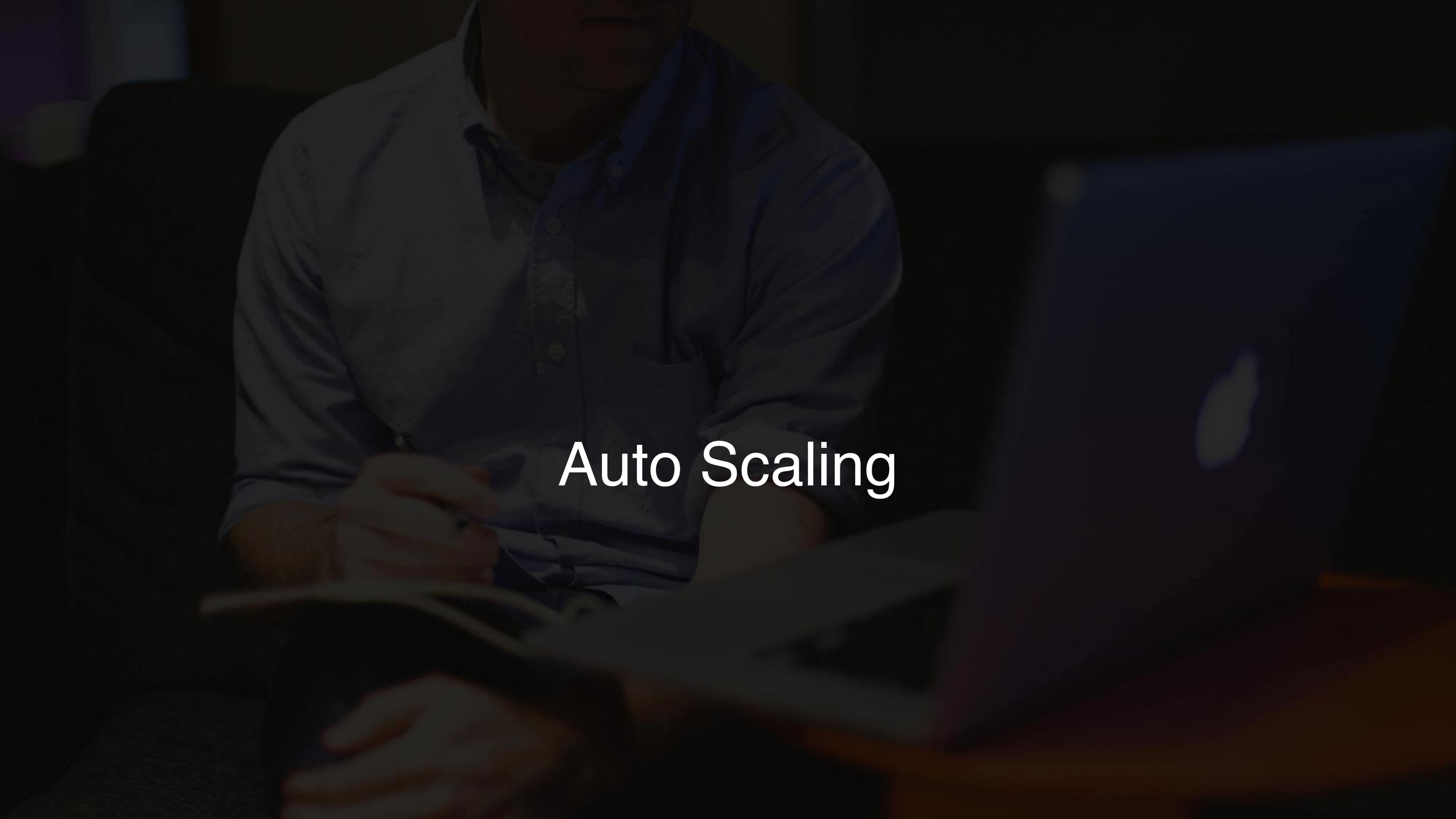
A CLOUD GURU

No Scaling	$46.08
Auto-Scaling	$32.26
Savings	$13.82

Scaling Economics

A CLOUD GURU

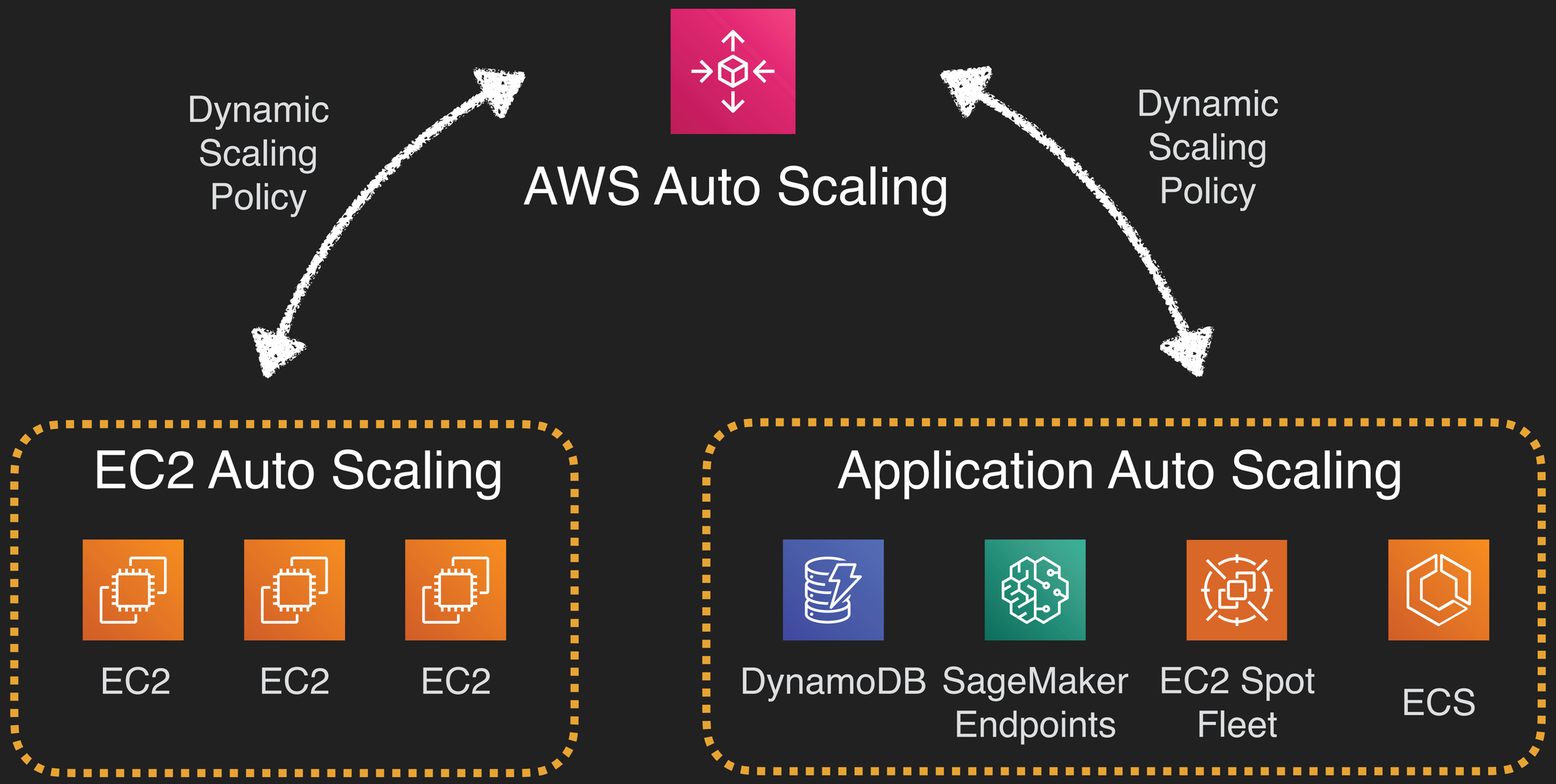
No Scaling	$46.08
Auto-Scaling	$32.26
Savings	$13.82



Auto Scaling

Types of Auto-Scaling

A CLOUD GURU



Types of Auto-Scaling

A CLOUD GURU

Name	What	Why
Amazon EC2 Auto Scaling	Like the name says…focused on EC2 and nothing else.	Setup scaling groups for EC2 instances; health checks to remove unhealthy instances
Application Auto Scaling	API used to control scaling for resources other than EC2 like Dynamo, ECS, EMR	Provides a common way to interact with the scalability of other services
AWS Auto Scaling	Provides centralized way to manage scalability for whole stacks; Predictive scaling feature	Console that can manage both of the above from a unified standpoint

EC2 Auto-Scaling Groups

A CLOUD GURU

- Automatically provides horizontal scaling (scale-out) for your landscape.
- Triggered by an event or scaling action to either launch or terminate instances.
- Availability, Cost and System Metrics can all factor into scaling.
- Four scaling options:

  Maintain - Keep a specific or minimum number of instances running

  Manual - Use maximum, minimum, or specific number of instances

  Schedule - Increase or decrease instances based on schedule

  Dynamic - Scale based on real-time metrics of the systems

Auto-Scaling

Launch Configuraitons

A CLOUD GURU

- Specify VPC and subnets for scaled instances
- Attach to a ELB
- Define a Health Check Grace Period

aws Services Resource Groups scott @ pletcher-a

1. Configure Auto Scaling group details | 2. Configure scaling policies | 3. Configure Notifications | 4. Configure Tags | 5. Review

Create Auto Scaling Group

Launch Configuration My Web Servers

Group name My Web Server Group

Group size Start with 1 instances

Network vpc-bd100ec4 (172.31.0.0/16) (default) Create new VPC

Subnet Create new subnet

Each instance in this Auto Scaling group will be assigned a public IP address.

Advanced Details

Load Balancing Receive traffic from one or more load balancers Learn about Elastic Load Balancing

Health Check Grace Period 300 seconds

Monitoring Amazon EC2 Detailed Monitoring metrics, which are provided at 1 minute frequency, are not enabled for the launch configuration My Web Servers. Instances launched from it will use Basic Monitoring metrics, provided at 5 minute frequency.
Learn more

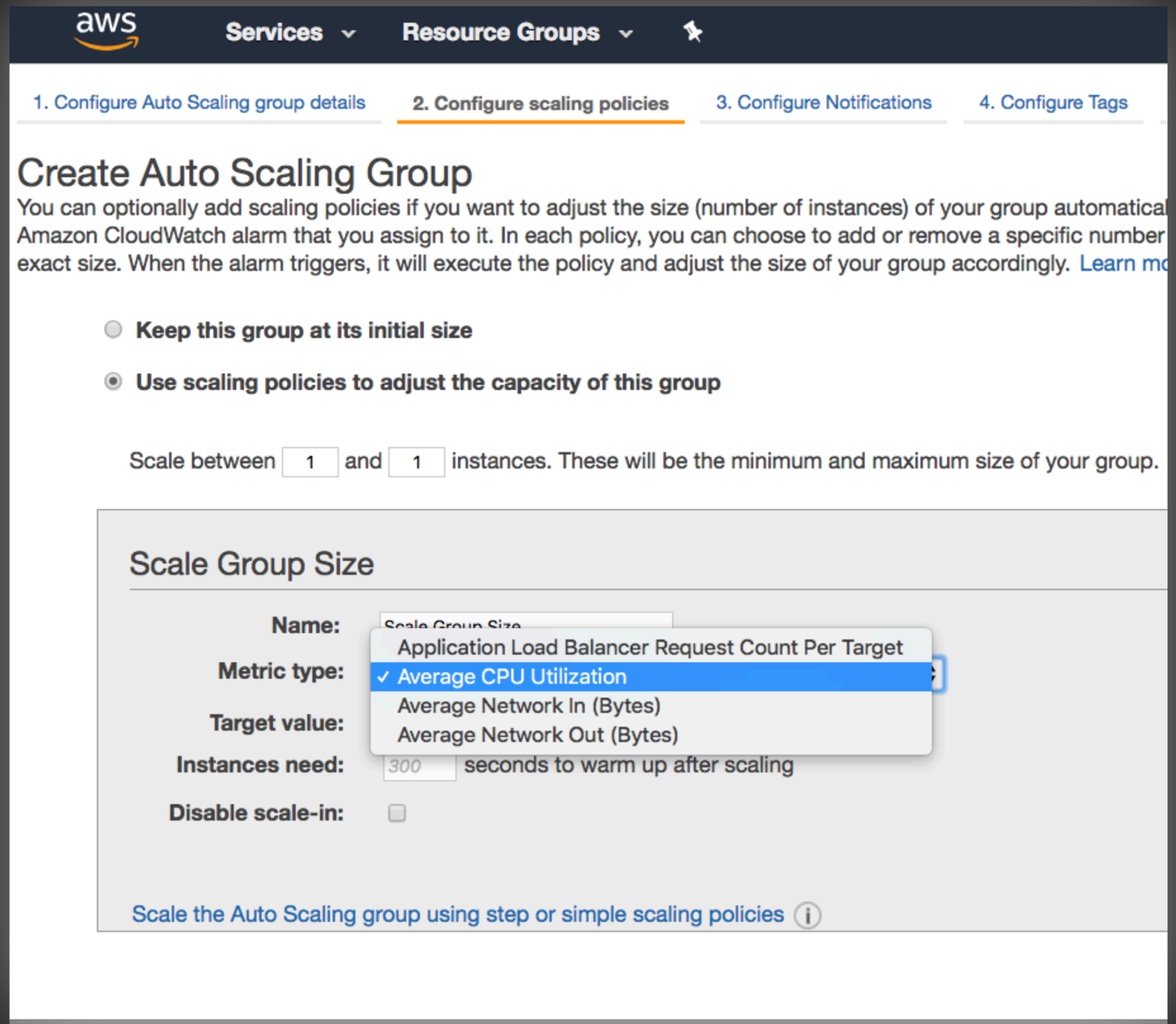
Instance Protection

Service-Linked Role AWSServiceRoleForAutoScaling

The default role does not exist. It will be automatically created on your behalf.

Launch Configuraitons

A CLOUD GURU



- Define size of group to stay at initial size.
- Or use scaling policy which can be based from metrics

Scaling Types

A CLOUD GURU

Scaling Type	What	When
Maintain	Hands-off way to maintain X number of instances	“I need 3 instances always.”
Manual	Manually change desired capacity via console or CLI	“My needs change so rarely that I can just manually add and remove”
Scheduled	Adjust min/max instances based on specific times	“Every Monday morning, we get a rush on our website.”
Dynamic	Scale in response to behavior of elements in the environment	“When CPU utilization gets to 70% on current instnaces, scale up.”

EC2 Auto Scaling Policies

A CLOUD GURU

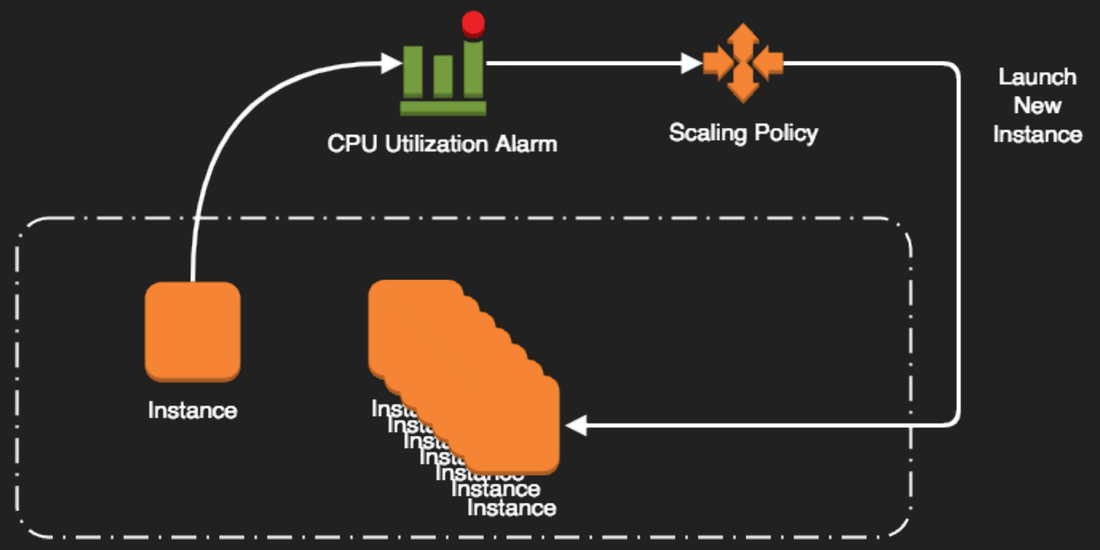
Scaling	What	When
Target Tracking Policy	Scale based on a predefined or custom metric in relation to a target value	“When CPU utilization gets to 70% on current instances, scale up.”
Simple Scaling Policy	Waits until health check and cool down period expires before evaluating new need	“Let’s add new instances slow and steady.”
Step Scaling Policy	Responds to scaling needs with more sophistication and logic	“AGG! Add ALL the instances!”

Scaling Cooldown Concept for EC2

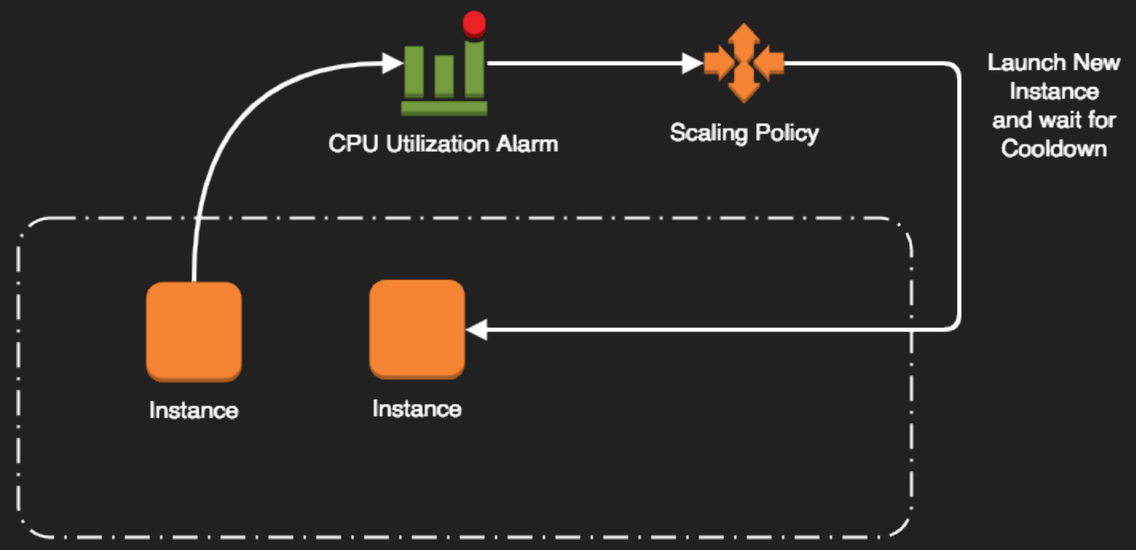
A CLOUD GURU

- Configurable duration that gives your scaling a chance to “come up to speed” and absorb load.
- Default cooldown period is 300 seconds.
- Automatically applies to dynamic scaling and optionally to manual scaling but not supported for scheduled scaling.
- Can override default cooldown via scaling-specific cool down

No Cooldown

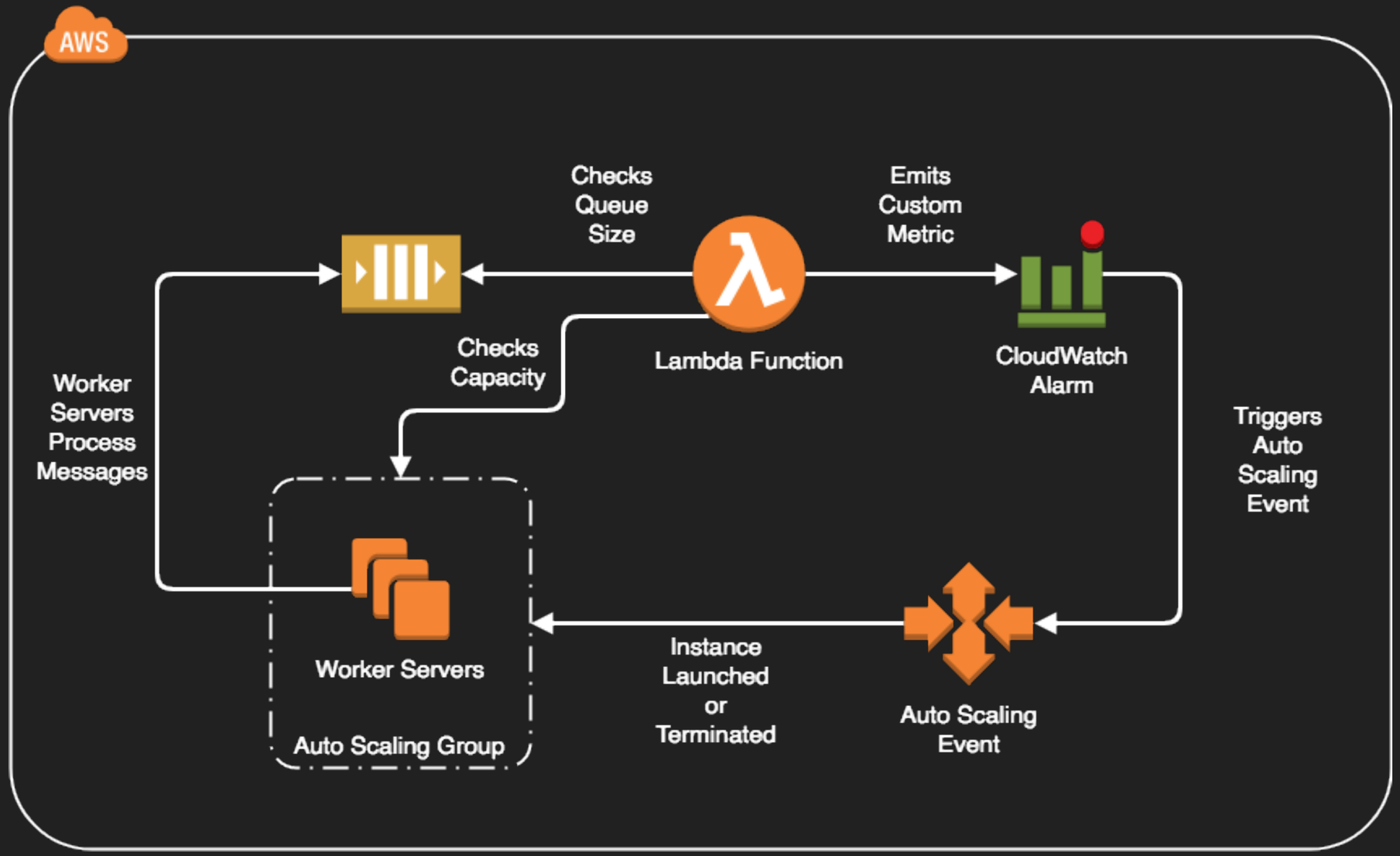


With Cooldown Period



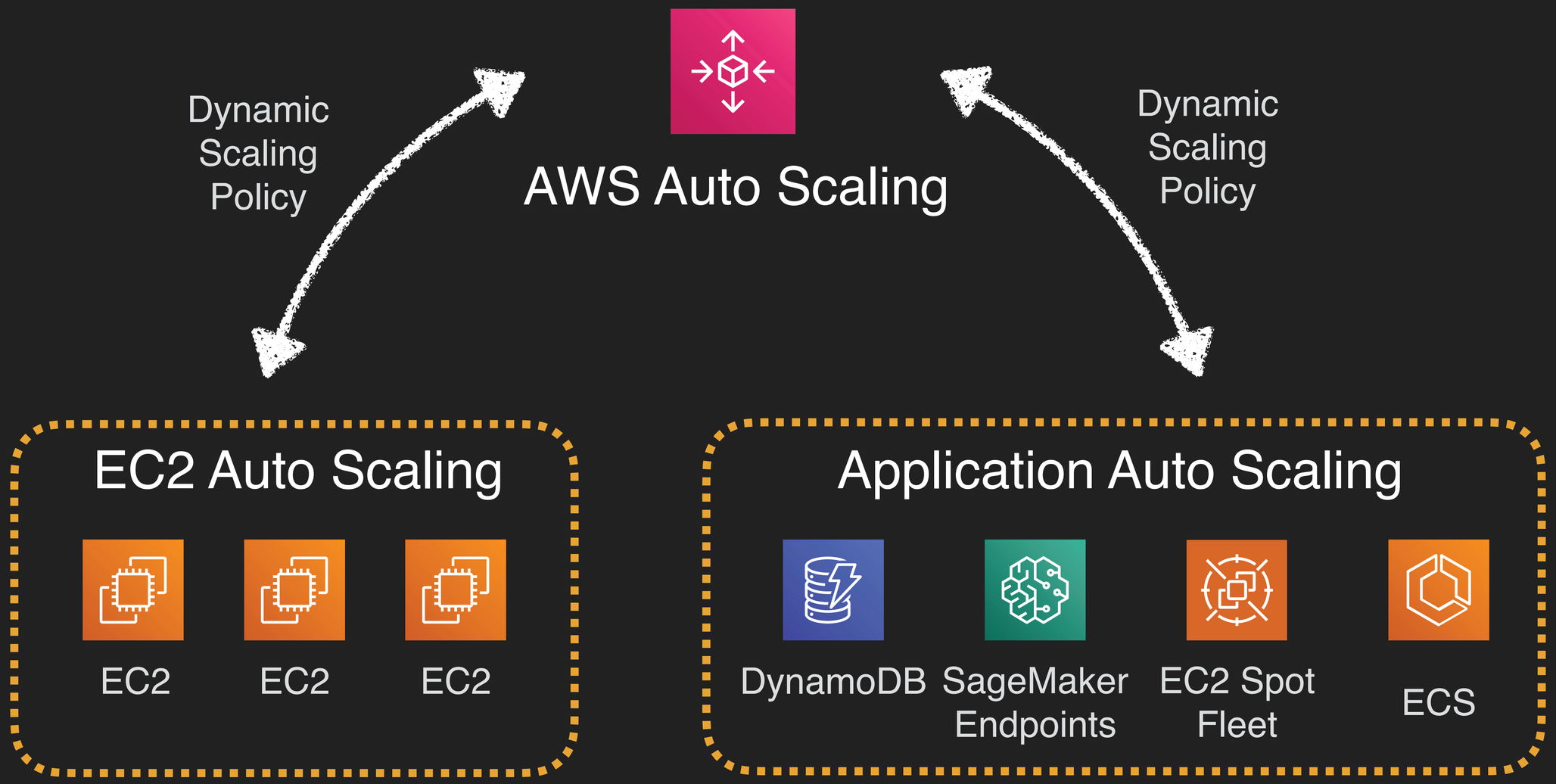
Scaling Based on Amazon SQS

A CLOUD GURU



Types of Auto-Scaling

A CLOUD GURU



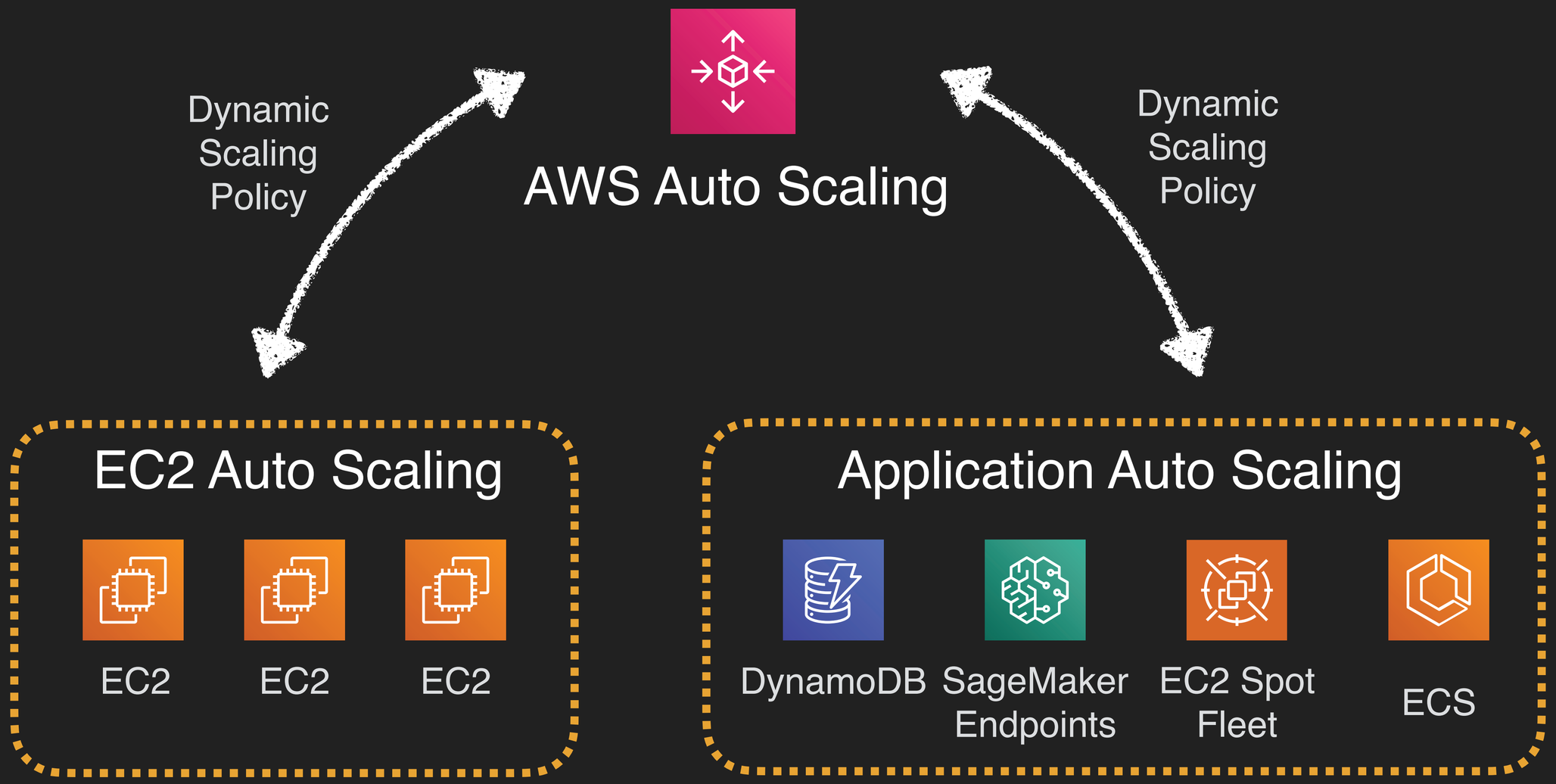
AWS Application Auto Scaling

A CLOUD GURU

Scaling	What	When
Target Tracking Policy	Initiates scaling events to try to track as closely as possible a given target metric.	“I want my ECS hosts to stay at or below 70% CPU utilization.”
Step Scaling Policy	Based on a metric, adjusts capacity given certain defined thresholds.	“I want to increase my EC2 Spot Fleet by 20% every time I add another 10,000 connections on my ELB”
Scheduled Scaling Policy	Initiates scaling events based on a predefined time, day or date.	“Every Monday at 0800, I want to increase the Read Capacity Units of my DynamoDB Table to 20,000”

Types of Auto-Scaling

A CLOUD GURU



AWS Auto Scaling

A CLOUD GURU

aws Services Resource Groups scott @ pletcher-acg-master Ohio Support

AWS Auto Scaling > Scaling plans > Create scaling plan

Step 1
Find scalable resources

Step 2
Specify scaling strategy

Step 3
Configure advanced settings (optional)

Step 4
Review and create

Find scalable resources

Automatically discover or manually choose resources to add to your scaling plan. Info

Choose a method

- Search by CloudFormation stack
  Search for resources provisioned by an AWS CloudFormation stack.
- Search by tag
  Search for resources by tags applied to them.
- Choose EC2 Auto Scaling groups
  Choose one or more Auto Scaling groups to include in your scaling plan.

AWS Auto Scaling

A CLOUD GURU

aws Services Resource Groups scott @ pletcher-acg-master Ohio Support

DynamoDB tables and indexes (2)

Specify a scaling strategy for 1 DynamoDB table and/or index. The read and write capacity units are configured separately.

☑ Include in scaling plan

Scaling strategy

The strategy defines the scaling metric and target value used to scale your resources.

- ◉ **Optimize for availability** — Keep the read and write capacity utilization of your DynamoDB tables and indexes at 40% to provide high availability and ensure capacity to absorb spikes in demand.
- ○ **Balance availability and cost** — Keep the read and write capacity utilization of your DynamoDB tables and indexes at 50% to provide optimal availability and reduce costs.
- ○ **Optimize for cost** — Keep the read and write capacity utilization of your DynamoDB tables and indexes at 70% to ensure lower costs.
- ○ **Custom** — Choose your own scaling metric, target value, and other settings.

☑ Enable dynamic scaling

Support your scaling strategy by creating target tracking scaling policies to monitor your scaling metric and increase or decrease capacity as you need it. Info

▶ **Configuration details**

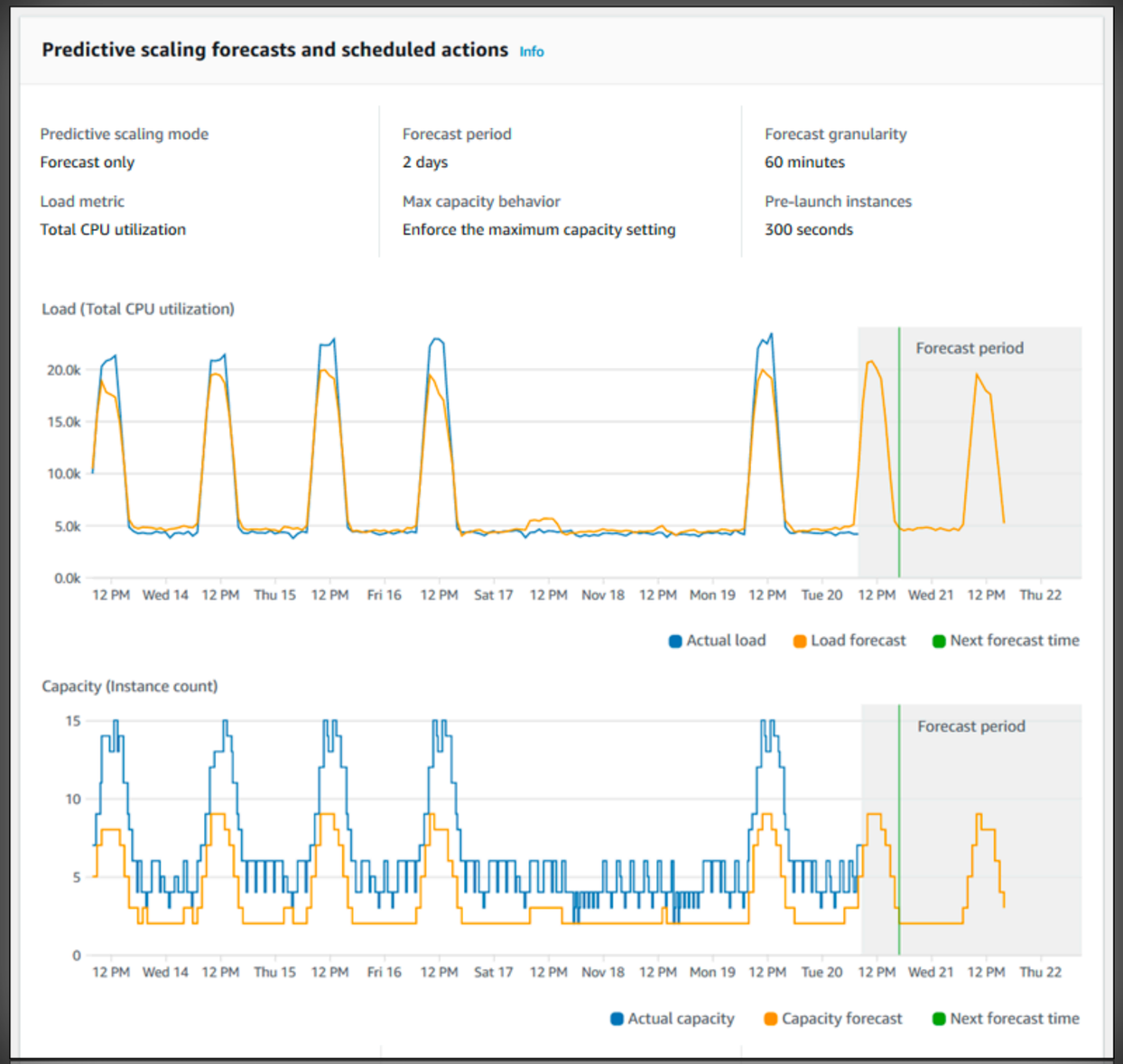
AWS Predictive Scaling

A CLOUD GURU

Predictive scaling can dynamically scale based on learning your load and calculating expected capacity.

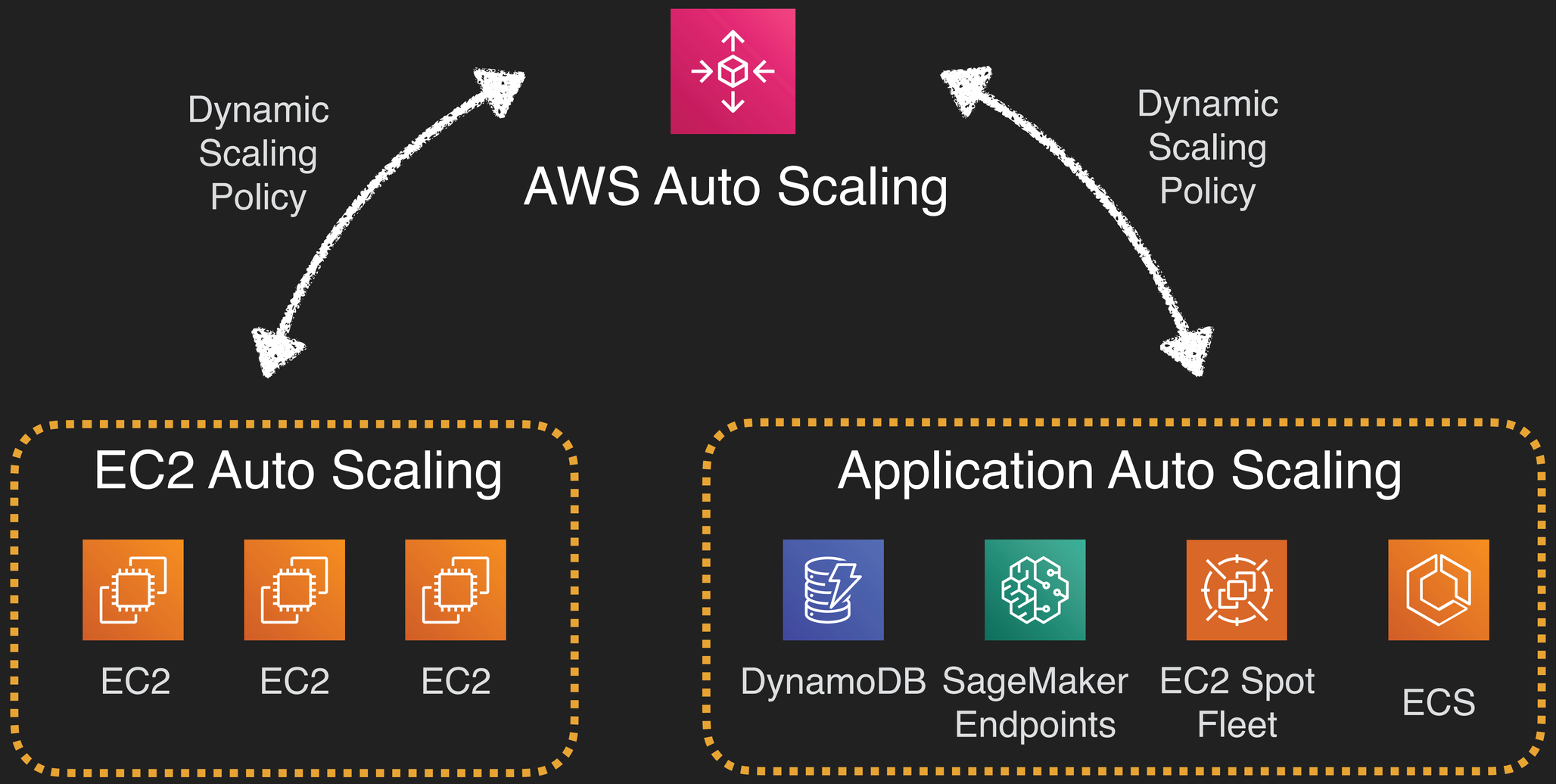
Without dynamic scaling, you can just use the data to adjust your own scaling policies.

You can also opt-out of this if you don't want AWS collecting this data.



Types of Auto-Scaling

A CLOUD GURU



Scalability and Elasticity

Kinesis

Kinesis

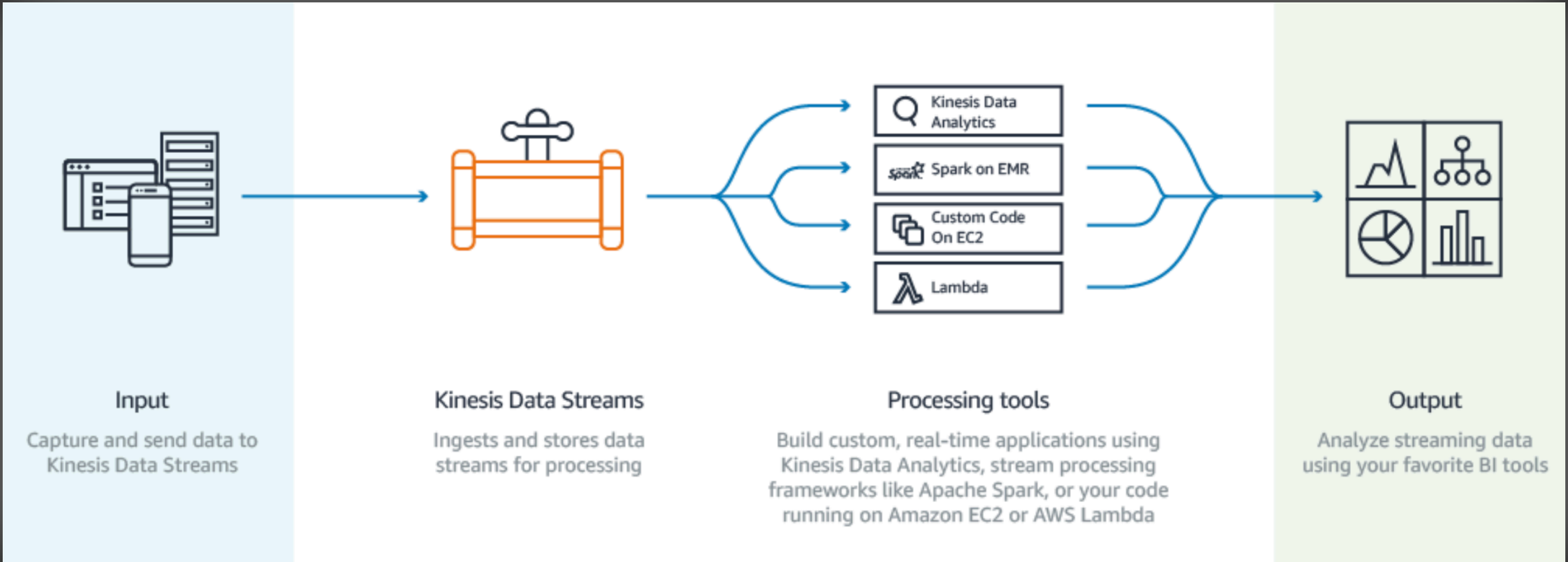
A CLOUD GURU

- Collection of services for processing streams of various data.
- Data is processed in “shards” — with each shard able to ingest 1000 records per second.
- A default limit of 500 shards, but you can request an increase to unlimited shards.
- Record consists of Partition Key, Sequence Number and Data Blob (up to 1 MB).
- Transient Data Store — Default retention of 24 hours, but can be configured for up to 7 days.

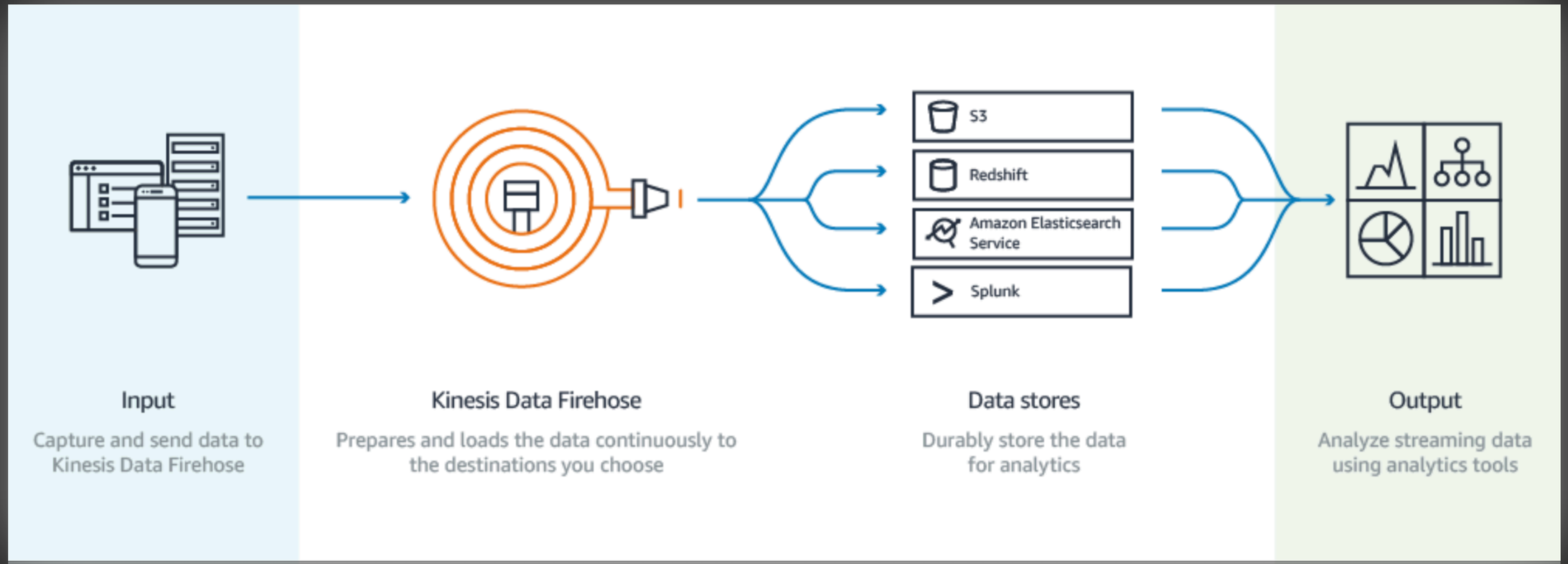
AWS Kinesis

Kinesis Video Streams

Kinesis Data Streams

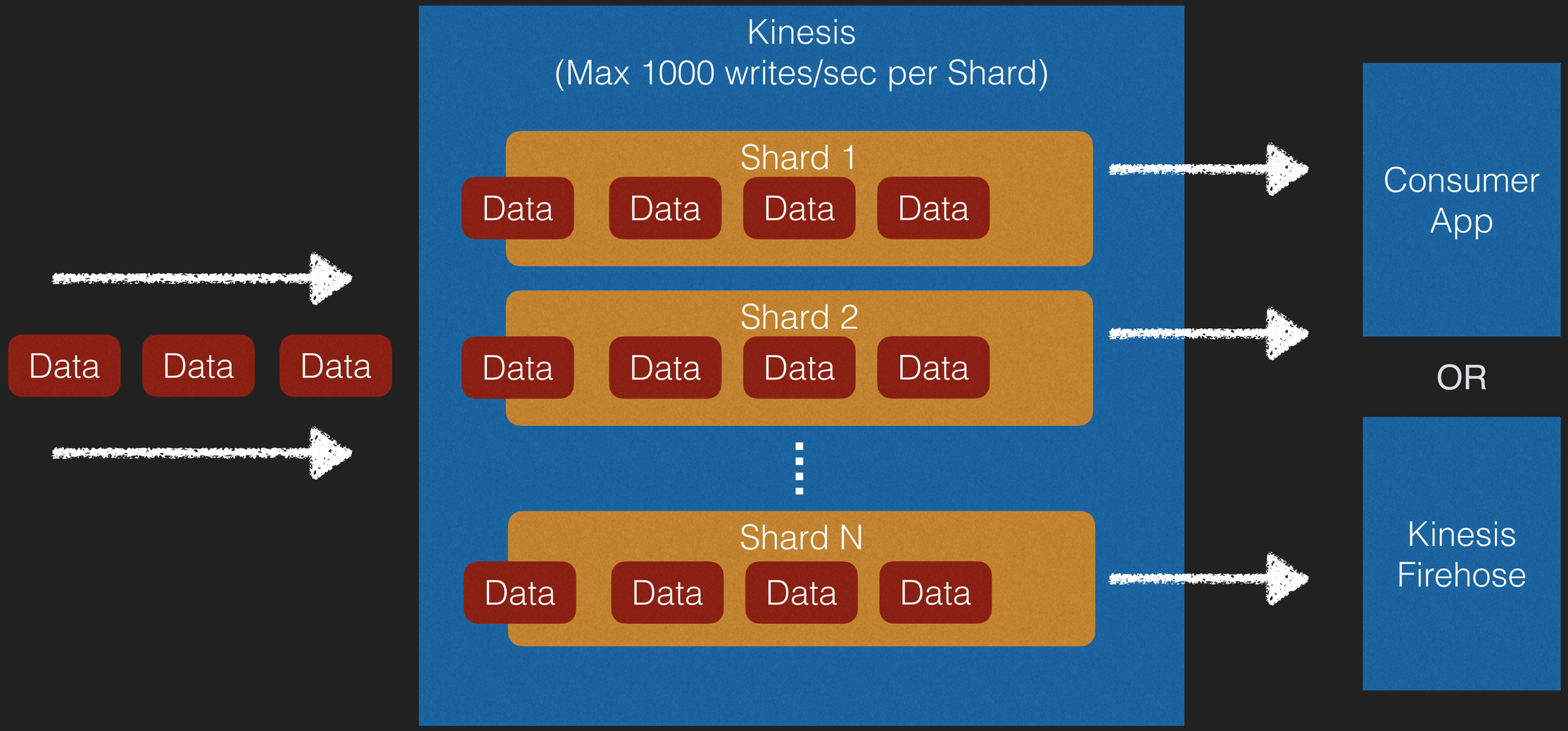


Kinesis Firehose



Kinesis Data Analytics

Kinesis Data Stream Key Concepts



Kinesis Data Stream Key Concepts

A CLOUD GURU

Kinesis

Shard 1

Partition Key: **e55fab…**
Sequence: **1**
Data: <some fancy data>

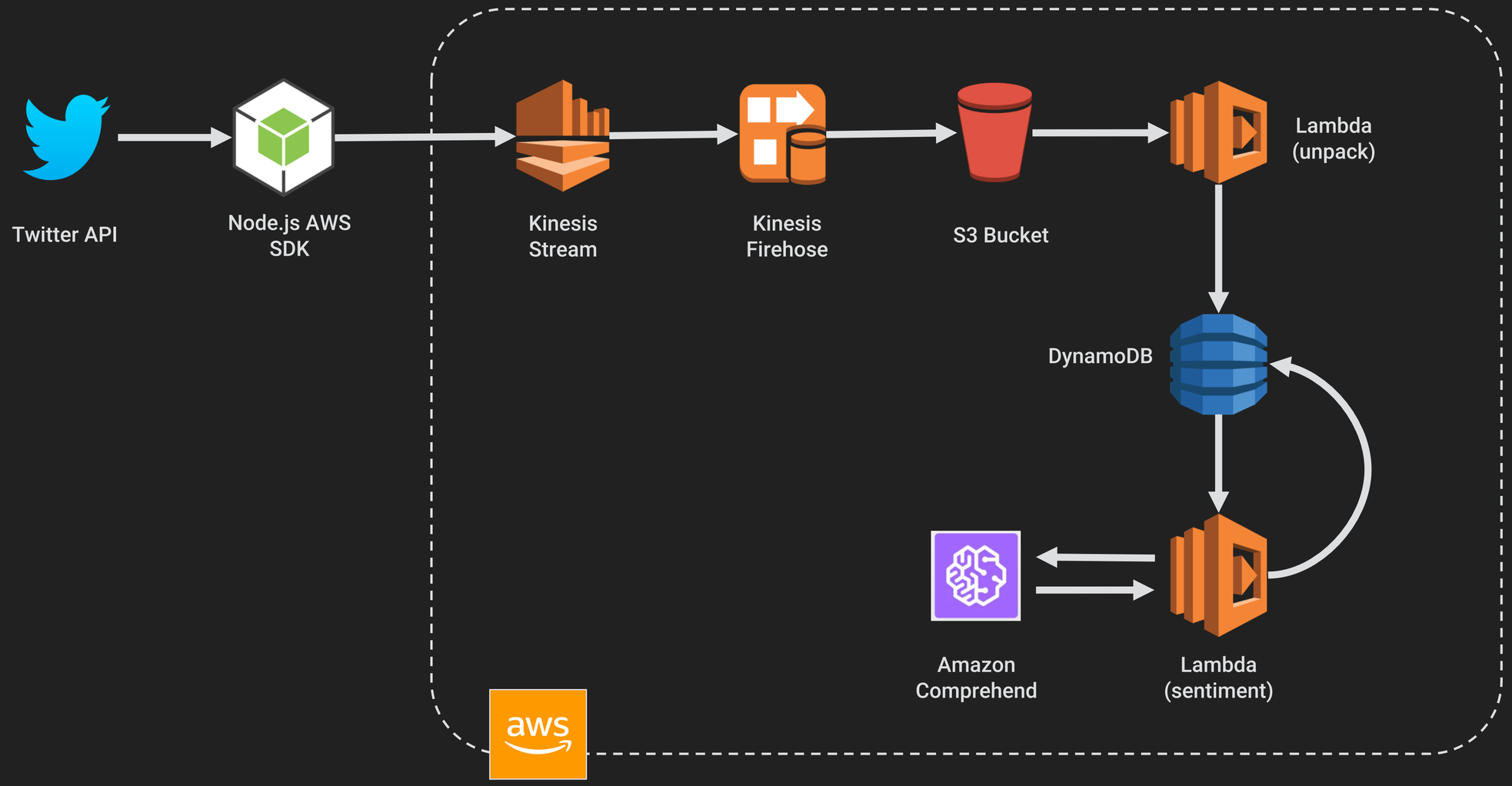
Partition Key: **e55fab…**
Sequence: **2**
Data: <some fancy data>

Shard 2

Partition Key: **45ab15…**
Sequence: **1**
Data: <some fancy data>

Partition Key: **45ab15…**
Sequence: **2**
Data: <some fancy data>

Kinesis Stream and Firehose



Scalability and Elasticity

DynamoDB Scaling

DynamoDB Scaling

A CLOUD GURU

Throughput

- Read Capacity Units
- Write Capacity Units

Size

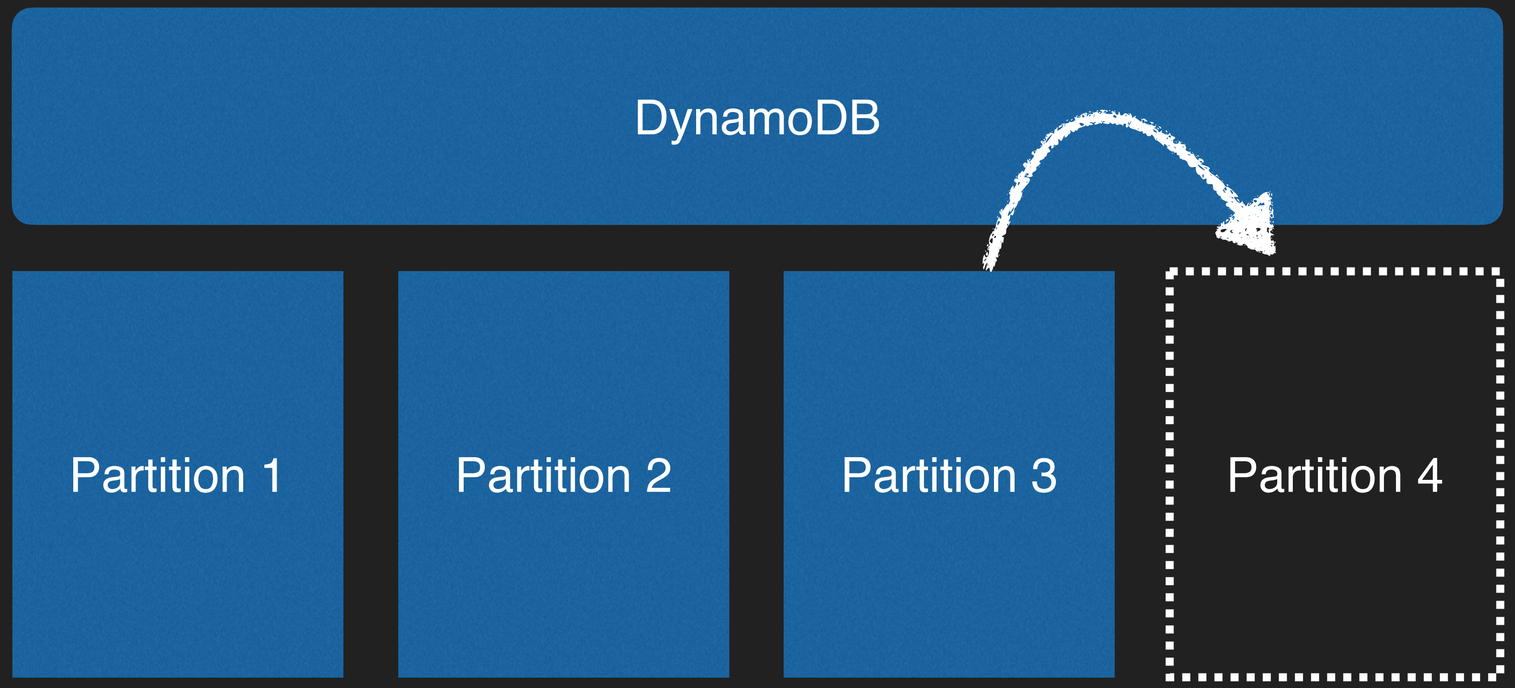
- Max item size is 400KB

DynamoDB Terminology

A CLOUD GURU

Term	What
Partition	A physical space where DynamoDB data is stored.
Partition Key	A unique identifier for each record; sometimes called a Hash Key.
Sort Key	In combination with a partition key, optional second part of a composite key that defines storage order; Sometimes called a Range Key

DynamoDB Under the Hood



DynamoDB Under the Hood

A CLOUD GURU

Partition Calculations	
By Capacity	(Total RCU / 3000) + (Total WCU / 1000)
By Size	Total Size / 10 GB
Total Partitions	Round Up for the MAX(By Capacity, By Size)

DynamoDB Under the Hood

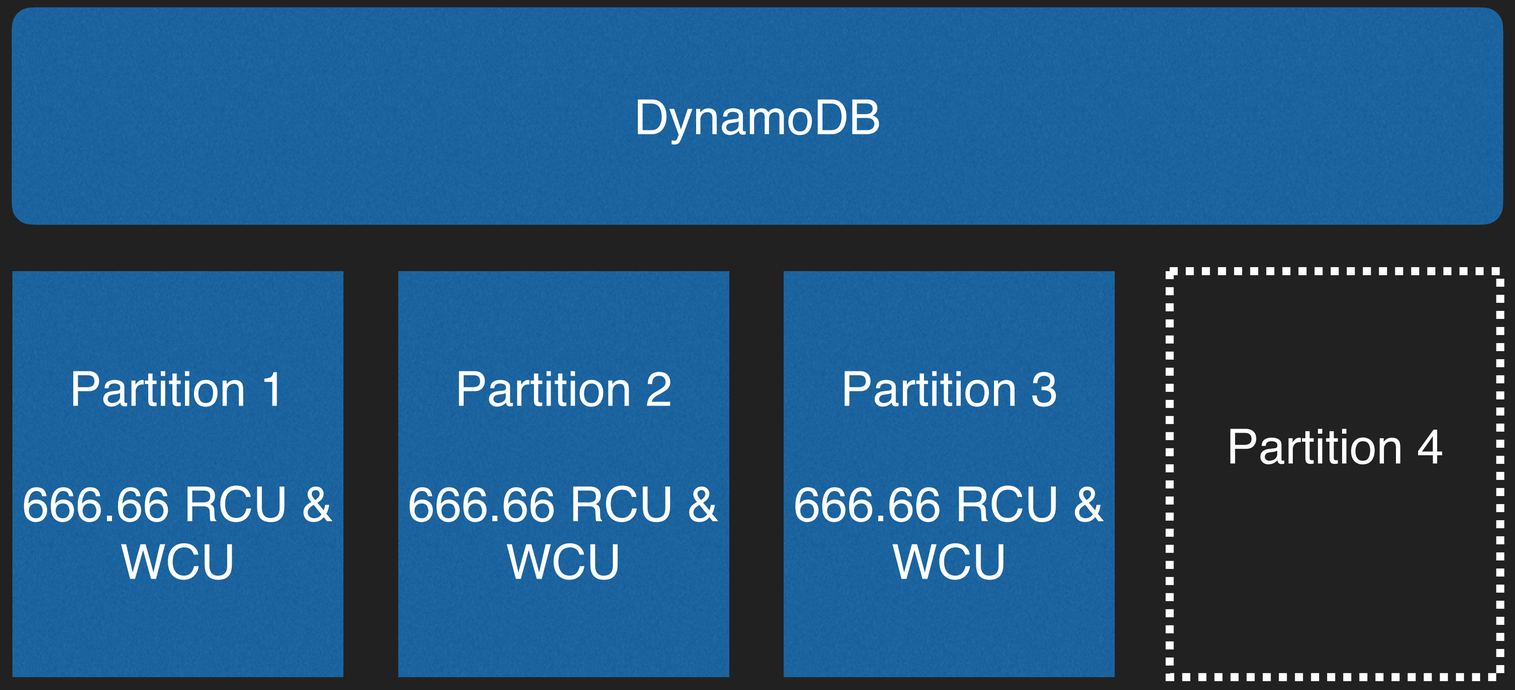
A CLOUD GURU

10GB Table, 2000 RCU and WCU

Partition Calculations	
By Capacity	(2000 RCU / 3000) + (2000 WCU / 1000) = 2.66
By Size	10 GB / 10 GB = 1
Total Partitions	MAX(2.66, 1) = 2.66 Round Up = 3 Partitions

DynamoDB Under the Hood

A CLOUD GURU



DynamoDB Under the Hood

A CLOUD GURU

DynamoDB

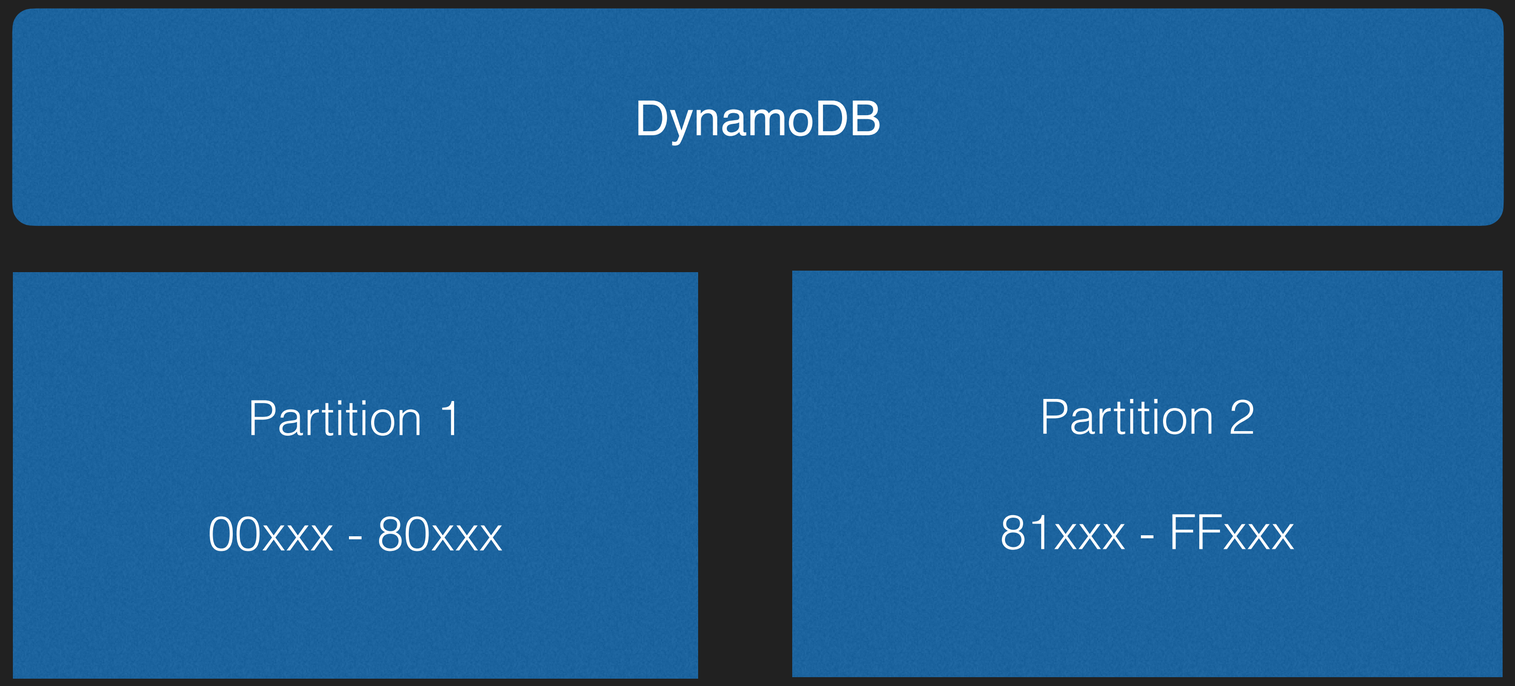
Partition 1

Hash 00xxx - FFxxx

DynamoDB Under the Hood

A CLOUD GURU

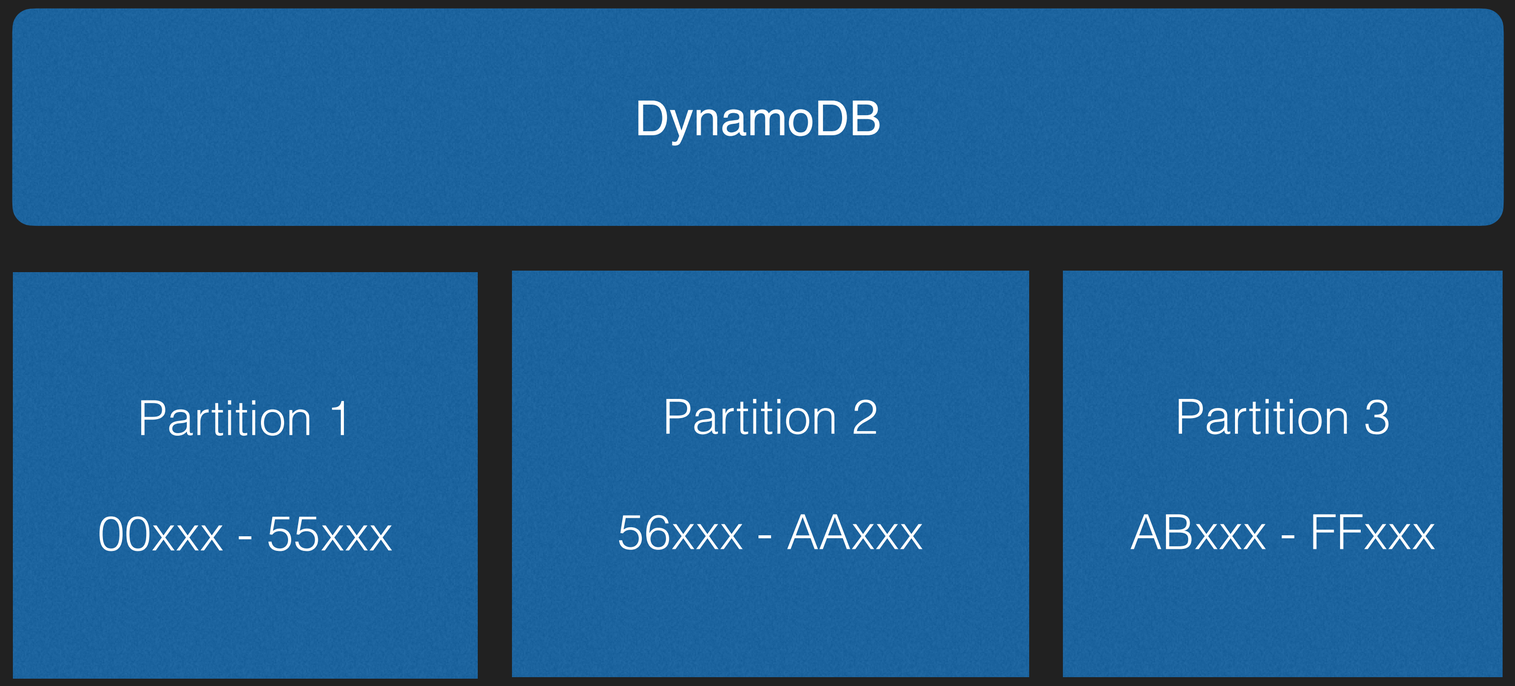
Increase RCU or WCU Provisioning or reach 10GB



DynamoDB Under the Hood

A CLOUD GURU

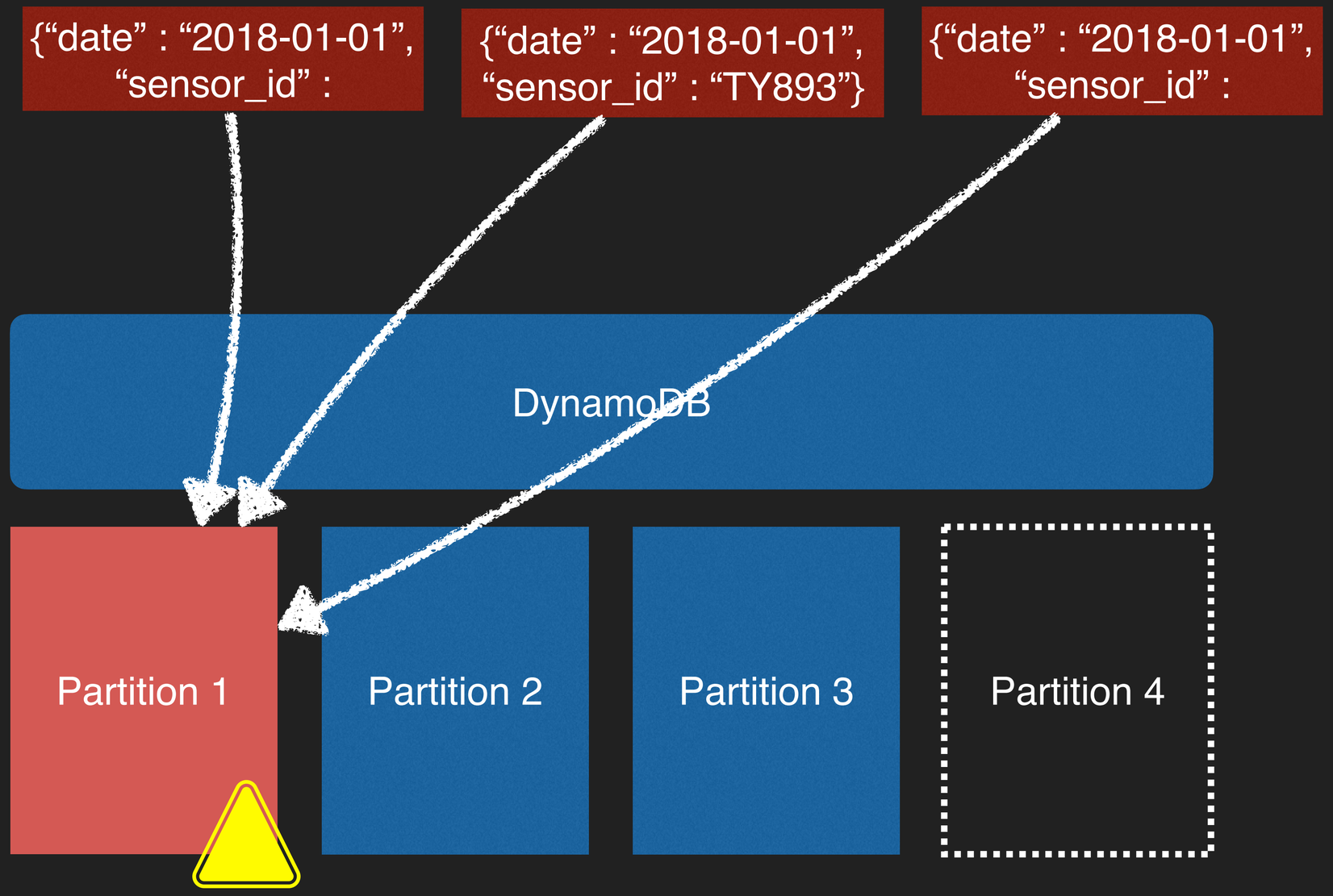
Increase RCU or WCU Provisioning or reach 20GB



DynamoDB Under the Hood

A CLOUD GURU

Partition Key =
date
Sort Key =
sensor_id

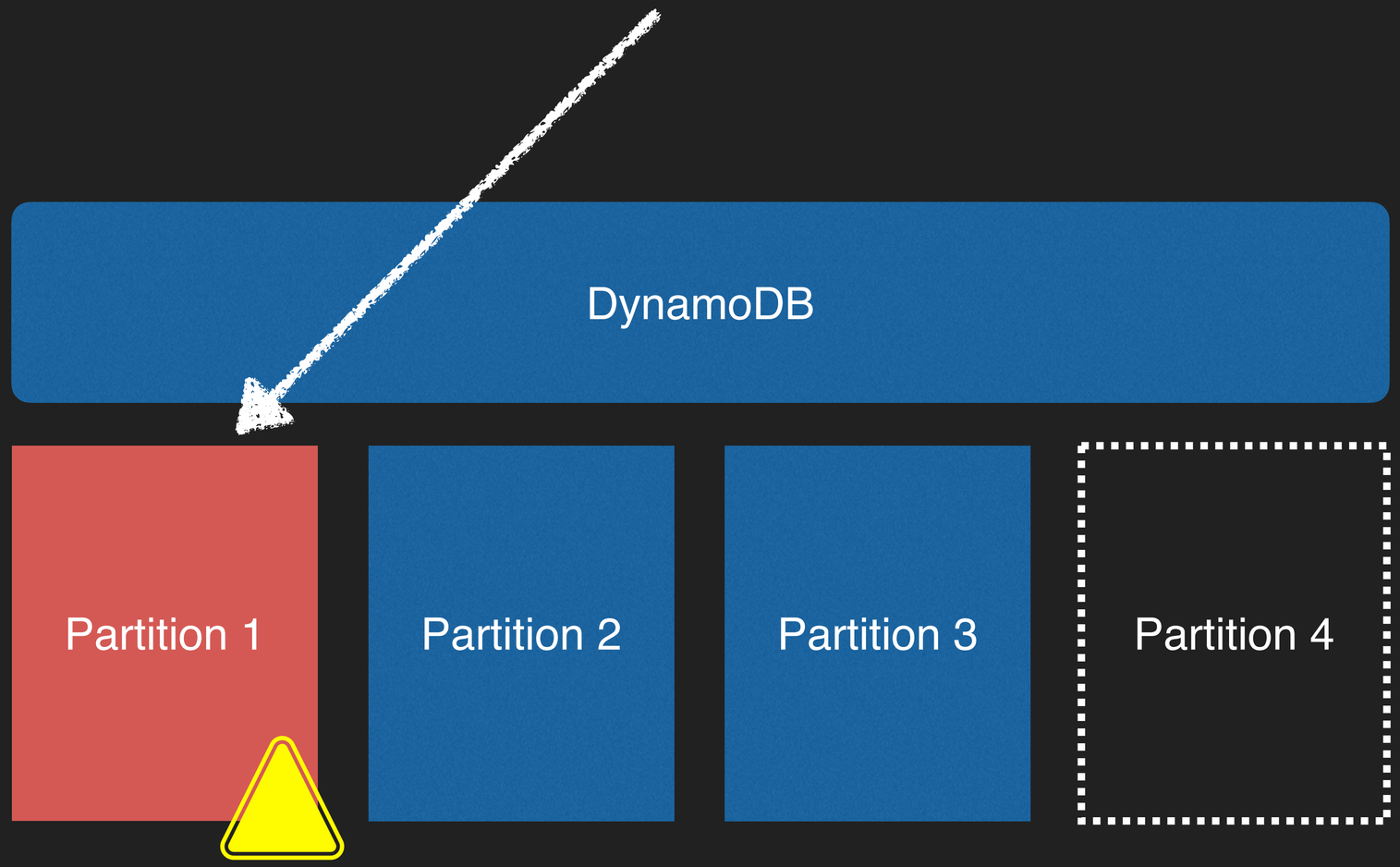


DynamoDB Under the Hood

A CLOUD GURU

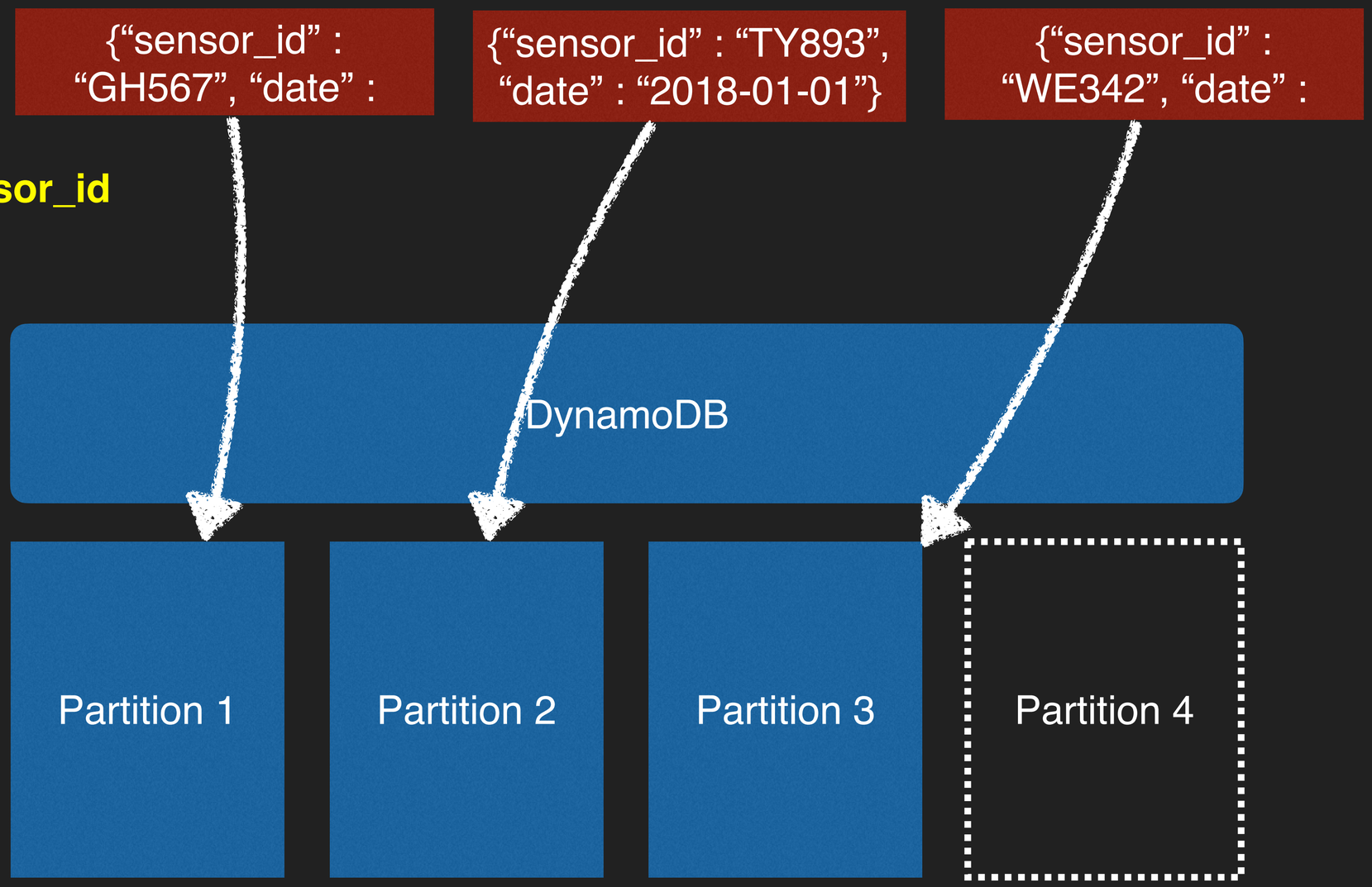
Partition Key =
date
Sort Key =
sensor_id

"Give me all the sensor readings for 2018-01-01."



DynamoDB Under the Hood

A CLOUD GURU

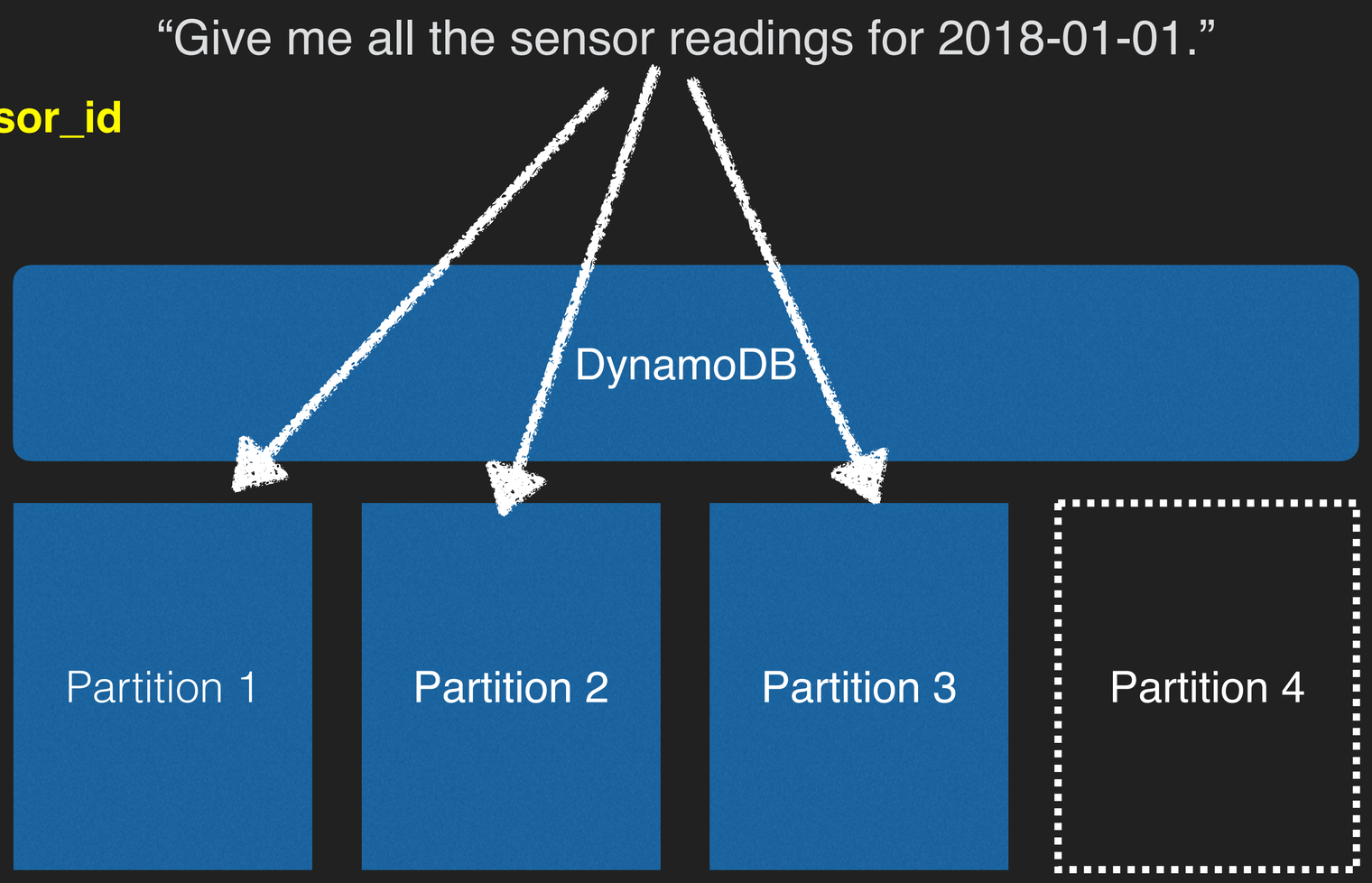


Partition Key = sensor_id
Sort Key = date

DynamoDB Under the Hood

A CLOUD GURU

Partition Key = sensor_id
Sort Key = date



Auto Scaling for DynamoDB

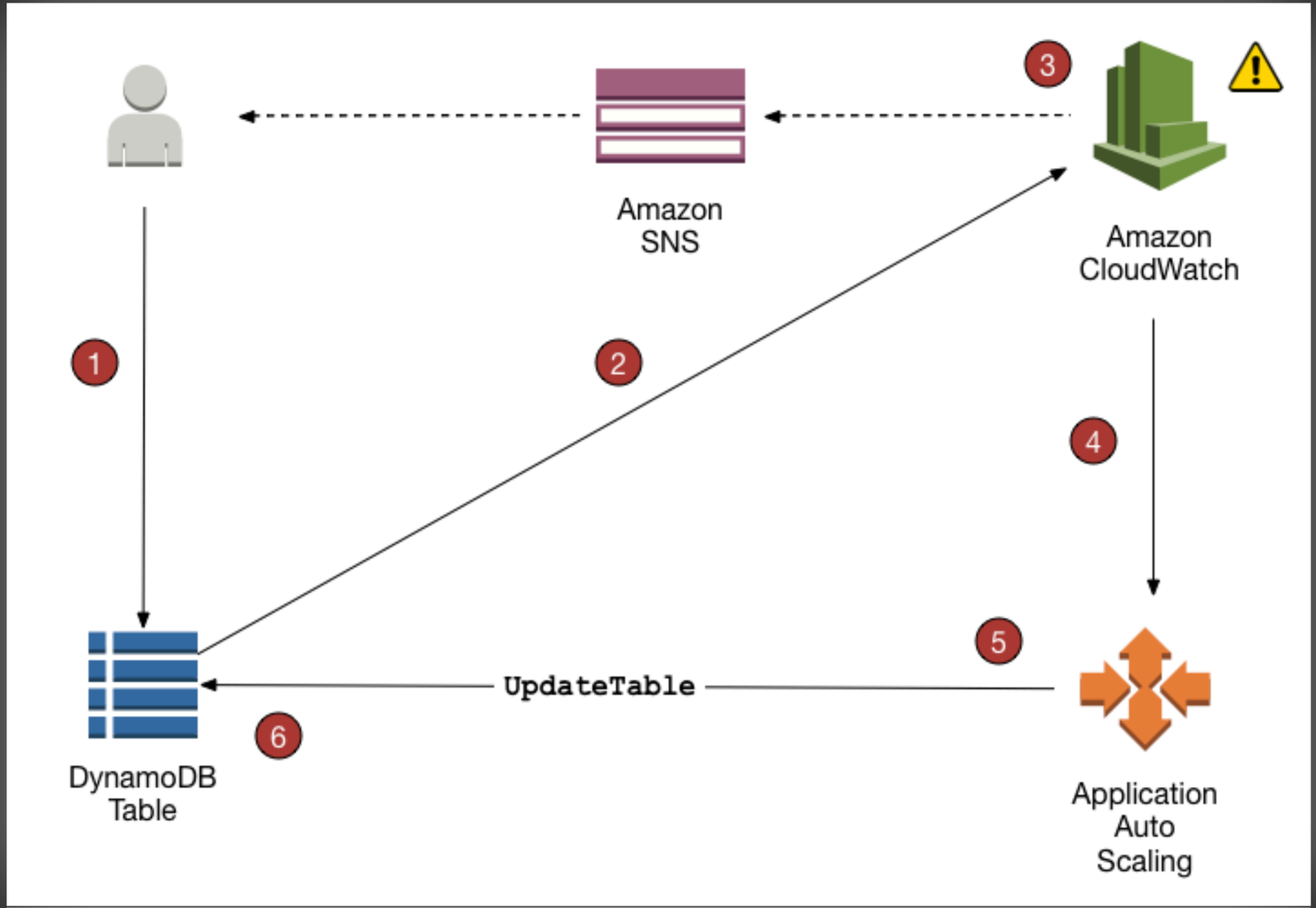
A CLOUD GURU

Auto Scaling

	Read capacity		Write capacity	
			Same settings as read	
Target utilization	50	%	50	%
Minimum provisioned capacity	5	units	5	units
Maximum provisioned capacity	1000	units	1000	units
	Apply same settings to global secondary indexes		Apply same settings to global secondary indexes	

Auto Scaling for DynamoDB

A CLOUD GURU

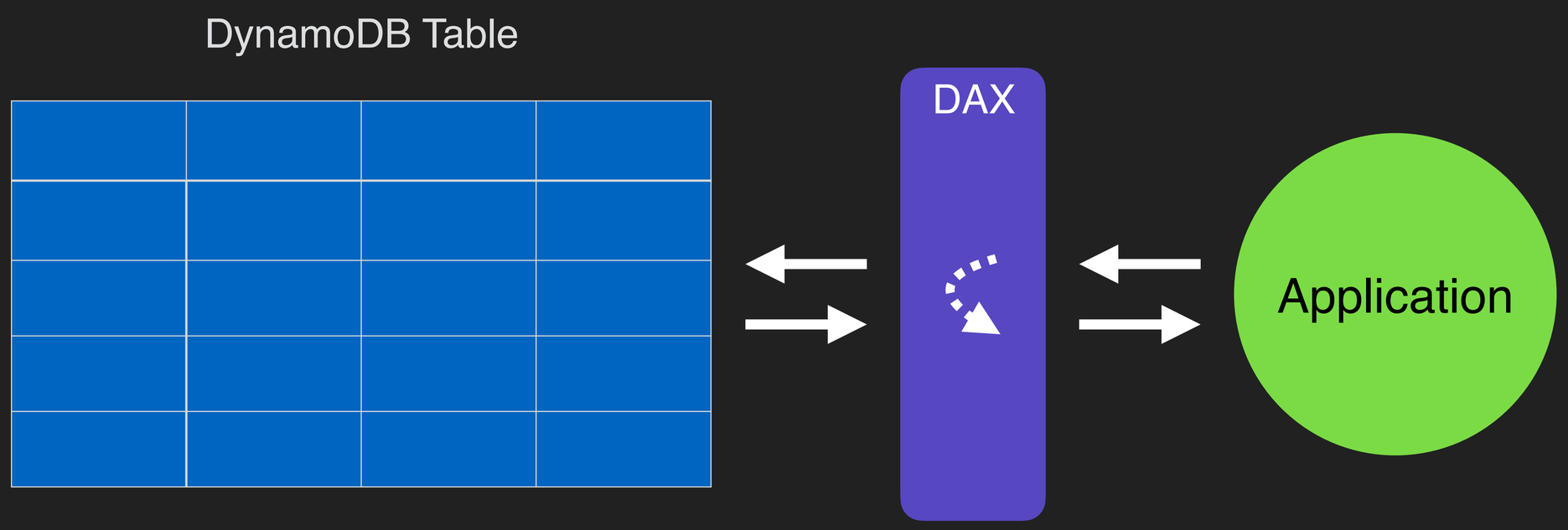


On-Demand Scaling for DynamoDB

A CLOUD GURU

- Alternative to Auto-Scaling
- Useful if you can’t deal with scaling lag or truly have no idea of the anticipated capacity requirements
- Instantly allocates capacity as needed with no concept of provisioned capacity
- Costs more than traditional provisioning and auto-scaling

DynamoDB Accelorator (DAX)



DAX Use Cases

A CLOUD GURU

Good Uses of DAX

- Require fastest possible reads such as live auctions or securities trading
- Read-intense scenarios where you want to offload the reads from DynamoDB
- Repeated reads against a large set of DynamoDB data

Bad Uses of DAX

- Write-intensive applications that don’t have many reads
- Applications where you use client caching methods

Scalability and Elasticity

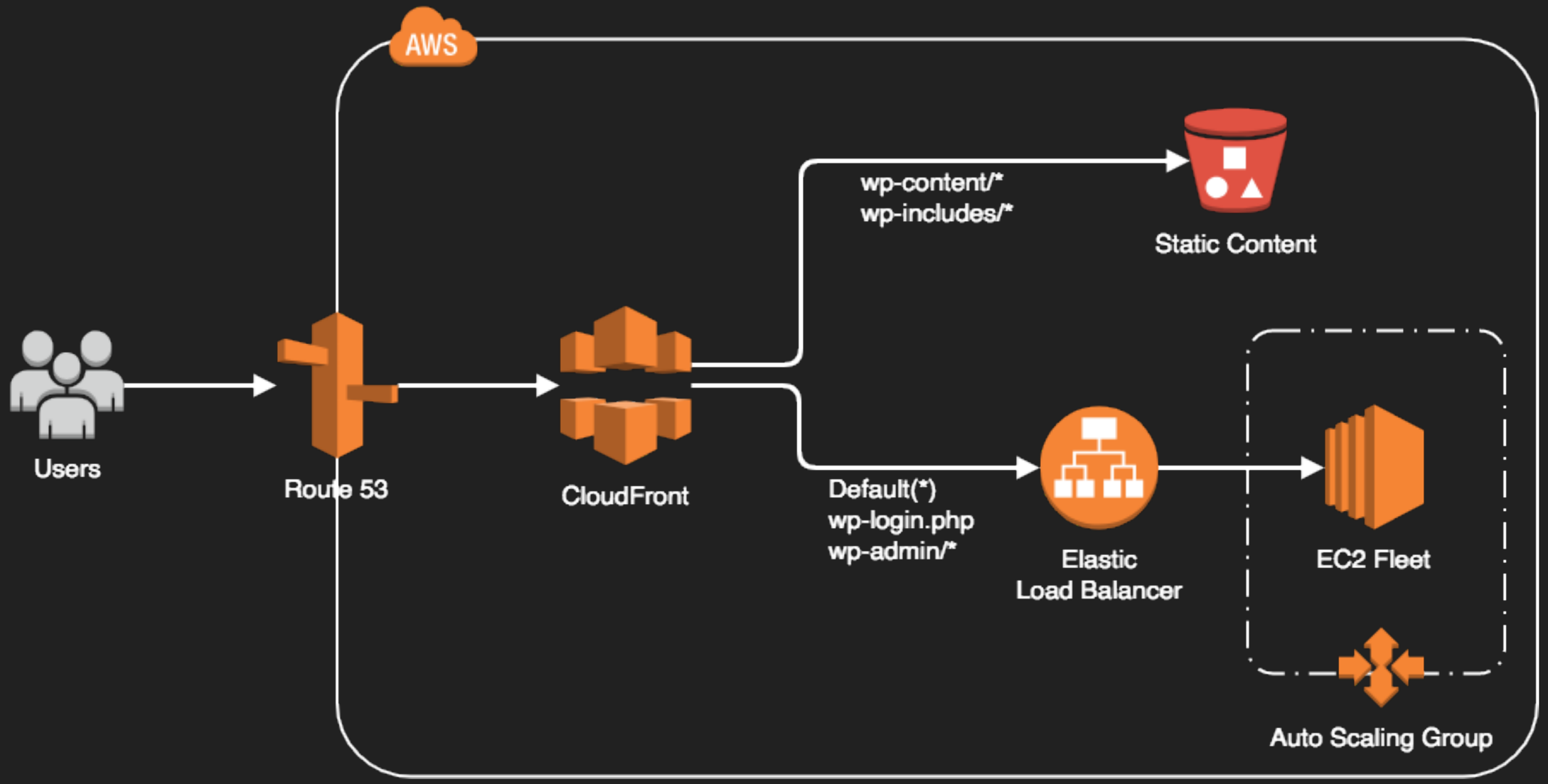
CloudFront Part 2

CloudFront Part Deux

A CLOUD GURU

- Can deliver content to your users faster by caching static and dynamic content at edge locations.
- Dynamic content delivery is achieved using HTTP cookies forwarded from your origin.
- Supports Adobe Flash Media Server's RTMP protocol but you have to choose RTMP delivery method.
- Web distributions also support media streaming and live streaming but use HTTP or HTTPS.
- Origins can be S3, EC2, ELB or another web server.
- Multiple origins can be configured.
- Use Behaviors to configure serving up origin content based on URL paths.

Origins and Behaviors



Invalidation Requests

A CLOUD GURU

1. Simply delete the file from the origin and wait for the TTL to expire.
2. Use the AWS Console to request invalidataion for all content or a specific path such as /images/*
3. Use the CloudFront API to submit an invalidation request
4. Use third-party tools to perform CloudFront invalidation (CloudBerry, Ylastic, CDN Planet CloudFront Purge Tool)

Zone Apex Support

A CLOUD GURU

Create Record Set

Name: scottpletcher.guru.

Type: A – IPv4 address

Alias: Yes No

Alias Target: d2hph8qaljbyi5.cloudfront.net

Alias Hosted Zone ID: Z2FDTNDATAQYW2

You can also type the domain name for the resource. Examples:

- CloudFront distribution domain name: d111111abcdef8.cloudfront.net
- Elastic Beanstalk environment CNAME: example.elasticbeanstalk.com
- ELB load balancer DNS name: example-1.us-east-1.elb.amazonaws.com
- S3 website endpoint: s3-website.us-east-2.amazonaws.com
- Resource record set in this hosted zone: www.example.com

Learn More

```
xray:~ scott$ nslookup scottpletcher.guru
Server:         10.2.2.1
Address:        10.2.2.1#53

Non-authoritative answer:
Name:   scottpletcher.guru
Address: 52.85.236.165
Name:   scottpletcher.guru
Address: 52.85.236.39
Name:   scottpletcher.guru
Address: 52.85.236.190
Name:   scottpletcher.guru
Address: 52.85.236.159
```

Geo-Restrictions

A CLOUD GURU

aws Services Resource Groups ACG Global Support

Edit Geo-Restrictions

Geo-Restriction Settings

Enable Geo-Restriction
- (•) Yes
- ( ) No

Restriction Type
- (•) Whitelist
- ( ) Blacklist

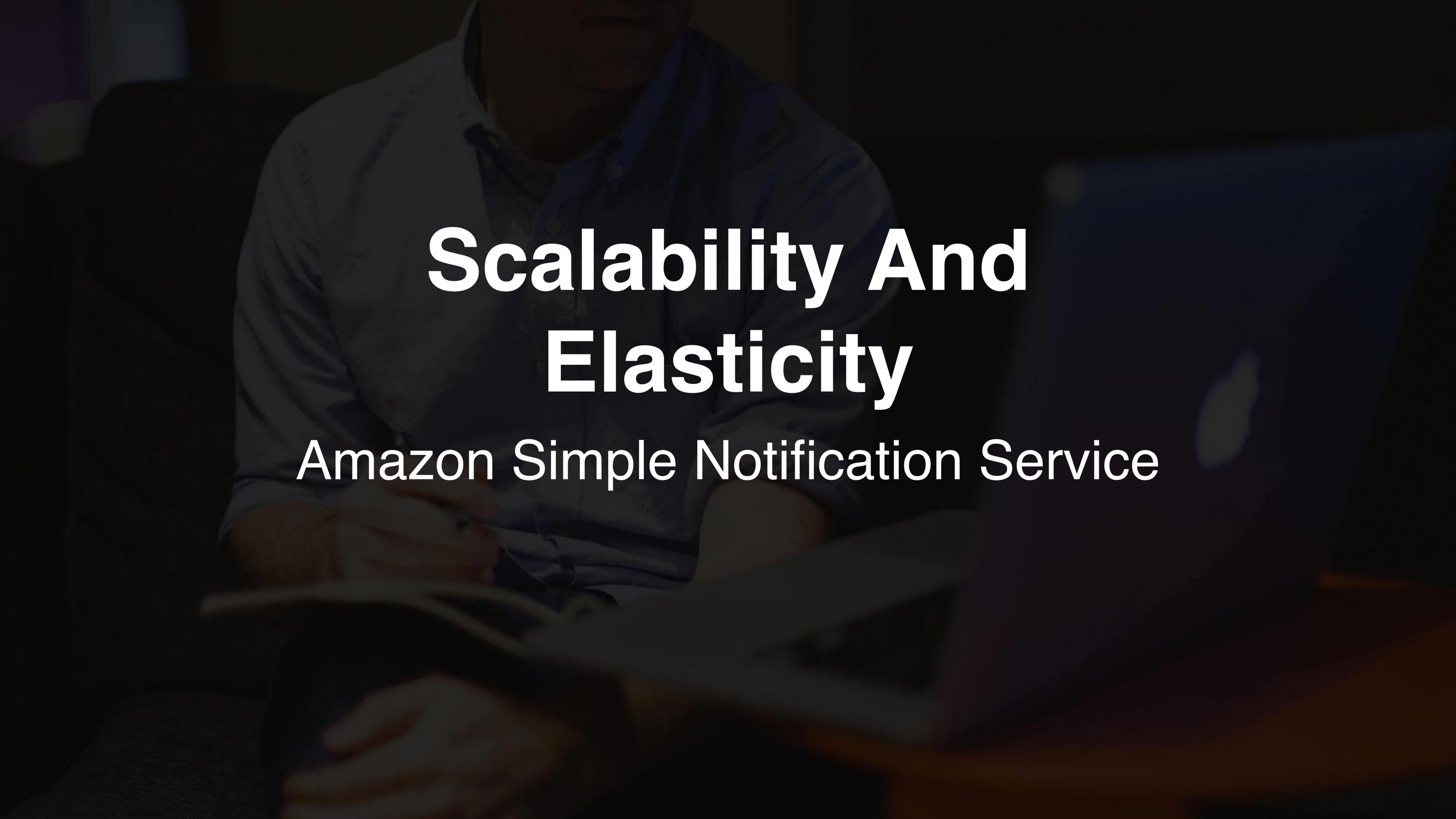
Countries

AF -- AFGHANISTAN
AX -- ALAND ISLANDS
AL -- ALBANIA
DZ -- ALGERIA
AS -- AMERICAN SAMOA
AD -- ANDORRA

Add >>

<< Remove

Cancel Yes, Edit



Scalability And Elasticity

Amazon Simple Notification Service

Amazon Simple Notification Service

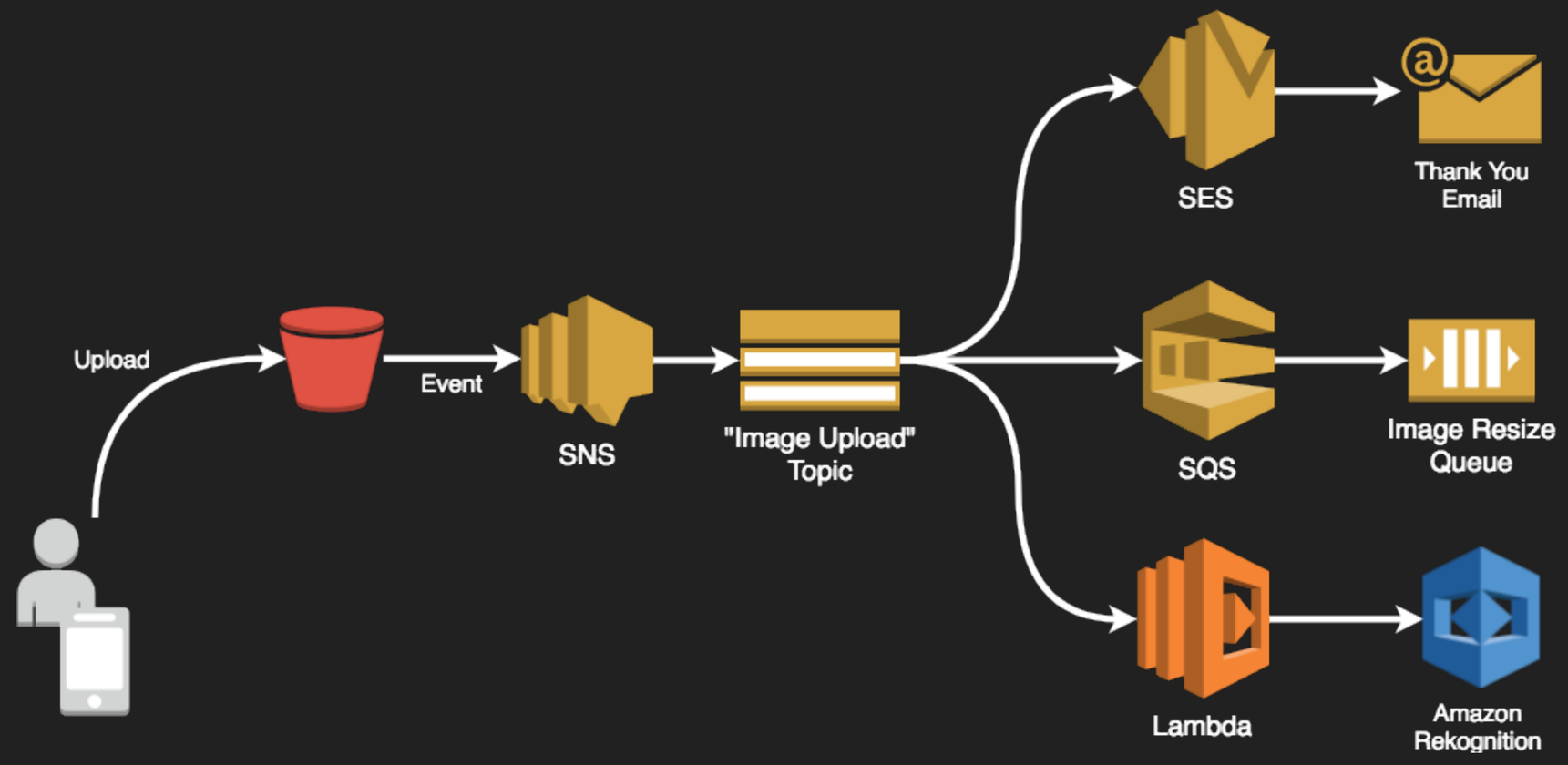
A CLOUD GURU

- Enables a Publish/Subscribe design pattern.
- Topics = A channel for publishing a notification
- Subscription = Configuring and endpoint to receive messages published on the topic
- Endpoint protocols include HTTP(S), Email, SMS, SQS, Amazon Device Messaging (push notifications) and Lambda

Amazon SNS

Fan Out Architecture

A CLOUD GURU



Amazon
SNS

Scalability and Elasticity

Amazon SQS

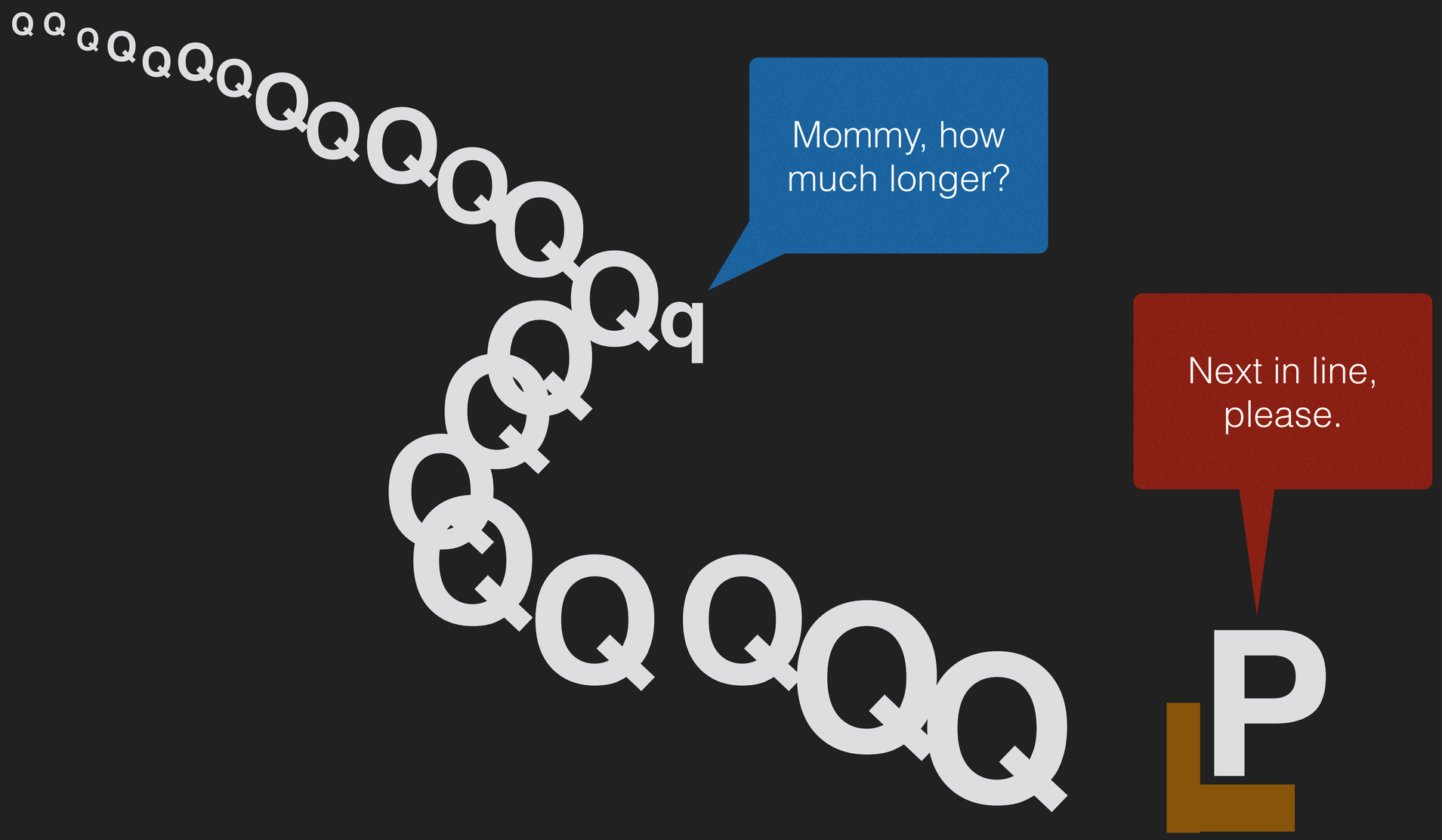
Amazon SQS

A CLOUD GURU

Q

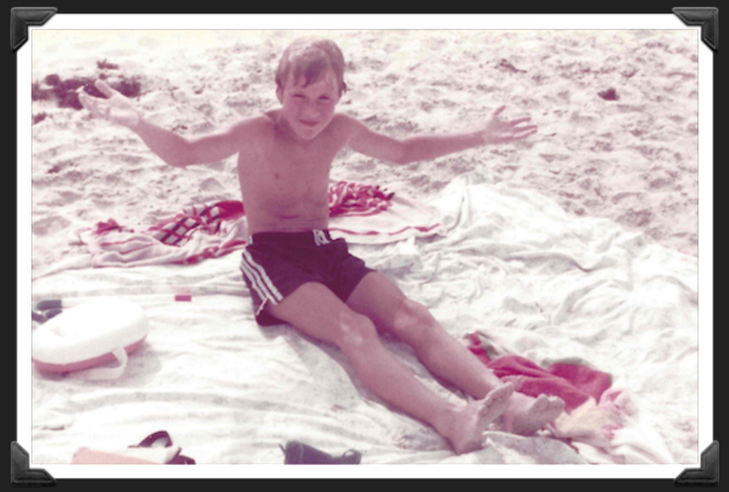
Amazon SQS

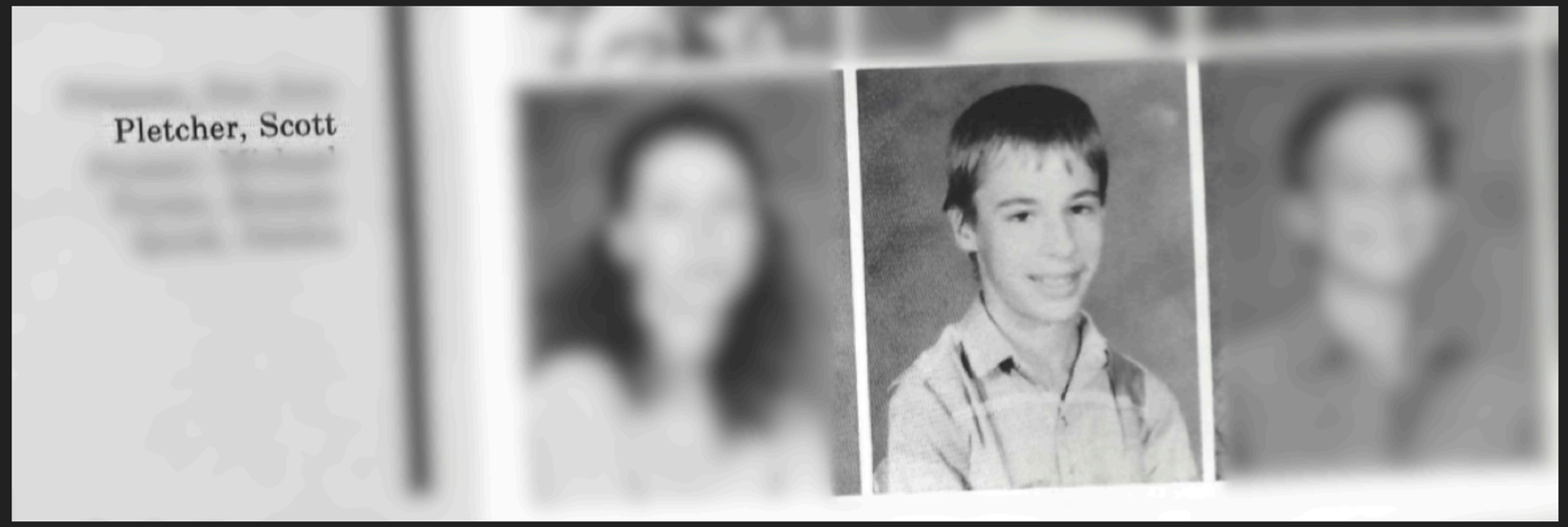
A CLOUD GURU



Honing My Craft

A CLOUD GURU



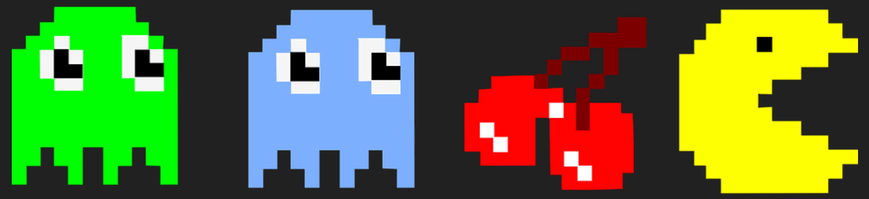


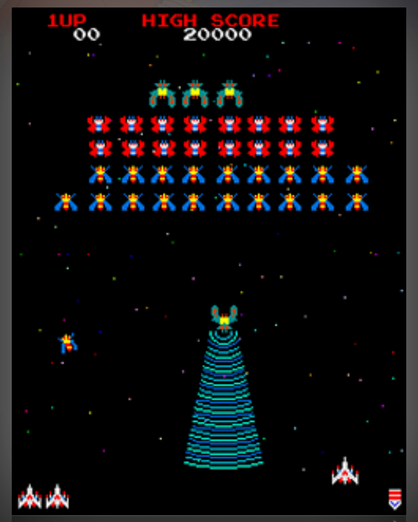
My Bike

A CLOUD GURU

The Arcade

A CLOUD GURU



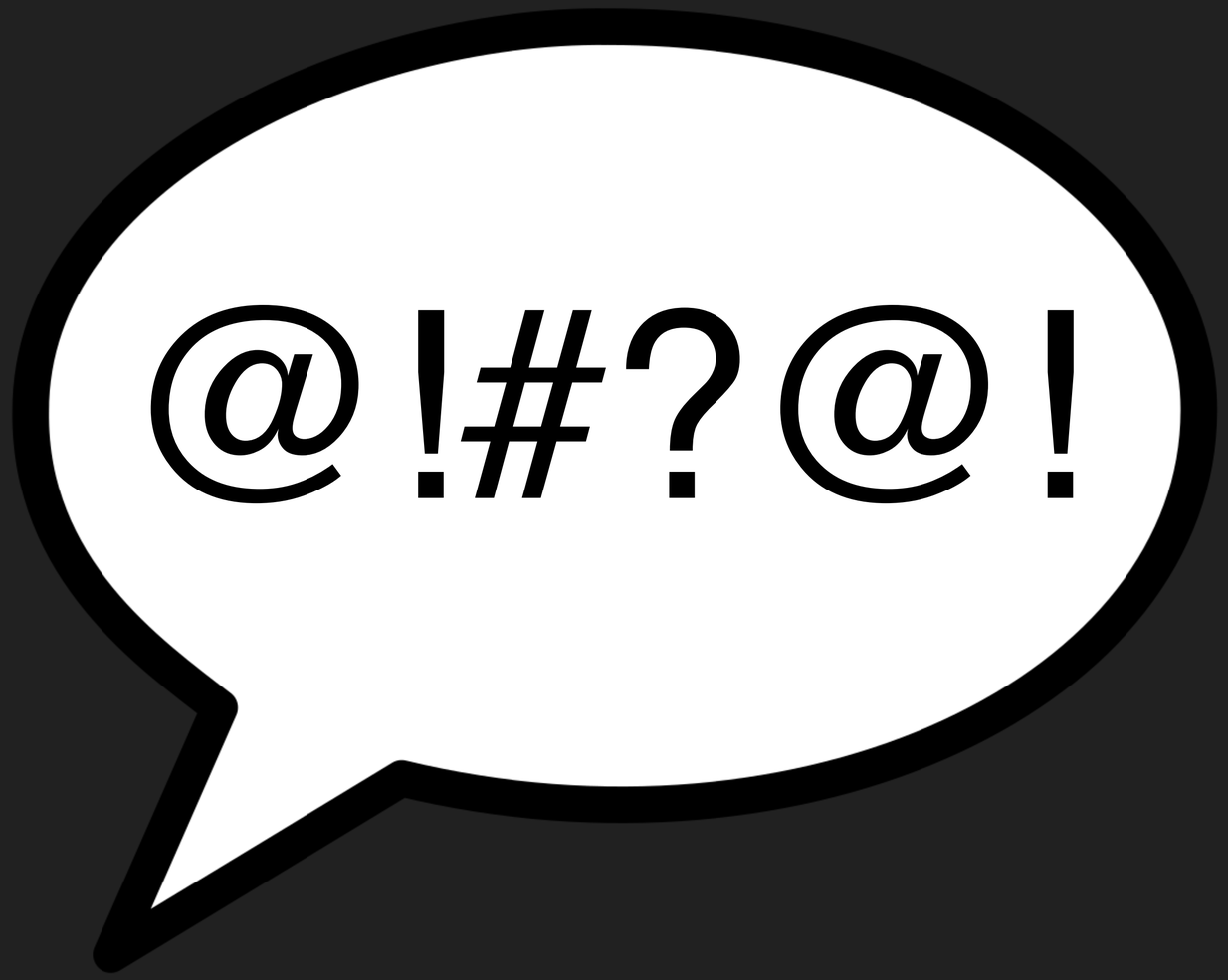


Still Has It…

A CLOUD GURU

No Substitute

A CLOUD GURU



Amazon SQS

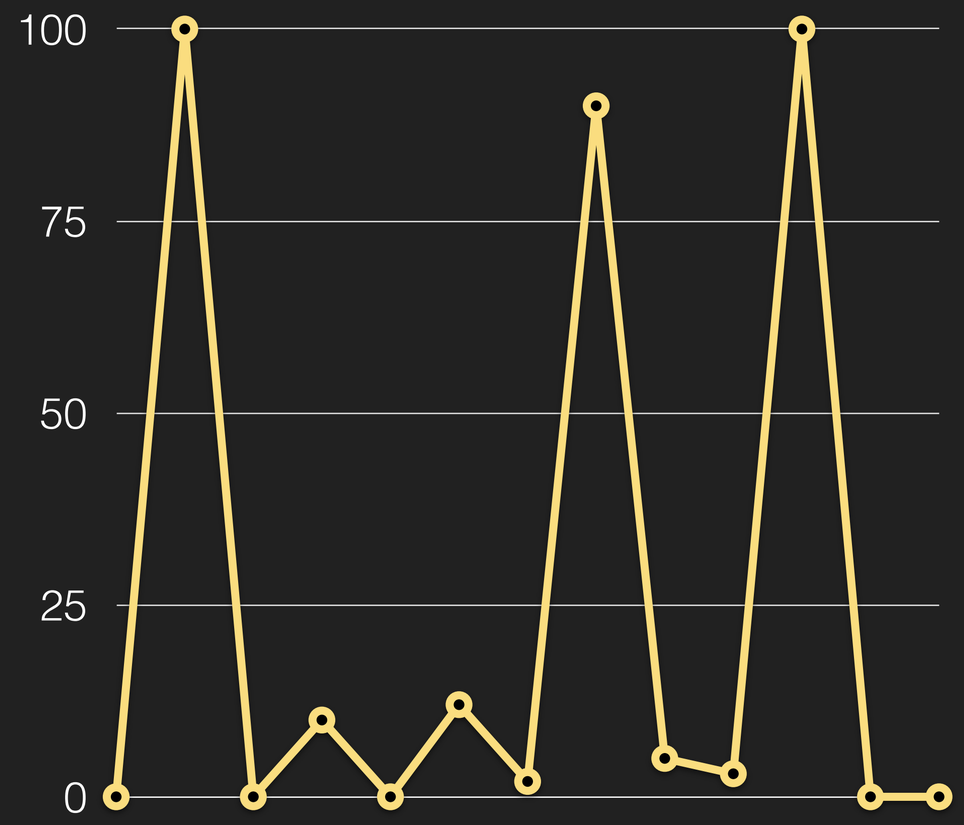
A CLOUD GURU

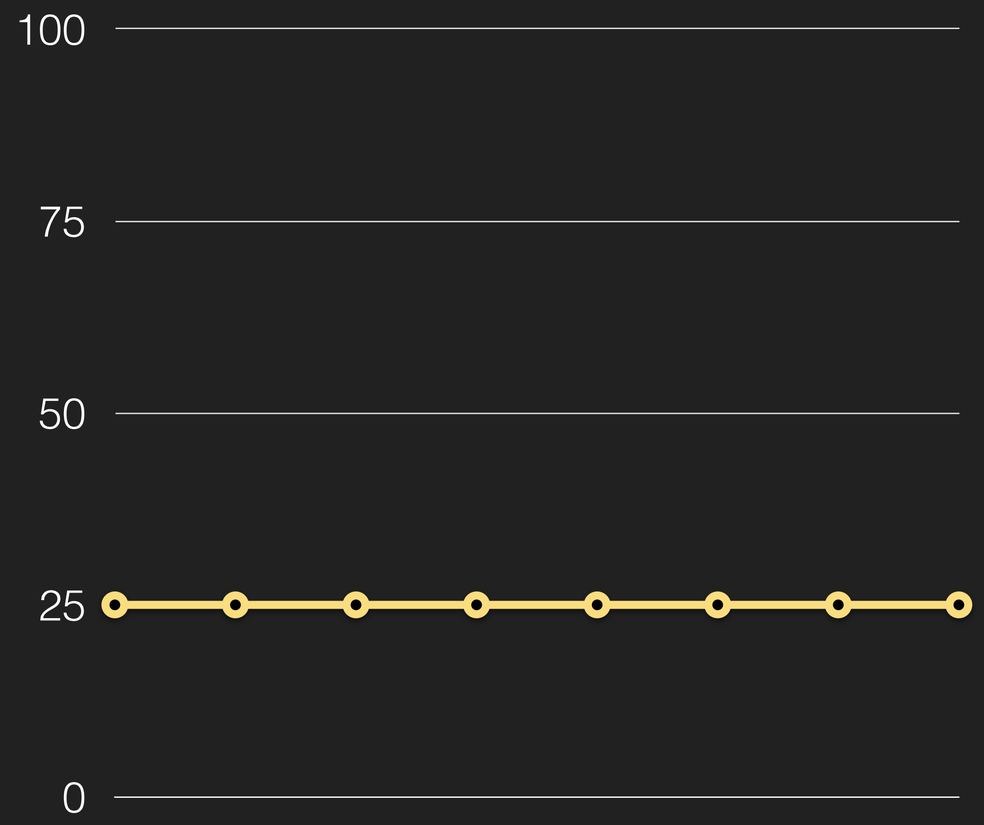
- Reliable, highly-scalable, hosted message queuing service.
- Available integration with KMS for encrypted messaging.
- Transient storage—default 4 days, max 14 days.
- Optionally supports First-In First-Out queue ordering.
- Maximum message size of 256KB but the using a special Java SQS SDK, you can have messages as large as 2GB.

Amazon SQS

Loosely Coupled

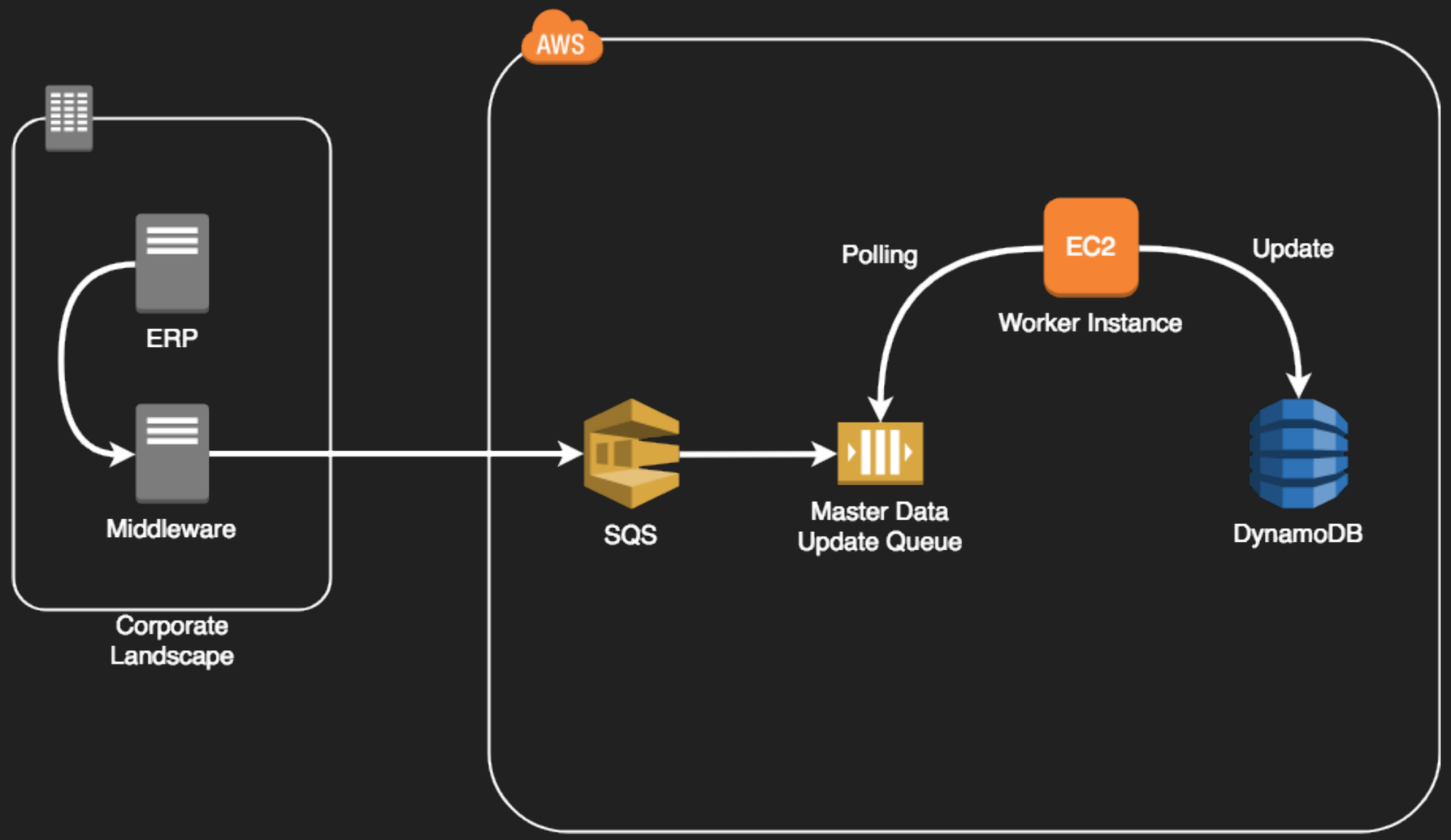
A CLOUD GURU





Amazon SQS

Loosely Coupled Architecture



Amazon SQS

Standard Queue

A CLOUD GURU

Message Stream

```
{   "update":"1",
    "orderStatus":"received"
}
```

```
{   "update":"2",
    "orderStatus":"sent to warehouse"
}
```

```
{   "update":"3",
    "orderStatus":"picked"
}
```

```
{   "update":"4",
    "orderStatus":"packed"
}
```

```
{   "update":"5",
    "orderStatus":"shipped"
}
```

Standard Queue

```
{   "update":"1",
    "orderStatus":"received"
}
```

```
{   "update":"5",
    "orderStatus":"shipped"
}
```

```
{   "update":"4",
    "orderStatus":"packed"
}
```

```
{   "update":"2",
    "orderStatus":"sent to warehouse"
}
```

```
{   "update":"3",
    "orderStatus":"picked"
}
```

Order of Update into DynamoDB

FIFO Queue

A CLOUD GURU

Message Stream

```
{   "update":"1",
    "orderStatus":"received"
}
```

```
{   "update":"2",
    "orderStatus":"sent to warehouse"
}
```

```
{   "update":"3",
    "orderStatus":"picked"
}
```

```
{   "update":"4",
    "orderStatus":"packed"
}
```

```
{   "update":"5",
    "orderStatus":"shipped"
}
```

FIFO Queue

```
{   "update":"1",
    "orderStatus":"received"
}
```

```
{   "update":"2",
    "orderStatus":"sent to warehouse"
}
```

```
{   "update":"3",
    "orderStatus":"picked"
}
```

```
{   "update":"4",
    "orderStatus":"packed"
}
```

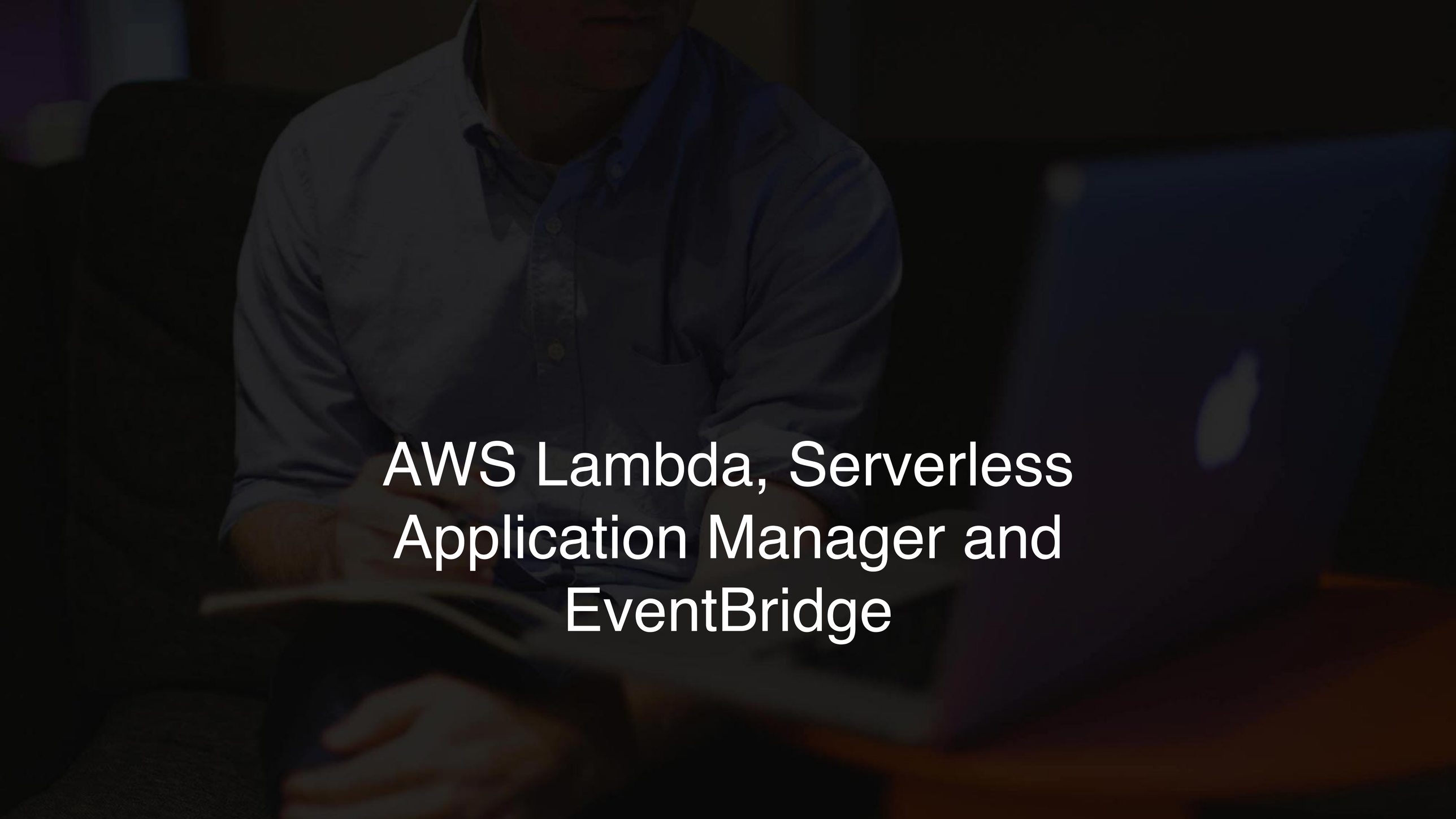
```
{   "update":"5",
    "orderStatus":"shipped"
}
```

Order of Update into DynamoDB

Amazon MQ

A CLOUD GURU

- Managed implementation of Apache ActiveMQ
- Fully managed and highly available within a region
- ActiveMQ API and support for JMS, NMS, MQTT, WebSocket
- Designed as a drop-in replacement for on-premise message brokers
- Use SQS if you creating a new application from scratch
- Use MQ if you want an easy low-hassle path to migrate from existing message brokers to AWS



AWS Lambda, Serverless Application Manager and EventBridge

AWS Lambda

A CLOUD GURU

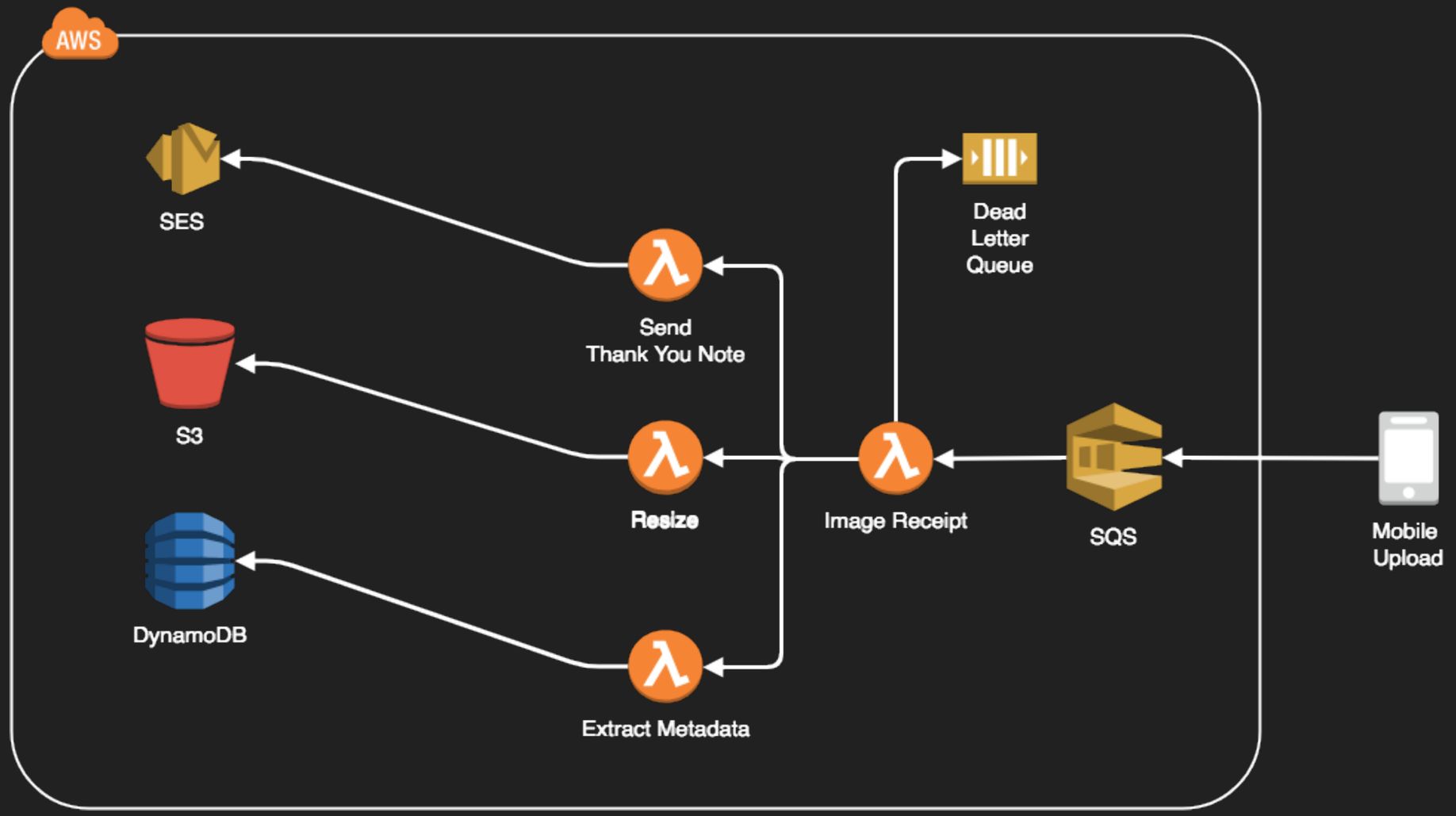
- Allows you to run code on-demand without the need for infrastructure.
- Supports Node.js, Python, Java, Go, and C#.
- Extremely useful option for creating serverless architectures.
- Code is stateless and executed on an event basis (SNS, SQS, S3, DynamoDB Streams, etc.).
- No fundamental limits to scaling a function since AWS dynamically allocates capacity in relation to events.

Lots of great in-depth content on Lambda and Serverless right here on ACG!

AWS Lambda

Lambda at Scale

A CLOUD GURU



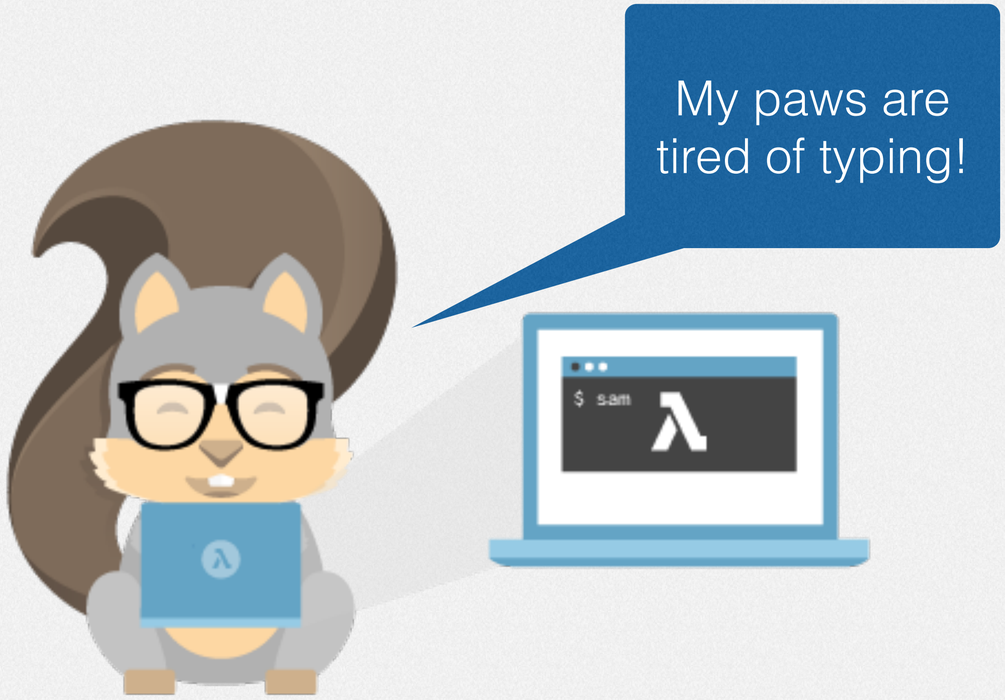
AWS Serverless Application Model

A CLOUD GURU

"Hi, I'm Sam the Squirrel. I left behind a cozy life in the trees to help AWS users build their server less applications effectively and more easily."

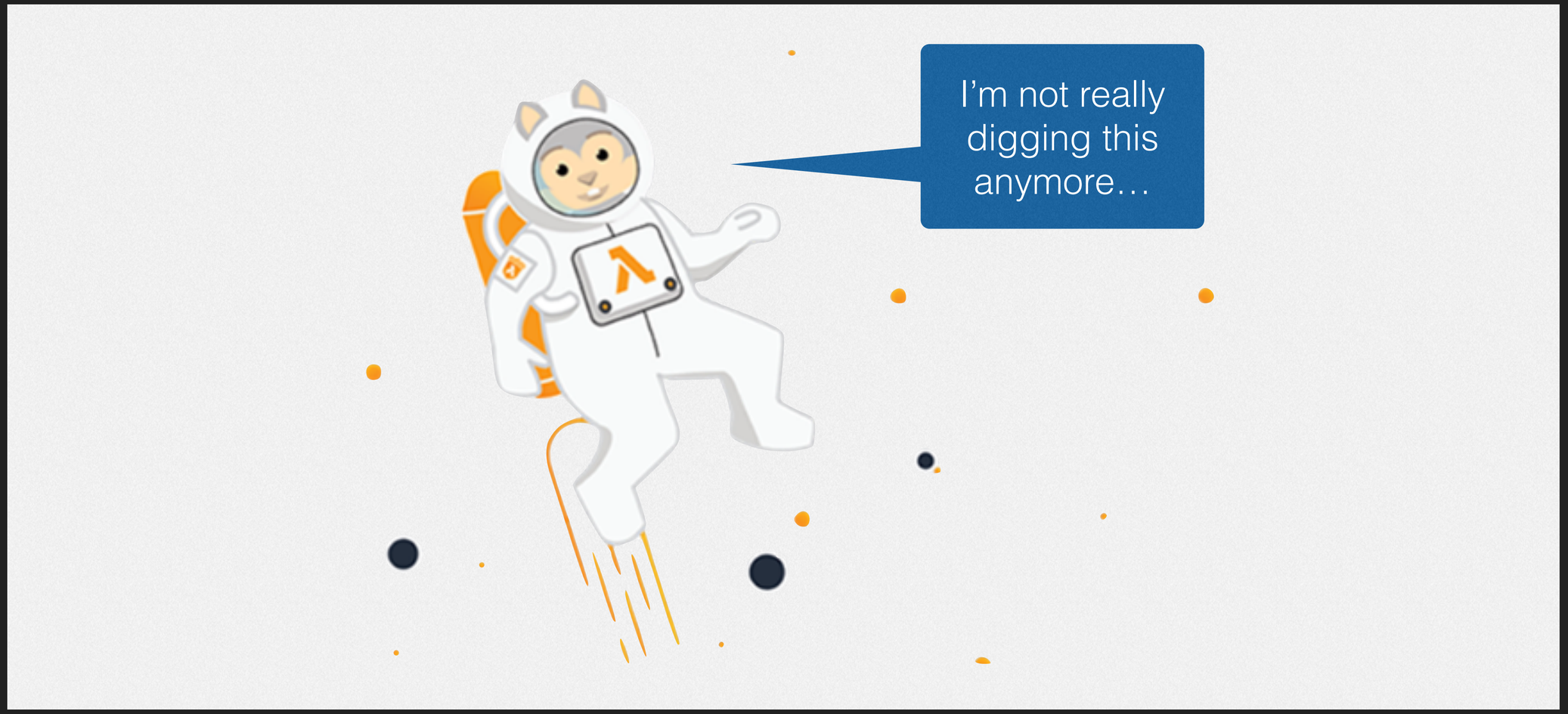
AWS Serverless Application Model

A CLOUD GURU



AWS Serverless Application Model

A CLOUD GURU

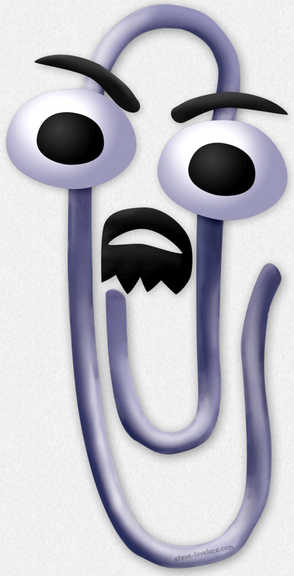


AWS Serverless Application Model

A CLOUD GURU

"I got out while I still could!

"Lemme tell ya kid…take it from me. Never let go of what makes you YOU!



AWS Serverless Application Model

A CLOUD GURU

- Open source framework for building serverless apps on AWS.
- Uses YAML as the configuration language.
- Includes AWS CLI-like functionality to create, deploy and update serverless apps using AWS services such as Lambda, DynamoDB and API Gateway.
- Enables local testing and debugging of apps using a Lambda-like emulator via Docker
- Extension of CloudFormation so you can use everything CloudFormation can provide by way of resources and functions.

AWS Serverless Application Model

A CLOUD GURU

Local System

AWS SAM CLI

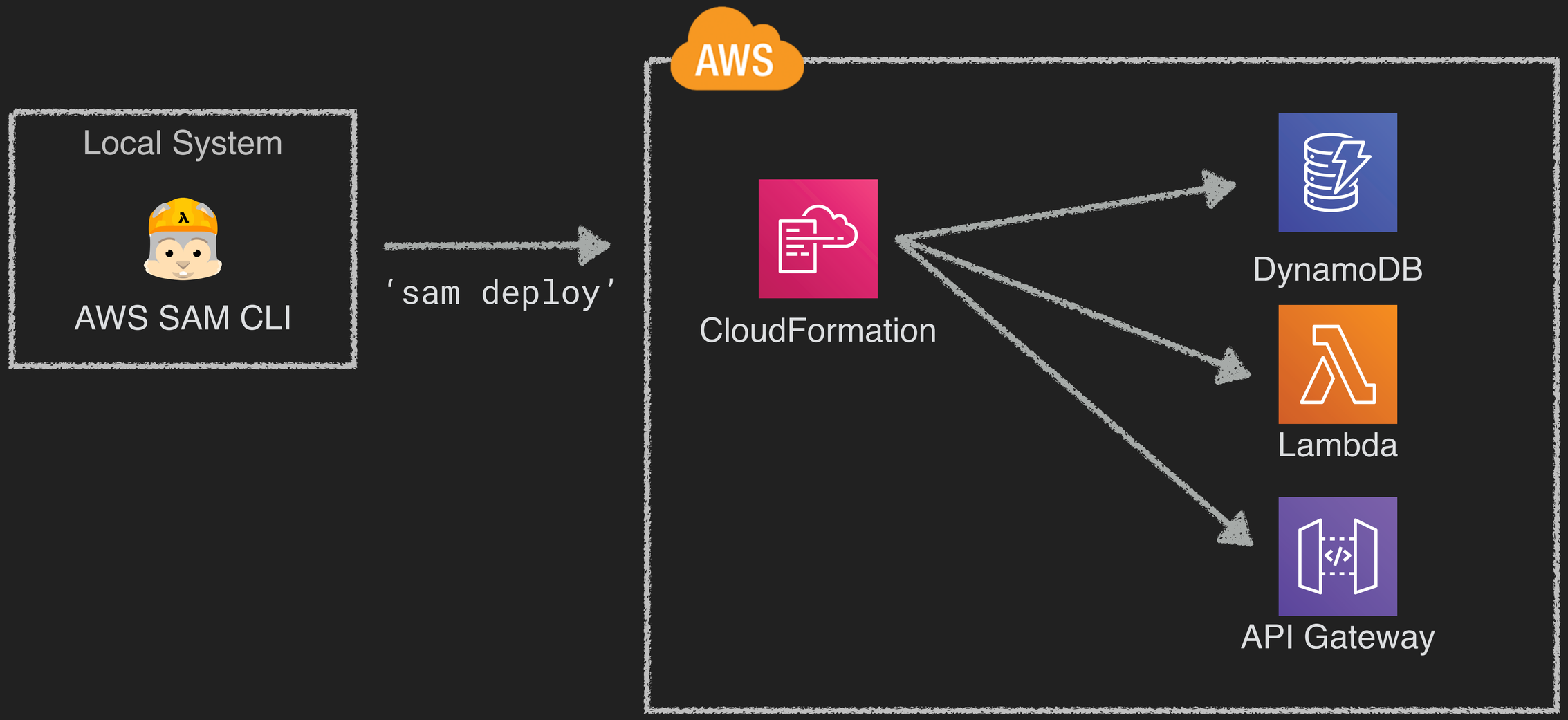
```yaml
AWSTemplateFormatVersion: '2010-09-09'
Transform: 'AWS::Serverless-2016-10-31'
Description: >-
  Uses ImageMagick to perform simple image processing operations, such as resizing.
Parameters:
    IdentityNameParameter:
      Type: String
Resources:
  imageprocessingservice:
    Type: 'AWS::Serverless::Function'
    Properties:
      Handler: index.handler
      Runtime: nodejs8.10
      CodeUri: .
      Description: >-
        Uses ImageMagick to perform simple image processing operations, such as resizing.
      MemorySize: 512
      Timeout: 30
      Policies:
        - SESSendBouncePolicy:
            IdentityName: !Ref IdentityNameParameter
```

AWS Serverless Application Model

A CLOUD GURU

‘sam local start-api’

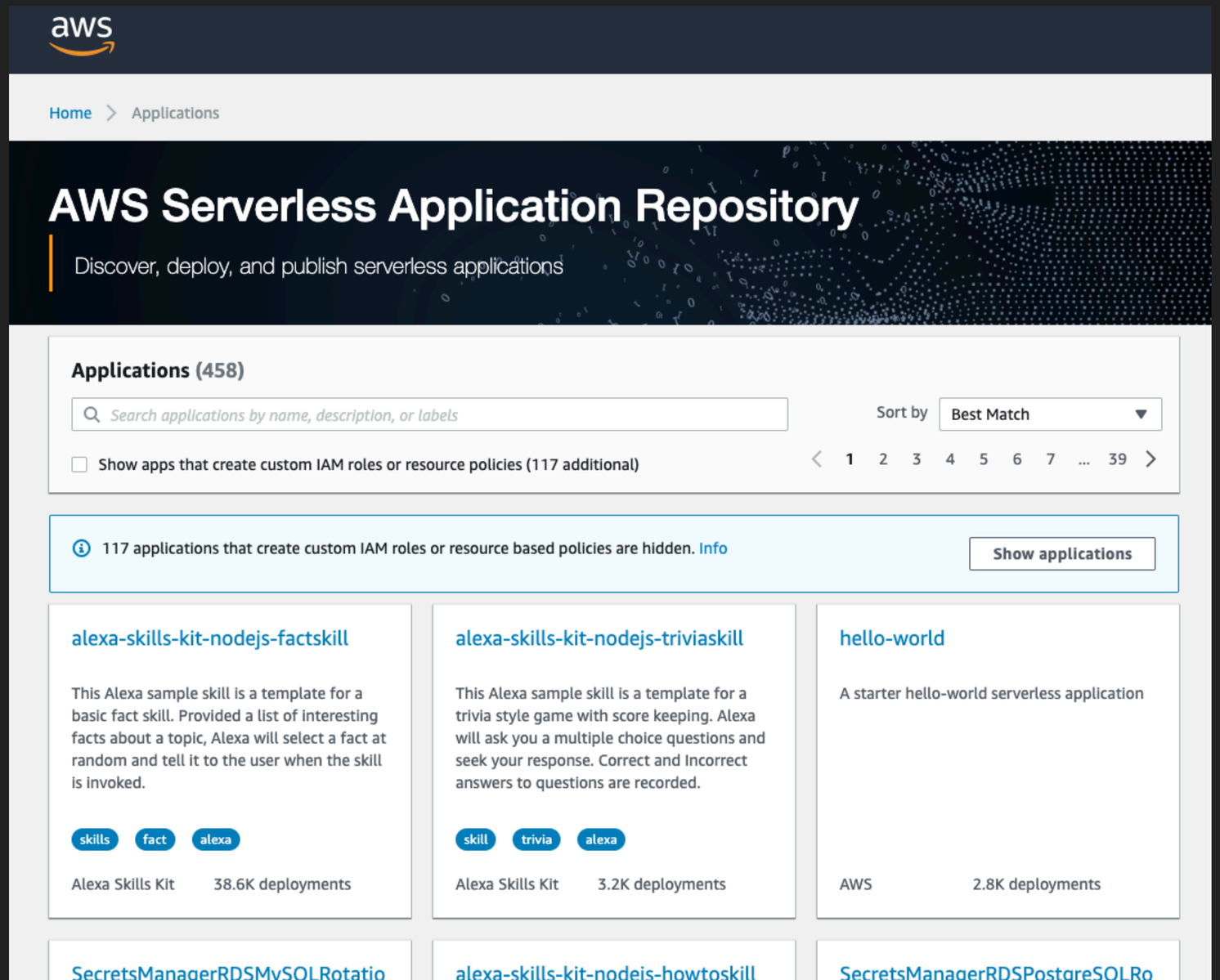
AWS Serverless Application Model



AWS Serverless Application Model

A CLOUD GURU

AWS Serverless Application Repository



AWS SAM vs Serverless Framework

A CLOUD GURU

AWS SAM

- Uses YAML for templates
- Purpose-built to help make developing and deploying serverless apps as efficient as possible
- Generates CloudFormation scripts

serverless

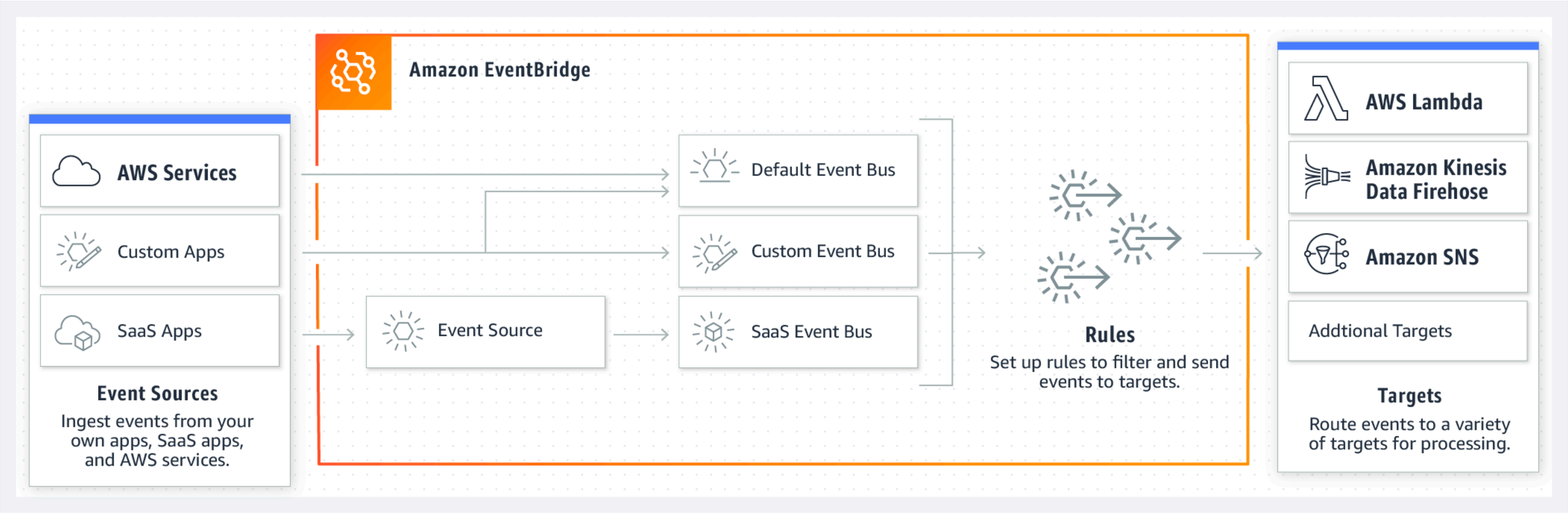
Framework

- Uses YAML for templates
- Purpose-built to help make developing and deploying serverless apps as efficient as possible
- Generates CloudFormation Scripts
- Supports many other cloud providers such as Azure, GCP, IBM, CloudFlare, etc.

Amazon EventBridge

A CLOUD GURU

- Designed to link variety of AWS and 3rd party apps to rules logic for launching other event-based actions



Amazon EventBridge

A CLOUD GURU

- Designed to link variety of AWS and 3rd party apps to rules logic for launching other event-based actions

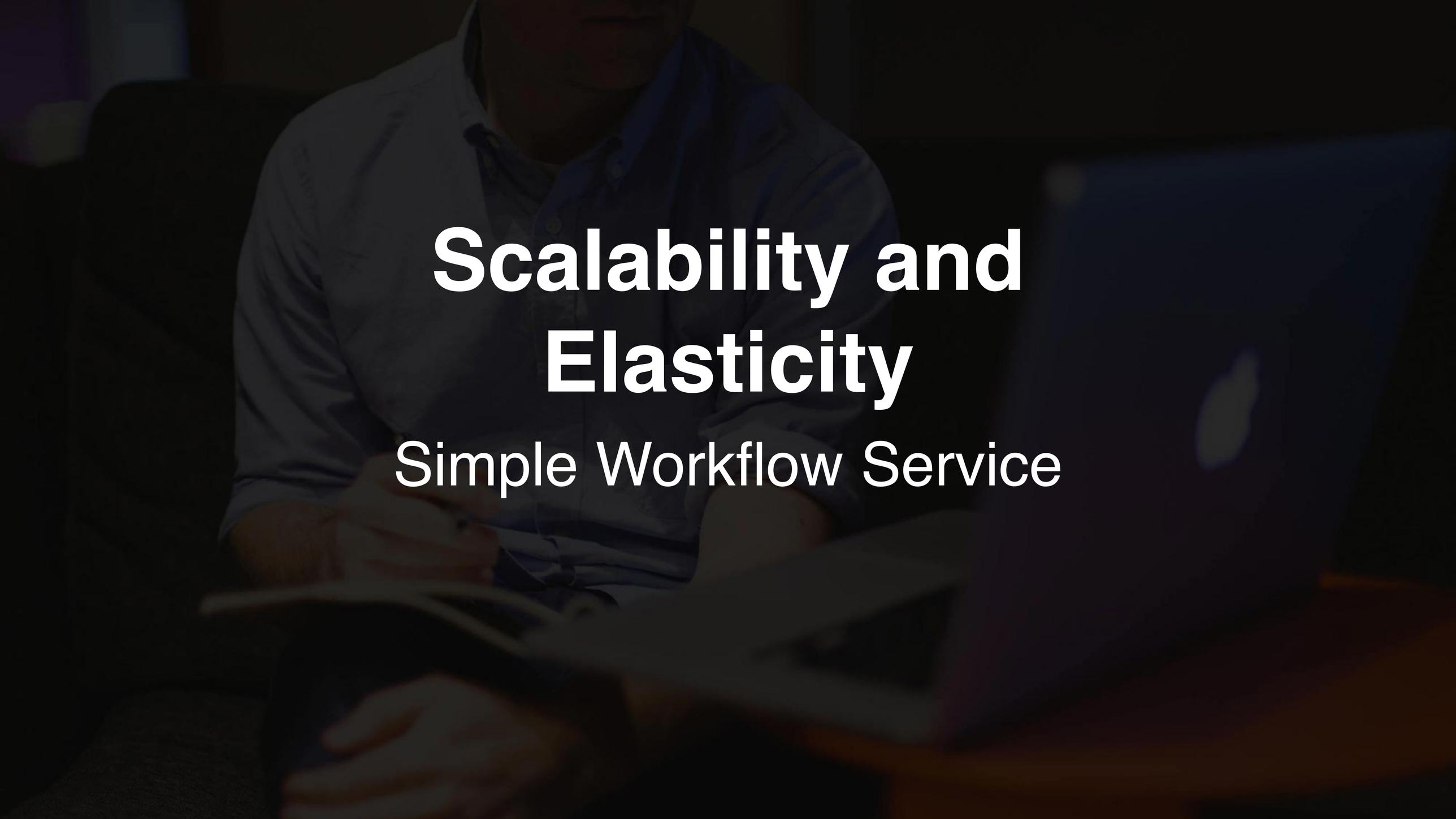
onelogin

DATADOG

pagerduty

SUGARCRM

zendesk



Scalability and Elasticity

Simple Workflow Service

Simple Workflow Service (AWS SWF)

A CLOUD GURU

- Create distributed asynchronous systems as workflows.
- Supports both sequential and parallel processing.
- Tracks the state of your workflow which you interact and update via API.
- Best suited for human-enabled workflows like a order fulfillment or procedural requests.
- AWS recommends new applications — look at Step Functions over SWF.

AWS SWF

AWS SWF

A CLOUD GURU

Activity Worker

A program that interacts with the AWS SWF service to get tasks, process tasks and return results

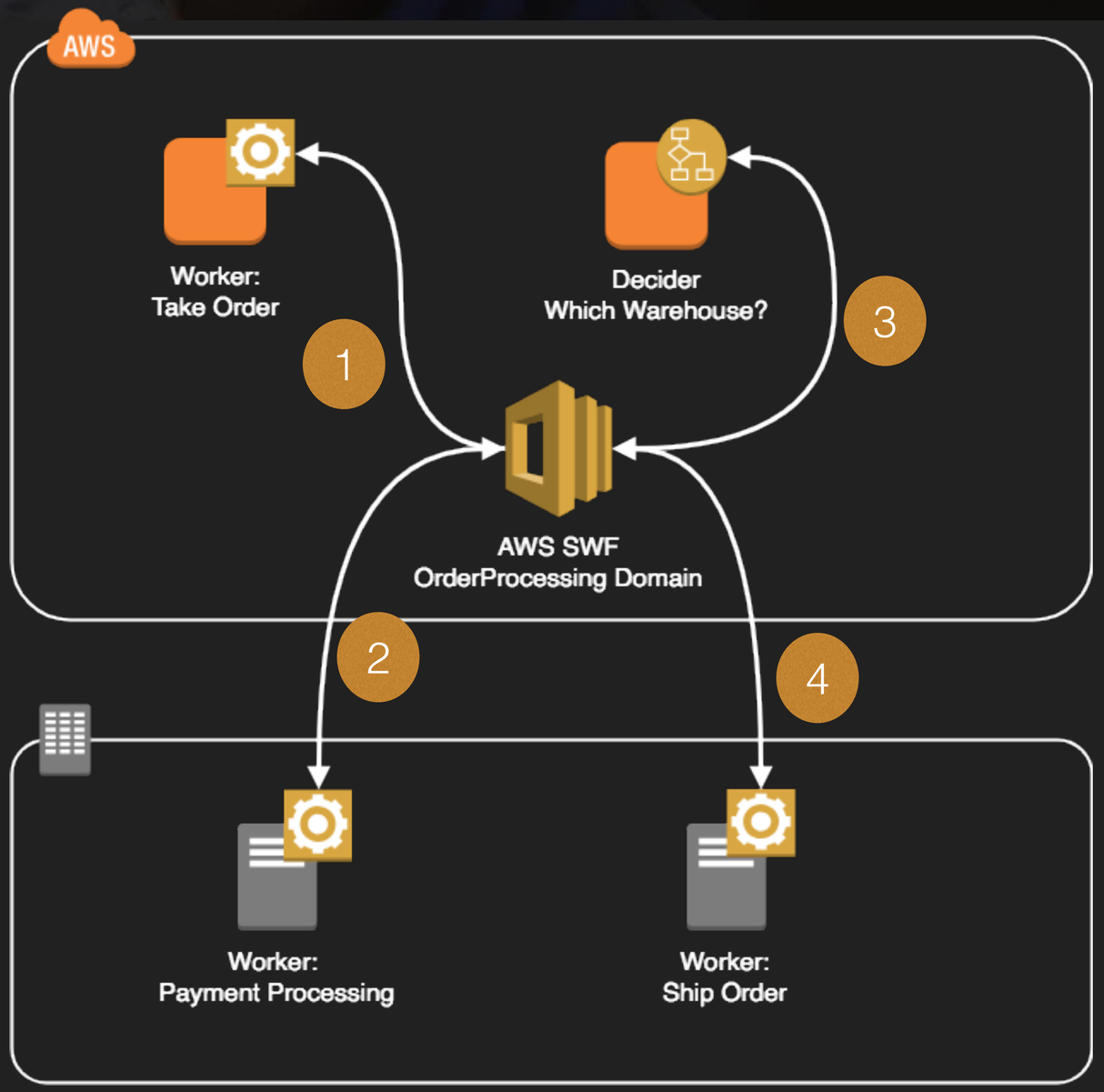
API Calls

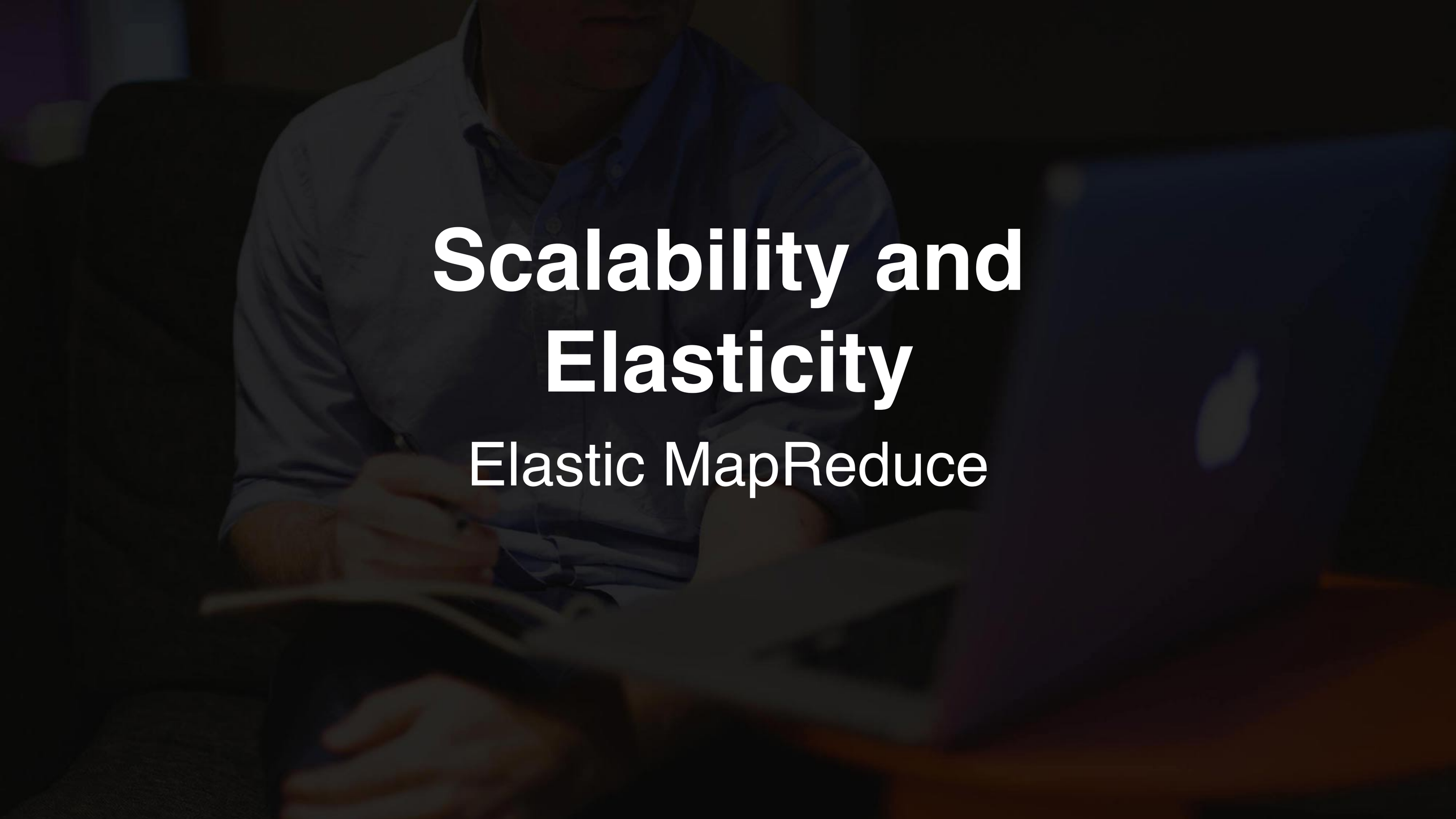
AWS SWF

Decider

A program that controls coordination of tasks such as their ordering, concurrency and scheduling

AWS SWF


AWS
Worker:
Take Order
Decider
Which Warehouse?
1
3
AWS SWF
OrderProcessing Domain
2
4
Worker:
Payment Processing
Worker:
Ship Order

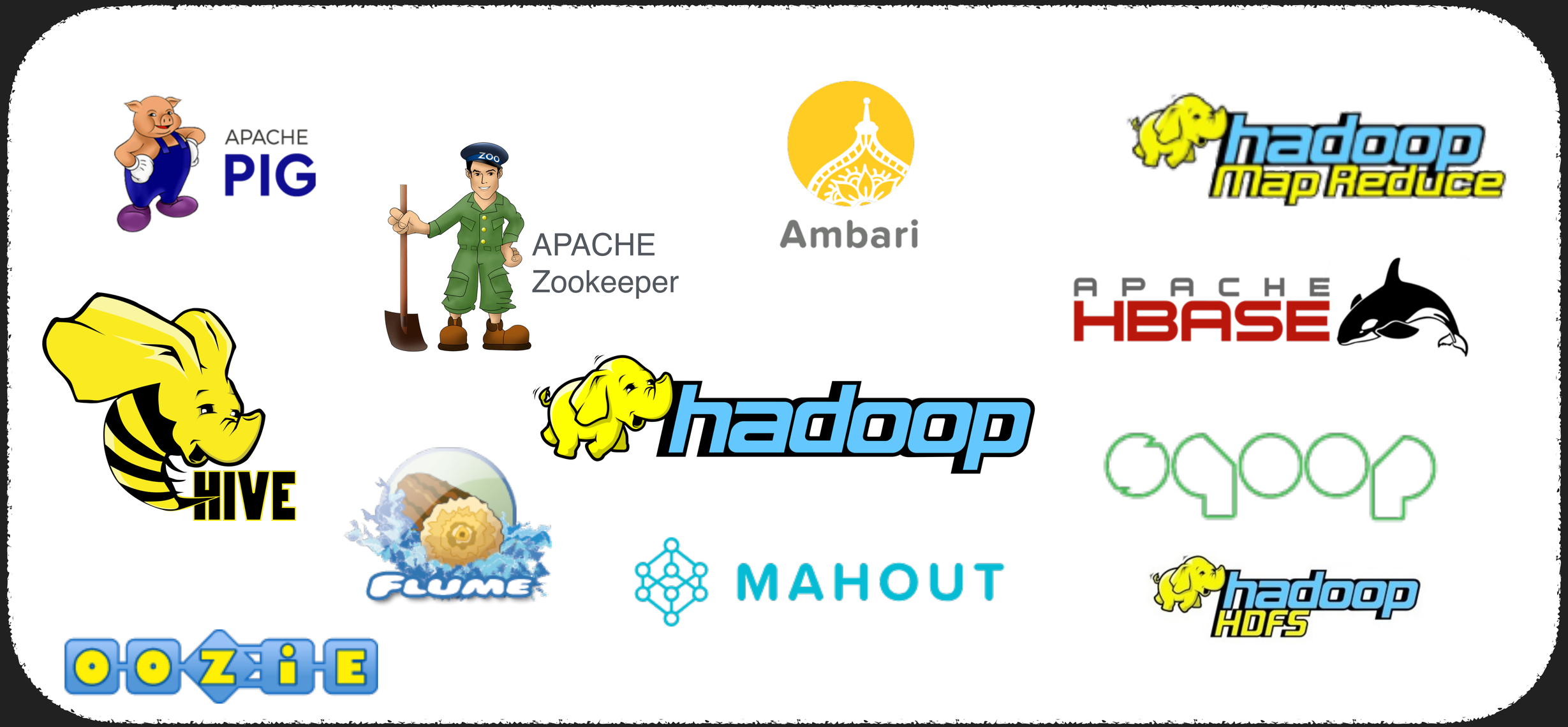


Scalability and Elasticity

Elastic MapReduce

The Zoo

A CLOUD GURU



The Zoo

A CLOUD GURU

Ambari
Management and Monitoring

Ambari

Sqoop

Sqoop
Data Transfer

ZooKeeper
Resource
Coordination

OOZIE

Oozie
Workflow

APACHE PIG

Pig
Scripting

HIVE

Hive
SQL

MAHOUT

Mahout
Machine
Learning

APACHE HBASE

HBase
Columnar
Datastore

Flume

Flume
Log Collection

hadoop MapReduce

Hadoop MapReduce
Distributed Processing

hadoop HDFS

Hadoop Distributed File System
Persistant Datastore

The Zoo

A CLOUD GURU

Ambari
Management and Monitoring

Ambari

Sqoop

Sqoop
Data Transfer

ZooKeeper
Resource
Coordination

OOZIE

Oozie
Workflow

APACHE PIG

Pig
Scripting

HIVE

Hive
SQL

MAHOUT

Mahout
Machine
Learning

APACHE HBASE

HBase
Columnar
Datastore

Flume

Flume
Log Collection

hadoop MapReduce

Hadoop MapReduce
Distributed Processing

hadoop HDFS

Hadoop Distributed File System
Persistant Datastore

The Zoo

A CLOUD GURU

Enterprise Support, Professional Services and Project Contributions

HORTONWORKS® POWERING THE FUTURE OF DATA™

cloudera®

Ambari

Ambari

Sqoop

Flume

ZooKeeper

Oozie

Pig

Hive

Mahout

HBase

Hadoop MapReduce

Hadoop Distributed File

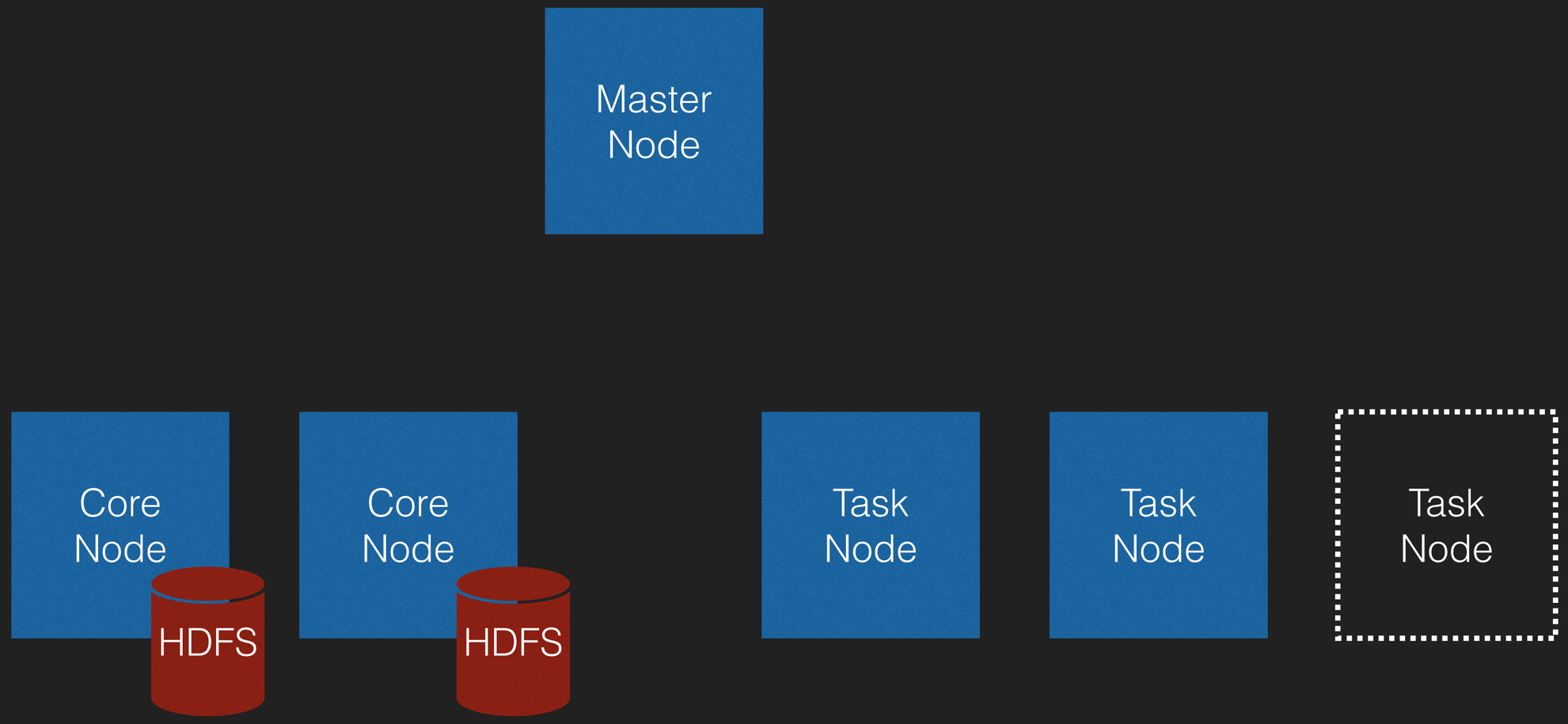
AWS Elastic MapReduce

A CLOUD GURU

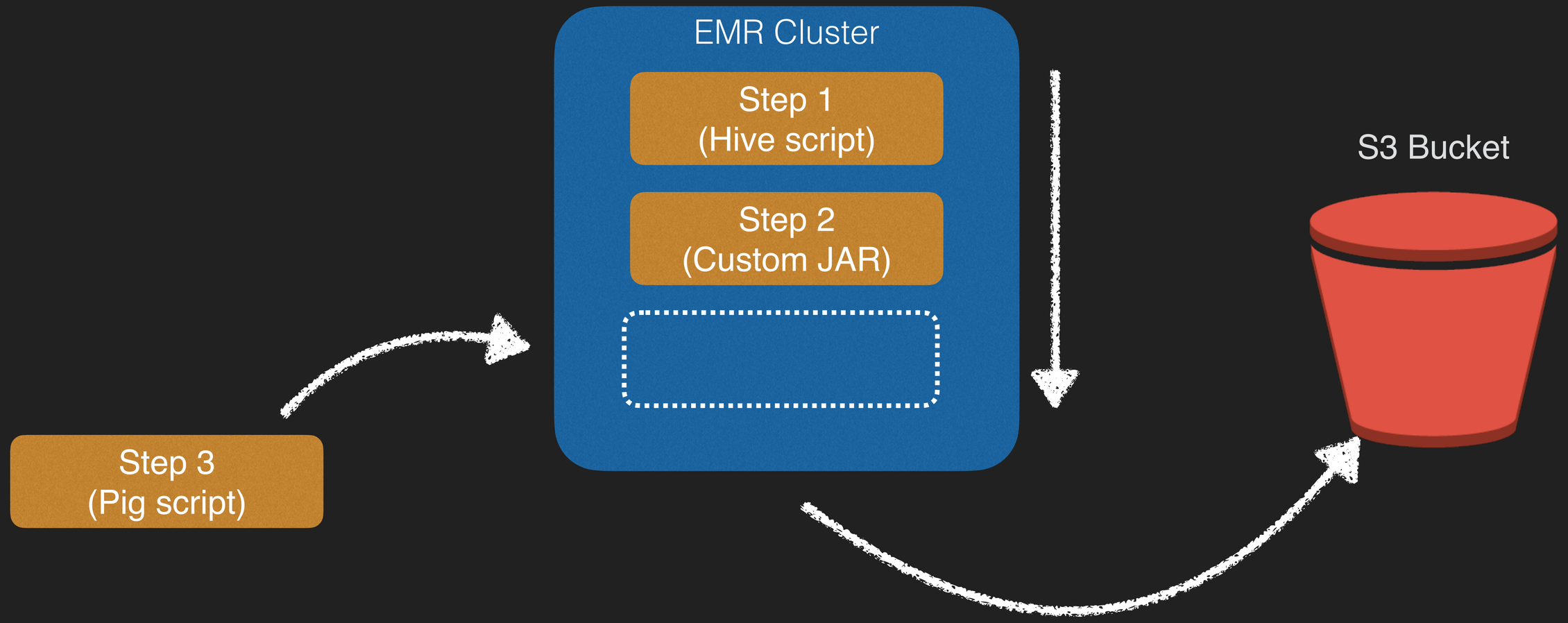
- Managed Hadoop framework for processing huge amounts of data
- Also supports Apache Spark, HBase, Presto and Flink
- Most commonly used for log analysis, financial analysis or extract, translate and loading (ETL) activities
- A Step is a programatic task for performing some process on the data (i.e. count words)
- A Cluster is a collection of EC2 instances provisioned by EMR to run your Steps.

AWS EMR

Components of AWS EMR



AWS EMR Process



AWS Step Functions and Amazon Batch

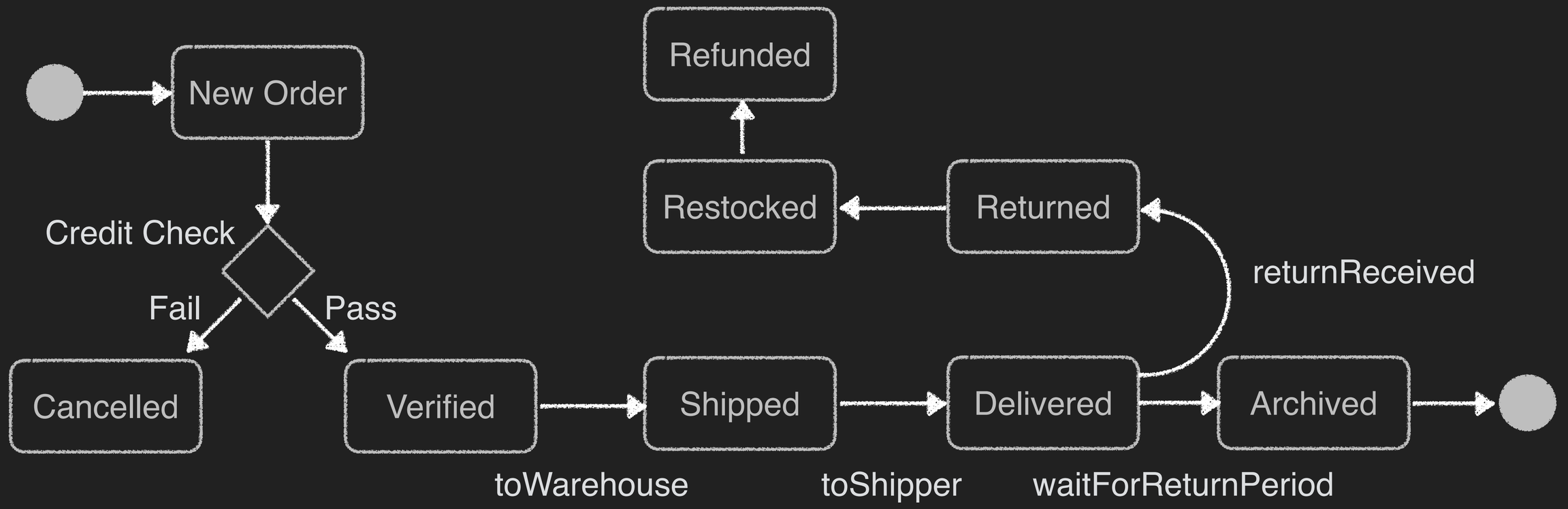
AWS Step Functions

A CLOUD GURU

- Managed workflow and orchestration platform
- Scalable and highly available
- Define your app as a state machine
- Create tasks, sequential steps, parallel steps, branching paths or timers
- Amazon State Language declarative JSON
- Apps can interact and update the stream via Step Function API
- Visual interface describes flow and realtime status
- Detailed logs of each step execution

AWS Step Functions

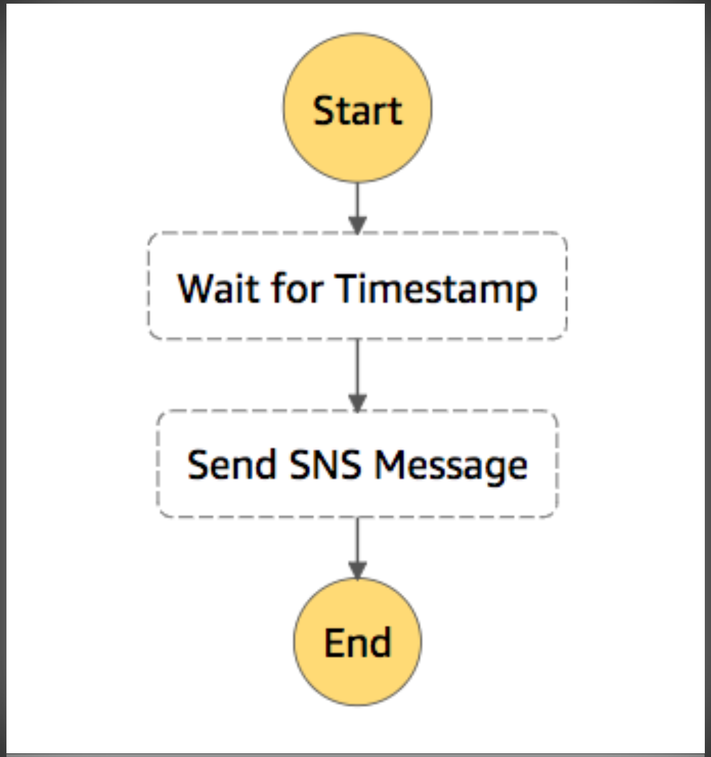
Finite State Machine



Task Timer Example

A CLOUD GURU

```
{
  "Comment": "A Task Timer example of the Amazon States Language scheduling a task",
  "StartAt": "Wait for Timestamp",
  "States": {
    "Wait for Timestamp": {
      "Type": "Wait",
      "SecondsPath": "$.timer_seconds",
      "Next": "Send SNS Message"
    },
    "Send SNS Message": {
      "Type": "Task",
      "Resource": "SNS_ARN",
      "Retry": [
        {
          "ErrorEquals": [
            "States.ALL"
          ],
          "IntervalSeconds": 1,
          "MaxAttempts": 3,
          "BackoffRate": 2
        }
      ],
      "End": true
    }
  }
}
```



AWS Batch

A CLOUD GURU

Management tool for creating, managing and executing batch-oriented tasks using EC2 instances.

1. Create a Compute Environment: Managed or Unmanaged, Spot or On-Demand, vCPUs
2. Create a Job Queue with priority and assigned to a Compute Environment
3. Create Job Definition: Script or JSON, environment variables, mount points, IAM role, container image, ec.
4. Schedule the Job

AWS Batch

Comparisons

A CLOUD GURU

	When	Use Case
Step Functions	Out-of-the-box coordination of AWS service components	Order Processing Flow
Simple Workflow Service	Need to support external processes or specialized execution logic	Loan Application Process with Manual Review Steps
Simple Queue Service	Messaging Queue; Store and forward patterns	Image Resize Process
AWS Batch	Scheduled or reoccurring tasks that do not require heavy logic	Rotate Logs Daily on Firewall Appliance

Architecting to Scale

Exam Tips

Exam Tips

A CLOUD GURU

Kinesis

- Exam is likely to be restricted to the Data Stream use cases for Kinesis such as Data Streams and Firehose.
- Understand shard concept and how partition keys and sequences enabled shards to manage data flow.

DynamoDB Auto Scaling

- Know the new and old terminology and concept of a partition, partition key and sort key in the context of DynamoDB.
- Understand how DynamoDB calculates total partitions and allocates RCU and WCU across available partitions.
- Conceptually know how data is stored across partitions.

Exam Tips

A CLOUD GURU

CloudFront Part 2

- Know that both static and dynamic content is supported.
- Understand possible origins and how multiple origins can be used together with Behaviors.
- Know invalidation methods, zone apex and geo-restriction options.

SNS

- Understand a loosely coupled architecture and benefits it brings.
- Know the different types of subscription endpoints supported.

Exam Tips

A CLOUD GURU

SQS

- Know the difference between Standard and FIFO queues.
- Know difference between a Pub/Sub (SNS) and Message Queueing (SQS) architecture.

Lambda

- Know what “Serverless” is in concept and how Lambda can facilitate such an architecture.
- Know the languages supported by Lambda.

Exam Tips

A CLOUD GURU

SWF

- Understand the difference and functions of a Worker and a Decider.
- Best suited for human-enabled workflows like order fulfillment or procedural requests.

Elastic MapReduce

- Understand the parts of a Hadoop landscape at a high level.
- Know what a Cluster is and what Steps are.
- Understand the roles of a Master Node, Core Nodes, and Task Nodes.

Exam Tips

A CLOUD GURU

Step Functions

- Managed workflow and orchestration platform considered preferred for modern development over AWS Simple Workflow Service.
- Supports tasks, sequential steps, parallel steps, branching, timers.

AWS Batch

- Ideal for use cases where a routine activity must be performed at a specific interval or time of day.
- Behind the scenes, EC2 instances are provisioned as workers to perform the batch activities then terminated when done.

Exam Tips

A CLOUD GURU

Serverless Application Model

- Open-source framework created to make serverless application development and deployment more efficient
- SAM CLI translates specific SAM YAML into CloudFormation scripts which is then used to deploy on AWS
- Serverless Application Repository contains lots of apps you can use as examples or elements of your own apps
- AWS SAM and the “Serverless Framework” are different things.

Exam Tips

A CLOUD GURU

Auto Scaling

- Be familiar with the positioning and different purposes of EC2 Auto Scaling vs. Application Auto Scaling vs. AWS Auto Scaling
- Know the different scaling options and policies.
- Understand the difference and limitations between horizontal and vertical scaling.
- Know what a cool down period is and how it might impact your responsiveness to demand.

Architecting to Scale

Pro Tips

Pro Tips

A CLOUD GURU

- Elasticity will drive most benefit from the cloud.
- Think Cloud-First designs if you’re in a Green Field scenario even if you’re deploying on-prem.
- If you’re in a “Brown Field” situation, create roadmaps for cloud-first enablers like distributed applicaitons, federated data and SOA.
- Be careful to not let elasticity cover for poor development methods.
- Microservice concepts help achieve scalability via decoupling, simplification and separation of concerns.

Architecting to Scale

Challenge 1

Challenge 1

A CLOUD GURU

You have been called in to consult with a Mobile App company around a challenge. Their app, SeeFood, allows people to take very high resolution multi-megabyte pictures of their food and derives the nutritional value from the photo. They are due to launch within 5 days and have to be ready to scale as the app is expected to be a hit.

Unfortunately, the machine learning portion of the app is not yet perfected so they have recruited a large number of people (analysts) to view the photos, lookup the nutritional value and return the info to the user.

How would you best architect for this scenario?

A. Create a pub/sub architecture, supplying the analysts with a custom program they can use to receive images pushed from SQS queues and process the images.

B. Use SNS to create a process which will email received images equally across a set of email accounts. Provide the analysts with a custom program to read those mailboxes and process the images.

C. Use Kinesis Data Stream and Kinesis Firehose to take in the image and save them off to S3. Write a custom program that analysts can use to view the S3 images and process them.

D. Use AWS SWF to receive the images from the mobile app. Use the weighted routing feature in SWF to forward the images equally across a pool of analysts to process the images.

E. Configure the app to upload the image to S3. Using S3 Events, trigger a message to an SQS queue linking to the image. Provide analysts with a custom program to poll the queue and process the images.

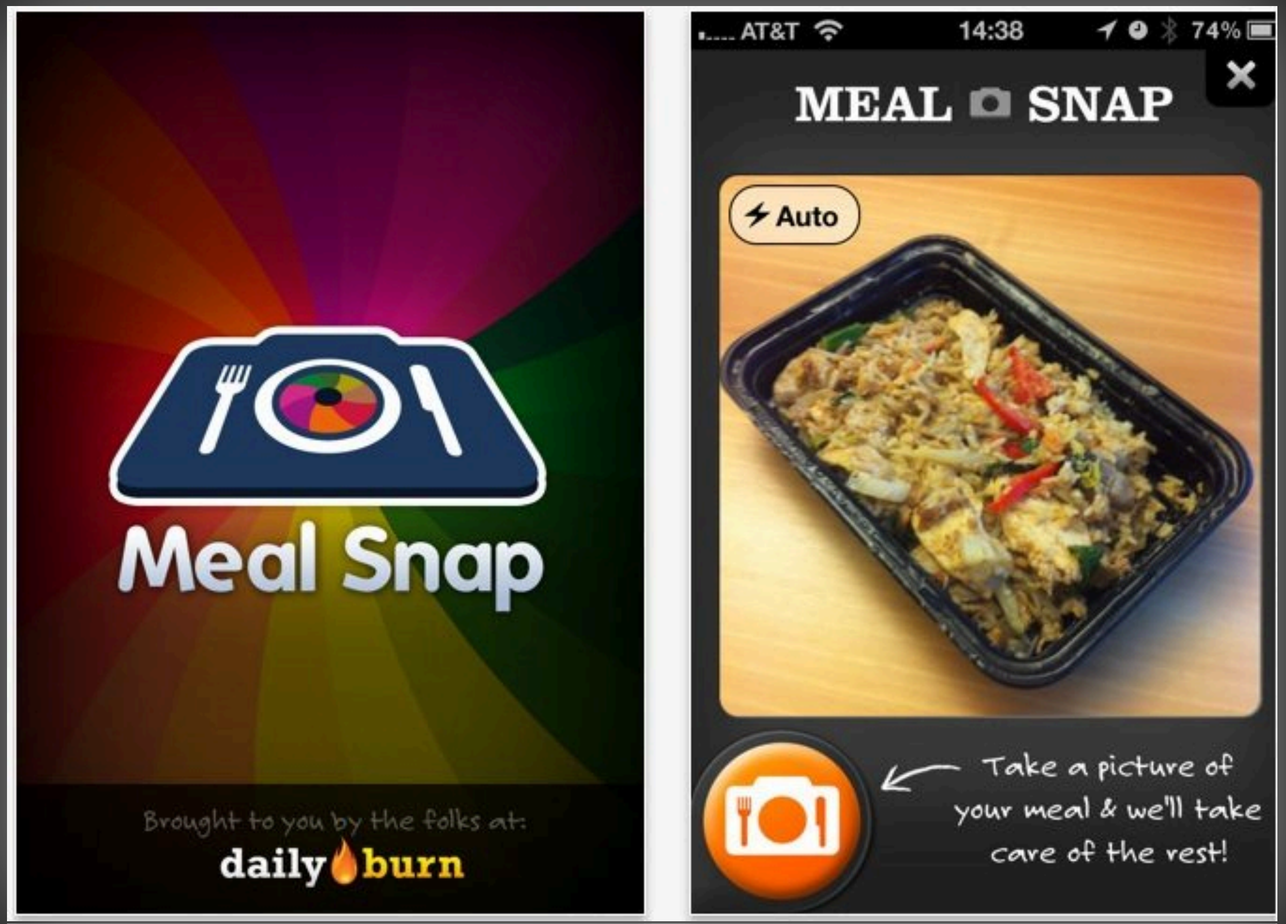
Challenge 1

A CLOUD GURU

DAILY BURN

Challenge 1

A CLOUD GURU



amazon
beta
mechanical turk

Challenge 1

A CLOUD GURU

You have been called in to consult with a Mobile App company around a challenge. Their app, SeeFood, allows people to take very high resolution multi-megabyte pictures of their food and derives the nutritional value from the photo. They are due to launch within 5 days and have to be ready to scale as the app is expected to be a hit.

Unfortunately, the machine learning portion of the app is not yet perfected so they have recruited a large number of people (analysts) to view the photos, lookup the nutritional value and return the info to the user.

How would you best architect for this scenario?

A. Create a pub/sub architecture, supplying the analysts with a custom program they can use to receive images pushed from SQS queues and process the images.

B. Use SNS to create a process which will email received images equally across a set of email accounts. Provide the analysts with a custom program to read those mailboxes and process the images.

C. Use Kinesis Data Stream and Kinesis Firehose to take in the image and save them off to S3. Write a custom program that analysts can use to view the S3 images and process them.

D. Use AWS SWF to receive the images from the mobile app. Use the weighted routing feature in SWF to forward the images equally across a pool of analysts to process the images.

E. Configure the app to upload the image to S3. Using S3 Events, trigger a message to an SQS queue linking to the image. Provide analysts with a custom program to poll the queue and process the images.

Challenge 1

A CLOUD GURU

You have been called in to consult with a Mobile App company around a challenge. Their app, SeeFood, allows people to take very high resolution multi-megabyte pictures of their food and derives the nutritional value from the photo. They are due to launch within 5 days and have to be ready to scale as the app is expected to be a hit.

Unfortunately, the machine learning portion of the app is not yet perfected so they have recruited a large number of people (analysts) to view the photos, lookup the nutritional value and return the info to the user.

How would you best architect for this scenario?

A. ~~Create a pub/sub architecture, supplying the analysts with a custom program they can use to receive images pushed from SQS queues and process the images.~~

B. Use SNS to create a process which will email received images equally across a set of email accounts. Provide the analysts with a custom program to read those mailboxes and process the images.

C. Use Kinesis Data Stream and Kinesis Firehose to take in the image and save them off to S3. Write a custom program that analysts can use to view the S3 images and process them.

D. ~~Use AWS SWF to receive the images from the mobile app. Use the weighted routing feature in SWF to forward the images equally across a pool of analysts to process the images.~~

E. Configure the app to upload the image to S3. Using S3 Events, trigger a message to an SQS queue linking to the image. Provide analysts with a custom program to poll the queue and process the images.

Challenge 1

A CLOUD GURU

You have been called in to consult with a Mobile App company around a challenge. Their app, SeeFood, allows people to take very high resolution multi-megabyte pictures of their food and derives the nutritional value from the photo. They are due to launch within 5 days and have to be ready to scale as the app is expected to be a hit.

Unfortunately, the machine learning portion of the app is not yet perfected so they have recruited a large number of people (analysts) to view the photos, lookup the nutritional value and return the info to the user.

How would you best architect for this scenario?

A. ~~Create a pub/sub architecture, supplying the analysts with a custom program they can use to receive images pushed from SQS queues and process the images.~~

B. Use SNS to create a process which will email received images equally across a set of email accounts. Provide the analysts with a custom program to read those mailboxes and process the images.

C. Use Kinesis Data Stream and Kinesis Firehose to take in the image and save them off to S3. Write a custom program that analysts can use to view the S3 images and process them.

D. ~~Use AWS SWF to receive the images from the mobile app. Use the weighted routing feature in SWF to forward the images equally across a pool of analysts to process the images.~~

E. Configure the app to upload the image to S3. Using S3 Events, trigger a message to an SQS queue linking to the image. Provide analysts with a custom program to poll the queue and process the images.

Architecting to Scale

Challenge 2

Challenge 2

A CLOUD GURU

You've implemented EC2 Auto Scaling for the web layer behind an ELB. User traffic is quite unpredictable and changes rapidly so you rely on auto-scaling to provide the best user experience and performance.

Weekly, you receive a report from a third-party web analytics company that shows how your website is performing from the perspective of a customer. Unfortunately, during moments of high demand, customer response is still slower than your target.

What might be some causes of this and possible corrective actions but yet remain as cost-efficient as possible?

Select 2.

A. The default cool down time for scaling is too long. Create a scaling-specific cool down that overrides the default time.

B. Your Internet Gateway has become saturated. Create a new VPC, ELB and auto-scaling group with its own Internet Gateway and route traffic equally to either ELB.

C. You don't have enough instances active during peak load times. Implement a scheduled scaling and increase the number of minimum instances during peak times.

D. Customer demand is not stable enough to properly trigger scaling thresholds of your existing scaling policies. Implement a step scaling policy configured around CPU utilization.

E. Your launch configuration is setup to launch undersized instances. Edit your current launch configuration to launch larger instances to better absorb the load.

F. You need more dynamic web capacity than your current configuration. Create a new launch configuration to launch smaller instances and implement a step scaling policy with a shorter cool down period.

Challenge 2

A CLOUD GURU

You've implemented EC2 Auto Scaling for the web layer behind an ELB. User traffic is quite unpredictable and changes rapidly so you rely on auto-scaling to provide the best user experience and performance.

Weekly, you receive a report from a third-party web analytics company that shows how your website is performing from the perspective of a customer. Unfortunately, during moments of high demand, customer response is still slower than your target.

What might be some causes of this and possible corrective actions but yet remain as cost-efficient as possible?

Select 2.

A. The default cool down time for scaling is too long. Create a scaling-specific cool down that overrides the default time.

B. Your Internet Gateway has become saturated. Create a new VPC, ELB and auto-scaling group with its own Internet Gateway and route traffic equally to either ELB.

C. You don't have enough instances active during peak load times. Implement a scheduled scaling and increase the number of minimum instances during peak times.

D. Customer demand is not stable enough to properly trigger scaling thresholds of your existing scaling policies. Implement a step scaling policy configured around CPU utilization.

E. Your launch configuration is setup to launch undersized instances. Edit your current launch configuration to launch larger instances to better absorb the load.

F. You need more dynamic web capacity than your current configuration. Create a new launch configuration to launch smaller instances and implement a step scaling policy with a shorter cool down period.

Architecting to Scale

Lab

Auto-Scaling

Auto-Scaling Experiment

A CLOUD GURU

You are consulting with a pharmaceutical company new to AWS. Every night, they have a series of calculations they run against data gathered by the day's research.

Some days they have lots of data and some days they only have a little data to process. They have some knowledge about scalability on AWS but are skeptical that it will work for them.

Since they are scientists, they want objective proof the scaling works. Design and conduct an experiment to show how scaling works in the most cost-effective way.

Auto-Scaling Experiment

A CLOUD GURU

Sequence	Task	Hint
1	Design an experiment to demonstrate autoscaling to the scientists using cost-effective methods.	
2	Build the scenario that simulates heavy computational load to trigger scale-out.	
3	Demonstrate scale-in as the load diminishes.	

Auto-Scaling Experiment

A CLOUD GURU

Expert Level

Pause now and go give it a try without hints.

Auto-Scaling Experiment

A CLOUD GURU

Intermediate Level

Keep going for some hints to get you started.

Auto-Scaling Experiment

A CLOUD GURU

Sequence	Task	Hint
1	Design an experiment to demonstrate autoscaling to the scientists using cost-effective methods.	X marks the Spot.
2	Build the scenario that simulates heavy computational load to trigger scale-out.	Sometimes, stress is good.
3	Demonstrate scale-in as the load diminishes.	Very interesting dynamic.

Auto-Scaling Experiment

A CLOUD GURU

Intermediate Level

Pause now and go give it a try.
Keep going to see how I did it.

Auto-Scaling Experiment

A CLOUD GURU

Walkthrough

Business Continuity

Introduction

A CLOUD GURU

Exam Blueprint

A CLOUD GURU

Domain 1: Design for Organizational Complexity

1.1. Determine cross-account authentication and access strategy for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

1.2. Determine how to design networks for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

1.3. Determine how to design a multi-account AWS environment for complex organizations (for example, an organization with varying compliance requirements, multiple business units, and varying scalability requirements).

Domain 2: Design for New Solutions

2.1. Determine security requirements and controls when designing and implementing a solution.

2.2. Determine a solution design and implementation strategy to meet reliability requirements.

2.3. Determine a solution design to ensure business continuity.

2.4. Determine a solution design to meet performance objectives.

2.5. Determine a deployment strategy to meet business requirements when designing and implementing a solution.

Exam Blueprint

A CLOUD GURU

Domain 3: Migration Planning

3.1. Select existing workloads and processes for potential migration to the cloud.

3.2. Select migration tools and/or services for new and migrated solutions based on detailed AWS knowledge.

3.3. Determine a new cloud architecture for an existing solution.

3.4. Determine a strategy for migrating existing on-premises workloads to the cloud.

Domain 4: Cost Control

4.1. Select a cost-effective pricing model for a solution.

4.2. Determine which controls to design and implement that will ensure cost optimization.

4.3. Identify opportunities to reduce cost in an existing solution.

Domain 5: Continuous Improvement for Existing Solutions

5.1. Troubleshoot solution architectures.

5.2. Determine a strategy to improve an existing solution for operational excellence.

5.3. Determine a strategy to improve the reliability of an existing solution.

5.4. Determine a strategy to improve the performance of an existing solution.

5.5. Determine a strategy to improve the security of an existing solution.

5.6. Determine how to improve the deployment of an existing solution.