# Buenas Prácticas en Arquitecturas de Analítica en la Nube

Carlos Paez – Solutions Architect
linkedin.com/in/carlospaez/

# Well-Architected Framework

# Well-Architected Framework

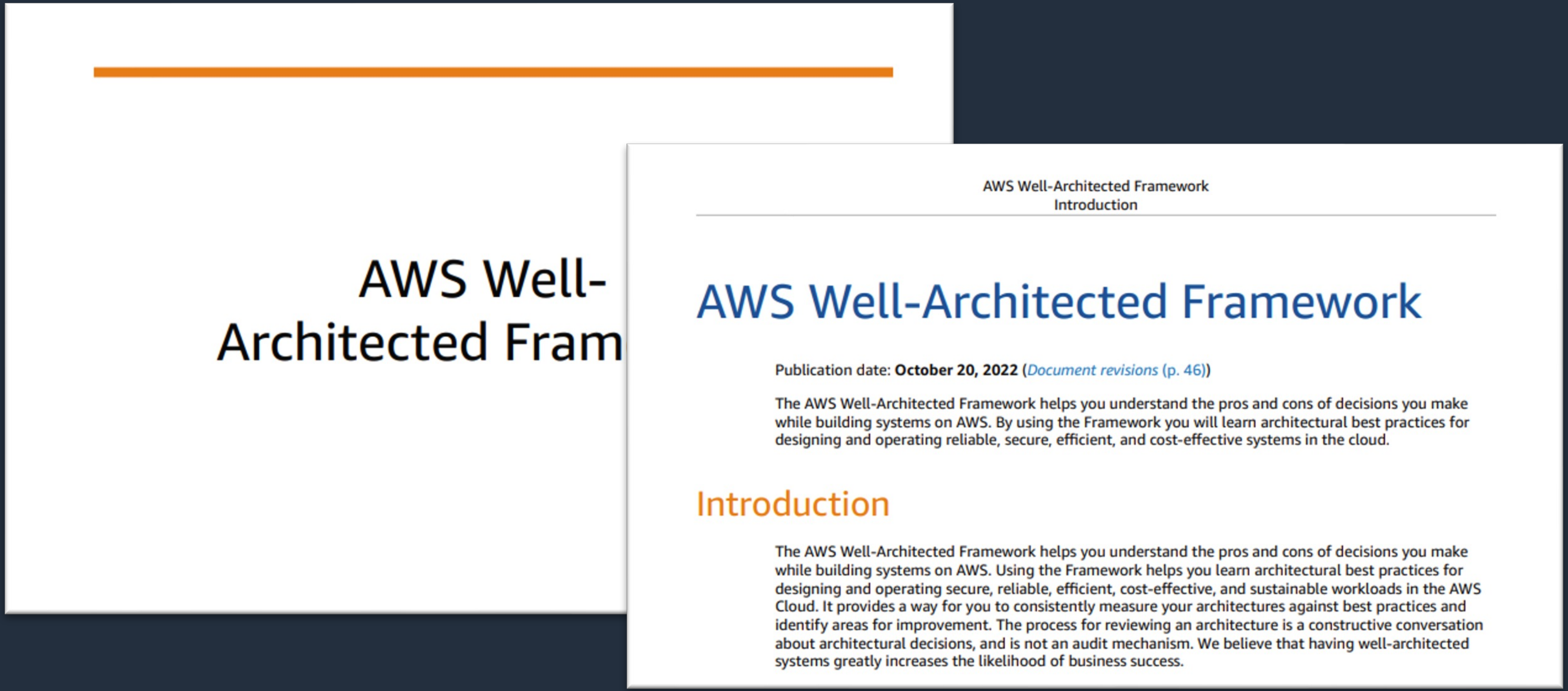
The AWS Well-Architected Framework helps you understand the pros and cons of decisions you make while building systems on AWS. By using the Framework you will learn architectural **best practices** for designing and operating **reliable**, **secure**, **efficient**, and **cost-effective** systems in the cloud.

- Pillars

  Design principles

- Lenses

  Best Practices



AWS Well-
Architected Fram

AWS Well-Architected Framework
Introduction

## AWS Well-Architected Framework

Publication date: **October 20, 2022** (*Document revisions* (p. 46))

The AWS Well-Architected Framework helps you understand the pros and cons of decisions you make while building systems on AWS. By using the Framework you will learn architectural best practices for designing and operating reliable, secure, efficient, and cost-effective systems in the cloud.

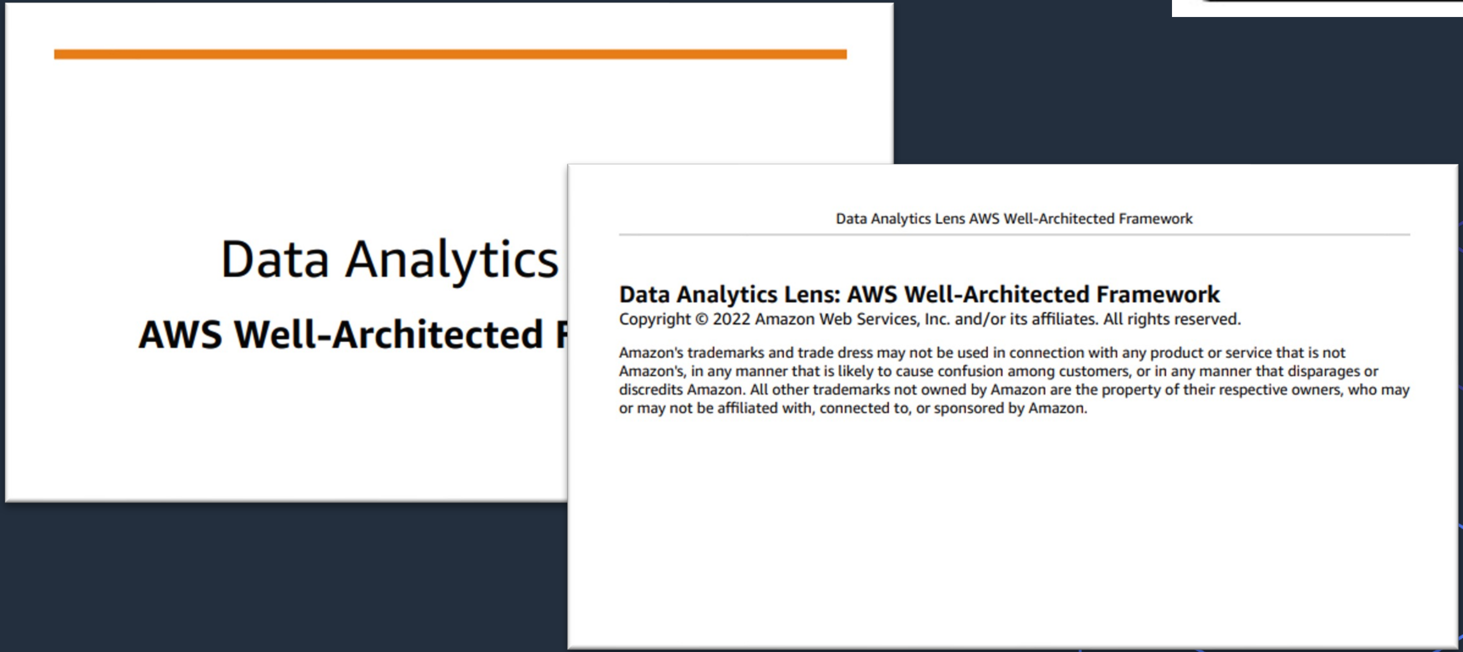
## Introduction

The AWS Well-Architected Framework helps you understand the pros and cons of decisions you make while building systems on AWS. Using the Framework helps you learn architectural best practices for designing and operating secure, reliable, efficient, cost-effective, and sustainable workloads in the AWS Cloud. It provides a way for you to consistently measure your architectures against best practices and identify areas for improvement. The process for reviewing an architecture is a constructive conversation about architectural decisions, and is not an audit mechanism. We believe that having well-architected systems greatly increases the likelihood of business success.

https://docs.aws.amazon.com/wellarchitected/latest/framework/welcome.html

# Well-Architected Data Analytics Lens

SCAN ME



Data Analytics

**AWS Well-Architected F**

Data Analytics Lens AWS Well-Architected Framework

**Data Analytics Lens: AWS Well-Architected Framework**

Copyright © 2022 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

# Analytics Lens

Defines the principles  and key architectural elements for designing Analytics Workload based on the 6 pillars of Well-Architected Framework.

- Operational Excellence
- Security
- Reliability
- Performance Efficiency
- Cost Optimization
- Sustainability

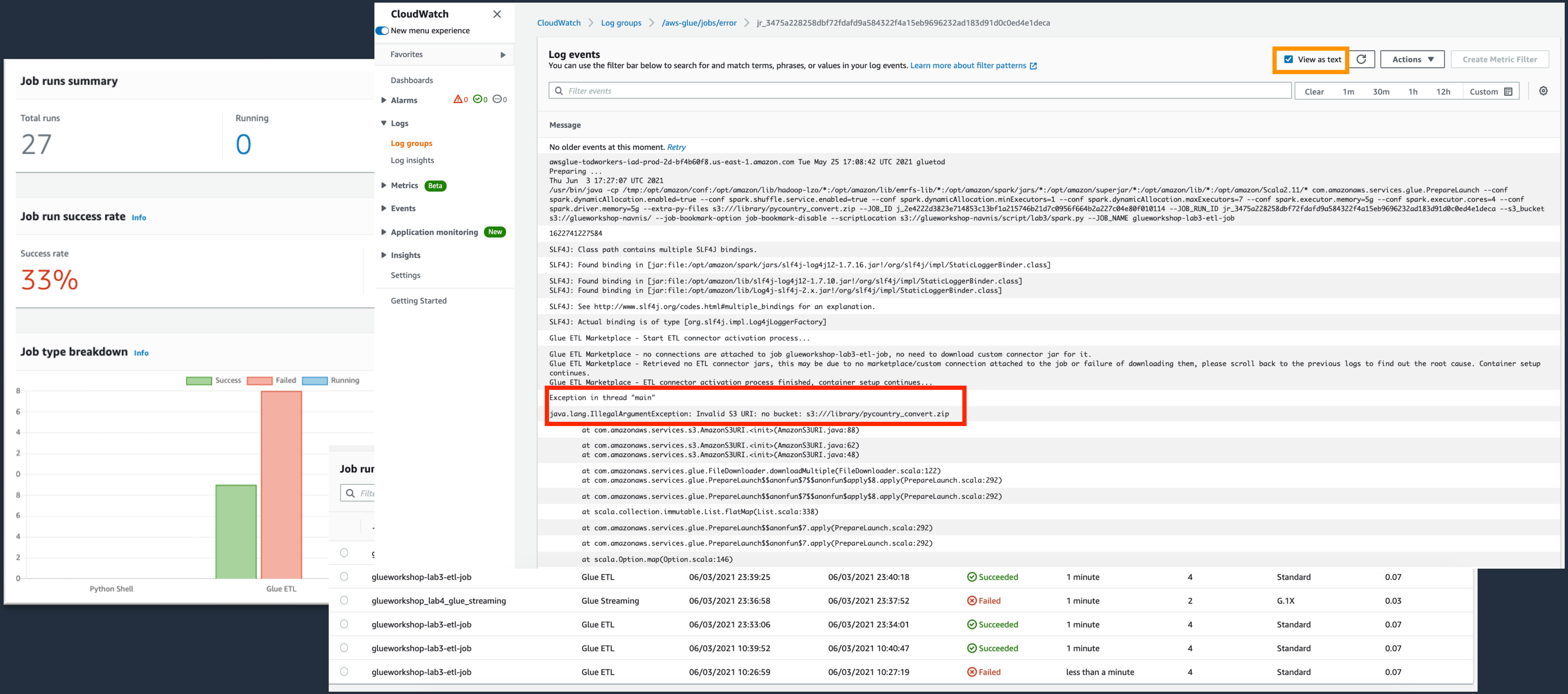
https://docs.aws.amazon.com/wellarchitected/latest/analytics-lens/analytics-lens.html

# Operational Excellence

- Modernize deployment of the analytics jobs and applications
- Build financial accountability models for data and workload usage
- Monitor the health of the analytics pipelines

https://docs.aws.amazon.com/wellarchitected/latest/analytics-lens/operational-excellence.html

# Monitor the health of the analytics pipelines



You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns

No older events at this moment. Retry

```
awsglue-todworkers-iad-prod-2d-bf4b60f8.us-east-1.amazon.com Tue May 25 17:08:42 UTC 2021 gluetod
Preparing ...
Thu Jun  3 17:27:07 UTC 2021
/usr/bin/java -cp /tmp:/opt/amazon/conf:/opt/amazon/lib/hadoop-lzo/*:/opt/amazon/lib/emrfs-lib/*:/opt/amazon/spark/jars/*:/opt/amazon/superjar/*:/opt/amazon/lib/*:/opt/amazon/Scala2.11/* com.amazonaws.services.glue.PrepareLaunch --conf
spark.dynamicAllocation.enabled=true --conf spark.shuffle.service.enabled=true --conf spark.dynamicAllocation.minExecutors=1 --conf spark.dynamicAllocation.maxExecutors=7 --conf spark.executor.memory=5g --conf spark.executor.cores=4 --conf
spark.driver.memory=5g --extra-py-files s3:///library/pycountry_convert.zip --JOB_ID j_2e4222d3823e714853c13bf1a215746b21d7c0956f664b2a227c04e80f010114 --JOB_RUN_ID jr_3475a228258dbf72fdafd9a584322f4a15eb9696232ad183d91d0c0ed4e1deca --s3_bucket
s3://glueworkshop-navnis/ --job-bookmark-option job-bookmark-disable --scriptLocation s3://glueworkshop-navnis/script/lab3/spark.py --JOB_NAME glueworkshop-lab3-etl-job

1622741227584

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/opt/amazon/spark/jars/slf4j-log4j12-1.7.16.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/opt/amazon/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/amazon/lib/Log4j-slf4j-2.x.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

Glue ETL Marketplace - Start ETL connector activation process...

Glue ETL Marketplace - no connections are attached to job glueworkshop-lab3-etl-job, no need to download custom connector jar for it.
Glue ETL Marketplace - Retrieved no ETL connector jars, this may be due to no marketplace/custom connection attached to the job or failure of downloading them, please scroll back to the previous logs to find out the root cause. Container setup
continues.
Glue ETL Marketplace - ETL connector activation process finished, container setup continues...

Exception in thread "main"

java.lang.IllegalArgumentException: Invalid S3 URI: no bucket: s3:///library/pycountry_convert.zip

        at com.amazonaws.services.s3.AmazonS3URI.<init>(AmazonS3URI.java:88)

        at com.amazonaws.services.s3.AmazonS3URI.<init>(AmazonS3URI.java:62)
        at com.amazonaws.services.s3.AmazonS3URI.<init>(AmazonS3URI.java:48)

        at com.amazonaws.services.glue.FileDownloader.downloadMultiple(FileDownloader.scala:122)
        at com.amazonaws.services.glue.PrepareLaunch$$anonfun$7$$anonfun$apply$8.apply(PrepareLaunch.scala:292)

        at com.amazonaws.services.glue.PrepareLaunch$$anonfun$7$$anonfun$apply$8.apply(PrepareLaunch.scala:292)

        at scala.collection.immutable.List.flatMap(List.scala:338)

        at com.amazonaws.services.glue.PrepareLaunch$$anonfun$7.apply(PrepareLaunch.scala:292)

        at com.amazonaws.services.glue.PrepareLaunch$$anonfun$7.apply(PrepareLaunch.scala:292)

        at scala.Option.map(Option.scala:146)
```

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ○ | glueworkshop-lab3-etl-job | Glue ETL | 06/03/2021 23:39:25 | 06/03/2021 23:40:18 | ✓ Succeeded | 1 minute | 4 | Standard | 0.07 |
| ○ | glueworkshop_lab4_glue_streaming | Glue Streaming | 06/03/2021 23:36:58 | 06/03/2021 23:37:52 | ⊗ Failed | 1 minute | 2 | G.1X | 0.03 |
| ○ | glueworkshop-lab3-etl-job | Glue ETL | 06/03/2021 23:33:06 | 06/03/2021 23:34:01 | ✓ Succeeded | 1 minute | 4 | Standard | 0.07 |
| ○ | glueworkshop-lab3-etl-job | Glue ETL | 06/03/2021 10:39:52 | 06/03/2021 10:40:47 | ✓ Succeeded | 1 minute | 4 | Standard | 0.07 |
| ○ | glueworkshop-lab3-etl-job | Glue ETL | 06/03/2021 10:26:59 | 06/03/2021 10:27:19 | ⊗ Failed | less than a minute | 4 | Standard | 0.07 |

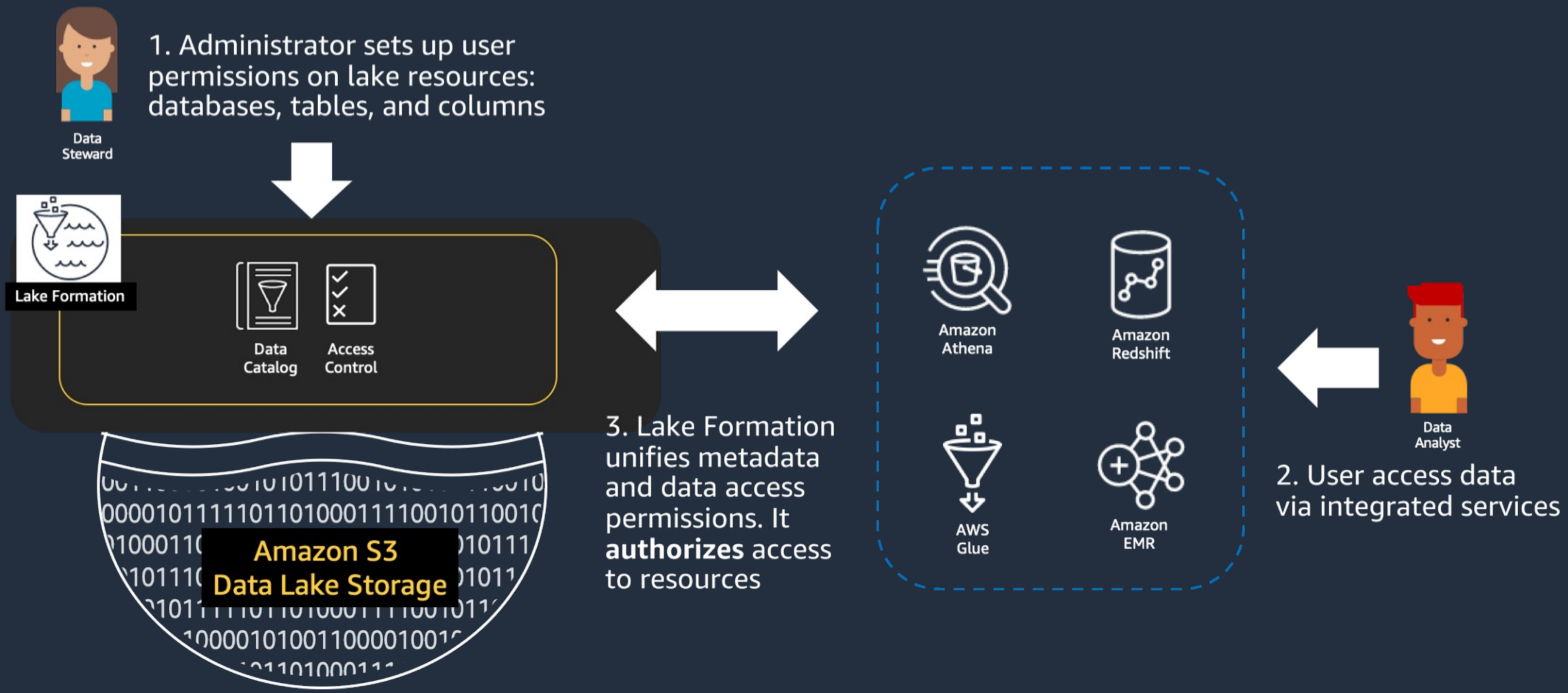
# Security

- Classify and protect data
- Control the data access
- Control the access to workload infrastructure

https://docs.aws.amazon.com/wellarchitected/latest/analytics-lens/security.html

# Control the access to workload infrastructure

## Centralized permissions



1. Administrator sets up user permissions on lake resources: databases, tables, and columns

2. User access data via integrated services

3. Lake Formation unifies metadata and data access permissions. It **authorizes** access to resources

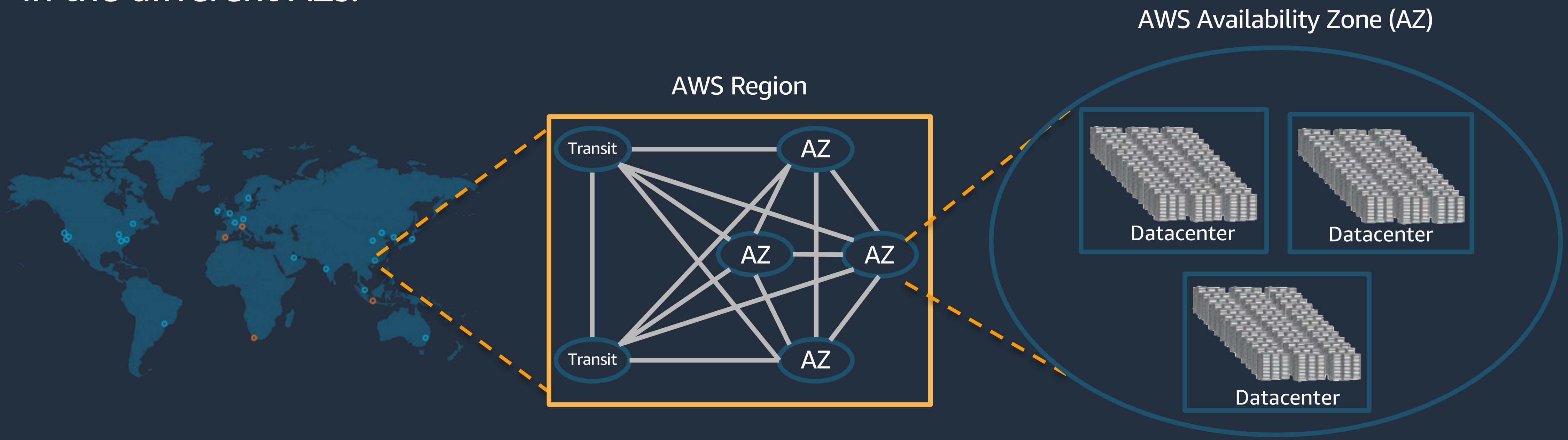
# Reliability

- Design resiliency for analytics workload
- Govern data and metadata changes

https://docs.aws.amazon.com/wellarchitected/latest/analytics-lens/reliability.html

# Design resiliency for analytics workload

AWS Regions are comprised of multiple AZs for high availability, high scalability, and high fault tolerance. Applications and data are replicated in real time and consistent in the different AZs.



**A Region** is a physical location in the world where we have multiple **Availability Zones.**

**Availability Zones** consist of one or more discrete data centers, each with redundant power, networking, and connectivity, housed in separate facilities.

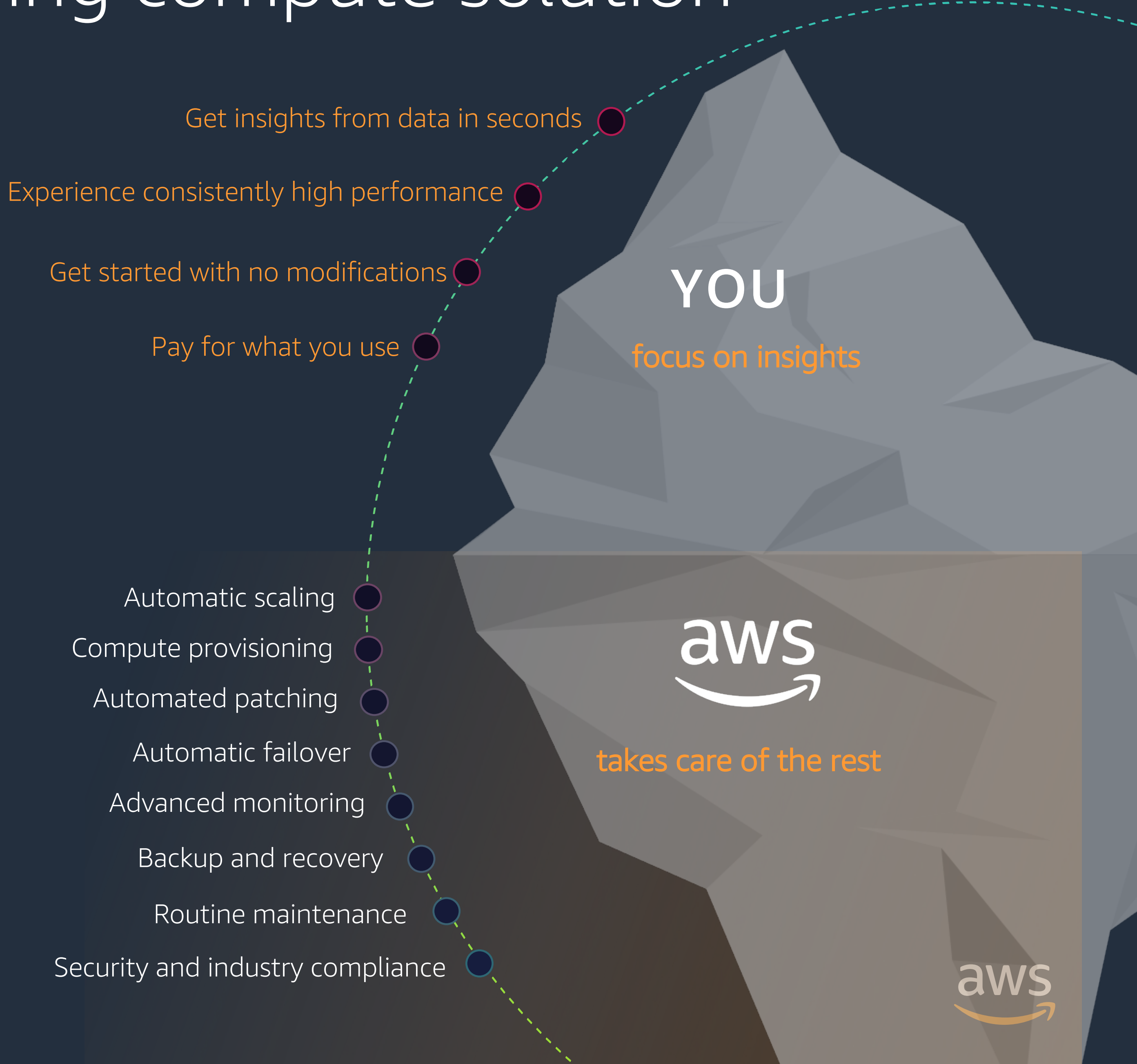
# Performance Efficiency

- Choose the best-performing compute solution
- Choose the best-performing storage solution
- Choose the best-performing file format and partitioning

https://docs.aws.amazon.com/wellarchitected/latest/analytics-lens/performance-efficiency.html

# Choose the best-performing compute solution

## Amazon Redshift Serverless



Get insights from data in seconds

Experience consistently high performance

Get started with no modifications

Pay for what you use

YOU

focus on insights

Automatic scaling

Compute provisioning

Automated patching

Automatic failover

Advanced monitoring

Backup and recovery

Routine maintenance

Security and industry compliance

aws

takes care of the rest

# Cost Optimization

- Choose cost-effective compute and storage solutions based on workload usage patterns

- Manage cost over time

- Use optimal pricing models based on infrastructure usage patterns

https://docs.aws.amazon.com/wellarchitected/latest/analytics-lens/cost-optimization.html

# Use optimal pricing models based on infrastructure usage patterns

| S3 Intelligent-Tiering | S3 Standard | S3 Standard-IA | **S3 Glacier Instant Retrieval** | S3 Glacier Flexible Retrieval | S3 Glacier Deep Archive |
|---|---|---|---|---|---|
| **Changing access patterns** | **Frequently accessed data** | **Infrequently accessed data** | **Rarely accessed data** | **Archive data** | **Long term archive data** |
| • Milliseconds access<br>• No retrieval charge<br>• **Archive Instant Access tier** | • Milliseconds access<br>• No retrieval charge | • Milliseconds access<br>• Per-GB retrieval charge | • Milliseconds access<br>• Per-GB retrieval charge | • Retrieval options from minutes to hours<br>• **Free bulk retrievals** | • Retrieval in hours |

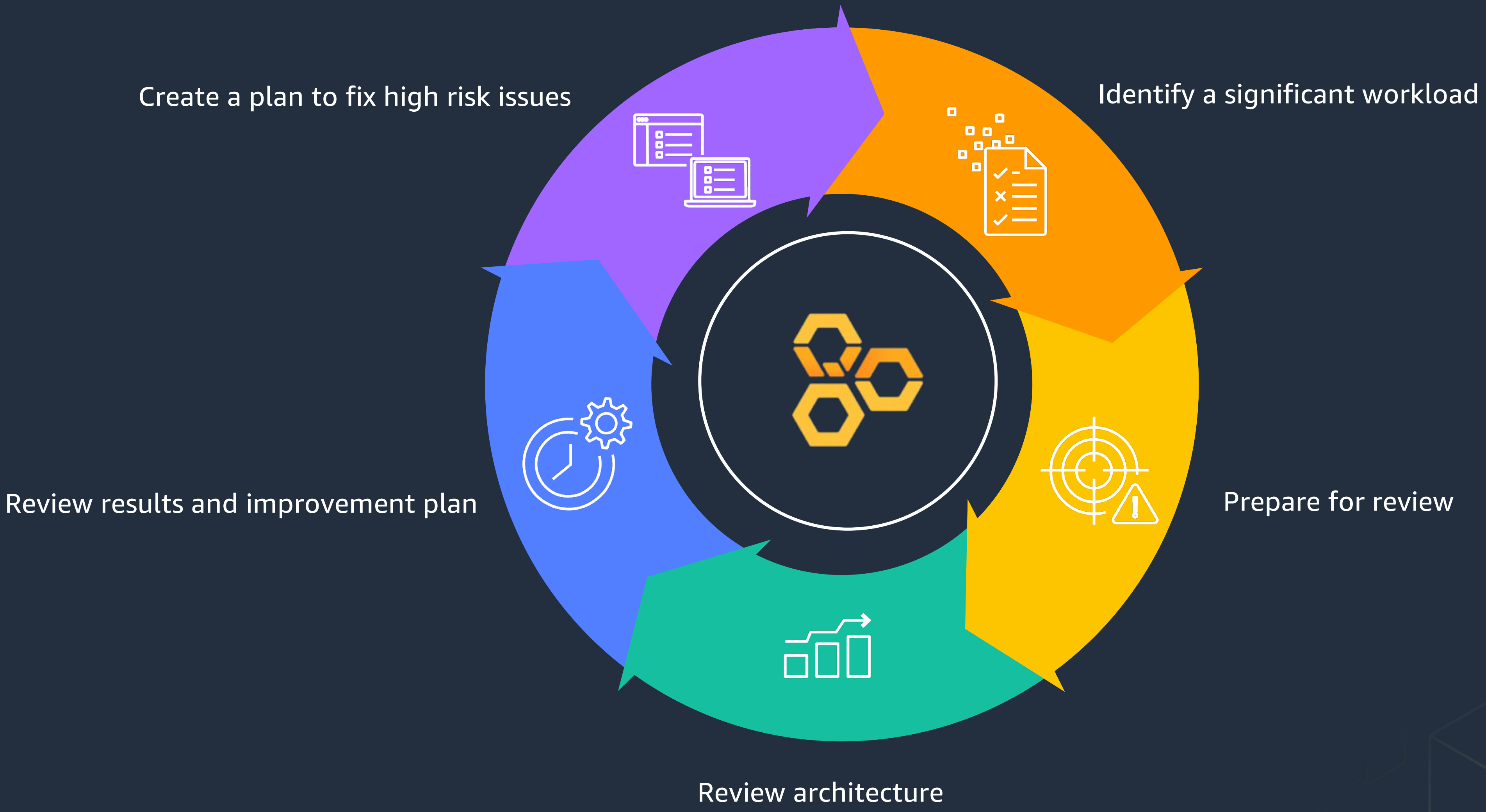
## AWS Region ≥ 3 Availability Zones

# Sustainability

To be released soon..

# Well-Architected Analytics Review

# Well-Architected Review Process



Identify a significant workload

Prepare for review

Review architecture

Review results and improvement plan

Create a plan to fix high risk issues

# Resumen

- Usar Well-Architected Framework – Analytics Lens como guía
- Aprovechar el review como self-assessment de cargas de trabajo
- Adecuar el análisis dependiendo del
    - Escenario
    - Casos de uso
    - Componentes
    - NFR

# Resources

- Well-Architected Framework (link)
- Data Analytics Lens white paper (link)
- Data Classification white paper (link)

- Analytics on AWS (link)
- Modern Data Architecture on AWS (link)

- AWS Global Infrastructure (link)
- AWS S3 Storage Classes (link)

SCAN ME

Carlos Paez – Solutions Architect
linkedin.com/in/carlospaez/

# Q&A

# ¡Gracias!