
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

SUS303

Delivering sustainable, high-performing architectures

Dr. Sam Mokhtari

Global lead, WA Sustainability Pillar
AWS

Thomas Coombs

Principal Technical Account Manager
AWS

Chie Hayashida

Data Analytics Specialist SA
AWS

aws

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

Agenda

- The urgency for environmental sustainability
- Amazon sustainability journey
- AWS as a technology partner
- Dive deep on delivering sustainable, high-performing architectures
- Modern data architecture
- Demo

aws

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

The urgency for environmental sustainability



IPCC: "Now or never" on 1.5°C warming limit

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

Source: https://www.ipcc.ch/report/sixth-assessment-report-working-group-3/

“**Technology** is a vital enabler of sustainability”

“Powered by these **twin engines** of growth, they’re **2.5x** more likely to outperform their peers.”

Accenture (April 2021, May 2022)



Source: https://www.accenture.com/us-en/insights/technology/uniting-technology-sustainability, https://www.accenture.com/us-en/blogs/technology-innovation/tech-sustainability-leadership

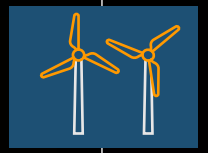
aws

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

Amazon sustainability journey



Sustainability report published annually since 2019



Amazon becomes largest corporate buyer of renewable energy globally



Ordered 100,000 Rivian electric delivery vehicles



HQ2: Building achieved LEED Platinum, highest certification for sustainable development



Reduced per-shipment packaging weight by 38% and eliminated over 1.5 million tons of packaging since 2015



Echo and Fire TV Low Power Mode, use of recycled materials, investment in renewable energy capacity equal to Echo, Fire TV, and Ring device use by 2025

Read more: https://sustainability.aboutamazon.com/

aws

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

Further and faster, together

September 19, 2019

Amazon and Global Optimism announced The Climate Pledge, a commitment to meet the goals of the Paris Agreement 10 years early—and achieve net-zero carbon by 2040. Amazon is a co-founder and first signatory of The Climate Pledge.





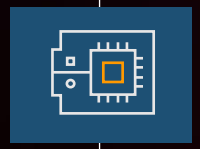


aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS sustainability journey



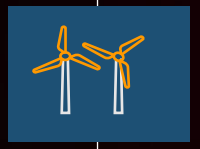
Water stewardship: evaporative cooling, recycled water, and onsite water treatment



Graviton processors provide better performance per watt than any other AWS processor



New data center construction incorporates use of low-carbon concrete



Renewable energy to power data centers



Launch of AWS Well-Architected Sustainability Pillar; launch of AWS customer carbon footprint tool



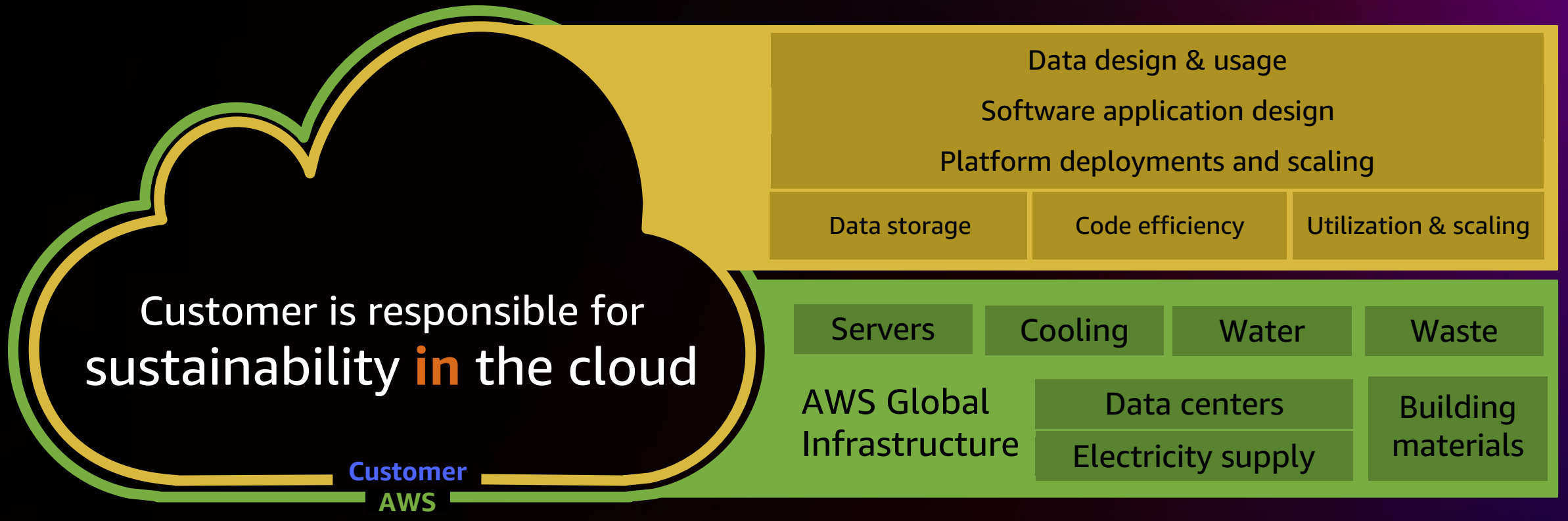
Amazon Sustainability Data Initiative (ASDI) provides free access to satellite data and climate models

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

AWS as your sustainability partner

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

Shared responsibility model with AWS



AWS is responsible for
sustainability of the cloud

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

Your sustainability journey and AWS



Migrate

Take advantage of the cloud and AWS efficiency



Transform

Leverage data, AWS professional services, and digital innovation



Optimize

Deliver sustainable, high-performing architecture

aws

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



Migrate to AWS

A carbon reduction opportunity

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

Carbon reduction opportunity

AWS can lower the carbon footprint of average on-premises data center workloads by nearly 80% today and up to 96% once AWS is powered with 100% renewable energy





Europe
Up to 5 times more energy efficient

US
3.6 times more energy efficient

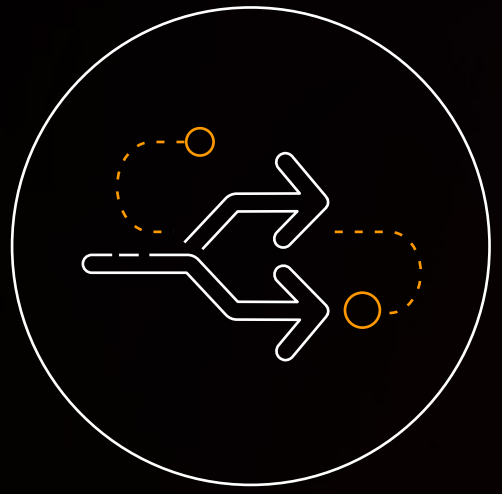


Asia Pacific
5 times more energy efficient

Find all the reports on https://aws.amazon.com/sustainability/resources

aws

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



Transform

Access and leverage sustainability data
AWS Professional Services Sustainability Practice
Build cloud-native solutions

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

Build sustainability solutions



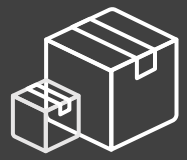
Carbon footprint framework



Sustainable building management



Energy analytics and forecasting



Sustainable packaging

Tap into industry-specific expertise, including:



Agriculture



Energy



Retail, logistics, and consumer packaged goods

aws

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

Sustainability data programs

Open Data Program

Amazon Sustainability Data Initiative (ASDI)

AWS Data Exchange

Data to address a wide range of sustainability challenges



Image from Landsat 8 satellite, courtesy of the U.S. Geological Survey

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

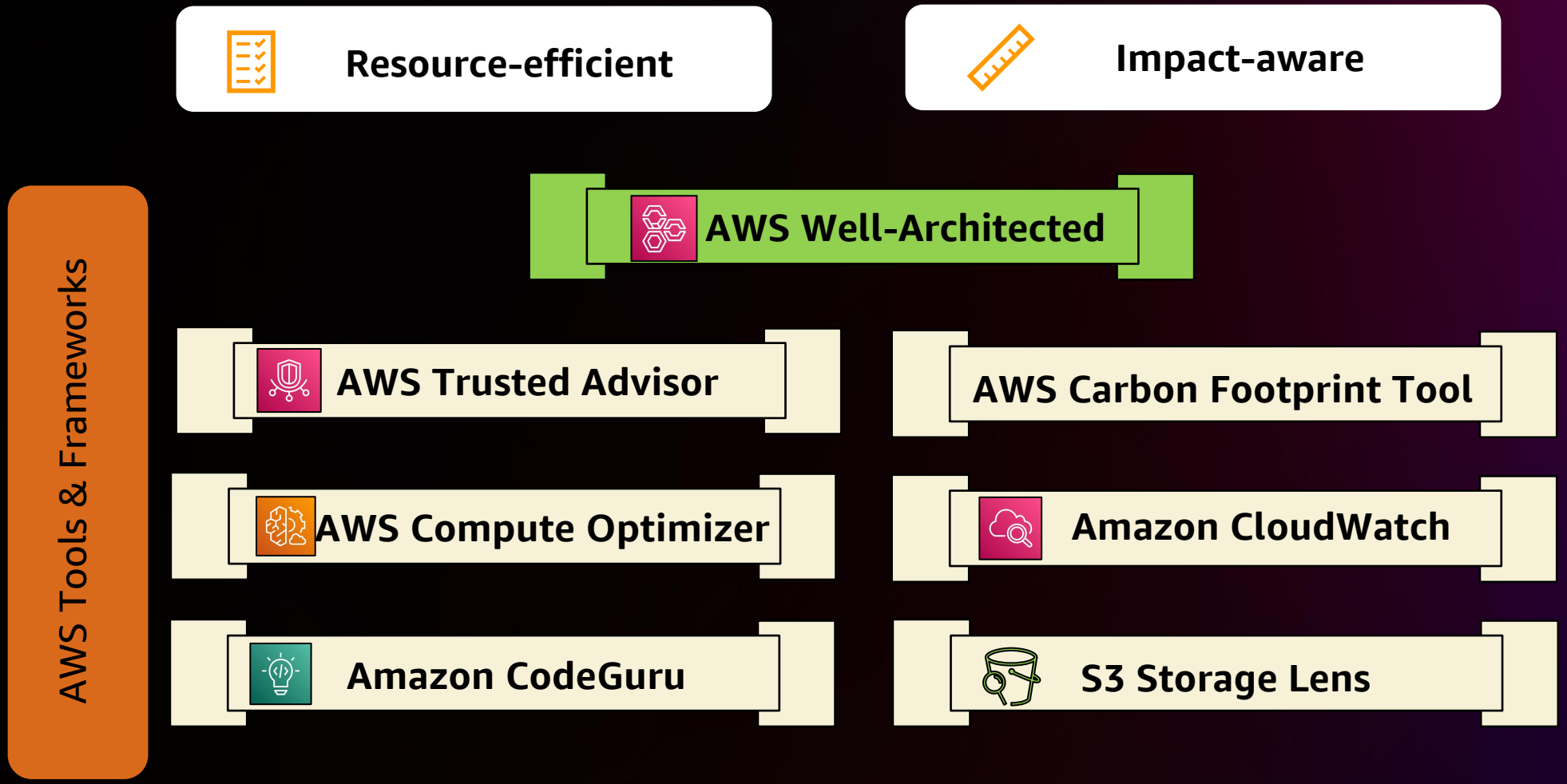


Optimize workloads

To deliver sustainable, high-performing architecture

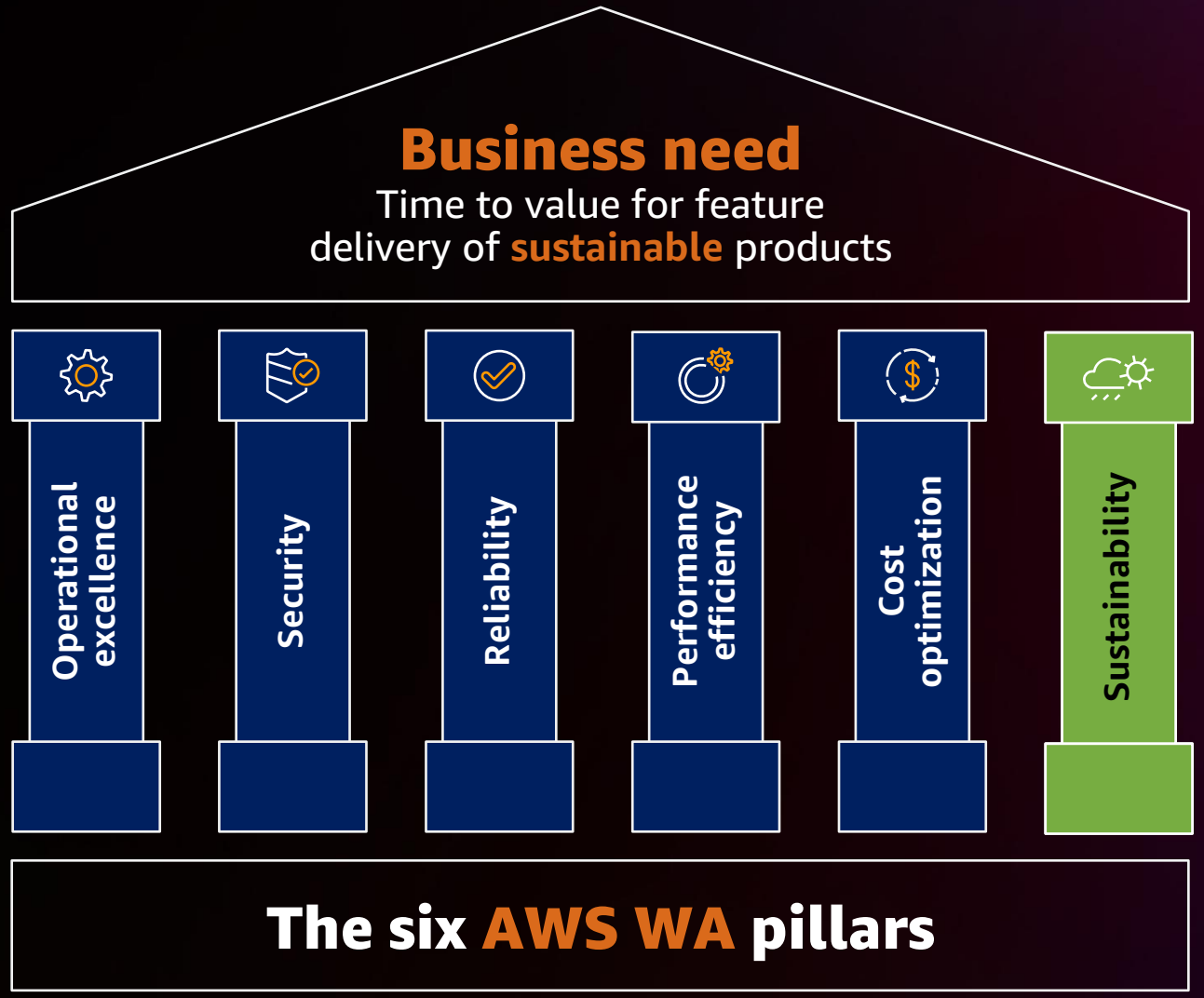
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Sustainable high-performing architectures



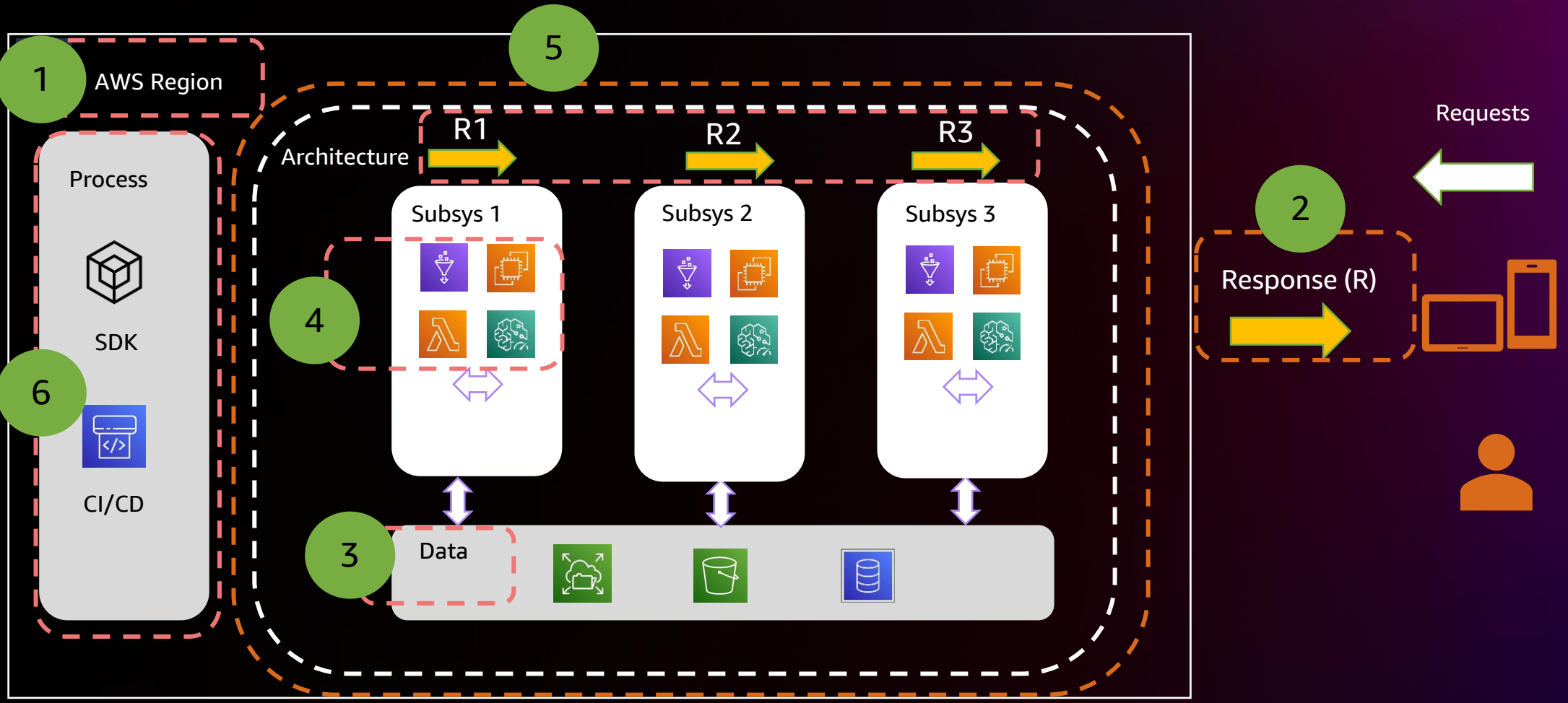
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Well-Architected



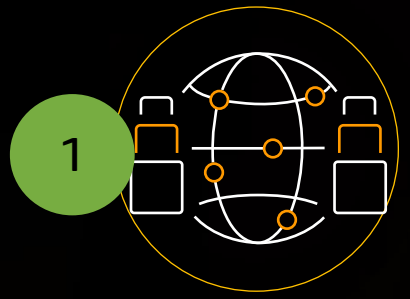
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

A typical cloud workload

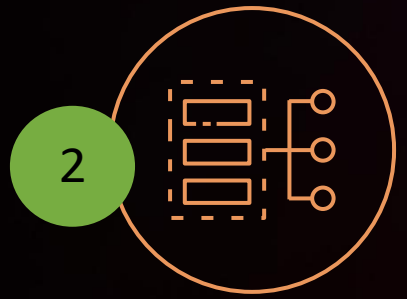


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Focus domains of AWS WA Sustainability pillar



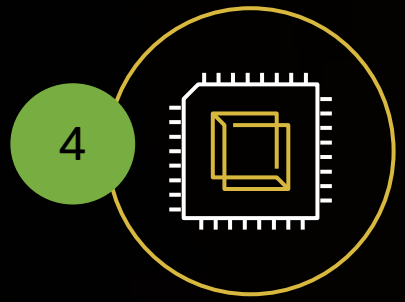
Region selection



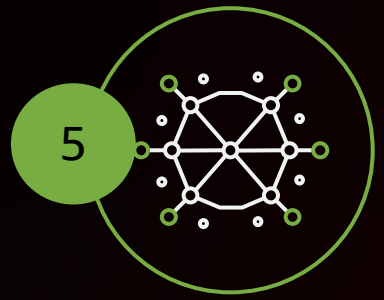
User behavior



Software & architecture



Hardware



Data



Development & deployment process

aws

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

Region selection

1. Assess and shortlist potential Regions for your workload based on your business requirements (e.g., compliance, latency, cost, and features)

2. Choose Region(s) near Amazon renewable energy projects or Regions where the grid has a published carbon intensity that is lower than others

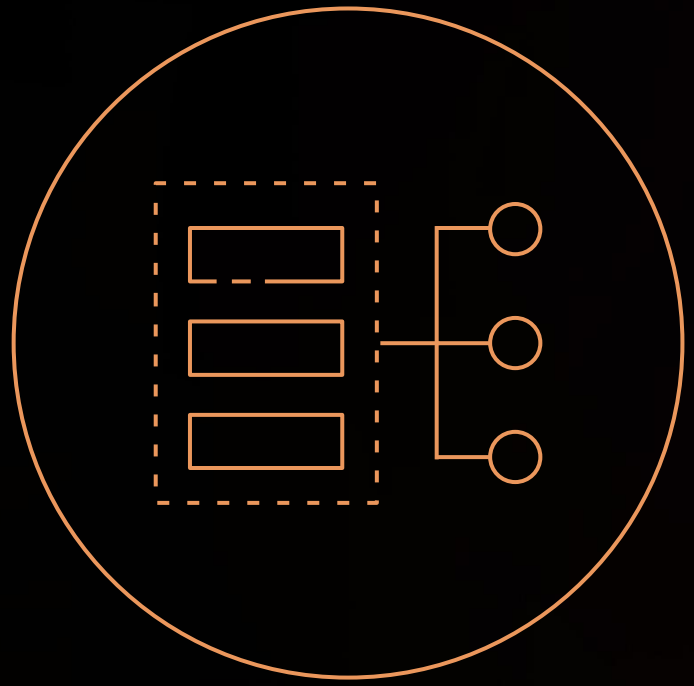
AWS Regions powered by over 95% renewable energy

https://sustainability.aboutamazon.com/environment/the-cloud?energyType=true

aws

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

User behavior

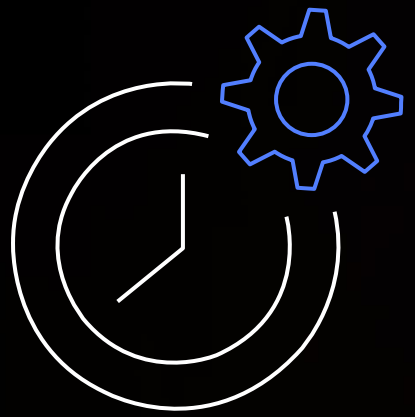


- Scale infrastructure with user load
- Optimize resources used by users for the activities performed
- Align service level agreements (SLAs) with sustainability goals

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

Architectural recommendation

NEGOTIATE IMPACT-FRIENDLY SLAS



Small increase
in response time

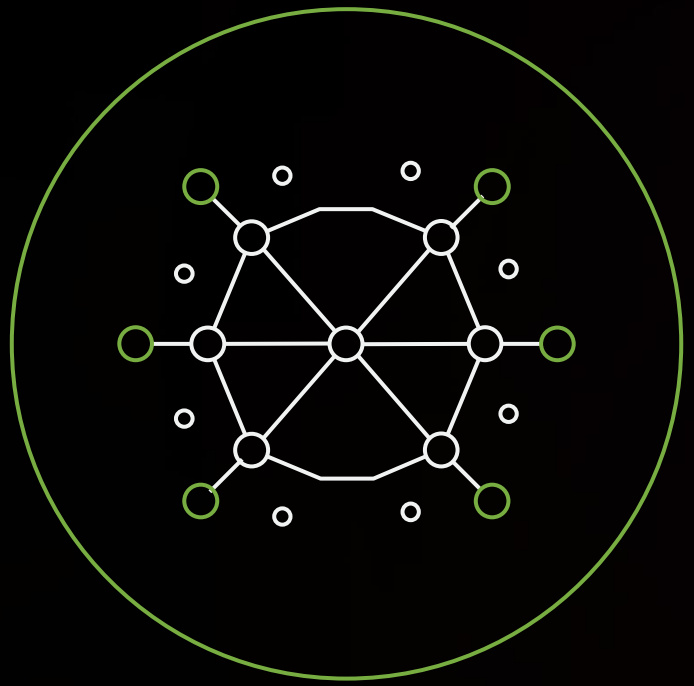


Reduction in
overall impact

aws

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

Data

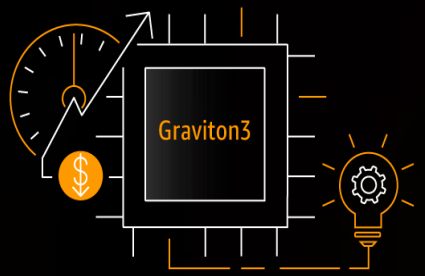


- Implement a data classification policy
- Use lifecycle policies to automatically delete unnecessary data
- Minimize data movement across networks
- Only back up data when it would be more impactful, or impossible, to re-create

aws

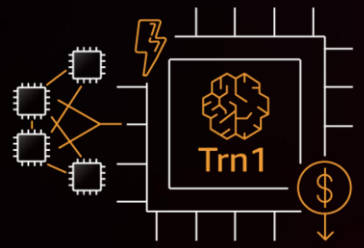
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use hardware with least impact



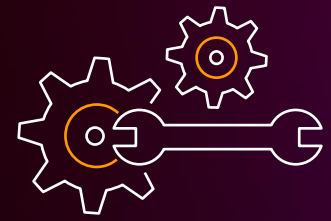
Switch to Graviton

Powerful and efficient, modern applications



Leverage specialized hardware

e.g., AWS Inferentia and AWS Trainium



Right-size resources

e.g., AWS Compute Optimizer

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

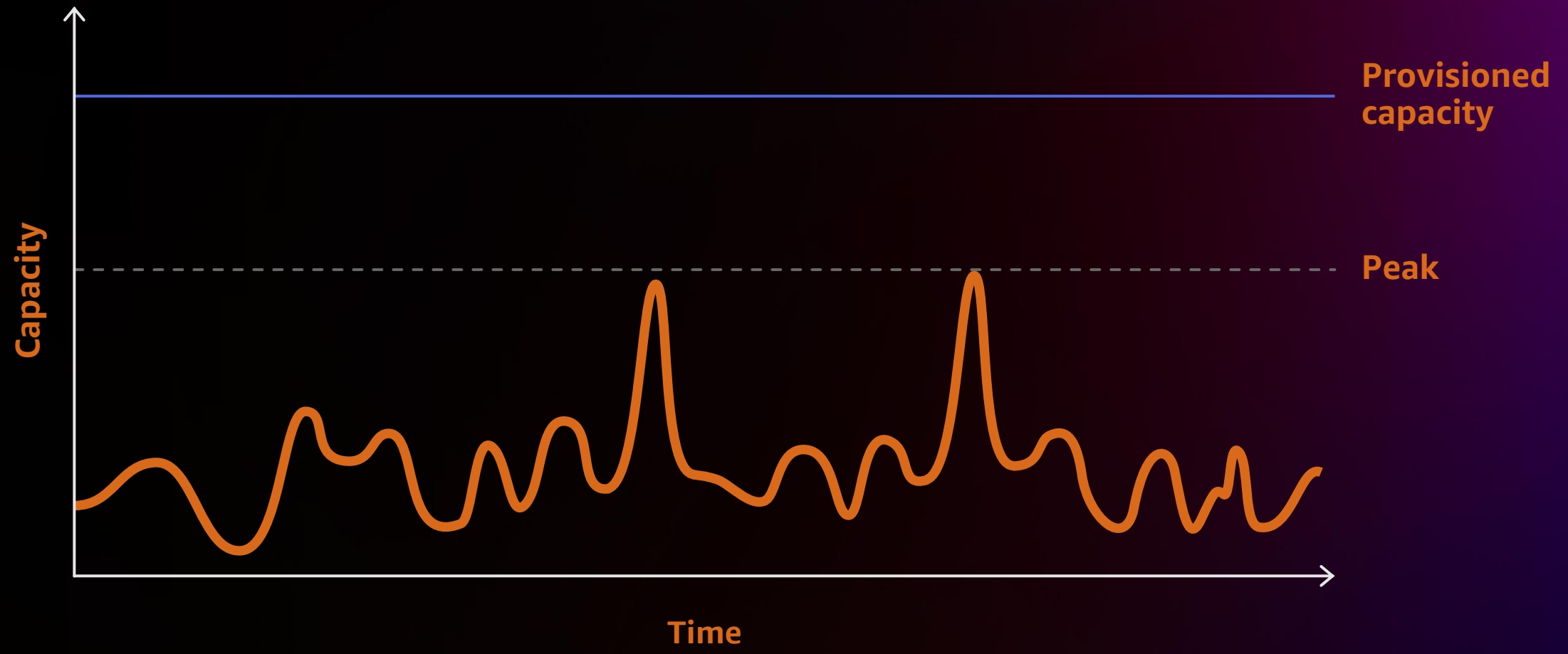
Software and architecture



- Remove or refactor workload components with low or no use
- Use software patterns and architectures that best support your data access and storage
- Enable sustainable scheduling in architecture

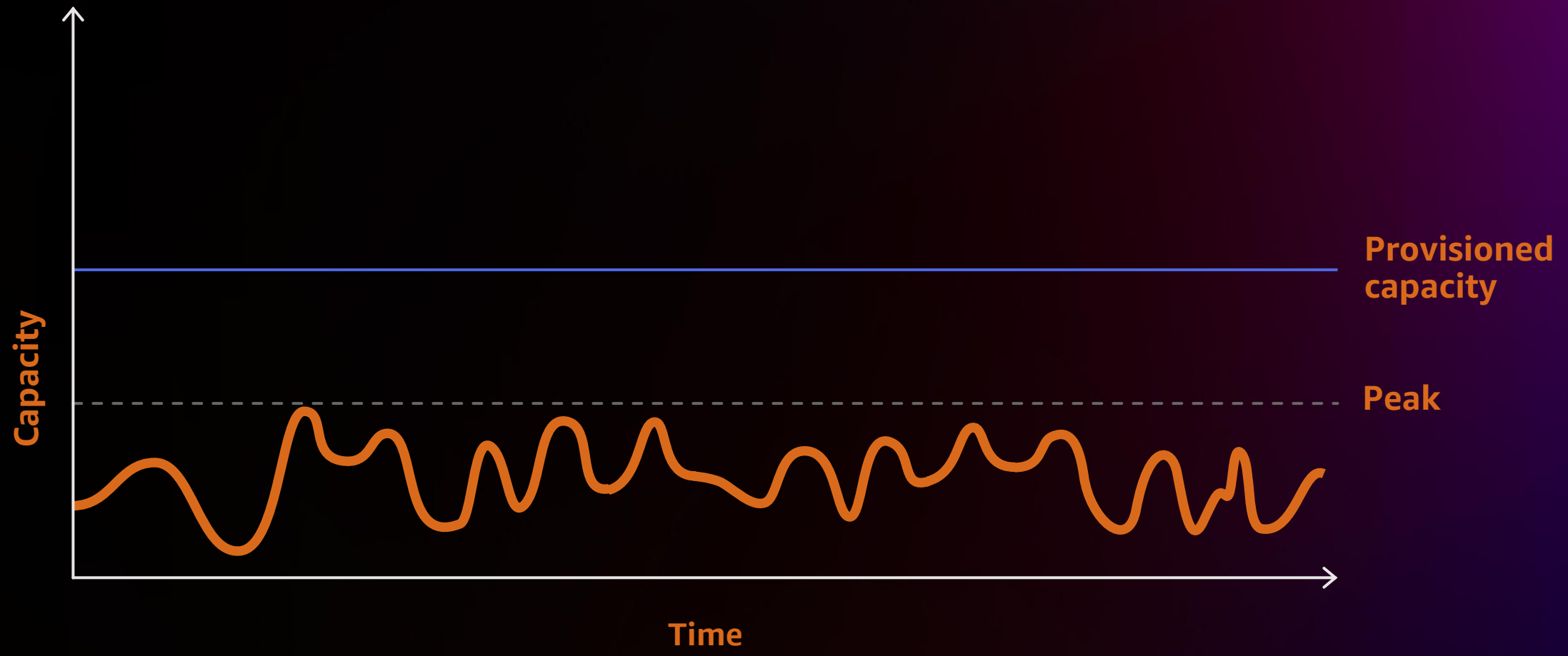
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Sustainable scheduling



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

Sustainable scheduling



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

Development and deployment process



Adopt more automation



Increase the utilization of build environments



Keep your workload up to date

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

AWS tooling to improve resource efficiency



AWS Compute Optimizer

Finds under-utilized compute resources



AWS Trusted Advisor

Makes recommendations to improve performance



AWS CodeGuru

Finds your least performant lines of code

aws

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

Why should architectures be impact-aware?

Improvement Process

Review | Identify → Prioritize | Validate → Deploy | Replicate

Improvement Goal


AWS Well-Architected


AWS Compute Optimizer


Amazon CodeGuru


AWS Trusted Adviser


AWS customer carbon footprint tool


Amazon CloudWatch


AWS Cost and Usage Report


Amazon S3 Storage Lens


Amazon Athena


Amazon QuickSight

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

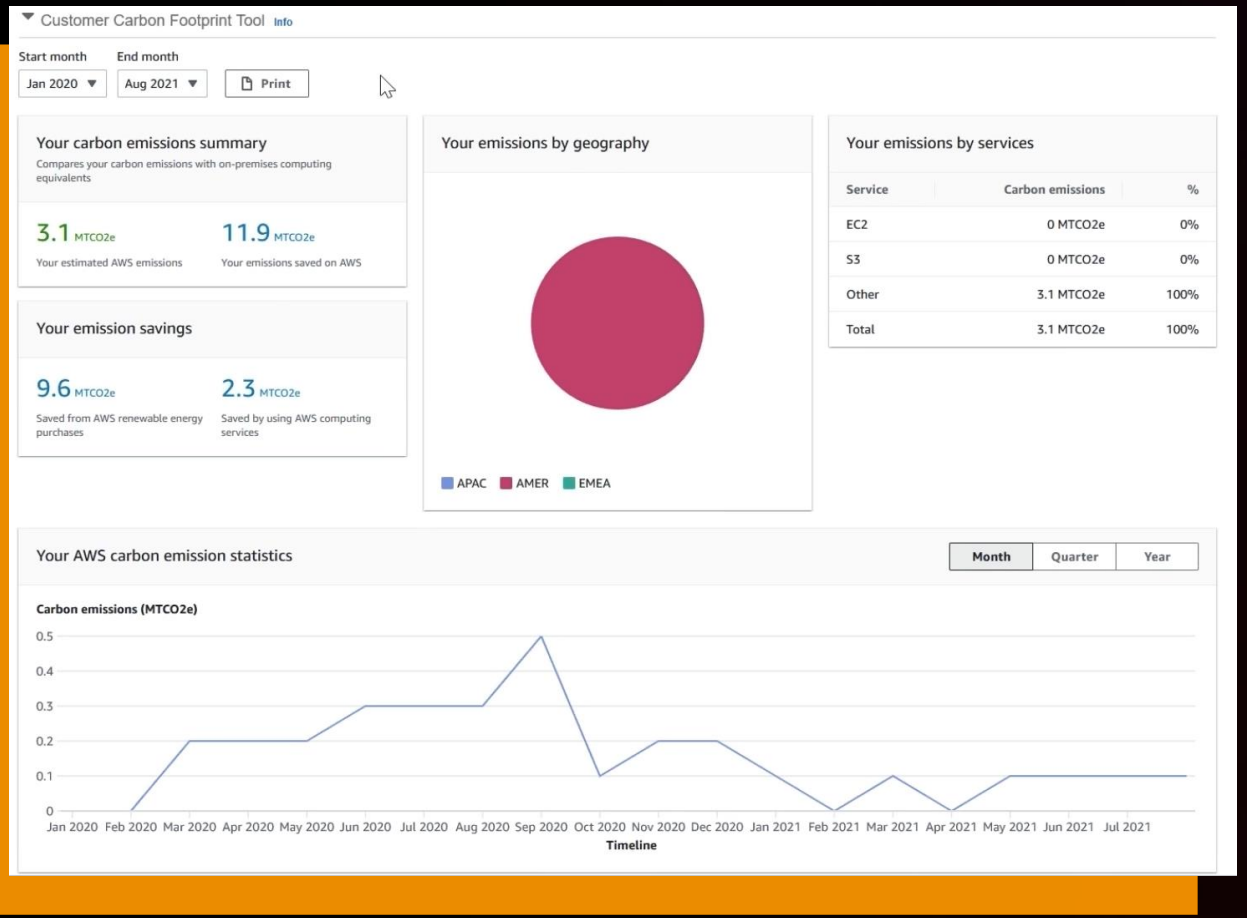
AWS customer carbon footprint tool



Track, measure, review, and **forecast** the carbon emissions generated from your AWS usage

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

AWS customer carbon footprint tool



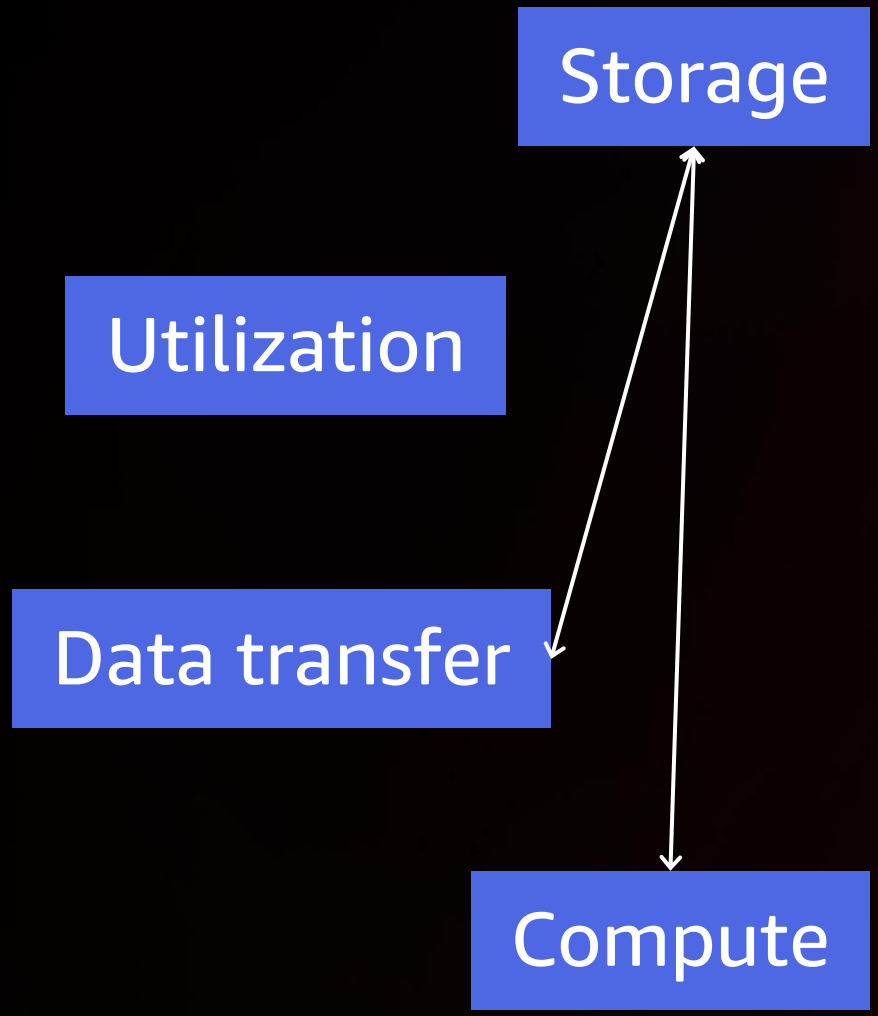
Calculate carbon emissions generated from your AWS workloads

Understand historical carbon footprint and review changes in emissions over time

Forecast changes as AWS moves forward on a path to 100% renewable power in data centers by 2025

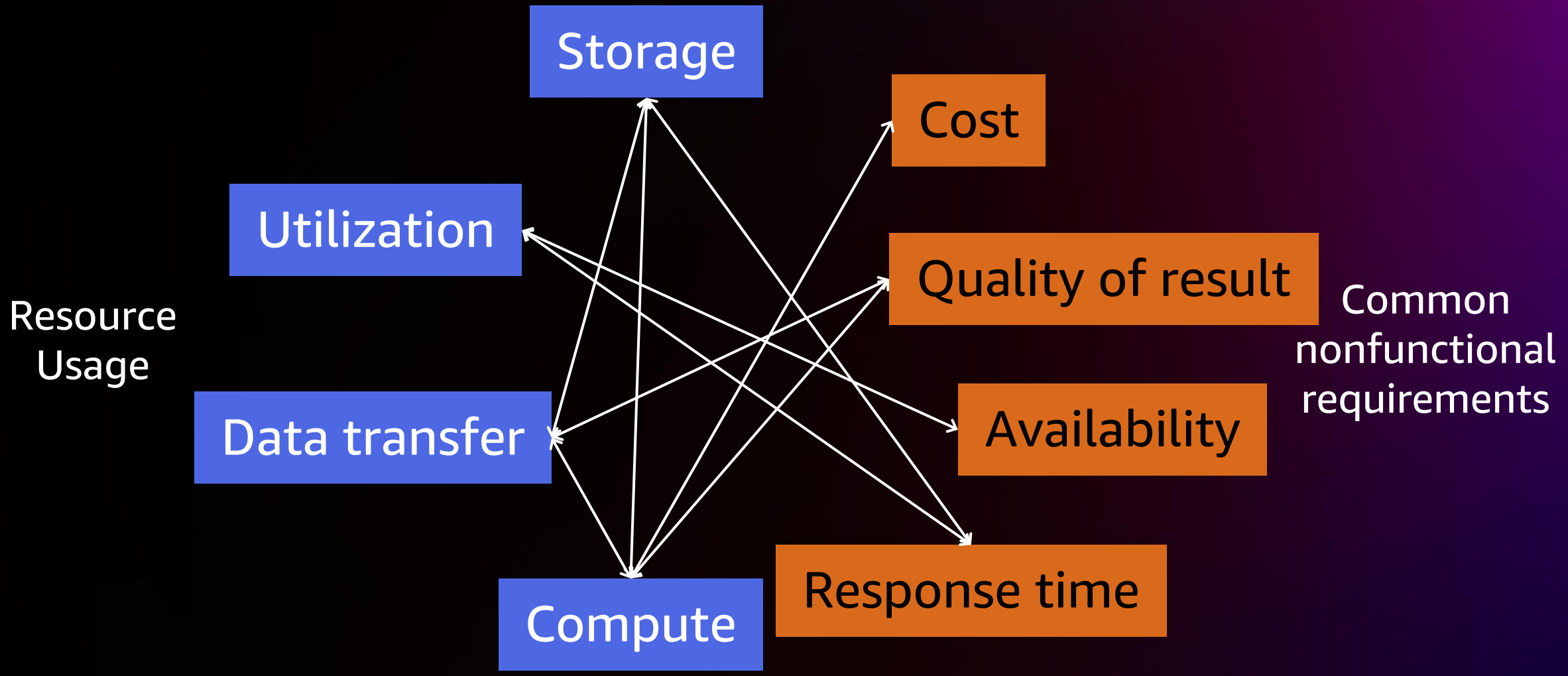
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Understand your impact through metrics



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

Sustainability is a nonfunctional requirement



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

$$KPI = \frac{\text{Proxy Metric}}{\text{Business Metric}}$$

Normalize sustainability KPIs to make them comparable

aws

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

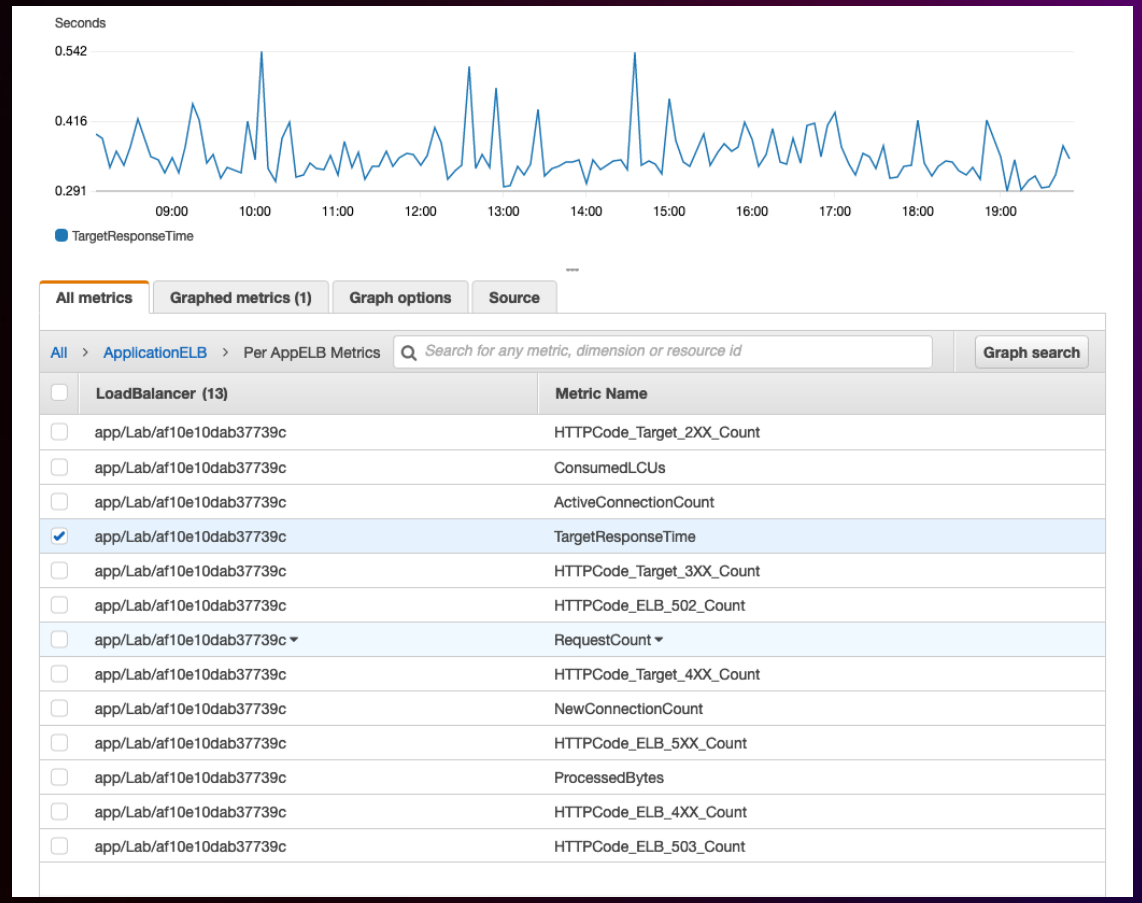
CloudWatch metrics

Collecting metrics is time consuming; Amazon CloudWatch allows you to collect default metrics from more than 70 AWS services, such as

- Amazon EC2
- Amazon DynamoDB
- Amazon S3
- Amazon ECS
- AWS Lambda
- Amazon API Gateway

CloudWatch Insights to get metric from logs

Custom metric for business metrics

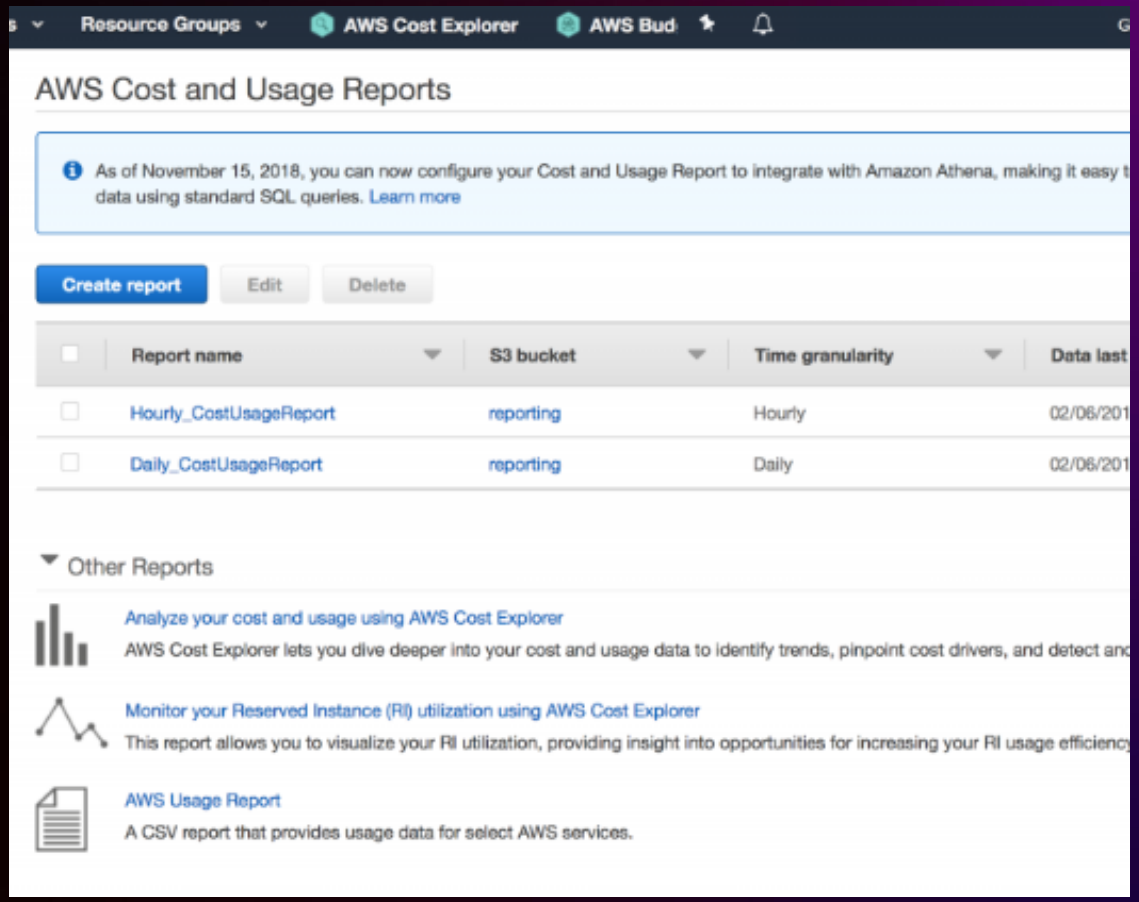


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cost and Usage Report

AWS Cost and Usage Reports (AWS CUR) contain the most comprehensive set of cost and usage data available

- Deliver report files to your Amazon S3 bucket
- Update the report up to three times a day
- Create, retrieve, and delete your reports using the AWS CUR API Reference
- We can use it as a data source for proxy metrics



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

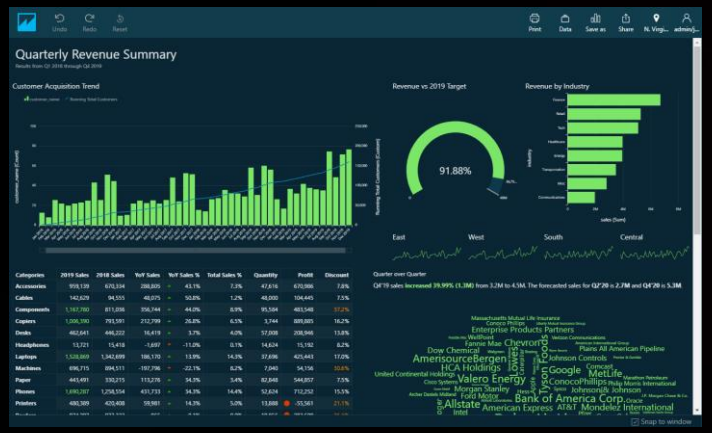
Architect workload to be impact-aware with metrics

Resource	Proxy metrics	Source
Compute	• vCPU minutes per transaction; normalized vCPU • Example: vCPU in CloudWatch*	CloudWatch or Cost and Usage
Storage	• GB per transaction • Example: BucketSizeBytes	CloudWatch
Network	• GB transferred or packets transferred • Example: NetworkIn in CloudWatch	CloudWatch/network logs (AWS Transit Gateway logs, for example)

vCPU in CloudWatch: https://www.youtube.com/watch?v=6HyhV26S_a0

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Evaluate improvement



Amazon QuickSight



Amazon Managed Grafana



AWS CloudWatch dashboard

Dashboards, visualizations, and purpose-built analyses to make sustainable decisions on best actions and improvement

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

Establish fine-grained, frequent KPIs at team level



Business metrics

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

What are the cloud intelligence dashboards?

A series of in-depth and customizable dashboards which provide the most comprehensive cost and usage details to help you optimize cost, track usage goals, and achieve operational excellence

Easy to use

- Self service
- Auto generated recommendations
- Actionable insights for entire organization in a single place
- Multi-payer customer support

Secure

- No agents running
- Data stays in your organization
- All AWS native services

In-depth

- Resource-level granularity
- Fully customizable
- ML-driven Insights

Cost-efficient

- Serverless: Pay as you go and scale on demand
- No additional charges – pay only for underlying services Amazon S3, Athena, QuickSight
- Estimated cost in a large AWS deployment less than $100 per month



Deploy today using the Cloud Intelligence Well-Architected Labs

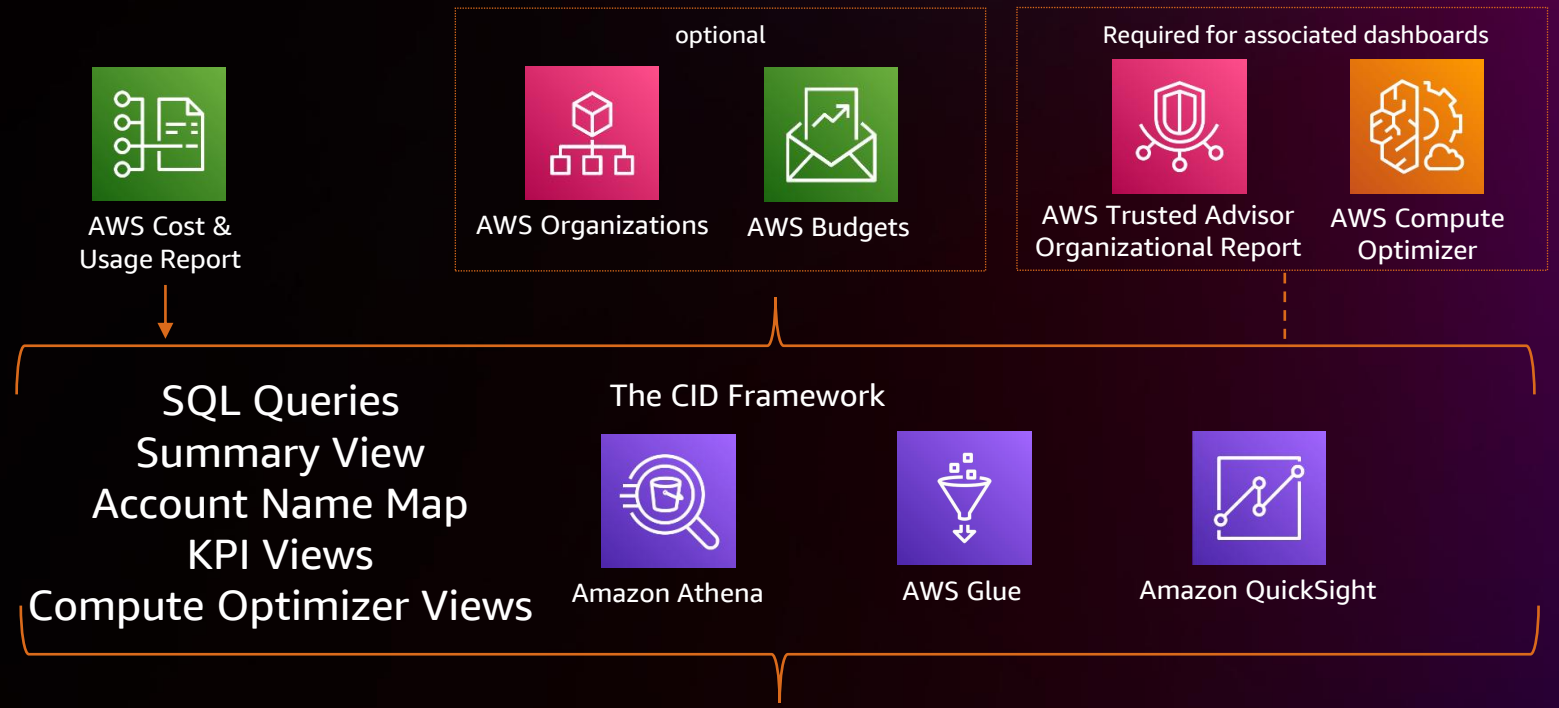
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The cloud intelligence dashboards framework

1. Start with your Cost and Usage Reports

2. Set up the framework

3. Access your dashboards





The Cost Intelligence Dashboard



CUDOS Dashboard



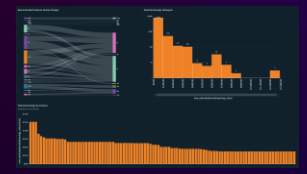
KPI Dashboard



Trends Dashboard



Trusted Advisor Organizational (TAO) Dashboard



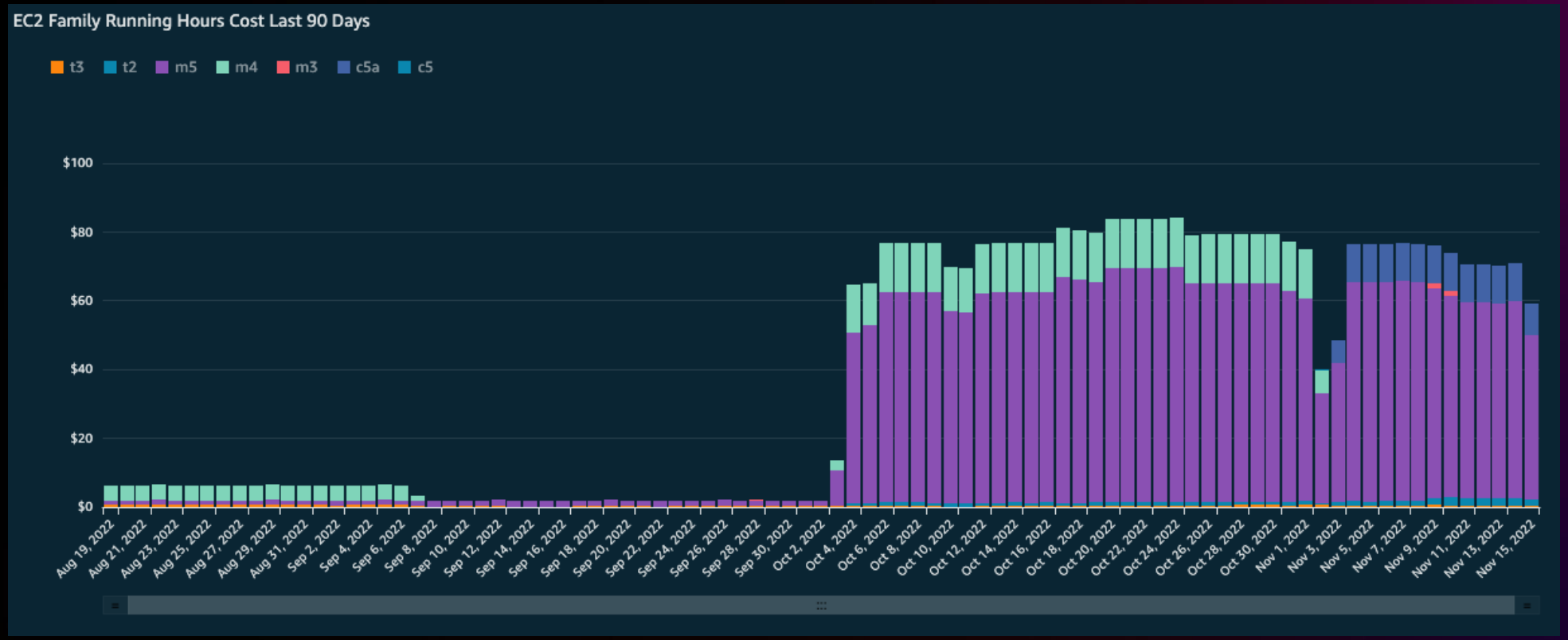
Compute Optimizer Dashboard

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

Cloud intelligence dashboards customization

The aim is to update this graph:

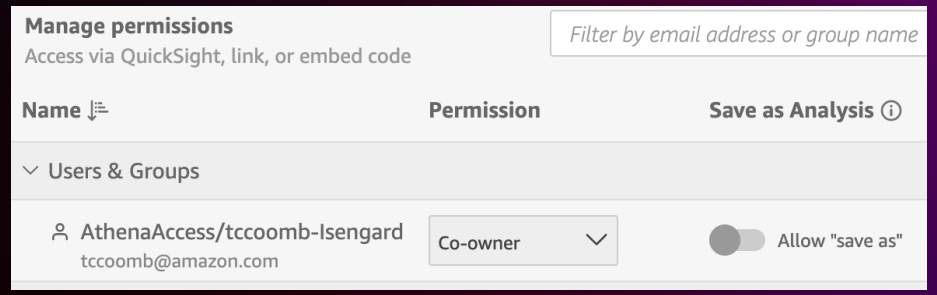
1) CPU usage, not cost
2) Scale the Y axis by a business KPI



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

Cloud intelligence dashboard customization

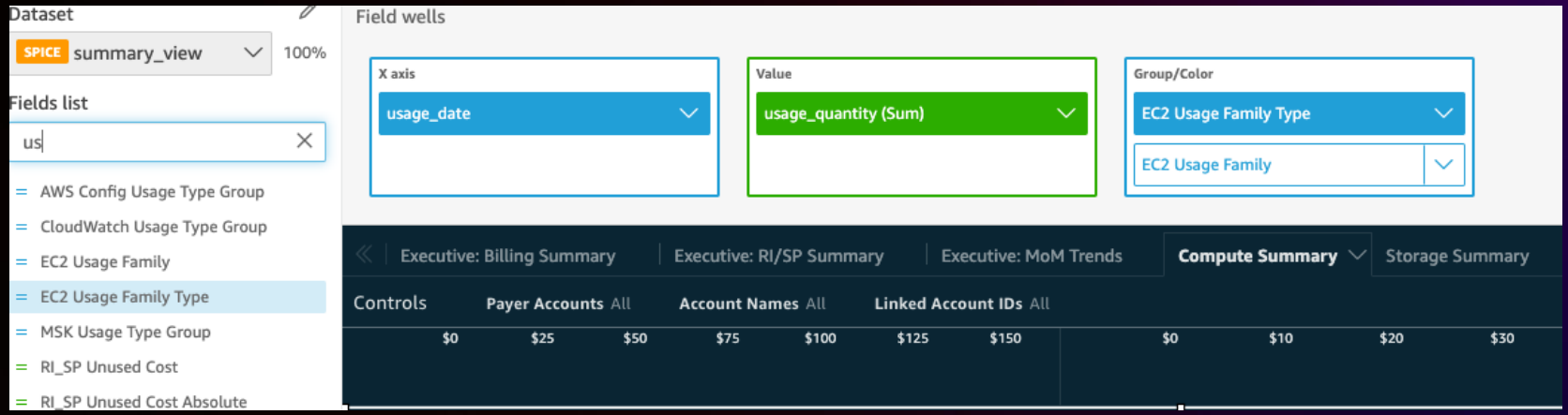
1) In the dashboard, select **share** to bring up the manage permissions page. Give **save as** permission



2) In dashboard settings; then **save as** at the top right. Save as a new analysis



3) Switch out the **cost** field for **usage_quantity** field

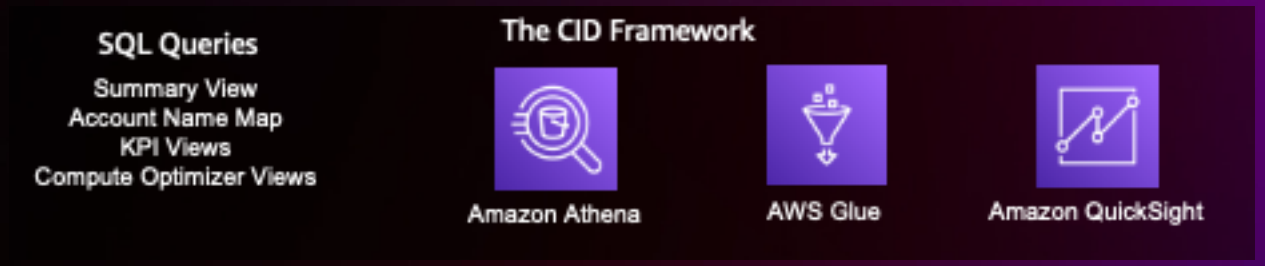


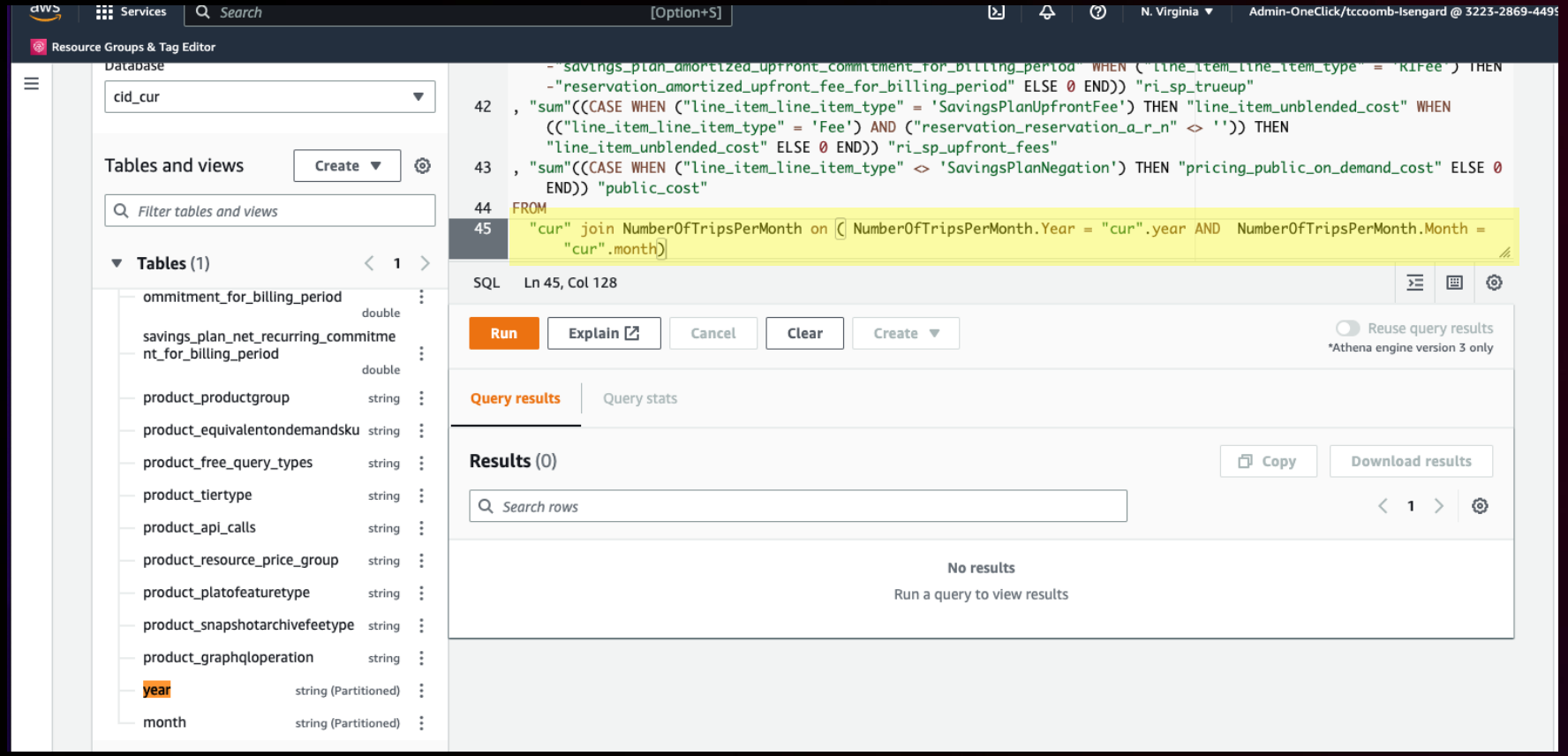
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cloud intelligence dashboard customization

The final aim is to scale the Y axis by a unit of work

We need to change the **summary view** in Athena





We join on the business data, adding that to the dataset

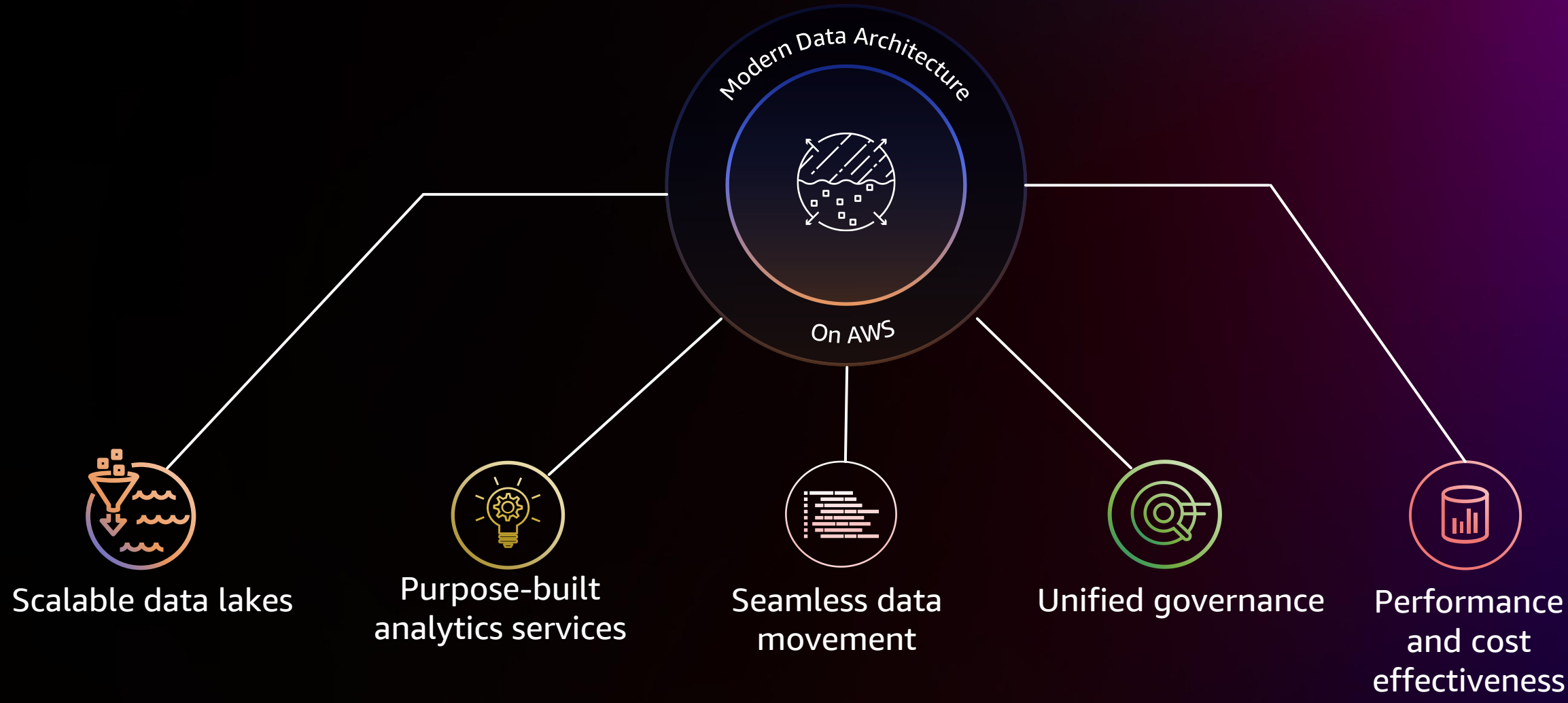
Then change the select to divide the resource value by this

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

Modern data architecture

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

Pillars within modern data architecture



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

AWS Glue

SERVERLESS DATA INTEGRATION FOR COMPLEX WORKLOADS USING APACHE SPARK



Serverless

No infrastructure to maintain; allocate needed compute power and run jobs

Data integration for every user

Development environments catered to different skillsets – visual ETL development for data engineers, notebook styled development for data scientists, and no code development for data analysts

Cost-effective

All-in-one pricing model is 55% cheaper than other cloud data integration options

Handles complex workloads

Connect to 30+ data sources, process petabytes of data in real time; includes batch- and event-driven modes

No lock-in

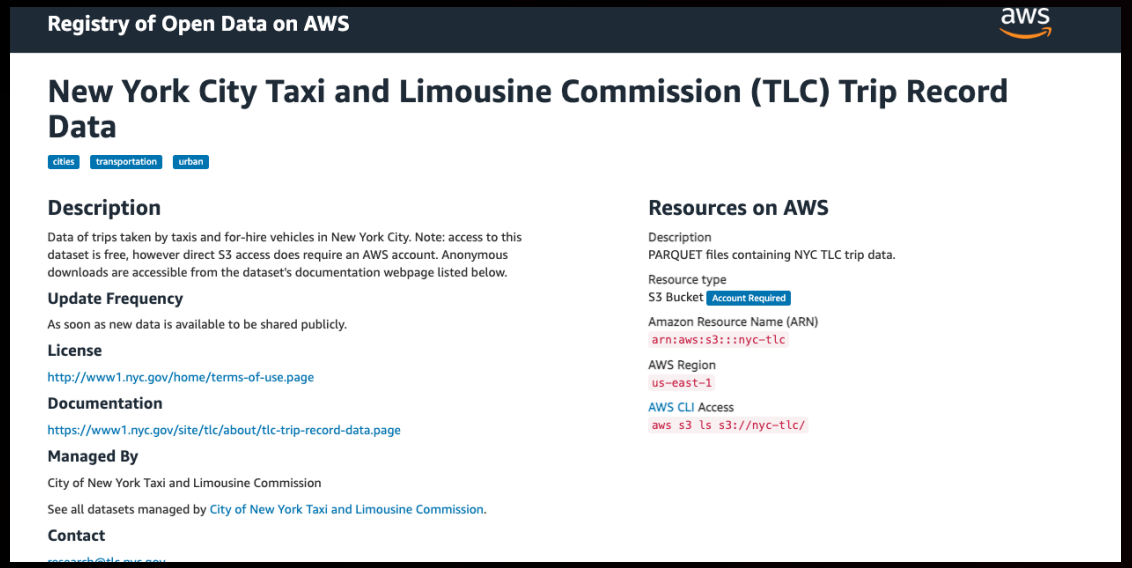
Develop data integration pipelines in open source SparkSQL, PySpark, and Scala

aws

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

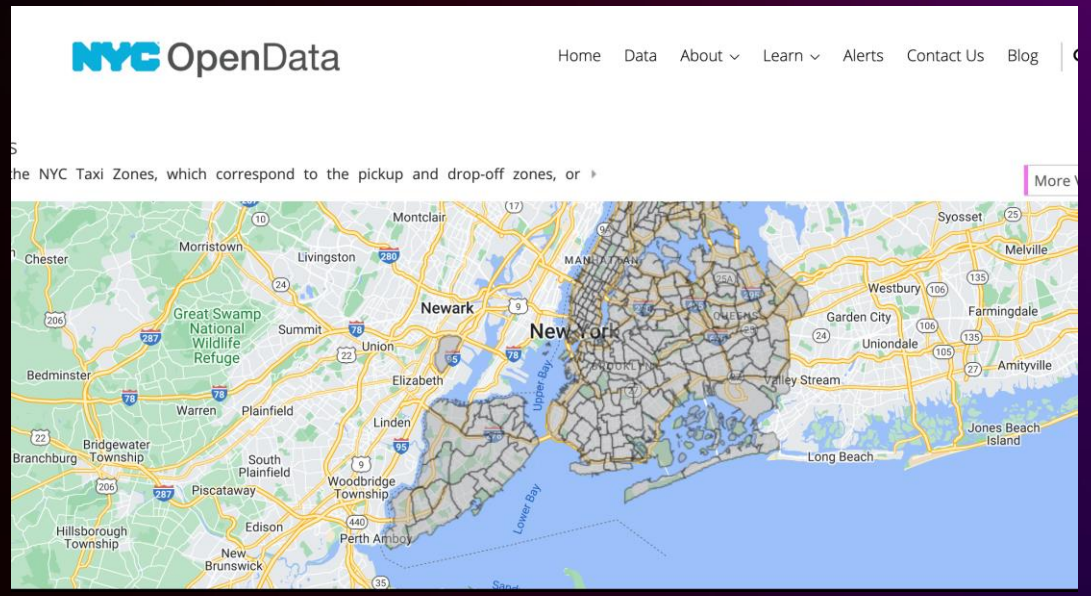
We will Use NYC taxi trip data

Taxi trip data



https://registry.opendata.aws/nyc-tlc-trip-records-pds/

Taxi zone mapping data



https://data.cityofnewyork.us/Transportation/NYC-Taxi-Zones/d3c5-ddgc

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

Data transformation in AWS Glue

Amazon S3

Taxi trip data

#	vid	pickedup_at	dropped_at	trip_distance	fare	tip
1	2	2016-09-19 15:11:09.000	2016-09-19 15:22:30.000	0.95	8.5	0.0
2	2	2016-09-19 15:24:02.000	2016-09-19 15:48:44.000	2.96	16.5	1.0
3	2	2016-09-19 15:52:36.000	2016-09-19 16:17:25.000	2.43	16.5	0.0
4	2	2016-09-19 15:19:46.000	2016-09-19 15:27:10.000	1.27	6.5	0.0
5	2	2016-09-19 15:27:59.000	2016-09-19 15:32:56.000	0.9	5.5	1.26
6	2	2016-09-19 15:35:52.000	2016-09-19 15:42:24.000	0.74	6.0	1.36
7	2	2016-09-19 15:45:28.000	2016-09-19 15:49:10.000	0.75	4.5	1.06
8	2	2016-09-19 15:50:26.000	2016-09-19 15:57:52.000	0.97	6.5	0.0
9	2	2016-09-19 15:59:44.000	2016-09-19 16:06:09.000	1.51	7.0	1.56
10	1	2016-09-19 15:41:04.000	2016-09-19 16:01:31.000	0.9	13.0	0.0

Taxi zone mapping data

#	objectid	shape_leng	the_geom	shape_area
1	1	0.116357453189	"MULTIPOLYGON (((-74.18445299999996 40.694995999999904	
2	2	0.43346966679	"MULTIPOLYGON (((-73.82337597260663 40.63898704717672	
3	3	0.0843411059012	"MULTIPOLYGON (((-73.84792614099985 40.87134223399991	
4	4	0.0435665270921	"MULTIPOLYGON (((-73.97177410965318 40.72582128133706	
5	5	0.0921464898574	"MULTIPOLYGON (((-74.17421738099989 40.5625680859999	
6	6	0.150490542523	"MULTIPOLYGON (((-74.06367318899999 40.60219816599994	
7	7	0.107417171123	"MULTIPOLYGON (((-73.90413637799996 40.767520316999864	
8	8	0.0275906911574	"MULTIPOLYGON (((-73.92334041500001 40.775128911999936	
9	9	0.0997840924705	"MULTIPOLYGON (((-73.78502434699996 40.76103651599987	
10	24	0.0469999619287	"MULTIPOLYGON (((-73.95953658899997 40.7987185259999	

AWS Glue

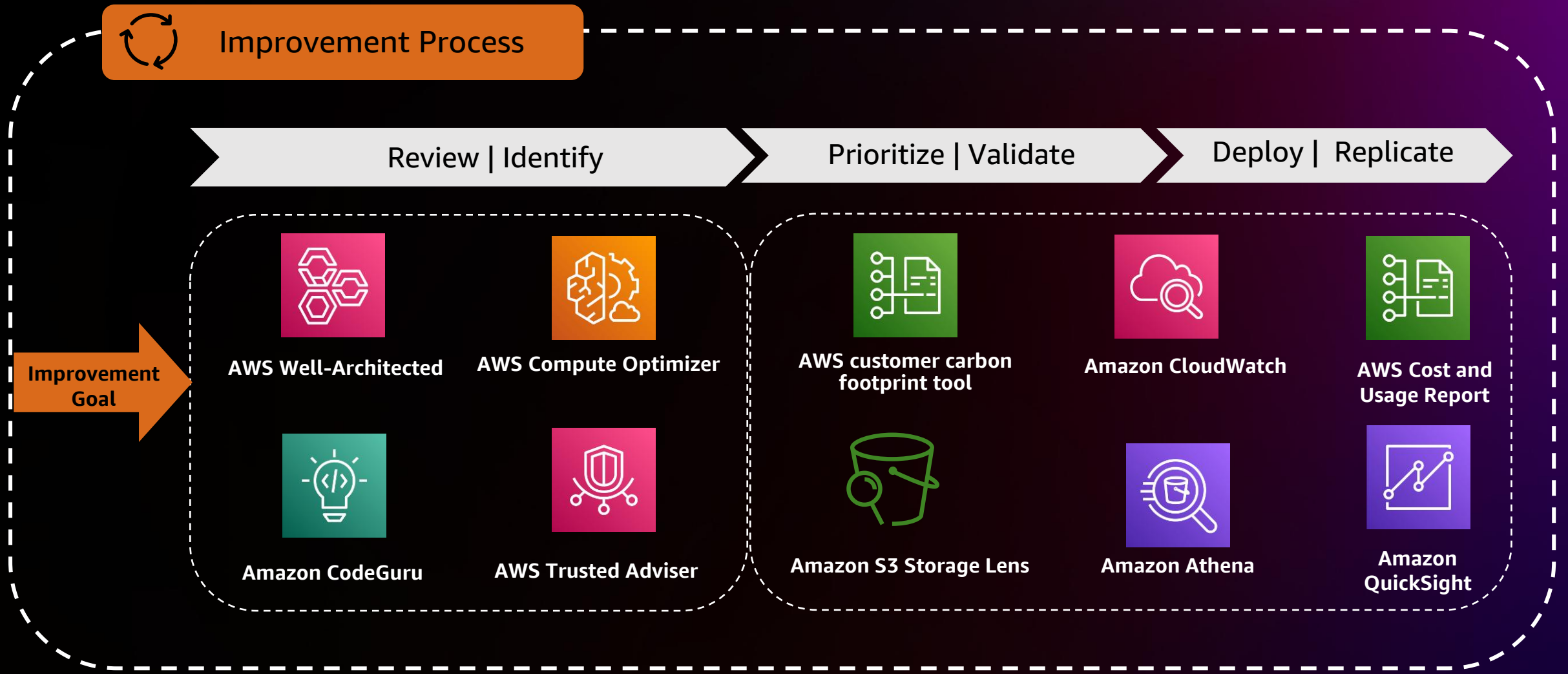
Average fare per taxi zone against Jun 2016 taxi ride data

Amazon S3

#	zone	average_total_amount	pickup_count
1	Bellerose	31.397662835249026	783
2	Manhattan Beach	31.180336700336692	594
3	South Beach/Dongan Hills	30.608468468468473	111
4	Starrett City	30.541820809248552	346
5	Coney Island	29.878358078602627	1145
6	Bath Beach	29.83827357237715	753
7	Marine Park/Mill Basin	29.64398137369033	859
8	Corona	29.528714632174566	9896
9	Gravesend	26.334874999999986	720
10	Bayside	26.142337662337653	693

aws ©

Improvement process – Reminder



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

Step 1
Start with sustainability goal

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

Improve **workload resource efficiency** by 50%

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

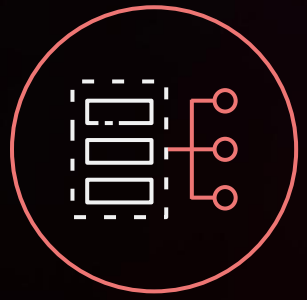
Step 2
Review and identify targets for improvement

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

Review best practices



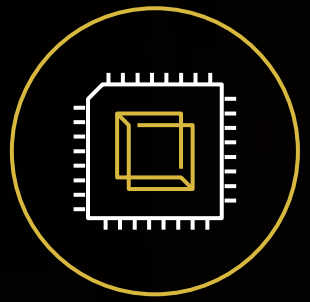
Region selection



User behavior



Software & architecture



Hardware



Data



Development & deployment process

aws

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

Identify targets for improvement



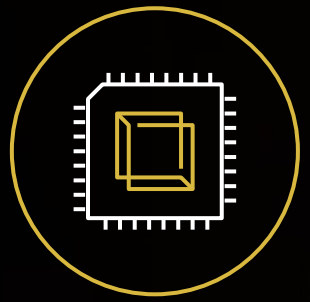
Region selection

Use Auto Scaling
to improve compute utilization

User behavior



Software & architecture



Hardware

Change JOIN mechanism
to reduce data shuffling

Change file format
to reduce scanned data

Data

Upgrade Spark version
to adopt efficient technology

Development & Deployment process

aws

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

Step 3
Prioritize targets for improvement

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

Prioritize targets for improvement

Improvement target	Potential	Effort	Risk
Use Auto Scaling	High	Medium	Medium
Upgrade Spark version	Medium	Low	Low
Change JOIN mechanism	Low	Medium	Medium
Change file format	Very High	High	Low

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Improvement targets after prioritization

1. Upgrade Spark version

2. Change file format

3. Use Auto Scaling

4. Change JOIN mechanism

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

Step 4
Validate specific improvements

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

$$KPI = \frac{\text{Proxy Metric}}{\text{Business Metric}}$$

Normalize sustainability KPIs to make them comparable

aws

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

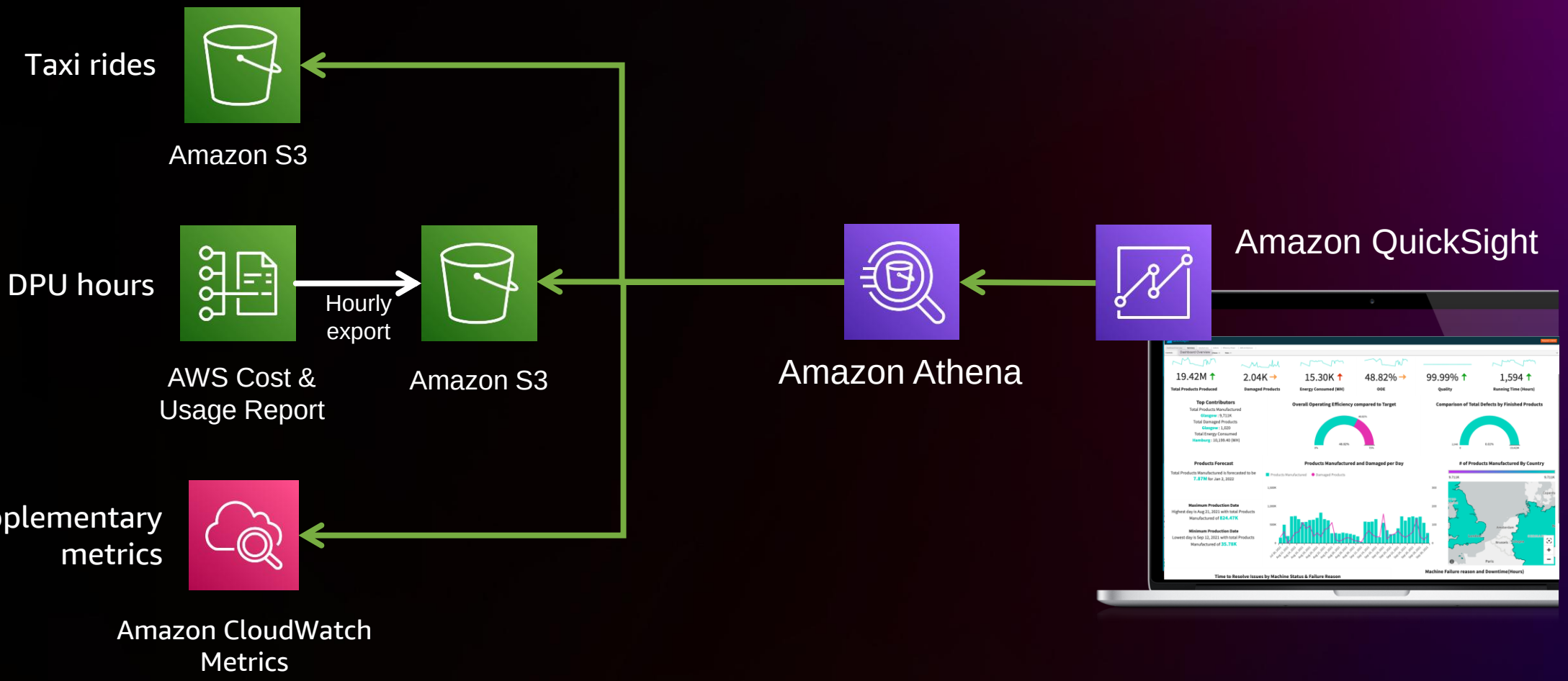
$$KPI = \frac{\text{DPU Hours}}{\text{Taxi Rides}}$$

Normalize sustainability KPIs to make them comparable

aws

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

Collect and visualize metrics for validation

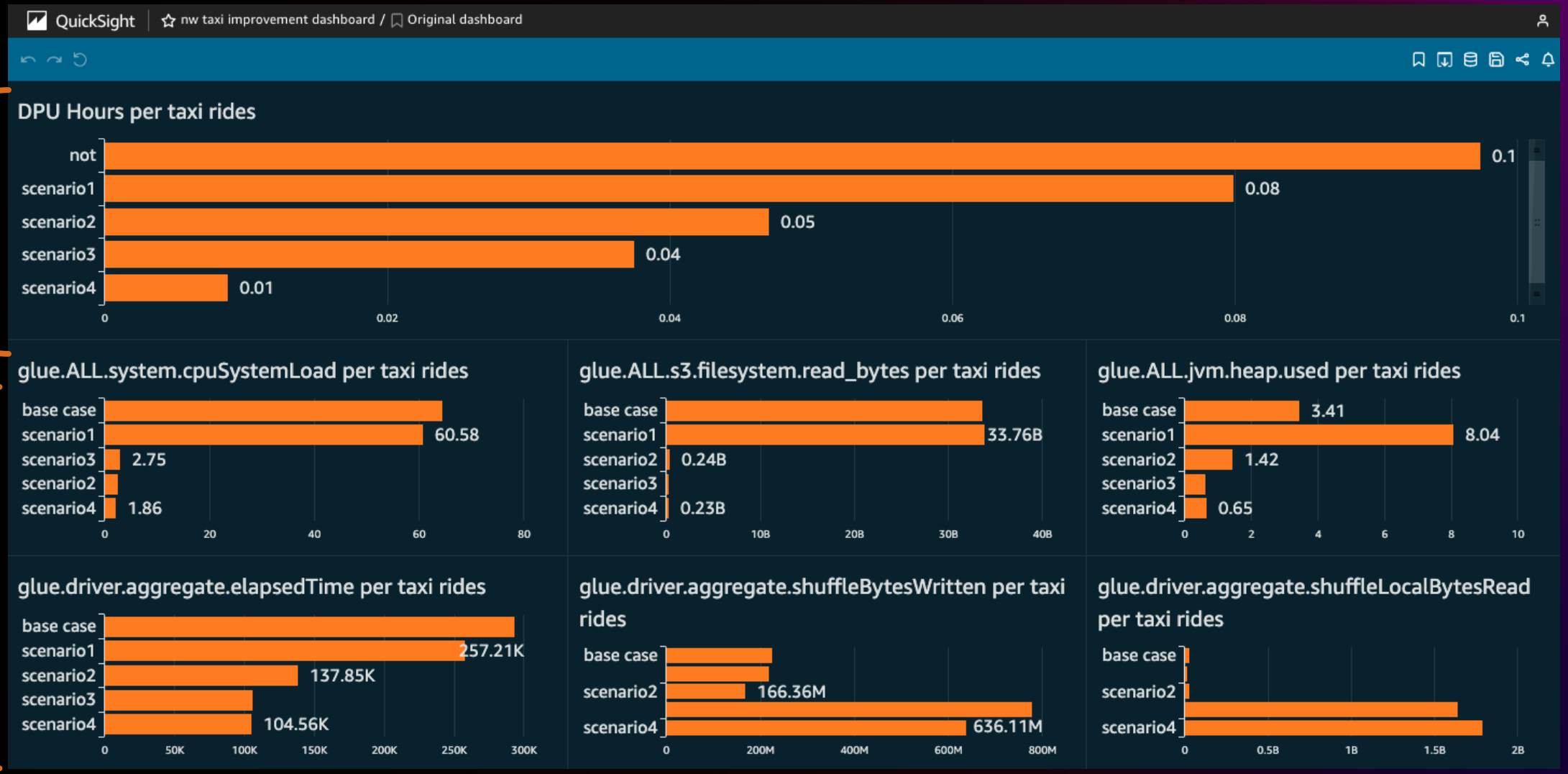


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

QuickSight dashboard for the demo workload

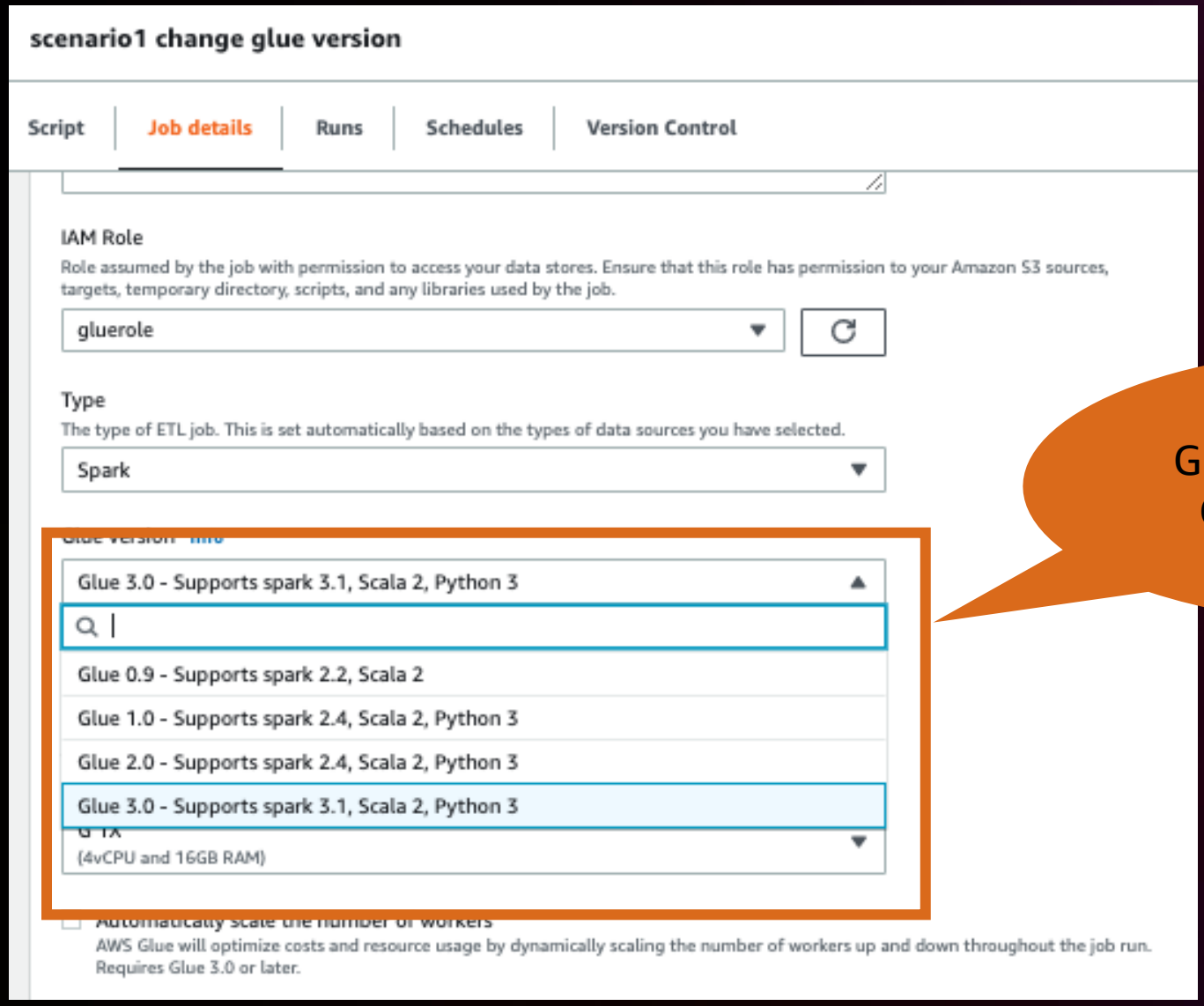
KPI

Metrics



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

Scenario 1: Upgrade Spark version



Glue 2.0 (Spark 2.4) to
Glue 3.0 (Spark 3.1)

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

Scenario 1: Upgrade Spark version

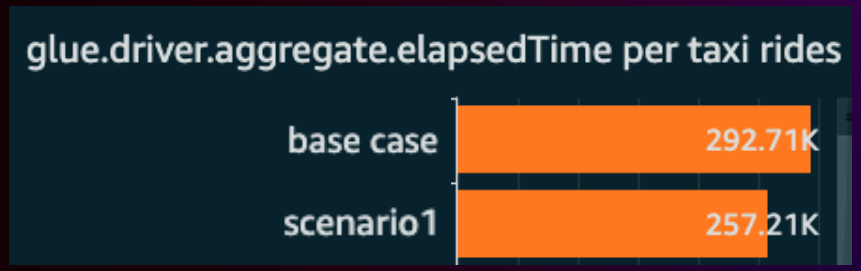
Outcome

- Sustainability KPI **dropped by 20%**



Why

- New feature in Apache Spark 3 optimize query plan dynamically



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

Scenario 2: Change file format

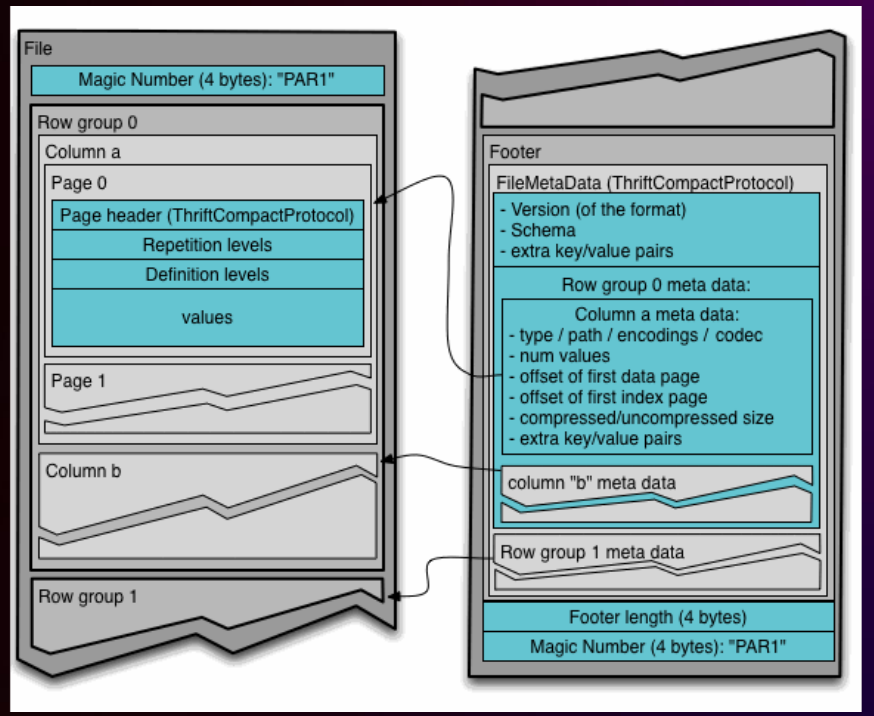
CSV

Text format without any schemas and metadata

```
VendorID,tpep_pickup_datetime,tpep_dropoff_datetime,passenger_count,trip_distance,pickup_longitude,pickup_la
titude,RatecodeID,store_and_fwd_flag,dropoff_longitude,dropoff_latitude,payment_type,fare_amount,extra,mta_t
ax,tip_amount,tolls_amount,improvement_surcharge,total_amount
2,2016-06-09 21:06:36,2016-06-09 21:13:08,2,.79,-73.983360290527344,40.760936737060547,1,N,-73.9774627685546
88,40.753978729248047,2,6,0.5,0.5,0,0,0.3,7.3
2,2016-06-09 21:06:36,2016-06-09 21:35:11,1,5.22,-73.981719970703125,40.736667633056641,1,N,-73.981636047363
281,40.670242309570313,1,22,0.5,0.5,4,0,0.3,27.3
2,2016-06-09 21:06:36,2016-06-09 21:13:10,1,1.26,-73.994316101074219,40.751071929931641,1,N,-74.004234313964
844,40.742168426513672,1,6.5,0.5,0.5,1.56,0,0.3,9.36
2,2016-06-09 21:06:36,2016-06-09 21:36:10,1,7.39,-73.98236083984375,40.773891448974609,1,N,-73.9294662475585
94,40.851539611816406,1,26,0.5,0.5,1,0,0.3,28.3
2,2016-06-09 21:06:36,2016-06-09 21:23:23,1,3.10,-73.987106323242187,40.733173370361328,1,N,-73.985908508300
781,40.766445159912109,1,13.5,0.5,0.5,2.96,0,0.3,17.76
2,2016-06-09 21:06:36,2016-06-09 21:19:21,1,2.17,-73.995201110839844,40.739490509033203,1,N,-73.993202209472
656,40.762641906738281,1,10.5,0.5,0.5,2.36,0,0.3,14.16
2,2016-06-09 21:06:36,2016-06-09 21:30:13,5,6.02,-73.980476379394531,40.741680145263672,1,N,-73.902961730957
031,40.748741149902344,2,21.5,0.5,0.5,0,0,0.3,22.8
1,2016-06-09 21:06:37,2016-06-09 21:16:47,1,1.40,-73.985702514648437,40.746742248535156,1,N,-73.982505798339
844,40.762771606445313,1,8.5,0.5,0.5,1.95,0,0.3,11.75
1,2016-06-09 21:06:37,2016-06-09 21:15:44,1,1.20,-73.984275817871094,40.775009155273438,1,N,-73.968498229980
469,40.766429901123047,2,8,0.5,0.5,0,0,0.3,9.3
1,2016-06-09 21:06:37,2016-06-09 21:23:57,2,1.90,-73.977920532226563,40.750091552734375,1,N,-73.987236022949
219,40.749359130859375,1,12,0.5,0.5,3.33,0,0.3,16.63
1,2016-06-09 21:06:37,2016-06-09 21:10:51,1,.70,-73.991203308105469,40.739662170410156,1,N,-73.9829406738281
25,40.745071411132812,1,5,0.5,0.5,1.55,0,0.3,7.85
2,2016-06-09 21:06:37,2016-06-09 21:27:12,5,3.12,-73.967498779296875,40.763149261474609,1,N,-73.992439270019
531,40.724838256835938,1,15,0.5,0.5,3.26,0,0.3,19.56
2,2016-06-09 21:06:37,2016-06-09 21:08:27,2,.44,-73.976333618164062,40.785850524902344,1,N,-73.9717330932617
19,40.791969299316406,2,3.5,0.5,0.5,0,0,0.3,4.8
2,2016-06-09 21:06:37,2016-06-09 21:17:33,1,2.09,-73.993583679199219,40.724590301513672,1,N,-74.008049011230
469,40.740341186523438,1,10,0.5,0.5,2.26,0,0.3,13.56
2,2016-06-09 21:06:37,2016-06-09 21:17:35,1,1.86,-73.977958679199219,40.748809814453125,1,N,-74.000762939453
125,40.738880157470703,1,9,0.5,0.5,2,0,0.3,12.3
1.2016-06-09 21:06:38.2016-06-09 21:27:14.1.3.70.-73.994224548339844.40.751262664794922.1.N.-73.952278137207
```

Parquet

Compressed with schema and various metadata for performance efficiency

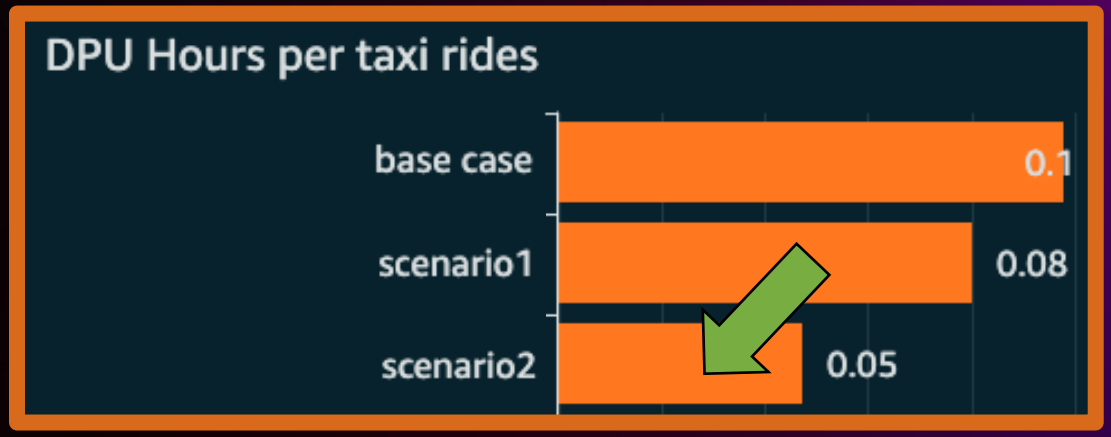


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Scenario 2: Change file format

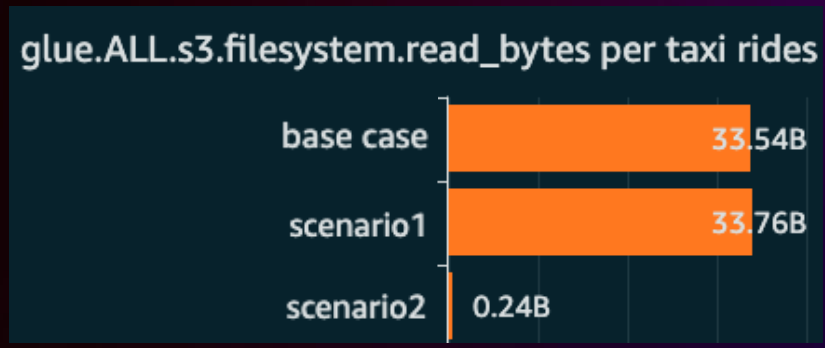
Outcome

- Sustainability KPI **dropped by another 30%** compared to base case



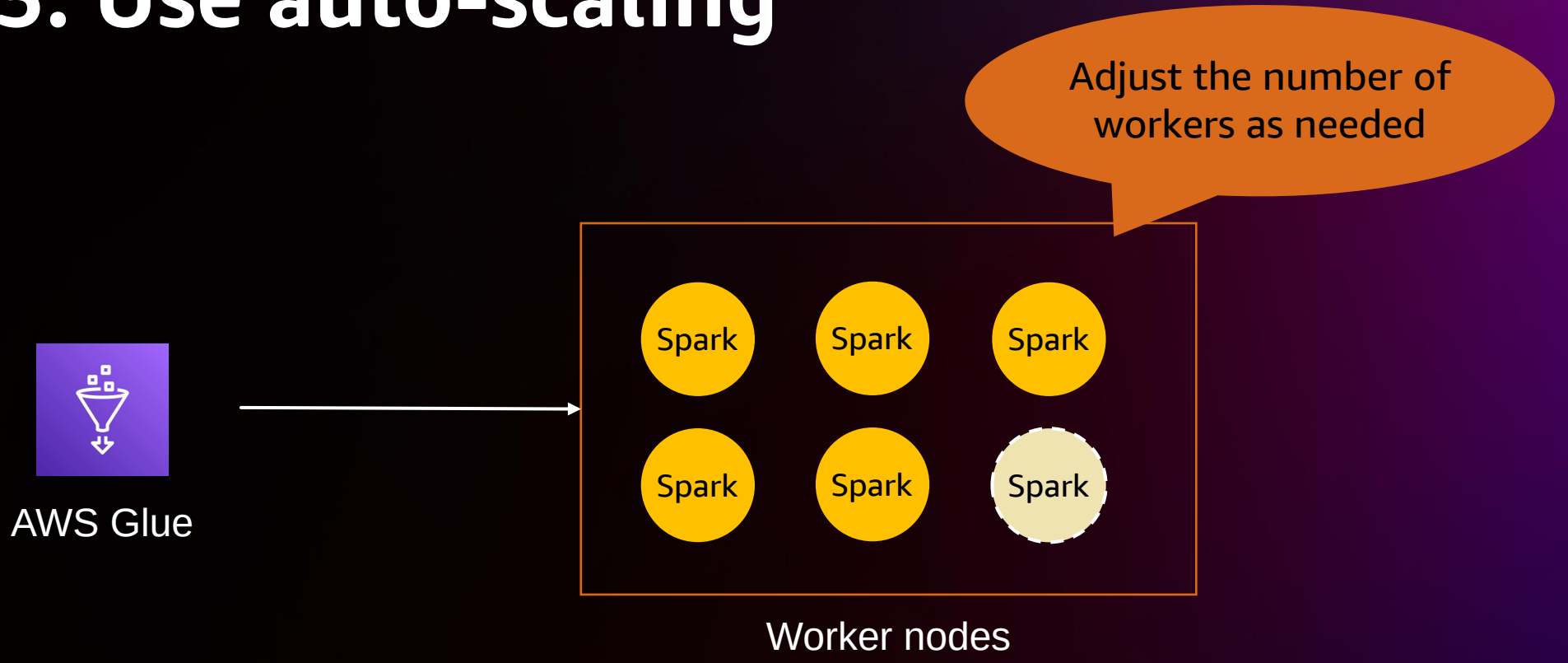
Why

- Only the necessary data was scanned using metadata with proper compression



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Scenario 3: Use auto-scaling



Automatically scale the number of workers

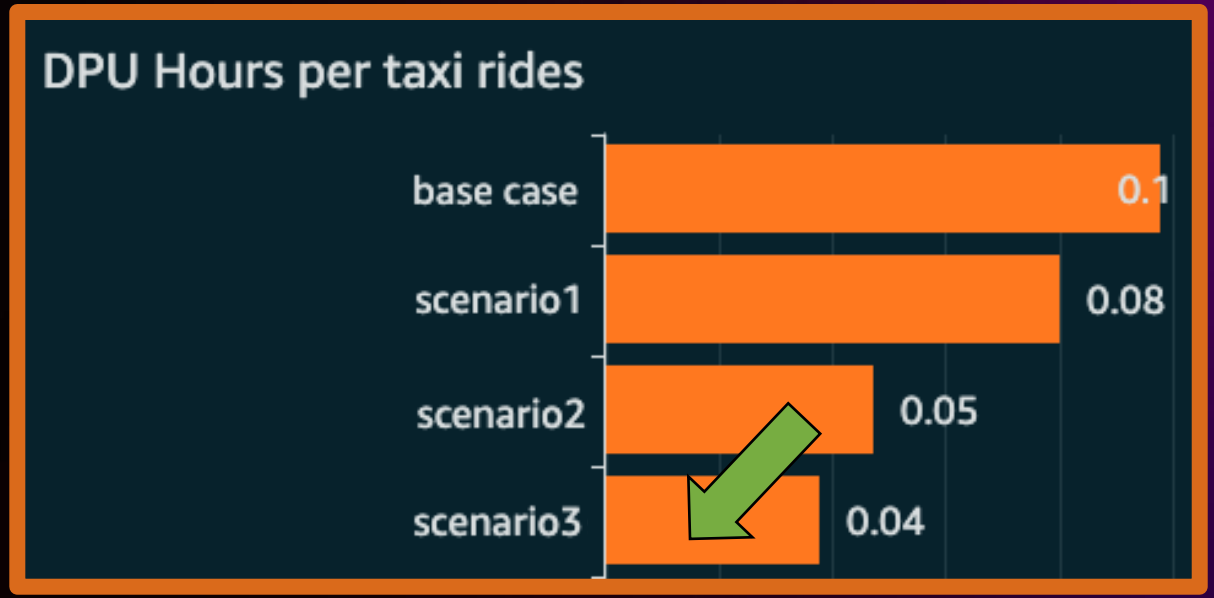
AWS Glue will optimize costs and resource usage by dynamically scaling the number of workers up and down throughout the job run. Requires Glue 3.0 or later.

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

Scenario 3: Use auto-scaling

Outcome

- Sustainability KPI **dropped by another 10%** compared to base case



Why

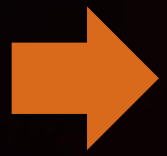
- AWS Glue auto-scaling adjusts capacity in real-time which reduced idle resources

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

Scenario 4: Change JOIN mechanism

Sort merge join

Efficient if both datasets are large



Broadcast join

Efficient if one of the two datasets is small

```
trip_with_area = trip_df \
    .join(broadcast(taxi_zones), trip_df.PULocationID == taxi_zones.LocationID)
```

aws

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

Scenario 4: Change JOIN mechanism

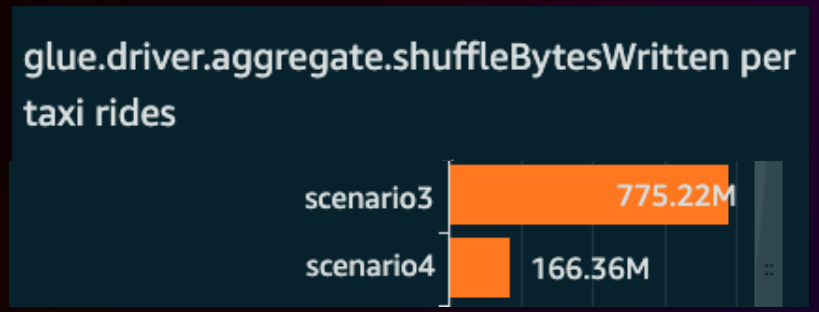
Outcome

- Sustainability KPI **dropped by another 30%** compared to base case

Why

- Changes to the join mechanism have minimized network transfers between workers





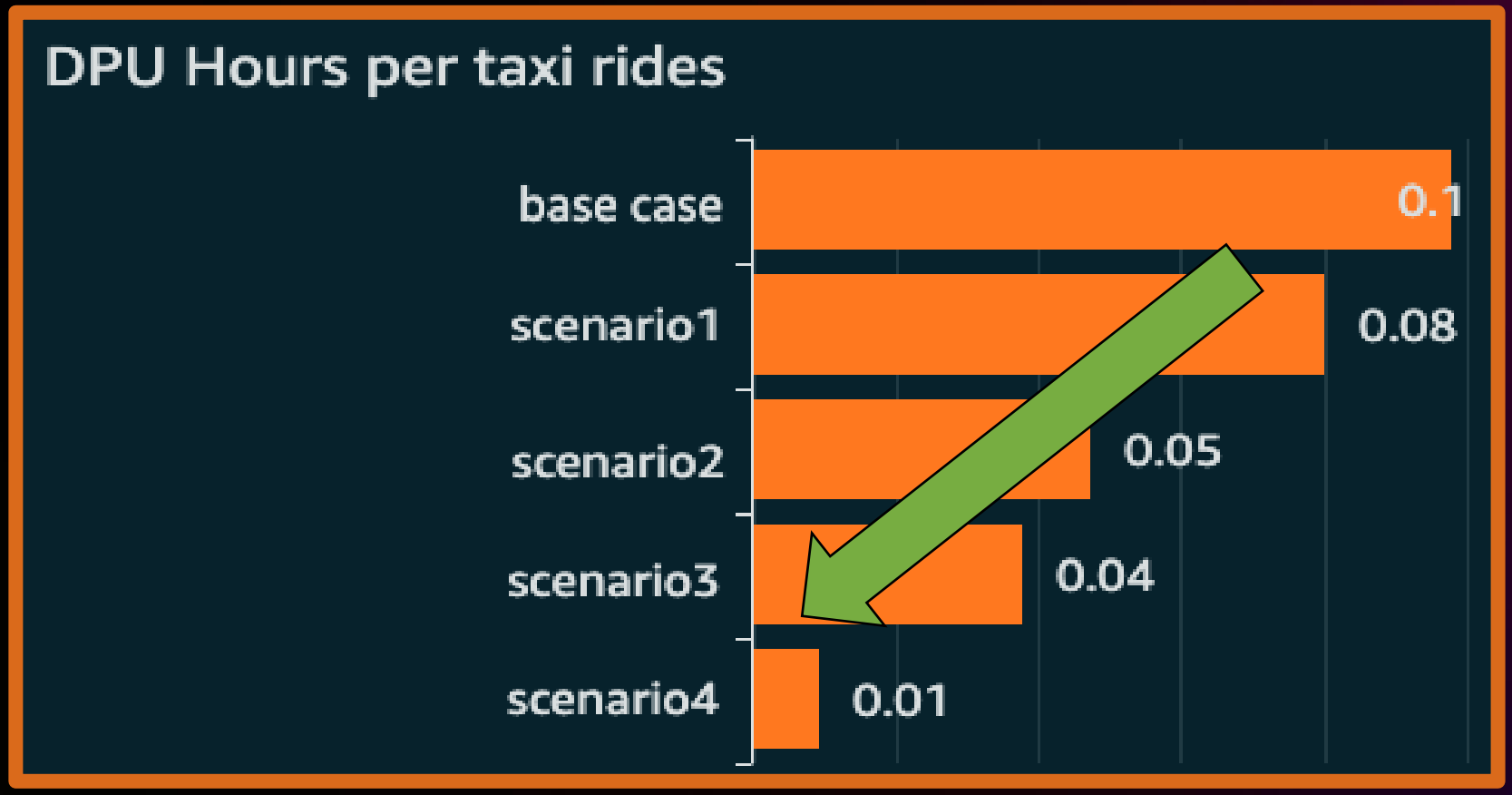
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Step 5
Measure results and replicate success

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

Overall improvement

- 90% reduction in Sustainability KPI



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

Key takeaways

1. How AWS can help you in your Sustainability journey
2. AWS Frameworks and Tooling to deliver sustainable high-performing architectures
3. Improvement process to deliver sustainable high-performing architectures

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

Sustainability breakout sessions

SUS204: Sustainability in AWS Global Infrastructure

SUS205: Architecting sustainably and reducing your AWS carbon footprint

SUS206: Sustainability and AWS Silicon

SUS208: Utilizing sustainability data at scale

SUS210: Modeling climate change impacts and risks at scale

SUS211: Water stewardship and renewable energy at scale

SUS212: Accelerating decarbonization & sustainability transformation

SUS301: Sustainable machine learning for protecting natural resources

SUS303: Delivering sustainable, high-performing architectures

SUS312: How innovators are driving more sustainable manufacturing

STP213: Scaling global carbon footprint management

aws

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

Sustainability workshops and chalk talks

SUS201: Detecting deforestation with geospatial images and Amazon SageMaker

SUS202: Using Amazon SageMaker, AutoGluon, and ASDI to predict air quality

SUS203: Carbon emissions reporting using modern data architecture on AWS

SUS207: Patterns for obtaining insights from sustainability data

SUS302: Optimizing architectures for sustainability

aws

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

Thank you!

Dr. Sam Mokhtari
ghm@amazon.com

Thomas Coombs
tccoomb@amazon.ch

Chie Hayashida
haychie@amazon.co.jp



Please complete the session survey in the **mobile app**

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