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

BOA302-R

The serverless future of cloud data analytics

Wojciech Gawronski (he/him)
CEE Senior Developer Advocate
AWS

Lukasz Panusz (he/him)
CEE Senior Solutions Architect
AWS

Scope of work

Quick workshop introduction

Serverlesspresso business case

Coding time

Workshop summary

General housekeeping

Workshop will take **2 hours**

- Workshop **cannot** be completed with a tablet
- Have **Chrome or Firefox** web browsers (not compatible with Safari)
- Workshop is 100% browser-based; do NOT use your local IDE or terminal

We have an **AWS account** for you

Questions? We are here to help!

The AWS services you will use today

Developer tools

AWS Cloud9

Workshop focus

Amazon MSK Serverless

Amazon EMR Serverless

Amazon Redshift Serverless

Environment setup

AWS Glue

Amazon S3

Amazon CloudWatch

Amazon DynamoDB

serverlesspresso

Serverlesspresso: Coffee business at scale

aws


serverlesspresso
COFFEE BAR
SERVERLESSLAND.COM

AWS re:Invent 2021

1,920 orders
71 drinks per hour

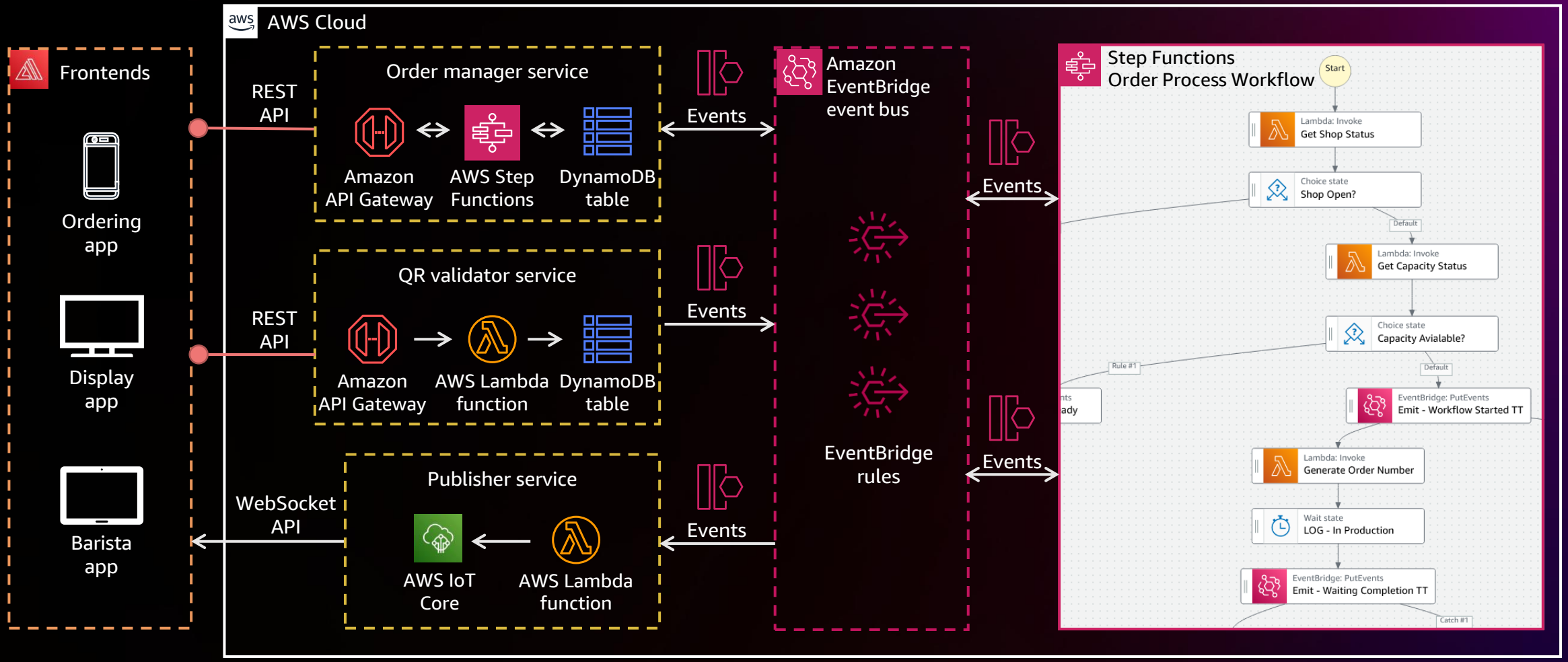


LEAP Saudi Arabia

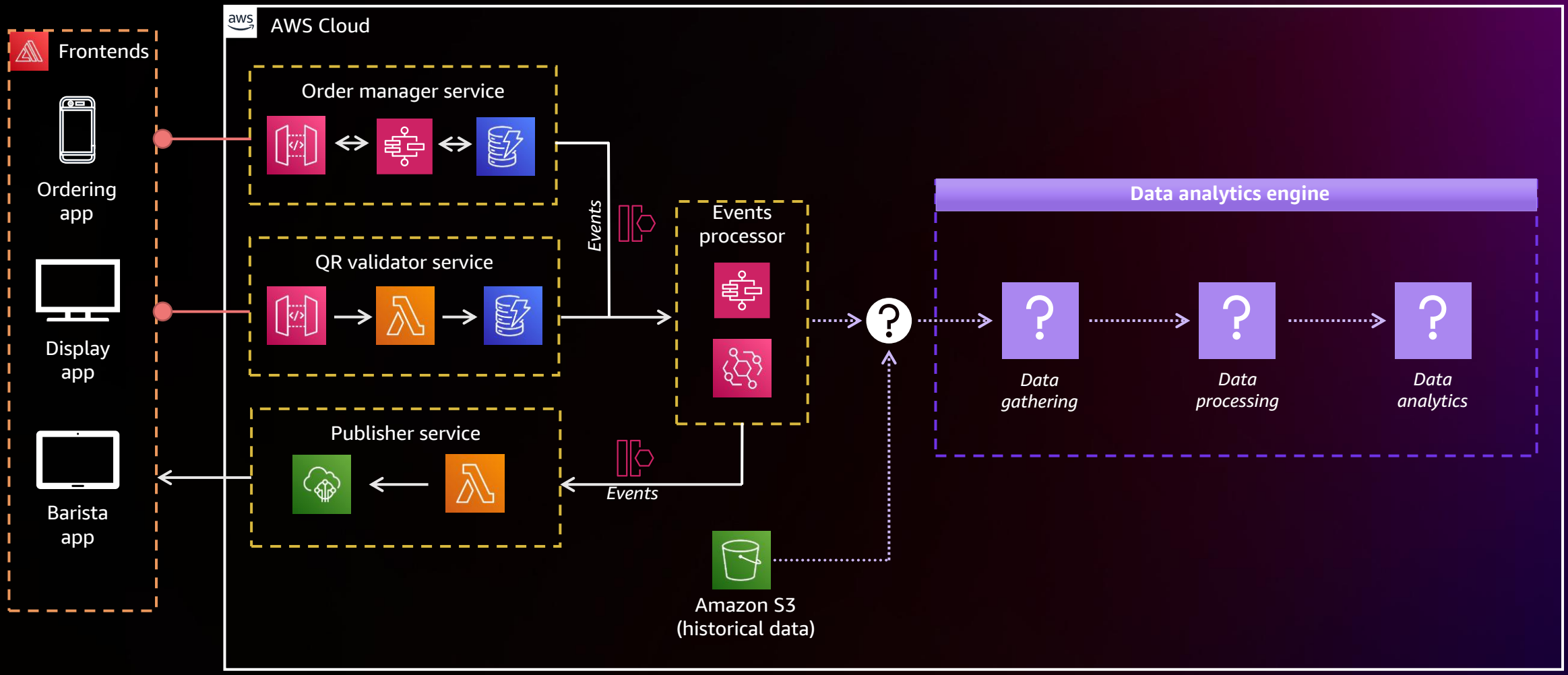
1,310 orders
63 drinks per hour

Informed decisions: Power of data analytics

Existing Serverlesspresso system



Serverlesspresso data analytics engine



What do we want to build?

Setup

Login to prepared AWS account

Module 1: Serverless data acquisition

Build a connection to data sources to capture new orders and save them to data lake

Module 2: Serverless data transformation

Create scripts to trigger data transformation and enrichment

Module 3: Serverless data analytics

Configure data warehouse to use newly enriched data and start analysis of the orders using queries

What do we want to build?

Setup

Login to prepared AWS account

Module 1: Serverless data acquisition

Build a connection to data sources to capture new orders and save them to data lake

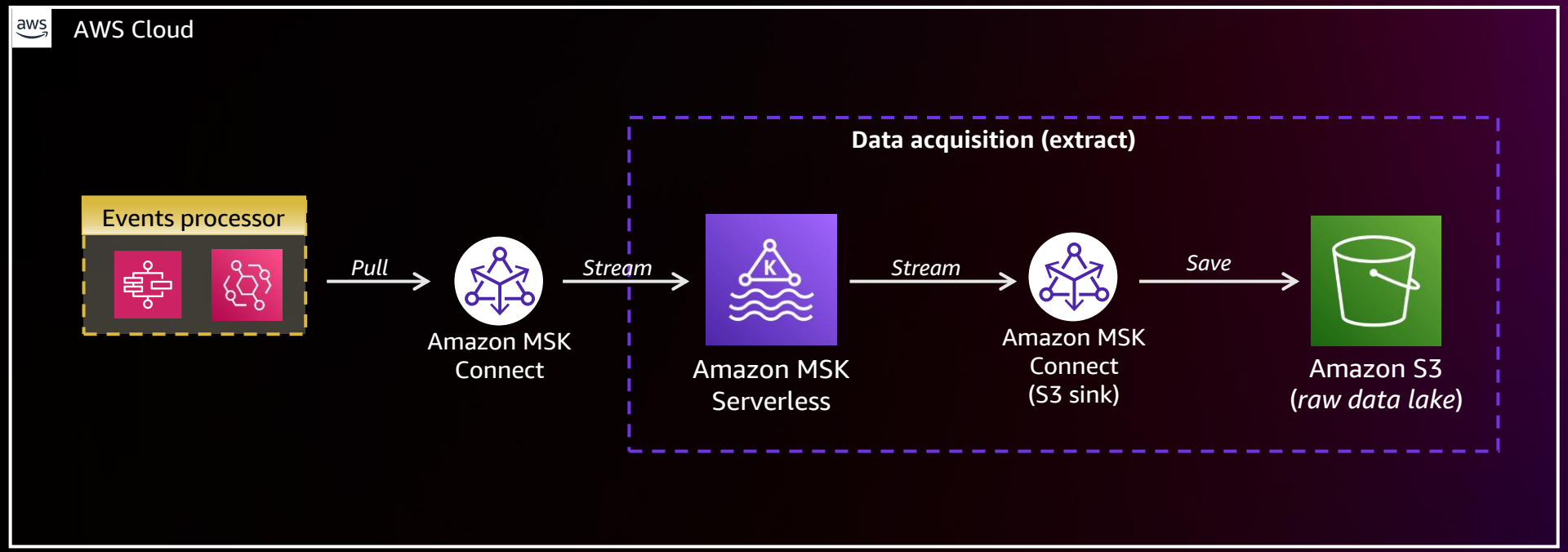
Module 2: Serverless data transformation

Create scripts to trigger data transformation and enrichment

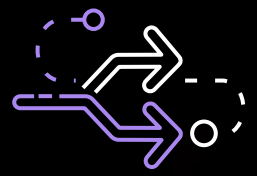
Module 3: Serverless data analytics

Configure data warehouse to use newly enriched data and start analysis of the orders using queries

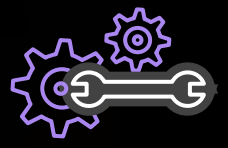
Data acquisition – high-level view



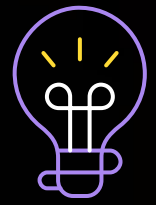
Short introduction to Amazon MSK Serverless



Allows developers to **focus more on application development** and less on infrastructure management



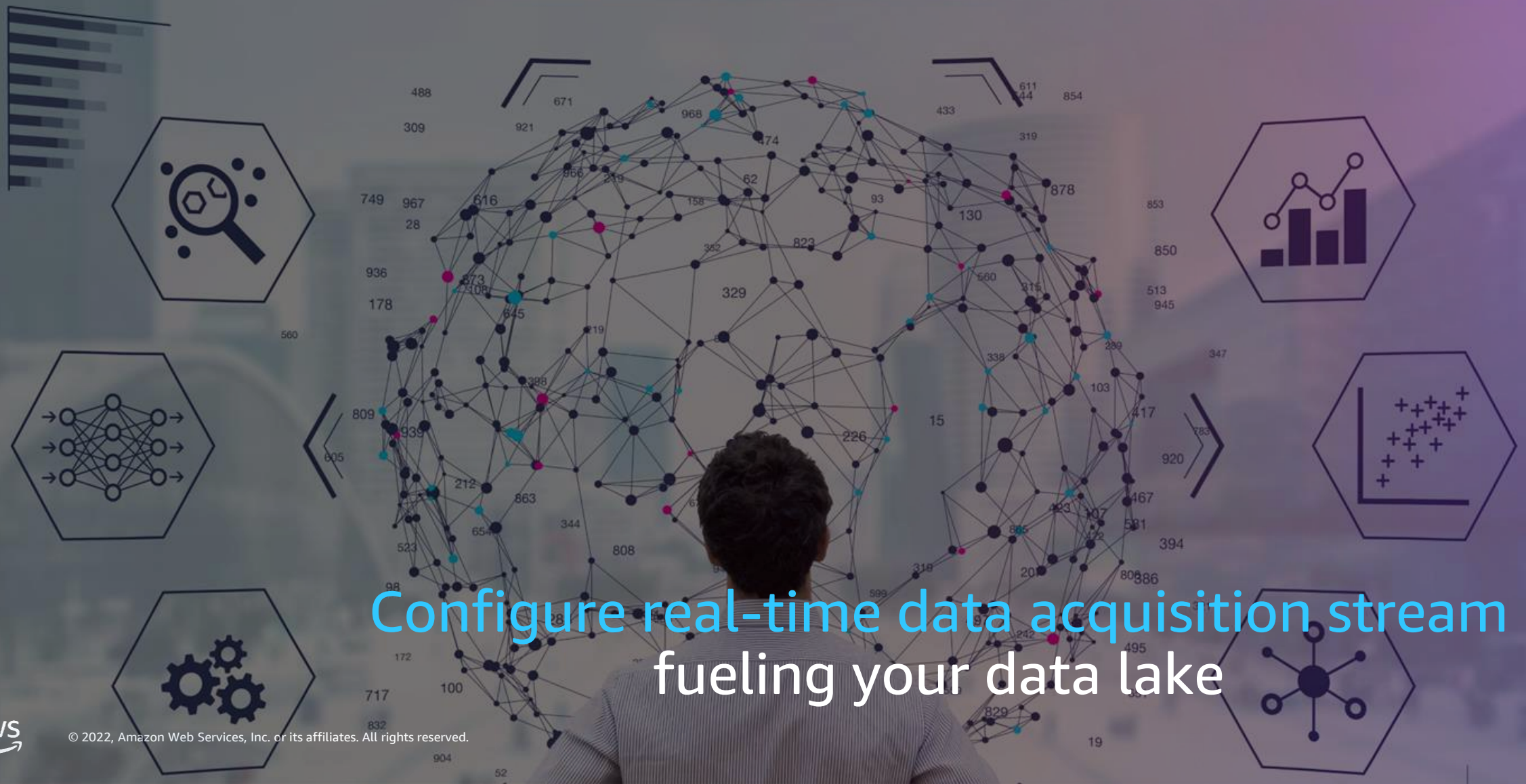
Eliminates rightsizing, scaling, partition management, and instantly scales I/O without worrying about scaling capacity or partitions reassigning



Makes Apache Kafka more **secure, highly available, and accessible** to your organization

Pay for **the data volume you stream and retain** with throughput-based pricing

Acquiring data from each coffee order



Configure real-time data acquisition stream
fueling your data lake

What do we want to build?

Setup

Login to prepared AWS account

Module 1: Serverless data acquisition

Build a connection to data sources to capture new orders and save them to data lake

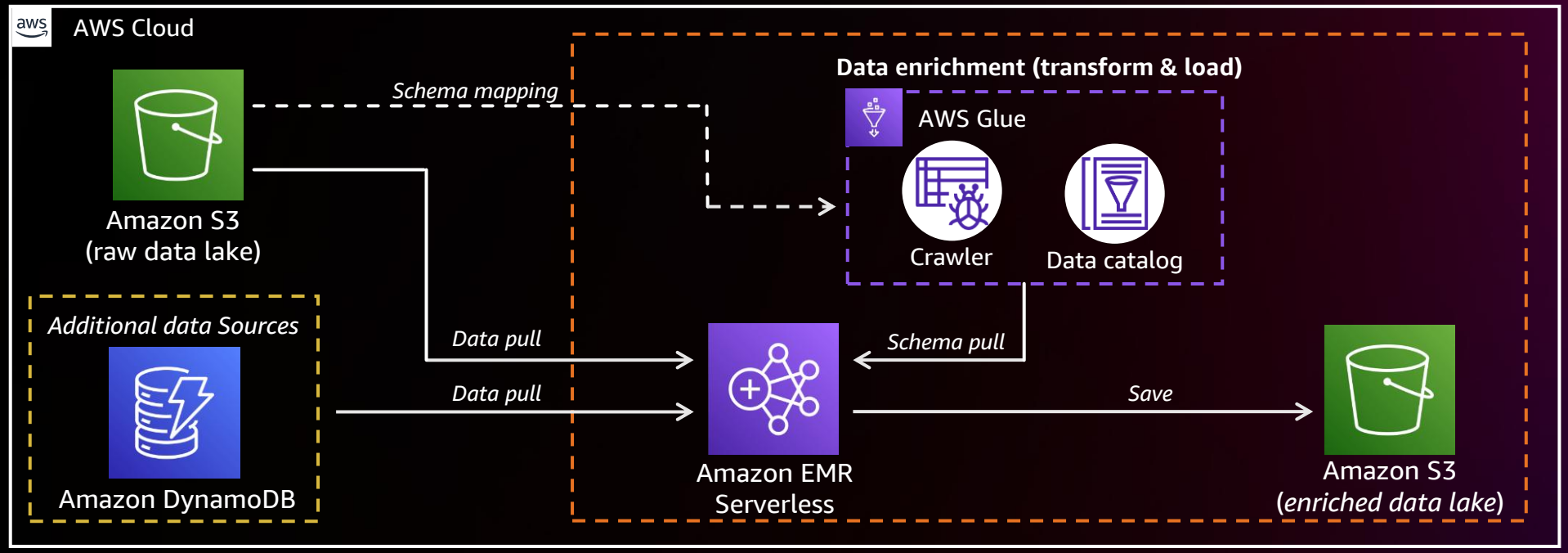
Module 2: Serverless data transformation

Create scripts to trigger data transformation and enrichment

Module 3: Serverless data analytics

Configure data warehouse to use newly enriched data and start analysis of the orders using queries

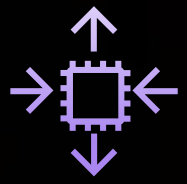
Data processing – high-level view



Short introduction to Amazon EMR Serverless

No servers to manage. EMR Serverless provisions, configures, and dynamically scales the compute and memory resources needed at each stage of data processing

Eliminates rightsizing and scaling. Performance-optimized runtime is fully compatible with, and faster than, standard open source



Use EMR Studio with notebooks and familiar open source tools to easily develop, visualize, and debug applications

Pay only per-second billing for worker execution time, aggregated resources allocation (vCPU, memory, and storage) and dimensions

Consolidating data for analysis at scale

Transform and **enrich** the orders data
and save it in the data lake

What do we want to build?

Setup

Login to prepared AWS account

Module 1: Serverless data acquisition

Build a connection to data sources to capture new orders and save them to data lake

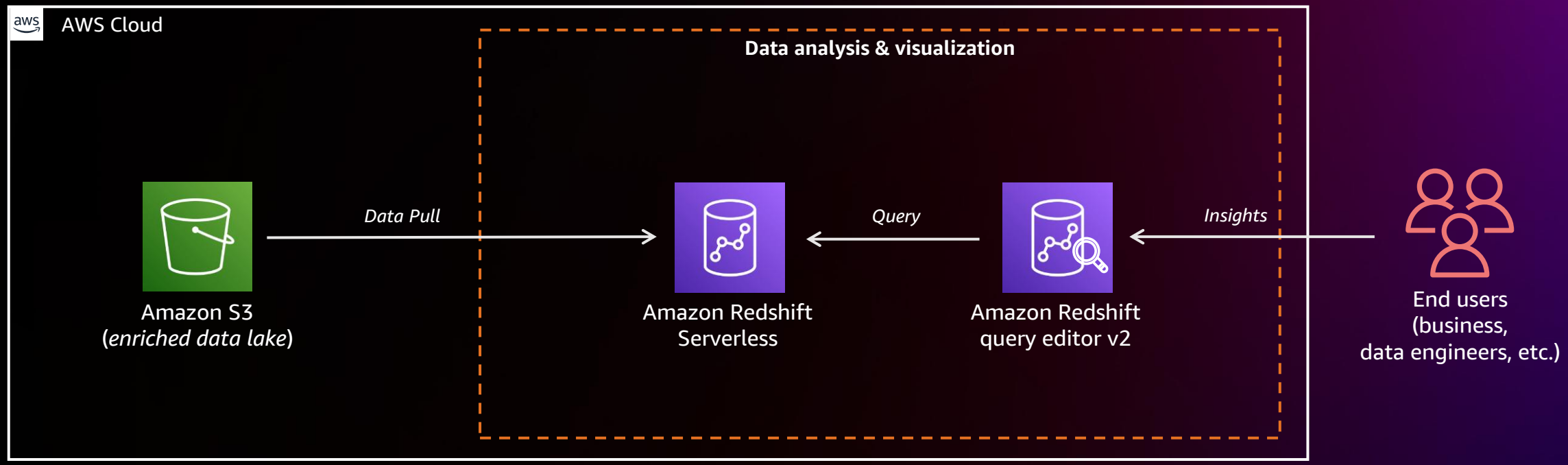
Module 2: Serverless data transformation

Create scripts to trigger data transformation and enrichment

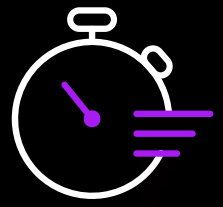
Module 3: Serverless data analytics

Configure data warehouse to use newly enriched data and start analysis of the orders using queries

Data analytics – high-level view



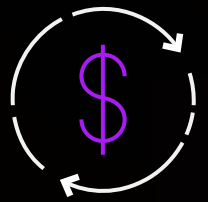
Short introduction to Amazon Redshift Serverless



Focus on getting to insights by getting started quickly and running real-time or predictive analytics on all your data. Don’t worry about managing data warehouse infrastructure

Automatically scales data warehouse capacity up or down to deliver consistently fast performance for even the most demanding and unpredictable workloads

Powerful analytics with a simple-to-use experience, without any impact on existing applications or your usage of advanced capabilities like machine learning



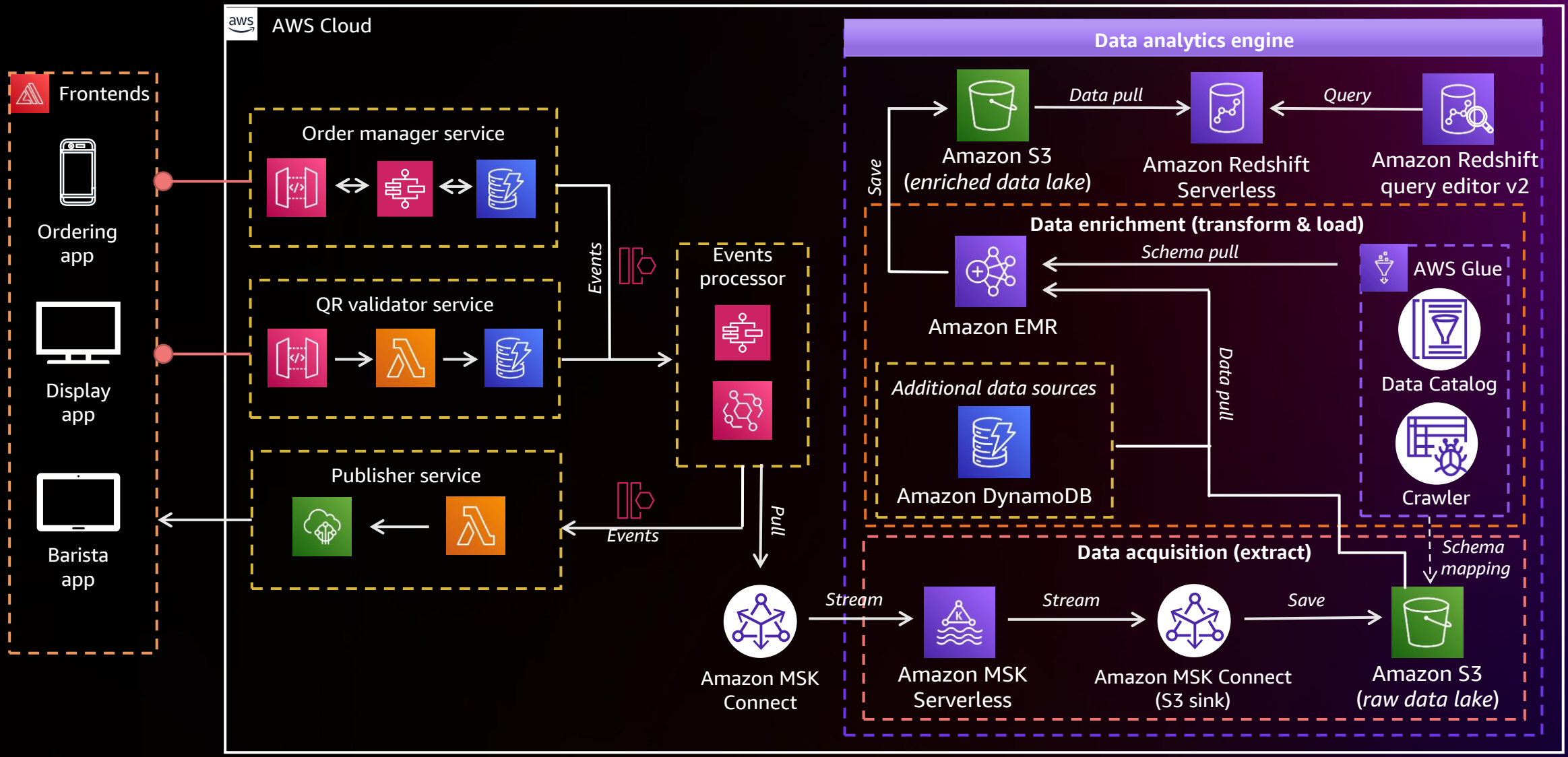
Pay only for what you use – on a per-second basis. When the data warehouse is idle, you pay nothing. Ability to set spend limit

Business insights from data

Query the data and get the insights from the business
to make informed decisions and earn loyal customers

Workshop summary

High-level architecture of complete solution



Learn in-demand AWS Cloud skills

AWS Skill Builder

Access **500+ free** digital courses and learning plans

Explore resources with a variety of skill levels and **16+** languages to meet your learning needs

Deepen your skills with digital learning on demand

Train now

AWS Certifications

Earn an industry-recognized credential

Receive Foundational, Associate, Professional, and Specialty certifications

Join the **AWS Certified community** and get exclusive benefits

Access **new** exam guides

Thank you!

Wojciech Gawronski

wojgaw@amazon.com

/in/afronski

@afronski

Lukasz Panusz

pnszlp@amazon.com

/in/lukasz-panusz

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