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

API305

Integrating Amazon MWAA across containers and serverless

Uma Ramadoss (she/her)

Solutions Architect, Specialist
AWS

Parnab Basak (he/him)

Senior Solutions Architect
AWS

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

Agenda

Introduction to Amazon MWAA, including Airflow

Effective usage of Airflow workers

Offload compute by externalizing tasks

Choices to offload compute – containers and Lambda

Demo

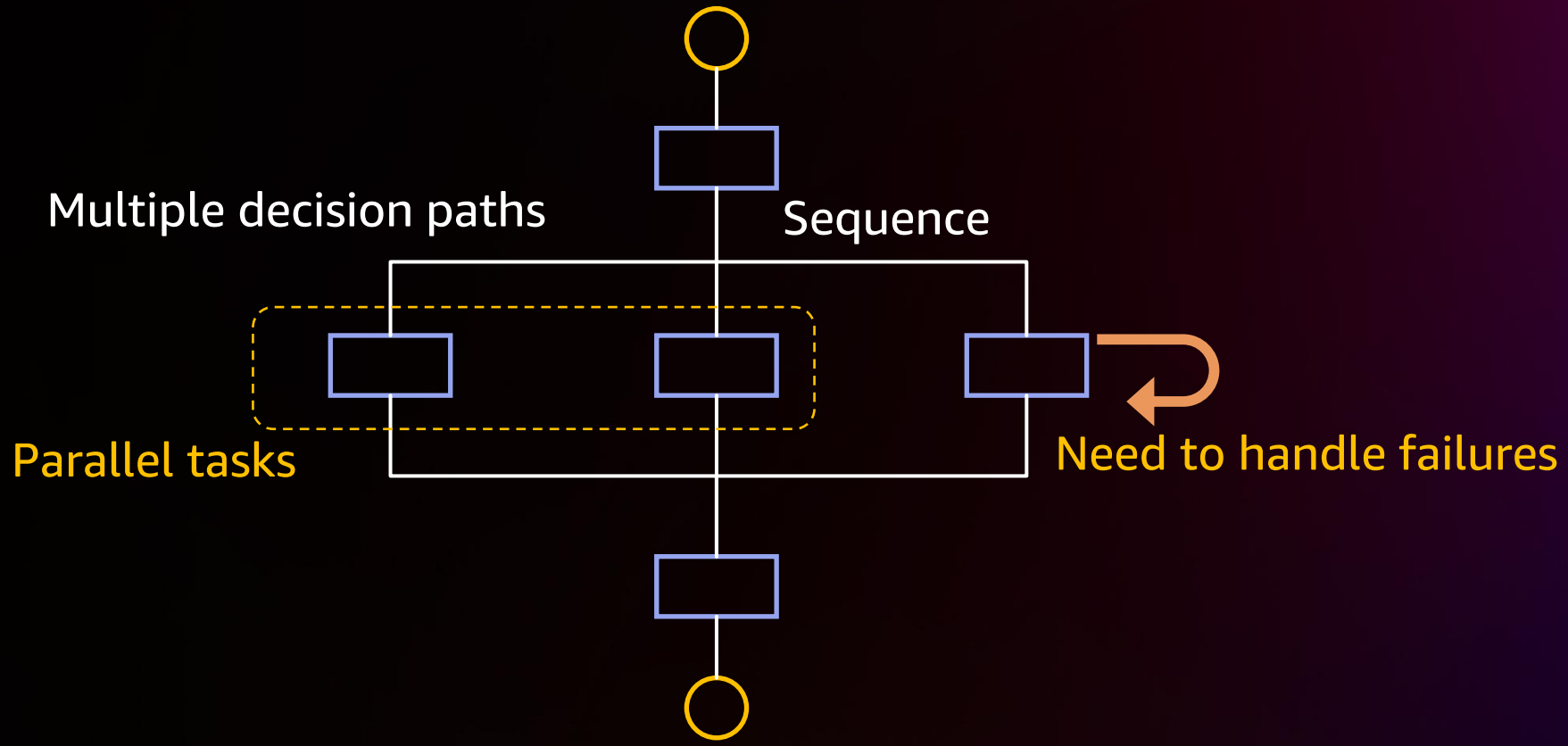
Getting started

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

Data orchestration

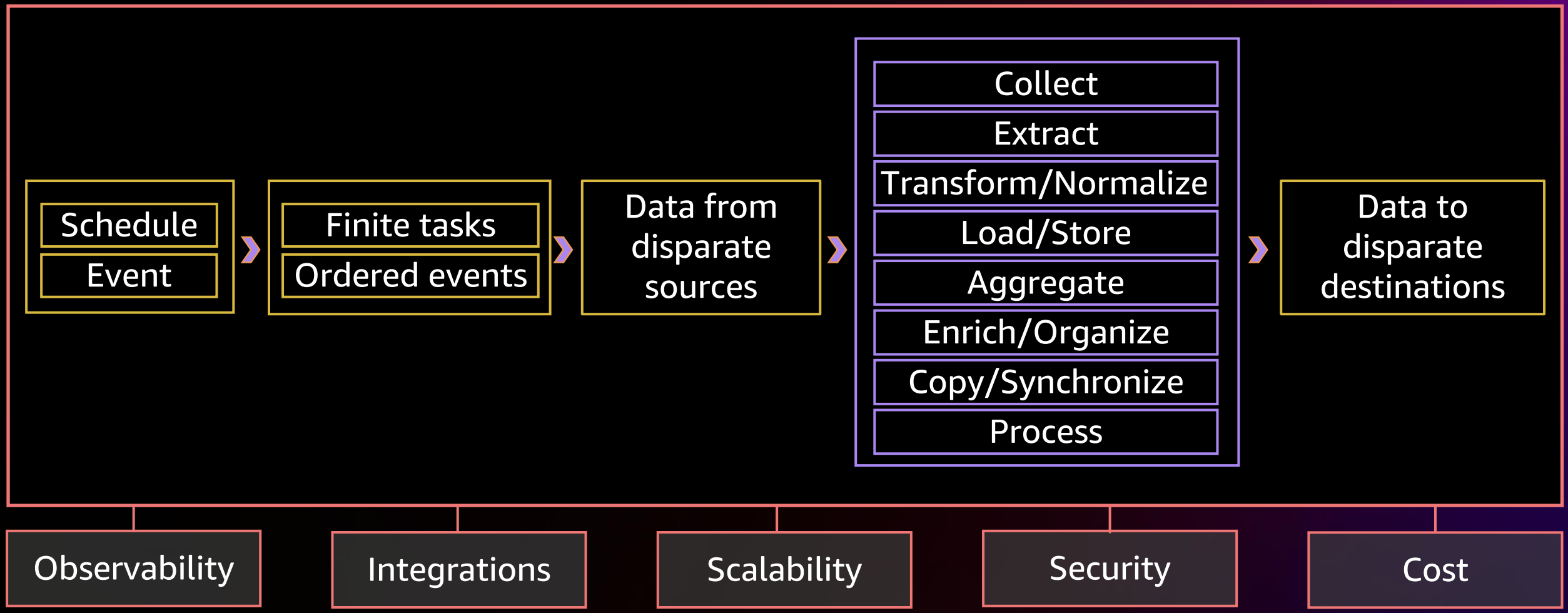
Data **orchestration** is the process of taking siloed data from multiple data storage locations, combining and organizing it, and making it available to serve business needs

Break down
Data silos



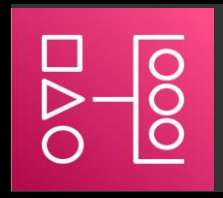
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data orchestration



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

Today's focus



Amazon Managed Workflows for Apache Airflow (Amazon MWAA)

What

- Fully managed
- Open-source Airflow minus operational responsibilities
- Large open-source community
- Data orchestration

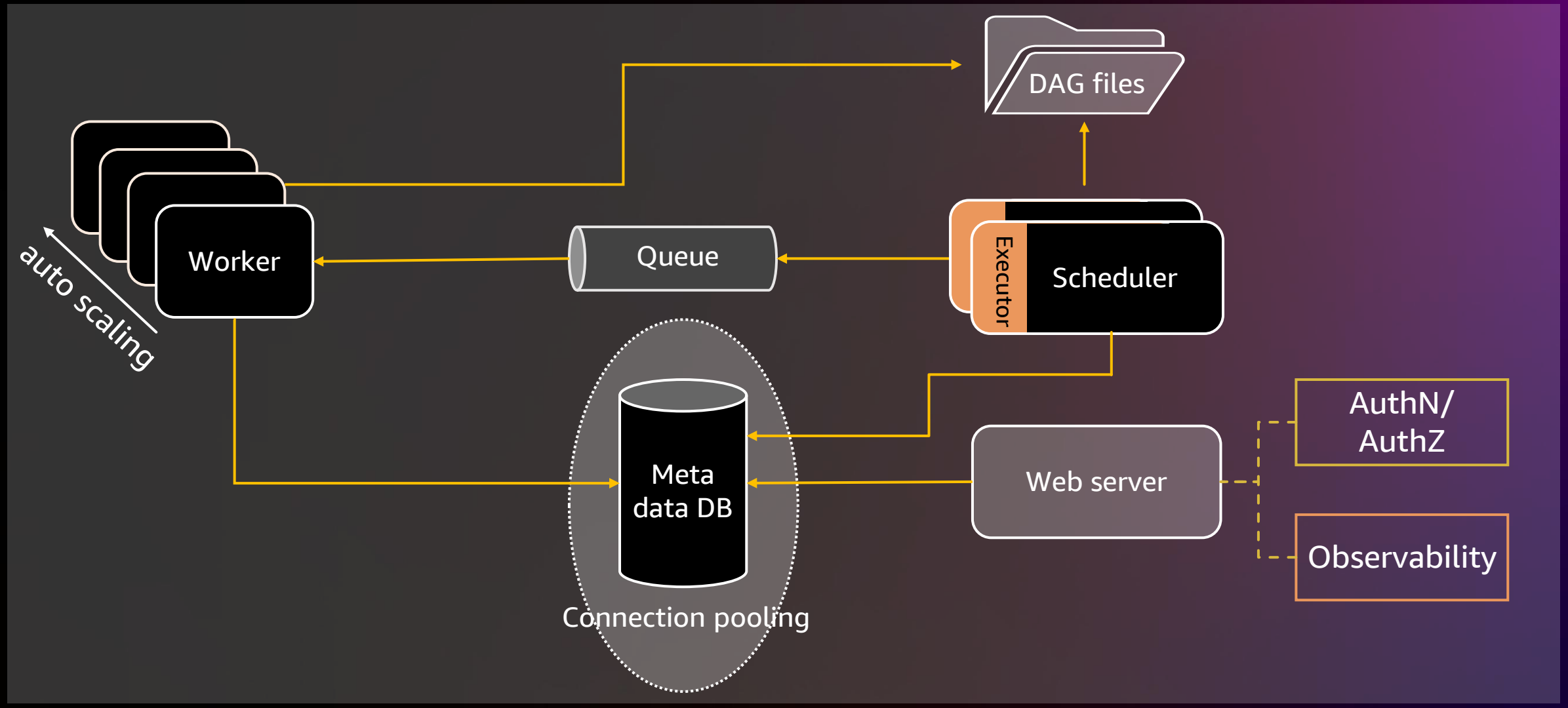
Why

- Deploy Airflow rapidly at scale
- Reduce operational costs
- Run Airflow with built-in security
- Use a preexisting plugin or use your own

aws

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

Amazon MWAA deployment architecture of Apache Airflow



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

To learn more about schedulers

API304 – Amazon MWAA at scale
Wednesday (November 30) at 2:30 PM – Caesars Forum 104

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

MWAA worker capacity

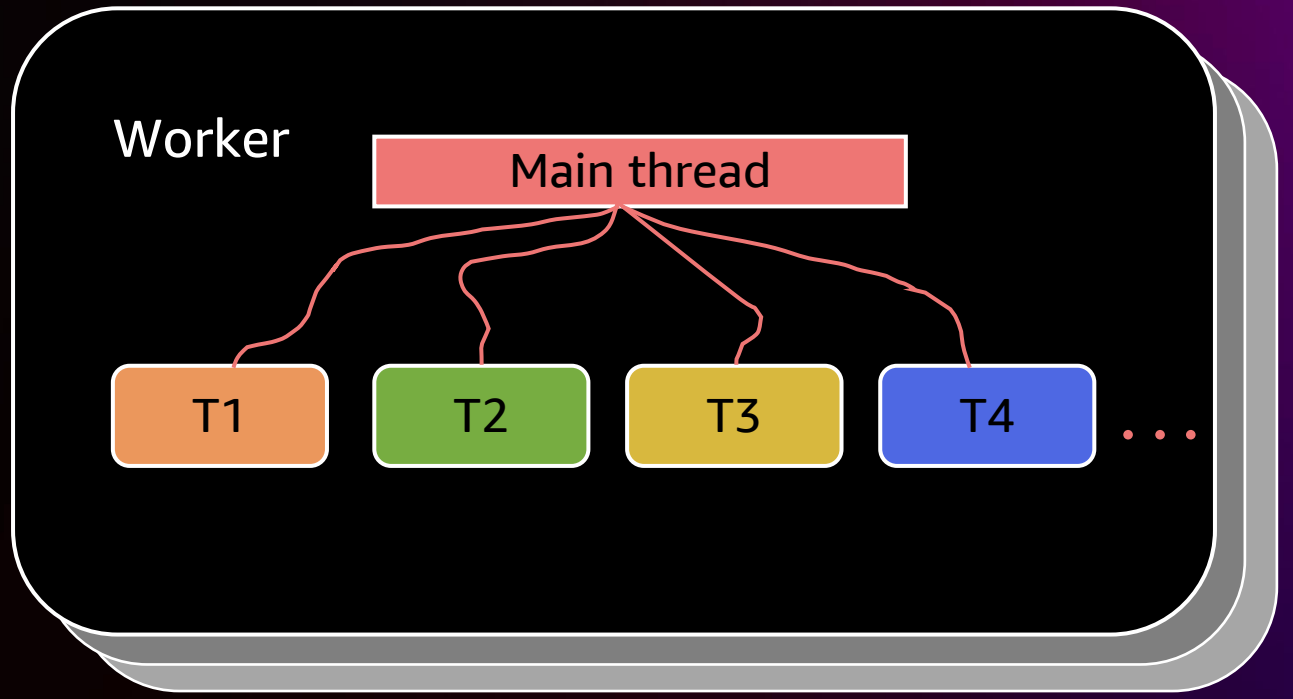
Class	Worker vCPU	Worker RAM (GB)	Task concurrency Number
mw1.small	1	2	5
mw1.medium	2	4	10
mw1.large	4	8	20

aws

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

Workers

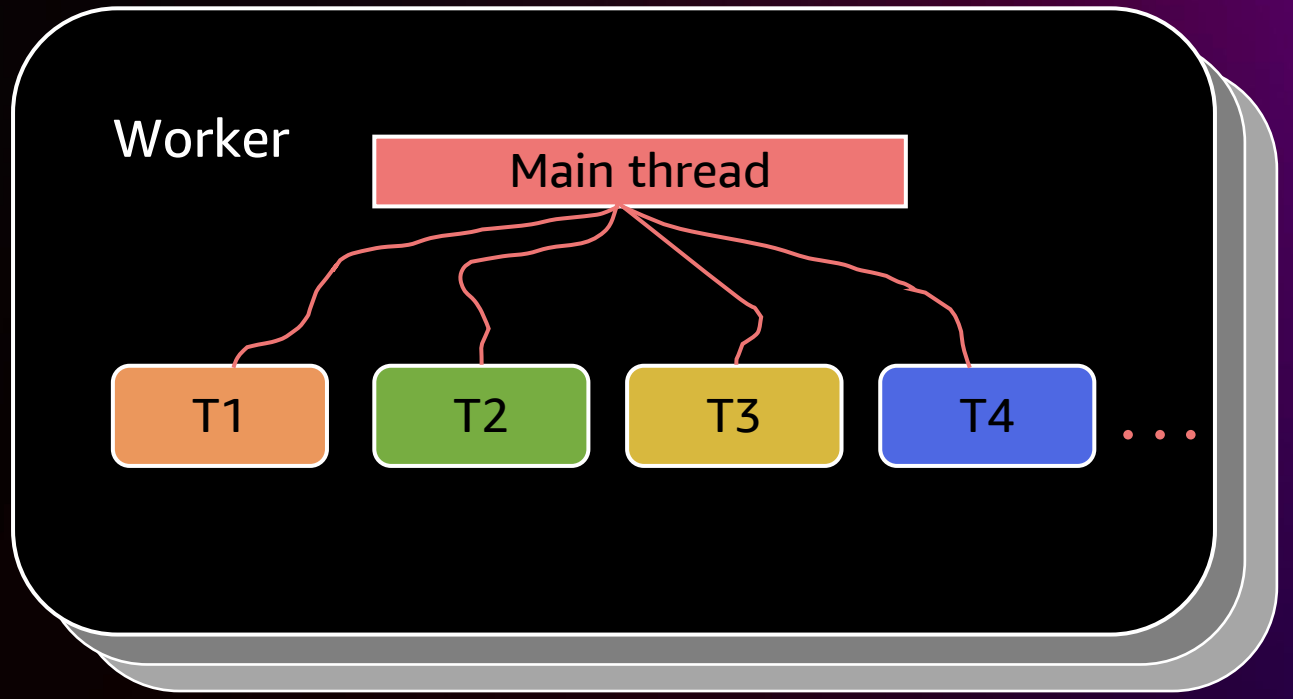
- Consumes tasks
- Parses Dag file with every single task
- Runs multiple tasks concurrently
- Updates state



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

Workers

- Consumes tasks
- Parses Dag file with every single task
- Runs multiple tasks concurrently
- Updates state

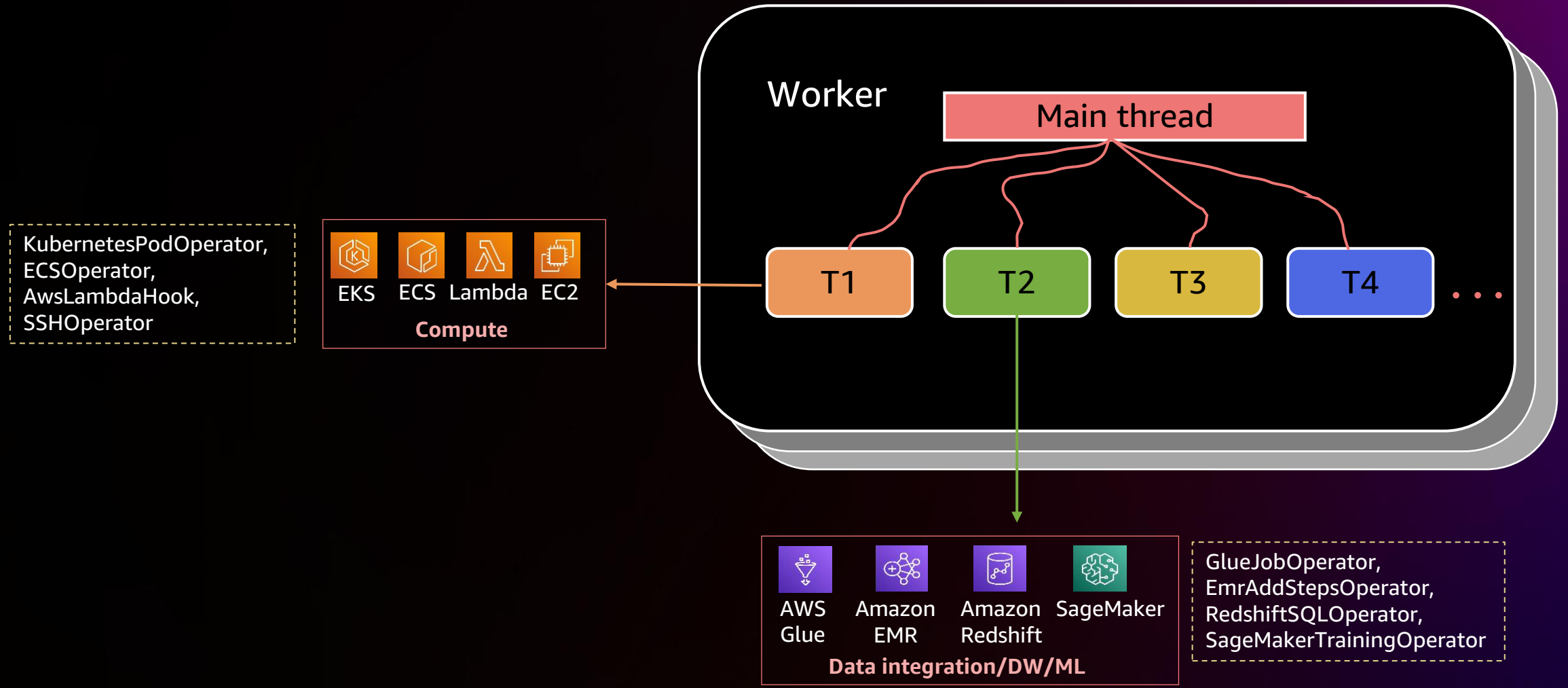


Large environment, min workers = 9, max workers = 15

Min concurrency = 9 * 20 = 180
Max concurrency = 15 * 20 = 300

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

Externalize tasks



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

Choices to offload compute

Containers



Amazon Elastic Container Service (Amazon ECS)



Amazon Elastic Kubernetes Service (Amazon EKS)

Serverless



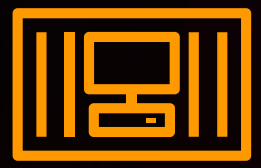
Lambda



Fargate

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

Managed containers



Container services to store, orchestrate, and run compute

aws

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

Why managed containers?



Multiple choices of compute
Powerful simplicity with Fargate
Open flexibility with EKS
Run Anywhere



Focus on business value
Managed control plane and
managed data plane with AWS Fargate



Per-second billing
Compute-based pricing model



Integration with existing tool chains
Familiarity and preference

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

AWS Lambda



AWS Lambda

Event-driven
serverless compute

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

Why AWS Lambda?



Focus on business value

No infrastructure provisioning



Flexible deployment

Deploy as Docker or bundle as zip



Pay for use

Request-based pricing model



Faster to spin up

Creation and invocations in seconds

aws

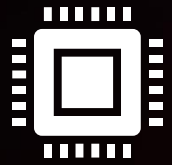
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architectural considerations

TECHNICAL CONSTRAINTS



Running time



CPU type



Package size



Network modes



Readiness

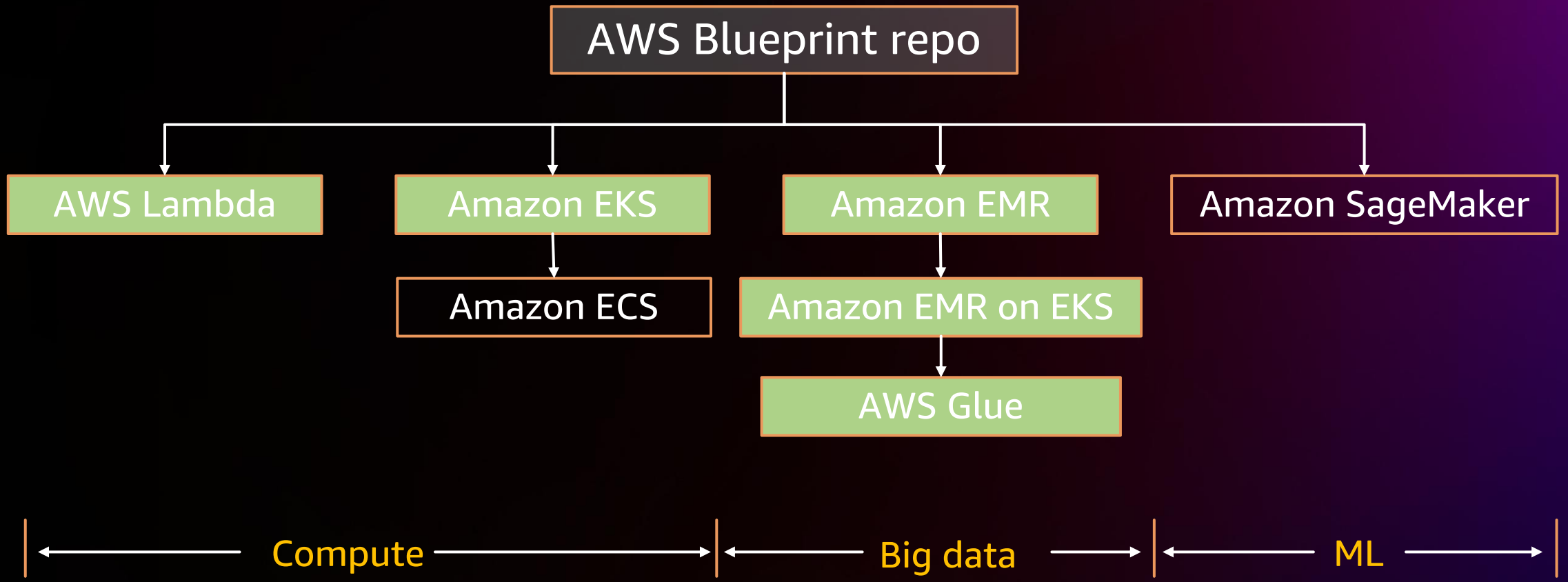
aws

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

Getting started . . .

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

Blueprints



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

Call to action

1. Identify computational tasks that can/should be offloaded
2. Evaluate the appropriate compute to run tasks
3. Dive deeper into AWS Lambda and Amazon container service features
4. Choose Blueprints and samples
5. Reach out to your account team

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

Check out these other sessions

API304: Amazon MWAA at scale

Wednesday (November 30) at 2:30 PM – Caesars Forum 104

BOA312: Orchestrating hybrid workflows with Apache Airflow

Thursday (December 1) at 11:00 AM – Caesars Forum, Alliance 311

SVS203: Getting started with serverless

Monday (November 28) at 1:45 PM – Mandalay Bay, Level 1 North

CON309: Build your application easily & efficiently with serverless containers

Thursday (December 1) at 2:00 PM – Encore Level 1, Encore Ballroom 5

aws

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

Thank you!



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

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