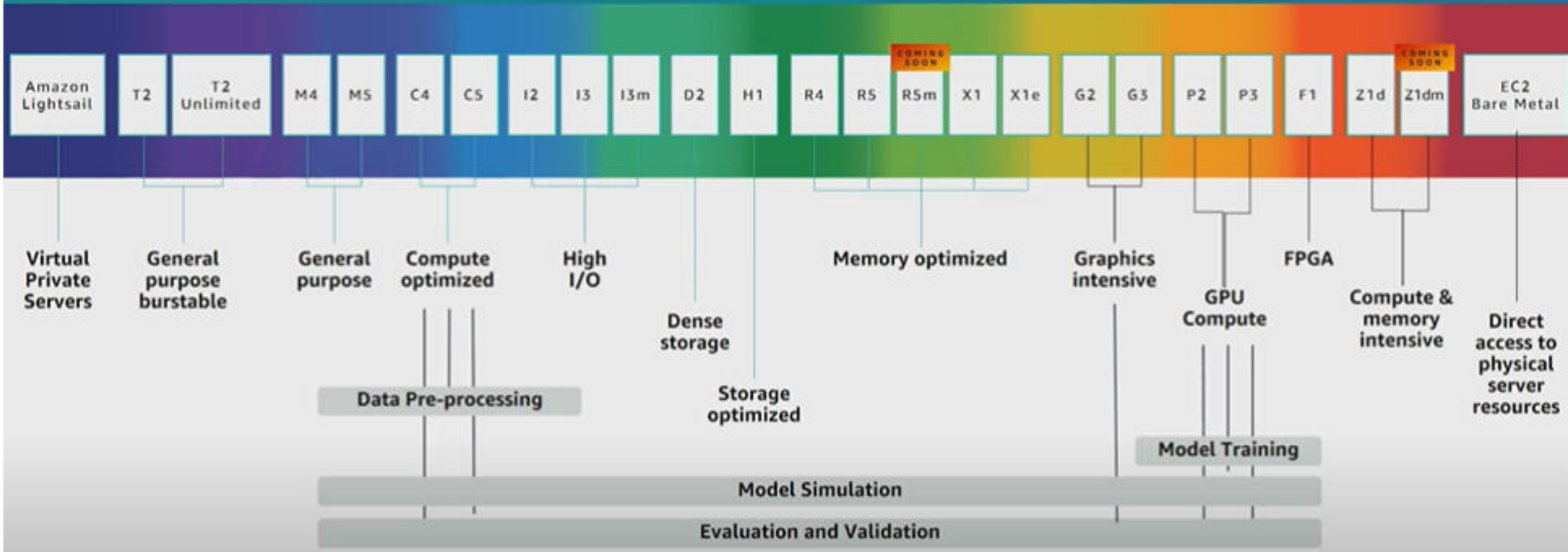
Wide Range of Compute Instances for HPC/ML Workloads

aws training and certification



EC2 Elastic GPUs

Graphics acceleration for EC2 instances

EC2 Fleet

- Simplified provisioning
- Massive scale
- Flexible capacity allocation

Comprehensive Portfolio of High Performance Storage Options

aws training and certification

Block storage

Amazon EBS

High performance attached storage with 99.999% availability. Tune size and performance with elastic volumes

File storage

Amazon EFS

Petabyte, elastic file storage sharable across applications, instances and servers

Object storage

Amazon S3

Low cost, highly scalable cloud storage with 99.999999999% durability

AWS Batch

aws training and certification

Introducing AWS Batch

aws training and certification

Fully Managed

No software to install or servers to manage. AWS Batch provisions, manages, and scales your infrastructure

Integrated with AWS

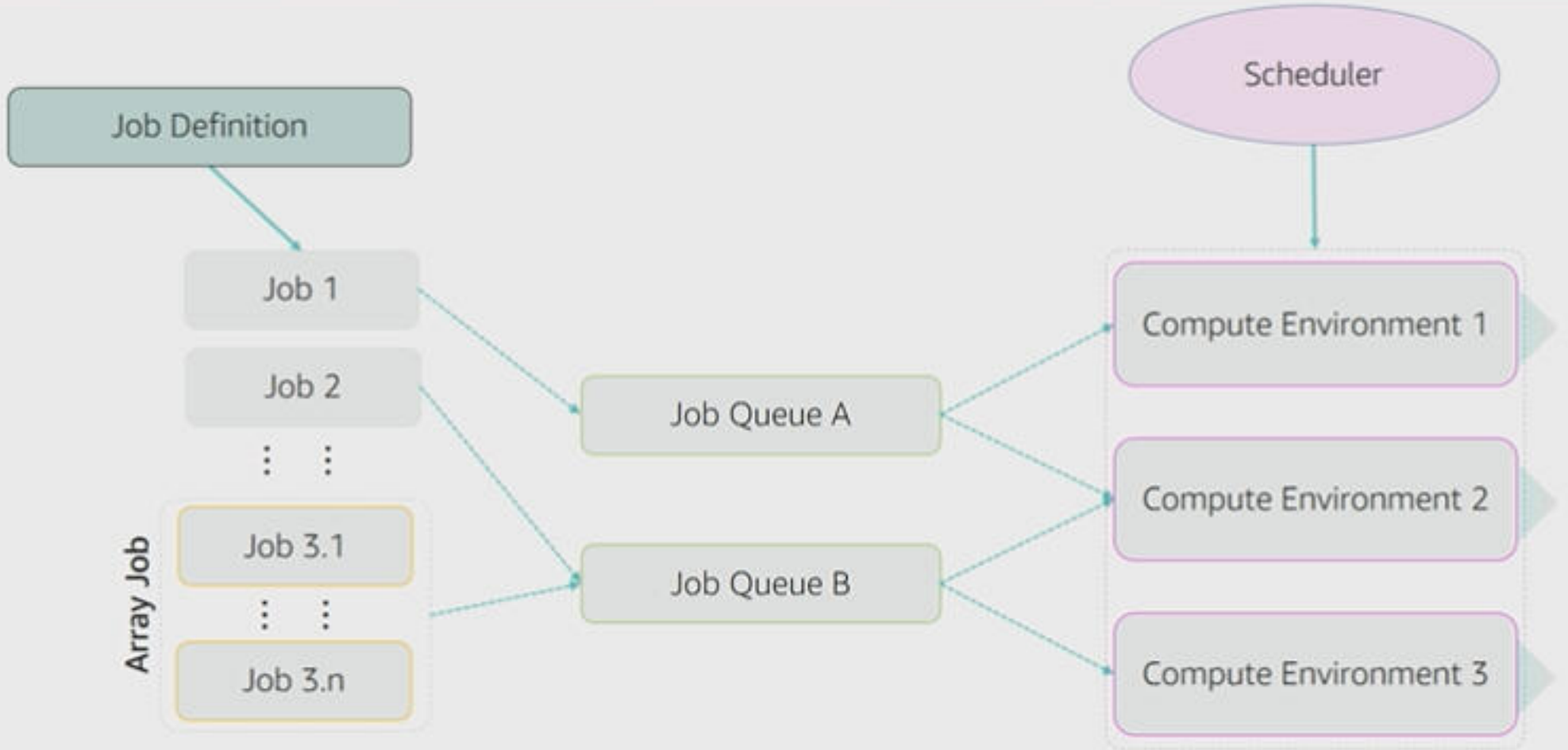
Natively integrated with the AWS Platform, AWS Batch jobs can easily and securely interact with services such as Amazon S3, Amazon DynamoDB, and Amazon Rekognition

Cost-optimized Resource Provisioning

AWS Batch automatically provisions compute resources tailored to the needs of your jobs using Amazon EC2 and EC2 Spot

AWS Batch Main Components

aws training and certification



AWS Batch Concepts

aws training and certification

- The Scheduler evaluates when, where, and how to run jobs that have been submitted to a job queue.
- Jobs run in approximately the order in which they are submitted as long as all dependencies on other jobs have been met.

AWS Batch Concepts

- The Scheduler evaluates when, where, and how to run jobs that have been submitted to a job queue.

- Jobs run in approximately the order in which they are submitted as long as all dependencies on other jobs have been met.

Jobs

aws training and certification

- Jobs are the unit of work executed by AWS Batch as containerized applications running on Amazon EC2.
- Containerized jobs can reference a container image, command, and parameters or users can simply provide a .zip containing their application and we will run it on a default Amazon Linux container.

```
aws batch submit-job --job-name sim-variation-1 \
                     --job-definition sim-sensors \
                     --job-queue high-mem-and-cpu
```

Jobs

aws training and certification

- Jobs are the unit of work executed by AWS Batch as containerized applications running on Amazon EC2.
- Containerized jobs can reference a container image, command, and parameters or users can simply provide a .zip containing their application and we will run it on a default Amazon Linux container.

```
aws batch submit-job --job-name sim-variation-1 \
                     --job-definition sim-sensors \
                     --job-queue high-mem-and-cpu
```

Jobs States

aws training and certification

- Jobs submitted to a queue can have the following states:

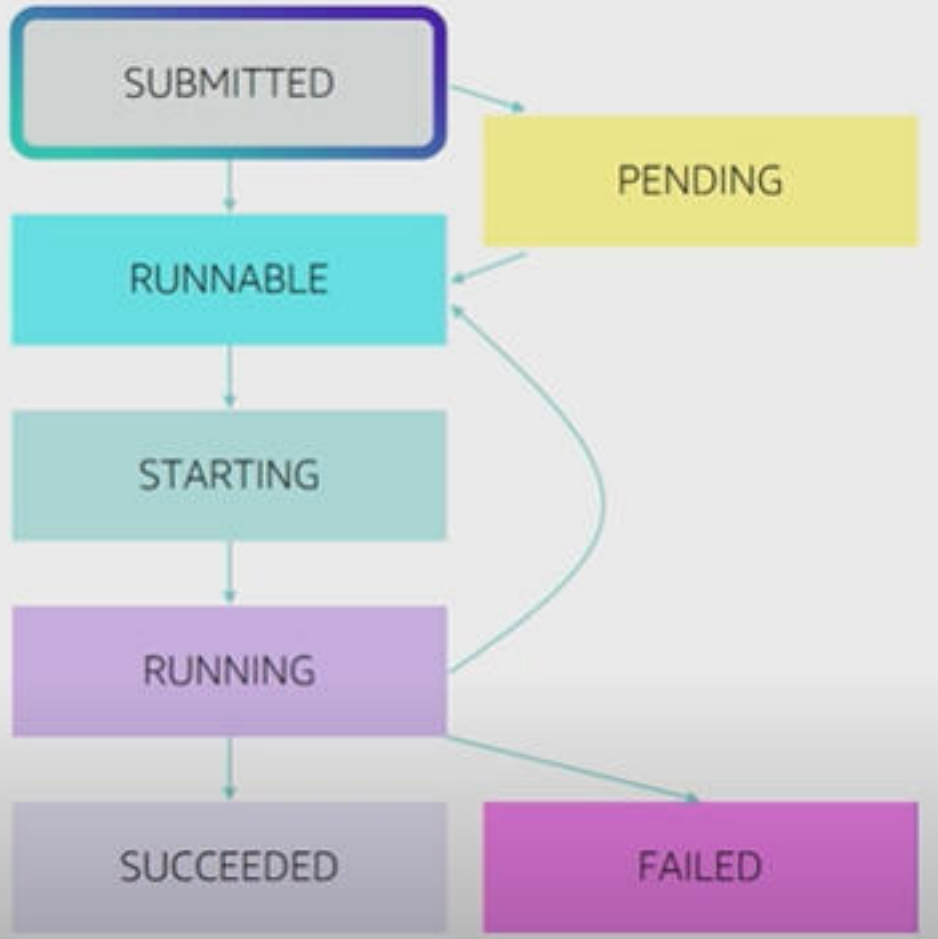
  - SUBMITTED: Accepted into the queue, but not yet evaluated for execution
  - PENDING: Your job has dependencies on other jobs which have not yet completed
  - RUNNABLE: Your job has been evaluated by the scheduler and is ready to run
  - STARTING: Your job is in the process of being scheduled to a compute resource
  - RUNNING: Your job is currently running
  - SUCCEEDED: Your job has finished with exit code 0
  - FAILED: Your job finished with a non-zero exit code or was cancelled or terminated.

Jobs States

aws training and certification

- Jobs submitted to a queue can have the following states:

  - SUBMITTED: Accepted into the queue, but not yet evaluated for execution
  - PENDING: Your job has dependencies on other jobs which have not yet completed
  - RUNNABLE: Your job has been evaluated by the scheduler and is ready to run
  - STARTING: Your job is in the process of being scheduled to a compute resource
  - RUNNING: Your job is currently running
  - SUCCEEDED: Your job has finished with exit code 0
  - FAILED: Your job finished with a non-zero exit code or was cancelled or terminated.

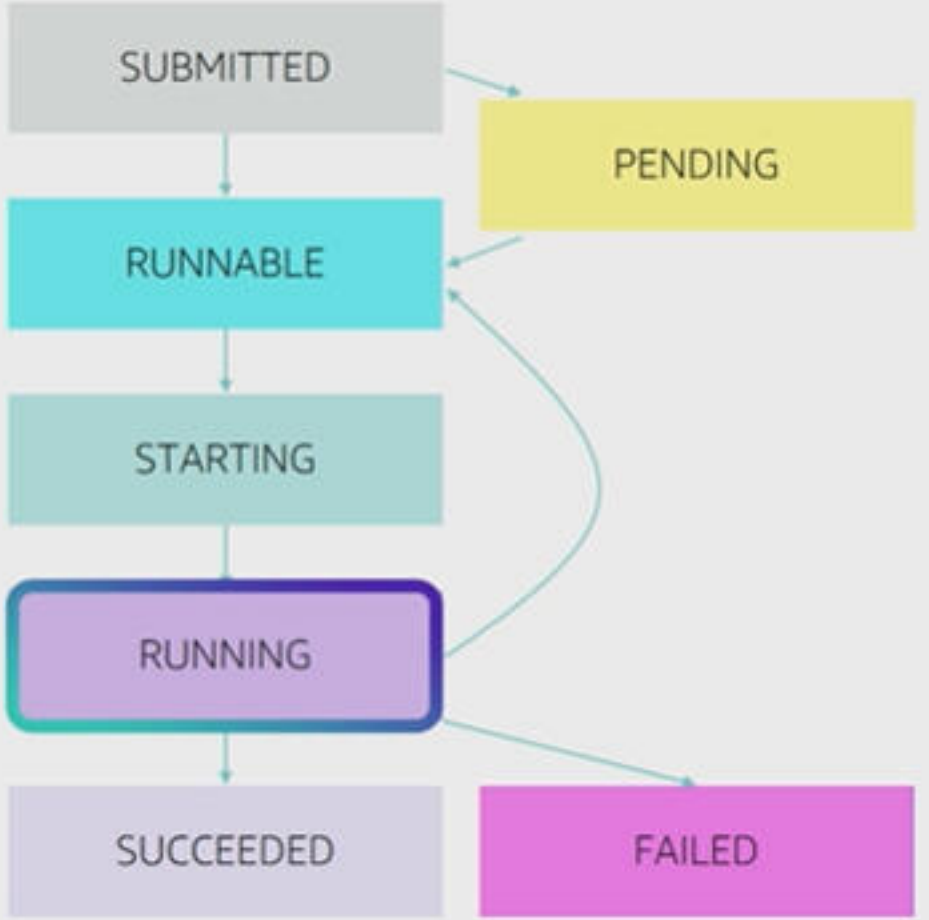


Jobs States

aws training and certification

- Jobs submitted to a queue can have the following states:

  - SUBMITTED: Accepted into the queue, but not yet evaluated for execution
  - PENDING: Your job has dependencies on other jobs which have not yet completed
  - RUNNABLE: Your job has been evaluated by the scheduler and is ready to run
  - STARTING: Your job is in the process of being scheduled to a compute resource
  - RUNNING: Your job is currently running
  - SUCCEEDED: Your job has finished with exit code 0
  - FAILED: Your job finished with a non-zero exit code or was cancelled or terminated.

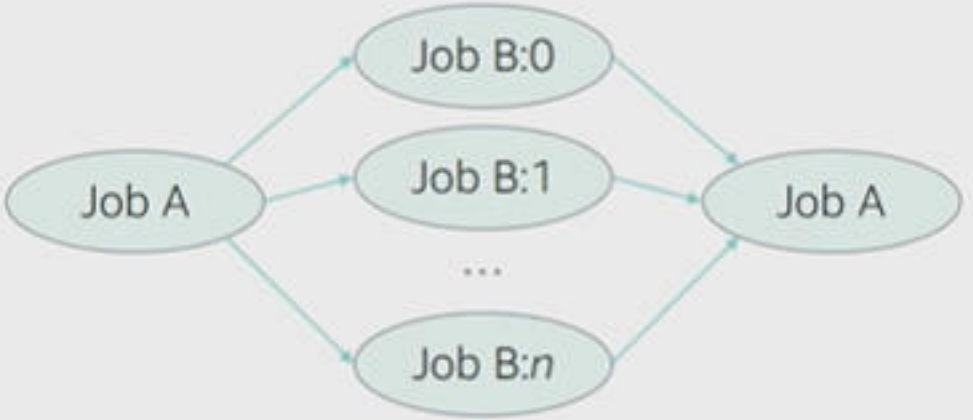


Array Jobs

aws training and certification

- Instead of submitting a large number of independent "simple jobs", we also support "array jobs" that run many copies of an application against an array of elements.

- Array jobs are an efficient way to run:
  - Parametric sweeps
  - Monte Carlo simulations
  - Processing a large collection of objects




```
aws batch submit-job --job-name sim-variation-1 \
                     --job-definition sim-sensors \
                     --job-queue high-mem-and-cpu \
                     --array-properties "{"size": 300}"
```

Job Queues

aws training and certification

- Jobs are submitted to a Job Queue, where they reside until they are able to be scheduled to a compute resource. Information related to completed jobs persists in the queue for 24 hours.

```
aws batch create-job-queue --job-queue-name simulation-high-priority \
                           --state ENABLED
                           --priority 500
                           --compute-environment-order \
                              order=1,computeEnvironment="cpu-high"
```

Compute Environments

aws training and certification

- Job queues are mapped to one or more Compute Environments containing the EC2 instances used to run containerized batch jobs.
- Managed compute environments enable you to describe your business requirements (instance types, min/max/desired vCPUs, and EC2 Spot bid as a % of On-Demand) and we launch and scale resources on your behalf.
- You can choose specific instance types (e.g. c5.18xlarge), instance families (e.g. C5, M5, P3), or simply choose "optimal" and AWS Batch will launch appropriately sized instances from our latest C/M/R instance families.

Compute Environments

aws training and certification

- Job queues are mapped to one or more Compute Environments containing the EC2 instances used to run containerized batch jobs.
- Managed compute environments enable you to describe your business requirements (instance types, min/max/desired vCPUs, and EC2 Spot bid as a % of On-Demand) and we launch and scale resources on your behalf.
- You can choose specific instance types (e.g. c5.18xlarge), instance families (e.g. C5, M5, P3), or simply choose "optimal" and AWS Batch will launch appropriately sized instances from our latest C/M/R instance families.

Compute Environments

aws training and certification

- Alternatively, you can launch and manage your own resources within an Unmanaged compute environment. Your instances need to include the ECS agent and run supported versions of Linux and Docker.

```
aws batch create-compute-environment --compute-environment-name
unmanagedce --type UNMANAGED ...
```

- AWS Batch will then create an Amazon ECS cluster which can accept the instances you launch. Jobs can be scheduled to your Compute Environment as soon as your instances are healthy and register with the ECS Agent.

Job Definitions

aws training and certification

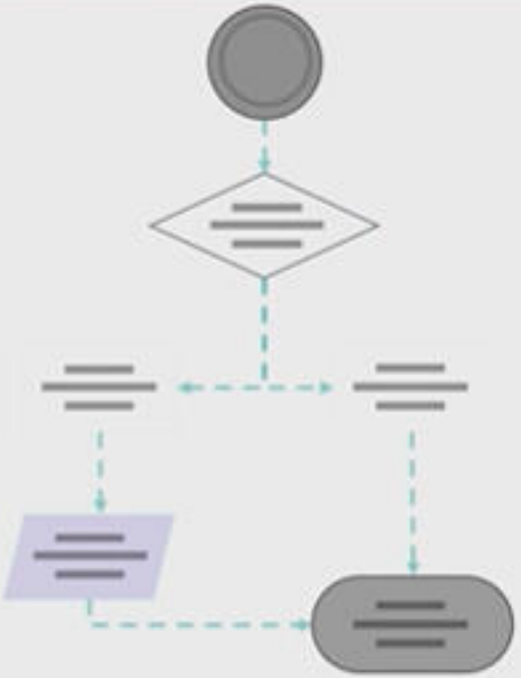
- Similar to ECS Task Definitions, AWS Batch Job Definitions specify how jobs are to be run. While each job must reference a job definition, many parameters can be overridden.
- Some of the attributes specified in a job definition:
  - IAM role associated with the job
  - vCPU and memory requirements
  - Retry strategy
  - Mount points
  - Container properties
  - Environment variables

```
aws batch register-job-definition --job-definition-name sim\
                                  --container-properties ...
```

Workflows, Pipelines, and Job Dependencies

aws training and certification

- Jobs can express a dependency on the successful completion of other jobs or specific elements of an array job.
- Use your preferred workflow engine and language to submit jobs. Flow-based systems simply submit jobs serially, while DAG-based systems submit many jobs at once, identifying inter-job dependencies.

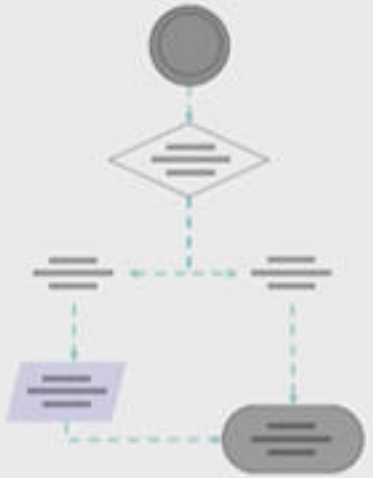


```
aws batch submit-job --depends-on 606b3ad1-aa31-48d8-92ec-f154bfc8215f
```

Workflows, Pipelines, and Job Dependencies

aws training and certification

- Jobs can express a dependency on the successful completion of other jobs or specific elements of an array job.
- Use your preferred workflow engine and language to submit jobs. Flow-based systems simply submit jobs serially, while DAG-based systems submit many jobs at once, identifying inter-job dependencies.



```
aws batch submit-job --depends-on 606b3ad1-aa31-48d8-92ec-f154bfc8215f
```

Array Job Dependency Models

aws training and certification

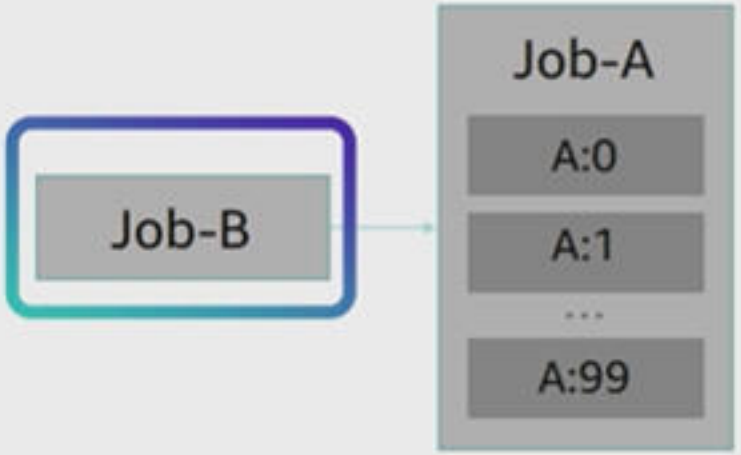
Job Depends on Array Job


```
$ aws batch submit-job -cli-input-json file://./Job-A.json

<Job-A.json>
{
   "jobName": "Job-A",
   "jobQueue": "ProdQueue",
   "jobDefinition": "Job-A-Definition:1",
   "arrayProperties":
   {
        "copies": 100
   }
}

$ aws batch submit-job -cli-input-json file://./Job-B.json

<Job-B.json>
{
   "jobName": "Job-B",
   "jobQueue": "ProdQueue",
   "jobDefinition": "Job-B-Definition:1",
   "dependsOn": [
        {"jobId": "<job ID for Job A>"  }
   ]
}
```

Array Job Dependency Models

aws training and certification

Array Job depends on Job




```
$ aws batch submit-job --job-name Job-A --job-queue
ProdQueue --job-definition job-A-Definition:1
{
        "jobName": "sequential-stress-10",
        "jobId": "7a6225f0-a16e-4241-9103-192c0c68124c"
}

<Job-B.json>
{
   "jobName": "Job-A",
   "jobQueue": "ProdQueue",
   "jobDefinition": "Job-A-Definition:1",
   "arrayProperties":
   {
        "copies": 100
   },
   "dependsOn": [
         {"jobId": "7a6225f0-a16e-4241-9103-192c0c68124c"
}
   ]
}
```

Array Job Dependency Models

aws training and certification

Array Job Depends on Self, a.k.a. Sequential Job


```
$ aws batch submit-job --job-name Job-A --job-queue
ProdQueue --job-definition job-A-Definition:1 --
array-properties size=10 --depends-on type=SEQUENTIAL
{
      "jobName": "Job-A",
      "jobId": "7a6225f0-a16e-4241-9103-192c0c68124c"
}
```

Model Example

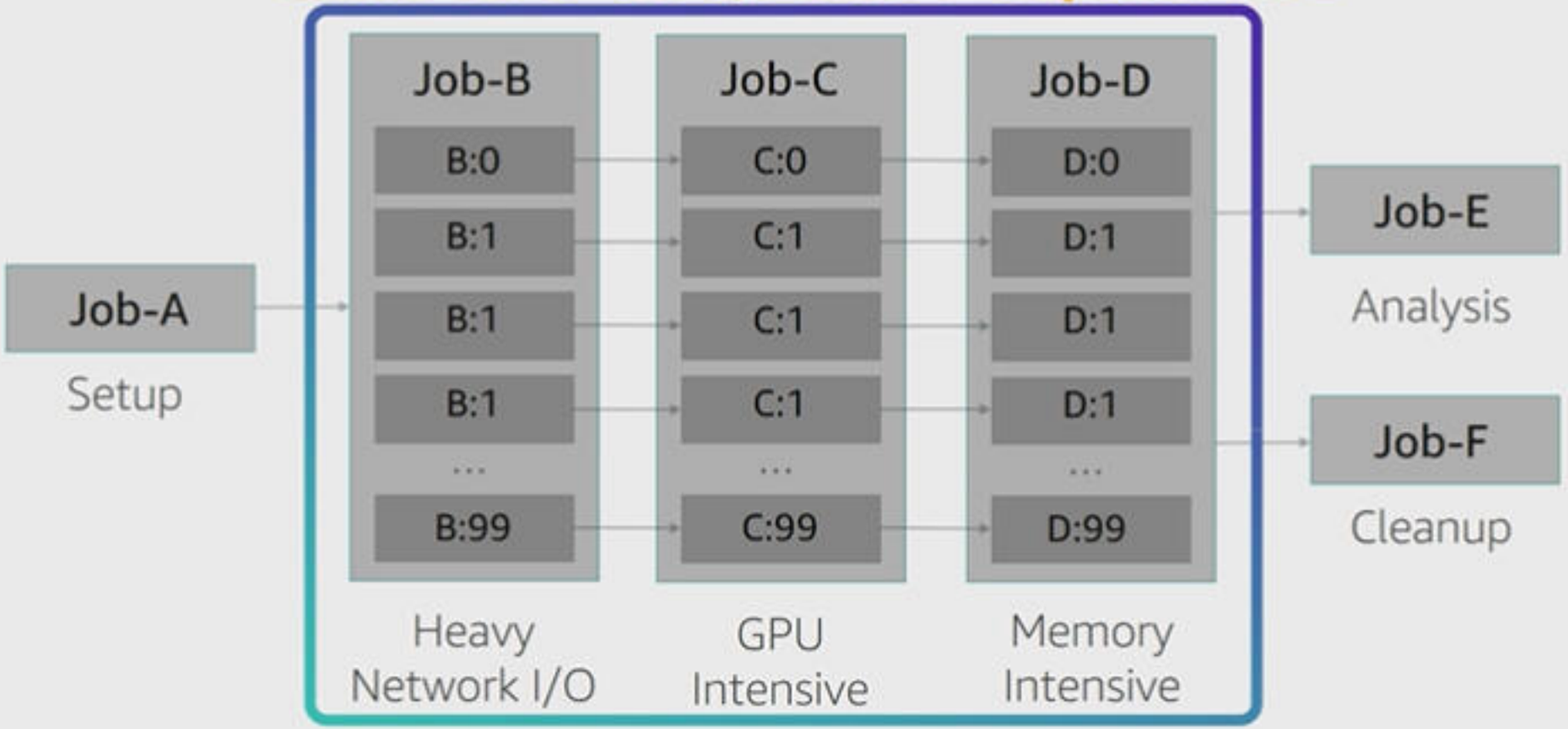
aws training and certification

C is dependent on A, C N_TO_N dependency on B, same for D on C, E and F depend on D

Model Example

aws training and certification

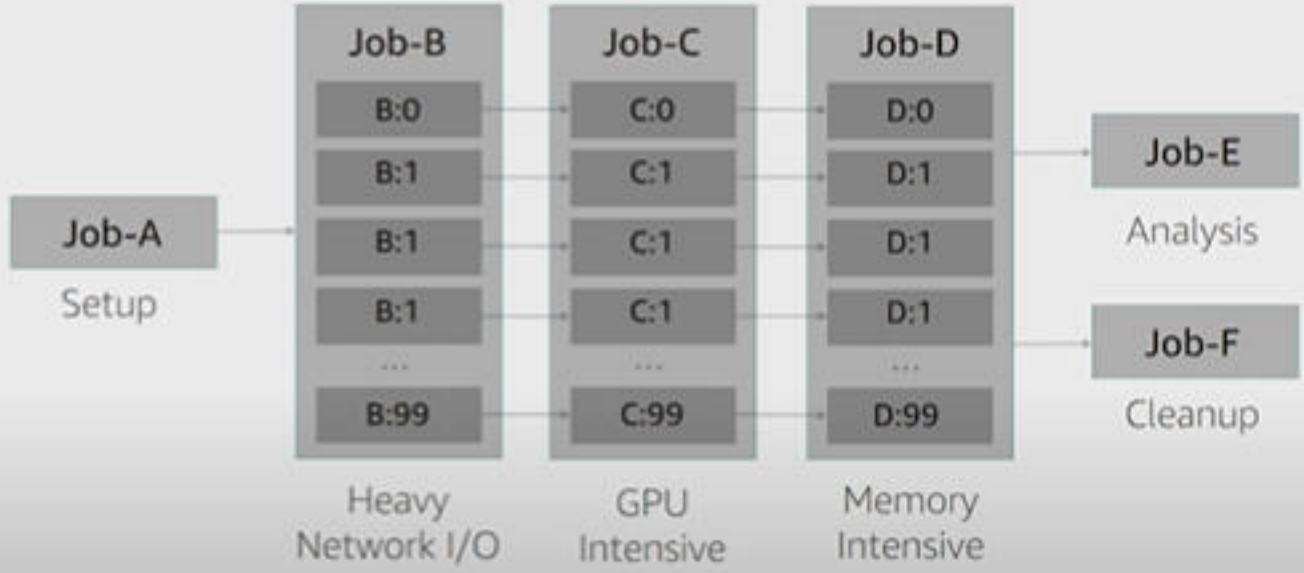
C is dependent on A, C N_TO_N dependency on B, same for D on C, E and F depend on D

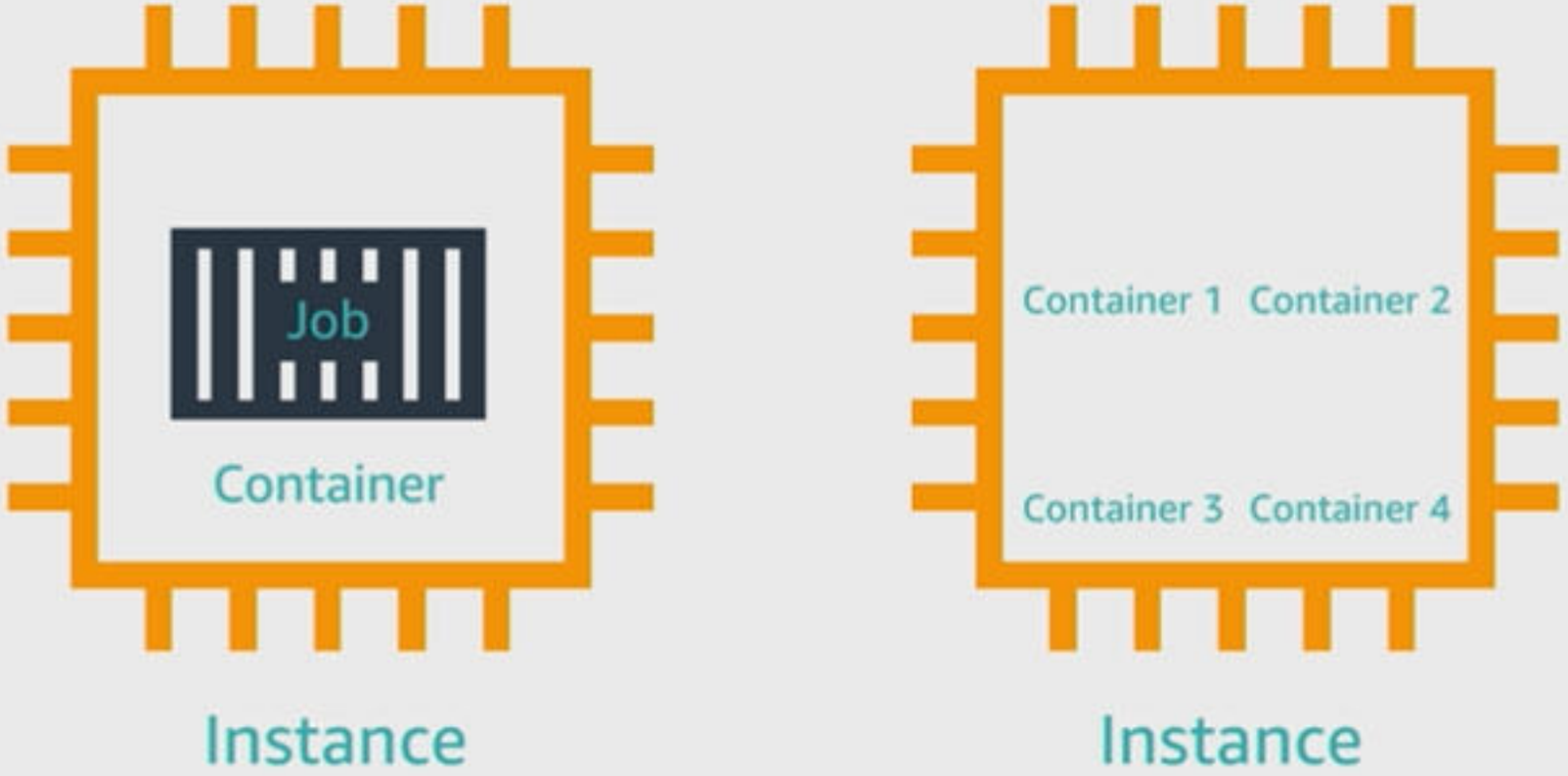


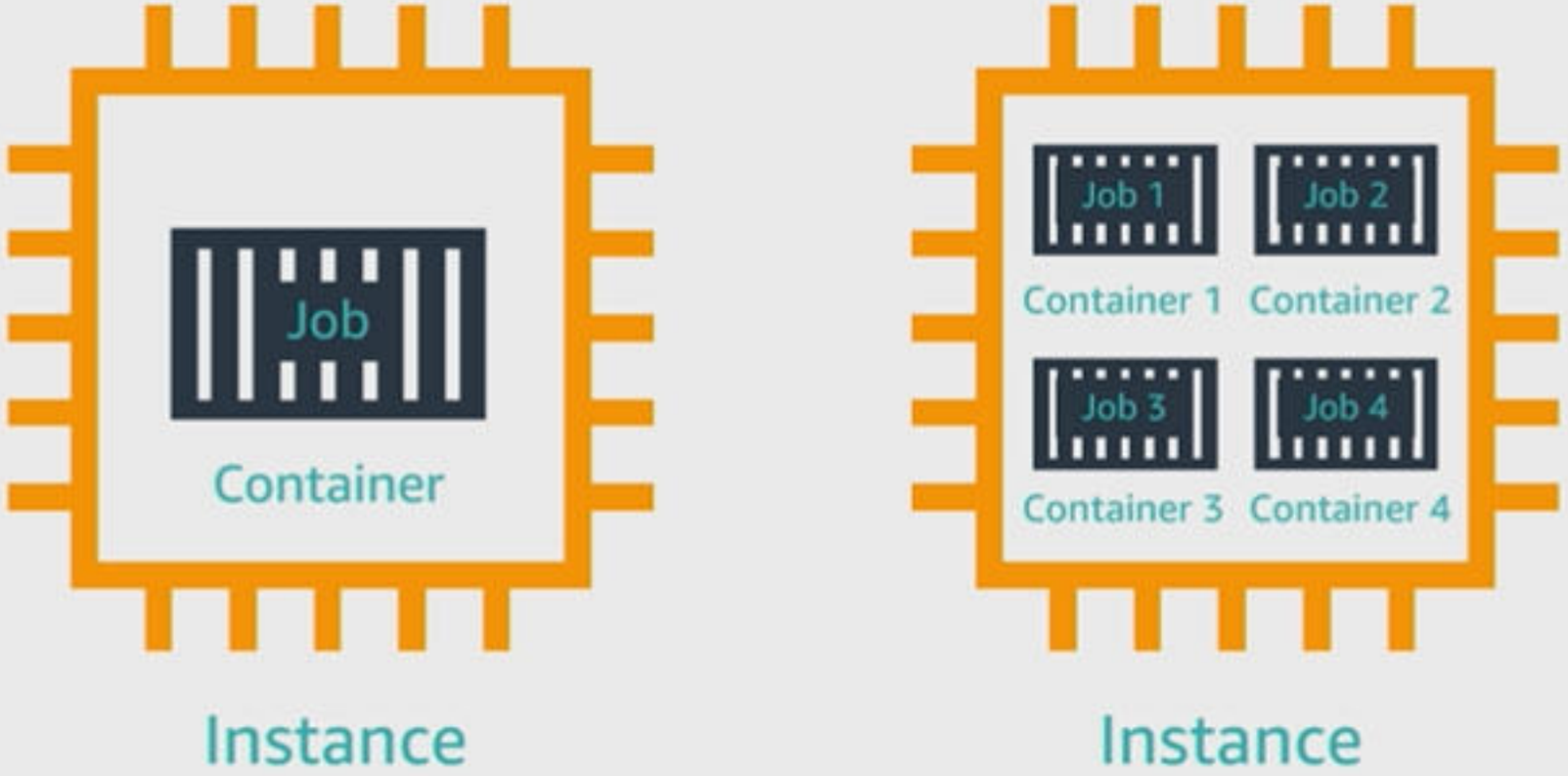
Model Example

aws training and certification

C is dependent on A, C N_TO_N dependency on B, same for D on C, E and F depend on D







Service Comparisons

aws training and certification

AWS Batch is one of many complementary services:

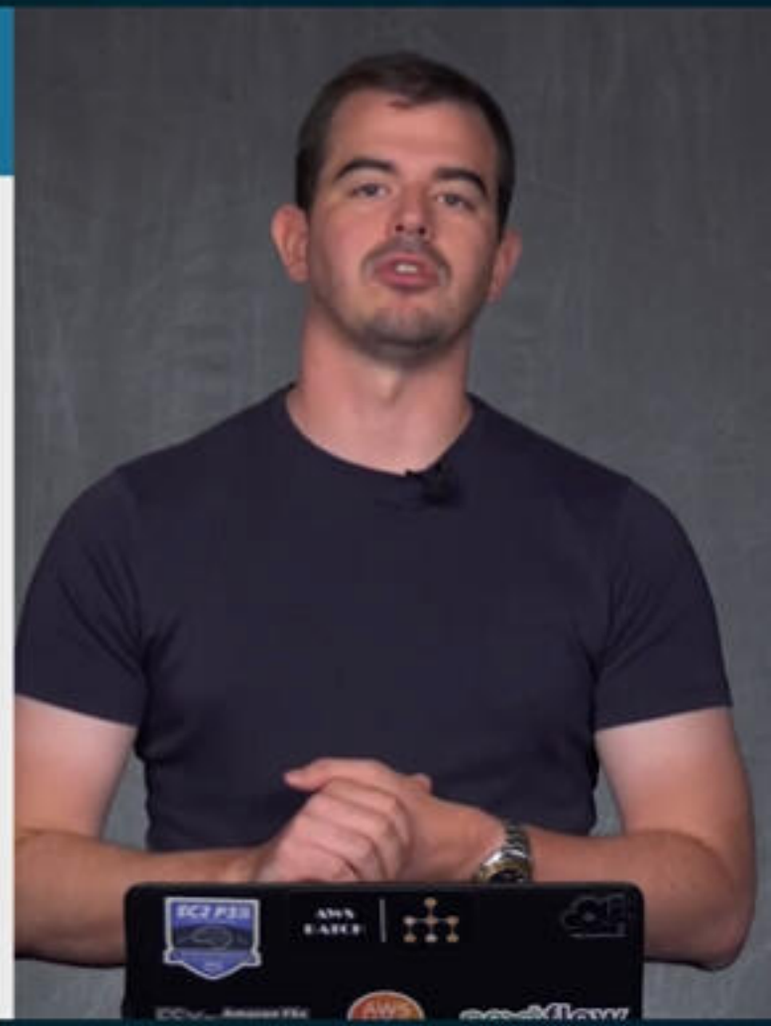
- ParallelCluster – Elastic HPC cluster that is ideal for tightly-coupled, latency sensitive applications, or when customers would like to use an OSS or commercial job scheduler
- Glue/DataPipeline – ETL to/from relational databases with known schemas
- EMR – Managed MapReduce clusters using Hadoop/Spark for large-scale data processing
- Lambda – Run short duration functions without provisioning or managing servers
- Step Functions/SWF – Design and orchestrate workflows, with support for branching and callouts to other AWS services. Often used in conjunction with AWS Batch.

Service Comparisons

aws training and certification

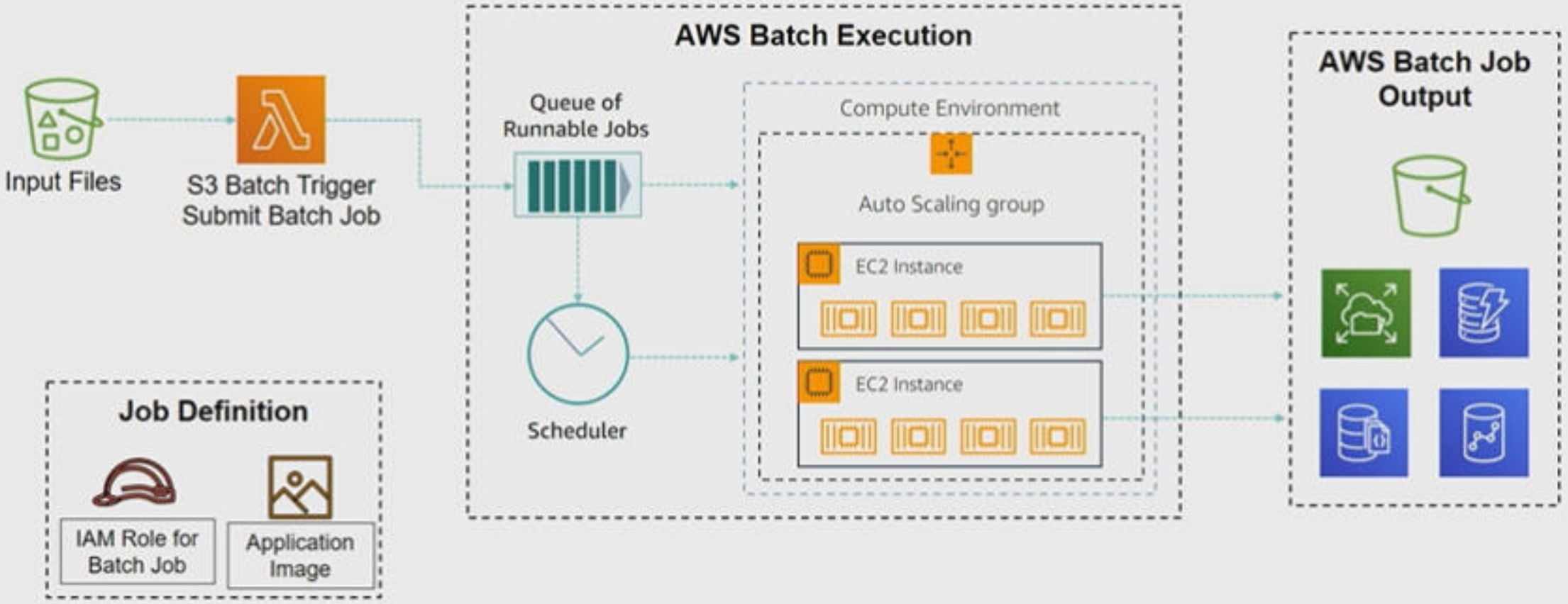
AWS Batch is one of many complementary services:

- ParallelCluster – Elastic HPC cluster that is ideal for tightly-coupled, latency sensitive applications, or when customers would like to use an OSS or commercial job scheduler
- Glue/DataPipeline – ETL to/from relational databases with known schemas
- EMR – Managed MapReduce clusters using Hadoop/Spark for large-scale data processing
- Lambda – Run short duration functions without provisioning or managing servers
- Step Functions/SWF – Design and orchestrate workflows, with support for branching and callouts to other AWS services. Often used in conjunction with AWS Batch.



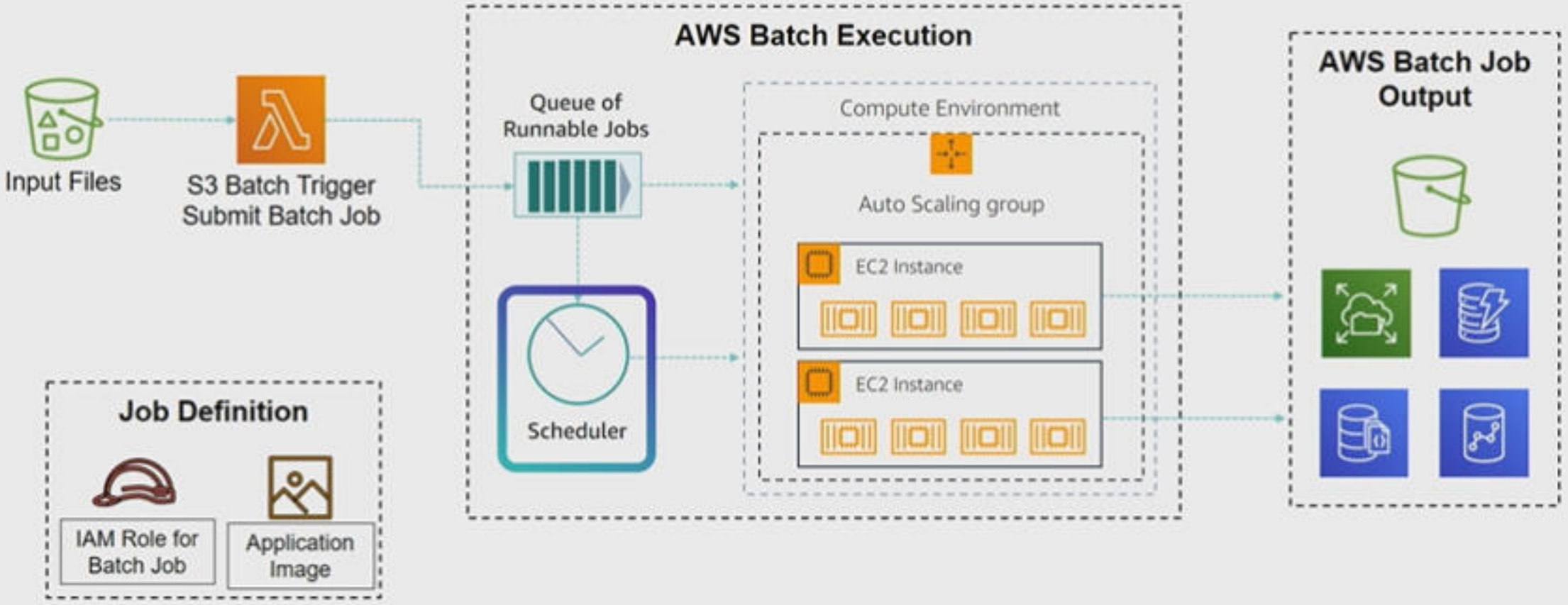
Typical Batch Architecture

aws training and certification



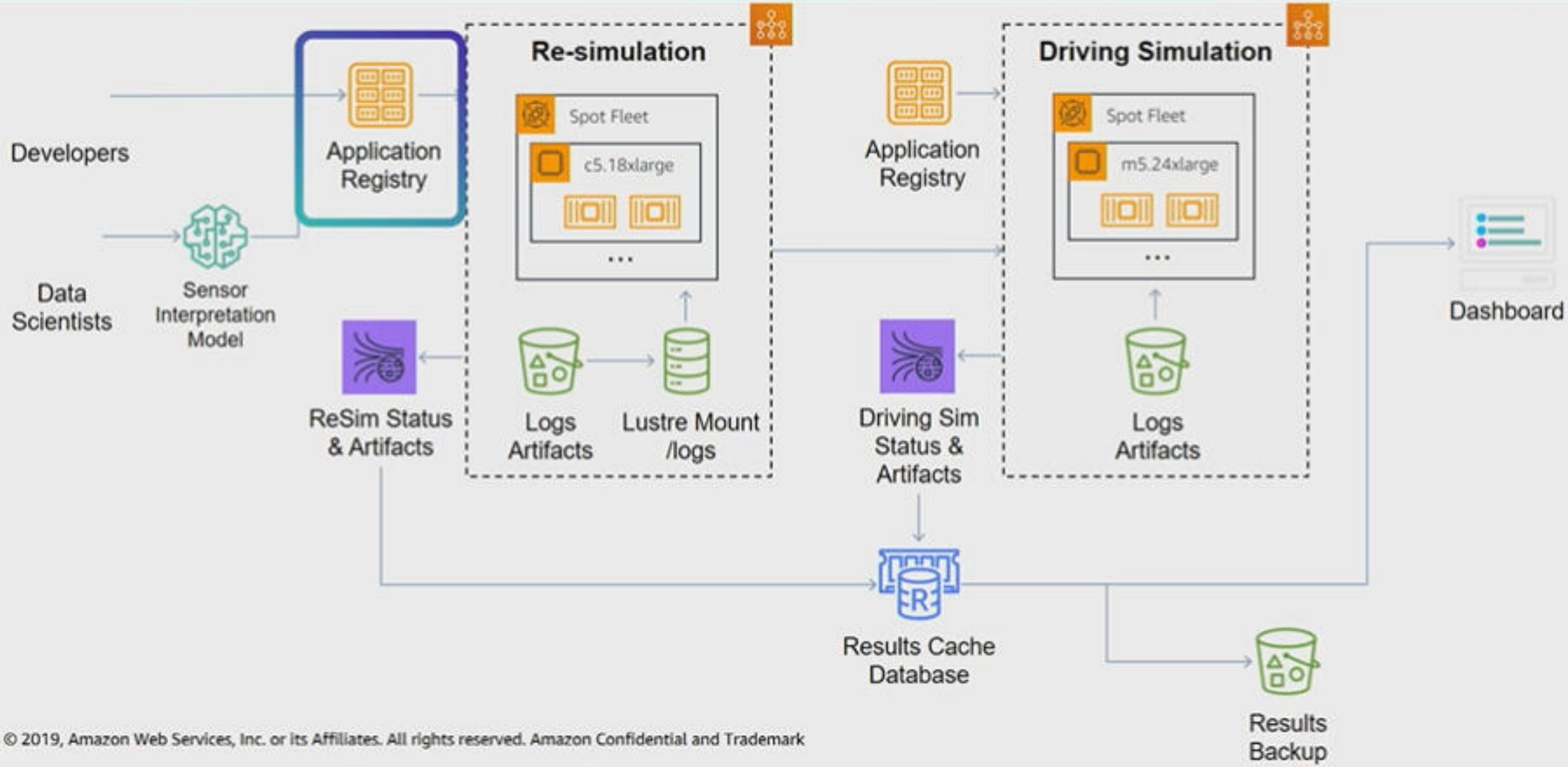
Typical Batch Architecture

aws training and certification



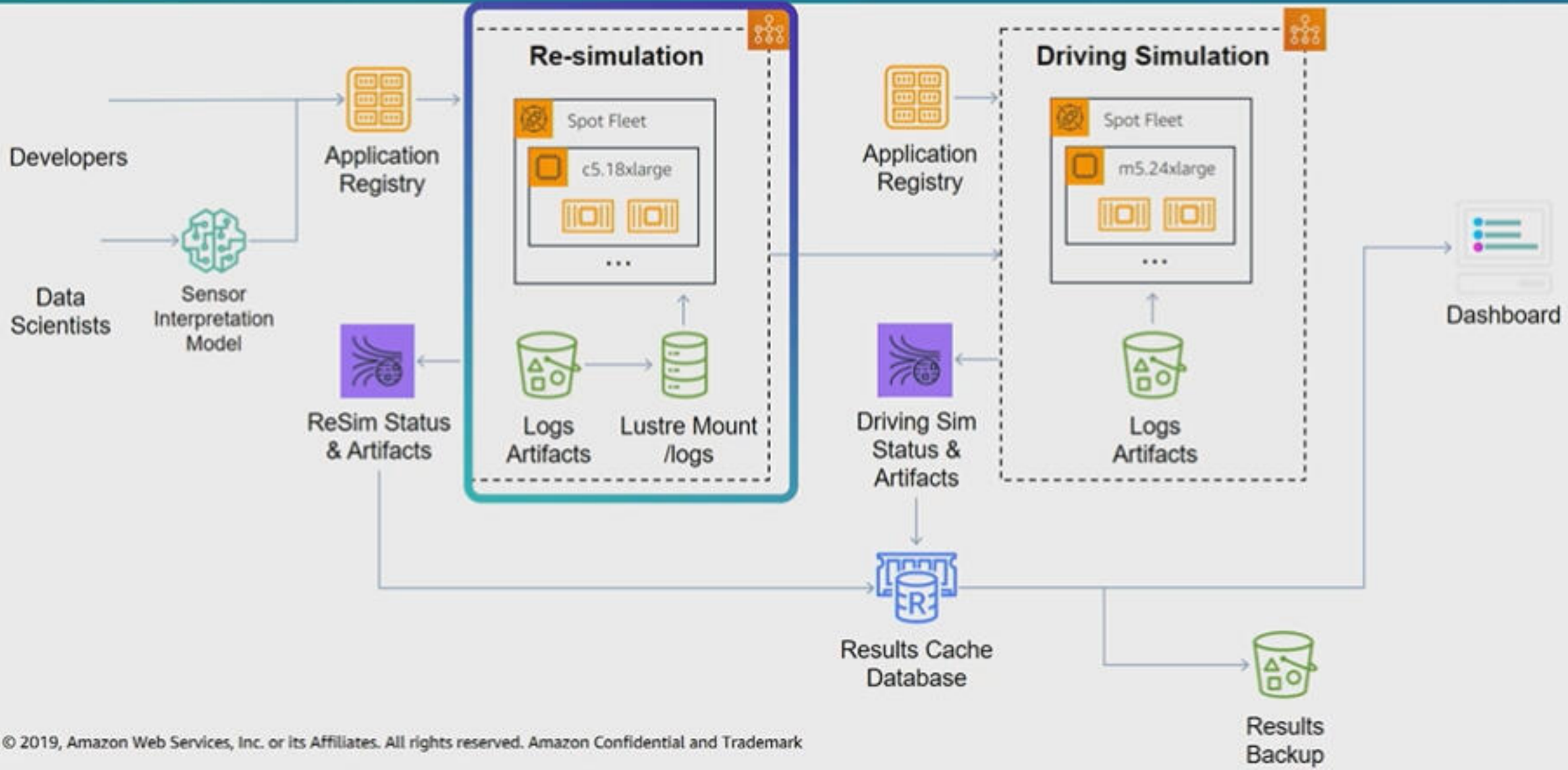
Autonomous Vehicle Simulations Workflow with Batch

aws training and certification



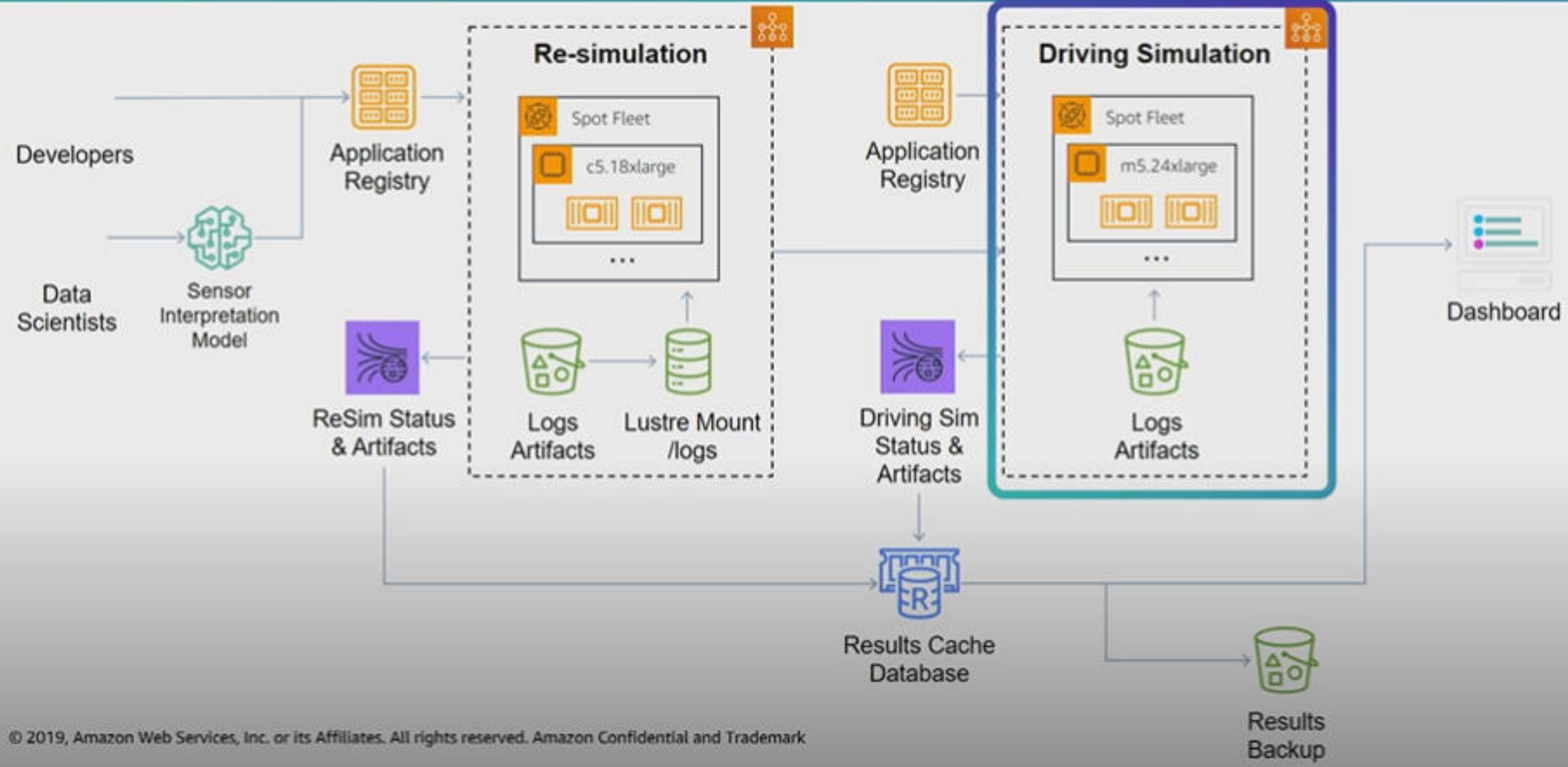
Autonomous Vehicle Simulations Workflow with Batch

aws training and certification



Autonomous Vehicle Simulations Workflow with Batch

aws training and certification

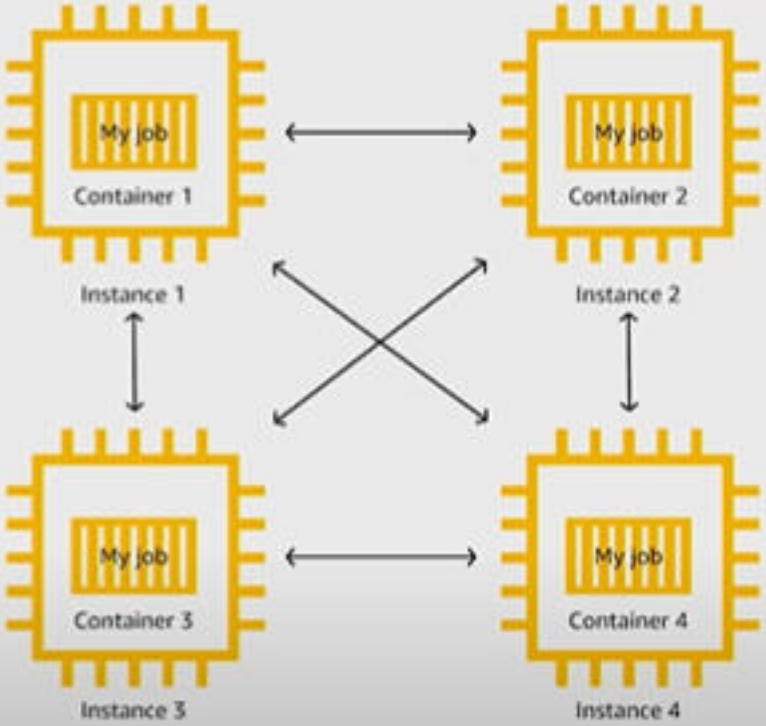


Multinode Parallel Jobs on AWS Batch

aws training and certification

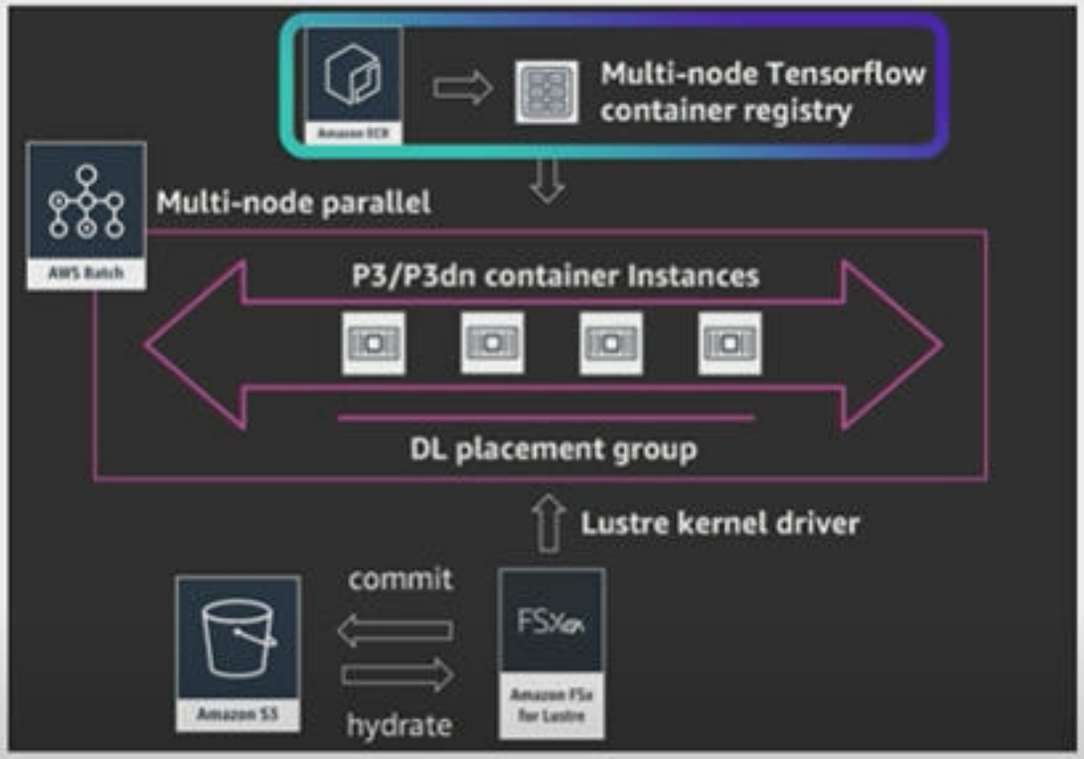
A MultiNode Parallel (MNP) Job enables AWS Batch to run single jobs which span multiple EC2 instances

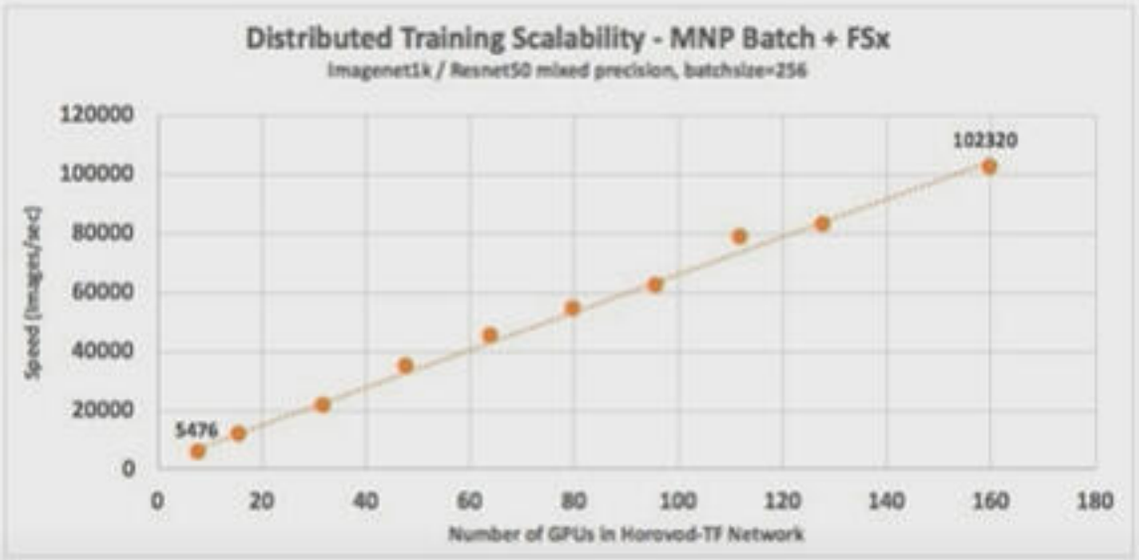
Integrated with the Elastic Fabric Adaptor for low latency between nodes



Deep Neural Network Training

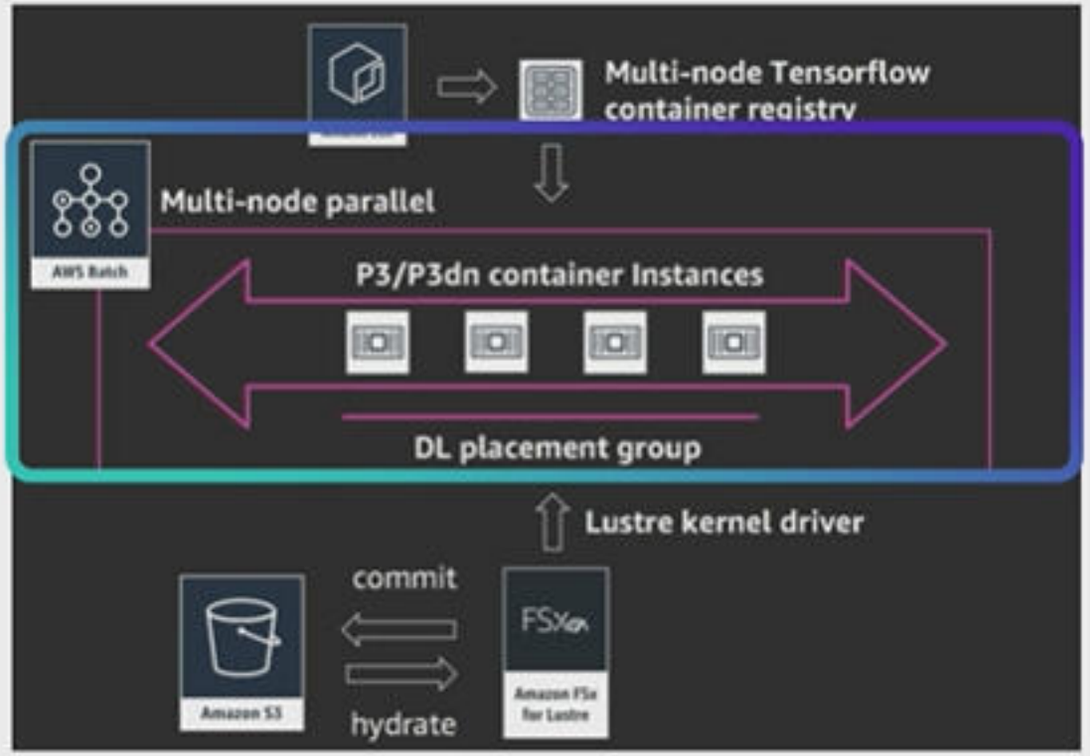
aws training and certification

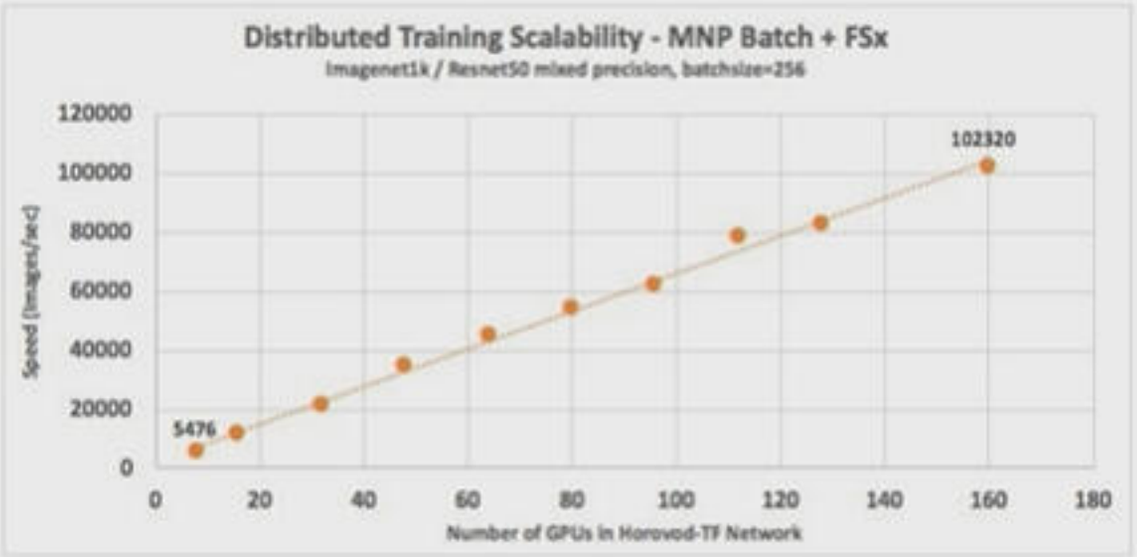




Deep Neural Network Training

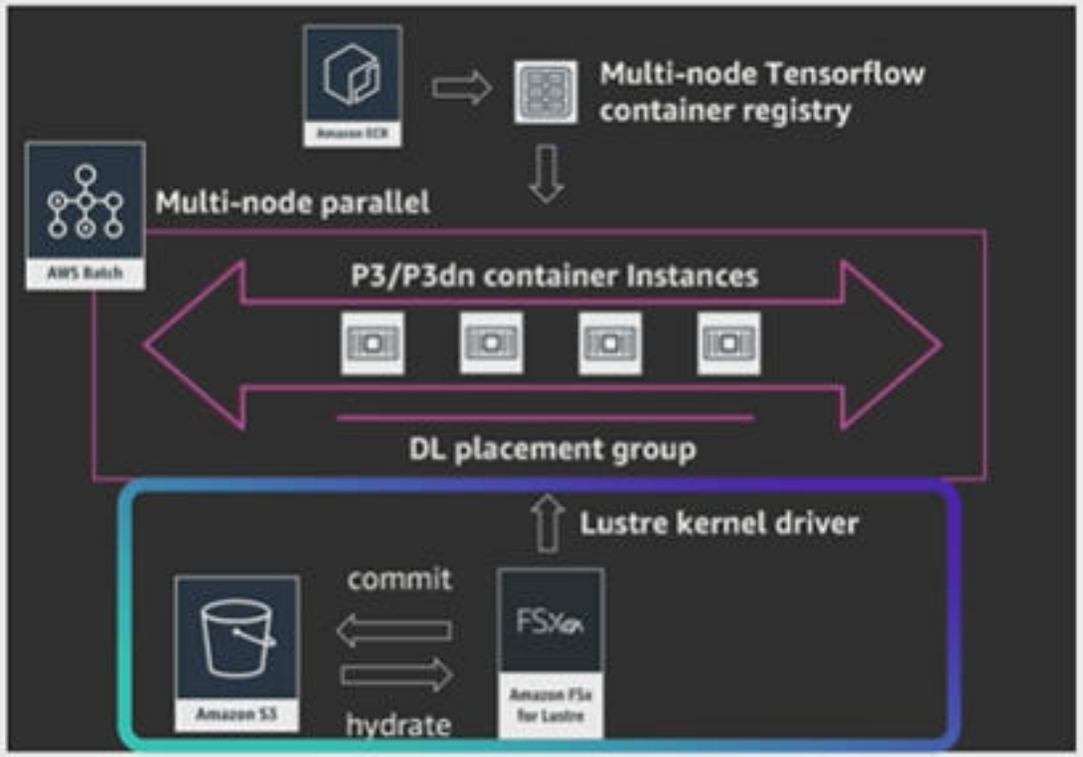
aws training and certification

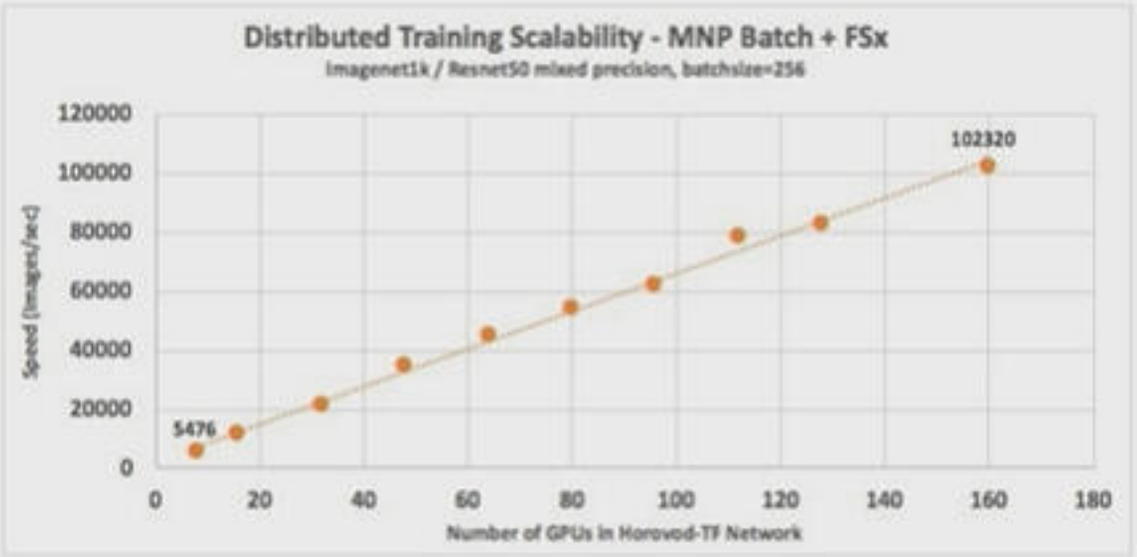




Deep Neural Network Training

aws training and certification





aws training and certification

Certificate of Completion

Hem Bahadur Gurung

Has successfully completed

Machine Learning in the Cloud with AWS Batch

Director, Training and Certification

30 minutes

Duration

10 September, 2021

Completion Date

Questions:

- What are the made assumptions?
- What is your learning target?
- What type of ML problem is it?
- Why did you choose this algorithm?
- How will you evaluate the model performance?
- How confident are you that you can generalize the results?

When is Machine Learning an Option?

- If the problem is persistent
- If the problem challenges progress or growth
- If the solution needs to scale
- If the problem requires personalization in order to be solved

What Does a Successful ML Solution Require?

- People
- Time
- Cost