AWS Batch

AWS Batch is used to manage and run batch computing workloads, running multiple jobs in parallel.

Example use cases:

- Analyzing financial risk models
- Media transcoding
- Engineering simulations

Lots of jobs running on vast resources

What is batch computing?

- Used for batch processing, requiring vast amounts of compute across a cluster of compute resources
- Outside of a cloud environment this is difficult to do
- Easily create a scalable, load-balanced cluster of compute resources, managed by AWS

AWS Batch Components

A job is a unit of work

Jobs

Units of work to be run by AWS Batch (for example a .exe file, an app, ECS Cluster or shell script).

They are run as containerized apps on EC2 instances and can have different states ('Submitted', 'Pending', 'Running', 'Failed', etc)

Parameters for running Jobs

Job Definitions

Parameters for jobs, determining how they will run and their configuration.

Examples:

- **How many vCPUs to use for a container**
- **Which data volumes to use**

Ordered List of scheduled jobs

Job Queues

Consist of **scheduled jobs**. You can have multiple queues with different priorities.

On-demand and Spot instances are supported, and AWS Batch can bid on your behalf for Spot instances.

Job Scheduling

When a Job should run and from which environment. Typically **first-in-first-out**

Contain compute resources. Can be managed or unmanaged

Compute Environments

Contain compute resources to carry out the Job.

Managed Environments

- AWS Batch manages instances based on your config parameters
- Created as Amazon ECS Cluster

Unmanaged Environments

• Provisioned, managed and maintained by you
• Greater customization, but requires more admin • You have to create the Amazon ECS cluster yourself