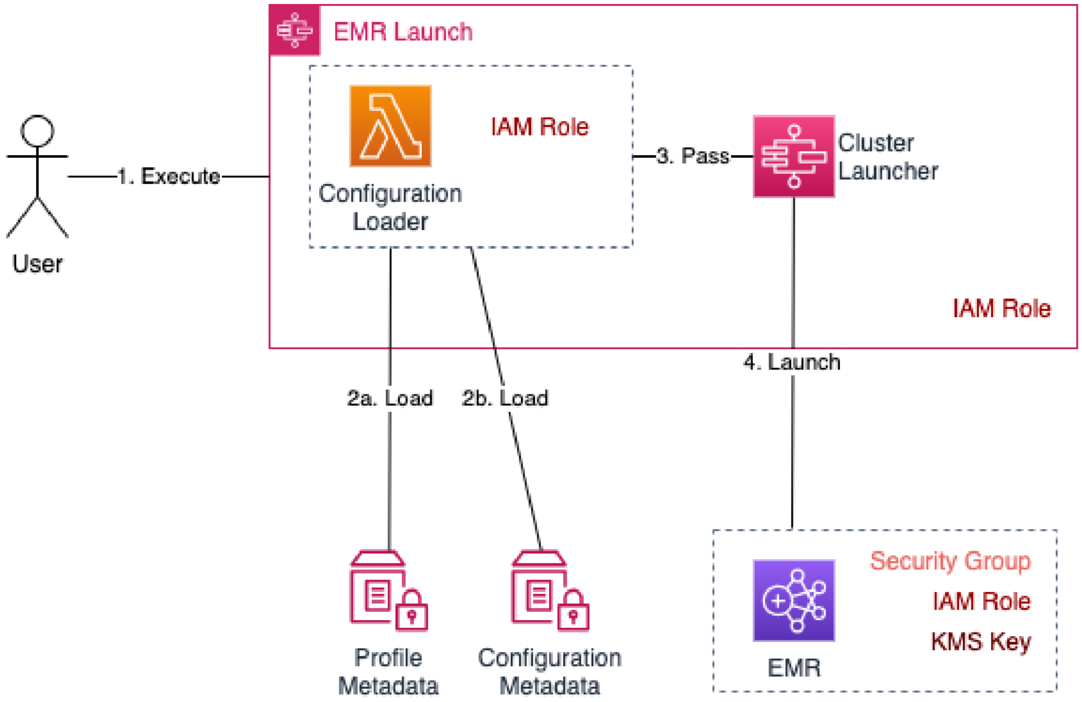
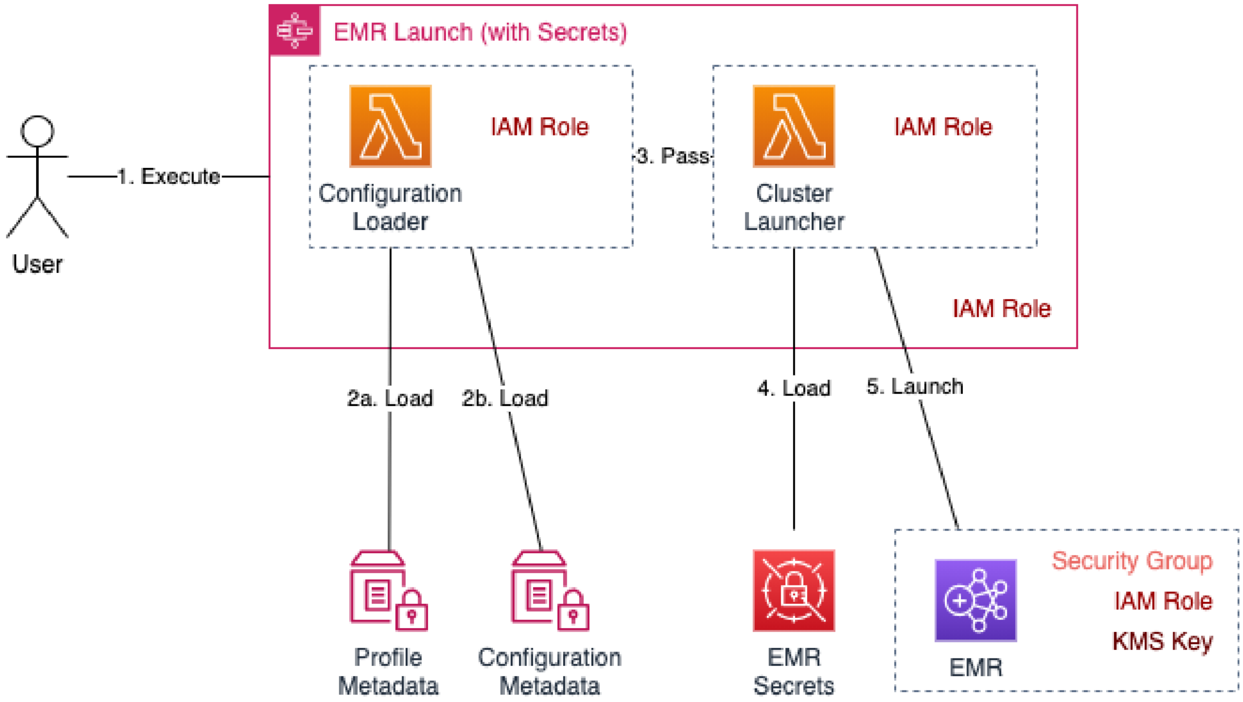
AWS EMR Launch Threat Model

Intro

AWS EMR Launch provides Step Functions preconfigured to launch EMR Clusters with predefined Profiles (IAM Roles, Instance Profiles, KMS Keys, Security Groups) and Configurations (Node Types/Counts, Configuration Files, EMR Versions, etc). Each Step Function is configured to load specific Profile and Configuration Metadata, ensuring that Users authorized to execute the Step Function are restrited to this Cluster Definition.

Two types of EMR Launch are supported: with and without Secrets. Secrets are configuration parameters like database/metastore credentials, Kerberos parameters, etc. Rather than storing the Secrets in Step Function definition, these are kept in Secrets Manager and loaded dynamically when the Cluster is launched.





Launching a Cluster (without Secrets)

1. The EMR Launch Step Function is executed. The IAM User/Role must be authorized the execute the Step Function
2. The Step Function utilizes a Lambda Function to load Profile and Configuration Metadata from the Parameter

Store

   a. This is a dedicated Lambda Function with an Execution Role granted access to only these specific Parameter Store values
   b. The Step Function Execution Role is granted execute on only this specific Lambda Function

3. Metadata is combined and passed to a Step Function EMR Integration Task
4. The EMR Integration Task launches the EMR Cluster
   a. The Step Function Execution Role is granted PassRole to only the specific IAM Role/Instance Profile defined in the Profile Metadata

Launching a Cluster (with Secrets)

1. The EMR Launch Step Function is executed. The IAM User/Role must be authorized the execute the Step Function
2. The Step Function utilizes a Lambda Function to load Profile and Configuration Metadata from the Parameter Store
   a. This is a dedicated Lambda Function with an Execution Role granted access to only these specific Parameter Store values
   b. The Step Function Execution Role is granted execute on only this specific Lambda Function
3. Metadata is combined and passed to a Cluster Launcher Lambda Function
4. The Cluster Launcher Lambda Function loads Secrets from the Secrets Manager, combines them with the Cluster Definition.
   a. This is a dedicated Lambda Function with an Execution Role granted access to the Secrets.
5. The Cluster Launcher Lambda Function launches the EMR Cluster
   a. The Lambda Execution Role is granted PassRole to only the specific IAM Role/Instance Profile defined in the Profile Metadata

Deployment

The library is used to create resources that can launch EMR Clusters into existing environments. Prerequisites include:

- VPC
- Administrative Role for creating launch resources

Dependencies

- AWS IAM
- Amazon S3
- Amazon EMR
- AWS Lambda
- AWS Step Functions
- AWS Systems Manager

Assets

Customer Assets

1. Customer data: None

2. Customer code: Python code used to define the CDK and EMR Launch Constructs. This is managed by the customer, preferably in a source control system.
3. Metadata: Environment metadata is stored in SSM Parameter Store values

Security Assets

1. KMS Keys: the solution utilizes the customers KMS Keys for encryption at rest.
2. IAM Roles: the solution can either create minimal IAM Roles used by the EMR Cluster and allow customers to attach their own Managed Policies, or let users declare the IAM Roles the want to use for the clusters.

Threats and Mitigations

Objective: Use a Launch Function to create an EMR Cluster with unauthorized permissions
Threat: A User can use the Launch Function to create a cluster with an EC2 Instance Profile/Role which they have not been authorized to use.
Mitigation: Execution of the Launch Function is controlled by IAM Policies/Roles. Users must be granted StartExecution for the specific Launch Function.

Objective: Modify Profile metadata to enable creation of EMR Clusters with alternate IAM Instance Profiles/Roles
Threat: By modifying the metadata in the SSM Parameter Store metadata, a User could modify the EC2 Instance Profile an EMR Cluster is launched with.
Mitigation: Each Launch Function uses a dedicated Lambda Function to access the SSM Parameter Store metadata. This Lambda is granted access to a only a specific Parameter Store value. In addition, the Launch Function is granted PassRole to only the EC2 Instance Profile/Role originally defined in the Parameter Store metadata.