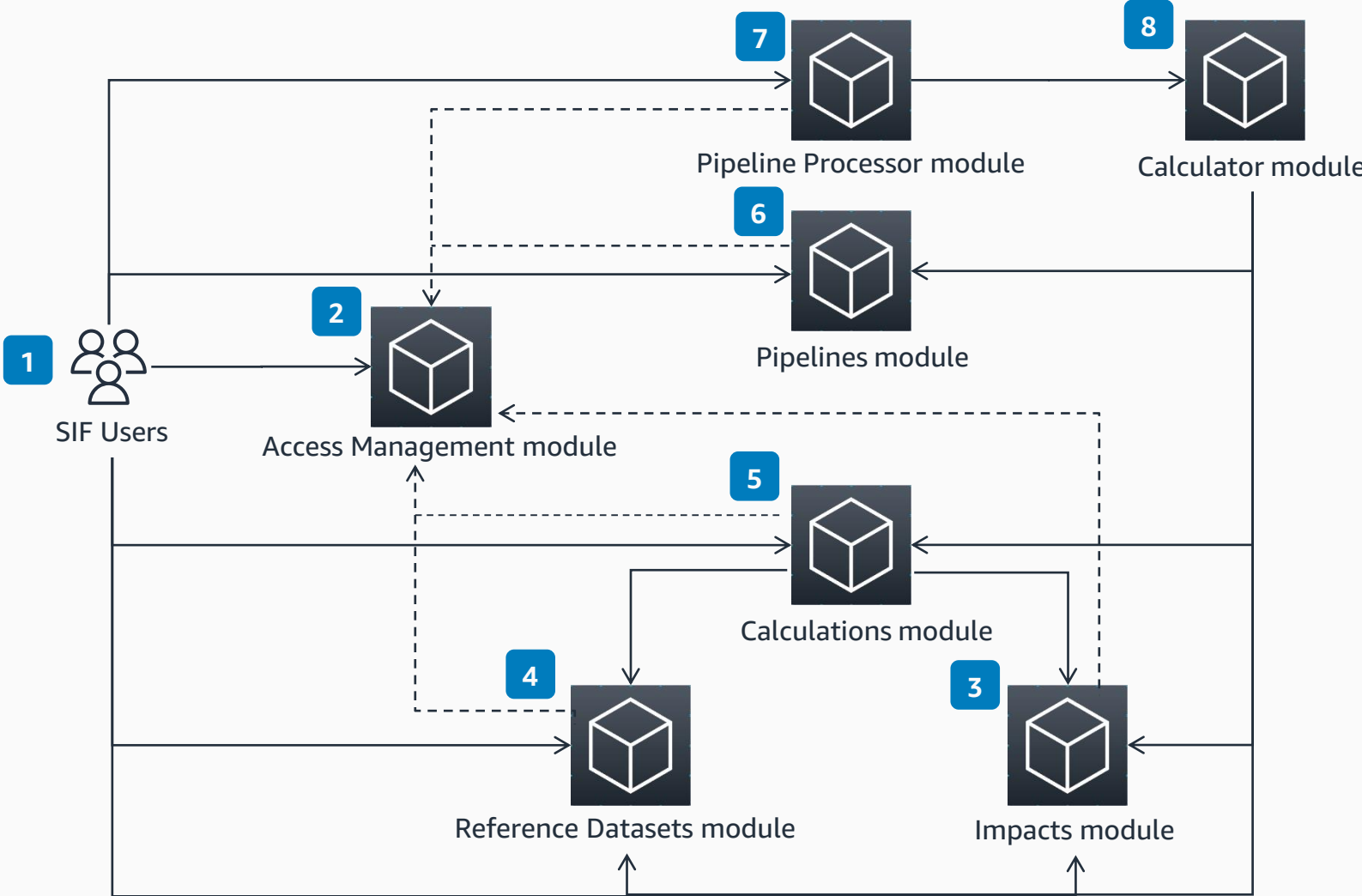


Guidance for Sustainability Insights Framework on AWS

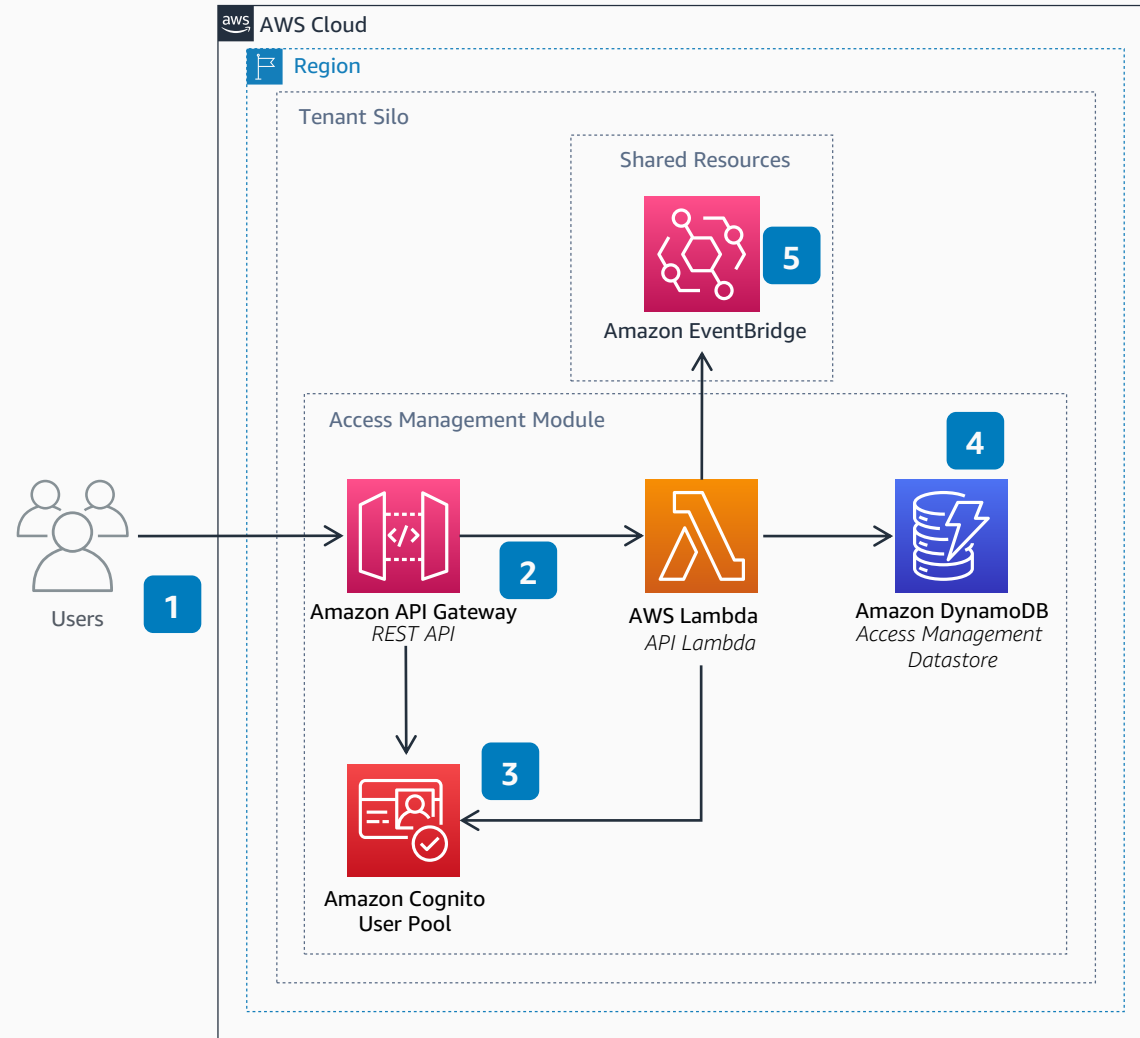
This is a top-level view of how the modules of the Sustainability Insights Framework (SIF) interact. For a detailed view of the services and resources that support each module, see the diagrams that follow.



Overview: The SIF is composed of a suite of modules focusing on a specific set of features. This conceptual architecture shows these modules and their interactions.

- 1** Users interact with SIF through REST APIs.
- 2** The Access Management module is where users and permissions are managed and resources are separated by groups.
- 3** When performing data processing calculations, the Impacts module enables users to manage resources, such as impact factors, that can be referenced from within the Calculations and Pipelines modules.
- 4** The Reference Datasets module enables users to manage datasets, such as lookup tables. These datasets can be referenced from within the Calculations and Pipelines modules.
- 5** The Calculations module enables users to define and manage equations or functions which can be referenced in other modules to perform data processing calculations.
- 6** The Pipelines module enables users to configure data processing pipelines used to perform calculations.
- 7** The Pipeline Processor module is responsible for orchestrating pipelines and performing pipeline aggregations.
- 8** The Calculator module is a backend component that runs operations in a pipeline. This can include arithmetic or the lookup of resources.

Access Management Module

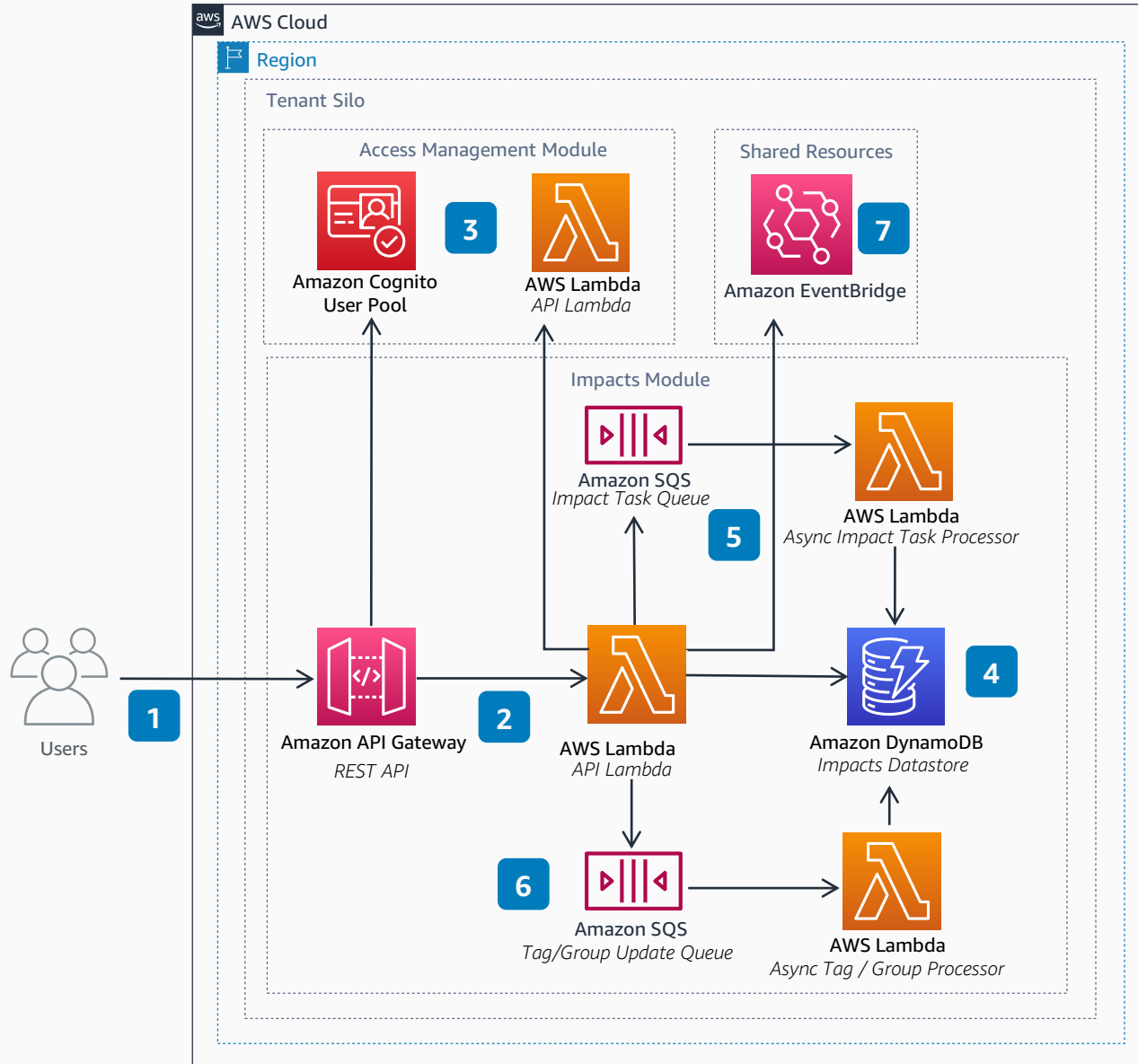


Overview: The Access Management Module uses the concepts of users and groups to allow for permissions management and segregation of resources within SIF. SIF users can define users and groups through an external REST API. Internally, other SIF modules make calls to the Access Management module to validate permissions. Each tenant has their own copy of the Access Management infrastructure.

- 1** Users of SIF interact with the Access Management module through an externally available API.
- 2** The externally available API consists of a REST API in **Amazon API Gateway**. The application logic is deployed in **AWS Lambda**.
- 3** User authentication is done through tokens received from **Amazon Cognito**.
- 4** The Access Management data is stored in an **Amazon DynamoDB** table.
- 5** Access Management resource changes emit events to a message bus in **Amazon EventBridge**. Events can be tracked to update other components of the framework.



Impacts Module

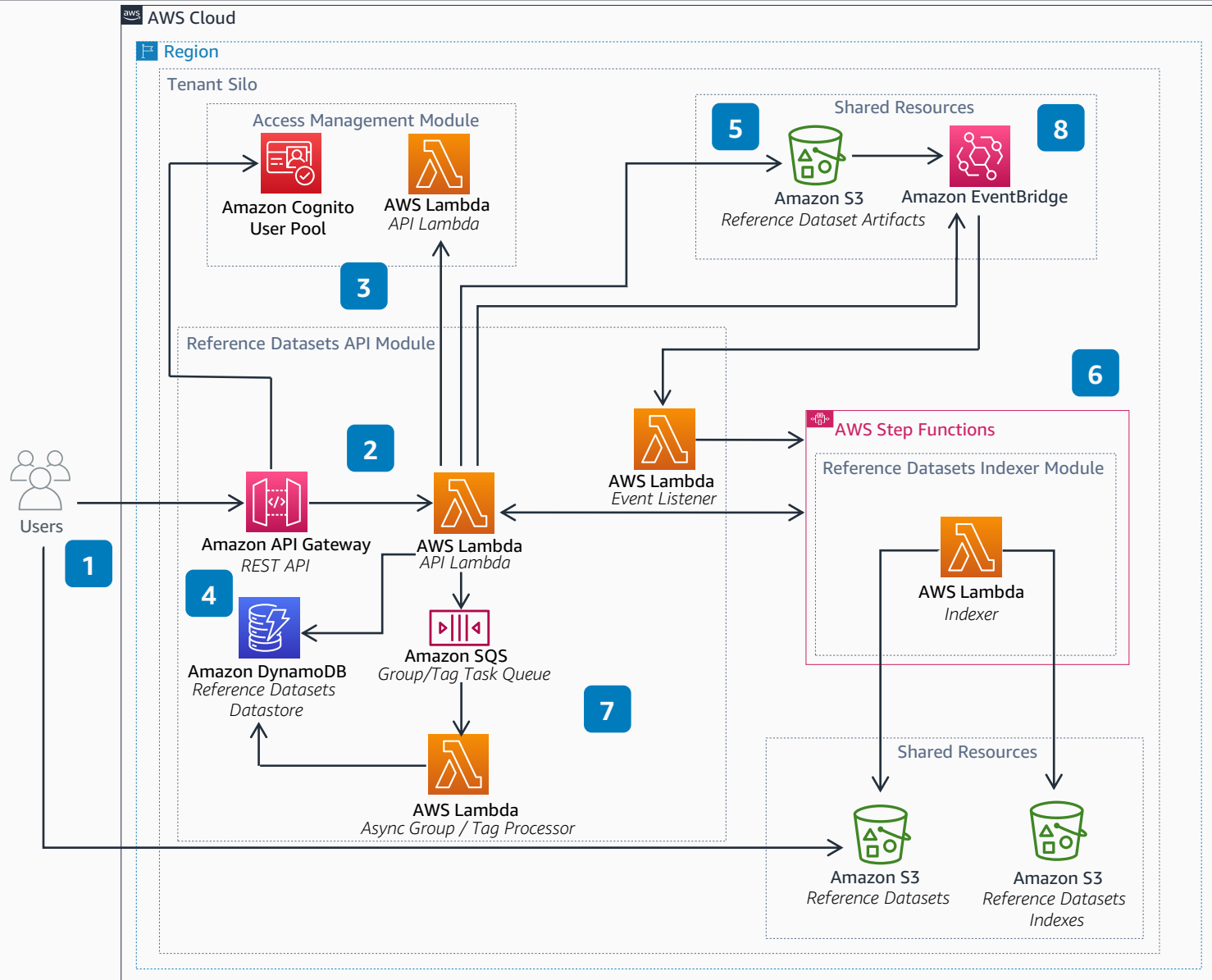


Overview: The Impacts Module enables users to manage impact-related resources. These resources can be referenced from within the Calculations and Pipelines modules when performing data processing calculations, such as emissions. An example Impact could be the carbon dioxide equivalent (CO₂e) of a specific activity, such as mobile diesel fuel consumption. The Impacts module has the capability to create many Impact resources in bulk through an Impact Tasks API. Impacts are versioned to provide traceability.

- 1 Users of SIF interact with the Impacts module through an externally available API.
- 2 The externally available API consists of a REST API in **API Gateway**. The application logic is deployed in **Lambda**.
- 3 User authentication is done through tokens received from **Amazon Cognito**. Authorization is done through the Access Management module.
- 4 Impact data is stored in a **DynamoDB** table.
- 5 **Amazon Simple Queue Service (Amazon SQS)** is used along with a **Lambda Impact Task Processor** to orchestrate bulk Impact creation tasks.
- 6 **Amazon SQS** asynchronously processes metadata updates to resources, such as adding searchable tags.
- 7 Impacts resource changes emit events to a message bus in **EventBridge**. Events can be tracked to update other components of the framework.



Reference Datasets Module

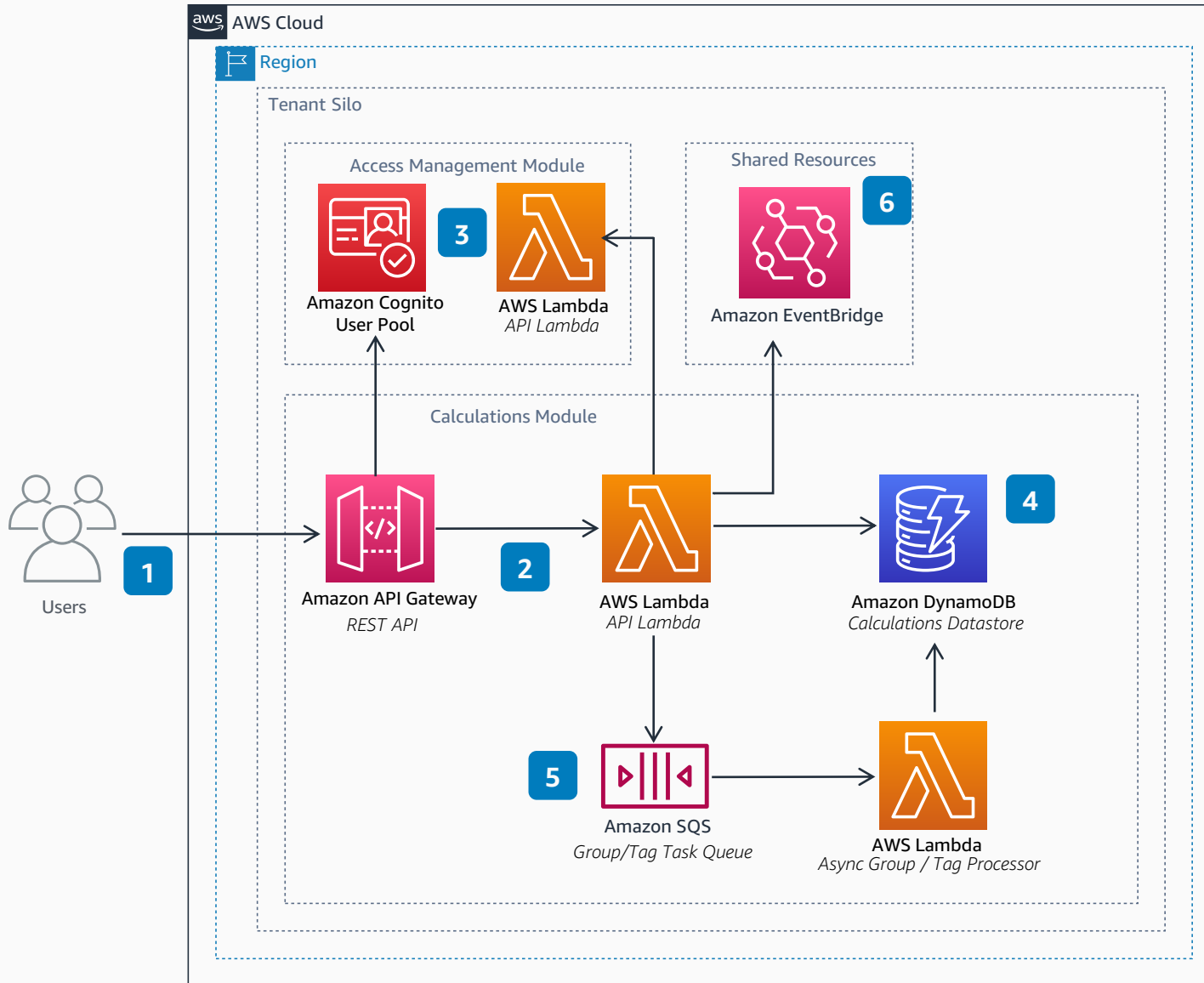


Overview: The Reference Datasets Module enables users to manage datasets, such as lookup tables. These datasets can be referenced from within the Calculations and Pipelines modules when performing data processing calculations, such as emissions. An example Reference Dataset is a table that enables lookup of the mix of electricity generation (such as coal, nuclear, wind) for a particular location. Reference Datasets are versioned to provide traceability.

- Users of SIF interact with the Reference Datasets module through an externally available API and file uploads to **Amazon Simple Storage Service (Amazon S3)** through a pre-signed URL.
- The externally available API consists of a REST API in **API Gateway**. The application logic is deployed in **Lambda**.
- User authentication is done through tokens received from **Amazon Cognito**. Authorization is done through the Access Management module.
- Data (such as dataset names) and metadata (such as tags) are stored in a **DynamoDB** table.
- The dataset is stored in **Amazon S3**.
- The dataset is indexed on Create/Update using **AWS Step Functions**.
- Amazon SQS** asynchronously processes metadata updates to resources, such as adding searchable tags.
- Reference Dataset resource changes emit events to a message bus in **EventBridge**. Events can be tracked to update other components of the framework.



Calculations Module

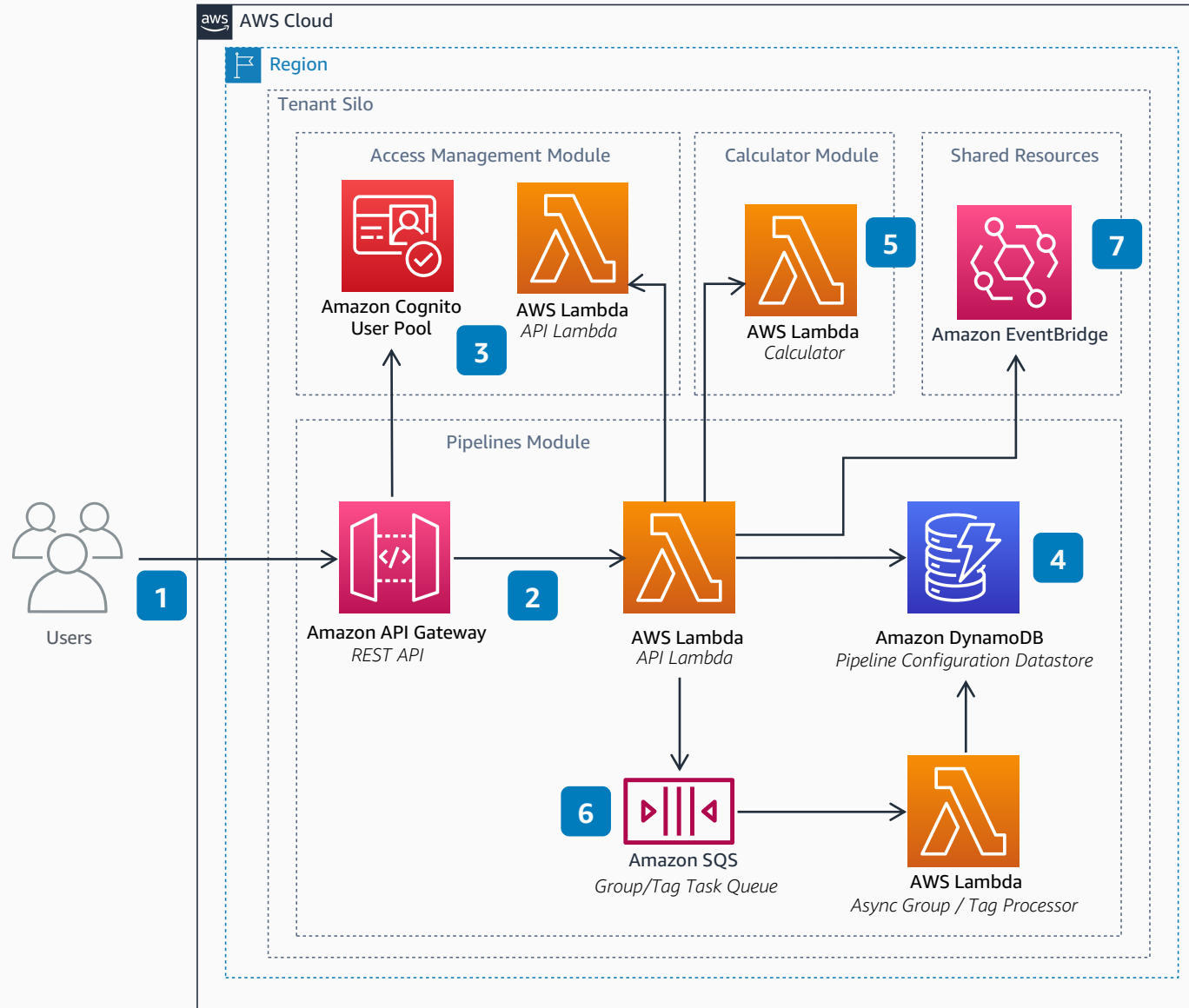


Overview: The Calculations Module enables users to define and manage equations or functions. These equations or functions can then be referenced in other Calculations or Pipelines modules when performing data processing calculations, such as emissions. Example calculations could be simple, such as unit conversions, or complex, such as the agreed upon calculation for emissions by the business. Calculations are versioned to provide traceability.

- 1** Users of SIF interact with the Calculations module through an externally available API.
- 2** The externally available API consists of a REST API in **API Gateway**. The application logic is deployed in **Lambda**.
- 3** User authentication is done through tokens received from **Amazon Cognito**. Authorization is done through the Access Management module.
- 4** Calculations data is stored in a **DynamoDB** table.
- 5** **Amazon SQS** asynchronously processes metadata updates to resources, such as adding searchable tags.
- 6** Calculations resource changes emit events to a message bus in **EventBridge**. Events can be tracked to update other components of the framework.



Pipelines Module

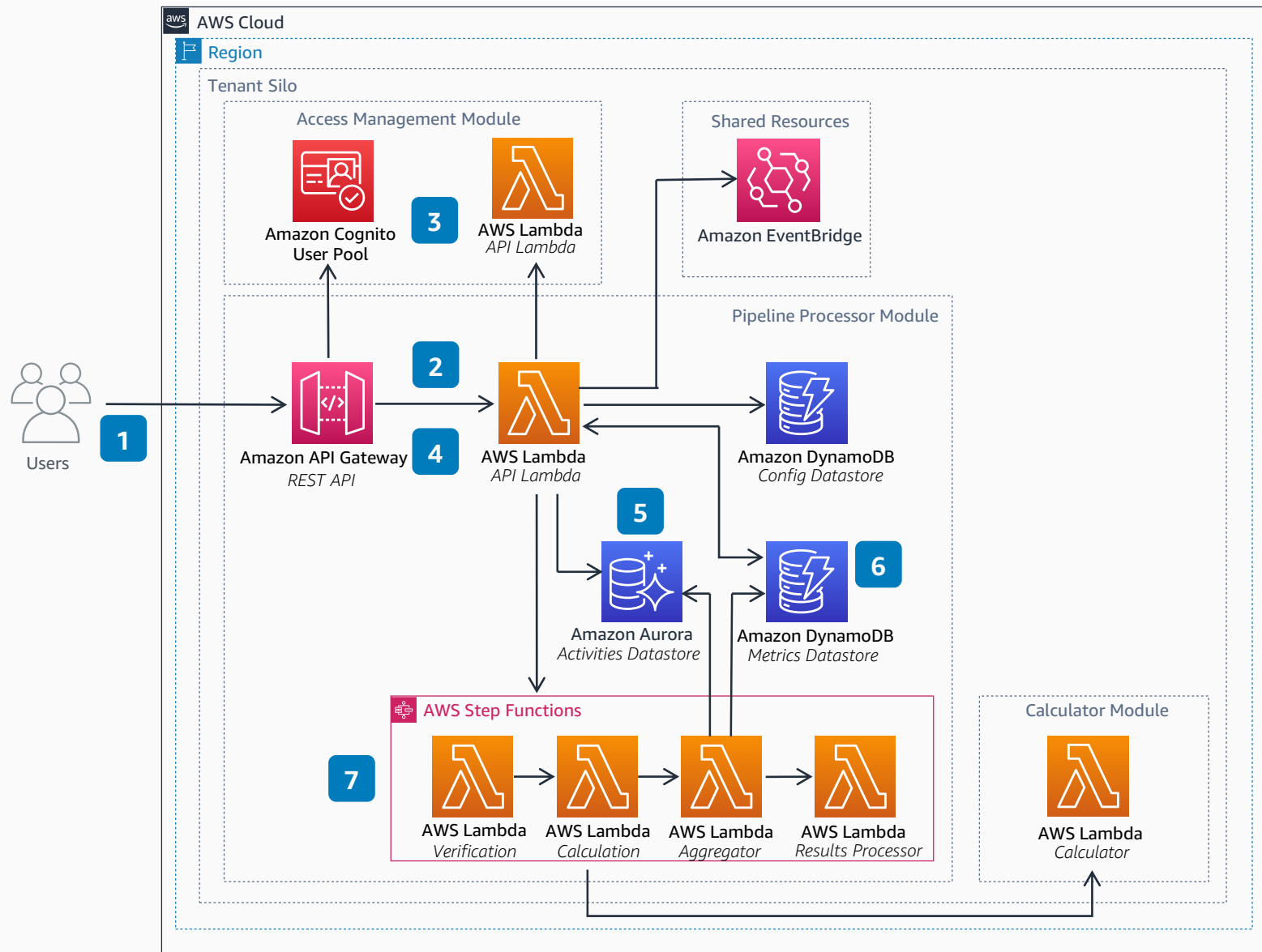


Overview: The Pipelines Module enables users to manage Pipeline configurations. These configurations define data processing pipelines used to perform calculations, such as emissions. A Pipeline can be configured to aggregate outputs across executions and groups into metrics. Metrics capture key performance indicators (KPIs), such as total emissions over time. A user can request a dry run of a Pipeline configuration to have the configuration processed by the Calculator and to check for errors and validate before creation. Pipeline configurations are versioned to provide traceability.

- 1 Users of SIF interact with the Pipelines module through an externally available API.
- 2 The externally available API consists of a REST API in **API Gateway**. The application logic is deployed in **Lambda**.
- 3 User authentication is done through tokens received from **Amazon Cognito**. Authorization is done through the Access Management module.
- 4 Pipeline configuration is stored in a **DynamoDB** table.
- 5 The Pipelines module can directly invoke the Calculator module to dry run a pipeline configuration.
- 6 **Amazon SQS** asynchronously processes metadata updates to resources, such as adding searchable tags.
- 7 Pipelines resource changes emit events to a message bus in **EventBridge**. Events can be tracked to update other components of the framework.



Pipeline Processor Module

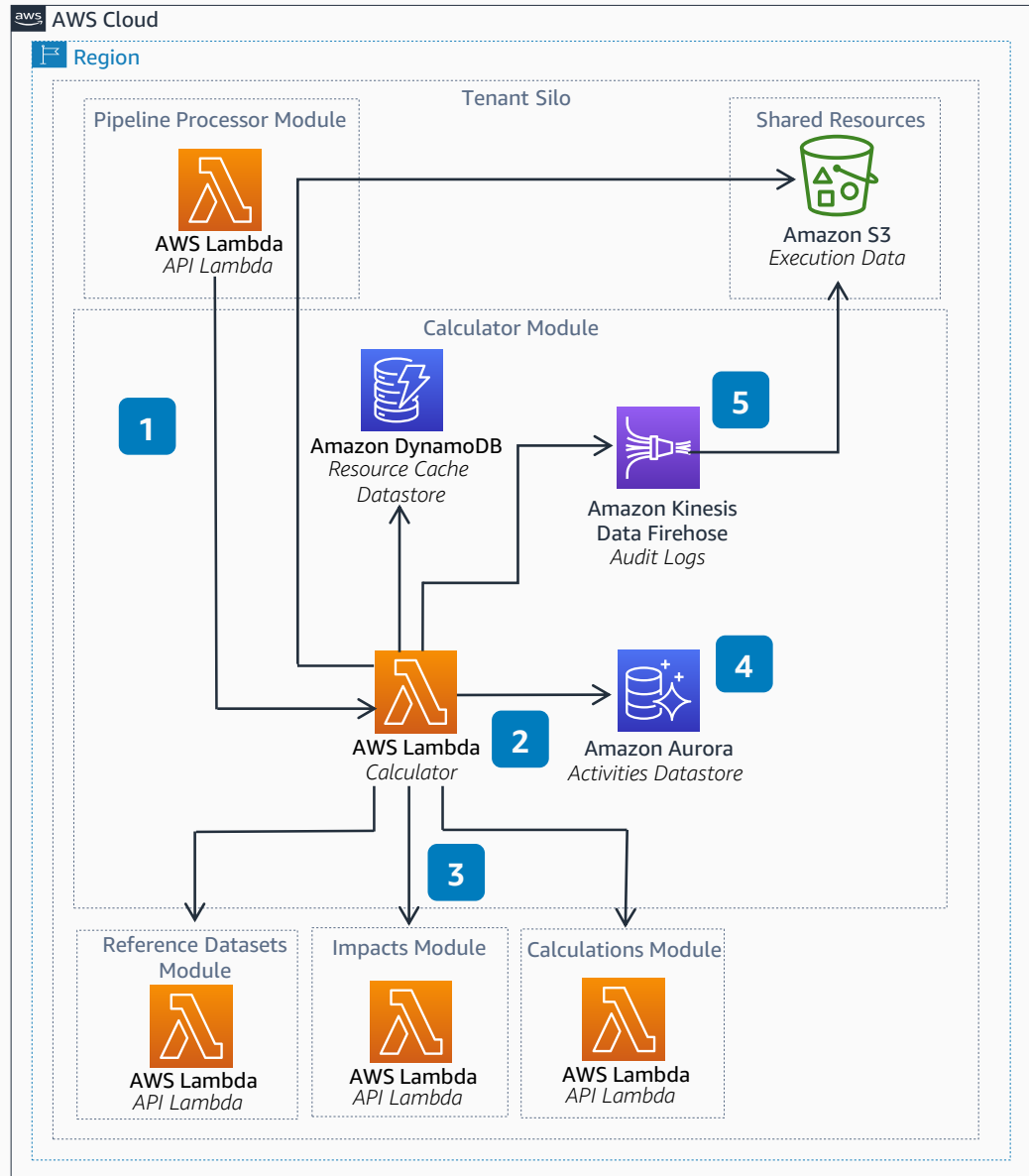


Overview: The Pipeline Processor Module is responsible for the orchestration of Pipelines. This includes starting a pipeline execution in response to input files provided by a user and performing any aggregations defined in the pipeline configuration. The Pipeline Processor module also provides the status of pipeline executions.

- 1 Users of SIF interact with the Pipeline Processor module through an externally available API.
- 2 The externally available API consists of a REST API in **API Gateway**. The application logic is deployed in **Lambda**.
- 3 User authentication is done through tokens received from **Amazon Cognito**. Authorization is done through the Access Management module.
- 4 The REST API allows a user to query the execution status of a pipeline, query for activities processed by a pipeline, and query for metrics aggregated from activities.
- 5 Activity data processed by a pipeline is stored in an **Amazon Aurora** Serverless v2 database.
- 6 Metrics data processed by a pipeline are stored in a **DynamoDB** table.
- 7 Pipelines execution is done through tasks defined in **Step Functions**. This verifies a pipeline and input data, performs calculations by invoking the Calculator, performs aggregations on Calculator outputs, stores aggregations as metrics, and records the status of the execution.



Calculator Module



Overview: The Calculator Module is a backend component which parses and executes the operations defined within a pipeline. This can include arithmetic operations or lookups of resources, such as Reference Datasets and Impacts. The Calculator also captures an audit log of all operations performed in the pipeline, such as input values, and the version of each resource (for example, Reference Datasets, Impacts, Calculations) used in the execution.

- 1** The Calculator module is invoked through a **Step Functions**, defined in the Pipeline Processor module.
- 2** The Calculator uses the Pipeline configuration to execute all of the operations in the configuration.
- 3** These operations may be lookups in Reference Datasets, retrieving Impacts, or retrieving functions defined in the Calculations module. This is done by invoking the **Lambda** APIs for each module. Retrieved resources can be cached to **DynamoDB**.
- 4** Outputs for each activity processed as part of a pipeline are written to the activity data in an **Aurora v2** Serverless database.
- 5** Audit logs are written to an output location in **Amazon S3** through writes to **Amazon Kinesis Data Firehose**.