Organizing Data

Deployment Options

Three Types of data collection

Local Data Collection

1. Collector is installed on **all target hosts**
2. Sends log data produced on those target hosts directly to Sumo Logic Backend via HTTPS connection

Centralized Data Collection

1. Collector is installed on a set of **dedicated machines**
2. Collects log data from target hosts through various remote mechanisms
3. Forwards data to Sumo Logic Backend

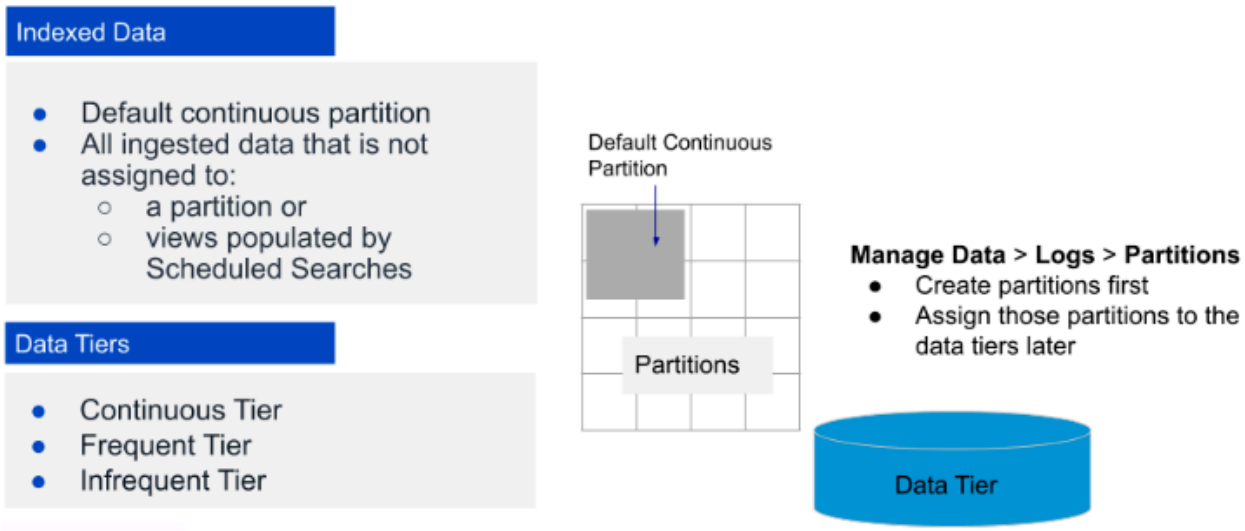
Hosted (Cloud) Data Collection

1. The cloud service generates most data in the cloud

2. Collects data through Sumo Logic cloud integrations

Organizing Data in Sumo Logic

- Sumo Logic provides **partitions** and **data tiers**

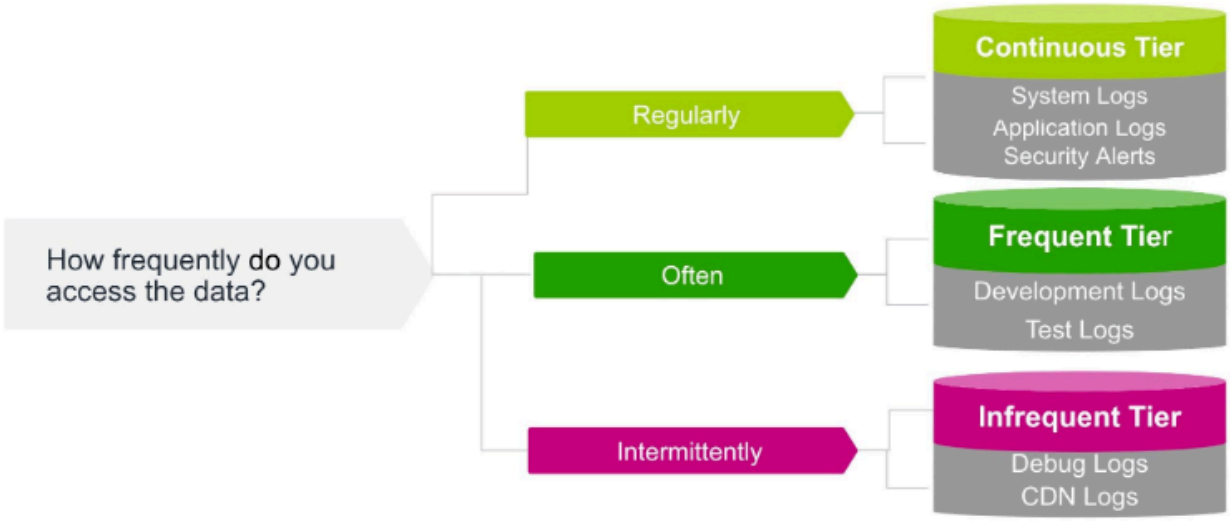


Partition (How data is organized)

- The Default Continuous Partition Contains all ingested data that is not assigned to a partition or to views populated by Scheduled Searches
- Default retention period is 30 days.
- Partitions allows you to **improve search performance** by searching over a smaller number of messages.
- **Manage Data** > **Logs** > **Partitions**
  - Need Admin role
- Sumo Logic enables you to slice and dice usage based on metadata, types of data, and tiering solutions
  - Need to identify the metadata
    - Select the type of collector
    - Add source, source name

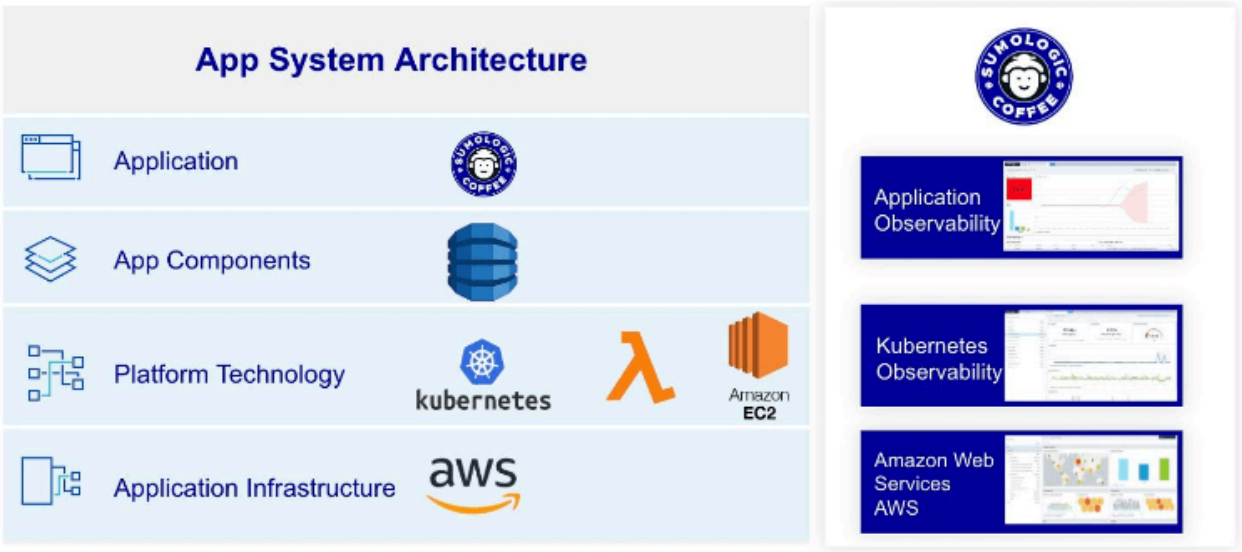
        - Define source category, and the source host.
    - Identify which logs, metrics or traces you need to ingest
    - Set that data type into the **proper data tier** or **partition**

Which Tier to use?



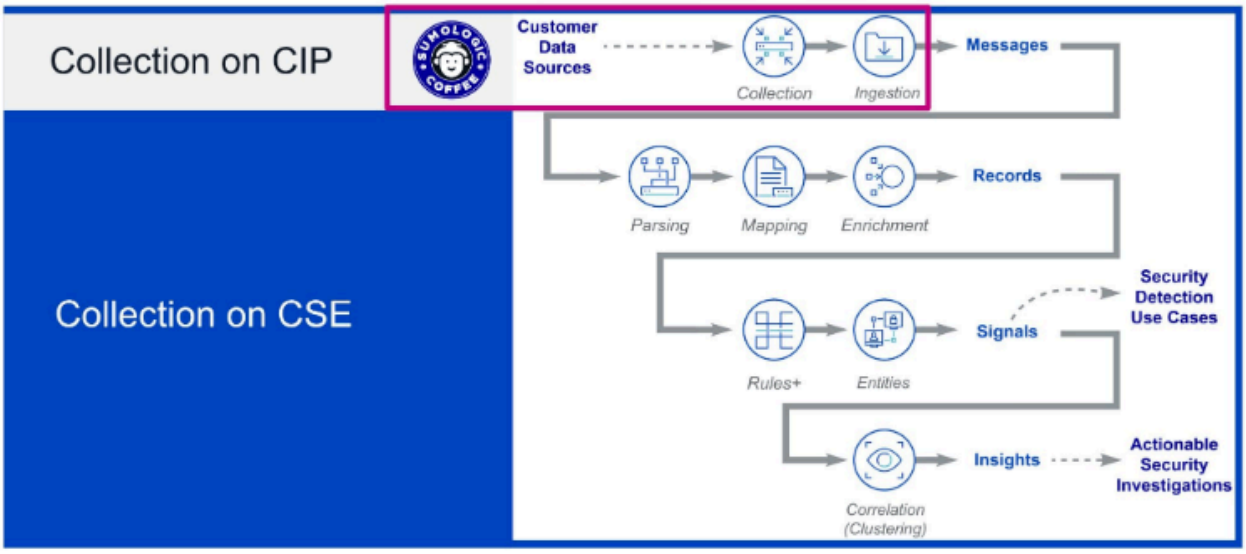
- Depends how **frequently** do you want to **access** the data
- **Continuous Tier**
    - Data you use to
        - Monitor and troubleshoot production applications
        - To ensure the security of your applications
- **Frequent** or **Infrequent**
    - Depends on how frequently you need to access the data
    - ex 1) For a large development team with hundreds of developers, it is better to send development and test logs to the Frequent Tier
    - ex 2) Debug or other verbose log sources are used to troubleshoot very specific issues that occur infrequently.

Ingesting Data for Observability



- **Observability**
  - Being able to ask any question you need about your application and how customers are interacting with it.
  - Sumo Logic offers Sources to collect from many AWS products.
    - Amazon CloudFront Source
    - Amazon CloudWatch Source for Metrics
    - Amazon Path Expressions
    - Amazon S3 Audit Source
    - AWS CloudTrail Source
    - AWS Elastic Load Balancing Source
    - AWS Metadata (Tag) Source
    - AWS S3 Source

Ingesting Data for Security



- Cloud SIEM takes millions of log messages and funnels them down into a handful of actionable security insights
  - First, logs are collected and ingested through environment on the CIP into messages.
  - They are *parsed, mapped, and enriched* into Cloud SIEM records.
  - These records are **compared to rules**.
    - If a rule is triggered, an entity is extracted, a severity score is assigned, and a signal is created.
    - If enough signals with the same entity cluster together, they become an **insight**.