Log Aggregation Explained

Log aggregation is the process of collecting logs from multiple computing systems, parsing them, extracting structured data, and putting them together in a format that is easily searchable and explorable by modern data tools.

There are four common ways to aggregate logs, and many log aggregation systems combine multiple methods. These include:

Syslog

- A standard logging protocol. Network administrators can set up a Syslog server that receives logs from multiple systems, storing them in an efficient, condensed format which is easily queryable. Log aggregators can directly read and process Syslog data.

Event Streaming

- Protocols like SNMP, Netflow and IPFIX allow network devices to provide standard information about their operations, which can be intercepted by the log aggregator, parsed and added to central log storage.

Log Collectors

- Software agents that run on network devices, capture log information, parse it and send it to a centralized aggregator component for storage and analysis.

Direct Access

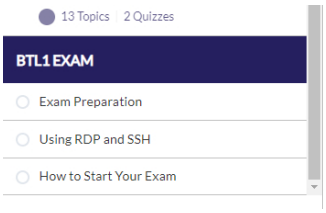
- Log aggregators can directly access network devices or computing systems, using an API or network protocol to directly receive logs. This approach requires custom integration for each data source.

DATA TYPES

When considering the data that is being pulled into a SIEM platform, there are two categories; Structured data, and Unstructured data.

- **Structured data:** These are usually logs for Apache, IIS, Windows events, Cisco logs, and some other manufacturers. They have clearly-defined fields (such as "`src_ip`") and are similar to other structured logs, making them relatively easy to parse and normalize.
- **Unstructured data:** This type of logging typically comes from a custom-built application where each message can be printed differently in different operations and the event itself can span multiple lines with no defined event start point, or event end point, or both. This is likely to be the majority of the data being sent to the SIEM.

In order to get all logs to follow a similar format to make it easier to perform searches across a large set of different logs, where possible, we can use normalization techniques, which we will cover in the next section of this domain.



Previous Topic
Mark Complete
Back to Lesson