

8 Burning Questions About Log Management



1 How does SAFE collect logs?

SAFE can collect logs from virtually all IT infrastructure elements.

Log sources that push their logs out to network addresses are captured directly without deployment of any agents outside of SAFE. The push protocols include: syslog, both UDP and TCP, and snmp, both v1 and v2. Logs received from syslog forwarders or syslog concentrators are also supported and they are automatically processed as if they were received directly.

Log sources that maintain their own repository and allow the logs to be pulled are supported through polling technologies that are part of the SAFE product. SAFE does not require deploying agents outside of SAFE. The pull protocols include: ODBC for database releases [Oracle, Sybase, MS-SQL Server, MySQL]; IBM's UDB; OPSEC/LEA for Checkpoint firewalls; SDEE for Cisco's IDS/IPS; and eStreamer for Sourcefire product suite.

Finally, there are products that log their event to files that can then be sent to SAFE for processing. Typically this is done when the log file rotates so that duplicate events are not sent to SAFE. Any script on any hosts can be used to securely copy the log files over to SAFE.

2 Does SAFE require agents on all log sources?

No, SAFE is primarily agent-less. This means that organization do not have to install agents on their IT infrastructure components to collect logs which are then forwarded to SAFE. The benefits are clearly a lower total cost of ownership and minimal disruption to the managed environment.

All SAFE appliances have a huge capacity for acquiring pushed logs directly. All SAFE appliances host the custom pulling protocols described above and can pull from any source that is visible on the network.

Some organizations choose to implement the technology that supports the pull protocols as an "agent". This is appropriate in situations where the infrastructure is distributed. Again, this is an option; the pull technology is part of the one appliance required to implement SAFE.

3 Does SAFE delete the logs from the source upon collection?

No, SAFE does not delete logs upon collection.

First, log sources that push logs out generally do not keep a copy of those logs. Some organizations choose to fork the pushed logs to alternate storage. Regardless, SAFE can accept the logs directly or via a forwarder.

SAFE does not alter the source in cases where SAFE is receiving a file from a repository or pulling logs from a repository. When pulling from a repository SAFE tracks the last read event and only pulls events that came in since that last read event. Most sources that maintain data in these ways are configured to rotate through and re-use the log storage space based on time or size interval. The suggested configuration for the rotation would allow SAFE enough time to pull all logs; ideally there would be consideration for any network outage that might occur between the repository and the SAFE appliance.

4 What processing does SAFE do on acquired logs?

SAFE processes acquired logs in multiple steps and the results of each step are stored in the embedded Security Data Warehouse (SDW) for different purposes. There is a technical brief explaining the advantages of the SDW in the resource center of the Intellitactics website. This brief also includes more detail on storage.

These are the steps in the process:

Step 1: SAFE compresses the unprocessed logs, calculates an MD5 hash value and stores the logs in the SDW raw store. The files are compressed and stored in form in which they were received by SAFE; the streamed/pushed logs (syslog and snmp) are written to files in one-minute chunks. The MD5 hash value is stored elsewhere in the SDW to support non-repudiation claims if the raw log is used for forensic analysis.

Step 2: SAFE parses and normalizes the event data in the log, then generates a tab-delimited parsed file and stores that in the SDW parsed store. A parsed file is created for each minute of log data based on the timestamp of when the event is generated; it is stored in a compressed format. The SDW maintains a smart index into the content of each file for efficient retrieval. SAFE is differentiated on its speed of access or retrieval. Only SAFE provides continuous parsing using patented technology.

Step 3: In this step the SDW takes over, regularly summarizing the parsed events based on hourly, daily, and monthly time boundaries. These summaries provide for a consistently fast response time for generating the 1400 charts and reports that are pre-defined with SAFE.

Every few minutes the SDW evaluates the data against the notifications configured in SAFE and sends snmp or email alerts as defined by the customers' configuration.

Once a day the SDW evaluates the expiry configuration and deletes all log data in the SDW raw store and the SDW parsed store that is older than the configured retention policy. The date used for determining expiry is the date that the event was generated; not when the log was acquired.

5 | When does SAFE delete or expiry old log data?

The log retention policy is customer configurable on a per log source basis using separate values for days, for the SDW raw store and the SDW parsed store. The special retention period of 0 days implies no expiry of logs and the logs are never removed from the embedded SDW.

Once a day the SDW evaluates the expiry configuration and deletes all log data in the SDW raw store and the SDW parsed store that is older than the time specified in the retention policy. The date used to determine expiry is the date that the event was generated; not when it was acquired.

Alternatively, If longer term storage is required it is suggested that these logs be backed up or archived before the expiry period specified in the configuration. (See question below on log archival and back-up).

6 | How does SAFE act when the embedded storage fills up?

SAFE appliances can be purchased in 5 different sizes that vary based on the amount of embedded storage (1.5TB to 8TB of raw storage). Customers discuss their retention policy and review the devices they are monitoring with their sales engineer and are typically able to select a right sized appliance that obviate a situation where embedded storage might fill up.

The general guidance for determining the required storage is to assume a 75% compression of the raw, uncompressed log data being sent to SAFE every day. For example, if customers are sending 10G of MS-Windows log data that they wish to retain for 365 days and 40G of router log data that they wish to retain for 30 days then their minimum storage requirement can be calculated as: $365 * 10G * 0.25 + 30 * 40G * 0.25 = 1212.5GB$ or 1.3TB of embedded SDW storage.

In the event that the embedded storage fills up, the built-in systems monitor and alarm manager in SAFE alerts the administrator if free disk space is less than a specified threshold. Corrective action is then taken and assistance is available from support@intellitactics.com

7 How can the acquired log data be backed up?

The daily volume of log data is so large that traditional applications for back-up solutions are not effective. Many organizations address this by deploying a redundant SAFE appliance. The redundant appliance is often placed in a different physical data center to provide for business continuity. The redundant appliance can be fed directly from the log sources or the primary SAFE appliance can be configured to forward all acquired logs to the redundant appliance.

The SDW stores are accessible via SCP and the stores have a straight-forward directory structure based on data; any commercial solution or home-grown script can be pointed to the SDW. This enables the use of an external archival or back-up solution. When backing up, we suggest backing up nightly and for archiving we suggest that the archival solution be configured to match the expiration policy. Today, the archival and back-up implementations for SAFE are automated using custom scripts that write to SAN or WORM repositories.

Customers that archive or back-up the log data externally generally choose between the SDW raw format and the SDW parsed format based on their use case. If the use case is forensic then the raw format is preferred since it is unprocessed and has an MD5 hash value. If the use case is for safety of review and research then the parsed format is preferred since it protects the investment of processing. The parsed format also contains a raw event as one of the keys. More details are available from support@intellitactics.com

Committed to Your Success

Intellitactics features a low total cost of ownership. Primarily agentless data collection reduces the burden on the infrastructure. The unique Security Data Warehouse, which combines an embedded, self-managing relational database requiring no DBA and compressed stores of raw logs, is easy on the storage budget. The product architecture grows with you as you add more data sources or evolve your risk policies. Intellitactics features 'security know-how' in packaged reports, metrics and correlations. The Customer Center, located at www.intellitactics.com, features instant access to new reports and metrics, and automates support functions for faster response to all inquiries.



intellitactics.com

1800 Alexander Bell Drive
Reston, Virginia 20191
703 620 3800 | 877 746 7658

Copyright © 2008 Intellitactics, Inc. All rights reserved.
All trademarks are the property of their respective owners.

8 How can the archived or backed up log data be reloaded?

A file that has been archived or backed up from the SDW raw store needs to be re-processed by the SAFE appliance. The filename contains all the necessary encoding for SAFE to process the log appropriately when it is copied into a special inbox, i.e. file directory, on the SAFE appliance. The time to reprocess a file, from a SAFE capacity perspective, is the same as when it was initially acquired.

A file archived or backed up from the SDW parsed store is simply re-loaded into the embedded SDW of the SAFE appliance. The filename contains all the necessary encoding for SAFE to reload the parsed logs appropriately when the file is copied into a special inbox, i.e. file directory, on the SAFE appliance. The time to reload a file, from a SAFE capacity perspective, is approximately 20% of the time it takes to reprocess the raw version of the same logs.

Reprocessing or reloading is generally a manual task which is done on demand. Old data should only be reloaded if it is not present on the SAFE appliance anymore; either because of catastrophic failure of the appliance or expiration of the originally acquired data. For this reason, some customers deploy the smallest SAFE appliance for non-production use in their lab and use it expressly for restoring and searching on old data. More details are available from support@intellitactics.com