

Sawmill for IronPort

VERSION 7.3.2

USER GUIDE

for Web Security Appliances



COPYRIGHT

Copyright © 2010 by IronPort Systems®, Inc. All rights reserved.

Part Number: 421-0531(B)

Revision Date: March 15, 2010

The IronPort logo, IronPort Systems, SenderBase, and AsyncOS are all trademarks or registered trademarks of IronPort Systems, Inc. All other trademarks, service marks, trade names, or company names referenced herein are used for identification only and are the property of their respective owners. Portions of this work are copyright © Flowerfire, Inc.

This publication and the information contained herein is furnished “AS IS” and is subject to change without notice. Publication of this document should not be construed as a commitment by IronPort Systems, Inc. IronPort Systems, Inc., assumes no responsibility or liability for any errors or inaccuracies, makes no warranty of any kind with respect to this publication, and expressly disclaims any and all warranties of merchantability, fitness for particular purposes and non-infringement of third-party rights.

Some software included within IronPort AsyncOS is distributed under the terms, notices, and conditions of software license agreements of FreeBSD, Inc., Stichting Mathematisch Centrum, Corporation for National Research Initiatives, Inc., and other third party contributors, and all such terms and conditions are incorporated in IronPort license agreements.

The full text of these agreements can be found at https://support.ironport.com/3rdparty/AsyncOS_User_Guide-1-1.html. Portions of the software within IronPort AsyncOS is based upon the RRDtool with the express written consent of Tobi Oetiker.

IRONPORT SYSTEMS®, INC. CONTACTING IRONPORT CUSTOMER SUPPORT

IronPort Systems, Inc.
950 Elm Ave.
San Bruno, CA 94066

If you have purchased support directly from IronPort Systems, you can request our support by phone, email or online 24 hours a day, 7 days a week. During our office hours (24 hours per day, Monday through Friday excluding US holidays), one of our engineers will contact you within an hour of your request. To report a critical issue that requires urgent assistance outside of our office hours, please call us immediately at the numbers below.

U.S. Toll-free: 1 (877) 641-IRON (4766)

International: [www.ironport.com/support/
contact_support.html](http://www.ironport.com/support/contact_support.html)

Support Portal: www.ironport.com/support

If you have purchased support through a reseller or another entity, please contact them for support of your IronPort products.

Table of Contents

Preface	vii
Audience	vii
Sawmill Information	vii
Command and Keyboard Conventions	vii
Formatting conventions	viii
Contacting IronPort Customer Support	viii
IronPort Welcomes Your Comments	viii
 1. Introduction to Sawmill for IronPort	 1
Introduction	2
Profiles	2
Reports	3
IronPort Log Format Plug-In	4
Sawmill for IronPort Reports	4
Sawmill for IronPort Log Filters	7
Deployment Planning	9
Network Planning	9
Choosing an SCP or FTP Server	11
Sawmill Architecture Overview	12
Log Importer	12
Sawmill Database	12
Reporting Interface	12
Administrative Interface	13
Web Server	13
Command Line Interface	13
What's New in Sawmill for IronPort	14
What's New in Sawmill for IronPort Version 7.3.2	14
What's New in Sawmill for IronPort Version 7.3.1	14
What's New in Sawmill for IronPort Version 7.3.0	14
Fixed Issues, Known Issues, and Limitations	15

Fixed: Inaccurate Page View Statistics	15
Fixed: Log Entries Containing Spaces in Usernames Do Not Appear in Reports	15
Fixed: HR Profile Erroneously Displays a “FIELD: Auth User” Report	15
Fixed: Scheduling Database Updates in the Sawmill CLI Can Cause Sawmill for IronPort to Crash	
15	
Fixed: Importing Access Logs Can Cause Sawmill for IronPort to Crash	15
Fixed: Updating the Database in Sawmill for IronPort May Throw an Application Error at the End of	
the Update	15
Fixed: Sawmill for IronPort Discards Session that are Over Two Hours Long	15
Fixed: Sawmill for IronPort Does Not Allow Users to Enable WBRS in the Summarized Logs Report	
16	
Fixed: Sawmill for IronPort Incorrectly Reports Page View Calculations Due to MIME Type Entries	
16	
Multiple Log Formats in Same Directory Not Supported.	16
“Date Offset” Supports Whole Hours Only	16
Generating “All Reports” Does Not Always Work if the Database Is Large	16
Zoomed Reports Show Zero Results When Using Global Filters in Some Cases	16
Custom URL Category Number Erroneously Appears as Access Policy Name	16
When Running on a 32-Bit Operating System, Sawmill for IronPort May Have Difficulty Handling	
Access Logs Larger than 2G	17
Typo in the Daylight Savings Time Log Filter in the Online Help	17
Disabling the Rewrite URL Feature Causes Sawmill for IronPort to Stop Functioning Properly . .	17
The Default Pathname Suggested in the New Profile Wizard is Problematic	17
Sawmill for IronPort Incorrectly Reports CONNECT Log Entries	17
2. Installation and Configuration	19
Overview	20
Installing and Configuring Sawmill for IronPort	21
Step 1. Install Sawmill for IronPort	21
Step 2. Configure Sawmill for IronPort Setup	21
Step 3. Log In and Create a Profile	21
Working with Time Zones	24
Accessing Windows Mapped Drives	25
3. Navigating Sawmill for IronPort	27
Sawmill Web Interface Overview	28
Reports Page	29
Reports Header	29
Reports Toolbar	30
Reports Menu	30
The Report	30
Config Page	33
Profile Summary	33
Log Data	33

Database	34
DNS Lookup	34
Manage Reports	34
Rebuild Database Button	34
Update Database Button	34
Administrative Page	35
Profiles	35
Licensing	35
Users	35
Scheduler	35
Preferences	36

4. Reading Access Logs with Sawmill for IronPort. 37

Working with Profiles	38
Creating a Profile	38
Working with Reports	40
Creating Reports	40
Saving Reports	41
Printing Reports	41
Creating a PDF of a Report	41
Working with Report Filters	43
Global Filters	44
Date/Time Filters	46
Zoom Filters	49
Using Log Filters	51
How Log Filters Work	51
Hits vs. Page Views	52
The Log Filter Editor	52
Working with Users	54
Scheduling Sawmill Tasks	55
Using the Scheduler	55
Using Extra Options	56
Debugging the Scheduler	57
Removing Database Data	57

5. Guidelines and Tips 59

Optimizing Sawmill for IronPort	60
Database Tuning Options	60
Log Processing Options	64
Working with Multiple CPUs	64
Removing Data Before Processing	66
Filtering Out Old Data	67
Creating a New Filter	67

Scheduling Guidelines	68
AntiVirus Scanning	69
Choosing Hardware Resources	70
Rolling Over Access Logs	71
Transferring Access Logs	72
Working with Multiple Web Security Appliances	73
Customizing URL Category Classifications in Compliance Reports	74
Custom URL Categories	77
Date Reports and Individual Fields	78
Logging into the Sawmill Web Interface	79
Showing Full URLs	80
Viewing Particular Detailed Information	81
Website Activity on a Per Client Basis, Highlighting Top Websites	81
URL Category Details by Website	81
List All Users for a Particular Website or URL Category	81
List All Websites Visited by Anyone	81
View Client Malware Risk Detailed by Malware	81
View Detailed Information About Which Script or File was Blocked	81
View Which Content Was Blocked	82
Backup and Recovery	83

Preface

The Sawmill for IronPort *User Guide* provides instructions for setting up and administering Sawmill for IronPort.

AUDIENCE

This guide is designed for experienced system administrators who are familiar with computer networks, logging, Web technology, and HTTP services.

SAWMILL INFORMATION

This document was created in reference to the following software versions:

Sawmill for IronPort version 7.3.2

Note — Sawmill for IronPort version 7.3.2 is based on FlowerFire Sawmill version 7.2.18 Enterprise.

In addition, some of the content used in this document was provided by Flowerfire, Inc. at <http://www.sawmill.net>.

COMMAND AND KEYBOARD CONVENTIONS

When describing key combinations, this guide uses the plus sign (+) to separate individual keys. For example, “Ctrl+D” means pressing the “Control” and “D” keys simultaneously.

This guide uses the term “type” to mean pressing one or more keys on the keyboard. It uses the term “enter” to mean pressing one or more keys and then Enter. Also, this guide uses the term “Enter” to refer to the key that generates a carriage return, although the key is named “Return” on some keyboards.

This guide uses capitalization and some abbreviations to refer to the keys on the keyboard. (The keys on your keyboard might not be labeled exactly as they are in this guide.)

FORMATTING CONVENTIONS

The following table describes typographical conventions used in this guide.

Formatting Convention	Type of Information
<i>Italic type</i>	Words or characters that require special attention. Placeholders for information you must supply. For example, if the guide says to enter the <code>arp -d hostname</code> command, you enter the characters “arp -d” followed by the actual name of the host. Book titles in cross-references.
<code>Monospaced font</code>	Command and process names. Information displayed on the system console or other computer monitors. The contents of files.
Bold monospaced font	Words or characters you type. What you type is always shown in lowercase letters, unless it must be typed in uppercase letters to work properly.

CONTACTING IRONPORT CUSTOMER SUPPORT

You can request our support by phone, email or online 24 hours a day, 7 days a week.

During our office hours (24 hours per day, Monday through Friday excluding US holidays), one of our engineers will contact you within an hour of your request.

To report a critical issue that requires urgent assistance outside of our office hours, please call us immediately at the numbers below.

U.S. Toll-free: 1 (877) 641-IRON (4766)

International: www.ironport.com/support/contact_support.html

Support Portal: www.ironport.com/support

IRONPORT WELCOMES YOUR COMMENTS

We are interested in improving our documentation and welcome your comments and suggestions. You can email your comments to us at:

docfeedback@ironport.com.

Please include the document’s part number (421-0531(B)) of your document in the subject of your email.

Introduction to Sawmill for IronPort

This chapter contains the following information:

- “Introduction” on page 2
- “IronPort Log Format Plug-In” on page 4
- “Deployment Planning” on page 9
- “Sawmill Architecture Overview” on page 12
- “What’s New in Sawmill for IronPort” on page 14
- “Fixed Issues, Known Issues, and Limitations” on page 15

INTRODUCTION

Welcome to Sawmill for IronPort, IronPort's centralized reporting and tracking solution for the IronPort Web Security appliance. You can use Sawmill for IronPort for:

- Centralized reporting
- Centralized end user tracking
- Detailed end users reporting

Sawmill for IronPort includes an IronPort log format plug-in that processes Web Security appliance access logs to help you understand what is going on in your network. The IronPort log format plug-in allows you to create multiple types of profiles.

When Sawmill for IronPort processes Web Security appliance access logs, it uses a profile you create to perform the following steps:

1. Reads access logs from the location specified.
2. Parses the data according to the IronPort log format plug-in included in Sawmill for IronPort and populates the Sawmill database with the parsed data based on the profile type you create.
3. Analyzes the data in the database and generates reports.

You must create at least one profile for Sawmill to read and parse log data. The profile you create uses the IronPort log format plug-in included in Sawmill for IronPort to know how to parse the data. The profile type you choose when you create the profile determines the data that gets loaded into the database and the reports that are generated. For more information about the plug-ins, see “IronPort Log Format Plug-In” on page 4.

First, you must choose how to deploy Sawmill for IronPort in your network. In particular, you must choose how to transfer the access logs from the Web Security appliances to a location where Sawmill can access them. For more information, see “Deployment Planning” on page 9.

Note — This document is aimed at users who install, configure, and use Sawmill for IronPort. This is not intended to be a substitution for the documentation provided by Flowerfire, makers of the Sawmill product. For more information, consult your IronPort Systems Engineer or visit <http://www.sawmill.net>.

Profiles

A Sawmill profile is a collection of options that defines a view into the data that Sawmill collects and analyzes. The license you purchase determines how many profiles you can create. You must create at least one profile to analyze each type of data.

You might *want* to create multiple profiles for any of the following reasons:

- Different departments in your organization want different customized reports created from the same log data.

- The administrator for each department in an organization needs to view data for his/her department only.
- You have multiple Web Security appliances to analyze and you want to report on each one individually.

You might *need* to create multiple profiles for any of the following reasons:

- You need to report on date ranges and have reports for each range, such as 2007, 2008, or Q1 2007, Q2 2007, etc.

When you create a profile, you must specify the log source. The log data can come from one or more Web Security appliance access logs. The files can be local, or Sawmill for IronPort can download them from an FTP server. Sawmill for IronPort automatically detects the IronPort log format when it reads Web Security appliance access logs. After the profile is created, you can view the reports defined by IronPort on the Reports screen.

For more information on profiles, see “Working with Profiles” on page 38.

Reports

Sawmill reports present Web Security appliance access log file information in an attractive and easily navigable format. You can “drill down” on links in the reports to zoom in on particular subsets of data in the report.

For a list of the reports included in Sawmill for IronPort, see “Sawmill for IronPort Reports” on page 4.

IRONPORT LOG FORMAT PLUG-IN

Sawmill for IronPort includes configuration information, known as a log format plug-in, that allows Sawmill to work with Web Security appliance access logs. The IronPort log format plug-in defines the following:

- **How to recognize the access log format.** The plug-in has the access log format defined so it can recognize valid access logs to parse.
- **Which log filters to use and apply to the access logs.** The plug-in defines log filters to filter some data that does not provide useful information and takes up too much space in the database. For example, it filters out all rows with server responses in the 500 range which are server errors. For more information on the log filter, see “Sawmill for IronPort Log Filters” on page 7.
- **How to parse access logs.** The plug-in parses the logs and applies the log filters to the data before instructing Sawmill to load the data into the database.
- **How to convert abbreviated URL categories to their full category names.** Access logs abbreviate all URL category names which can be difficult for humans to read and understand. The plug-in uses a file that converts the abbreviated names to their full names.

Note — You can edit this file to include custom URL categories for conversion. For more information on how to do this, see “Custom URL Categories” on page 77.

- **Predefined reports.** The plug-in defines different reports you can view to get a better understanding of the web activity on your network. For a list of the different reports available, see “Sawmill for IronPort Reports” on page 4.

The IronPort log format plug-ins allows you to create the following profile types:

- **Security Operations (Sec Ops).** The Sec Ops profile type creates a profile that contains the most detail about web activity stored in the access logs. Profiles created from this profile type include the most number of predefined reports. Use the Sec Ops profile type to create analytical profiles for security related and operational requirements.
- **Human Resources (HR).** The HR profile type creates a profile that contains much less data and fewer reports than the Sec Ops profile type. Because this profile type filters out more data when parsing the access logs, most profile related actions, such as importing log files and report generation, see greater performance than a Sec Ops profile. Use the HR profile type to create tracking profiles if your organization wants to track and report on users web activity.

Sawmill for IronPort Reports

When you create a profile, Sawmill for IronPort generates different types of reports depending on the profile type used.

Figure 1-1 shows the types of reports displayed in the Sawmill for IronPort web interface for the Sec Ops profile type.

Figure 1-1 Report Types for Sec Ops Profile Type

• Overview
▸ Security
▸ Resource
▸ Date Reports
▸ Individual Fields
Summarized Logs
Log detail
Single-page Summary

Figure 1-2 shows the types of reports displayed in the Sawmill for IronPort web interface for the HR profile type.

Figure 1-2 Report Types for HR Profile Type

Overview
▸ Compliance
▸ Resource
▸ Date Reports
▸ Individual Fields
• Summarized Logs
Log detail
Single-page Summary

Note — You can create your own reports for a profile and save them. For more information on how, search for “The Report Editor” section in the Sawmill documentation at <http://www.sawmill.net>.

Table 1-1 describes the reports included in the Sec Ops profile type.

Table 1-1 Reports in the Sec Ops Profile Type

Report Type	Description
Overview	The Overview report shows a very basic high level view of the log data.
Security	The Security reports display information about malware detection on the network.
Resource	The Resource reports display information about network resource usage, such as proxy cache usage. The individual reports included in the Resource reports is different for the Sec Ops and HR profile types.
Date Reports	The Date Reports show web requests for different date ranges. These reports are useful for cross referencing from other reports so you can easily zoom on particular date ranges.

Table 1-1 Reports in the Sec Ops Profile Type (Continued)

Report Type	Description
Individual Fields	The Individual Fields reports show web requests for different fields in the access logs. These reports are useful for cross referencing from other reports so you can easily zoom on particular fields.
Summarized Logs	The Summarized Logs report is a more human readable version of the access logs that does not include all access log fields. You might want to use this report to show executive level management a summary of the network traffic. This report only includes rows for page views and not the images on a page.
Log Detail	The Log Detail report is a human readable version of the access logs that includes more fields than the Summarized Logs report. It is geared toward a technical audience that needs to see a lot of the data in the access logs in a more readable format.
Single Page Summary	<p>The Single Page Summary includes <i>every report</i> and combines them all onto one page in the web interface. You might want to view the Single Page Summary to print it out or email it to a manager. For example, if you zoom in on data for a particular user in your organization, you can then view the Single Page Summary and send that report to the person or his/her manager.</p> <p>Note: The Single Page Summary can take a very long time to process depending on the amount of data currently zoomed in on in Sawmill. IronPort recommends only viewing the Single Page Summary when you are zoomed in to a small subset of data, such as a single person, department, or particular time range.</p>

Table 1-2 describes the reports included in the HR profile type.

Table 1-2 Reports in the HR Profile Type

Report Type	Description
Overview	The Overview report shows a very basic high level view of the log data.
Compliance	The Compliance reports display information about how users are complying with acceptable use policies.
Resource	The Resource reports display information about network resource usage, such as proxy cache usage. The individual reports included in the Resource reports is different for the Sec Ops and HR profile types.
Date Reports	The Date Reports show web requests for different date ranges. These reports are useful for cross referencing from other reports so you can easily zoom on particular date ranges.

Table 1-2 Reports in the HR Profile Type (Continued)

Report Type	Description
Individual Fields	The Individual Fields reports show web requests for different fields in the access logs. These reports are useful for cross referencing from other reports so you can easily zoom on particular fields. The HR profile type includes a subset of the Individual Fields reports compared to the Sec Ops profile type.
Summarized Logs	The Summarized Logs report is a more human readable version of the access logs that does not include all access log fields. You might want to use this report to show executive level management a summary of the network traffic. This report only includes rows for page views and not the images on a page.
Log Detail	The Log Detail report is a human readable version of the access logs that includes more fields than the Summarized Logs report. It is geared toward a technical audience that needs to see a lot of the data in the access logs in a more readable format.
Single Page Summary	<p>The Single Page Summary includes <i>every report</i> and combines them all onto one page in the web interface. You might want to view the Single Page Summary to print it out or email it to a manager. For example, if you zoom in on data for a particular user in your organization, you can then view the Single Page Summary and send that report to the person or his/her manager.</p> <p>Note: The Single Page Summary can take a very long time to process depending on the amount of data currently zoomed in on in Sawmill. IronPort recommends only viewing the Single Page Summary when you are zoomed in to a small subset of data, such as a single person, department, or particular time range.</p>

Sawmill for IronPort Log Filters

The IronPort log format plug-in use log filters to specify how to process, categorize, and filter data from the access logs *before* it populates the database. Using log filters to filter out data before loading data into the database allows Sawmill to reduce storage and processing power needs to analyze your log data.

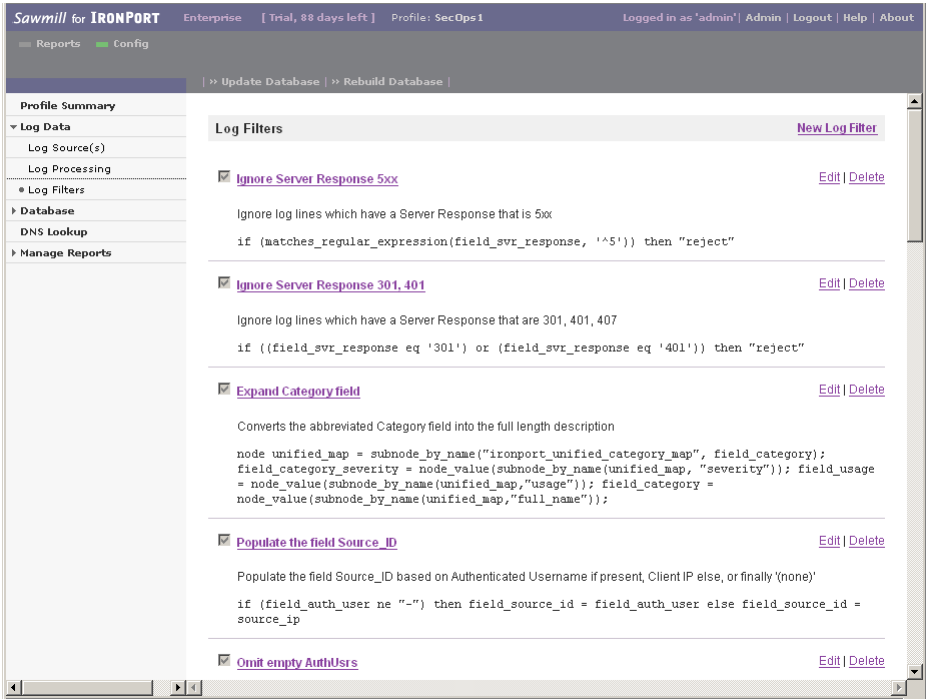
By default, a profile enables all log filters specified in the IronPort log format plug-in (except for the “Ignore log lines older than 45 days” filter, which is disabled by default). However, you can choose to modify, disable, or delete some of the log filters. You can also add additional log filters depending on the organization’s needs.

Each profile type includes different log filters. For example, the Sec Ops profile type include log filters that affect the different malware related fields, and the HR profile type includes a log filter that saves the server URL into the database.

To view the log filters in a profile, go to the Config page and then go to Log Data > Log Filters.

Sawmill for IronPort evaluates the log filters on the access log data in order, starting at the top of the list of filters. Figure 1-3 shows the log filters included in the Sec Ops profile type.

Figure 1-3 Sawmill for IronPort Log Filters—Sec Ops Profile Type



For more information on how to use log filters, see “Using Log Filters” on page 51.

Unlogged Fields

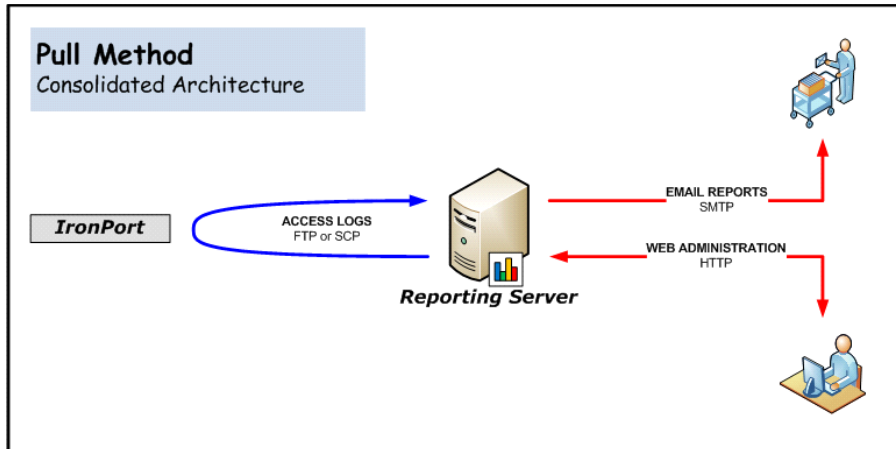
The IronPort Web Security appliance will populate all unused fields with a hyphen (-) to ensure that every line conforms to the same field format.

DEPLOYMENT PLANNING

Network Planning

Small Deployments

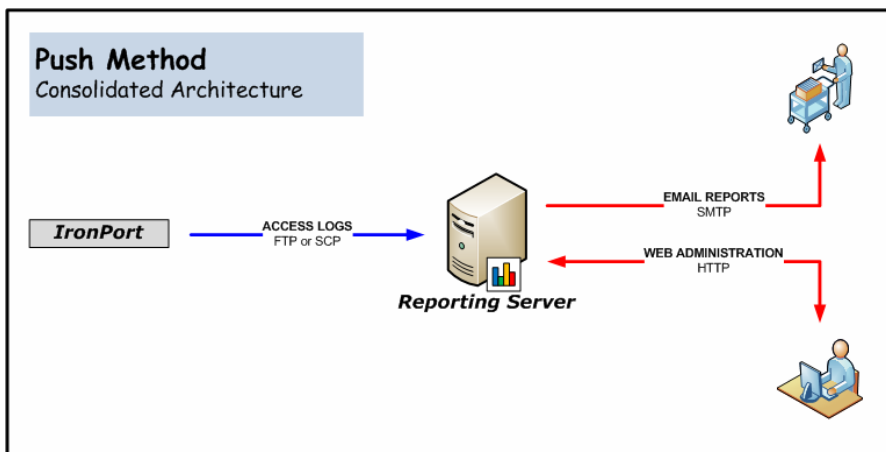
For IronPort evaluations or single appliance deployments there are two methods of data transfer available, but the Push Method is the preferred and the one intended to scale to large network environments.



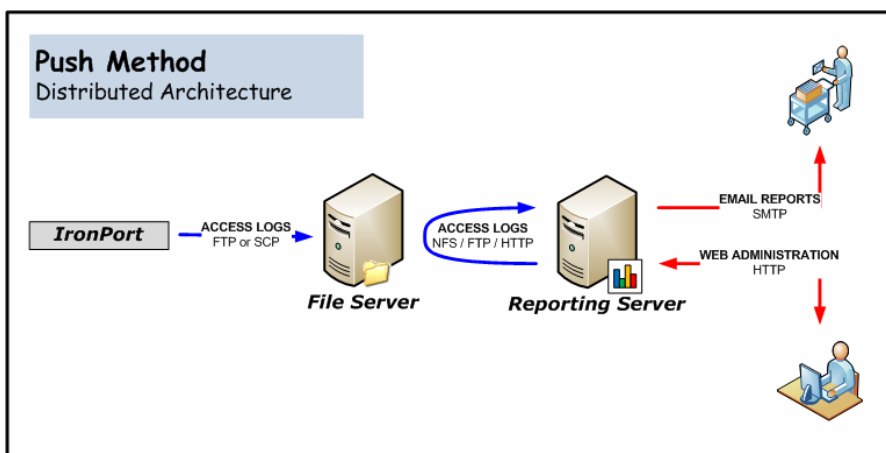
- **Pull Method.** Consolidated Architecture: Without Log Repository.
 - **Pro:** Requires less Reporting Server disk space.
 - **Con:** Once the Web Security appliance rotates its logs, the original logs will be lost, destroying forensic evidence and preventing Reporting database rebuilds.
- **Push Method.** Consolidated Architecture: With Log Repository.
 - **Pro:** The original logs will be available as forensic evidence and for Reporting database rebuilds.
 - **Con:** Reporting Server will require additional disk space.

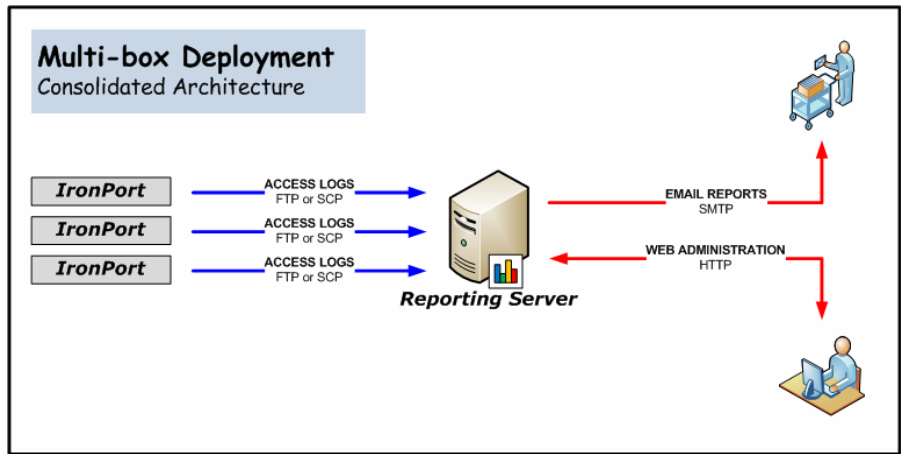
Large Deployments

- **Consolidated Push Method:** Have each Web Security appliance push its logs to a single off box repository where Sawmill is installed locally.



- **Distributed Push Method:** Have each Web Security appliance push its logs to a single off box repository and Sawmill is installed on another server.





Choosing an SCP or FTP Server

Assuming the Operating System is installed, and depending on the deployment method chosen, you may need to set up an SCP or FTP server for the reception of logs from the IronPort Web Security appliance. You can choose whatever server your organization has standardized on, but here are some different vendors to consider if needed:

Enterprise SCP Server

VShell® 3.0 Server for Windows and UNIX (price varies)

<http://www.vandyke.com/products/vshell/index.html>

Enterprise FTP Server

3Com 3CDaemon (free)

http://support.3com.com/software/utilities_for_windows_32_bit.htm

SAWMILL ARCHITECTURE OVERVIEW

This document provides a high-level overview of the internal architecture that is specific to Sawmill.

- **Log Importer.** A component which reads log data from a log source, and puts it into the Sawmill database.
- **Sawmill Database.** A database which keeps a copy of the log data, and is queried to generate reports.
- **Reporting Interface.** A web interface providing access to dynamic reports.
- **Administrative Interface.** A web interface for administrating profiles, users, tasks, and more.
- **Web Server.** A built-in web server providing a graphical interface for administrating and viewing reports.
- **Command Line Interface.** An extensive command-line interface that can be used to manage profiles, generate reports, and more.

Log Importer

Sawmill is a log analyzer. It reads text log data from a log source (usually log files on the local disk, a network mounted disk, or FTP), parses the data, and puts it in the Sawmill Database. Log Filters can be used to convert or filter the data as it is read, or to pull in external metadata for use in reports.

Sawmill Database

The Sawmill Database stores the data from the log source. The Log Importer feeds data into the database, and the Reporting Interface queries the database to generate reports.

Update Database vs. Rebuild Database

Update Database updates the database, by adding any new log data in the log source (data which is in the log source but not in the database) versus Rebuild Database which rebuilds the database from scratch. Needless to say, rebuilding the database should only be done under rare circumstances.

Reporting Interface

The Reporting Interface is an HTML interface delivered through the Web Server, to any compatible web browser. Reports are generated dynamically by querying the Sawmill Database. The Reporting Interface allows arbitrary filters to be applied to any report, including Boolean filters. Reports can be created or edited through the Administrative Interface.

Administrative Interface

The Administrative Interface is an HTML interface delivered through the Web Server, to any compatible web browser. The Administrative Interface is used to create and edit profiles, users, scheduled tasks, and more.

Web Server

The Web Server is an HTTP server built into Sawmill. It serves the Reporting Interface and the Administrative Interface. It is possible to use an external web server like Apache or IIS to serve the Sawmill interface, but IronPort Customer Support does not support this. You can read the Sawmill documentation at <http://www.sawmill.net> for more information.

Command Line Interface

Sawmill includes an extensive command-line interface to perform some tasks, such as creating profiles and building or updating the database. You can read the Sawmill documentation at <http://www.sawmill.net> for more information.

WHAT'S NEW IN SAWMILL FOR IRONPORT

This section describes new features and enhancements added in Sawmill for IronPort.

Sawmill for IronPort version 7.3.2 is based on FlowerFire Sawmill version 7.2.18 Enterprise. For more information about what is new and fixed in FlowerFire Sawmill version 7.2.18, go to:

http://www.sawmill.net/version_history.html

What's New in Sawmill for IronPort Version 7.3.2

Sawmill for IronPort version 7.3.2 does not have any new features.

What's New in Sawmill for IronPort Version 7.3.1

New Feature: New Log Filter

Sawmill for IronPort version 7.3.1 includes a new log filter, "Ignore log lines older than 45 days," which filters out log lines older than 45 days when enabled for both HR and Sec Ops profiles.

Note — The filter is disabled by default.

What's New in Sawmill for IronPort Version 7.3.0

New Feature: New Profile Types

In previous versions of Sawmill for IronPort, you could only create one type of profile. Now, Sawmill for IronPort version 7.3.0, contains two profile types, Security Operations (Sec Ops) and Human Resources (HR). The Sec Ops profile type is similar to the previous profile type, but not identical.

Enhanced: Supported Platforms

You can now install Sawmill for IronPort version 7.3.0 on the 64-bit version of Red Hat Linux Enterprise as well as on Microsoft Windows, both 32-bit and 64-bit.

New Feature: Support Tools

Sawmill for IronPort version 7.3.0 includes multiple tools created by IronPort sales engineers to ease deployment and use of Sawmill for IronPort. The tools are located in the LogAnalysisInfo\tools directory and are divided into subdirectories for each operating system. Example tools include a script for automatically fetching log files, and a batch file that checks if an anti-virus software application is configured to scan the Sawmill for IronPort installation directory. For more information on each tool, read its associated "readme" file.

FIXED ISSUES, KNOWN ISSUES, AND LIMITATIONS**Fixed: Inaccurate Page View Statistics**

The Page View calculations have been improved in Sawmill for IronPort version 7.3.1. Previously, the Page View statistics within various reports included data that was not necessary to make an accurate count of page views, such as entries for all responses. Now, Page View statistics use a filter to ignore all dropped or blocked client requests as well as server responses 1xx, 3xx, 4xx, and 5xx, none of which return data to the requested client. However, image files are incorporated into the Page View calculations. [Defect ID: 50257]

Fixed: Log Entries Containing Spaces in Usernames Do Not Appear in Reports

Previously, Sawmill for IronPort ignored log entries containing usernames with two or more elements with space between them. For example, log entries for the user “barack 2” would not appear in reports. This issue has been resolved. [Defect ID: 50075]

Fixed: HR Profile Erroneously Displays a “FIELD: Auth User” Report

Although the HR profile does not contain an Auth User field, the profile displays a FIELD: Auth User report. Please ignore this report in the HR profile. [Defect ID: 49697]

Fixed: Scheduling Database Updates in the Sawmill CLI Can Cause Sawmill for IronPort to Crash

Using the Sawmill CLI to schedule database updates can cause Sawmill for IronPort to crash.

Workaround: Use the Sawmill GUI to schedule database updates.

[Defect ID: 48356]

Fixed: Importing Access Logs Can Cause Sawmill for IronPort to Crash

Importing or updating the access logs can cause Sawmill for IronPort to crash.

Workaround: Import or update the access logs with a single processor, or set the Log Processing Threads in the Configuration Master file to 0.

[Defect ID: 55020]

Fixed: Updating the Database in Sawmill for IronPort May Throw an Application Error at the End of the Update

When you update the database, Sawmill for IronPort may throw an application error at the end of the update. This occurs regardless of the Log Processing Thread settings. However, the update has been successful, and no data corruption has occurred.

[Defect ID: 52227]

Fixed: Sawmill for IronPort Discards Session that are Over Two Hours Long

Sawmill for Ironport discards all work sessions that are over two hours long. The default session length for Sawmill for Ironport is a maximum of two hours.

[Defect ID: 55242]

Fixed: Sawmill for IronPort Does Not Allow Users to Enable WBRS in the Summarized Logs Report

Sawmill for Ironport is not allowing a user to enable the WBRS variable from the Table Options menu in the User Interface. Additionally, you cannot go to the profile.cfg for Summarized Logs, and change the field_wbbs_value to True.

[Defect ID: 55664]

Fixed: Sawmill for IronPort Incorrectly Reports Page View Calculations Due to MIME Type Entries

Page view calculations include log lines that have MIME types. However, in some cases, MIME types show up empty or with a '-'. In Sawmill for IronPort version 7.3.1, these log lines were not included in the page view calculations resulting in under-reported data. In Sawmill for IronPort version 7.3.2, this has now been corrected. As a result some of the calculations will show better accuracy resulting in different "time spent" numbers.

[Defect ID: 66678]

Multiple Log Formats in Same Directory Not Supported

If the log directory contains both logs in W3C format and logs in Squid format, reports may display incorrect data. Workaround: Avoid mixing log formats in the log directory; put the W3C logs in a separate directory. Sawmill for Cisco IronPort officially supports the Squid format. [Defect ID: 53890]

"Date Offset" Supports Whole Hours Only

If you enter fractional hours (for example, 9.5) when manually adjusting the time zone ("Date offset") in the Sawmill Profile, Sawmill for IronPort reports do not display the correct time.

[Defect ID: 43693]

Generating "All Reports" Does Not Always Work if the Database Is Large

Attempting to generate all reports at once sometimes causes errors if the database is large. Workaround: Schedule reports to be generated individually. [Defect ID: 49122]

Zoomed Reports Show Zero Results When Using Global Filters in Some Cases

When you zoom in on data using a Global Filter that uses an expression with an OR clause, the data sometimes erroneously shows zero results.

Workaround: Rewrite the expression so it does not use an OR clause.

Custom URL Category Number Erroneously Appears as Access Policy Name

In versions of AsyncOS for Web before version 6.0, when an Access Policy allowed a transaction due to an allowed custom URL category, the custom URL category number appears as the Access Policy name in Sawmill for IronPort. For example, the Web Security

appliance access logs show “ALLOW_CUSTOMCAT-AccessPolicy1-1090519042” as part of the ACL decision tag. In this example, “1090519042” appears as the Access Policy name instead of “AccessPolicy1.” [Defect ID: 45336]

When Running on a 32-Bit Operating System, Sawmill for IronPort May Have Difficulty Handling Access Logs Larger than 2G

When running on a 32-bit operating system, Sawmill for IronPort may have difficulty handling access logs that are larger than 2G.

Workaround: Set the maximum file size for access logs to 2G.

[Defect ID: 48407]

Typo in the Daylight Savings Time Log Filter in the Online Help

The Daylight Savings Time log filter in the Sawmill for IronPort online help incorrectly refers to `v.date_time_year` as `v_date_time_year` in a couple of places. All references in the filter should read `v.date_time_year`; otherwise, a syntax errors occurs if you attempt to use the filter. [Defect ID: 49693]

Disabling the Rewrite URL Feature Causes Sawmill for IronPort to Stop Functioning Properly

Disabling the Rewrite URL feature causes Sawmill for IronPort to stop functioning properly and throws a ‘Running out of Memory’ error.

Workaround: Do not disable the ReWrite URL filter in Sawmill for Ironport Plugin. The Rewrite URL feature must remain enabled.

[Defect ID: 54042]

The Default Pathname Suggested in the New Profile Wizard is Problematic

The default pathname suggested when selecting a Log Source in the New Profile Wizard is `logs/*.gz, logs/*, logs/access.log`. The `logs/*` causes Sawmill for IronPort to import duplicate entries into the database.

Workaround: Replace the suggested default string with `logs/.*(s|c)$` so that duplicate database entries do not occur.

[Defect ID: 55303]

Sawmill for IronPort Incorrectly Reports CONNECT Log Entries

Page view identification logic was modified in 7.3.2. As a result, in 7.3.2, some loglines for HTTPS CONNECT request that were not identified as page view previously, may be counted as page view against an '(empty)' domain. Summarized Logs report will show the log lines that are identified as page views.

[Defect ID: 67385]

Installation and Configuration

This chapter contains the following information:

- “Overview” on page 20
- “Installing and Configuring Sawmill for IronPort” on page 21
- “Working with Time Zones” on page 24
- “Accessing Windows Mapped Drives” on page 25

OVERVIEW

You can install Sawmill for IronPort on the following platforms:

- Microsoft Windows, 32-bit
- Microsoft Windows, 64-bit
- Red Hat Linux Enterprise, 64-bit

Sawmill for IronPort runs as a web server. You can access it by opening a web browser to the following URL:

`http://<Sawmill_hostname_or_IP_address>:8987/`

Note — On Windows, Sawmill for IronPort runs as the SYSTEM user by default, which is the most secure approach, but restricts access to network shares or mapped drives. If Sawmill for IronPort needs to access some mapped drives and have different network privileges, you can run it as a different Windows user. For information on how to do this, see “Accessing Windows Mapped Drives” on page 25.

Installing and configuring Sawmill for IronPort is a multiple step process. You must complete the following steps:

- **Configure any antivirus software.** If the machine that runs Sawmill for IronPort also runs some antivirus software, then you need to configure the antivirus software to not interfere with the Sawmill database files. To do this, configure the antivirus software to *not scan* the LogAnalysisInfo directory in the Sawmill for IronPort installation directory. This can potentially increase Sawmill for IronPort performance and can prevent inconsistent behavior when the antivirus software tries to “cleanse” some of the Sawmill database files.

WARNING: If an antivirus program scans the Sawmill for IronPort installation directory, Sawmill for IronPort performance can be severely impacted and the main database might become corrupt. Verify no antivirus program scans this directory.

- **Install and configure Sawmill for IronPort.** Run the Sawmill installer and configure Sawmill for your log data. For more information, see “Installing and Configuring Sawmill for IronPort” on page 21.

INSTALLING AND CONFIGURING SAWMILL FOR IRONPORT

To install and configure Sawmill for IronPort, complete the following steps:

- “Step 1. Install Sawmill for IronPort” on page 21
- “Step 2. Configure Sawmill for IronPort Setup” on page 21
- “Step 3. Log In and Create a Profile” on page 21

Step 1. Install Sawmill for IronPort

To install Sawmill for IronPort, complete the following steps:

1. Open a web browser and log in to the IronPort Customer Support Portal.
2. In the Support Portal, navigate to the Web Security Appliances page and download the archived Sawmill installer that is appropriate for your operating system.
3. Unzip the Sawmill installer and then run it. Take all default options.

Note — For more information on installing on Linux, see the Sawmill documentation provided by Flowerfire.

4. After installation is complete, Sawmill Setup opens in a web browser.

Step 2. Configure Sawmill for IronPort Setup

After you run the installation, it opens up Sawmill Setup in a web browser. Complete the following steps to configure Sawmill for IronPort:

1. Click **Next** in the Sawmill Setup.
2. Accept the terms of the license agreement, and click **Next**.
3. On the Licensing screen, enter the license key from IronPort, and click **Next**.
4. On the Administrative User screen, enter a user name and password for the administrator user that accesses Sawmill. Click **Next**.
5. On the Automated Feedback Agent screen, choose whether or not to allow Sawmill to send information to the Sawmill development team at Flowerfire about the types of devices your Sawmill installation analyzes. Click **Next**.

Note — Sawmill does not send your log data or any other personally identifiable information to Flowerfire or IronPort.

6. On the Complete Setup screen, click **Finish**.

The Sawmill Login screen appears.

Step 3. Log In and Create a Profile

After you install and configure Sawmill for IronPort, complete the following steps to create a Sawmill Profile and its corresponding Reports:

1. On the Sawmill Login screen, enter the user name and password you chose in step 4 on page 21 under Step 2. Configure Sawmill for IronPort Setup .

When a Sawmill installation includes no Profiles, it prompts you to create a profile.

2. Click **Start Here** to create a Profile using the IronPort configuration files.

The New Profile Wizard appears in a separate browser window.

3. On the Log Source screen, enter the following information:

Field	Description
Log source	<p>Specify how Sawmill should access the access log file:</p> <ul style="list-style-type: none">• Local disk. Click Browse to locate the path on the local machine.• FTP server. Enter the host name of the FTP server and the user name and password used to access the server.• HTTP server. Enter the host name of the HTTP server. <p>The source type you choose depends on how you choose to deploy Sawmill for IronPort. For more information choosing how to deploy Sawmill for IronPort, see “Deployment Planning” on page 9.</p>
Pathname	<p>Enter the directory path on the machine hosting the access log file that contains the access log file. If the access log file is on a local or network drive, you can use the Browse button to navigate to the directory.</p> <p>You can also enter a pattern to specify multiple log source file names. For example, to specify all files with a .log extension in the logs directory, you can enter the text using the following types of patterns:</p> <ul style="list-style-type: none">• Wildcard: C:\logs*.log• Regular expression: C:\logs\^.*log\$ <p>If you enter a regular expression in the Pathname field, you must enable the “Pattern is a regular expression” option.</p>
Process subfolders	<p>Choose whether or not to also process data in subdirectories of the path specified in the pathname field. This field only applies to log files stored on the machine hosting Sawmill.</p> <p>If you enter a pattern in the Pathname field, this option searches for that pattern in all subdirectories.</p>
Pattern is a regular expression	<p>Enable this option if you entered a regular expression in the Pathname field.</p>

4. Click **Next**.

Sawmill reads the log files in the specified path and tries to detect the log format.

5. On the Log Format Detected screen, select the log format for the type of profile you want to create (HR or Sec Ops).

6. Verify “Continue with the above detected log format” is selected, and click **Next**.
7. On the Numerical Field Options screen, verify all check boxes are selected and click **Next**.
8. On the Database Options screen, choose to use the internal database created by Sawmill, and click **Next**.
9. On the Profile Name screen, enter a name for this Profile, and click **Finish**.
10. After the Profile has been saved, click **Close**.
11. Return to the browser window where you are logged into Sawmill.

Note — You might need to configure the time zone in the Sawmill Profile. For more information, see “Working with Time Zones” on page 24.

WORKING WITH TIME ZONES

The Web Security appliance access logs record log times in UNIX time, which uses no time zone information. The times recorded in the access logs are independent of the time zone setting on each Web Security appliance.

Therefore, if you want the reports in Sawmill for IronPort to display the local time on the Web Security appliance from which the access logs came, you need to manually adjust the time zone in the Sawmill Profile. Do this in the “Date offset” field on the Config page > Log Data > Log Processing.

For example, if the Web Security appliance is in the Pacific Standard Time zone, enter “-8” in the “Data offset” field. Note that the Web Security appliance uses the POSIX-style method of indicating the time zone using a GMT offset. So, for the Pacific Standard Time zone, the Web Security appliance is configured for GMT+8, and the Sawmill Profile is configured for “-8.”

ACCESSING WINDOWS MAPPED DRIVES

Sawmill runs as the SYSTEM user by default, which is the most secure approach, but restricts access to network shares or mapped drives. If Sawmill for IronPort needs to access some mapped drives, such as H:\, you must complete some additional steps to access data on the mapped drives.

To access Windows mapped drives:

1. Change the Sawmill for IronPort Windows service to run as an administrative user, and restart the service.
(Open Services from the Windows Control Panel, double-click the Sawmill for IronPort service, click the Log On tab, choose "This account," and enter the user name and password for the administrative user. Restart the service.)
2. If the shared drive is password protected, authenticate to it using the **Network Drives** button of the Sawmill file browser.
3. When you create a profile, enter the Universal/Uniform Naming Convention (UNC) path in the Pathname field for the log source. Do *not* enter a mapped drive letter, such as S:\. The UNC path format on Windows is \\ComputerName\SharedFolder\directories\file.

Navigating Sawmill for IronPort

This chapter contains the following information:

- “Sawmill Web Interface Overview” on page 28
- “Reports Page” on page 29
- “Config Page” on page 33
- “Administrative Page” on page 35

SAWMILL WEB INTERFACE OVERVIEW

The Sawmill web interface has the following main areas where you can configure the software or reports:

- **Reports page.** Use this page to view the different reports generated for a particular profile. You can zoom in on data in a report and print out reports. For more information, see “Reports Page” on page 29.
- **Config page.** Use this page to configure options for a particular profile. You configure options specific to a profile, such as the log source, log filters, and which reports are generated. For more information, see “Config Page” on page 33.
- **Administrative page.** Use this page to configure global options that apply to all profiles, such as licensing tasks, general preferences, users, and scheduled tasks. For more information, see “Administrative Page” on page 35.

Figure 3-1 shows how you access the different pages in the Sawmill web interface.

Figure 3-1 Sawmill Web Interface Pages

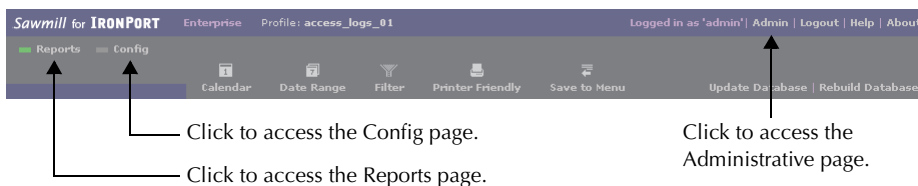
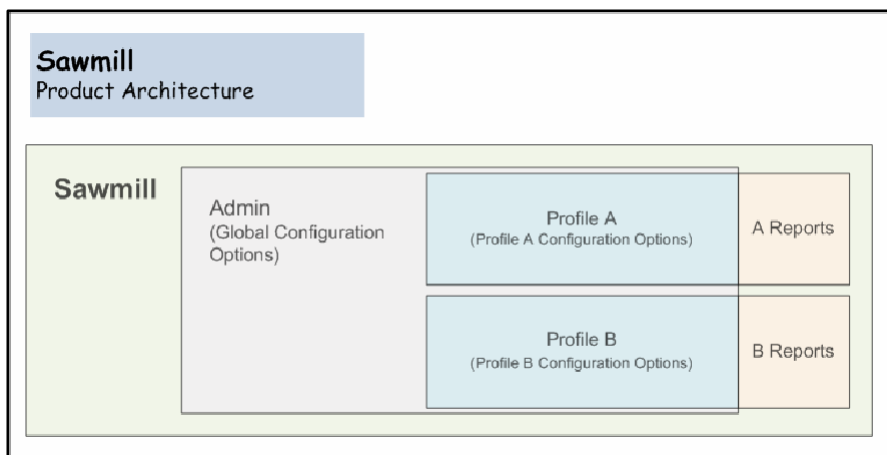


Figure 3-2 shows how the different parts of the Sawmill web interface related to each other in the Sawmill architecture.

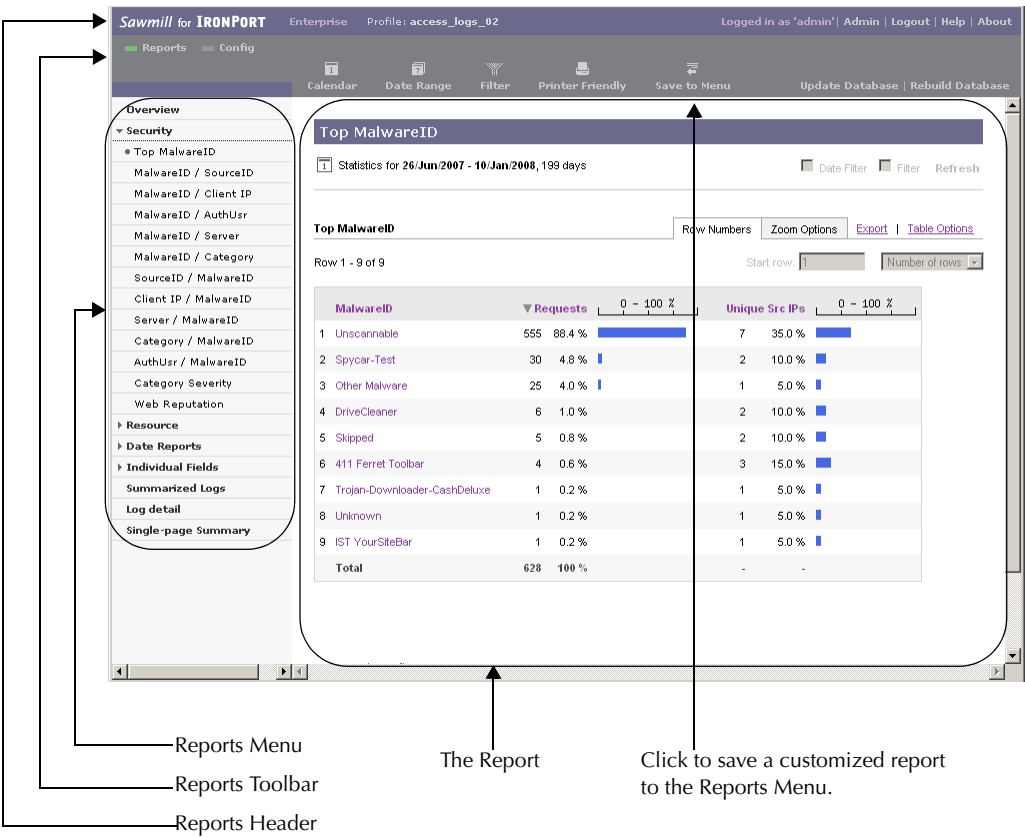
Figure 3-2 Sawmill Architecture



REPORTS PAGE

Sawmill Reports present Web Security appliance access log file information in an attractive and easily navigable format. Figure 3-3 shows the Reports page where you view and configure reports.

Figure 3-3 Reports Page



This section describes the different components of the Reports page.

Reports Header

The header of the Reports page is a bar containing the following:

- **Profile Name.** The name of the active profile, the profile whose reports are being displayed
- **Admin.** A link to the administrative interface, the profiles list, and other administrative functions

- **Logout.** A link to log out of Sawmill
- **Help.** A link which opens a new window containing the Sawmill documentation provided by Flowerfire

Reports Toolbar

Below the header is a bar which contains the following objects:

- **Reports.** A link to the Reports, which is selected when you are currently viewing the Reports page. Click this link to get to the Reports from another page.
- **Config.** A link to the Config page, where you can perform profile actions, such as rebuilding the database, and change profile options. For more information, see “Config Page” on page 33.
- **Calendar.** Click this to open the Calendar window, where you can select a single day, month, or year to use as the date/time filter. When you have selected an item in the Calendar, *all reports* show only information from that time period until the date/time filter is removed by clicking “Show All” in the Calendar.
- **Date Range.** Click this to open the Date Range window, where you can select a range of days to use as the date/time filter. When you have selected a range in the Date Range, *all reports* show only information from that time period until the date/time filter is removed by clicking “Show All” in the Calendar.
- **Filter.** Click this to open the Global Filter window, where you can create filters for any fields in any combination. Filters created here dynamically affect all reports. Once you set a Global Filter, *all reports* show only information for that section of the data. Global Filters remain in effect until you remove them in the Global Filter window.
- **Printer Friendly.** Click this to open a separate browser window with the current report displayed suitably for printing. After the new window appears, use the browser’s Print function to print it. Printing in this way gives better results than printing the original report directly.
- **Save to Menu.** Click this button to save the current report with all of its zoom settings to the Reports Menu.

Reports Menu

At the left of the window is the Reports Menu, which lets you select the report to view. Clicking a category will expand or collapse that category. Clicking a report name changes the report display to show that one.

Note — Clicking a report name removes any Zoom filters, but does not remove Global Filters or Date/Time Filters.

The Report

The main portion of the window displays the report itself. This is a view of the data selected by the filters (global filters, date/time filters, and zoom filters). This provides one breakdown of

the data specified by the filters. You can select another report in the Reports Menu to break down the same data in a different way.

This section describes the different parts of a report.

The Report Bar

At the top of the report is a bar containing the report label and the current global and date/time filters, if any.

The Zoom Display

The Zoom Display shows what you are zoomed in on, if you are using Zoom filters. For instance, if you have clicked on item X in a table for field F, you will see “Zoomed in on F: > X”. You can apply a zoom filter by clicking on a linked table item. Zoom filters disappear when you click a new report name in the Reports Menu at the left.

The Zoom To Menu

The Zoom To Menu shows the name of the report that will be displayed when you zoom. For instance, if you select R from the menu, and then click an item X in a table for field F, it will zoom you in on X, and simultaneously switch to report R. This can be useful if you want to break down each item in a table by some other report.

If you are not zoomed on anything, there will be no immediate effect when you select a new item in the Zoom To Menu. Instead, it will just change the menu selection. But when you click, the selection in the Zoom To Menu will be used to determine which report to display.

The Report Graph

Some reports have a graph above the table. The graph’s characteristics, such as size and type, vary from report to report. You can also edit the graph’s characteristics in the profile config. The graph displays the same information as the table below it.

The Report Table

The Report Table contains the main information of the report. It displays one row per database field item, with the aggregated numerical values in columns next to it. It may also include columns showing bar graph representations of the numbers, and/or percentages.

You can configure the following options:

- **Table Options link.** Located above and to the right of the table, this link can be used to change which columns are visible, the sort order, and other aspects of the report.
- **Number Of Rows menu.** Located above and to the right of the table, this menu can be used to change the number of rows that are displayed in the table.
- **Sort order.** You can change the sort order by clicking a column name. Click once to sort by that column, or again to sort in reverse.
- **Zoom.** You can zoom in on a particular item by clicking it. Once you click an item, Sawmill switches you to the report specified in the Zoom menu and sets a Zoom filter for that item.

Filters

There are many levels of filters when viewing reports:

- **Global Filters.** These remain in effect until they are removed in the Global Filters page.
- **Date/Time Filters.** These remain in effect until they are removed in the Calendar page.
- **Zoom Filters.** These remain in effect until they are removed in the Calendar page.
- **Report Filters.** These are set per report in the profile config, and cannot be removed from the web interface.
- **Report Element Filters.** These are set per report-element in the profile config, and cannot be removed from the web interface.

All these filters are “ANDed” together. That is, an item is in the total filter set if it is selected by the Global Filters AND by the Date/Time Filters AND by the Zoom Filters AND by the Report Filters AND by the Report Element Filters. For instance, if the Global Filters show events during 1am-2am, and the Zoom Filters show events on January 1, then the table will show events on January 1, during 1am-2am.

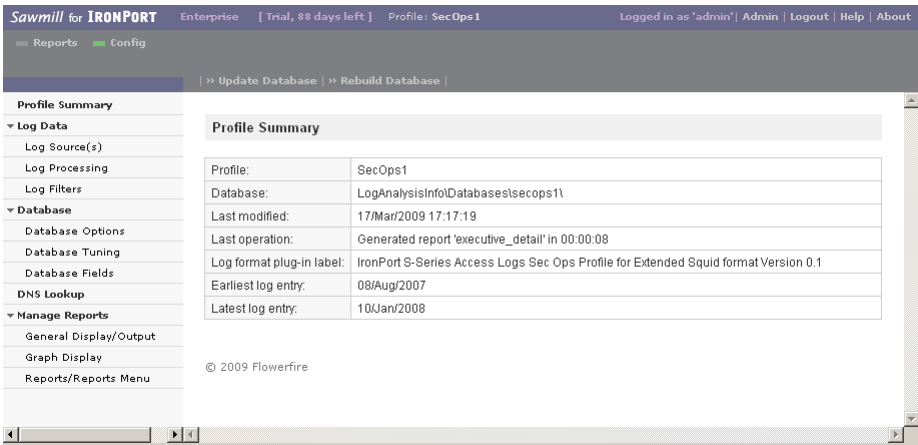
CONFIG PAGE

The Config page provides a graphical interface for editing most aspects of profiles, including the log source, log filters, database fields, and other options. You can access the Config page in the following ways:

- In the profile list, click **View Config**.
- On the Reports page, click **Config**.

The Config page is divided into different sections, with a link to each section in the left column. The Config page also includes buttons to manage the Sawmill database. Figure 3-4 shows the Config page where you configure a profile.

Figure 3-4 Config Page



This section describes the different components of the Config page.

Profile Summary

This is an information page, providing information about the profile, including the location and type of the database, the log format, details of the last operation, and the date range.

Log Data

The Log Data section contains the following subsections:

- **Log Source(s).** This is the Log Source Editor, which describes the log sources for the profile. A log source specifies the location of log data to analyze in this profile, or the method of acquiring the log data.
- **Log Processing.** This section contains the log processing options, which control how log data is read from the log sources.

- **Log Filters.** Log Filters perform translations, conversions, or selective inclusion (“filter out”) operations. For instance, a log filter could be used to reject (exclude) all log entries from a particular IP, or all log entries during a particular time. Log Filters could also be used to convert user names to full names, or to simplify a field, such as truncating the end of a URL, which is sometimes necessary to analyze a large proxy dataset efficiently.

Log Filters are written in The Configuration Language, which provides full programming language flexibility, including the use of if/then/else clauses, and/or/not expressions, loops, and more. For more information, see the Sawmill documentation.

Database

The Database section contains the following subsections:

- **Database Options.** This section contains the general database options, including the type and location of the database.
- **Database Tuning.** This section contains options for tuning database performance.
- **Database Fields.** This is an information section, displaying information about the database fields.

DNS Lookup

This section contains the DNS options used to look up IP addresses in the log data.

Manage Reports

The Manage Reports section contains the following subsections:

- **General Display/Output.** This section includes general report output options, including headers and footers.
- **Graph Display.** This section includes general graph options, including default graph sizes.
- **Reports/Reports Editor.** This section includes the Reports Editor and the Reports Menu editor, which can be used to create custom reports.

Rebuild Database Button

This button appears at the top of the profile config editor. Clicking it rebuilds the database from scratch.

WARNING: This option erases the entire contents of the current database and builds it from the current contents of the log sources. If the database contains data that is no longer in the log sources, that data will be permanently lost.

Update Database Button

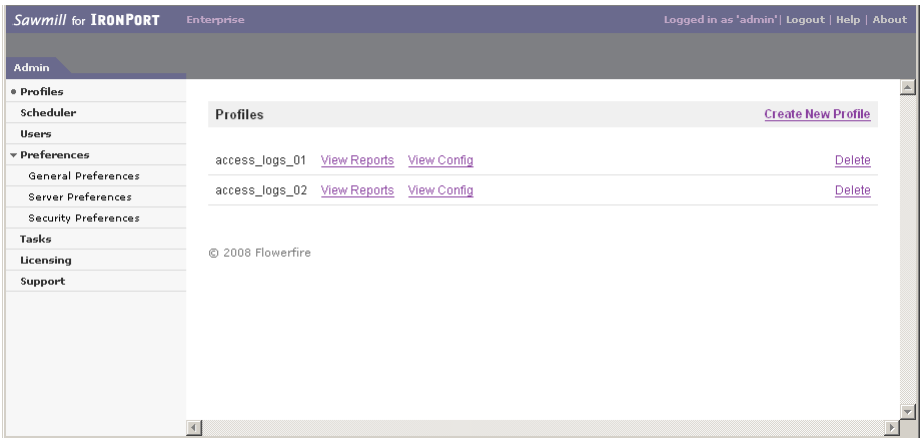
This button appears at the top of the profile config editor. Clicking it updates the database by adding any new log data in the log source (data which is in the log source but not in the database).

ADMINISTRATIVE PAGE

The Administrative menu appears at the left of the main administrative page. It provides basic administrative functions. This section describes the different components of the Administrative page.

Figure 3-5 shows the Administrative page where you administer Sawmill for IronPort.

Figure 3-5 Administrative Page



Profiles

The Profiles section lists the profiles created in this instance of Sawmill for IronPort. From here you can view the reports for a particular profile, or configure or delete the profile. You can also create new profiles by clicking the Create New Profile link. Clicking the View Reports link takes you to the Reports page for that profile, and clicking View Config takes you to the Config page for that profile.

Licensing

Clicking this link will show the licensing page. In this page, you can add and remove licenses.

Users

Clicking this link will show the User Editor. In this page, you can add and remove users, and change the options for each user. For example, you can specify which users have administrative access, and which profiles they are permitted to view. For more information, see “Working with Users” on page 54.

Scheduler

Clicking this link will show the Scheduler. You can create, delete, and edit scheduled tasks in this section. For instance, you can create a task to update all your databases every night, or to

send a report of the previous month by email on the 1st of each month. For more information, see “Scheduling Sawmill Tasks” on page 55.

Preferences

Clicking this link will show the Preferences editor. This lets you change global preferences, including server IP and port, language, charset, and more.

Reading Access Logs with Sawmill for IronPort

This chapter contains the following information:

- “Working with Profiles” on page 38
- “Working with Reports” on page 40
- “Working with Report Filters” on page 43
- “Using Log Filters” on page 51
- “Working with Users” on page 54
- “Scheduling Sawmill Tasks” on page 55

WORKING WITH PROFILES

You can create a profile at any time. The number of profiles you can create is determined by the license you purchase. Some of the profile options define the following:

- **Log source to analyze.** A single profile can analyze one or more access logs.
- **Log database.** Profiles create and use a database containing the processed log data. Use the internal Sawmill database with Sawmill for IronPort.
- **Reports.** Each profile includes reports that display the data using tables and graphs. Sawmill for IronPort uses reports defined by IronPort, and the included reports depend on the profile type, either Sec Ops or HR. You can create new and modify existing reports, if necessary.

After you create a profile, you can view its reports on the Reports page. Sawmill reads and processes the log data to build its database, and then displays the reports. You can also configure the different profile options on the Config page.

Creating a Profile

To create a profile:

1. Go to the Admin page.
2. Click **Create New Profile**.

The New Profile Wizard appears in a separate browser window.

3. On the Log Source screen, enter the following information:

Field	Description
Log source	<p>Specify how Sawmill should access the access log file:</p> <ul style="list-style-type: none">• Local disk. Click Browse to locate the path on the local machine.• FTP server. Enter the host name of the FTP server and the user name and password used to access the server.• HTTP server. Enter the host name of the HTTP server. <p>The source type you choose depends on how you choose to deploy Sawmill for IronPort. For more information choosing how to deploy Sawmill for IronPort, see “Deployment Planning” on page 9.</p>
Pathname	<p>Enter the directory path on the machine hosting the access log file that contains the access log file. If the access log file is on a local or network drive, you can use the Browse button to navigate to the directory.</p>
Process subfolders	<p>Choose whether or not to also process data in sub directories of the path specified in the pathname field. This field only applies to log files stored on the machine hosting Sawmill.</p>

Field	Description
Pattern is a regular expression	Do not enable this option. The IronPort log format plug-in does not use regular expressions to read the access logs.

4. Click **Next**.
Sawmill reads the log files in the specified path and tries to detect the log format.
5. On the Log Format Detected screen, select the log format for the type of profile you want to create (HR or Sec Ops).
6. Verify “Continue with the above detected log format” is selected, and click **Next**.
7. On the Numerical Field Options screen, verify all check boxes are selected and click **Next**.
8. On the Database Options screen, choose to use the internal database created by Sawmill, and click **Next**.
9. On the Profile Name screen, enter a name for this Profile, and click **Finish**.
10. After the Profile has been saved, click **Close**.
11. Return to the browser window where you are logged into Sawmill.

WORKING WITH REPORTS

When you create a profile, Sawmill for IronPort generates different types of reports depending on the profile type used. You can view reports at any time by clicking Show Reports next to the name of a profile in the administrative profile list. You can also switch to the reports when you are editing the profile options by clicking the Reports link in the upper left.

For more information on the reports Sawmill for IronPort defines for each profile, see “Sawmill for IronPort Reports” on page 4.

Creating Reports

You can report on information in the following ways:

- **Create a new report.** You can use the Config page for a profile to create reports in addition to the ones Sawmill for IronPort defines when you create a profile. For more information, see “Creating a New Report” on page 40.
- **Drill down in an existing report.** You can “drill down” on links in existing reports to zoom in on particular subsets of data in the report. This is the most common way to create reports on the data you are interested in. For more information, see “Drilling Down in an Existing Report” on page 41.

Creating a New Report

You can use the Config page for a profile to create reports in addition to the ones Sawmill for IronPort defines when you create a profile. This section includes instructions for creating a new report that four particular fields.

For more details on creating reports, see the Sawmill documentation at <http://www.sawmill.net>.

To create a report with the MalwareID, Server, Auth User, and Client IP fields:

1. Navigate to the Config page for the profile where you want to create the report.
2. In the Reports Menu, choose Manage Reports > Reports/Reports Menu, and then click **New Report**.
3. On the Report Options tab of the New Report dialog box, enter a menu name. This will be the name that appears in the Reports Menu area.
4. Click the Report Elements tab, and click **New Report Element**.
5. In the Report Element Type field, select Log detail.
6. On the General tab of the New Report Element dialog box, enter a report element name. You might want to enter a name that indicates the fields this report will show.
7. In the Report element filter field, enter:
`not (field_malware_id matches_regexp '^(empty')`
8. Click the Fields tab and remove all fields except for the following:
MalwareID, Server, Auth User, Client IP

Note — To create a report with different fields, choose the desired fields in the Active Fields section on the Fields tab.

9. Click **OK**.
10. Click **Save and Close** in the New Report dialog box.
11. Navigate to the Reports page. The new report appears at the bottom of the Reports Menu.

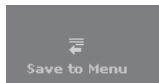
Drilling Down in an Existing Report

You can “drill down” on links in existing reports to zoom in on particular subsets of data in the report. This is the most common way to create reports on the data you are interested in. You drill down on the data in a report using any of the different report filters. For more information on report filters, see “Working with Report Filters” on page 43.

Note — For information on how to view particular information by drilling down in a report, see “Viewing Particular Detailed Information” on page 81. For example, this section includes information on viewing website activity per client.

Saving Reports

When viewing a report, you can zoom in on the detail using any of the filters provided with Sawmill for IronPort. You can save any report with its current filters applied by clicking the Save to Menu button in the Reports Toolbar.



The report name you specify appears at the bottom of the Reports Menu on the left side of the web interface.

Printing Reports

You can send a report to the printer. However, reports from the Reports page are not optimized for printing. Before you print a report, click the **Printer Friendly** button in the Reports Header.



The report appears in a separate browser window and is optimized for printing. Use the browser’s print function to print the report.

Creating a PDF of a Report

If you can create a PDF of an HTML file in your web browser using some third party software, such as the Acrobat PDF Writer, you can create a PDF of a report. Before you print the report

to a PDF file, click the **Printer Friendly** button in the Reports Header, and then print from the new browser window.

WORKING WITH REPORT FILTERS

There are many filters available to you when viewing reports. The report filters that are applied to the report determine what statistics you see. The filters let you “zoom in” on one part of your data. You can use the filters to get information about a particular day, a particular directory, a particular domain, or more.

- **Global Filters.** These remain in effect until they are removed in the Global Filters page. See “Global Filters” on page 44.
- **Date/Time Filters.** These remain in effect until they are removed in the Calendar page. See “Date/Time Filters” on page 46.
- **Zoom Filters.** These remain in effect until they are removed in the Calendar page. See “Zoom Filters” on page 49.

All of these filters are combined when used together; i.e., an item is included if it is selected by the Global Filters AND by the Date/Time Filters AND by the Zoom Filters. For instance, if the Global Filters show events during 1am-2am, and the Zoom Filters show events on January 1, then the table will show events from January 1, during 1am-2am.

If there are no filters in place, that means you are looking at your complete data; all available data is represented by the graphs and tables shown. If the The Report Bar shows that there are filters active, then you are not seeing your entire data; you are seeing only a portion of it. The portion you are looking at depends on the filters. For example, if the only filter is a /dir1/ filter on the page field, then the data displayed shows only those hits which were on /dir1/ or pages contained in /dir1/ (or in other directories contained in /dir1/, or pages in them, etc.). If you have 1000 hits on your site, and 500 of them were inside /dir1/, then if there are no filters active, you will see 1000 hits in the tables and graphs, or all the hits on your site. But if there is a Filter /dir1/ on the page field, you will see 500 hits, or only those hits in /dir1/.

The filters are an extremely powerful way of getting detailed information about your site. If you want to know what day you got the most hits on /dir1/, you can do that by adding /dir1/ as a filter, and then changing to the “Years/months/days” view. With /dir1/ as a filter, you will see only those hits on /dir1/ (500 of them, in the example above), and you will see how those 500 hits break down by date and time. You can add an additional filter to the date/time field if you want to examine just the hits on /dir1/ on a particular day. This gives you almost infinite flexibility in how you want to examine your data.

Another way to change filters is to click on something in the statistics. For instance, clicking on a directory name in a page table will “zoom in” on that directory by adding it as the page field filter. Clicking on a month in the calendar view will “zoom in” on the month by adding it as the date/time field filter. If the “filter checkboxes and menu” option is turned on (using the Show menu; see below), you can click the check box next to any item and choose a view from the menu below the table, to set that item as a filter, and then change the view to the view you select. To answer the question, “What days did I get hits on this directory?” you can click the checkbox next to a particular directory name, and select “Show checked data in ‘Top days’ view”, this will add the directory as a page filter, and also change to the top days statistics view.

Global Filters

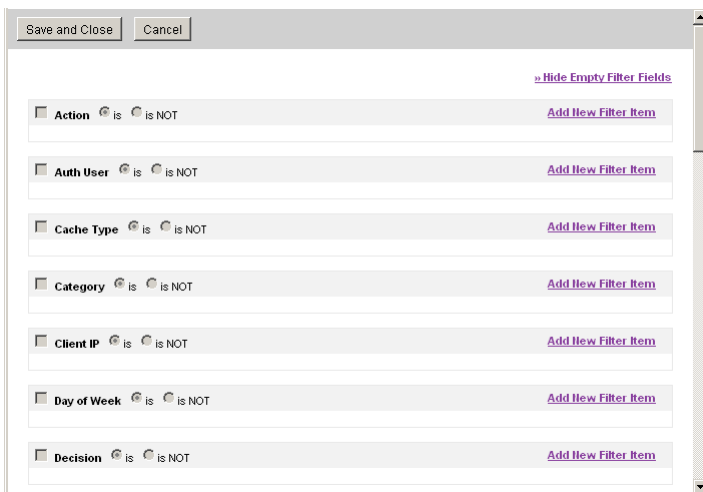
The Global Filters Editor lets you add, change, and remove report filters from any field available in the database. You can use Global Filters to filter on multiple values of any field, such as a range of IP addresses. When you configure Global Filters, you can define the values using the following methods:

- Literal string
- Wildcard expression
- Regular expression

Use Global Filters to filter data in ways you are not able to using Zoom Filters. For example, if you want to filter on two IP addresses, do that using Global Filters. You cannot filter on two IP addresses simultaneously using Zoom Filters.

To access the Filter Editor, click the **Filter** button in the Report Toolbar.

Figure 4-1 Global Filters Editor



To filter one of the report views, click the “Add New Filter Item” link next to the field you want to filter. You will see this:



The three radio buttons are the three options for the way in which Sawmill can match what you want to see in the reports. Select the one you want to use, type in the search string, and click **OK**. These are examples of what you will see:

The Name filter will match exactly what you type in, so if you type in “192.168.1.200” as the client IP address, Sawmill will look for that string and show you all matches for it in the report.

Edit Client IP filter item

☒ Name ☐ Wildcard Expression ☐ Regular Expression

OK

Cancel

☒ Client IP ☒ is ☐ is NOT [Add New Filter Item](#)

☒ 192.168.1.200 [Edit](#) [Delete](#)

The Wildcard Expression filter matches using the standard Windows wildcard characters (* and ? among others). If you type “192.168.1.2*” as the MIME type, Sawmill will return all MIME types that start with “192.168.1.2”.

Edit Client IP filter item

☐ Name ☒ Wildcard Expression ☐ Regular Expression

OK

Cancel

☒ Client IP ☒ is ☐ is NOT [Add New Filter Item](#)

☒ 192.168.1.2* [Edit](#) [Delete](#)

The Regular Expression filter matches using a regular expression processor. Regular expressions are more powerful and flexible than wildcard expressions.

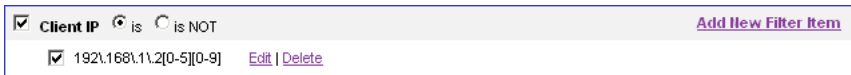
Consider the following regular expression:

Edit Client IP filter item

☐ Name ☐ Wildcard Expression ☒ Regular Expression

OK

Cancel



☒ **Client IP** ☒ is ☐ is NOT [Add New Filter Item](#)

☒ 192\168\1\2[0-5][0-9] [Edit](#) | [Delete](#)

This regular expression matches IP addresses from 192.168.1.200 to 192.168.1.255.

Once you have entered the search string and clicked **OK**, you can:

- Choose to turn this new filter on (or off if it is currently on) with the checkbox in the top left hand corner.
- Choose to use an 'is' or 'is NOT' match by selecting the radio buttons at the top.
- Add more filters.

The Global Filters Editor will save your filters even when they are not enabled, so you can come back the next time you look at this profile and enable these filters again, without having to enter them again. You can add multiple filters and enable some, none or all of them for any view. You can edit the current filters by clicking the edit link and delete them with the delete link.

Once complete, click **Save and Close** to save the filter. Once you have Global Filters active, you will see the active filters in the Report Bar and the Filter checkbox becomes active. You can enable and disable all filters from here.

Note — To enable the Global Filters from the report bar, the filters will need to be active from within the Global Filter Editor.

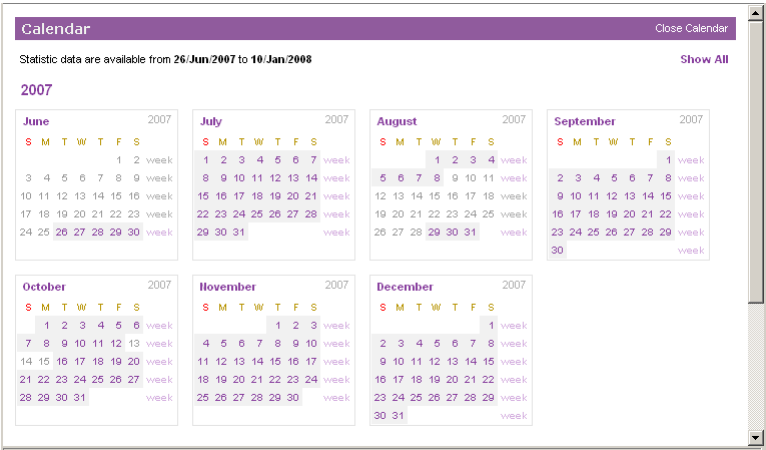
Date/Time Filters

Date/Time filters are controlled by The Calendar and The Date Range Selector. They remain active on all Reports until they are removed by Clicking the Show All link from within the Calendar.

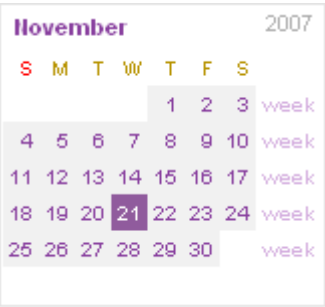
The Calendar

To access the Calendar, click the **Calendar** button in the Report Toolbar.

The Calendar shows the years, months, weeks and days during which there was traffic by displaying a clickable link for each day, week, month or year. All links that are not clickable have no data in the database. Clicking any day, week, month, or year adds Date/Time Filters to the reports for the selected period, thereby “zooming in” on that data.



Each day, week, month or year in the calendar that has data will be highlighted when you move the mouse over it. The following image is selecting one day, the 21st of November, 2007:



The following image is selecting the week of December 9th through the 15th:



The Calendar controls the date and time filtering in the report and once filtered, The Report Bar shows the time period that the report is displaying. The screenshot below shows an Overview report filtered by one week, December 9 to December 15, 2007.

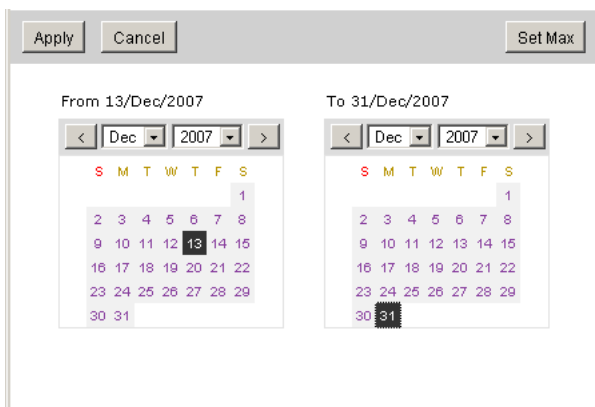
Overview		
<div> <div>1</div> <div>Statistics for 09 Dec/2007 - 15 Dec/2007, 7 days</div> <div> <input checked="" type="checkbox"/> Date Filter <input type="checkbox"/> Filter Refresh </div> </div>		
	All days	Average per day
Page Views	140,096	20,013.71
Requests	142,742	20,391.71
Size	2.64 G	385.62 M
Bandwidth Cost	\$0.20	\$0.03
Security Cost	\$21200.00	\$3028.57
Unique Src IPs	18	-

The Date/Time Filters can be removed by clicking the Show All link in the top right hand corner of the Calendar.

The Date Range Selector

The Date Range window is where you can select a range of days to use as the date filter. To access the Date Range window, click the **Date Range** button in the Report Toolbar.

Figure 4-2 Date Range Window



You can select any range by clicking on the dates each calendar. Use the calendar on the left to enter the start date, and the calendar on the right for the end date. You can change the month shown in each calendar using the arrow buttons or drop down menus. Click **Apply**, or use the **Set Max** button to select all available dates.

When you have selected a range in the Date Range window, all reports will show only information from that time period, until the date filter is changed. You can change the date filter by going into the Date Range window again. Or, you can remove the date filter by clicking **Show All** in the Calendar.

Zoom Filters

Zoom Filters are *temporary* filters that allow you to zoom into a report and see more detailed information. You activate a zoom filter by clicking on a table item. The Zoom Display (part of The Report Bar) shows what you have zoomed in on.

Note — Zoom Filters allow you to select one value to filter on. If you want to filter on multiple values, such as two MIME types, use Global Filters. For more information, see “Global Filters” on page 44.

For example, if you click on the item “Suspect/Threat URLs” in the Category Severity report, the report bar displays the Zoom filter as shown in Figure 4-3:

Figure 4-3 Zoomed Data



Zoom filters disappear when you click a new report.

The Zoom To Report menu shows the name of the report that will be displayed when you zoom. For example, after the user zoomed in on the Suspect/Threat URLs data in the Category Severity report (as shown in Figure 4-3 on page 49), the Zoom to Report menu appears.

This can be useful if you want to break down each item in a table by some other report. For example, Figure 4-4 shows traffic by client IP for the uncategorized URLs.

Figure 4-4 Data Zoomed in Further

FIELD: Client IP

1

Statistics for 08/Aug/2007 - 10/Jan/2008, 156 days

Date Filter

Filter

Refresh

Report is zoomed and shows data for

Severity: 1-Emergency

Category: Suspect/Threat URLs

Zoom to report >>

FIELD: Client IP

Row Numbers

Zoom Options

Export

Table Options

Row 1 - 3 of 3

Start row: 1

Number of rows

	Client IP	▼ Requests	Page Views	Unique Src IPs	Size
1	68.8.38.225	6	1	1	11.20 k
2	63.251.108.100	3	2	1	5.55 k
3	67.83.205.210	2	2	1	3.77 k
	Total	11	5	-	20.52 k

To get to Figure 4-4, choose “FIELD: Client IP” in the Zoom to Report menu from the report shown in Figure 4-3 on page 49.

You can change the Zoom menu view repeatedly, and filter each view from the Zoom to Report menu. You could also use Global Filters to add a filter to the entire report rather than using the Zoom Filters.

USING LOG FILTERS

Sawmill uses log filters to filter out data from your log source before populating the database. You might want to use log filters to:

- Selectively eliminate portions of your log data from the statistics.
- Convert values in log fields to a more meaningful value.

Log filters are written in Sawmill's configuration language. Log filters should not be confused with report filters that appear in reports. Log filters affect how the log data is processed, and report filters affect which parts of the database data are displayed. There are many reasons you might want to filter the log data, including:

- You may not be interested in seeing the hits on files of a particular type (for example, image files in web logs).
- You may not be interested in seeing the events from a particular host or domain (for example, web log hits from your own domain).
- You may not be interested in seeing hits which did not result in separate page views, such as 404 errors (file not found) or redirects.

The log filters included with Sawmill for IronPort perform the most common filtering. You may want to modify, disable, or remove those log filters. You may also want to create your own. For more information about the log filters included in Sawmill for IronPort, see "Sawmill for IronPort Log Filters" on page 7.

Note — Each profile type includes different log filters. For example, the Sec Ops profile type include log filters that affect the different malware related fields, and the HR profile type includes a log filter that only saves the server name into the database for each URL

How Log Filters Work

Log filters are arranged in a sequence, like a computer program, starting with the first filter and continuing up through the last filter. Each time Sawmill processes a log entry, it runs the filters in order, starting with the first one. Sawmill applies that filter to the log entry. The filter may accept the log entry by returning "done," in which case it is immediately selected for inclusion in the statistics. If a filter accepts an entry, the other filters are not run. Once a filter accepts, the acceptance is final. Alternately, the filter may reject the entry by returning "reject," in which case it is immediately discarded, without consulting any filters farther down the line. Finally, the filter may neither accept nor reject, but instead pass the entry on to another filter (by returning nothing). In this case, and only in this case, another filter is run.

In other words, every filter has complete power to pass or reject entries, provided the entries make their way to that filter. The first filter that accepts or rejects the entry ends the process, and the filtering is done for that entry. A filter gets to see an entry only when every filter before it in the sequence has neither accepted nor rejected that entry. So the first filter in the sequence is the most powerful, in the sense that it can accept or reject without consulting the

others. The second filter is used if the first has no opinion on whether the entry should be accepted or rejected.

Note — The last log filter in Sawmill for IronPort is called “Mark as event.” This Log Filter instructs Sawmill for IronPort to add the access log entry as a row of data in the database. Do not delete, disable, or modify this log filter.

Hits vs. Page Views

Sawmill for IronPort distinguishes between “hits” and “page views” for Web Security appliance access logs. A hit is one access to a web server, such as one request for a file which may not actually result in the transfer of a file, as in the case of a redirect or an error. A page view is an access to a page rather than to an image or a support file like a style sheet.

For some web sites and some types of analysis, image files, .class files, .css file, and other files are not as important as HTML pages—the important number is how many pages were accessed, not how many images were downloaded. For other sites and other types of analysis, all accesses are important.

By default, Sawmill for IronPort only tracks page views. It determines whether or not an event is a page view by the file type. To learn more about this log filter, go to Config page > Log Data > Log Filters and read the filter expression for the “Detect page views” and “Strip non-page-views” log filters.

The Log Filter Editor

The easiest way to create log filters is in the Log Filter Editor, in the Log Filters section of the Config. To access the Log Filters Editor, go to Config page > Log Data > Log Filters and then click the Edit link for one of the log filters.

Figure 4-5 shows the Log Filter Editor for the Rewrite URL log filter.

Figure 4-5 Log Filter Editor

Save and Close Cancel

Name: Rewrite URL ☒ Filter active

Comment Filter SortFilters

Filter type: Advanced expression syntax

```
url = field_url_scheme . '://' . field_url_server . '/';
```

[Show Examples](#)

Available **log fields** to be used in the filter expression:

- city
- country

The Log Filters Editor lets you create new filters using a user-friendly graphical interface, without having to write advanced filter expressions. However, some filtering operations cannot be performed without advanced filter expressions, so the Log Filter Editor also provides an option to enter an expression. This option is found in the Filter Type menu under the Filter tab.

Note — All log filters included in Sawmill for IronPort use advanced expressions to define the filtering action.

WORKING WITH USERS

The first time you run Sawmill through a web browser, it will prompt you to choose an administrative user name and password. This information specifies the first user and is stored in the users.cfg file in the LogAnalysisInfo directory. After that you can create additional users through the User Editor, see this section of the The Administrative Menu.

The administrative users(s) must be logged in to perform any tasks where security is an issue, for instance, when specifying which log data to process, or when creating a profile. If your administrator changes or you forget the name and password, you can reset it by removing the users.cfg file.

You can check the “save my password” box when you enter the password. If you do that, Sawmill will save your user name and password on the machine running the web browser, and will not ask you for the password if you are using that machine. Be careful though -- anyone else using that machine will also be able to get into Sawmill without a password.

SCHEDULING SAWMILL TASKS

Sawmill includes a built-in scheduler which you can use to schedule regular database builds, database updates, database expirations, off-line HTML page generation, emailed reports, and more. For example, you can use it to schedule a profile's database to be updated every day at midnight, so the statistics are never more than a day old. You could also schedule Sawmill to generate HTML pages from your statistics every week, so you have an off-line browseable HTML snapshot of your weekly statistics. The Scheduler can also update the database every day, and after that, send you the statistics by email.

A scheduled item consists of the following components:

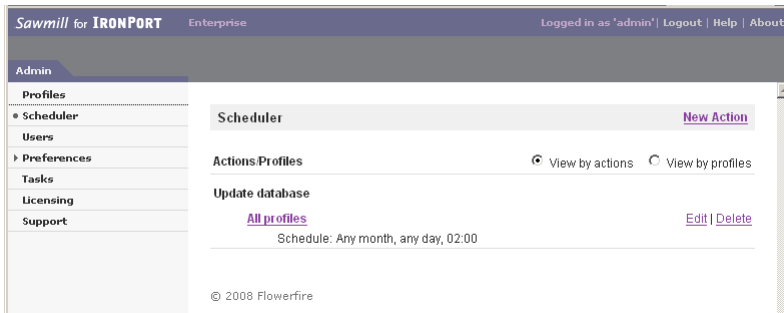
- **Action to perform.** You can choose one of the following actions to perform on a profile:
 - Generate report files.
 - Remove database data. For more information, see “Removing Database Data” on page 57.
 - Send report by email.
 - Update database.
- **Profile name.** Choose on which profile to perform the action.
- **Date and time.** You can choose the date and time down to the minute. You can choose “any” for any of the time fields so the scheduled task runs at each interval. For example, when you choose “any” for the Day field, the task runs everyday at the time and month specified.
- **Extra options (optional).** This field accepts any options available on the Sawmill Command Line. You can apply Report Filters by using the -f and -df options.

Sawmill performs the operation on the profile at the date and time specified in the scheduled task.

Note — By default, when you install Sawmill for IronPort, it includes a scheduled task to Update the database for all profiles at 02:00 every day.

Using the Scheduler

To access the Scheduler, click the “Admin” link in the upper right corner of either the Reports or Config page. Then once you are in the Administrative page, click the Scheduler link in the the Administrative menu.



From here, you can create, delete, and edit scheduled tasks. For example, you can create a task to update all your databases every night, or to send a report of the previous month by email on the first day of each month.

To create a new scheduled task, click **New Action**.

Figure 4-6 shows the Sawmill for IronPort Scheduler where you create new scheduled tasks.

Figure 4-6 Sawmill for IronPort Scheduler

Using Extra Options

Sawmill for IronPort calls itself from the command line to perform scheduled tasks. The extra options, if specified, are added to the end of the command line. This makes it possible to perform complex scheduled actions by overriding the default options on the command line. For example, without any extra options, sending email will send the default view by email to the default address, without any filters. But if the options are set to the following:

```
-ss smtp.example.com -rea IT@example.com -rca HRManager@example.com -rn
single_page_summary -f recentdays:30
```


then the Single-page summary of the past 30 days will be sent to HRManager@example.com, from IT@example.com, using the SMTP server at smtp.example.com. A list of available reports can be displayed by running Sawmill from the command line with the “-p profilename -a lr” options.

Debugging the Scheduler

Sawmill creates a file called TaskLog, in the LogAnalysisInfo directory, which contains a log of all scheduled tasks that have run, as well as logs of all other actions Sawmill has taken, such as view statistics or build database from the web interface. This log can be very helpful if you are trying to debug a problem with the Scheduler.

Removing Database Data

When Sawmill for IronPort removes database data, it performs the following actions:

- **Copies the main database table.** Sawmill for IronPort creates a copy of the main database table with some entries removed. If very little data is being removed, then the database temporarily takes up almost two times the disk space that it normally does. This step is relatively fast, especially if there is enough space on the storage device.
- **Rebuilds the indices and cross references.** After it finishes copying the main database table and removing data, Sawmill for IronPort rebuilds the table’s indices and cross references. These steps take about as long as they do during a database rebuild.

Removing database data can last about 60% - 70% as a rebuild database operation if only a small amount of data is being removed. The more data that is removed, the faster this operation is. Therefore, IronPort recommends to not remove data from the database more often than necessary. For example, do not remove data daily.

Depending on the size of the database and the amount of data to remove, it might be faster to rebuild the database without importing the older data you do not want instead of removing the old data from the current database.

Guidelines and Tips

This chapter contains the following information:

- “Optimizing Sawmill for IronPort” on page 60
- “Removing Data Before Processing” on page 66
- “Filtering Out Old Data” on page 67
- “Scheduling Guidelines” on page 68
- “AntiVirus Scanning” on page 69
- “Choosing Hardware Resources” on page 70
- “Rolling Over Access Logs” on page 71
- “Transferring Access Logs” on page 72
- “Working with Multiple Web Security Appliances” on page 73
- “Customizing URL Category Classifications in Compliance Reports” on page 74
- “Custom URL Categories” on page 77
- “Date Reports and Individual Fields” on page 78
- “Logging into the Sawmill Web Interface” on page 79
- “Viewing Particular Detailed Information” on page 81

OPTIMIZING SAWMILL FOR IRONPORT

Most organizations using the Web Security appliance produce very large access logs. And if the organization needs to analyze access logs from multiple Web Security appliances, the amount of data is even greater. Most of the default configuration options that come with Sawmill for IronPort are fine for relatively smaller data sets. However, if your organization processes a lot of data, consider making the changes described in this section.

Note — To increase performance during database creation, Sawmill for IronPort needs a lot of RAM. How much RAM is required depends on the data being processed and the options you configure on the Database Tuning and Log Processing pages.

Database Tuning Options

The Database Tuning page in Sawmill for IronPort includes a lot of features which can improve performance when it builds the main database table. Sawmill for IronPort builds the main database table during the following tasks:

- Profile creation
- Rebuild the database
- “Remove database data” scheduled task
- “Build database” scheduled task

This section includes the following subsections:

- “Database Tuning Options” on page 60
- “Optimizing the Database Tuning Options” on page 63

Database Tuning Options

Before describing some of the Database Tuning options, let us review how Sawmill for IronPort builds and stores the database.

Normally, when Sawmill for IronPort processes log data to build the main database table, it performs the following steps:

1. Builds the main database table.
2. Builds the indices of the main database table.
3. Builds the cross reference tables of the main database table.

However, you can configure Sawmill for IronPort to perform some of these steps simultaneously. Performing these steps simultaneously can increase performance during database creation, but it also requires a lot of RAM to do so. If the machine does not have enough RAM, do not configure Sawmill for IronPort to perform these steps simultaneously.

The main database table is stored as separate segments on the disk. The size of each segment is configurable in the Profile’s configuration. Each database table segment has its own index. The larger the dataset to store in the database, the greater the number of database

table segments and indices. When the main database table is stored in a very large number of segments, the index for each segment gets proportionately much larger. This can increase the amount of hard disk space required for the main database and the amount of time required to build the database. For example, if you have 200 GB of data to store, the database size might be four times that when the segment size is left at the default value. Therefore, for very large datasets, you should increase the size of each segment to improve performance and decrease the amount of space the database (especially the indices) uses on disk.

Figure 5-1 shows the Database Tuning page where you can optimize the database settings.

Figure 5-1 Database Tuning Page

Database Tuning

Edit Database Tuning

Initial size of database table:	4096
Expansion factor for database table:	2
Surplus factor for database table:	5
Maximum main table segment size:	100 MB
Maximum cross-reference table segment size:	100 MB
List cache size:	100 MB
Maximum main table segment size to merge:	10 MB
Maximum xref segment size to merge:	10 MB
<input checked="" type="checkbox"/> Build all indices simultaneously	
<input checked="" type="checkbox"/> Build indices during log processing	
<input checked="" type="checkbox"/> Build all cross-reference tables simultaneously	
<input checked="" type="checkbox"/> Build cross-reference tables and indices simultaneously	
<input checked="" type="checkbox"/> Build cross-reference tables during log processing	
<input checked="" type="checkbox"/> Build cross-reference tables in threads	
<input checked="" type="checkbox"/> Build indices in threads	
<input checked="" type="checkbox"/> Build indices in memory	

Table 5-1 on page 61 describes the options you can configure on the Database Tuning page.

Table 5-1 Database Tuning Page Options

Option	Description
Maximum main table segment size	<p>The amount of space on the hard disk taken up by each database table segment. Each processor on the machine requires this much RAM to build the database.</p> <p>IronPort recommends increasing this value to the amount of free memory (RAM) divided by the number of processors. For example, if the machine has 2 GB of free memory and a single processor, enter 2 GB. If the machine has 8 GB of free memory and two processors, enter 4 GB.</p>

Table 5-1 Database Tuning Page Options (Continued)

Option	Description
Maximum main table segment size to merge	IronPort recommends setting this value to 10% of the "Maximum main table segment size" value.
Build all indices simultaneously	Indicates that Sawmill for IronPort should build all indices at the same time instead of sequentially. Enabling this option can increase database creation performance, but also requires more RAM.
Build indices during log processing	Indicates that Sawmill for IronPort should build all indices when it processes the access logs to build the main database table. Enabling this option can increase database creation performance, but also requires more RAM.
Build all cross-reference tables simultaneously	Indicates that Sawmill for IronPort should build all cross reference tables at the same time instead of sequentially. Enabling this option can increase database creation performance, but also requires more RAM.
Build cross-reference tables and indices simultaneously	Indicates that Sawmill for IronPort should build all cross reference tables at the same time as building all indices instead of sequentially. Enabling this option can increase database creation performance, but also requires more RAM.
Build cross-reference tables during log processing	Indicates that Sawmill for IronPort should build all cross reference tables when it processes the access logs to build the main database table. Enabling this option can increase database creation performance, but also requires more RAM.
Build cross-reference tables in threads	Indicates where Sawmill for IronPort builds the cross reference tables when it is configured to use multiple log processing threads (one for each processor). When Sawmill for IronPort is configured to use multiple log processing threads: <ul style="list-style-type: none"> • Option is enabled. Each process builds cross reference tables for the table segments in that process. • Option is disabled. The main process builds the cross reference tables for the main database table after the table segments for each process have been merged into the main table.

Table 5-1 Database Tuning Page Options (Continued)

Option	Description
Build indices in threads	Indicates where Sawmill for IronPort builds the indices when it is configured to use multiple log processing threads (one for each processor). When Sawmill for IronPort is configured to use multiple log processing threads: <ul style="list-style-type: none">• Option is enabled. Each process builds indices for the table segments in that process.• Option is disabled. The main process builds the indices for the main database table after the table segments for each process have been merged into the main table.
Build indices in memory	Indicates whether indices are managed as buffers allocated in memory or as memory mapped files on disk. Building indices in memory requires more RAM, but is typically faster.

Optimizing the Database Tuning Options

To optimize the Database Tuning options:

1. Navigate to the Config page for the Profile you want to edit.
2. In the Reports Menu, expand the Database category, and click Database Tuning.
3. Click **Edit Database Tuning**.
4. In the “Maximum main table segment size” field, enter the amount of free memory available on the machine and divide it by the number of processors (threads).

For example, if the machine has 2 GB of free memory and a single processor, enter 2 GB. If the machine has 8 GB of free memory and two processors, enter 4 GB.
5. In the “Maximum main table segment size to merge” field, enter a number that is 10% of the amount you entered in step 4.

For example, if you entered 2 GB in step 4, enter 200 MB in this field.
6. Enable all check boxes.

Note — When you enable all check boxes here, Sawmill for IronPort builds indices and cross reference tables at the same time it builds the main database table. This requires a lot of RAM. If the machine does not have a lot of RAM, leave the default values for all check boxes.
7. Click **Save and Close**.

Log Processing Options

You can optimize Sawmill for IronPort when it processes access log files. You do that on the Log Data > Log Processing page.

Figure 5-2 shows the Log Processing page.

Figure 5-2 Log Processing Page

Log Processing

[Edit Log Processing](#)

Date offset:	0 hours
Log processing threads:	0
Thread data block size:	1 MB
Log reading block size:	100 KB
Log entry pool size:	1000
<input type="checkbox"/> Allow empty log source	
<input type="checkbox"/> Skip processed files on update by pathname	
<input type="checkbox"/> Skip most recent file	
<input checked="" type="checkbox"/> Look up location with GeoIP	
<input type="checkbox"/> Convert log data charset	

The most common settings to change on this page include the following fields:

- **Log processing threads.** When the machine hosting Sawmill for IronPort has multiple processors, you can configure Sawmill for IronPort to take advantage of the extra processing power. For more information, see “Working with Multiple CPUs” on page 64.
- **Skip processed files on update by pathname.** Sawmill for IronPort skips files which have already been processed during a database update or add operation. It considers the log file pathname to determine which files to skip. Whether or not you should enable this setting depends on where the access logs are stored:
 - **On the Web Security appliance.** Do not enable this option. The access log file is changing on the appliance, and this option should not be enabled when the log source is changing.
 - **On an FTP server.** Enable this option to increase performance.

Working with Multiple CPUs

If the machine hosting Sawmill for IronPort has multiple CPUs or CPU cores, consider configuring *each profile* to take advantage of each CPU or CPU core. To do this, go to Config > Log Data > Log Processing. In the “Log processing threads” field, enter the number of CPUs or CPU cores.

For one CPU or core, leave the default value of zero (0). Entering “1” to indicate one CPU or core decreases performance.

Note — Changing the number of processors impacts the amount of RAM required. The machine hosting Sawmill for IronPort must have enough memory to cover the number of processing threads configured multiplied by the “Maximum main table segment size” property. For more information, see “Log Processing Options” on page 64.

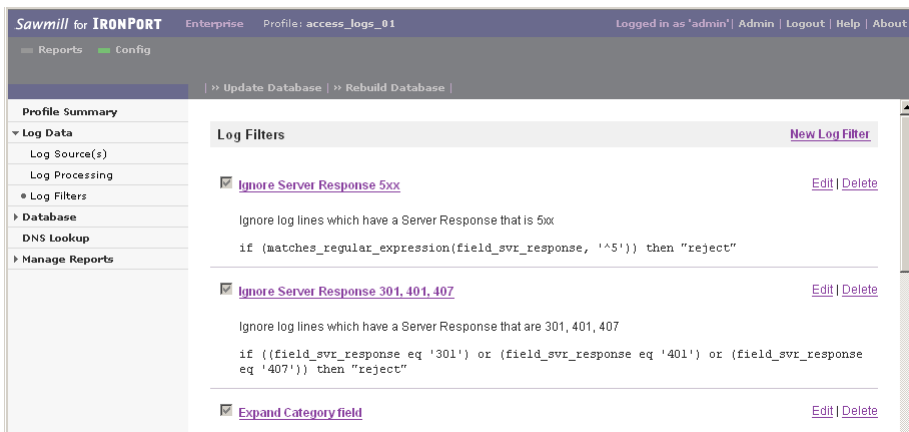
REMOVING DATA BEFORE PROCESSING

To increase performance when building or updating the database, consider removing unnecessary rows or unnecessary fields (columns) from the access logs before processing the log data into Sawmill for IronPort. Delete rows and fields by creating or editing log filters. Access the Log Filters page under the Log Data category from the Config page.

By default, Sawmill for IronPort already removes some rows from the access logs using Log Filters. For example, it removes all rows where the server responded with a 5xx response, such as “504 Gateway Timeout.” It does this with the “Ignore Server Response 5xx” log filter.

Figure 5-3 shows the “Ignore Server Response 5xx” log filter.

Figure 5-3 Using Log Filters to Remove Unnecessary Data



You can create new log filters to remove entire access log entries, like the “Ignore Server Response 5xx” log filter, or to remove a field from all access log entries. For example, if your organization does not need to know which policy groups were assigned to a transaction, you can create a log filter that removes that data before Sawmill for IronPort imports the data to create the database.

For more information on working with log filters, see “Using Log Filters” on page 51.

Note — Never delete, edit, or disable the Log Filter called “Mark as event.” This Log Filter instructs Sawmill for IronPort to add the access log entry as a row of data in the database. If you delete or disable this Log Filter, the database will contain no data.

FILTERING OUT OLD DATA

Sawmill for IronPort includes the “Ignore log lines older than 45 days” log filter, which filters out data older than 45 days when enabled for both the HR and Sec Ops profile types. You can also create a new filter specifying how many days to retain data.

WARNING: It is highly recommended that you use one of these log filters. Otherwise, the database will quickly grow to an infinite size.

Creating a New Filter

A new filter that limits how long Sawmill for IronPort retains data must be configured for each profile.

To create a new filter:

1. From the profile, navigate to Config > Log Data > Log Filters > New Log Filter.
2. Rename the filter to “Only the Last *<number of days to retain data>* Days.”

Note — Ensure that the number of days to retain data is within the sizing guidelines of Sawmill for IronPort and your server’s hardware. Use the Sizing Calculator provided on the Support Portal for basic guidelines.

3. Select the Filter tab and click New Condition.
4. Edit the conditions as follows, and then click **OK**:

Field	Value
Log field	Date/time
Operator	is older than number of days
Value	<i><number of days to retain data></i>

5. Click New Action.
6. Select “Reject log entry” as the action and click **OK**.
7. Select the Sort Filters tab.
8. Select the new filter “Only the Last *<number of days to retain data>* Days,” and click “Up [+]” to move the filter to the top of the list.
9. Click **Save and Close**.

SCHEDULING GUIDELINES

When you create a scheduled item, Sawmill for IronPort updates all profiles by default and processes them sequentially.

You can create scheduled tasks that build or update the database, and you can create a scheduled task that can remove data from the database. You must verify that any scheduled task to remove data does not run at the same time as a different scheduled item that builds or updates the database. If this happens, Sawmill for IronPort performance can get very slow or break, or you can get inconsistent data in the database.

For more information on scheduling, see “Scheduling Sawmill Tasks” on page 55.

ANTIVIRUS SCANNING

If the machine that runs Sawmill for IronPort also runs some antivirus software, then you need to configure the antivirus software to not interfere with the Sawmill database files. To do this, configure the antivirus software to *not scan* the LogAnalysisInfo directory in the Sawmill for IronPort installation directory. This can potentially increase Sawmill for IronPort performance and can prevent inconsistent behavior when the antivirus software tries to “cleanse” some of the Sawmill database files.

Note — If an antivirus program scans the Sawmill for IronPort installation directory, Sawmill for IronPort performance can be severely impacted and the main database might become corrupt. Verify no antivirus program scans this directory.

CHOOSING HARDWARE RESOURCES

Run Sawmill for IronPort on a dedicated machine when possible. Sawmill needs a lot of CPU power and RAM to analyze the log data, and a lot of hard drive space to store the data.

One way to speed up building or updating of the database is to use multiple processors. Sawmill for IronPort has the ability to split database builds across multiple processes, building a separate database with each processor from part of the dataset, and then merging the results. This can provide a significant increase in speed. A 65% increase in speed is typical when using two processors.

When the machine hosting Sawmill for IronPort has multiple processors, you must configure Sawmill for IronPort to use them. For more information, see “Log Processing Options” on page 64.

Increasing the speed of your processor is also a very good way to increase the speed of database building. Database builds are primarily CPU-bound, so disk speed, memory speed, and other factors are not as important as the speed of the processor.

Note — The IronPort Support Portal contains a sizing calculator to help estimate CPU, RAM, and disk space requirements based on the average number of access log lines per day and the number of days of log data your organization requires to process.

ROLLING OVER ACCESS LOGS

Configure the access logs so they automatically roll over by maximum time instead of maximum file size. To do this, make sure that the configured maximum file size is large enough so that the access log only rolls over by time and not file size. Leave the maximum file size parameter as the default value or the largest value allowed in the web interface.

TRANSFERRING ACCESS LOGS

When possible, use the FTP push method described in “Deployment Planning” on page 9. Configure each Web Security appliance to automatically transfer its access log to a remote server using FTP. The server to which it sends the access logs can be a dedicated file server or the machine hosting Sawmill for IronPort. This ensures that Sawmill for IronPort always has all data stored in the access logs.

WORKING WITH MULTIPLE WEB SECURITY APPLIANCES

If Sawmill for IronPort analyzes access logs from multiple Web Security appliances, place the access logs for each appliance in its own directory underneath a single directory. For example, if you have 3 appliances, you might have a directory called “WSA_access_logs” and underneath that directory have three directories called “WSA01,” “WSA02,” and “WSA03.”

```
WSA_access_logs
  WSA01
  WSA02
  WSA03
```

When you place each access log in a separate subdirectory, verify that the profile processes data in all subdirectories. You can specify this when you create a profile, or at any time by going to Profile > Log Data > Log Source and enabling the “Process subfolders (local folders only)” field.

CUSTOMIZING URL CATEGORY CLASSIFICATIONS IN COMPLIANCE REPORTS

Sawmill for IronPort uses four Compliance reports in the HR profile type: Productivity Loss, Legal Liability, Internet Tools, and Business Usage. Each report contains several URL categories. You can customize the classifications of these URL categories. For example, you might remove the Real Estate category from the Productivity Loss report.

Table 5-2 URL Categories in Compliance Reports

Compliance Report	URL Categories
Productivity Loss	Advertisements, Advertisements & Popups, Alcohol & Tobacco, Alcohol and Tobacco, Arts, Arts and Entertainment, Blogs & Forums, Cheating and Plagiarism, Computer Security, Cults, Dating, Dining and Drinking, Entertainment, Fashion & Beauty, File Transfer Services, Food & Dining, Freeware and Shareware, Games, Government, Government and Law, Hacking, Hobbies & Recreation, Job Search, Job Search & Career Development, Kids Sites, Lottery and Sweepstakes, Motor Vehicles, Mobile Phones, Motor Vehicles, Nature, News, Non-sexual Nudity, Online Communities, Online Storage and Backup, Online Trading, Paranormal and Occult, Peer File Transfer, Peer-to-Peer, Personals & Dating, Philanthropic & Professional Orgs., Photo Searches, Politics, Proxies & Translators, Real Estate, Religion, Ringtones/Mobile Phone Downloads, Safe for Kids, Sex Ed and Abortion, Sex Education, Shopping, Social Networking, Social Science, Society & Culture, Society and Culture, Software Updates, Spiritual Healing, Sports, Sports and Recreation, Streaming Media, Suspect/Threat URLs, Tattoos, Transportation
Legal Liability	Adult, Adult/Sexually Explicit, Child Porn, Criminal Activity, Gambling, Hate Speech, Illegal Activities, Illegal Drugs, Intimate Apparel & Swimwear, Intolerance & Hate, Lingerie and Swimsuits, Porn, Tasteless & Offensive, Tasteless or Obscene, Violence, Weapons
Internet Tools	Education, Finance, Finance & Investment, Health & Medicine, Health and Nutrition, Hosting Sites, Internet Telephony, Search Engines, Search Engines and Portals, Travel, Web Hosting, Web Page Translation

Table 5-2 URL Categories in Compliance Reports

Compliance Report	URL Categories
Business Usage	<p>The Business Usage report divides all URL categories into four sets:</p> <p>Business Business, Business and Industry, Computers and Internet, Computing & Internet, Reference, Science and Technology</p> <p>Borderline Automatic Updating, Search Engines, Search Engines and Portals, Travel, Finance, Finance & Investment, Chat, Instant Messaging, Web-based Chat, Web-based Email</p> <p>Unknown Filter Avoidance, Infrastructure, URL Filtering Bypassed, Uncategorized URLs, Impossible (error generated)</p> <p>Personal All other categories</p>

To customize URL category classifications in Productivity Loss, Legal Liability, and Internet Tools reports:

1. On the Config page, navigate to Manage Reports > Reports/Reports Menu > Edit > Report Elements > Edit.
2. To remove a URL category from a report, delete the category string from the report element filter. For example, to remove the Real Estate category from the Productivity Loss report, delete `Real Estate` from the following filter:

```
((field_object_page matches_regexp 'Page') and (field_category matches_regexp 'Alcohol & Tobacco|Arts|Entertainment|Food & Dining|Games|Government|Hobbies & Recreation|Kids Sites|Motor Vehicles|News|Personals & Dating|Philanthropic & Professional Orgs.|Photo Searches|Politics|Real Estate|Religion|Sex Education|Shopping|Society & Culture|Sports'))
```

Note — Whenever you modify log filters, you must rebuild the affected databases (profiles). Click **Rebuild Database**.

To customize URL category classifications in the Business Usage report:

1. Locate the `ironport_unified_category_map.cfg` file. (For Windows, this file is located at `C:\Program Files\Sawmill for IronPort\LogAnalysisInfo\ironport_unified_category_map.cfg`.) This file contains a stanza for each abbreviated URL category.

2. Change the usage attribute in the stanza that you want to edit. For example, in the following Real Estate stanza, change the word `Personal` to `Business` if you want to classify the Real Estate URL category as business usage:

```
Real = {  
  full_name = "Real Estate"  
  usage = "Personal"  
  severity = "5-None"
```

CUSTOM URL CATEGORIES

The Web Security appliance truncates the custom URL category names in the access logs by following this logic: *C_<first four letters of the custom URL category name>*. For example, the custom URL category “Allow list” is logged as “C_Allo.”

When you use custom URL categories, do the following:

1. Make sure that the first four letters of each custom URL category name is unique.
2. Modify the IronPort plug-in (`ironport_sseries_accesslog__XSQUID_sec-ops-profile.cfg` for the Sec-Ops profile, and `ironport_sseries_accesslog__XSQUID_hr-profile.cfg` for the HR profile) as follows. In the log filter named `logfilter_expand_url_category`, delete the following phrase:

```
or starts_with(field_category, 'C_')
```

The code should now look as follows:

```
value = `if (contains(field_category, '.') or  
starts_with(field_category, 'IW_')) then (
```

Note — You can apply this modification only to the new profile.

3. Edit the category mapping file that Sawmill for IronPort uses to convert the abbreviated URL category names to the full name. Open the `<Sawmill_directory>\LogAnalysisInfo\ironport_unified_category_map.cfg` file in a text editor and add an entry for each custom URL category. For example, if you have a custom URL category called “Allow list,” add the following entry to the `ironport_unified_category_map.cfg` file:

```
C_Allo = {  
    full_name = "Allow list"  
    usage = "Unknown"  
    severity = "5-None"  
}
```

Note — If a custom URL category has no mapping, Sawmill for IronPort may crash during log parsing.

DATE REPORTS AND INDIVIDUAL FIELDS

Use the reports listed under the Date Reports and Individual Fields report categories to zoom in on the other main report categories: Security, Compliance, and Resource.

LOGGING INTO THE SAWMILL WEB INTERFACE

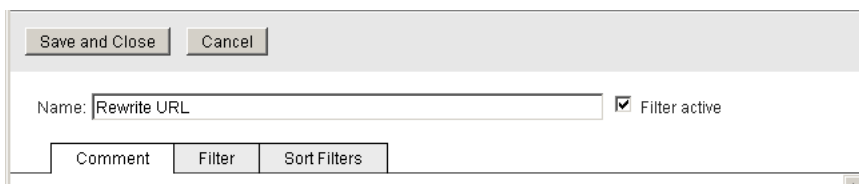
Verify that all users logging into the Sawmill for IronPort web interface each use a unique user name. When two users are logged into the web interface with the same user name concurrently, one person's work might cause report errors or interfere with the other person's filters.

SHOWING FULL URLS

In an effort to balance performance and reporting, detail the full URL is omitted by default. If reporting the full URL is required, you can disable the “Rewrite URL” log filter.

WARNING: Disabling the “Rewrite URL” log filter can cause performance problems and may prevent Sawmill from functioning properly.

1. Go to Config page > Log Data > Log Filters for the profile you want to edit.
2. Scroll down and click **Edit** for the “Rewrite URL” log filter.



The screenshot shows a configuration dialog box for the 'Rewrite URL' log filter. At the top, there are two buttons: 'Save and Close' and 'Cancel'. Below these, the 'Name' field is set to 'Rewrite URL'. To the right of the name field is a checked checkbox labeled 'Filter active'. At the bottom of the dialog, there are three tabs: 'Comment', 'Filter', and 'Sort Filters'. The 'Filter' tab is currently selected.

3. Disable the Filter Active check box, and click **Save and Close**.

VIEWING PARTICULAR DETAILED INFORMATION

This section includes instructions on how to access and view particular detailed information stored in the access logs by zooming in on data in a Sec Ops Profile.

Website Activity on a Per Client Basis, Highlighting Top Websites

1. In the Profile, choose the Individual Fields > FIELD: SourceID report.
2. In the report, click the client you want to view.
3. On the Overview report page in the Zoom To Report menu, select the FIELD: URL: Server report.

Note — You can highlight different categories for the client website activity by choosing a different field level report in step 3. For example, you can view website activity per client highlighting particular file extensions downloaded by choosing the FIELD: File Extension report in step 3.

URL Category Details by Website

1. In the Profile, choose the Resource > Top URL Categories report.
2. In the report, click the URL Category you want to view.
3. On the Overview report page in the Zoom To Report menu, select the FIELD: URL: Server report.

List All Users for a Particular Website or URL Category

1. In the Profile, choose the Resource > Top Web Sites report or the Resource > Top URL Categories report.
2. In the report, click the website or URL category you want to view.
3. On the Overview report page in the Zoom To Report menu, select the FIELD: SourceID report.

List All Websites Visited by Anyone

In the Profile, choose the Resource > Top Web Sites report.

View Client Malware Risk Detailed by Malware

In the Profile, choose the Security > MalwareID / SourceID report.

View Detailed Information About Which Script or File was Blocked

1. In the Profile, choose the Individual Fields > FIELD: Action report.
2. In the report, click the TCP_DENIED link.
3. On the Overview report page in the Zoom To Report menu, select the FIELD: URL: Server report.

Note — Sawmill for IronPort does not import any script or file name information.

View Which Content Was Blocked

1. In the Profile, choose the Individual Fields > FIELD: Action report.
2. In the report, click the TCP_DENIED link.
3. On the Overview report page in the Zoom To Report menu, select the FIELD: MIME Type report.

BACKUP AND RECOVERY

Backing Up Your Sawmill Profile:

`<Sawmill_installation_directory>\LogAnalysisInfo\profiles\profile_name.cfg`

Backing Up Your Sawmill Database:

`<Sawmill_installation_directory>\LogAnalysisInfo\Databases\profile_name\`

Note — The Databases folder is the largest subdirectory in the LogAnalysisInfo directory. If you choose to back up the Database directory, you might want to back up the entire LogAnalysisInfo directory.

Resetting the Administrative Password

Question: I forgot the password I chose for Sawmill when I first installed it. How can I reset it?

Short Answer: Delete the users.cfg file in the LogAnalysisInfo directory.

Long Answer: For security reasons, Sawmill requires an administrative user name and password whenever you use it, otherwise, anyone could use it to access your computer, since Sawmill is normally accessible by anyone on your network. You choose this user name and password when you first run Sawmill, and it asks you for it whenever you run it again. If you forget the user name or password you originally chose, you can reset your password by deleting the users.cfg file, which is in the LogAnalysisInfo directory of the Sawmill installation directory.

Note — This will delete all users from Sawmill. If you just want to delete the administrator while leaving the other users intact, you can edit users.cfg with a text editor and delete that administrator's group from the file. Once you have deleted users.cfg, access Sawmill again through a web browser, and you will be prompted to choose a new administrative user name and password.