Cisco Service Control Solution Guide

Cisco Service Control Business Intelligence Solution Guide, Release 3.6.x

Revised: November 8, 2010, OL-23868-01

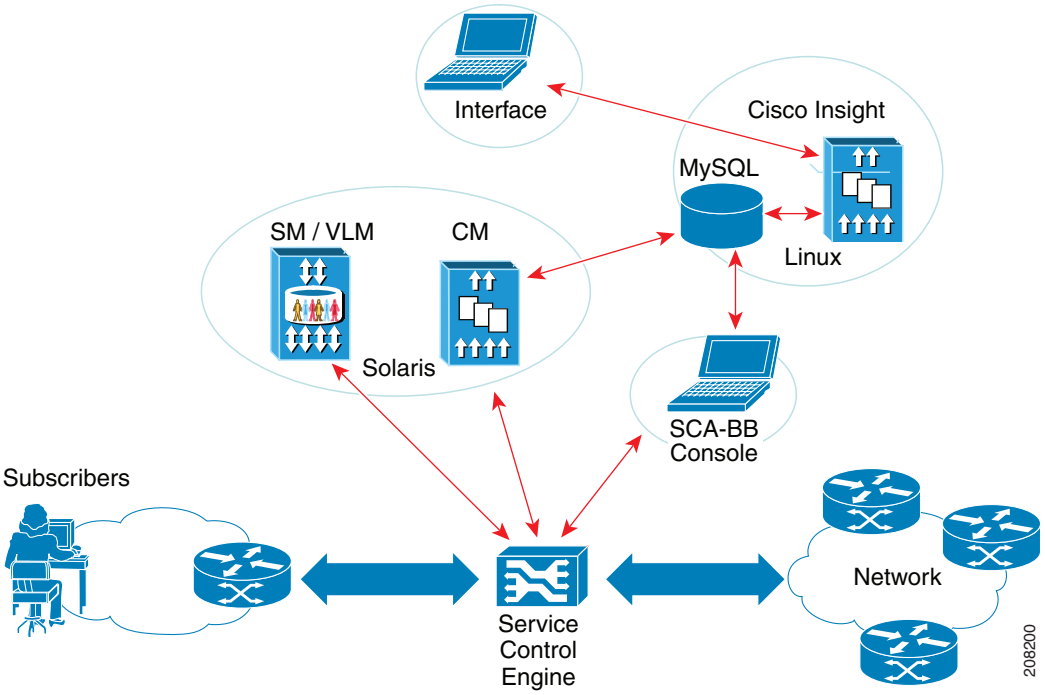
Note This document supports all 3.6.x releases.

1 Overview

The Cisco Service Control Business Intelligence (BI) solution provides the enhanced analysis and reporting of subscriber network traffic. The infrastructure of the Cisco Service Control BI solution consists of features that enable the user to analyze the behavior of subscribers within the network. For example, the traffic analysis can include volume consumption, access patterns, top content providers, and usage trends. The Cisco Service Control BI solution offers tools for trend analysis, traffic comparison, and enhanced data reporting. It also provides ways to access this data over longer periods of time (up to 1 year).

Within the Cisco Service Control BI solution infrastructure, the Cisco Service Control Engine (SCE) sends Raw Data Records (RDR) to the Collection Manager (CM). The CM then uses software modules to perform various data aggregations and writes the data to an external database (for example, MySQL). Like the Cisco Service Control Application for Broadband (SCA-BB) Reporter, the Cisco Insight Reporter processes the data from the database and presents the data in web reports. Figure 1 illustrates the network topology of the Cisco Service Control BI solution.

Figure 1 Cisco Service Control BI Solution Network Topology



Related Publications

Use this guide in conjunction with the following Cisco documentation:

- Cisco Service Control Management Suite Collection Manager User Guide
- Cisco Service Control Application for Broadband User Guide
- Cisco Service Control Application for Broadband Reference Guide
- Cisco Service Control Application Reporter User Guide

2 Features

Within the Cisco Service Control BI solution infrastructure, there exists a large amount of data that needs to be collected, processed, and stored. Reports on this data need to be generated within a reasonable response time to meet user requirements. The need for reporting on both detailed and long-term trends can create challenges. To address these challenges, the Cisco Service Control BI solution provides these features:

CM Features

The CM uses several complementary features to provide a compromise between level of detail and long-term storage of the data:

Focus on Significant Data

The term ClickStream describes the actual Hypertext Transfer Protocol (HTTP) requests that a particular user triggers. The SCE filters out irrelevant URLs that secondary HTTP requests trigger. The ability to classify HTTP requests as belonging to the ClickStream of the user allows an accurate and effective extraction of the web browsing habits of the user. The ClickStream of the user contains significant data about the access pattern of the user. Because ClickStream events make up only 1 to 5 percent of the total amount of HTTP requests, it reduces the amount of data to be analyzed.

The user can set the volume threshold within the reports based on the ClickStream, and transactions for particular applications that exceed the threshold are filtered out. This significantly reduces the amount of data to be processed without compromising accuracy.

Database Table Partitioning

The CM partitions records of the same type into separate tables within the database. These partitions are calculated according to time-stamp ranges. A rolling window mechanism is then used to periodically delete the oldest partition.

String Removal

Because strings require storage space and have limited use in data mining, eliminating such strings frees up space for more storage. The Java Database Connectivity (JDBC) adapter of the CM can replace strings with empty strings based on the *dbtables.xml* configuration file. Eliminating strings results in a smaller overall database.

To modify a string field in the configuration file, use the <*options*> subtag in the <*field*> tag to overwrite the string with an empty value.

The following reports become unusable if *INFO_String* is removed:

- Top Email Account Owners.xml
- Top Email recipients.xml
- Top Email senders.xml
- Top Newsgroups.xml
- Top Subscriber To Newsgroup.xml
- Top Peer-to-Peer (P2P) File Extensions.xml
- Top Session Initiation Protocol (SIP) Domains.xml

The following reports become unusable if *ACCESS_String* is removed:

- Real-Time Streaming Protocol (RTSP) Host Distribution by Subscriber Packages.xml
- Top HTTP Streaming Hosts.xml
- Top RTSP Hosts.xml
- Top Web Hosts.xml
- Web Host Distribution by Subscriber Packages.xml

Aggregation of Usage Data

The aggregation process is run per table as a database stored procedure. Aggregation of usage data significantly reduces memory requirements by freeing space previously used to store more granular data points. Aggregation is disabled by default.

The xUR (where x is a particular type of Usage RDR) table records are displayed over time. With little effect to the overall accuracy of the report, usage data can be aggregated for the following timeframes:

- No aggregation for the first day
- 15-minute aggregations up to the 3-month mark
- 1-hour aggregations up to the 1-year mark
- 1-day aggregations if over 1 year

Adaptive Frequency Counts

Long-term trend reports are based on frequency counts, that is, how frequently a host was accessed. Because performing frequency counts on an infinite data stream could be a complex computational task, the CM uses an adaptive frequency count method where data for each of the most frequent events is aggregated and then stored. Aggregation periods are configured by hour, day, or week.

SCE Features

SCE data classification features include:

- Signature support—Support for over 700 application signatures.
- Zero-day detection—Heuristic approaches for classification of application categories (VoIP, P2P, gaming).
- Protocol packs—Updates the latest protocol signatures to the Signature Engine every 2 months.
- External signature editor—Signature utility for creating L7 classification.

When a flow does not match a protocol signature, advanced classification mechanisms are used:

- Behavioral classification—Flows of certain application categories usually have a distinct behavioral pattern (sparse or dense, unidirectional or interactive).
- Classification based on recent history—Adjacent flows with similar source or destination are classified together because these flows usually belong to the same application.
- Multistage classification—Accurate classification might require several packets. Therefore, temporary classification is used when immediate policy decision is needed.

3 Reports

The Cisco Service Control BI solution includes enhanced reports that provide a comprehensive view of the network activity collected and processed by the Topper/Aggregator (TA) and Real-Time Aggregating (RAG) adapters in the CM:

- Bandwidth
- Capacity
- Subscriber demographics
- Server activity
- Traffic comparisons between regions or geographies, types of traffic, and so on
- Traffic trends

The real-time Subscriber Usage RDR (SUR) is periodically generated at user-configured time intervals for each subscriber that has real-time monitoring enabled. SURs are generated per subscriber and describe the traffic generated by that subscriber for a defined time interval.

The Transaction Usage RDR (TUR) is generated for each transaction, where a transaction is a single event detected in network traffic.

The real-time Flow Usage RDR (FUR) is periodically generated based not on time intervals but on open flow traffic. The content of the FUR is similar to the TUR but contains additional fields (flow ID and session ID).

A user can proactively drill down through the reports with the ability to display open subscriber flows, that is, display top services, then display subscribers per service, and then display detailed information about a specific subscriber. Figure 2 illustrates the data flow between components during a drill-down to flow-per-subscriber data.

The drill-down follows this data flow:

Step 1 Upon drill-down to a specific subscriber, Cisco Insight Reporter triggers a login operation to the Subscriber Manager (SM) via proprietary remote procedure call (PRPC) over the SM application programming interface (API), which sets the *monitor* property value.

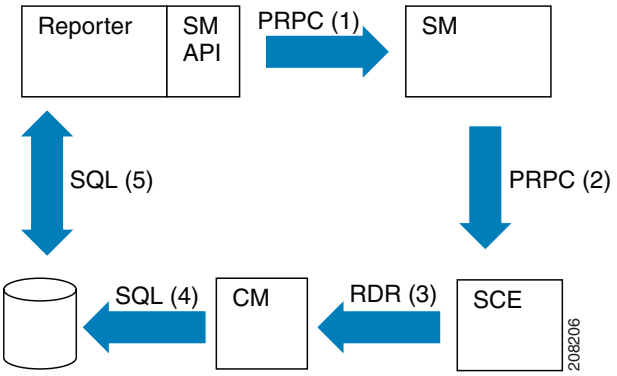
Step 2 The SM forwards this update to the SCE.

Step 3 The SCE generates flow-related RDRs for the subscriber and sends them to the CM.

Step 4 The CM JDBC adapter inserts the flow-related RDRs into the FUR table of the external database.

Step 5 Cisco Insight polls the FUR table in the external database and displays the flow information.

Figure 2 Data Flow of Drill-Down to Flow-Per-Subscriber



Video and HTTP Reports

The Cisco Service Control BI solution includes enhanced reports on Video and HTTP domains:

- Service-related reports—Video service distribution, top Flash video hosts
- Subscriber-related reports—Top video consumers, top browsing consumers
- Provider-related reports—Top video providers, top web hosts
- Trend reports—Changes over time

The SCE sends Video and HTTP TURs to the CM. The CM then routes these TURs to its RAG adapter. The RAG adapter aggregates these TURs to the domains at the first aggregation level and periodically populates the top domains into a new table in the database schema. A scheduled database server aggregation job periodically aggregates the first-level aggregation data into the second aggregation level. A reporter queries the database to generate the Video and HTTP reports.

Cisco and the Cisco Logo are trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1005R)

Any Internet Protocol (IP) addresses used in this document are not intended to be actual addresses. Any examples, command display output, and figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses in illustrative content is unintentional and coincidental.

© 2010 Cisco Systems, Inc. All rights reserved