Cisco Connected Analytics for Mobility (Wi-Fi) Overview

Cisco Connected Analytics for Mobility (Wi-Fi) (CAM) is a application built on top of Cisco Connected Streaming Analytics (CSA) (formerly known as Cisco Prime Analytics), which is a real-time big data analytics platform that allows you to connect to continuous streams of structured data and complete subsequent processing through derived streams to answer real-time business questions. The platform is made up of two main components—a Data Mediation server and a Data Warehouse server.

The following topics provide an overview to CAM features and architecture:

Features and Functions

CAM provides the following features and functions:

- Streaming data—CAM can process large streaming data volumes, which allows decisions to be made based upon current data trends and analysis.
- Continuous queries—To interpret continuous data streams, continuous queries are required. CAM continuous queries live for as long the client issuing the continuous query. New results are generated whenever new data arrives. The data is routed through the active query set and new results are published to downstream subscribers. Queries associated with derived stream queries, on the other hand, are active for as long as the system is on.
- Windows—CAM windows allow continuous data streams to be divided into segments so that results can be presented at meaningful points. Windows can be based on time intervals or by the number of arriving records.
- Streaming views—CAM allows you to create views of streaming data for higher-level real-time data analysis.
- SQL support—CAM supports SQL for real-time data stream queries. Supported SQL functions include stored procedures, user-defined functions, user-defined aggregates, joins, and other functions.
- Parallel processing—To handle continuous query performance demands, CAM uses parallel processing to maximum advantage and ensure performance degradation does not occur.

- Adaptability—The CAM continuous query engine is adaptive. Continuous queries can be added or removed without requiring a system restart.
- Support for data streams and data tables—CAM is a data stream engine built on top of the PostgreSQL database. It offers relational database functionality with the added functionality of stream processing .
- User Console—The CAM BI platform, based on Pentaho Enterprise Edition, allows you to create and manage reports and dashboards for stored and real-time data.

CAM Architecture Overview

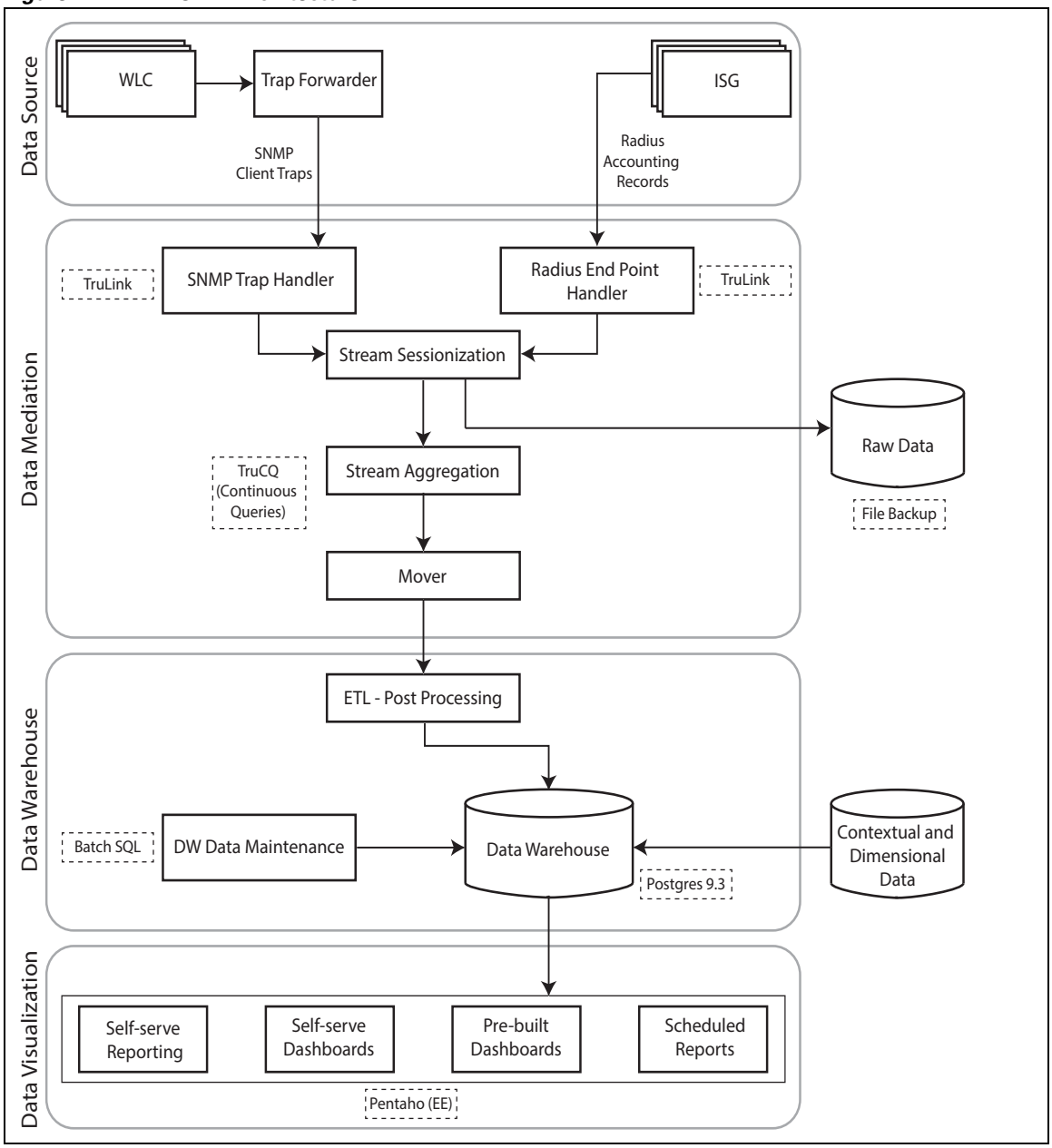
Note This section is written assuming you are familiar with the concepts of Cisco Connected Streaming Analytics, which CAM is built upon.

CAM is a typical reporting application and follows very similar architecture to other OLAP (OnLine Analytics Processing) applications. The key differentiator is that network elements in a Wi-Fi network, such as APs, WLCs and ISGs, serve as the high-volume streaming data sources. The data is mediated in real-time to update a data warehouse. The visualization follows a self-serve model, meaning that you have the control to build your own reports and dashboards. The three main areas of focus in this architecture include:

- Accuracy—CAM accurately translates machine-generated data stream into actionable insights, and no information is lost in the process.
- Scalability—CAM enables high-volume real-time data processing simultaneously across the network. It is built to scale linearly with increasing network demand.
- Flexibility—CAM offers predefined dashboards and reports, as well as a self-serve portal for ad-hoc reporting and custom dashboards. It also allows you to integrate custom data sources and provides long-term data retention.

The two main systems are Data Mediation (DM) and Data Warehouse (DW), as shown in Figure 1-1. The sections that follow provide further architectural details.

Figure 1-1 CAM Architecture



Data Mediation

The Data Mediation (DM) system consists of data handlers, stream processing, and a mover.

Data Handlers

The CAM system consumes two types of data sources: 1) SNMP client traps sent by WLCs and 2) RADIUS accounting records received from the ISGs that give information about user authentication into the Wi-Fi network.

SNMP Traps

The WLC generates SNMP client traps when clients associate and disassociate with APs. While the WLC generates many types of traps, CAM consumes four traps:

- ciscoLwappDot11ClientAssocTrap
- ciscoLwappDot11ClientDeAuthenticatedTrap
- ciscoLwappDot11ClientMovedToRunStateNewTrap
- ciscoLwappDot11ClientMobilityTrap

RADIUS Accounting Records

RADIUS accounting records from multiple ISGs are forwarded to CAM by configuring the ISGs.

CAM supports three types of RADIUS records:

- Start—Corresponds to a user authenticating on the network. This type of record can be either a MAC based authentication or a Web Portal based authentication. These two types of authentication need to be distinguished through the record attributes.
- Stop—Corresponds to user logging off the network.
- Interim Update—For long running sessions, interim records are sent updating the usage in the time period.

A TruLink handler for RADIUS acts as an end point that consumes accounting records. The handler will convert these accounting records into stream records pumped into a TruCQ raw stream.

Note The system relies on SNMP traps and AAA radius records, which use UDP as the transport layer. UDP is not reliable and may result in packet loss due to network configuration or issues. In such a scenario, the system may not be able to calculate the metrics very accurately in case there is significant loss. However, the system handles some combination of errors (e.g., missing association trap in a client session).

Stream Mediation

Stream mediation is done within TruCQ. The primary purposes of the stream processing include:

- The handlers join the multiple input raw streams (i.e., RADIUS and SNMP) to create a single stream.
- Most of the metrics and measures that are used in the system are based on user associated and authenticated sessions.

- The stream processing performs the necessary aggregates on the joint stream, and the results are saved to the disk in the form of the data-model fact table.

The stream mediation engine reduces the data volume from the raw streams into aggregated data and has to be sized appropriately to handle the load.

The raw joint stream created as part of the stream processing represents raw events that happened in the system, and as per application requirement, it needs to be saved into archive. This archive is periodically dumped into a backup file system and kept for a period of one year.

Mover

Aggregated data is saved into the archive on the TruCQ side in real-time. The mover transfers this data from the corresponding archive table to the Data Warehouse for ETL post-processing. The data flow is as follows:

TruCQ Side	Mover	Data Warehouse Side
client_ap_session_archive	→	client_ap_session_archive
ap_aggs_hourly_archive	→	ap_aggs_hourly_archive
client_usage_minute_archive	→	client_usage_minute

Data Warehouse

The Data Warehouse system is comprised of the database, ETL process, and BI platform for data visualization.

Database

The Postgresql 9.3 database is the Data Warehouse platform for persisting data. The data model is predefined and complete. The data model supports pre-aggregation of metrics and measures over often-requested dimensions. The data partition is built into the data model to improve query performance. The data model also supports custom metric definition based on existing measures and metrics. The key aspects for optimization for data model are accuracy, query performance and flexibility if new data dimensions are added.

ETL

Extract, Transform, and Load (ETL) is the process that moves data out of the stream processing, reformats the data, and cleans the data before persisting it to the data model in the Data Warehouse. This process is responsible for keeping the data model current.

ETL also interacts with any external data sources that are necessary to enrich the data stream. For example, contextual data is an external data source that includes account-specific information, billing information, or device-specific information you can provide to enrich the data model.

Dimension Data

Some of the dimensions in the system, such as the list of APs, might not be provided when you set up the application to run. There is a mechanism to update these dimensions offline through direct update to dimension tables, if customer decides to load data directly, and for dimensions that cannot be discovered from streaming data. There will be support for slowly changing dimensions, such as AP Location and Venue.

The mechanism is discussed in Contextual Data, page 1-6.

Contextual Data

Contextual information is provided in tabular format either as files or in a database accessible through standard means like JDBC. You provide the contextual data. Integration with custom contextual data sets will require a service engagement.

Data Visualization

The CAM business intelligence platform consists of a commercial application developed by Pentaho to create a web-based client that can generate dynamic dashboards and reports from continuous query data. The BI platform includes the following features:

- Default dashboards and reports
- Ability to build custom dashboards
- Ability to schedule reports
- Drag and drop reporting and data visualization capability, meaning you can pick dimensions and measures from the data model and create a report or chart
- Interactive reports and report publishing capability
- Ability to modify reports and publish through email and/or save in standard formats like PDF, CSV, and XLS

Security

CAM security functions include:

- Role-based password-protected access for multiple users
- Web-based user management

In addition, you can enable SSL for added security between client web browsers and the CAM BI platform.