International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V6I4P123
Eureka Vision Publication | Volume 6, Issue 4, 154-158, 2025

Original Article

ETL and TR-069 integrated into SaaS enabled Broadband Market Expansion

Amey Deshpande
Sr. Cloud Delivery Manager Calix Inc. McKinney, TX, USA.

Received On: 05/11/2025 ***Revised On: 27/11/2025*** ***Accepted On: 04/12/2025*** ***Published On: 13/12/2025***

Abstract - A major Telecom Cooperative in alliance with a Local service provider, are together a Broadband Service Provider (BSP) located in the Midwest. This BSP provides one of the best communications services at a fair price; employing modern technology and up-to-date services; offering people choices to be successful and enhance their quality of life. They believe online job opportunities and the expansion of local businesses allow their economies to thrive. A reliable internet is the bridge tying their current success to future growth. The schools in these towns rely on a range of online resources to create vibrant classrooms. Older students can connect to remote college classes, and those already in the workforce can use online resources to advance their careers. For all the reasons and more, the BSP had a mission to market advanced network access and managed services to their subscriber base, as well as expand and engage to prospective subscribers in current and neighboring towns. The mission statement was bringing powerful internet services to as many communities as possible at a very fast pace. This BSP spanned about 3,300 square miles, serving several thousand subscribers.

Keywords - Broadband Expansion, SaaS Telecom Solutions, ETL Integration, TR-069 Device Management, Digital Transformation in Telecom, Rural Broadband Connectivity, Managed Broadband Services

1. Introduction

With a BSP growth mission on the roster, the Telecom Cooperative partnered with the Local Service Provider, planned to extend the network another 90 square miles to significantly increase the number of people with access to critical digital resources. They planned to expand in 2 counties, growth that's vital to the BSP's success. These communities were outside the area traditionally served by the co-op, and adding these new members would provide a range of benefits. Not only would their network serve the residents of these communities, but it would also create benefits for all their members.

2. Existing Calix SaaS Product

The alliance functioned as a BSP in the Midwest, employing a Business Support System (BSS) provider to manage subscriber billing records. The access network infrastructure was based on Calix platforms, supported by corresponding hardware and monitoring software. This design enabled the BSP to deploy Calix Optical Network Terminals (ONTs) within the access network and connect Calix Residential Gateways or Routers as Customer Premises Equipment (CPE).

As part of this configuration, the BSP deployed Calix's Saas (Software as a Service) Support product that gave their customer service team comprehensive insights into each subscriber's connectivity, devices, alarms, operational status, and bandwidth usage. Using this product the BSP was able to effectively activate and manage services within its existing service areas. The CPE data was reported into cloud-based product via TR-069, a CPE WAN Management Protocol (CWMP). The CWMP provides support for a CPE to make available information that the ACS may use to monitor the CPE's status and performance statistics. It also defines a set of mechanisms that allow the CPE to actively notify the ACS of changes to its state [6].

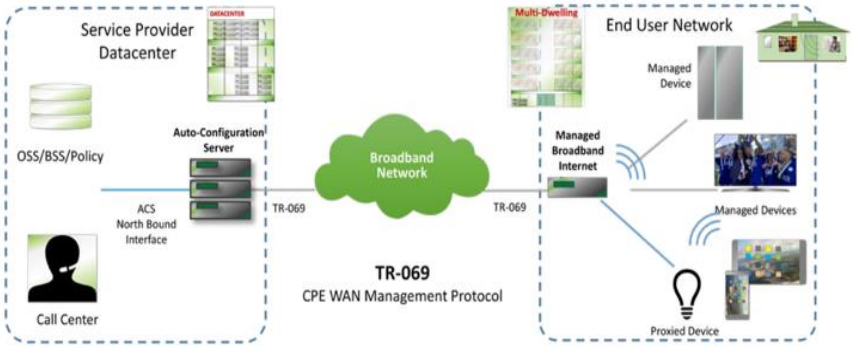

Figure 1. Tr-069 Protocol in the End-To-End Architecture [6]

The ACS is a server that resides in the network and manages devices in or at the subscriber premises. The CPE WAN Management Protocol may be used to manage both DSL B-NTs and other types of CPE, including stand-alone routers and LAN-side client devices. It is agnostic to the specific access medium utilized by the service provider, although it does depend on IP-layer connectivity having been established by the device [6]. Core or edge routers in the BSP's network were configured to transmit IP flow data and lastly the BSS supplied subscriber billing data. All the data streams converged to the Calix SaaS product and were ingested via an ETL (Extract Transform Load) data pipeline.

2.1. ETL Pipeline

An average billing data file in a CSV (Comma Separated Values) format exported from their BSS provider was not the cleanest collection of data. It included multiple rows of duplicate entries that encompassed all recorded details of a subscriber. The data typically included PII (Personal Identifiable Information) about the subscriber such as name, phone, email and service address. It contained network data such as exchange, market, technology, service codes, download and upload speeds, etc. It also contained device hardware details such as mac address, serial number and the make and model of the CPE. The data pipeline extracted data using a criterion on business logic and transformed values with filters based on user defined conditions. In other words, by using ETL. It is typically a three-step process.

- Extract raw data (in various formats like CSV, JSON, TXT, etc.) from different sources such as a database, files, and message queues.
- Transform the data using formatting, cleaning, normalization, and other techniques to convert it to a useful form for analysis and reporting.
- Load the data store, generally a warehouse or a columnar database. The goal of data pipeline was to ingest a raw .CSV file and write a deduplicated version of clean data collective of all attributes of each subscriber [1].

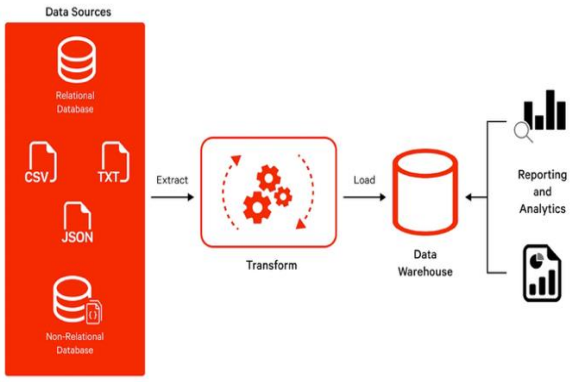

Figure 2. A Logical Overview of the Batch ETL Process [1]

3. A new Calix SaaS Product

The collaboration that supported the BSP's current utilization of the Calix Cloud product prompted their desire to expand their network. Achieving this expansion was challenging as it demanded robust marketing capabilities. The BSP was aiming to market and upsell to existing subscribers as well as attract new subscriber engagements. Acquiring new subscribers particularly those already subscribed to other broadband services presented a significant task at hand. Entering unfamiliar markets without insights into local service offerings or competitors further complicated the marketing team's efforts.

To address such challenges, the BSP chose to invest in the Calix SaaS Marketing product, designed to engage current and prospective subscribers. The billing data from the BSS and IP Netflows from the BSP edge routers were integrated into the Calix cloud database. Hence, adopting the Engagement Cloud was a logical first step.

3.1. Onboarding

The onboarding process for the Engagement Cloud product was led by the Calix Cloud Delivery Manager (CDM). The CDM initiated the project with the technical and marketing teams at Telecom Cooperative. This phase provided an overview of the product's features, outlined the RBAC (Role Based Access Control) roles to established for BSP employees. This framework ensures robust security and optimized data handling through granular access control mechanisms based on predefined roles [2]. The CDM highlighted key marketing concepts at project kickoff as essential for successful SaaS adoption.

The BSP shared a comprehensive list of all zip codes encompassing both current service areas and regions targeted for subscriber acquisition. Upon receiving the list, the CDM proceeded with updating the Calix backend databases to establish a foundational territory for the BSP. This pre-defined area would serve as the primary marketing region for pursuing new customer acquisitions and enhancing upsell and cross-sell opportunities among existing subscribers.

The Calix CDM acquired several technical details of the BSP's offerings such as the type of technology that delivered the services, the upload and download speeds on offer and the classification of the subscriber services being for commercial use or personal.

3.2. Data Integration

The Calix ETL pipeline utilized Regex with Python under the hood for data cleansing and text extraction from the raw data received from the BSS. Regex (Regular Expressions) is a powerful tool that is embedded inside Python which is a sequence of characters that define search patterns. Regex allows one to do string searching, string matching and manipulating strings based on the search pattern to do the operations like text extraction, data validation and search and replace functions [3].

The current integration presented a baseline of all subscriber records in the SaaS support product that could be supported solely on the base of having a Calix CPE. The marketing product however, needed an expanded subscriber base including to cover the full network served by the BSP.

Accurate representation of a subscriber record in the marketing SaaS product stood on the mapping of three essential data points:

- Subscriber record.
- IP flow traffic.
- IP endpoint, creating a direct association between subscribers and their respective IP flows.

The Calix CDM divided the subscribers into two categories, namely, managed and non-managed. Managed subscribers were those deploying a Calix CPE and the data mapping was straightforward – subscriber details originated from the BSS, IP flow data encompassed all BSP subscribers, and 1:1 mapping were established by the Calix CPE via the TR-069 protocol. This enabled seamless data correlation for managed subscribers in the marketing product.

In contrast, non-managed subscribers using third-party CPE plugged behind a Calix ONT could not communicate over the TR-069 protocol. Consequently, additional network infrastructure data was necessary to implement the three required mappings for such subscribers.

3.3. IP address Endpoint Mapping

The CDM added a milestone for this implementation of non-managed subscribers. It was necessary to obtain supplementary endpoint mapping data to establish a precise 1:1 correlation between each subscriber and their IP address. The Calix platforms deployed by this BSP included the Network Management application. Its primary role was to facilitate service activation on Calix ONTs in the access network. The CDM deep dived into the network where the Calix Optical Line Terminal (OLT) shelves were deployed in the network. Each OLT carried GPON cards in the shelf, to which fiber fed ONTs were connected and configured.

Table1. Record of the IP and ONT Serial Stored In the E7 OLT Switch.

Title	IP address
JohnSmith_987654_CXNK00A1B2C3	130.46.20.3

The locally connected ONTs were catalogued in the Calix OLT in a mac address table. This table recorded the subscriber's name as input by the BSP, the subscriber ID typically generated during service provisioning and the ONT serial number together forming the title. The table also linked this title to a corresponding IP address of the CPE, the only TCP/IP (Transmission Control Protocol/Internet Protocol) layer-3 device that was plugged behind the ONT [Table1]. When a device has an IP address that is on a remote network, the Ethernet frame cannot be sent directly to the destination device. Instead, the Ethernet frame is sent to the MAC address of the default gateway, the router. An Ethernet switch examines its MAC address table to make a forwarding decision for each frame [4].

The Network Management application came with a function to aggregate this information from all such OLTs in the network, consolidate it into a centralized server in the BSP's network. The server had the ability to export data to a designated endpoint via REST API. A REST API can be a gateway to transmit or share the data via HTTP to another application [5]. The CDM utilized this function to transfer data to cloud where the marketing SaaS product was hosted using a unique URL, username, and password. This accounted for half the data mapping required to complete 1:1 correlation between subscriber record and IP endpoint.

The other half was derived from the CSV billing file, which had to include ONT serial numbers as part of its routine uploads to Calix. This was not a requirement for the support product, however, Calix CDM collaborated closely with the BSS to ensure ONTs were accurately listed under a distinct parent header, while CPE serial numbers appeared under its separate header for CPE equipment. By synchronizing exports from the network management application and the ONT serials listed in the billing data, an enhanced layer of IP endpoint mapping was achieved, thereby enabling comprehensive correlation of non-managed subscribers within the marketing SaaS product.

After all the subscribers were set up to correlate to a unique IP endpoint, the next big step was to take the new data integration live for a daily ingest. The Calix CDM, in pursuit of a preliminary data Quality Analysis (QA) performed a dry run using the ETL data pipeline. The results were, however, unfavorable since the actual subscribers being written were about 11% less than the expected count.

4. Root Cause Analysis (RCA)

After sharing results with the BSP the CDM confirmed the result of the dry run was falling short by 11%.

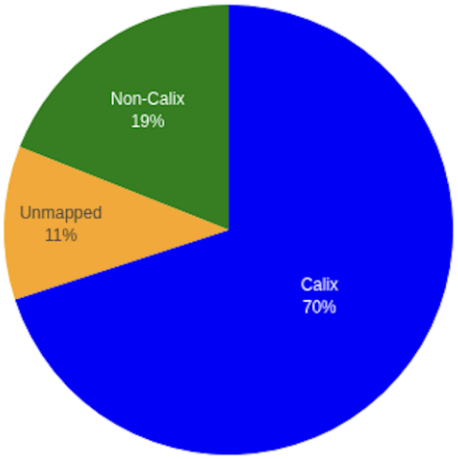

Figure 3. A Pie Chart Representing the Ratio of Mapped Subscribers and their Classification Based on Network Equipment

A root cause analysis (RCA) was necessary to determine if there were any unwanted filters in the Regex written in the data pipeline, or if any of the service codes were missing. Upon deep diving and carefully vetting the data pipeline no issues were found and was eliminated as being the root cause.

Next the CDM queried the service codes loaded in Calix cloud to ensure that they covered all service offerings of the BSP. The service codes were found to be sufficient and were eliminated as being the root cause.

Finally the BSS billing dataset was taken under scrutiny to locate anomalies in the manually entered data. For data validation the Calix CDM employed Python Pandas library to analyze the exported CSV files. Pandas library wrks with different types of files with extensions like xlsx, csv, pdf format [7].

By loading the data into a python Data Frame, Calix compared the expected patterns and value types in each column. Using pandas' functions such as "str.contains()" and "apply()" few anomalies were identified. This programmatic check revealed that a few hundred serial numbers and corresponding Calix CPEs that were typically in the CPE headers were found to be listed under the ONT headers. Post a collaborative session with the BSS, Calix established that this was not a system error but a manual entry in the incorrect field. This enabled targeted corrections before further processing of the data pipeline.

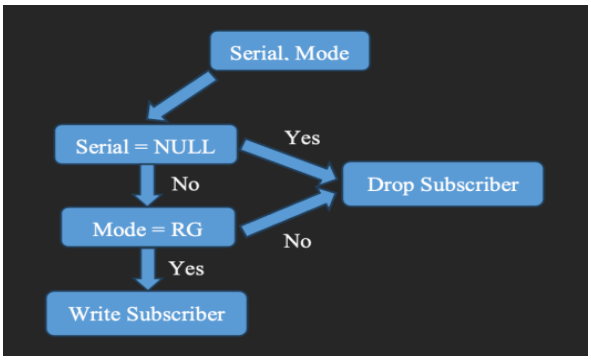

Figure 4. Graphical Flowchart of the ETL Process before RCA

As a fix, the regex used in ETL was relaxed to the point where a subscriber with no value in the CPE device header would still be written to the cloud as long all other criteria such as service codes were matched.

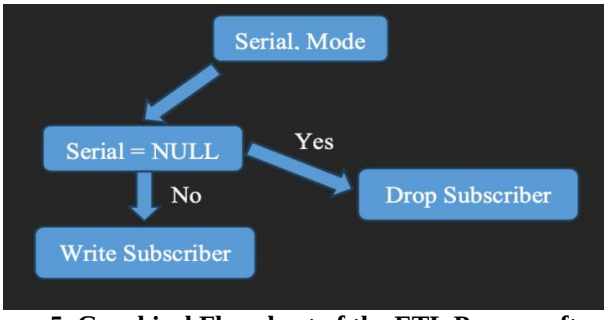

Figure 5. Graphical Flowchart of the ETL Process after RCA

This ensured that any manual errors made during provisioning of services and the device serial for an CPE accidentally entered in the ONT header will not be dropped but written to the cloud as a valid associated device.

The RCA was completed and the new correlation ratio for Calix grew up to 80%, with non-Calix staying at 19%. The remaining 1% were gaps and inaccuracies found in data that BSS entered in the billing system. Accurate record-keeping is essential for tracking service history and identifying recurring issues [8]. Systematic field practices and a one-time clean up on the BSP's end would bring the mapping correlation to a 100%.

5. Adoption and go to market

With RCA successfully completed the Calix CDM performed final dry run and the results were satisfactory to the BSP. The new ETL data pipeline was put in production and data integration now included not only subscriber data from the BSS, but IP Netflows from the BSP's network and additional IP endpoint data from the BSP's Network Management application. BSP adopted the marketing SaaS product with their business goals in sight. The product was leveraged to generate multiple marketing campaigns. The campaigns made for existing subscribers enabled them with better services based on their usage and requirements. As an example, a family of 5 living in a house were served with more bandwidth in their service, while a small business subscriber was provided multiple service drops and wireless repeater devices to accommodate all employees and visiting customers. The marketing was also targeted to create new offerings to new areas depending on the infrastructure laid down by the Telecom Alliance. Fiber fed areas were designed to offer the high-speed services, while more remote areas served via DSL or fixed wireless got better deals for their service. Riding on reachability and demographic data provided by the SaaS product based on the zip codes, the BSP was able to increase acquisition in new markets, while making sure there was no churn in their served areas.

6. Conclusion

The acquisition of new internet subscribers was a critical factor in the overall success of the initiative undertaken by the Cooperative. The return on investment (ROI) was primarily realized through a reduction in the per-member cost, as the expenses associated with maintaining the BSP network were distributed across a larger subscriber base. This expansion not only enhanced the average purchasing power of individual subscribers but also enabled access to services at more competitive rates. Although marketing efforts aimed at attracting new subscribers involved considerable risk, the investments ultimately proved beneficial by delivering increased broadband value to end users and strengthening the BSP's brand within the community. Subscribers were able to identify more closely with the brand, fostering loyalty and long-term engagement with the BSP.

References

[1] https://www.redpanda.com/guides/kafka-tutorial-streaming-etl

[2] Role-Based Access Control (RBAC) Enabled Secure and Efficient Data Processing Framework for IoT Networks Jaibir Singh, Suman Rani 2, Vipin Kumar

[3] “Regex in Action: Practical Examples for Python Programmers” By Prasanna Chandran Melnatami Krishnaram, Aug. 19, 25

[4] The MAC Address Table, Cisco Press https://www.ciscopress.com/articles/article.asp?p=3089352&seqNum=6

[5] A Secure Model based on Symmetric Encryption for REST API in Data Integration, Mohd Kamir Yosuf, Mustafa Man

[6] Broadband Forum, "TR-069 CPE WAN Management Protocol," Issue: Amendment 6 Corrigendum 1, June 2020. [Online]. Available: https://www.broadband-forum.org/pdfs/tr-069-1-6-1.pdf

[7] Data Analysis Using Pandas Library of Python (Acta Scientific, 2022): https://actascientific.com/ASCS/pdf/ASCS-04-0236.pdf

[8] Best Practices for WISP / ISP Field Technicians https://isprevolution.io/ensuring-stellar-customer-experiences-best-practices-for-wisp-isp-field-technicians/