SCTE·ISBE CABLE-TEC EXPO 2018 | ATLANTA, GA OCTOBER 22-25

2018 Fall Technical Forum SCTE·ISBE • NCTA • CABLELABS

Network Planning Automation Using Big Data

An Analytical Platform for Infrastructure Capital Planning

A Technical Paper prepared for SCTE•ISBE by

Ted Boone
Senior Director, Network Engineering
Cox Communications, Inc.
Atlanta, GA
404-269-4771
ted.boone@cox.com

Jignesh Patel
Principal Architect
Cox Communications, Inc.
Atlanta, GA
404-269-6898
jignesh.patel@cox.com

Rob Ames
Executive Director, Network Engineering
Cox Communications, Inc.
Atlanta, GA
404-269-5451
rob.ames@cox.com

Kyle Cooper, Cox Communications, Inc.

Chaitanya Vasamsetty, Cox Communications, Inc.

About Cox Communications

Cox Communications is a broadband communications and entertainment company, providing advanced digital video, Internet, telephone and home security and automation services over its own nationwide IP network. The third-largest U.S. cable company, Cox serves approximately 6 million residences and businesses. Cox Business is a facilities-based provider of voice, video and data solutions for commercial customers, and Cox Media is a full-service provider of national and local cable spot and digital media advertising. Cox is known for its pioneering efforts in broadband, voice and commercial services, industry-leading customer care and its outstanding workplaces. For eight years, Cox has been recognized as the top operator for women by Women in Cable Telecommunications; Cox has ranked among DiversityInc's Top 50 Companies for Diversity 13 times. More information about Cox Communications, a wholly owned subsidiary of Cox Enterprises, is available at www.cox.com and www.coxmedia.com.

Table of Contents

Table of Contents 3

Introduction 4

Network Planning Automation 4
Business Challenge 4
1. Big Data ecosystem 5
2. Data preparation 6
3. Analytics 6
4. Business Rules 7
5. Publication 7

Conclusion 7

Abbreviations and Definitions 8

List of Figures

Figure 1 - Network Planning using Big Data Infrastructure 6
Figure 2 - Network Transformation Visualization of select market 7

Introduction

Cox Communications is currently transforming its network infrastructure to support gigabit symmetrical speed offerings. This capital-intensive project needs a strategic plan providing prioritization using long term forecast accounting risk and macro-economic factors. Further, adding business constraints, budgetary restrictions, and other operational limitations we produce an optimal actionable plan.

The key business challenges with the current process of producing prioritized plan are scaling, repeatability, and traceability. Today, there are thousands of nodes in the network, but we will scale to a few hundred thousand nodes to meet network and customer growth. This increase made it difficult to continue with the prior manual process of building a strategic network plan. Data preparation relied on a manually extracting data from multiple sources to create a factual view of network bandwidth consumption. Business rules were applied in spreadsheets using macros.

Automation of this process had opportunities on several fronts, notably, providing consistency, repeat-ability, and modernization with the use of data science algorithms on an enterprise Big Data platform. Business requirements are made more transparent, and configurable, allowing planners to run multiple what-if scenarios faster to assist with the strategic decision process.

Network Planning Automation

Business Challenge

Last year, busy hour traffic on Cox's access network grew by almost a third. Downstream growth has softened a bit recently, dropping from near 50% in years past. Much of the drop coincided with re-encoding of video content by large over-the-top providers like Netflix since 2016. Upstream has been stable; only slightly off the 26% 5-year trend at 25% year over year growth.
Customer demand for higher internet speeds also continues to push higher fueled by service provider competition and newer technologies.

Most of this traffic is currently downstream. Currently 17 times as much data is delivered downstream during the busy hour as upstream. On the downstream side, traffic is predominantly over the top video sites like Netflix, YouTube, and Hulu. Future downstream growth drivers include the transition to higher definition video content like 4K/UHD which is about 3 times as much traffic as a high definition stream. We'll also be fueling some growth as we continue our own IPTV transition.

On the upstream side, traffic is uploading video to sites like Facebook and YouTube. Also, online storage sites like Google and Microsoft show up. Also, Cox Business subscribers drive a disproportionate amount of upstream traffic. (Not so much downstream). Upstream traffic growth drivers will include the growing Internet of Things streaming data to the cloud. With higher definition video content coming from a rapidly growing number of these devices. Ultra-High Def security cameras streaming continuously to the cloud is a trend we're watching.

Given all the network traffic growth drivers, Cox is in the midst of a network transformation that supports gigabit symmetrical speed offerings and manages long-term network congestion. This transformation will scale our network from thousands of nodes, to hundreds of thousands of nodes over the next 10 years. A capital-intensive project of this magnitude requires a comprehensive strategic planning capability that

balances demands from engineering and construction, manages congestion, delivers new product capabilities, incorporates geospatial, subscriber, business value, and budget data; and delivers an actionable plan that ensures we put the right investment in the right nodes at the right time to deliver the most value to the business, on budget.

Summarized Overview

Workflow

Challenges

• Storage shortage • Manual process	Improve using machine learning Apply AI technology	Manage rules in macros Spreadsheets unable to handle 20K nodes	Limited what-if analysis Unable to analyze sensitivity

Solution requirements

• Big Data platform • Real-time data • Telemetry from	Timeseries and Ensemble forecasting Optimization for prescriptive results	Use decision tables Fast updates to rules Multiple scenario runs More analysis time	Enterprise BI platform Multi user support Track forecast Realtime analysis

1. Big Data ecosystem

Cox's Enterprise Data Services team provides end-to-end application development and analytics infrastructure mostly using Hadoop ecosystem infrastructure.

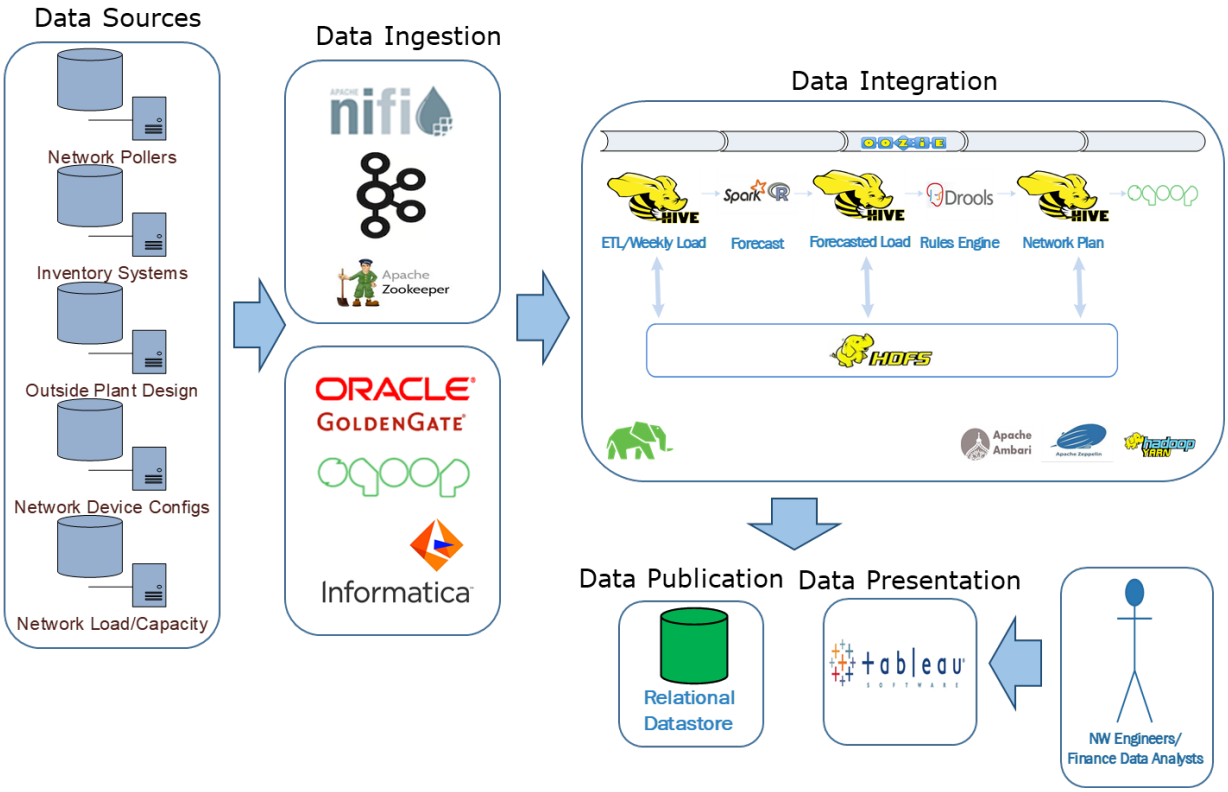

Figure 1 - Network Planning using Big Data Infrastructure

2. Data preparation

A network node-level capacity management and transformation plan requires several types of data such as including network topology, telemetry, physical, geographical, and subscriber data. Extracting, cleaning, preparing, and integrating these large data sets on a continuous basis can be the most time consuming and daunting task. The goal is to clean existing data for errors, combine data feeds from multiple systems, and create a single version of truth for analysis. Sqoop and HIVE are the primary Hadoop tools employed.

3. Analytics

An accurate node-level traffic load forecast is necessary to make savvy long-term business decisions about capacity management and technology transition. Accurate traffic forecasting must integrate long-term market level growth trends along with recent node-level dynamics. Combinations of exponential smoothing (Holt-Winters), ARIMA, and ensemble machine learning methods are compared, and combined, and the most accurate method selected.

Multivariant techniques, such as clustering, are employed with congestion, geospatial, subscriber, business value, and budget data to balance demands from engineering, outside plant construction, product and finance to ensure the right investment is made in the right nodes at the right time to deliver the most value to the business on budget.

Forecasting and clustering analytics automation is supported by Spark R on Apache Hadoop infrastructure.

4. Business Rules

Once the long-term node-level traffic load is determined, complex business rules engage to determine the optimal path for managing the technology and capacity evolution of the node. For example, spectrum updates, node-splits, N+0 build outs, and full-duplex DOCSIS steps must be qualified and sequenced.

Automated business rule application is enabled by the Drools core Business Rules Engine (BRE) using JBoss on the Apache Hadoop platform.

5. Publication

A capital-intensive project of this magnitude has numerous stakeholders, including engineering, outside plant construction, critical facilities, metro and backbone network, finance, field engineering, customer support, and executive steering. Data is transferred to data driven stakeholders via Sqoop. Data visualization is accomplished via Tableau.

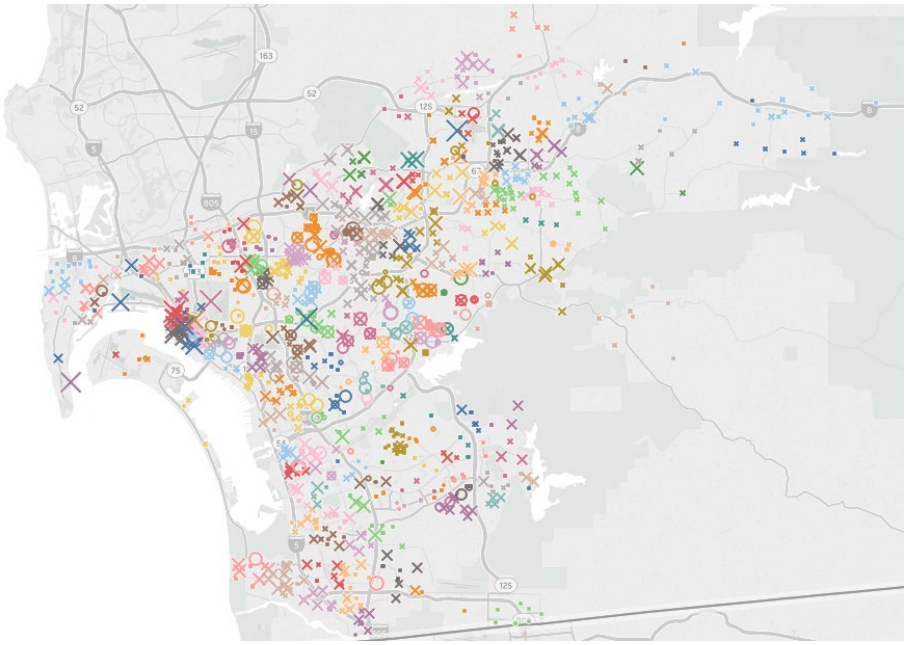

Figure 2 - Network Transformation Visualization of select market

Conclusion

This comprehensive end-to-end solution using a modern Big Data platform produces faster and repeatable results.

- Real-time input data processing and preparation combining multiple network topology and telemetry data sources (Hive, Spark on HDP Platform)
- Data Analysis (Forecasting and Clustering) using an advanced statistical programming language (Spark R on HDP)
- Enterprise business rule engine that can process declarative rule sets (JBoss Drools engine)
- Specialized data visualization dashboards to drill-down and assist detailed analysis (Tableau connected to Hive)

Abbreviations and Definitions

DOCSIS	Data Over Cable Service Interface Specification (DOCSIS) is a telecommunications standard used to provide Internet access via a cable modem. It is important because it is a key element in providing modem manufacturers and network service providers a common method for products to work together in a predictable manner.
Drools	Drools is a business rule management system (BRMS) with a forward and backward chaining inference-based rules engine, more correctly known as a production rule system, using an enhanced implementation of the Rete algorithm.
FDX	Full Duplex DOCSIS 3.1
HDP	Hortonworks Data Platform
HIVE	The Apache Hive ™ data warehouse software facilitates reading, writing, and managing large datasets residing in distributed storage using SQL. Structure can be projected onto data already in storage. A command line tool and JDBC driver are provided to connect users to Hive.
JBoss	Also, formerly known as JBoss Application Server, now known as WildFly, is an application server authored by JBoss, now developed by Red Hat. WildFly is written in Java and implements the Java Platform, Enterprise Edition specification. It runs on multiple platforms.
N+0	Node-plus-zero (N+0) architecture. N+0 means there are no amplifiers required between a node and a subscriber household
Spark	Apache Spark is a unified analytics engine for big data processing, with built-in modules for streaming, SQL, machine learning and graph processing.
Spark R	SparkR is an R package that provides a lightweight front end for using Apache Spark from R, supporting large-scale analytics on Hortonworks Data Platform (HDP) from the R language and environment.
Sqoop	Apache Sqoop (TM) is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases.
SCTE	Society of Cable Telecommunications Engineers