What It Takes to Automate Operations at Scale

Coupling Strategic Growth Analytics with Automated Methods for Real-Time Scalable Network Planning

Dr. Sung-eun Kim
Sr. Network Planning Engineer
Cox Communications, Inc.
Atlanta GA, 30328
sung-eun.kim@cox.com

Richard Brown
Sr. Manager Network Engineering
Cox Communications, Inc.
Atlanta GA, 30328
richard.brown@cox.com

Table of Contents

List of Figures

1. Introduction

Understanding the traffic consumption trend is one of the essential elements in network capacity planning. We have seen an exponential traffic growth for long periods and a huge step function increase during the COVID-19 pandemic period. In the first part of the paper, we present the historical traffic growth over the past years, especially emphasizing the impact of COVID shelter in place in various ways.

The next part of the paper, we will examine the ongoing analysis, innovation, strategic direction, and how all those factors and components materialize as actionable plans using our patent-pending planning tools. We will also examine how our scalable suite of tools was a key driver in staying in front of COVID, as we were able to make fast decisions with near-real-time data in the ever-evolving early days of COVID-19 lockdowns in the United States

2. Deep Analysis

Internet peak traffic has increased exponentially over the years. Figure 1 shows the internet peak traffic increase over the past 10 years. Peak traffic is calculated based on the aggregated 95th percentile peak traffic volume measured at the northbound Gigabit ethernet ports in each CMTS device. It includes residential and business customer traffic and a new customer growth on top of the organic traffic growth.

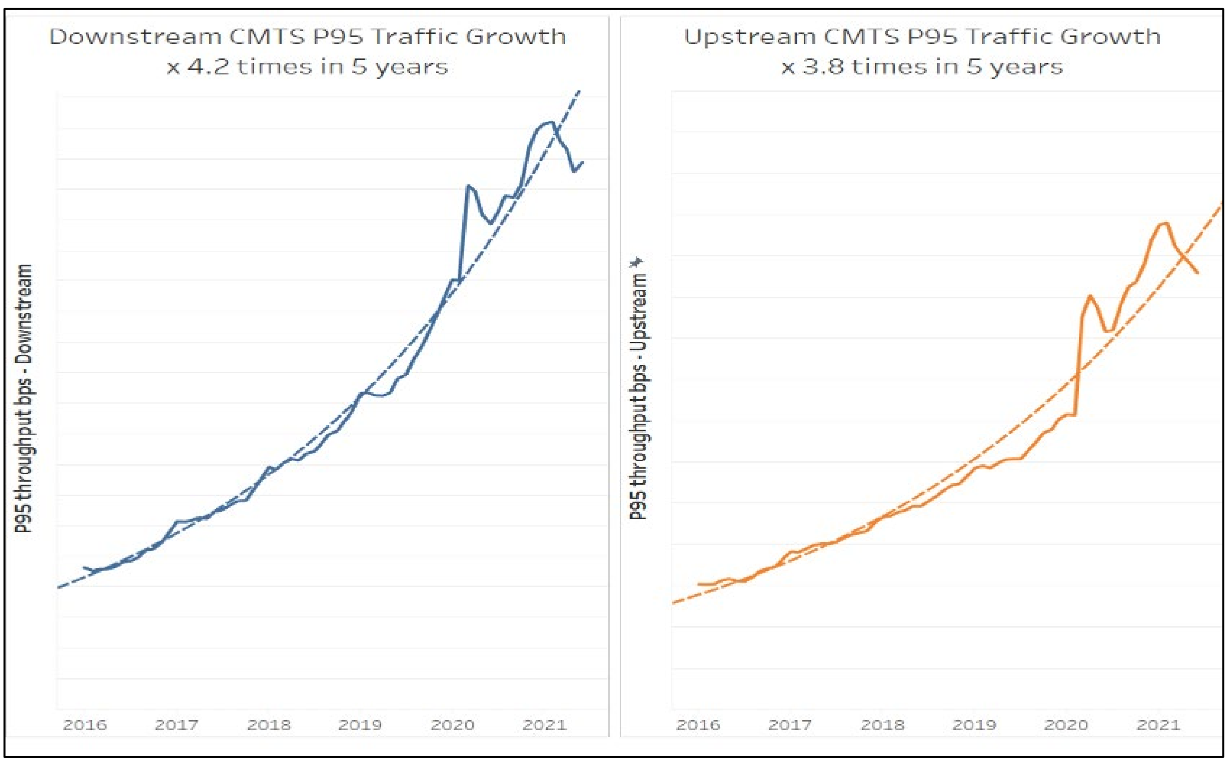

Figure 1 - Aggregated Peak Traffic Growth

Figure 2 shows the year over year (YoY) peak traffic growth rate. It had increased by 40-50% YoY growth until 2016 when major streaming service providers re-encoded the contents with optimized encoding algorithms and reduced the bandwidth usage. Since then, we have seen the traffic YoY growth rate to be about 30% until COVID-19 shelter in place started in March 2020.Traffic increased dramatically in a short period of time when shelter in place began. Upstream traffic growth is higher than downstream growth. Cox networks experienced 22% growth downstream and 34% growth upstream in 2 weeks after COVID-19 stay at home started.

As shown in Figure 2, upstream traffic grew about 70% and downstream traffic grew about 67% in April 2020 from April 2019. Downstream traffic growth continued to decline after it hit the highest point in April 2020. YoY growth rate since March 2021 significantly reduced because the latest value in the YoY calculation is compared to the surgical value after COVID-19 in 2020.

Figure 2 - YoY Peak Traffic Growth Rate

Figure 3 presents peak hour usage increase relative to pre-COVID-19 in February 2020. It compares the historical YoY growth trend and the actual growth rate. The recent downstream growth rate is smaller than the historical growth pattern. Upstream traffic continues to be growing and the recent trend shows 15% points above the historical YoY growth rate.

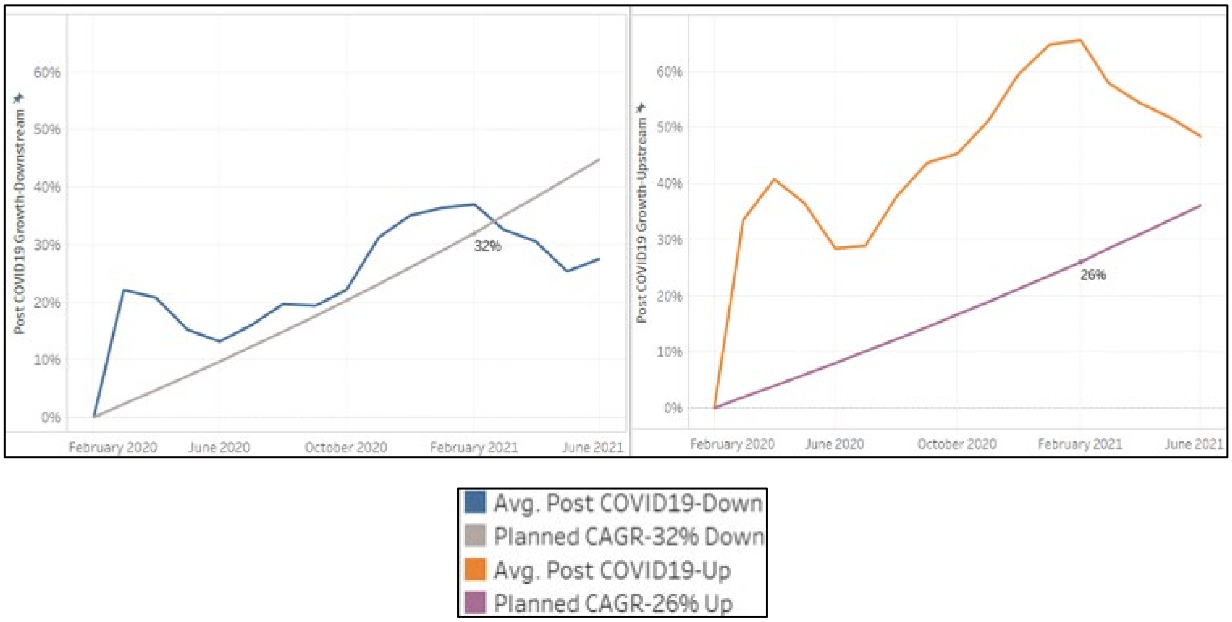

Figure 3 - Post COVID-19 Traffic Growth: Actual vs Expected

During COVID-19, the residential upstream traffic usage pattern has changed due to WFH (work from home) and SFH (study from home). In addition, upstream peak hours of residential customers have shifted to the daytime as shown in Figure 4. Recently it has returned to late evening, especially when the summer started, however, the daytime upstream usage pattern is still higher than pre-COVID-19.

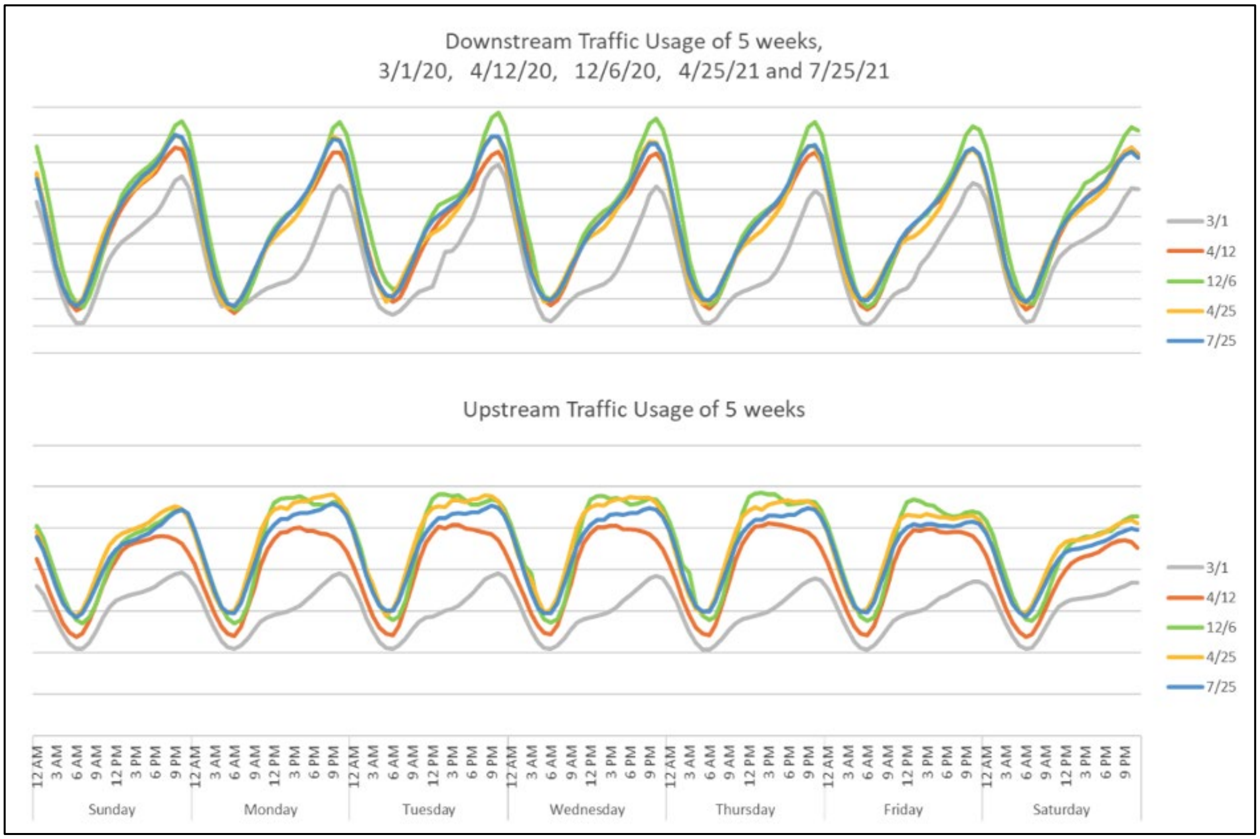

Figure 4 - Traffic Usage Pattern by Time of Day

The residential customers' peak traffic ratio of downstream to upstream is presented in Figure 5. It has grown to almost 19:1 since downstream traffic growth rate is higher; however, it reduced to 16:1 because upstream growth rate is higher after COVID-19 started. Currently the network has plenty of downstream capacity and relatively lower upstream capacity. Since Cox is building a mid-split plant, the capacity ratio of downstream to upstream will support the demand substantially.

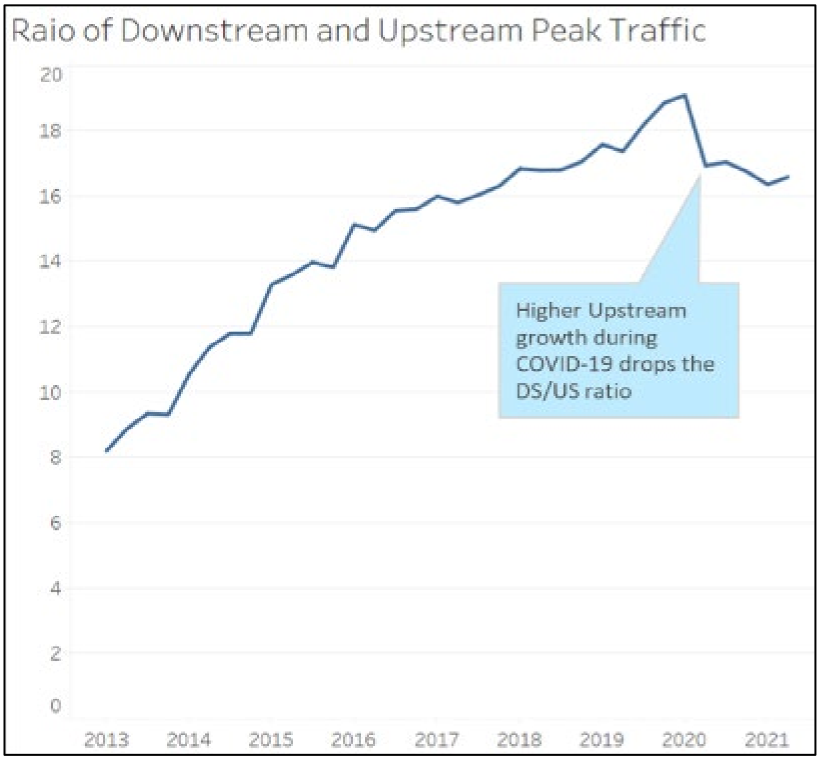

Figure 5 - Ratio of Downstream and Upstream Peak Traffic

Figure 6 displays the traffic category mix since March 2020. There has been a significant increase in WFH, and moderate increases in video, web, gaming and social since COVID started. The overall traffic remains dominated by video and web. Peak usage growth was still driven by traditional drivers such as OTT (over the top) video. In early 2021, WFH traffic increased up to 400% from the baseline of March 2020. Recently it went down to a 150% increase from the baseline.

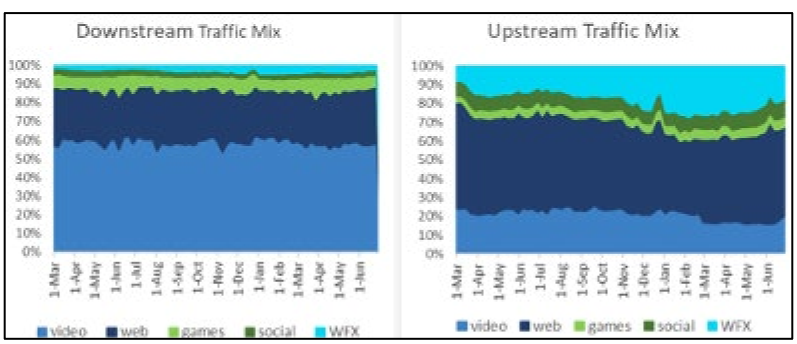

Figure 6 - Traffic Mix by Category

It would be interesting to see how business customers and residential customers consume network bandwidth differently. Figure 7 presents the average usage pattern of business and residential customers by the time of day. Business customers' usage peak hours are from 10AM to 4PM, while residential customers' peak hours are from 7PM to 10PM. Business customers consume less downstream traffic and more upstream than residential customers. Business customers' usage ratio of downstream to upstream is 5:1, while residential customers' usage ratio is about 16:1 as of time of authorship (August 2021).

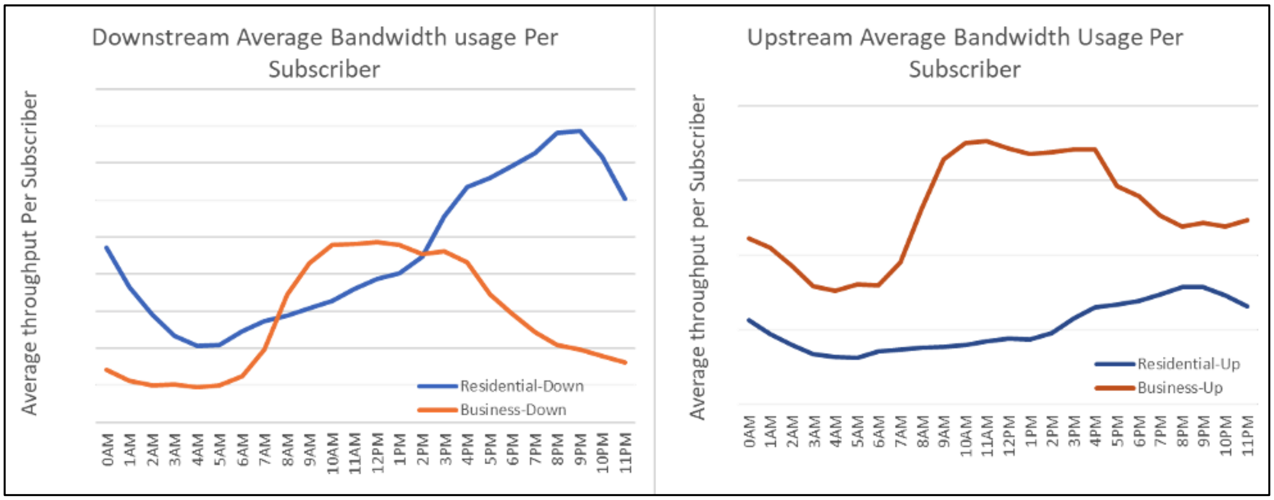

Figure 7 - Usage Pattern of Residential and Business Customers

Figure 8 shows the average usage throughput per customer by different tier speed, with the most popular tier, the first and second highest tier and the lowest speed tier. Basically, the average time of day usage patterns are almost the same across all of the different tiers. Figure 8 shows an example of the average throughput by speed Tier. The higher speed tier consumed more bandwidth, but not proportional to the maximum speed it can use.

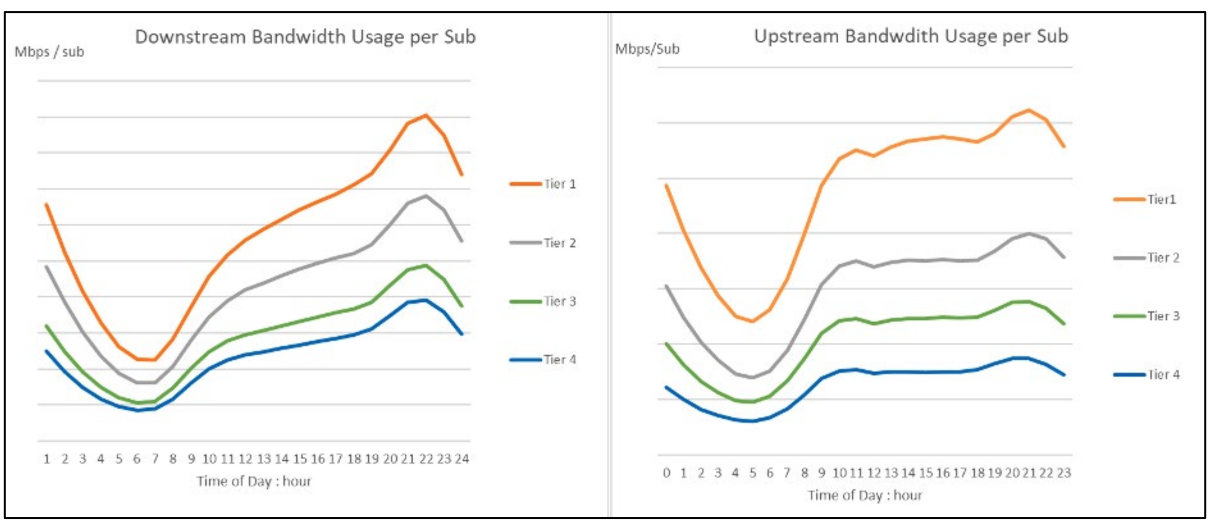

Figure 8 - Average Throughput of each Customer

3. Modeling and Planning

Translating network analysis into actionable information and data is the next facet of our teams. Implementing insights from our network analysis, in partnership with other industry experts such as CableLabs, has cultivated into a robust automated pipeline for the access network, and automated connectivity with other areas of network planning, such as backbone planning tools. These tools are developed on Cloudera, but built modularly for simpler migration as Cox's data strategy shifts

These modeling efforts were designed leveraging big data concepts and presented at SCTE in their infancy a few years back. As a recap, prior to the current methods, a series of manually generated Excel workbooks were used to infer a static CAGR-based growth rate for planning. These workbooks were first automated into workflows, connecting source data into a single schema, then logically joining and cleaning data. Within the workflow are tools designed to statistically identify anomalies, using kernel density and analysis of variance (ANOVA) to detect and alarm issues in the data. In addition to triggering external teams to research root cause and correct data issues, if possible, there are alternate "corrected" data tables generated. The corrected data uses interpolation methods to correct the gaps and errors until any corrected data loaded into the base (untouched) tables.

The final dataset, consisting of telemetry, topology, geospatial, upstream and downstream load, capacity, product and revenue data all on a time series basis, are loaded with node and market-level keying. That weekly utilization table is then used to infer node-level forecasts using an array of statistical and machine learning methods, specific to each node's specific growth pattern(s). From there, each node, with all its detailing attributes and forecasted load, are passed through a robust rules-based optimizer to determine the growth path, again, on a by-node basis.

This rules-based optimizer, dubbed the capacity response engine (or CRE) is one of the most important apprentices of the planning tools. The CRE can apply "if then" based business logic, simply applying business rules to grow node capacity, or it can use constraints such as budget, costs, manpower, ROI, customer counts, and other customizable variables to return strategic-based network growth views; and it can deploy combinations of business logic and constraint-based functions. With the data and information about the node, the CRE uses the current node topology and the various attributes of the node to determine the best series of node actions to both manage congestion, as well as selecting the shortest and/or most economic path to fulfil Cox's strategic network initiatives.

With the CRE completed, applying node actions to the 10 plus years of forecast, the results are used to create reports, as well as used to create the Volume Model (VM). The VM is our resource management model that feeds directly into our cost models. Here, we determine the license and physical equipment needed to grow the network. These outputs are also used to determine field headcount resources and critical facilities needed, with automations to link that data to their respective models, as well.

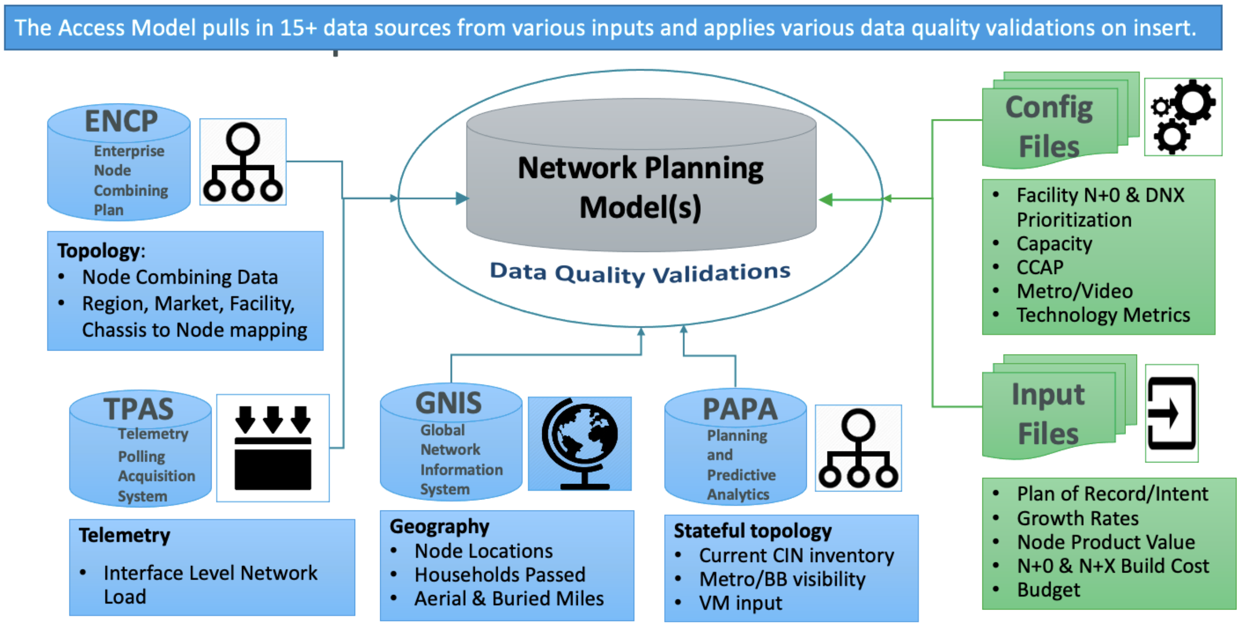

Figure 9 - General Overview of tooling

4. Business Continuity and COVID

Coupling our analysis and understanding of network and customer behavior has been vital for planning the network and keeping in front of the growth triggered by COVID. In the early days of COVID, in the beginning of lockdowns, we experienced a shift from the norm where customer-heavy nodes saw drastic decreases in usage while residential nodes saw a full year of growth, >30% year over year, in a few weeks' time frame (reference Figure 2).

With the robust pipeline of automated processes, we were able to marry the rapidly evolving data and analyses related to COVID and utilize insights from those analyses to quickly spin up forecasts to help us understand what "could" happen, as well as how we would need to grow the network to respond, thanks to the CRE. Data was used in creating hundreds of scenarios, and vital in determining courses of actions needed to address the elevated number of highly congested nodes* (see Figure 3). As well, leaders were able to assess costs near real time and make decisive decisions to keep up with this new network demand.

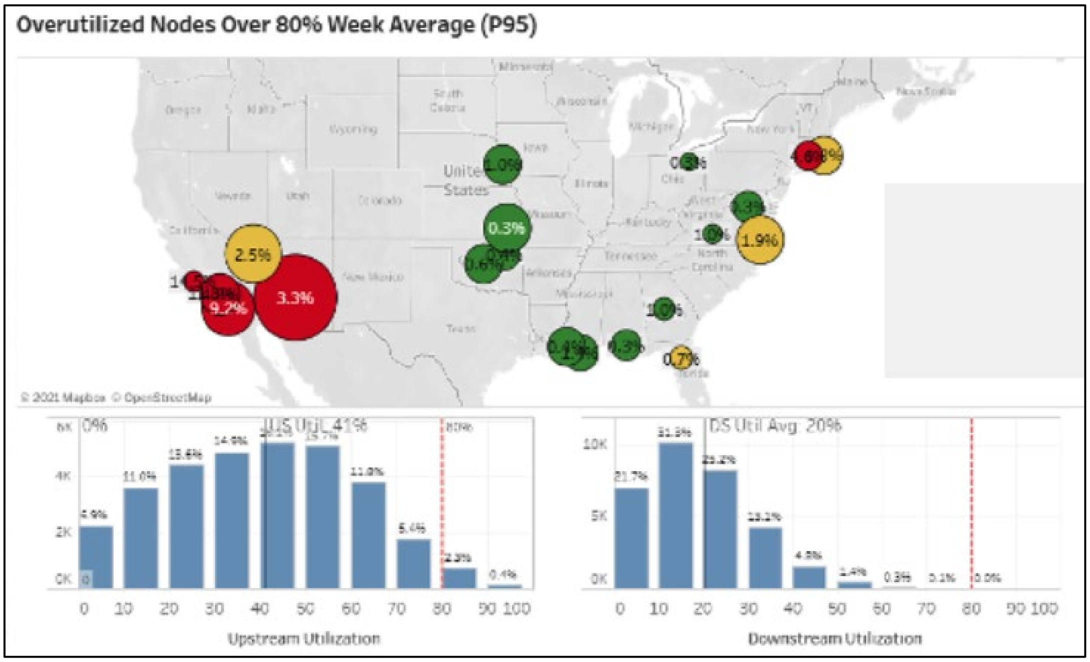

Figure 10 - COVID Related Network with High Congestion Nodes

Understanding the increased congestion was just a small portion of what was required. Next, we needed to be able to forecast and predict where we would see sustained growth, where we could expect less growth, and when we should see the effects of COVID decay from our growth curves. With our pipeline in place, we were able to increase the accuracy of our forecasts by implementing what we had learned so far, as well as reviewing and implementing assumptions regarding what we 'think' may happen. Of course, forecasting models weren't always 100% correct. Since we didn't have prior pandemic data to inform our models on the impact to network usage and node growth, we had to cycle input data in as fast as we could obtain it. From there, the models were updated on an ongoing basis. We were then able to add and control the COVID related variation by calculating a coefficient for the increased growth over what was expected (i.e. COVID growth over baseline expectation) then adding them into our Auto Regressive Integrated Moving Average models (ARIMA, expressed: $Y_t - Y_{t-1} - \mu = \phi_1(Y_{t-1} - Y_{t-2} - \mu) + a_t - \theta_1 a_{t-1}$) and controlling with future assumptions and COVID decay with indicator variables (sometimes referred to as dummy variables or one-hot encoding in computer science). Figures 11 – 14 show some of the evolution of our assumptions, as well as the increased accuracy as we integrated more and more data after the mass lockdowns started.

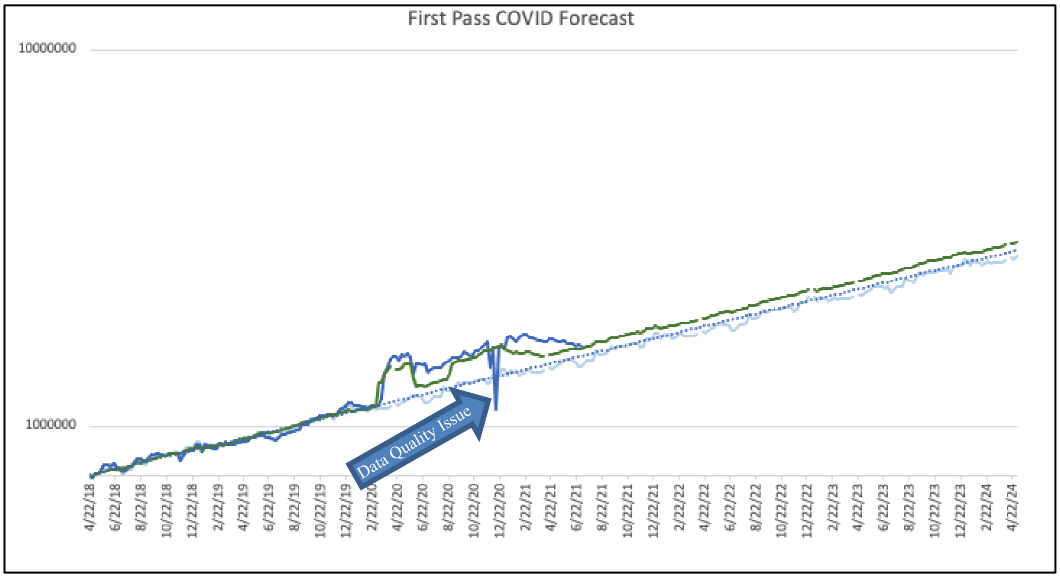

Figure 11 - Forecast post COVID W/O Forecasting COVID Effect

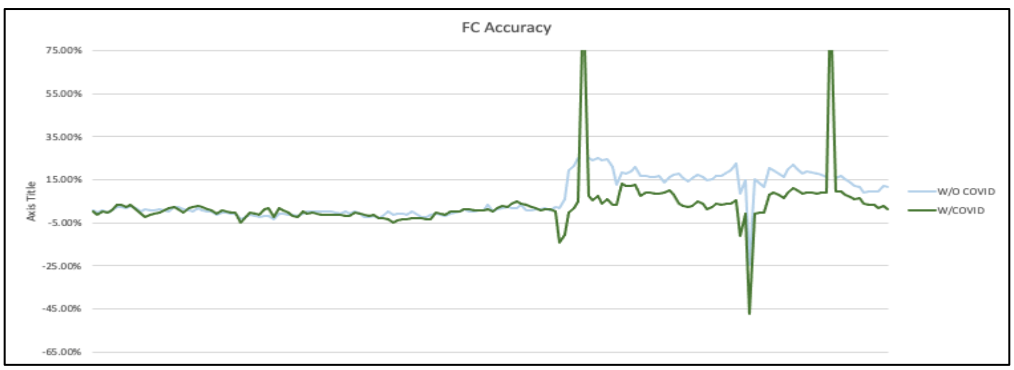

Figure 12 - First Pass Forecast post COVID

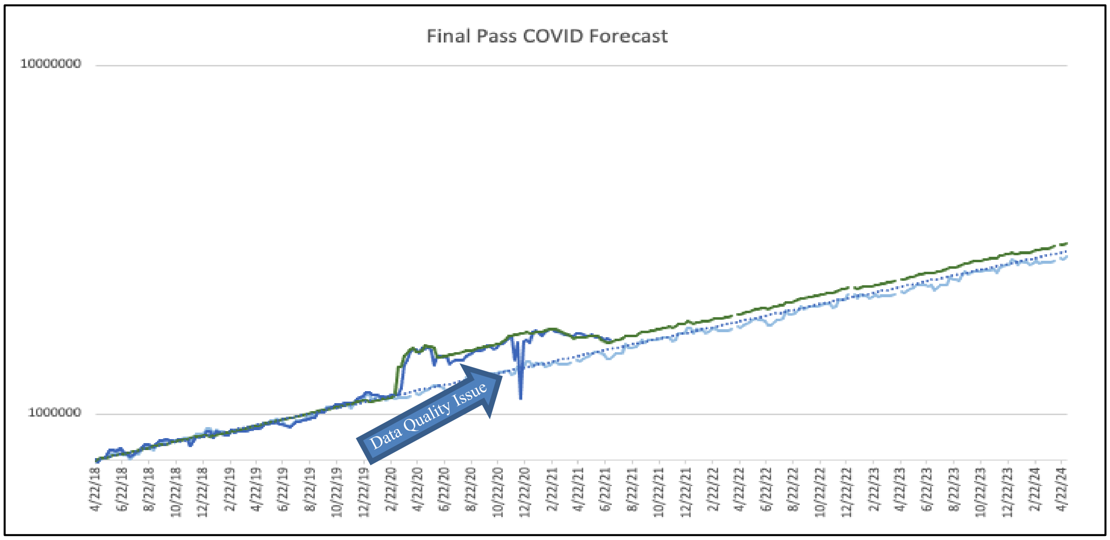

Figure 13 - Forecast post COVID with COVID Effect Forecasting

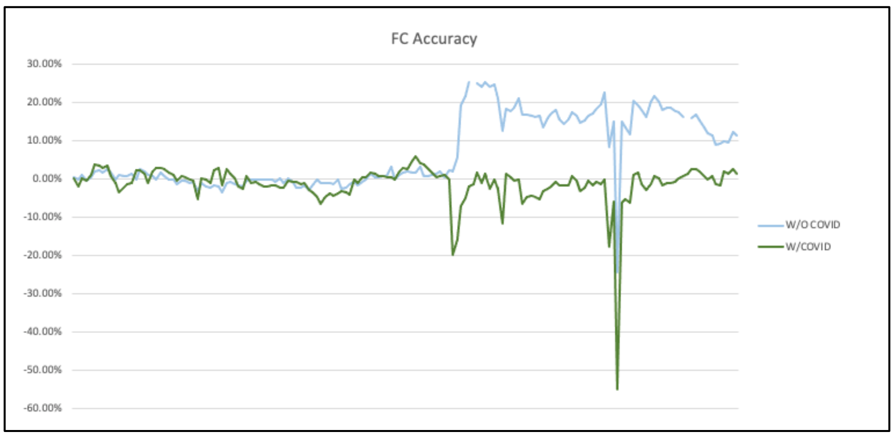

Figure 14 - Final Pass forecast with improved accuracy

With that, we were able to update and publish the models regularly. These published models are output into reports or used to generate costs and resource models, which are used in decision making about the direction of growing and planning the network. At first, we were updating the forecasts models weekly and communicating the findings to the broader teams. Despite the robust tasks of computing all the predictive models on a by-node basis, our scalable pipeline was always up to the task, and our planning teams were able to respond quickly. When we would see large shifts in the forecasts due to variations in the COVID related traffic, we would push the forecasts through the CRE to the outcome for our near-to-short-term plans.

This led to a need for complete network/model reworks of the full 2021 calendar year node growth plans, directly effecting outside plant and critical facilities executional plan of record. Utilizing our tools, we were able to re-work about year in just a few months, leaving more time for our field partners to adjust and react to the updates. To put this in perspective, our BAU (business as usual) policies require 18 months advanced planning "locks", worked one quarter at a time over a period of a quarter. Achieving a re-work of majority of the plan in the midst of the plan's deployment required solutions that not only work at scale, but can also be quickly adapted to unforeseen events, such as the COVID pandemic.

But just adjusting the plans also wasn't enough to address this congestion. Our engineering partners engineered solutions to address congestion; network solutions we had never modeled or deployed before. As these solutions were defined, we were able to model them. These solutions included spectrum actions, such as the implementation of 5th carrier and SOFA (sub split orthogonal frequency-division multiple access), to add fast and lower cost relief to highly utilized nodes. Being able to model these solutions as they were developed meant our leaders could measure the cost and benefits within days of the details being finalized so they could be integrated into planning for testing and deployment. Figure 15 below is an example of the output after a scenario run, implementing 5th carrier and SOFA to shoe the decreased need in other more invasive and expensive actions.

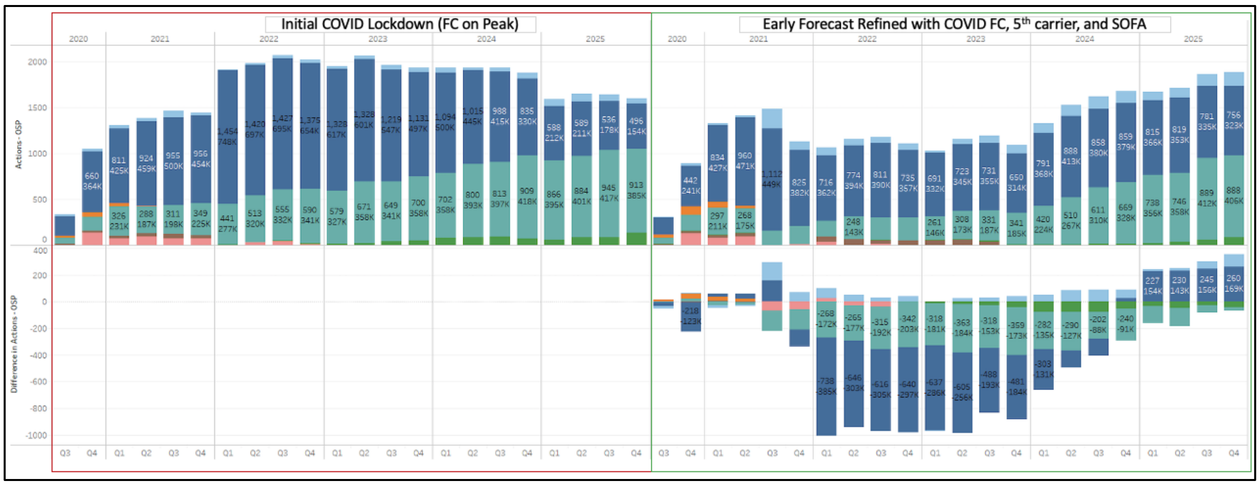

Figure 15 - Planning Scenario Implementing 5th carrier compared to predecessor

Conclusion

Cox's advanced network planning tools are where the rubber meets the road. A few years ago, the idea of these tools in their early developmental stages were presented at SCTE. Since then, we have been maturing the tools and increasing the capabilities and functionalities. Over the past few years, our planning teams have continuously developed and refined our tools to meet the needs of the business. In an extremely robust application, much of our analysis and planning components are automated, and automated with the flexibility to adapt quickly to changing business needs. The 2020 pandemic lookdowns tested our tools and infrastructure, but thanks to the scalable architecture, Cox was able to deliver refined plans to keep our customers connected.

Abbreviations and Definitions

ANOVA	analysis of variance

ARIMA	auto-regressive, integrated, moving-average
BAU	business as usual
CRE	capacity response engine
HDP	Hortonworks Data Platform
Highly Congest Node	Nodes with sustained congested p95 utilization >80%
OOT	over the top
SFH	school from home
SOFA	sub split orthogonal frequency-division multiple access
WFH	work from home
YOY	year over year

Bibliography & References

(1) *Texas A&M Department of Statistics, Dr. Simon Sheather, 2018*

