✎ Edit article
👁 View post

Not Normal Telecom
4,673 subscribers

✓ Subscribed

Can we go back to basics?

Telecom Cloud for Non-Engineers

Geoff Hollingworth 🔗
Not a normal CMO. Rakuten Whisperer. Rakuten Symphony

🔖 ⋯

October 24, 2022

Introduction

In my decades within telecom, I've not seen a more consequential technology shift as pervasive as the move to cloud. In conversations with non-engineer types, I'm sensing a widespread disconnect between what is happening, and the critical context and perspective required to make sense of AND act on these changes. This is equally true among new talent pouring into our industry, journalists covering this either cloud or this market for the first time, and those in management that have watched as the world around them has swiftly and drastically transformed. If you fall into one of these groups, this post is for you. It is long and detailed, but hopefully easy to consume. If you do not have time to read this then let me state the final summary principles at the start.

1. Each telecom needs to be a cloud operational expert in the same way each mobile operator needs to be a radio operational expert.

2. Design applications to be as cloud independent as possible. Tooling used by cloud XaaS offerings is increasingly offered independently, all internet companies are driving the same results for the same reasons.

3. As soon as a workload gains known predictable load, move it to private cloud for simple economic reasons.

4. New applications moving to cloud need a cloud that is very present to the user, both for latency/jitter as well as data volume processing reasons. This "edge cloud" does not exist today and the first

application with these requirements requiring scale are mobile network radio signal processing applications that exist today.

5. For all mobile network operators, build an edge cloud and be your own customer first. You will create the business case for a fully deployed edge cloud with yourself as the first customer. You will create a real time operations team that can manage your own uptime requirements, that then can be sold as an edge cloud service for others.

6. You don't lose control by using other peoples clouds, you lose control by not knowing how to use other peoples clouds you use.

To Begin...

This article assumes no prior computing or software knowledge.My goal is one of education. I am going to try to be more neutral than Switzerland here. If I fail, I welcome proposed improvements, clarifications and requests for updates. I only ask for no company driven agendas, which does no favors for the intended audience. With cloud, there is never one "right answer", only the "right answer for you."

By the end of this less than 10-minute read, I want readers to understand the basic strategic imperatives, requirements and implementation strategies involved in building a successful telecom cloud strategy, execution and operation. Telecom cloud is not an engineering problem. It is not an economics, operational or strategic problem. It is all of these things. I attempt to address these perspectives and how they interplay with one another.

We'll start with the basics, explaining what cloud is at its most basic level, technically and economically. We'll move onto how applications destined for cloud have different characteristics and performance needs impacting deployment decisions. Finally, we'll wrap with my opinion on where telecom should be focusing as we move from a centralized cloud view to an edge distributed cloud view.

Let's get started...

What is Cloud?

At its most basic cloud allow people to use computing capability as a service rather than having to build it themselves. It's primary characteristic and difference from previous generations of IT operations is its accessibility provided by software APIs (Application Programming Interfaces). There are still physical servers that are installed and need to be maintained but these are not what application developers and owners (users) directly use. Rather than controlling the actual hardware servers (in most instances), the users work with abstractions via these APIs. They can create virtual machines, that look like real servers to the user but are really software emulations of servers. They can create containers that also look like real servers but are more efficient representations, meaning they start faster, more can be fitted in real servers and they can run more efficiently. This abstraction is commonly known as Infrastructure as a Service (IaaS). You can also use Platform as a Service (PaaS) services that provide enablers to build applications on top of. And finally you can use end user services (Software as a Service).

We will focus on IaaS and containers in this article since this is where most of the basic conversation is now focused in telecom. Kubernetes,

commonly abbreviated as k8s - k + 8 letters + s. has become the de facto industry standard for container management. K8s is an open source project that was first released by Google in 2014.

Cloud users (developers, devOps people) create containers. Containers look like a stand alone server just for them. They deploy application software to the containers. The applications run. When finished the containers and applications can be deleted. The container capacity is returned to available cloud capacity and a different user and different application can use it.

Because of the accessibility, cloud is the most efficient form of IT infrastructure to date. It has transformed first how web 2.0 application developers could build applications from scratch. Application developers no longer have to buy physical servers and place them in hosting centers. Instead they can sign up to Amazon Web Services (for example), create "virtual servers" and just pay for them while they used them. If they need more, they create more, if they need less they give them back to AWS. No humans are involved in any of this process, it is all handled through the earlier described Application Programming Interfaces. Originally developers paid by the hour for the servers they used and it allowed very rapid development and testing. Now different pricing plans exists that we will cover later. This is the birth of what is called public cloud. Anybody with a credit card can sign up and create servers and new software apps. The growth of public cloud mirrors the growth of the smartphone app economy on on Apple and Android. The app on the phone, what we can see, is called the front-end. The app speaks to a "back-end" via APIs and send, receives data to make the app work as expected. The back-end lives on a cloud somewhere.

In parallel to rise of the public cloud phenomena, companies were also transitioning to private cloud operations, primarily with a company called VMware. This is where the company would keep its data center, still buy, provision and maintain servers, but offer its internal company users access to virtual servers rather than having them buy physical servers directly. In many cases, it was easier and faster for even company users to use public cloud rather than their own private cloud offerings and many did, primarily to avoid the internal bureaucracy and process that still existed and caused wasted time. This became commonly known as shadow IT and is still a big security and data privacy concern to many organizations.

The main drivers for cloud growth, especially public cloud growth, have been the massive convenience and speed of realization that the programmable accessibility provides, combined with the massive flexibility of seamlessly growing and shrinking the needed capacity when and if needed. Increasingly public cloud providers have increased the number of highly powerful additional services that they support above the basic infrastructure offerings, such as support for IOT, AI, databases, monitoring, CI/CD, everything.

From Pets to Cattle

Because of the increased accessibility and programmability of cloud infrastructure, it is possible to automate the management of thousands of servers the same way you automate one server. One of the most powerful changes that cloud brings to digital infrastructure is the opportunity to industrialize the whole operation. The effects of industrialization are always the same in any industry, increased

efficiency, increased output and faster ability to make changes at bigger scale. From the IT perspective this has been described as moving the mental model from pets to cattle. With traditional IT all servers were treated as pets. They were given unique names, they were individually cared for. If something went wrong with one (fell ill) then somebody spent the time investigating what was wrong, fixing it, and getting the server back to full health. This was because of the required investment required to bring each individual server online. And each individual server always had a specific purpose and role.

In cloud this is not true. All servers are treated as cattle. They are not given unique names, they are give identity numbers 0001, 0002, 0003, for example. They are provisioned and brought online in bulk. If there is something wrong with a server (output is decreased) then the server is replaced with a new server and the old server is recycled for parts. Key KPIs are not individual server health but herd effectiveness and output. You kill the sick cow and sell it for parts rather than nursing it back to health (sorry for the brutality, that is industrialization not me).

To understand more about this process see "**The Industrialization Cycle Explained**".

This is the resultant mindset change for the physical servers that the cloud runs on top of. This is also true for the virtual servers, containers and software applications running on top. It is easier to destroy a server and recreate a server than mend a server. A fundamental difference in the design of cloud native software versus traditional application software is that cloud native software assumes something will fail because it will, and designs to continue seamlessly when that happens. Traditional software design tried to mitigate the effects of failure through redundancy. However good the design, with the traditional model there will always be the potential for system outage.

What to Consider when Designing Telecom Cloud?

The above introduction hopefully gave some simple definitions and understanding to why cloud is such an important concept in the designing of any computing based operation. The more interesting discussion is how cloud should be utilized most appropriately to solve a specific industry need. Telecom is now full steam adopting cloud, calling it telecom cloud. The reality is that the industry is simply adopting cloud, it has specific needs, and the system design needs to most appropriately solve for these needs. The needs are highly distributed, highly latency sensitive real time workloads.

Earlier I introduced the two main differences that cloud has introduced – accessibility and the ability to increase and decrease capacity based on demand. I also introduced two main types of cloud, public clouds and private clouds. Public clouds are offered to all companies and the capacity can be shared across all companies and developers that use it. Private cloud is when a company invests in building and operating capacity to be shared just for that company.

When designing telecom cloud, we must consider performance requirements, as well as the costs for delivering such performance. Performance requirements are not a one size fits all proposition, especially in telecom, the needs are defined by application type and

should be modeled appropriately. I want to frame the discussion along two main axes with respect to application requirements.
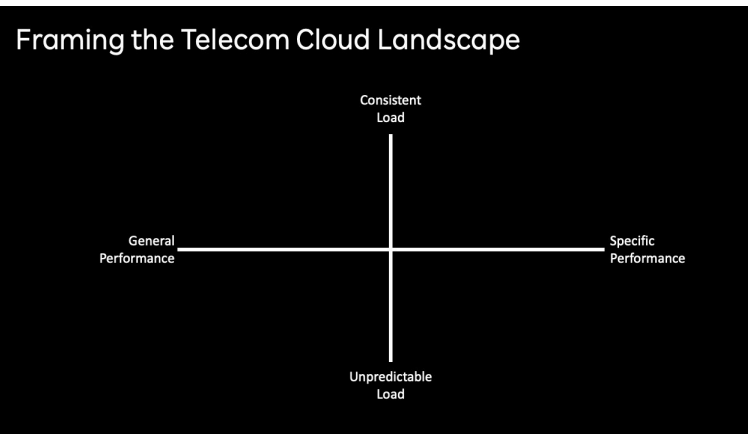
 On the x-axis, does an application require specific performance such as very high guarantees on latency? Obviously in telecom the distributed unit radio software falls into this category, requiring very low latency and also very low jitter. Other telecom applications, primarily outside the radio domain do not have such performance sensitivities. For example management applications like coverage planning, site design and maintenance management.

On the y-axis, what is the known base load for the application? This informs whether it is more economically beneficial to run on private infrastructure (private cloud) versus public infrastructure (public cloud). Let us look at example application types through this lens.
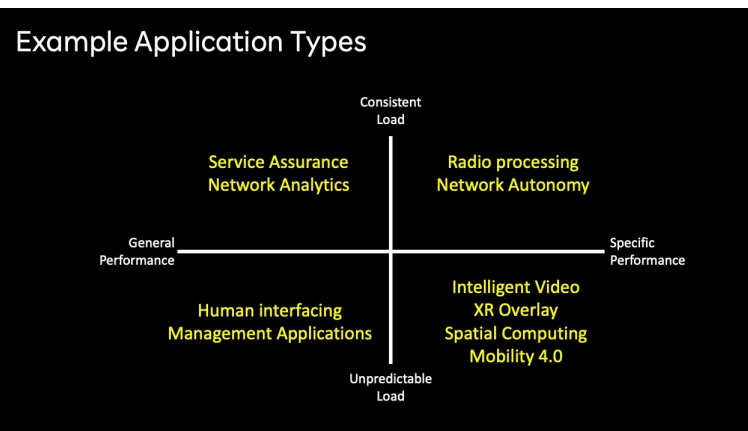
Telecom is the only existing industry that has highly distributed, high performance low latency requirements today. This is why telecom already runs such a highly distributed infrastructure, since distribution equates directly to ability to deliver. The radio software today has very specific performance requirements and it cannot function unless very precise single digit latency can be guaranteed. Mobile telephony is the only existing edge workload that operates at scale and this is why it is the perfect proxy for other applications being discussed. The second industry application type that is appearing is software that is enabling network autonomy operations. This requires real time insight into data streams, to allow the self driving of network operations. This is no different than what we see in autonomous driving challenges, except telecom is more like autonomous driving management and autonomous

fleet management in parallel. Both these workloads are highly predictable in terms of utilization and performance.

Intelligent Video, XR Overlay, Spatial Computing, Mobility 4.0 application types are all edge workloads that are still nascent in development. It is very unclear where and how quickly demand will develop and become mass market. These are all good application types that have high performance requirements but today, low predictability.

Human interfacing applications fall into the category of non-specific performance requirements and varying performance load. These systems are general management, workflow management systems.

Service Assurance and network analytics are good examples of applications that have general performance requirements but have known load requirements today. These application types also tend to be very data intensive and thus have sensitivity to data gravity, flow, cost of ingress/egress.

The next analysis is to consider public versus private cloud solutions for the different application types.
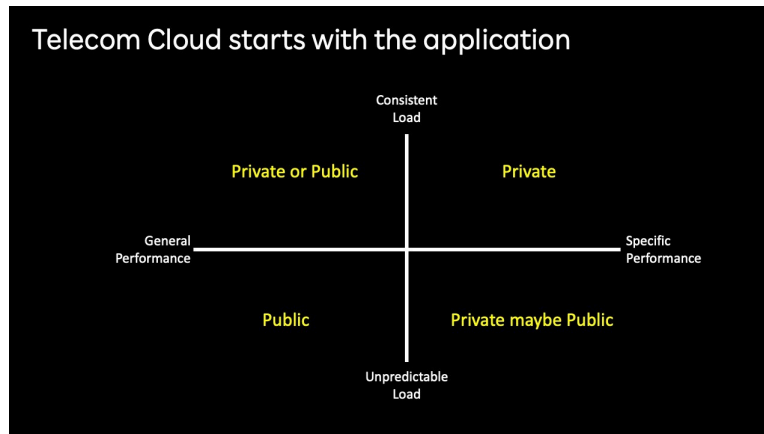
To be clear, this is not a static analysis. I fully expect the capabilities of public cloud to continue to develop and support more and more specific performance requirements. in the future. This also depends on the access network used. Fixed access networks reach a public cloud peering point with much greater immediacy than seen in mobile networks. This is because in mobile networks, traffic is routed to manage continuous mobility of the users, with needs to perform continuous handovers between cooperating radio towers.

Understanding Cloud Economics

Previously I presented the efficiencies from introducing the industrialization cycle into any industry. Cloud is the industrialization of IT. The companies who runs those clouds gather the rewards and decides who benefits from the efficiencies.

Public clouds are a business currently growing at 30-40% CAGR. They are not a charity. Internet companies know very well how expensive public cloud is once scale is achieved and load is predictable. Many internet companies find themselves in the tragic position of being married to a public cloud provider, with ever increasing expensive cloud

bills, and a highly complicated migration strategy away from the dependency. The consideration of business life cycle management versus costs is something that needs to be considered at the start.

It is also important to understand that there are many public cloud pricing models provided to lessen the cost, as predictability increases. Initially cloud capacity was rented by the hour. Now, cloud capacity can be reserved for longer periods, for less cost. Cloud capacity can be granted through auction type mechanisms when time sensitivity is not an issue (spot pricing) and even specific hardware machines can be rented, without any cloud abstraction, for instances that hardware isolation is a specific requirement or software licensing has still not modernized.

Even with long term pricing plans, leasing servers from another company at scale is becoming a very large known problem and the internet companies that have been touted as the beneficiaries of adopting cloud are now dealing with the brutality of ever increasing cloud bills while migration away from their cloud lock-in is a multi-year investment and engineering project in its own right. Examples are increasingly appearing in public.

How we reduced our annual server costs by 80%—from $1M to $200...

An interview with Zsolt Varga, the tech lead and general manager at Prerender

Medium

Why we're leaving the cloud

Basecamp has had one foot in the cloud for well over a decade, and HEY has been running there exclusively since it was launched two years ago....

world.hey.com

Uber recently celebrated the results of a multi-year project to engineer cloud independence. "New blog! Uber's infrastructure engineers deep dive into how they leverage infrastructure as code to manage hundreds of thousands of servers across multiple cloud and on-prem providers."

Crane: Uber's Next-Gen Infrastructure Stack

Uber has been on a multi-year journey to reimagine our infrastructure stack for a hybrid, multi-cloud world. The internal code name for this project is...

Uber Blog

Telecom is in the enviable position to avoid mistakes made by others, that are only just becoming visible now. Whatever choice is chosen, it is impossible to execute any strategy well unless there is a core highly competent cloud system engineering and operations team.

As David Heinemeier Hansson of 37signals says above "Now the argument always goes: Sure, but you have to manage these machines! The cloud is so much simpler! The savings will all be there in labor costs! Except no. Anyone who thinks running a major service like HEY or Basecamp in the cloud is "simple" has clearly never tried. Some things are simpler, others more complex, but on the whole, I've yet to hear of organizations at our scale being able to materially shrink their operations team, just because they moved to the cloud."

Closing Remarks

To summarize all of the above into some simple principles

1. Each telecom needs to be a cloud operational expert in the same way each mobile operator needs to be a radio operational expert.

2. Design applications to be as cloud independent as possible. Tooling used by cloud XaaS offerings is increasingly offered independently, all internet companies are driving the same results for the same reasons.

3. As soon as a workload gains known predictable load, move it to private cloud for simple economic reasons.

4. New applications moving to cloud need a cloud that is very present to the user, both for latency/jitter as well as data volume processing reasons. This "edge cloud" does not exist today and the first application with these requirements requiring scale are mobile network radio signal processing applications that exist today.

5. For all mobile network operators, build an edge cloud and be your own customer first. You will create the business case for a fully deployed edge cloud with yourself as the first customer. You will create a real time operations team that can manage your own uptime requirements, that then can be sold as an edge cloud service for others.

6. You don't lose control by using other peoples clouds, you lose control by not knowing how to use other peoples clouds you use.

Telecom cloud, or more preferably edge cloud in telecom, is not avoidable rather it is the most strategic future aspect of telecom to solve for. We are building edge clouds with radios on the end, the first tenant is mobile connectivity itself.

But herein lies the challenge of telecom. One that all newcomers and new learners alike should take to heart:

If telecom is no longer the industry that knows how to best run highly distributed, highly performant infrastructure (edge-cloud) and telecom is no longer the industry that is delivering the best end user experiences above connectivity on such infrastructure, then what IS telecom?

Those of you in the next generation of telecom doers, thinkers and leaders need to figure this out. We have to be honest to ourselves on both the problems we face and the opportunities we can meet. This is not a journey of 5G Dreams but one of hard love, being honest to ourselves first, so we can be honest to others. If we in telecom did not own licensed spectrum, would anyone care about us and would anybody want to do business with us?

The future requires real-time performing edge infrastructure that runs applications across industries. The only question is who enables this? Telecom is perfectly positioned to add value in this transition but telecom needs to choose to do hard things and choose what to be good at. By my estimation, telecom still has about two years to figure this out and then the next 15 to successfully execute.

The game is not over, it is just starting. What role will you play?
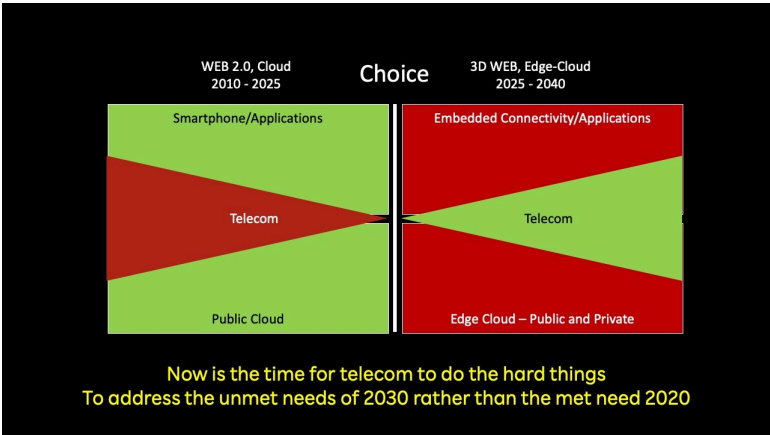
Comments

125 · 16 comments · 10 reposts

👍 Like   💬 Comment   ↗ Share

⚙

Add a comment…   🙂 🖼

Most recent ▾

Jesús Hernández · 1st   5mo •••
Telco Eng. | Cloud Core Network Solutions | Product Management …

Interesting redaction approach

Like | Reply

Per Treven 🛡 · 1st   2y •••
Chief Repair Officer - Home Repair Construction and Design

Thanks for a great article Geoff! Helpful and insightful.

Like | Reply

Nikhil Verma 🛡 · 2nd   2y •••
Product Owner | SAFe® Certified | Business Analyst | Ex-Rakuten

Amazing article and very well explained Geoff Hollingworth !!

Like · 👍 1 | Reply

Fredrik Almqvist 🛡 · 1st   2y •••
Innovative Business Engagement Leader

Great work Geoff! Would love to see telecom gearing up for the own edge cloud, but can also see it might go the "CDN" way. In 2 years time we know the answer if your prediction is correct.

Like · 👍 1 | Reply

Anthony Behan 🔗 · 1st   2y •••
Industry Managing Director @ Cloudera | Telecommunications, Medi…

Excellent piece Geoff, and well thought through and put together. There's a lot of hot air in the industry (it was always ever thus!) and every so often we need a reality check. The only thing I'd question is whether there's ever been 'a more consequential technology shift as pervasive as the move to cloud' - the shift to wireless, perhaps? ...more

Like · 👍 3 | Reply

Deepak Rana ☑ · 1st
CEO at ThinkScan - Agentic AI and applications | ex Intel (MCM, GM...

2y · · ·

Excellent writeup - as always!

Like | Reply

David Bradbury ☑ · 1st
Experimental Music Composer with a background in songwriting for ...

2y · · ·

Interesting thanks Geoff - as for the quote "You kill the sick cow and sell it for parts rather than nursing it back to health (sorry for the brutality, that is industrialization not me).
Maybe that's where the term "Server Farms" originates lol! ...more

Like · 👍 1 | Reply

Patrik Sporre ☑ · 1st
Digital Mobility | Strategy, Sales & Leadership at the Intersection of ...

2y · · ·

Brilliant!

Like | Reply

Mohsen Mirmoradi 💡 🌍 in · 1st
Driving Results in High-Stakes Projects | Freelance Program Manage...

2y · · ·

👏 👏 Great thoughts, nicely articulated!

Like | Reply

Stephane H. Maes, PhD. in · 1st
CTO & CPO (Product) @ IFS | IT, ITOM, ESM, Cloud, Mobile, Telecom...

2y · · ·

The problem with Telecom cloud instead of Cloud for Telecom is that may infer a technology vertical solution instead of telecom solutions build on horizontal clouds...

Supporting new use cases and requirements does not necessar ...more

Like · 👍 3 | Reply

⤡ Load more comments

Not Normal Telecom

Not Normal Telecom
Provoking conversation and breaking taboos

Brian, Mark and 2,300 connections are subscribed

4,673 subscribers

✓ Subscribed

More articles for you

‹ ›

THE BIG CLOUD MIGRATION

Telco Cloud

Telecom Cloud Market is Set to Fly High in Year to Come

Suraj A Sutar

6

The Big Cloud Migration: Opportunities, Challenges, and the Road Ahead

Daniel Junge

58 · 2 comments

Telco Cloud Evolutions

BeyondTelco

30 · 2 reposts

○ ○ ○ ○