Hyper Scalers and Energy Costs

How Rising Energy Costs is Accelerating ARM instances, and Oracle is an unlikely beneficiary

DOUG O'LAUGHLIN

NOV 01, 2022

♡ 32   💬 1   ⟳   Share

Hello - it's the midst of earnings season, and I wanted to write quickly about something I thought was interesting, and that's how energy costs are flowing through to hyperscaler operating costs. I think there's a bigger picture to draw from rising energy prices.

I couldn't help but notice the absolute energy cost headwinds from Microsoft and Amazon. Here's what they said on the call.

Microsoft on energy costs:

> Excluding the impact of the change in accounting estimate for useful lives, Microsoft cloud gross margin percentage decreased roughly 1 point driven by sales mix shift to Azure and lower Azure margin, **primarily due to higher energy cost**

> Therefore, with our first quarter results and lower expected OEM revenue for the remainder of the year as well as over **$800 million of greater-than-expected energy cost**, we now expect operating margins in U.S. dollars to be down roughly 1 point year-over-year

> We did not see as big of that. As I said, it's over $800 million for the year. Some of that was in Q1, but the majority of it will be in Q2 through 4. And I think if you want to think about it, **it's somewhere 250-ish a quarter**. It's not exact, but that would be a decent assumption for the remainder of the year.

Amazon mentioned energy costs being a 200 basis point headwind.

> We have seen inflation in our wages this year and particularly on our [Czech] employees is heavily concentrated in AWS. So that's one element of it. We're also seeing energy costs that are materially higher than they had in pre-pandemic,

electricity and the impact of natural gas pricing. **So those prices have up more than 2x over the last couple of years and contribute to about 200 basis point degradation versus 2 years ago**. So we're fighting through some of that as well, which is a new thing for the AWS business. But we'll continue to look for ways to optimize our operations to use less energy. And as we scale, we'll outrun that growth trajectory.

The public cloud is more beholden to energy costs than I thought! I want to discuss that and how to measure cloud energy efficiency. Google didn't quantify the impact at all, but luckily there is a datacenter agnostic metric we can refer to for energy efficiency for data centers, and it's called PUE.

Public Cloud and PUE (Power Usage Effectiveness)

PUE is a ratio of the data center's power efficiency spent on computing compared to overhead.

5.1.1 THE PUE METRIC

Power usage effectiveness (PUE) reflects the quality of the datacenter building infrastructure itself [67], and captures the ratio of total building power to IT power (the power consumed by the actual computing and network equipment, etc.). (Sometimes IT power is also referred to as "critical power.")

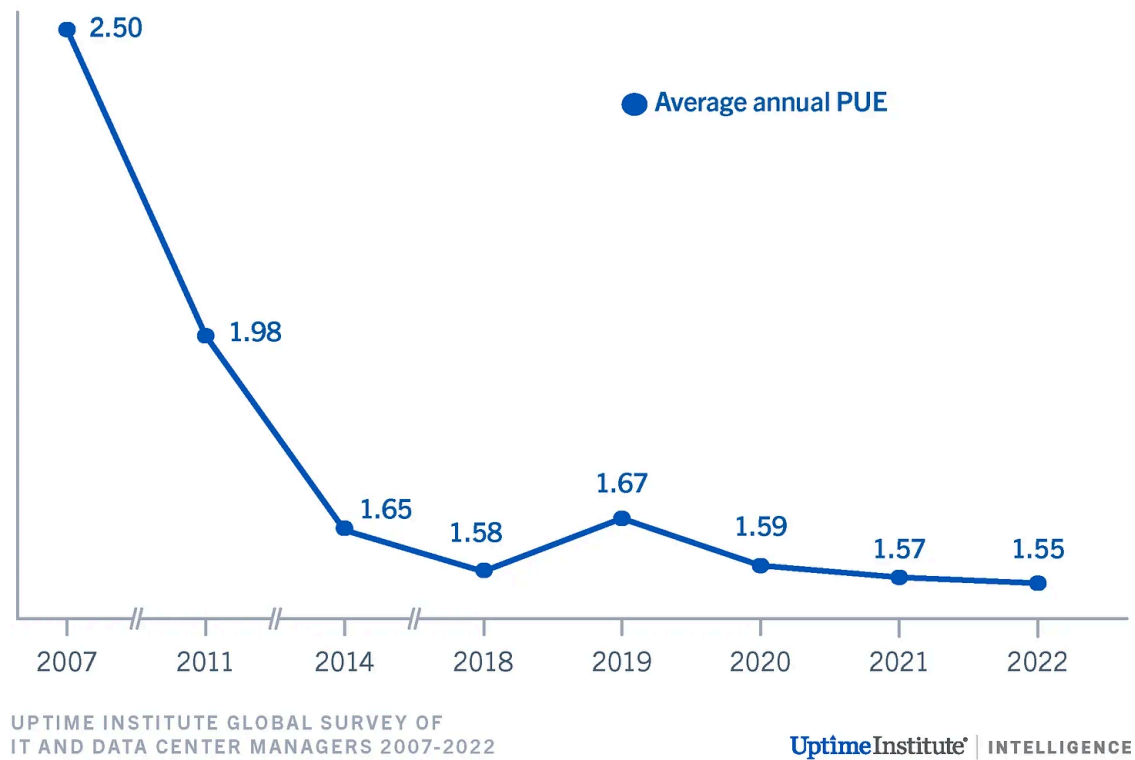PUE = (Facility power) / (IT Equipment power)

From the Datacenter as a Computer Textbook

It's a simple ratio — total energy/energy used for IT = PUE. So let's say that 50% of a data center's energy is used for the servers, and the other 50% is used to support the chips. That's a PUE of 2.0 (1.00 / .50) because of the total energy/equipment power.

We have a benchmark of this globally and, importantly, a good insight into what the hyperscalers perform at. Below is a chart from the Uptime Global Data Center Survey. PUEs are decreasing, but at a slower rate in the past, as Dennard's scaling is over.

PUE progress has stalled

What is the average annual PUE for your largest data center? (n=669)

● Average annual PUE

- 2.50 (2007)
- 1.98 (2011)
- 1.65 (2014)
- 1.58 (2018)
- 1.67 (2019)
- 1.59 (2020)
- 1.57 (2021)
- 1.55 (2022)

UPTIME INSTITUTE GLOBAL SURVEY OF
IT AND DATA CENTER MANAGERS 2007-2022
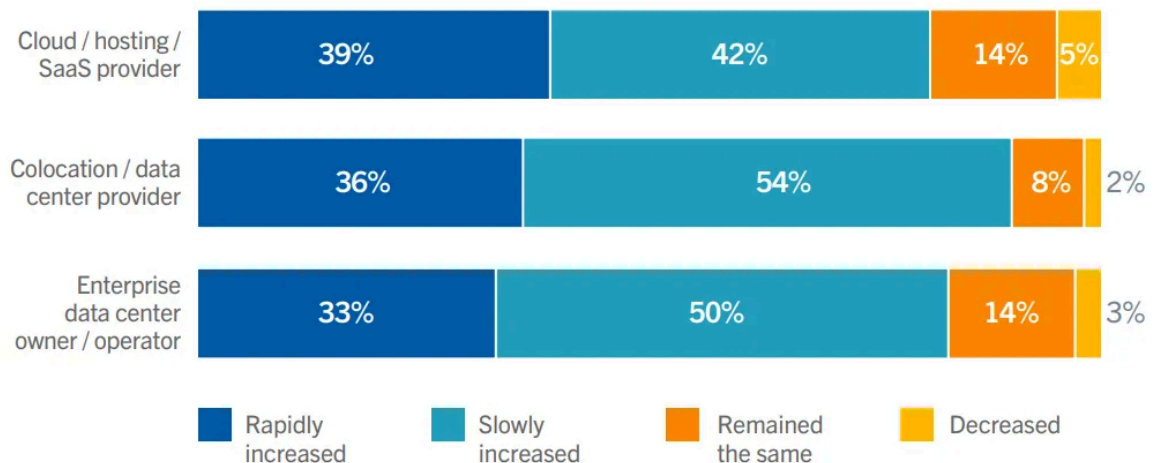
UptimeInstitute® | INTELLIGENCE

Uptime Institute Global Data Center Survey

One of the key drivers is that rack power density is increasing, and the leading edge chips are increasingly becoming more power-hungry, not less.

Figure 2 — Rack power density rising fast across major data center segments

Over the past three years, how has the most common (modal average) rack power density deployed in your organization / colocation data centers changed? (n=611)

Segment	Rapidly increased	Slowly increased	Remained the same	Decreased
Cloud / hosting / SaaS provider	39%	42%	14%	5%
Colocation / data center provider	36%	54%	8%	2%
Enterprise data center owner / operator	33%	50%	14%	3%

UPTIME INSTITUTE GLOBAL SURVEY OF
IT AND DATA CENTER MANAGERS 2022

UptimeInstitute® | INTELLIGENCE

I want to highlight that this is a huge advantage for the public cloud. Public cloud is a secular winner, and as history often shows, challenges tend to accelerate adoption towards secular winners and away from losers. In this case, the benefit goes to the

public cloud's fundamental better PUE ratio. Google reports ~1.1 PUE compared to the average of 1.55. Microsoft and AWS are likely similar.

> Here's the bottom line when comparing the PUE of Google, Amazon and Microsoft:
>
> - Google posts their ratios over the last few years at around 1.11 and now in 2020 at 1.10.[1] This is the best performer among the three cloud providers.
>
> - Amazon is vague and does not publish the exact PUE. It has as a footnote on one of its sites citing PUE at under 1.2. That was in 2014.[2] It is unclear of its past or current overall efficiency.
>
> - Microsoft, through its sustainability reports, does not publish its PUE other than citing that it is improving its ratio score.[3]

Source: [Which Cloud is the Most Environmentally Friendly?](#)

Colocation services like Equinix or private cloud cannot compete. Here's a very simple calculation of the net difference in Public cloud PUE versus the average PUE of a data center.

Public Cloud	PUE of Public Cloud	PUE of Datacenter
PUE	1.15	1.55
Total Energy	100	100
IT Energy	86.96	64.52
Overhead Ratio	13%	35%
100 MW Calc		
MW to Compute	86.96	64.52
MW to Overhead	13.04	35.48

The cost advantage accrues to the public cloud meaningfully. Public cloud companies are additionally using Purchase Power Agreements (PPAs) to create private, clean energy power agreements that are likely better than the market rate. [North American PPA prices rose approximately ~30% YoY](#), which corresponds with similar reports from Barclay's estimating Equinix's energy cost has gone up 30% YoY.

This should spur further cloud adoption, and Satya Nadella mentioned this.

> The best way to hedge against an energy cost and be, in fact, **more energy efficient is to move to the cloud**. So that's where I think a little bit of -- as you all think about what happens to cloud, for us, we look at this and say, **this is a period**

> **where cloud is going to gain share** because we're still in the early innings of adoption

Now I'm going to take a crack at the total energy costs of Microsoft and AWS based on the first comments just for fun.

Quantifying Energy in the Cloud

I'm making some quick assumptions here, but I used their callouts to guesstimate the energy costs of the clouds for Azure and Amazon.

Total Energy Cost	Microsoft
Cost Increase	30%
Quarterly Cost Impact	250
Total Implied Annual	**3,333**
% of Total Opex	*2.8%*

Total Energy Cost	Amazon
Cummulative Cost Increas	125%
LTM 200 Bps of Revenue	1,530
Cost Base	1,224
Total Cost	**2,754**
% of AWS Opex	*5.1%*

The problem is there are some real comparability issues. Amazon is giving you a trailing estimate, while *Microsoft is giving you a forward estimate.* I think these estimates roughly make sense if we account for the Amazon cost trailing while Microsoft is forward.

In the future, I think it's pretty fair to expect a 30-40% increase in this cost base for the hyperscalers. It's just inevitable that prices are rising higher in energy than elsewhere. These will be headwinds on the company's margins, but customers should be adopting cloud services faster, all else equal, given that private datacenters are less energy efficient. Funny how the scales balance out.

But looking at the companies from just PUE of datacenters and the total cost is slightly unrealistic. That's just the energy they purchase and use; what servers they use also matters quite. Think of everything I've mentioned so far as a source of energy analysis, but energy use is also important.

It's why AMD will have a meaningful share gain next year. [Dylan at Semianalysis does a great job breaking out the TCO difference between the higher core and lower TCO](#)

[Genoa / Bergamo generation over past generations](#). It's why AMD will take share next year, but that's just in x86. Because I have an interesting segue into ARM — surprisingly enough, it starts with Oracle's IaaS offering.

The ARM Adoption Acceleratant

One of the weirdest rabbit holes I have recently gone down is the Oracle IaaS offering. If you pay attention, it was the fastest-growing IaaS provider this quarter.

Oracle Announces Fiscal 2023 First Quarter Financial Results

- Q1 GAAP Earnings per Share $0.56, Non-GAAP Earnings per Share $1.03
- Q1 Total Revenue $11.4 billion, up 18% in USD, up 23% in constant currency
- Q1 Cloud Revenue (IaaS plus SaaS) $3.6 billion, up 45% in USD, up 50% in constant currency
- Q1 Cloud Infrastructure (IaaS) Revenue $0.9 billion, up 52% in USD, up 58% in constant currency
- Q1 Cloud Application (SaaS) Revenue $2.7 billion, up 43% in USD, up 48% in constant currency
- Fusion ERP Cloud (SaaS) Revenue, up 33% in USD, up 38% in constant currency
- NetSuite ERP Cloud (SaaS) Revenue, up 27% in USD, up 30% in constant currency

Oracle IaaS grew 58% in constant currency, faster than Azure (42% CC), GCP (38% in USD), and AWS (27% in USD). It's growing on a very small base, but Oracle accelerated meaningfully from last quarter's growth of ~39% in constant currency. What gives?

I think it's partially Oracle's raw pursuit of pricing over everything else and being all-in ([and a significant investor](#)) on Ampere. They are beating out pricing for AWS and Azure meaningfully by ~10-15% on their micro instances, and I presume the reason for this has to be Ampere and, by extension, ARM. AWS, for example, says that Graviton (their ARM-based CPU) uses 60% less energy than a comparable EC2 instance. This advantage should be more important now that energy costs are increasing as a total part of the pie.

Oracle is likely not winning higher value add services, but in bare metal instances and VMs that are containerized and can be deployed on any cloud, the flows should always go to the lowest cost. In this case, because of Ampere instances, it will be Oracle. This should further push the adoption of in-house ARM offerings at hyperscalers.

I'm not saying that Oracle is about to become a dark horse for 3rd place in the IaaS race, but I am saying that this is a great proving point for ARM. [Despite the increasing](#)

[licensing problems from ARM](#), it seems that this recent rise in energy costs *should accelerate* ARM-based instances in the cloud. ARM is getting a lucky tailwind.

What about RISC-V?

Whenever I talk about ARM, this inevitably always comes up — So I want to address it ahead of time. RISC-V will be something someday, but today it is the vast minority of chips produced today. **ARM was the ISA of the future ten years ago (it was one of my first stocks), but ten years later, it's still gaining traction**! Technology platform transition takes a long time, often much longer than we think.

For RISC-V to "be a huge winner," there needs to be some volume production of RISC-V in the datacenter today. That doesn't exist. When I see it, I'll gladly welcome the age of RISC-V, but even today, the industry is still primarily compromised of x86 datacenter chips. There's a lot of wood to chop for ARM, let alone RISC-V. That has some time yet.

If you enjoyed this free piece - I ask you to share or subscribe! Any bit of support makes a big difference to me. Thank you - I'll have my (behind paywall) post about this week's earnings up soon.
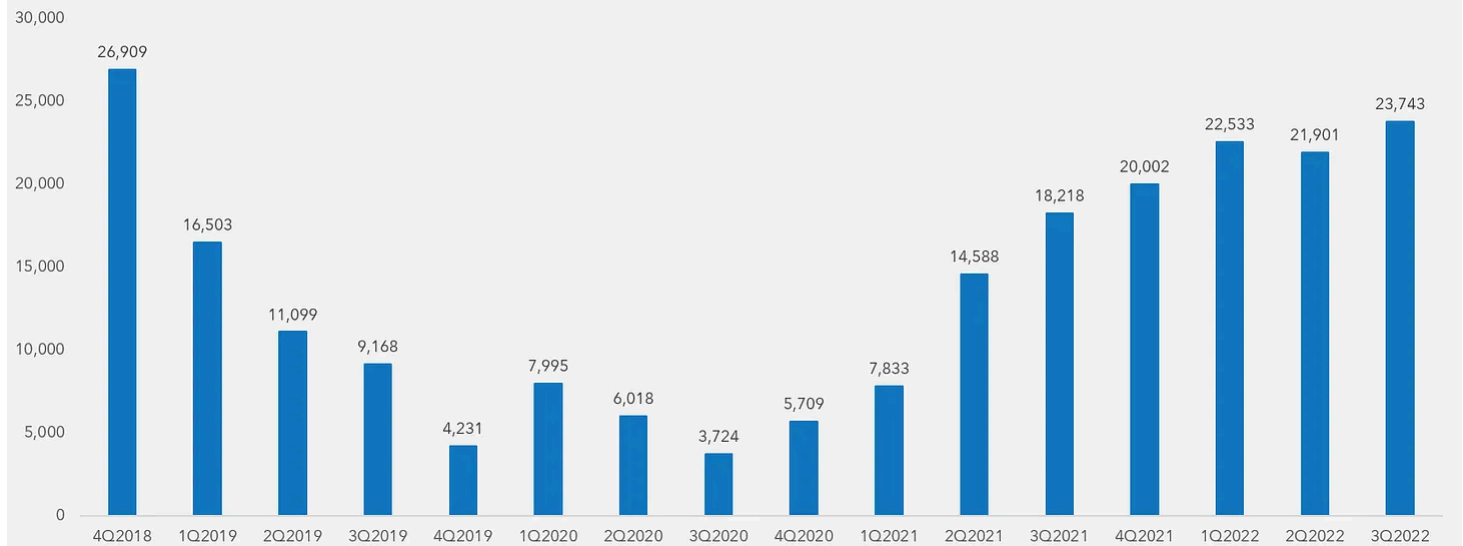
Type your email...	Subscribe

Since I updated the data sources, I thought I should share some charts.

Charts! and Tweets!

Hyperscaler YoY Growth of Capex %

Legend: AWS Capex, Google Capex, Microsoft Capex, Facebook Capex

Hyperscaler Total Capex – Dollars Added

Quarter	Value
4Q2018	26,909
1Q2019	16,503
2Q2019	11,099
3Q2019	9,168
4Q2019	4,231
1Q2020	7,995
2Q2020	6,018
3Q2020	3,724
4Q2020	5,709
1Q2021	7,833
2Q2021	14,588
3Q2021	18,218
4Q2021	20,002
1Q2022	22,533
2Q2022	21,901
3Q2022	23,743

Hyperscaler Capex Absolute Dollars Added (LTM – LTM 12 months Ago)

Legend: Amazon, Google, Microsoft, Facebook
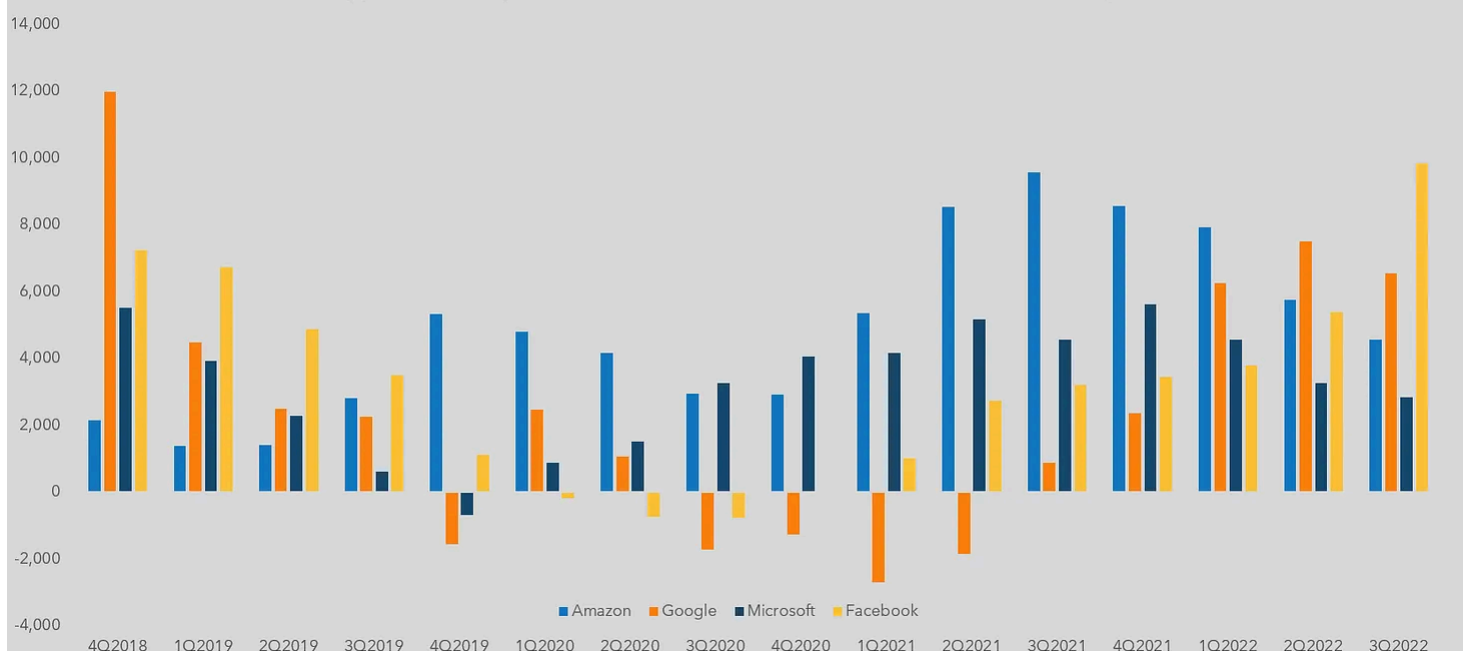
Last but not least, I wanted to recreate this legendary tweet that Modest Proposal made, noting that 30% more capex than the wildly ambitious Alphabet is pretty impressive.
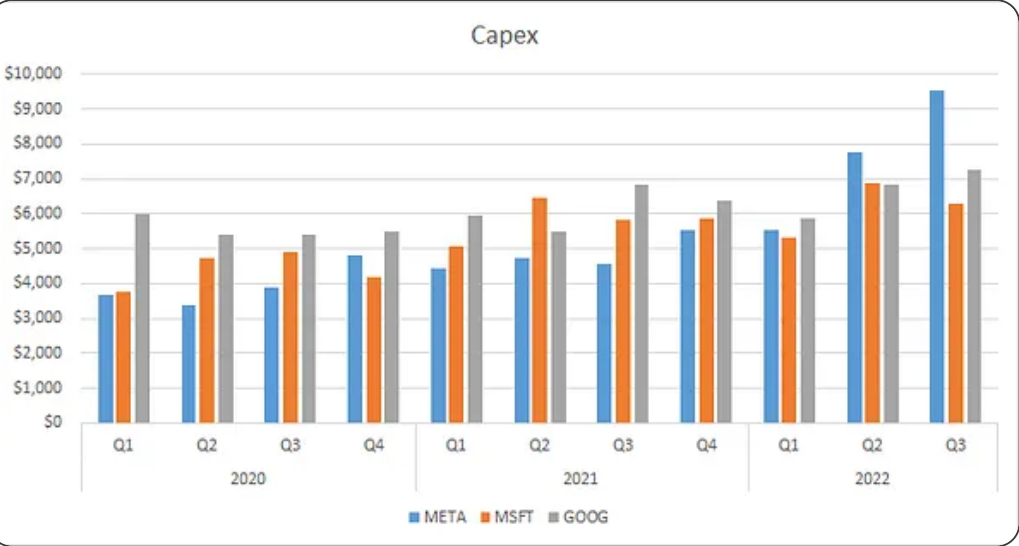

I also wanted to note that META might take the all-time quarterly spending record at the current rate.
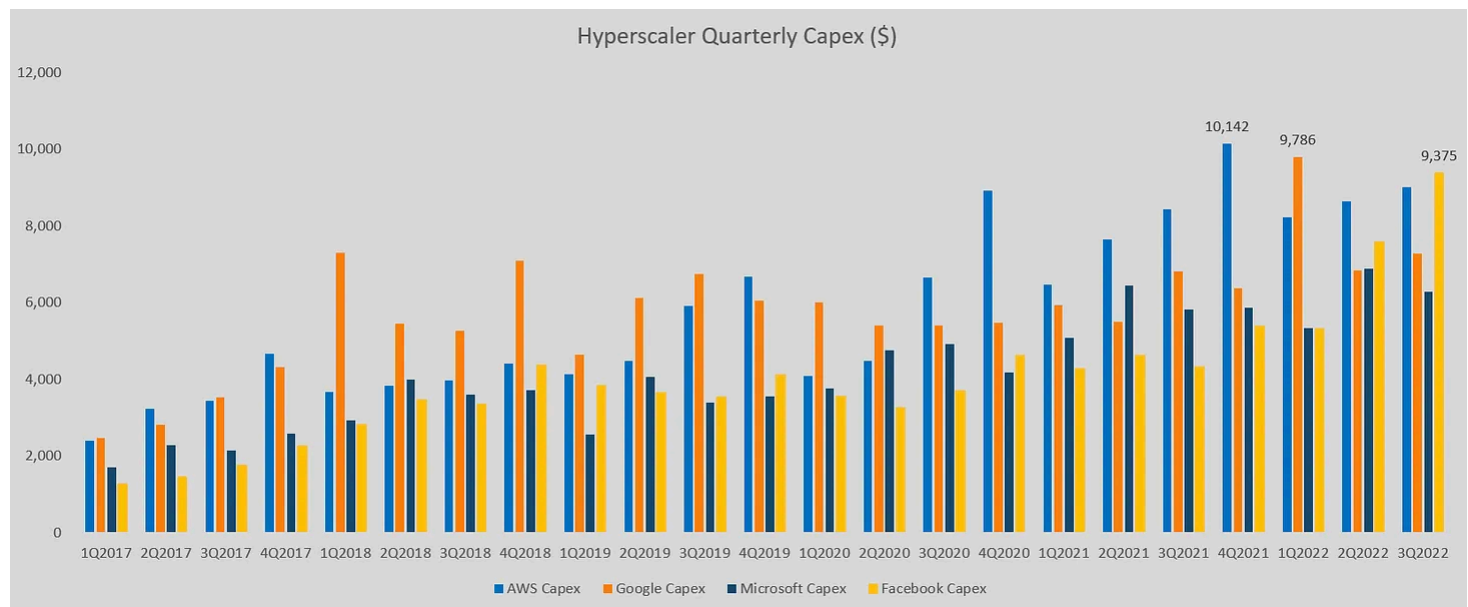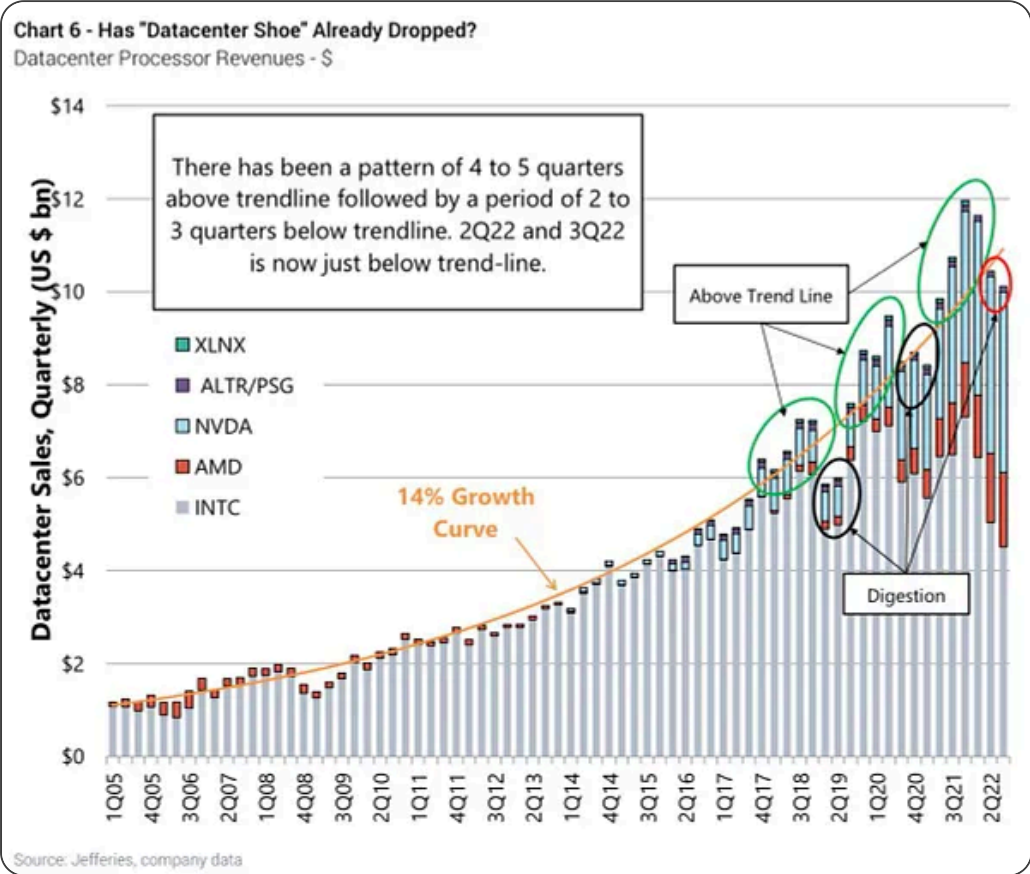
Last - my graph about the absolute dollars kind of hints at it, but it looks like the second derivative of data center spending is stopping out, and it's time for a digestion period. This great set of charts shared by @nuancerocket puts it into perfect perspective.

nuance rocket
@nuancerocket

data center digestion $intc $amd $nvda

Chart 6 - Has "Datacenter Shoe" Already Dropped?
Datacenter Processor Revenues - $

> There has been a pattern of 4 to 5 quarters above trendline followed by a period of 2 to 3 quarters below trendline. 2Q22 and 3Q22 is now just below trend-line.

Above Trend Line

- XLNX
- ALTR/PSG
- NVDA
- AMD
- INTC

14% Growth Curve

Digestion

Source: Jefferies, company data

12:18 AM · Oct 30, 2022

171 Likes   29 Retweets

32 Likes

Previous   Next

Discussion about this post

Comments   Restacks

Write a comment…

V P Nov 8, 2022   ...

♥ Liked by Doug O'Laughlin

Excellent article! From a long term point of view I worry about the rise in ARM for AMD. Sure they will win next year or the year after next as they gobble up share from Intel. But eventually x86 share will go down? Do you agree with that? With that in mind it worries me to be a AMD investor (longer term).

what is the best vehicle according to you to invest in ARM rise and RISC-V?

♡ LIKE (2)   💬 REPLY   ↥ SHARE