Time to redefine Hybrid HPC

how and why you can get the best of both worlds through HPCaaS

atnorth
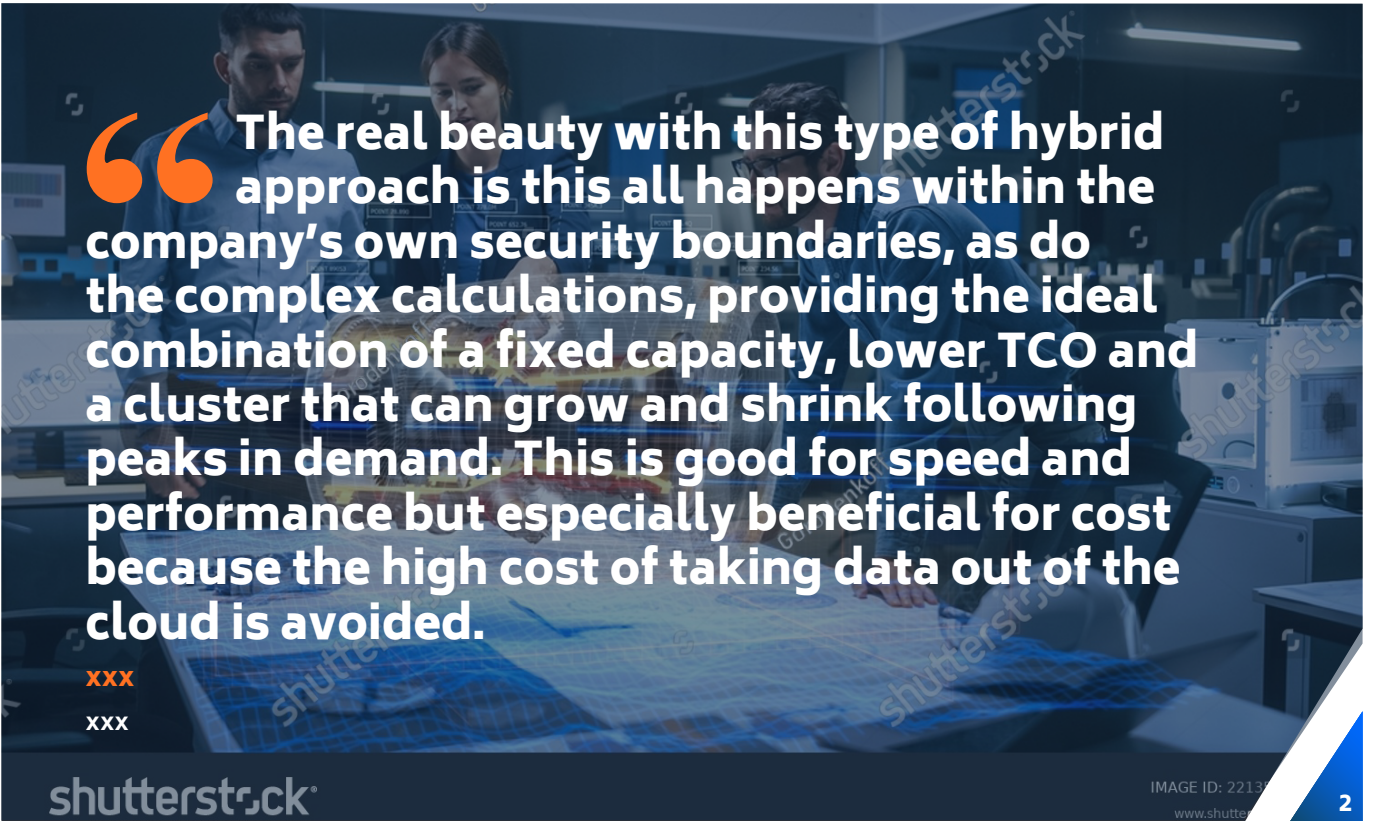
Meet the hosted HPC solution that feels like on-prem

High Performance Computing (HPC) is at the cornerstone of the new world of innovation and more and more combined with AI and machine learning systems that empower businesses to make datadriven decisions in real-time, while empowering engineers and scientists to test models and build a better tomorrow.

Businesses in every sector know they need the speed and power of high-performance computing and so the question has now moved on to not if, but rather how, they should be set up. Should clusters be kept inside the company's on-premises data center, for the ultimate control and performance? Or is the convenience of switching to the cloud, and paying monthly on usage, rather than upfront on equipment, too irresistible to avoid?

As with all decisions, the answer is never easy and there are pros and cons on either side. Many of these do not become apparent until a company has signed a deal to buy in new equipment or has signed up with a cloud provider. That is why it is crucial businesses carry out thorough research and, when they do, they will likely find that a new hybrid solution is the best choice.

The term hybrid is usually used to describe a solution that is split between the public cloud and onpremise. It promises the benefits of both, without the drawbacks, but all too often companies will find having a foot in both camps is expensive and inflexible, with a large impact on performance and security.

This is why the best hybrid solution is not to split on-premises and cloud but to instead to use a private cloud cluster that operates and feels like it is on-premise but has the benefit of being run by HPC experts, saving a business time and expense, while maintaining performance. Such a hybrid solution needs to offer complete flexibility so spikes in activity can be accommodated through extra capacity being made available in just minutes.

> **" The real beauty with this type of hybrid approach is this all happens within the company's own security boundaries, as do the complex calculations, providing the ideal combination of a fixed capacity, lower TCO and a cluster that can grow and shrink following peaks in demand. This is good for speed and performance but especially beneficial for cost because the high cost of taking data out of the cloud is avoided.**
>
> **xxx**
>
> **xxx**

Whitepaper: Time to redefine Hybrid HPC

The pros and cons of the public cloud

To truly understand why hybrid should now be defined as HPCaaS, companies need to examine thepros and cons of on-prem and public cloud. When the pros are combined, and the cons ditched, they will find themselves with only one solution – HPCaaS.

It is true that moving from on premises to the cloud shift cost from CapEx to OpEx, which many businesses will welcome. However, it is not always true that the monthly usage bill will be as predictable as thought, particularly if teams roll out new APIs, commission new servers and, most importantly for high performance use, if they need to move large datasets out of the cloud.

In fact, there is a long list of cost issues associated with a move into the public cloud:

Data Exchange: High performance computing applications and AI need a lot of data to be exchanged, it is difficult to predict how much. The cost of Ingress/egress can go through the roof very quickly. Data gravity becomes a bigger problem because it is easy to upload data but expensive to download.

Complex pricing: Finding the best optimized services from the vast choice on the public cloud requires a lot of experience or a lot of experimentation which can quickly get very expensive and is often unproductive.

No Standardisation: Each cloud provider uses its own proprietary way of how complex layered services need to be combined so it takes a great deal of multiplication of effort to implement multicloud or Hybrid cloud.

Cost control: Cloud is ideal for very unpredictable, sporadic projects but many small workloads, which need to scale capacity up and down, are not suited to how massive HPC workloads can run for days or weeks. As cloud is priced per minute/hour at a very high price, constant use becomes very expensive. The many items on a usage bill makes good cost control difficult.

Fast evolution of cloud portfolio: A lot of new services are constantly added which makes it difficult to keep up and choose which are optimal without a lot of experimentation or consulting experience being added.

Performance loss: To achieve fast speed in provisioning, almost all features are virtualized and so shared in the public cloud. This leads to unnecessary performance loss. As most resources are shared, very unpredictable performance bottlenecks can appear from other heavy users or workloads

There is a potential partial answer to some of these, but for those trying to build simpler, low-cost systems, it is not a welcome one. To control IaaS and PaaS cost, to guard against overspending, organizations must develop new financial management processes. For heavy computing services these need to include engineering, design and applications teams and possibly their cloud experts within an IT team, not to mention the multiple end users of each application of tool being used in the cloud.

Ironically, this can mean an attempt to simplify the management and IT function within a business can actually lead to new financial oversight requirements, creating more work for executives within a business.

Drilling down - Drawbacks of HPC in the 'public' cloud

There are some hidden costs of using cloud computing and these only become more obvious when businesses are involved in High Performance Computing really understand the day-to-day cost of operating in the cloud, and the potential performance issues.

Cost: Cloud HPC vs on premise

Equipment and running costs are very likely to be lower for a well-designed local cluster than consuming it from the cloud but offering only fixed capacity and no flexibility. How much so, depends on the type of cloud instances a business wants to create within its HPC Cluster. However, as anyone who has ever requisitioned equipment for a data center will tell you, it can take weeks or even months to purchase additional equipment for the short burstsof activity many HPC users need.

3

Scenario 1 - Bursting to the cloud

Typical HPC users will need to burst activity from time to time, perhaps to run a simulation for amends to a new car design or for some other intensive additional use. A typical approach here is to use on-demand instances to create a cluster in the cloud. When the job is finished, those instances can be deleted so they are no longer being paid for. Sounds cheap, but the cost can still be up to five times higher than that for a local cluster.

Pros – On demand allows flexibility to 'burst' into the cloud for extra capacity

Cons – It is up to five times as expensive as on-premises and a factor more expensive than scaling to more capacity in HPCaaS.
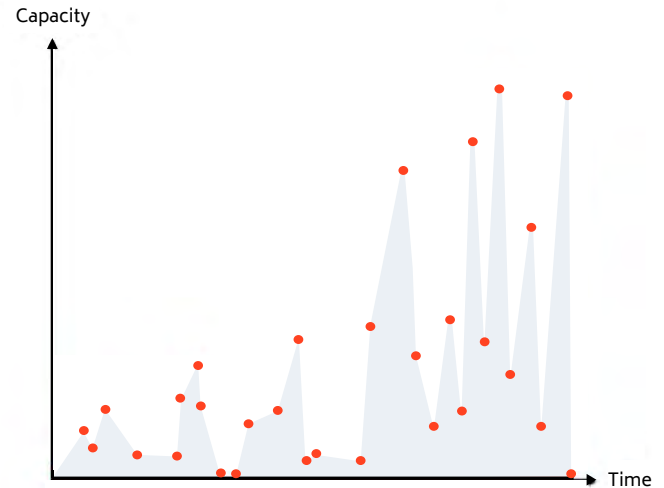
Scenario 2 - Sign a longer-term contract

Typical HPC users will need to burst activity from One potential answer to gaining extra capacity is to sign a long-term contract, for a year or more, with a cloud provider. However, this is still likely to be more expensive than a locally hosted cluster.

Pros – A long-term contract gives assurance of capacity for a fixed duration.

Cons – More expensive than an on-premises solution and more expensive than HPCaaS because the long term Public cloud pricing is based on a discounted highest price of On-demand which is a lot more expensive compared to HPCaaS.

Scenario 3 - Opt for a spot

The realization long term contracts can be expensive, and on-demand is even more costly, may tempt some to turn to the spot market. It is a cheaper alternative, and it avoids being tied into a lengthy deal.

The big drawback, though, is timing. Spot buyers must be prepared to wait for the type of HPC cluster that meet their requirements to become available. Even then, there is no reassurance if the extra resource will be available for a sufficient period. There is always the risk capacity to be run out before a major piece of work is completed.

On the positive side, the costs of spot instances are roughly two thirds lower than on-demand instances and closer to local cluster costs. However, the need to wait for the right type of capacity to open up, and the possibility use may be interrupted, means this option is rarely used for Enterprise HPC executives. When business critical compute requirements are very large and have to run uninterrupted, it simply does not make sense.

Pros – Spot capacity is roughly the same cost as on-premise (cheaper than ondemand cloud)

Cons – User must wait for the right capacity to become available and there is no guarantee it will be available for the entire duration of a work load

4

HPC data costs – Cloud vs On Premise

The huge amounts of data used in HPC, cannot be moved around easily. Even with the fastest possible cloud connections it takes time to download data for post-processing, and this would typically be for just one HPC job. The best solution is often to keep the data in the cloud and post-process on GPU cloud instances with Remote 3D access acceleration tools. That is because the sheer volume of data means it can be slow to move data back to a local on-premises network and, because of data egress, it is also very expensive.

Exporting data costs – Cloud vs On Premise

It is not always the first thing a business will come across in a sales brochure or proposal but the cost of moving data to and from the cloud back to a local, on-premise network is high. Customers will usually find the cost of putting data in the cloud is relatively low and so can then be surprised at the large bill for getting it back out again. This adds to the Data Gravity of the cloud which can be exponential for large HPC & AI data: once the data is in, it becomes more difficult to get it out. It is typically faster and cheaper to do post-processing in the cloud.
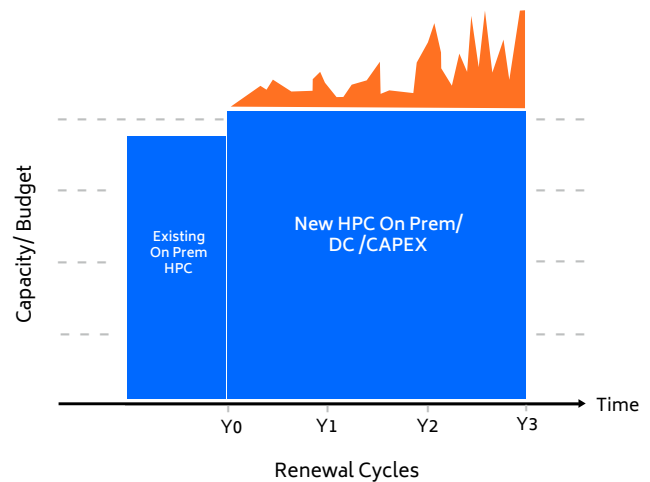
Performance, management and security – Cloud vs On Prem

Overheads cost dear - The performance of any HPC cluster will always rely on its configuration. Cloud installations are usually based on virtual machines (vCores) and are even used with hyperthreading turned on. It depends on the overhead of the virtual machine infrastructure, but most real life HPC benchmarks have shown that around 20-25% of the performance is lost to this virtualized overhead. There may be further performance limitations to factor in if a High-Speed Interconnect infrastructure is not available.

Time delays - Latency will typically be higher in a cloud cluster, meaning larger models with a higher number of cores will not scale as well over several nodes. This will not always be the case, because in some cloud offers high speed interconnect infrastructures are available or substitutes based on standard TCP with an optimized network stacks and additional hardware. However, businesses need to be aware latency can be an issue in the cloud.

ISV Licenses – terms and network access – Many companies run their HPC clusters through Independent Software Vendor (ISV) licenses. These can be expensive and can even cost more than the hardware they run on. It would be natural for a curious IT exec to wonder if they can get a better deal operating in the cloud. However, the trouble is, many ISV licenses have inflexible rules that can prevent a smooth roll out.

ISV licenses – splitting issues - There are also ISVs that have floating licenses, managed by a license manager. These can be used in a Cloud infrastructure, provided they always have a connection to the license manager via a secure network. Sometimes companies can also split a license pool in two parts, each with its own license manager: one sub-pool for use in the cloud and another for a local cluster. This may come at the cost of restricted flexibility and the cloud element often attracts a higher fee.

On-prem and cloud HPC – best or worst of both worlds?

It is clear there are pros and cons of using HPC in the Cloud. Often it is not attractive to move all HPC workloads to the public Cloud to take advantages of the benefits due to the security, cost (particularly taking data out of the cloud) and complex additional management oversight. One of the biggest reasons for not committing to the public cloud, especially when using the more affordable 'spot instances', is the inflexibility of only being able to scale up when the hardware, configured to a company's specific needs, become available – and still only for an unpredictable period.

This is what make a hybrid approach so appealing. Rather than choosing between the control and scalability of on-premises and the managed service paid for by OpEx in the cloud, hybrid appears to offer the best of both worlds. Combining the benefits of both approaches, yet in one secure location, it has the benefits of cloud and on-premises, without the drawbacks.

A typical approach is "Hybrid HPC": move part of the HPC workload to the Cloud most suitable for HPC resources in the cloud e.g. for occasional users and keep part of it as a stable well designed HPC cluster, storage and network well integrated with all your applications, and workflows that sit on an on-premises network.

Four problems with traditional hybrid approach

Security concerns – By putting some clusters in the public cloud and keeping others on-premise, security becomes a very obvious concern or needs specific attention. A standard technique is to a VPN (Virtual Private Network) to connect the cluster in the cloud to the on-prem solution. However, many security experts will agree that using ssh (secure shell) to login to the cluster not ideal and so many organizations find they usually have to also add end to end encryption. This can have serious implications for both cost and speed. It also requires complex networking policies to be put in place or set up the Cloud/hosted environment.

Extending brings problems - It looks attractive to extend the existing on-prem HPC cluster into the Cloud but there are several issues related to this approach. The first is related to dynamic enlarging and shrinking of the HPC cluster in the Cloud. The scheduler used in the HPC Cloud environment typically needs to be able to cope with dynamic re-sizing of the cluster to give a team extra capacity for a large job, so called 'cloud bursting'.

File and data sharing problems - In a local data center, a company will most likel have a shared file system that can be used by all HPC jobs. When a job requires input data, it simply gets it from that file system. Results are then written back to the file system at speed because the connections within an internal data center are fast. Conversely, the network bandwidth connecting to an HPC Cluster in the Cloud is much slower and the cost to synchronize data can be high. This decrease in performance, when sharing a file system into the cloud, is a major challenge that puts many companies off making the move.

Cloud issues still persist – We have already discussed the pros and the cons of using the public cloud and, of course, these still apply when it is part of a ecosystem that connects to on-premises. Due to data transfer complexities, license integration, dynamic allocation of cloud resources into the HPC cluster, not to mention day-to management of two systems, the result can be companies need to put in twice the effort, for having two installations, and that is reflected in cost too. Rather than solve a problem, companies can simply add to them.

Why sustainable HPCaaS is the answer

It is clear that the greatest control, the best performance at a cost-effective price comes from onpremises. But this requires upfront CapEx budgets to be deployed to systems that are subsequentlyhard to scale up.

Bursting into the cloud, or moving the entire HPC cluster into the cloud, might seem the answer but there are security and licensing issues to deal with and a degradation in performance that puts off many businesses.

Trying to co-locate the two options, to build what is traditionally called hybrid would appear to make sense. However, it simply combined the challenges offered by both options. Gaining extra capacity in the cloud for big jobs is unpredictable, time-consuming and expensive.

The clear answer has to be hybrid, but not as most people know it. Instead of relying on the public cloud, with all its scaling and performance issues, the smart new answer is a private cloud – but one that is in a very sustainable place, very energy efficient, using 100% renewable energy and treated as if it is on-premises from a security point of view.

That means all the compute power is set up in what is effectively an off-premise energy efficient data center so performance and speed is not compromised. The big jobs are simply run on this remotely hosted data center, or private secure cloud if you prefer, and then executives can log on to its servers to check on results.

This gets around the large cost of exporting data from the cloud back to corporate networks as the HPCaaS can sit literally on your private secure corporate network within your firewalls. If the big computing jobs are done in the HPC cluster and the results can be stored there for further processing, so there is no need to sign endless checks for data egress. Crucially, users can be set up to access the private cloud with the same credentials they use to log onto their local, corporate network. It saves on delays in logging in to different systems, making operations seamless.

One of the major deciding factors, though, can be a combination of both security and flexibility. Normally, the two are not close bed fellows but with a private cloud solution, the company concerned can set up its security, which it updates and maintains. Then, when it need extra capacity, the systems atNorth uses can simply extend their firewall and security measures to a new cluster.

The hosted company is in charge here. Requests can be made and fulfilled in a matter of minutes and there is no fear the extra capacity will be taken away until the job is finished and the additional cluster can be decommissioned.

Whitepaper: Time to redefine Hybrid HPC

Conclusion – Embrace the new 'hybrid'

The big picture is important for a sustainable data It is time to embrace a new meaning for hybrid cloud which is separate from most peoples' understanding of combining on-prem with the public cloud especially for HPC & AI. Rather than solving challenges, it just compounds and combines the performance, cost, licensing, impact and security issues of both options.

Instead, the hybrid solution companies are now looking for, particularly when they hit a critical mass for their HPC usage, is a private cloud, hosted on their behalf, which integrates seamlessly with their corporate network in a secure way. It combines the benefits of being remotely hosted and expertly managed with the performance, security and reassurance of an on-prem solutions.

On demand capacity

Base capacity

Capacity

Time
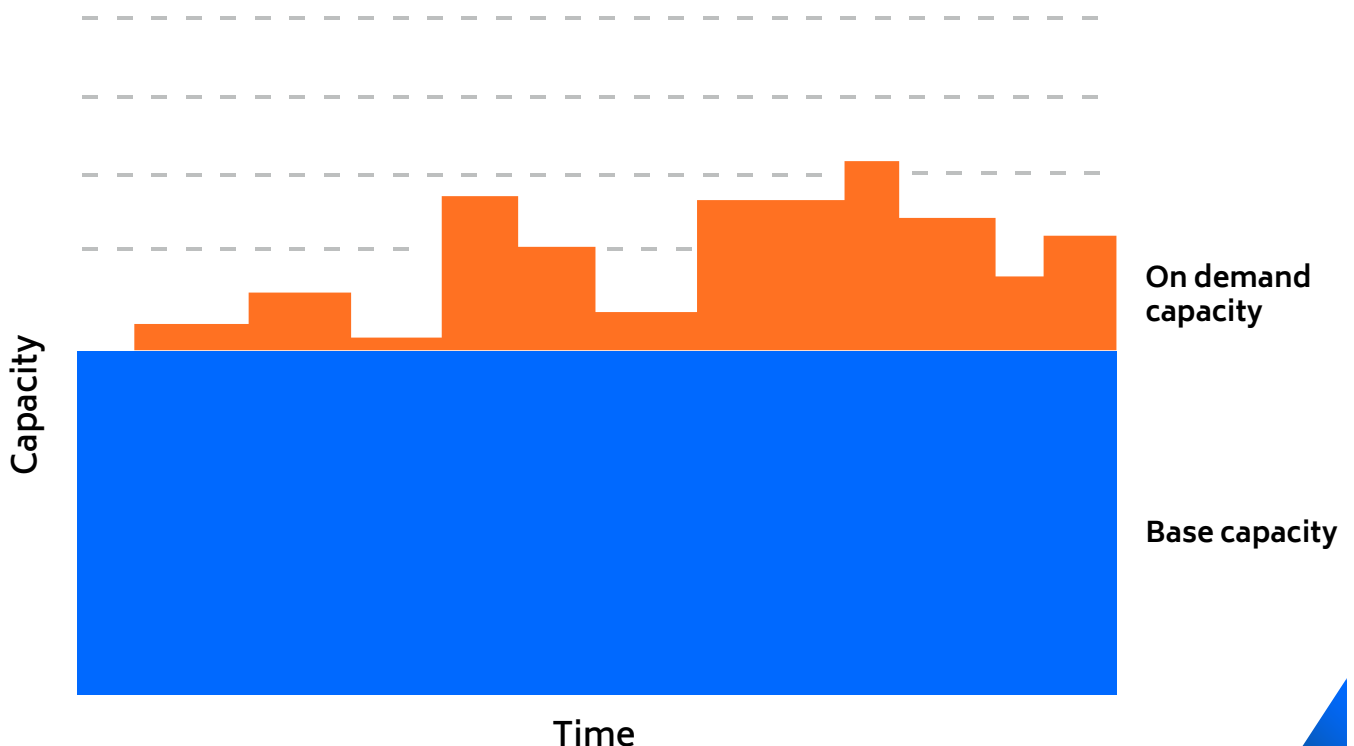
At a glance
Benefits of HPCaaS

Billing – High up-front CapEx on internal clusters is avoided and replaced by a monthly fee paid through OpEx budgets

Cost - High data egress charges are avoided, no need for long term contracts for additional capacity

Performance – Large compute jobs are started and completed within the HPC stack so performance does not suffer from slow connections between on-prem and the cloud Benefits of HPCaaS provided by atNorth

Security – atNorth HPCaaS allows companies to use their security on their cluster so they can be assured it is safe and its firewall can be extended at speed to deliver additional capacity

Flexibility – No need to wait on the spot market for capacity to become available, it can simply be commissioned in minutes

Environment – atNorth data centers are located in cold locations where cooling requirements are low and provided by renewable energy. Any heat generated through the cooling process is then fed into local housing systems to provide free heating.

atNorth is a leading Pan-Nordic data center services company offering sustainable, cost-effective, and scalable co-location and high-performance computing services across Iceland, Sweden and Finland. The company operates six data centers in strategic locations across the Nordics, with a seventh site to open in Finland in 2024.

With sustainability at its core, atNorth's data centers are powered by energy from renewable sources and support circular economy principles. All atNorth sites leverage innovative design, power efficiency, and intelligent operations to provide long-term infrastructure and flexible colocation deployments.

atNorth is headquartered in Reykjavik, Iceland, and is trusted by industry-leading organizations to operate their most critical workloads. The business was founded in 2009 and acquired by Partners Group in 2022.

To find out more, get in touch at:

atNorth Sales

sales@atnorth.com

+354 539 3282

atnorth.com