HPC in the Cloud

How can it help with library development

Arthur Petitpierre – Amazon Web Services
arthurpt@amazon.com – EMEA HPC Specialist Solutions Architect

Who am I ?

HPC Specialist Solutions Architect @AWS

Based out of Paris

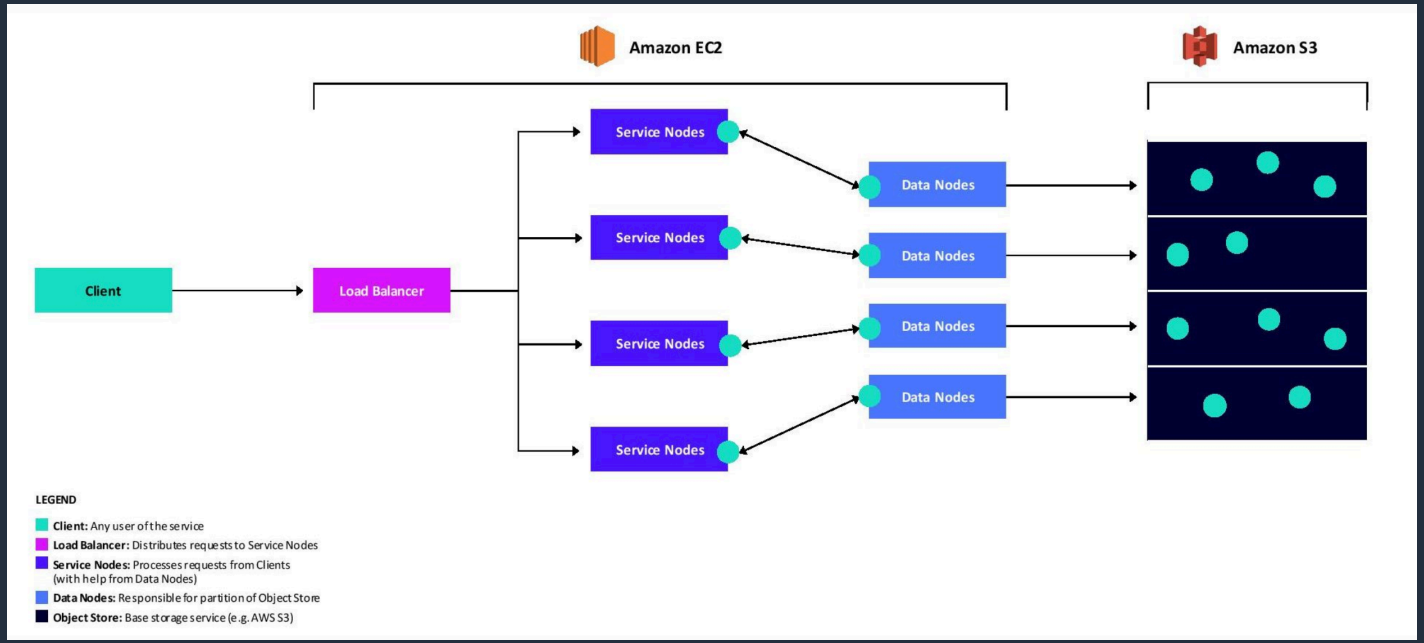
Previously:

- HPC Services CTO @ATOS
- HPC Support Mgr @Bull
- And a few other stuffs…

Enjoy moving Snowballs around Paris on a bicycle

HDFGroup & AWS

HDFGroup Kita architecture – 24K$ AWS credits



Source: https://www.hdfgroup.org/wp-content/uploads/2018/08/HDF-Kita-Architecture-e1533785700337.jpg

Open Data on AWS

Power from wind: Open data on AWS

by Caleb Phillips, Caroline Draxl, John Readey, and Jordan Perr-Sauer | on 20 MAR 2018 | in Amazon EC2, Amazon Simple Storage Services (S3), AWS Big Data | Permalink | Comments | Share

Data that describe processes in a spatial context are everywhere in our day-to-day lives and they dominate big data problems. Map data, for instance, whether describing networks of roads or remote sensing data from satellites, get us where we need to go. Atmospheric data from simulations and sensors underlie our weather forecasts and climate models. Devices and sensors with GPS can provide a spatial context to nearly all mobile data.

In this post, we introduce the WIND toolkit, a huge (500 TB), open weather model dataset that's available to the world on Amazon's cloud services. We walk through how to access this data and some of the open-source software developed to make it easily accessible. Our solution considers a subset of geospatial data that exist on a grid (raster) and explores ways to provide access to large-scale raster data from weather models. The solution uses foundational AWS services and the Hierarchical Data Format (HDF), a well adopted format for scientific data.

The approach developed here can be extended to any data that fit in an HDF5 file, which can describe sparse and dense vectors and matrices of arbitrary

Planning and siting for wind energy requires detailed information about long term historical weather trends and patterns. Wind turbines paired with agriculture is an increasingly common sight in the Midwestern and Central United States.
Image Credit: National Renewable Energy Laboratory (NREL)

500TB open weather model dataset

Built with HDF5

Can be accessed with h5pyd lib and a REST API

https://aws.amazon.com/blogs/big-data/power-from-wind-open-data-on-aws/

AWS Public Data Sets

AWS Public Dataset Program

The AWS Public Dataset Program covers the cost of storage for publicly available high-value cloud-optimized datasets. We work with data providers who seek to:

1. Democratize access to data by making it available for analysis on AWS.
2. Develop new cloud-native techniques, formats, and tools that lower the cost of working with data.
3. Encourage the development of communities that benefit from access to shared datasets.

You can see examples of datasets supported by the AWS Public Dataset Program on the Registry of Open Data on AWS.

Requirements

To share your dataset through the AWS Public Dataset Program, you must agree to the AWS Public Dataset Program Terms and Conditions, which are available at:
https://aws.amazon.com/public-datasets/terms/

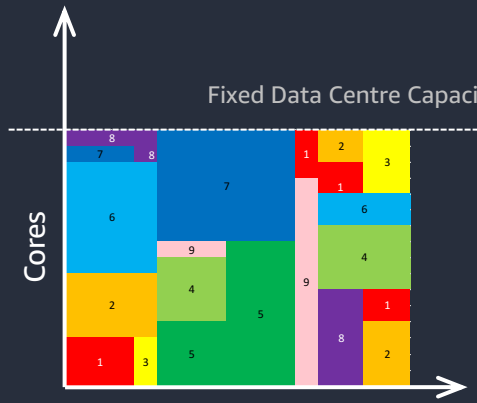
https://aws.amazon.com/opendata/public-datasets/

aws

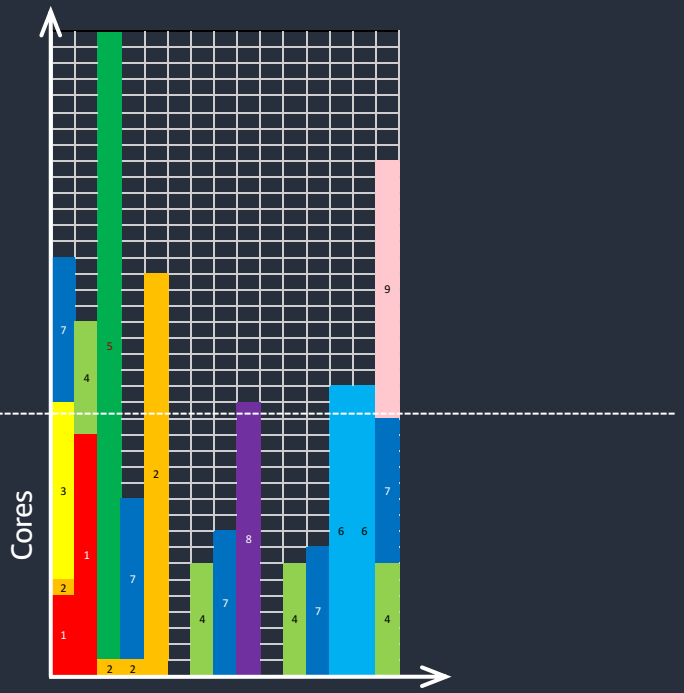
HPC in the public Cloud

aws

The metric for success should be time-to-results



Finite capacity, usually with long queues to wait in.

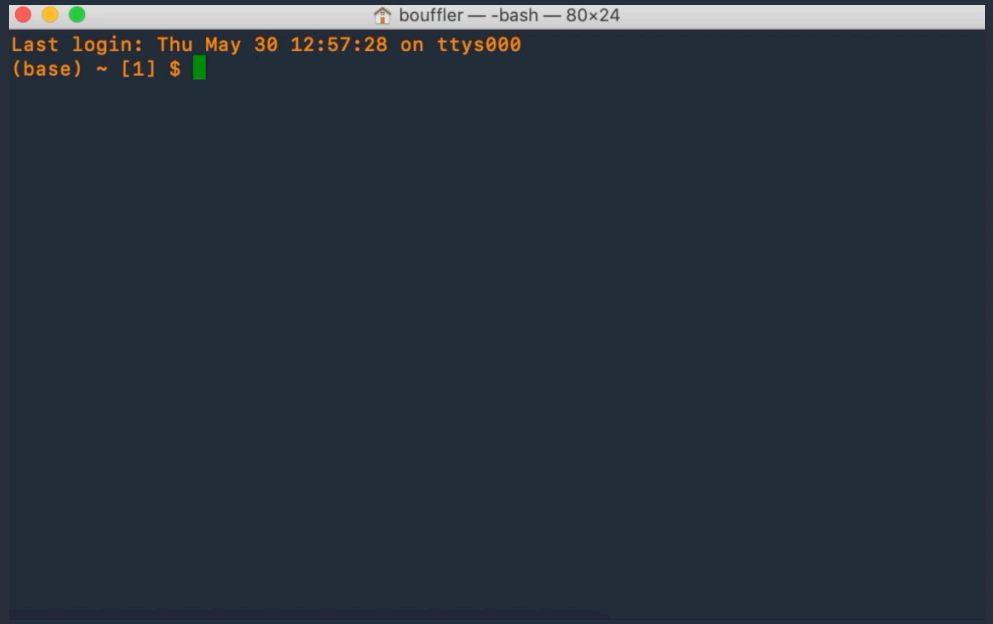


Massive capacity when needed to speed up time to results, and agile environment when additional hardware and software experimentation is needed.

Source: Hyperion ROI Study (http://www.hyperionresearch.com/roi-with-hpc/)

Infrastructure is code.

Not a 5-yearly refresh

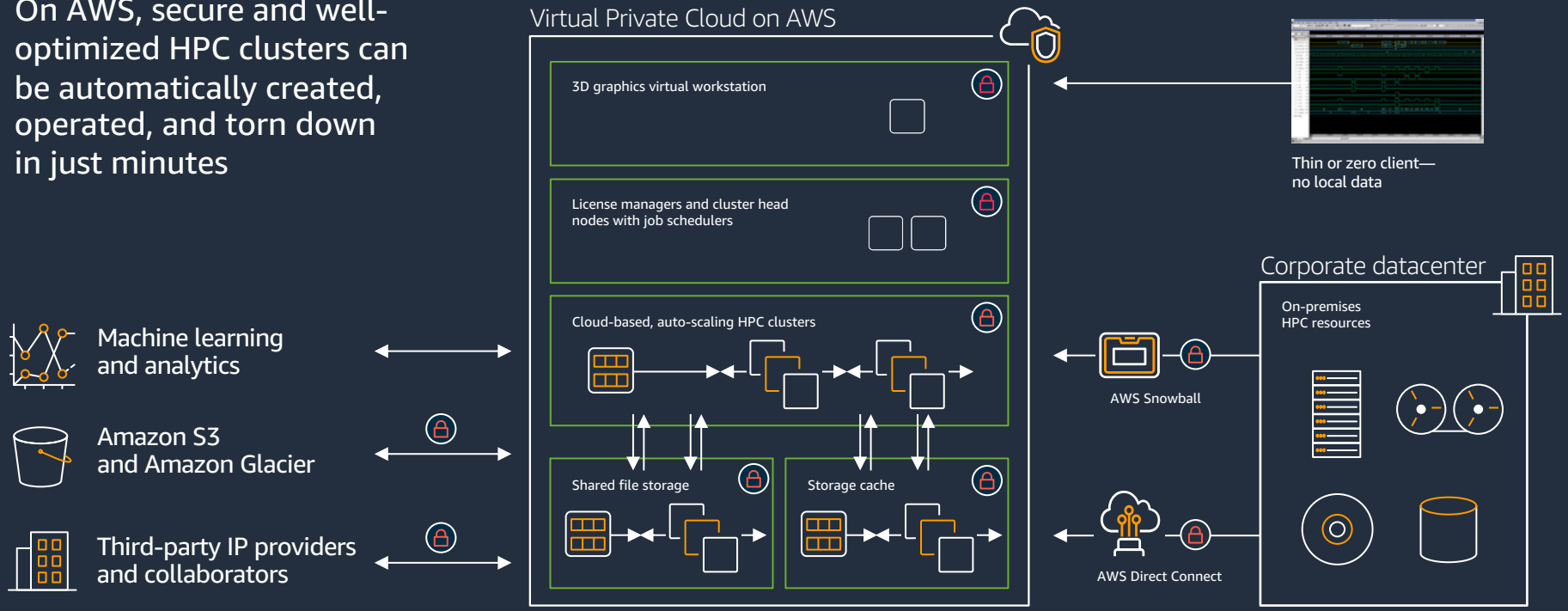


- **Iteratively** decide on the best CPU, GPU, memory or I/O architecture for your workload.
- Test multiple options in **parallel** rather than sequentially.
- **Dispose** of what you don't need (mercilessly, and without harming any animals :-)
- Make **CI/CD** part of your HPC practice.

Link to Tutorial <INSERT>

High Performance Computing (HPC) on AWS

On AWS, secure and well-optimized HPC clusters can be automatically created, operated, and torn down in just minutes



Broadest and deepest platform choice

Categories

General purpose

Burstable

Compute intensive

Memory intensive

Storage (High I/O)

Dense storage

GPU compute

Graphics intensive

Capabilities

NEW! Choice of processor
(AWS, Intel, AMD)

Fast processors
(up to 4.0 GHz)

High memory footprint
(up to 12 TiB)

Instance storage
(HDD and NVMe)

Accelerated computing
(GPUs and FPGA)

NEW! Networking
(up to 100 Gbps)

Bare Metal

Size
(Nano to 32xlarge)

Options

Elastic Block Store

Elastic Graphics

NEW! Elastic Inference

200+

instance types

for virtually every workload and business need

aws

What is Elastic Fabric Adapter (EFA)

Scale tightly-coupled HPC applications on AWS

EFA

Elastic Fabric Adapter, best for large HPC workloads

High data throughput

100 Gbps network bandwidth

Congestion control for cloud scale and rapid packet loss recovery.

Lower latency for message passing and more effective application-layer comms.

aws

Scalable Reliable Datagram (SRD)

A reliable high-performance lower-latency network transport

Inspired by Infiniband Reliable Datagram, without the drawbacks

- No limit on the number of outstanding messages per context

Out-of-order delivery – no head-of-line blocking

- Messages are independent in many cases, application/middleware can restore ordering only if/when needed
- Same motivation as weak/relaxed memory ordering

Packet spraying over multiple ECMP paths

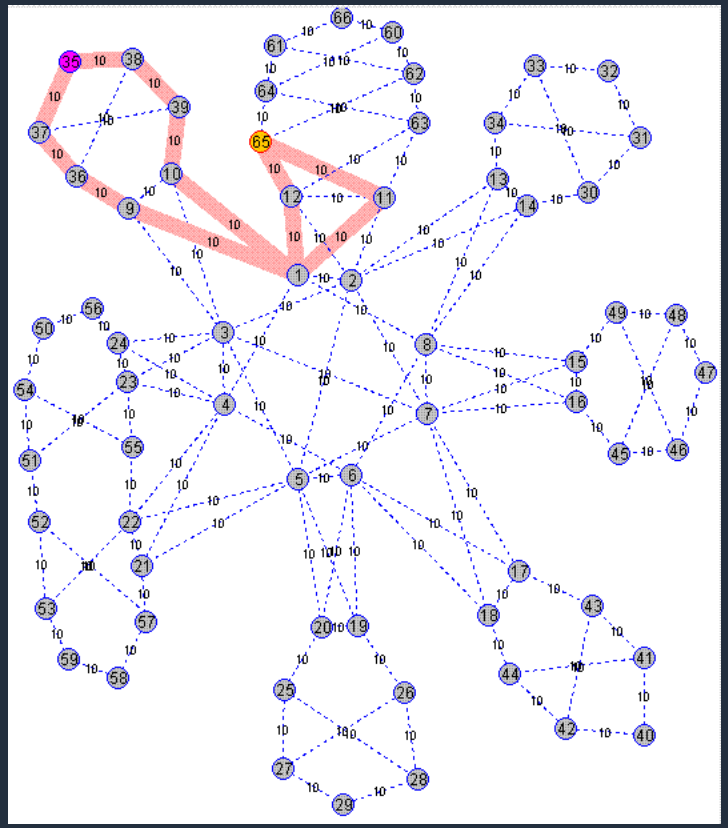
- Rapidly adapt to hot-spots
- Fast and transparent recovery from network failures

Congestion control designed for large-scale cloud

- Maintains high throughput in the face of packet drops
- Minimize latency jitter

aws

Multipath Routing



Equal-cost multi-path routing (**ECMP**) is a routing strategy where next-hop packet forwarding to a single destination can occur over **multiple "best paths"**. This can substantially increase bandwidth by load-balancing traffic over multiple paths.

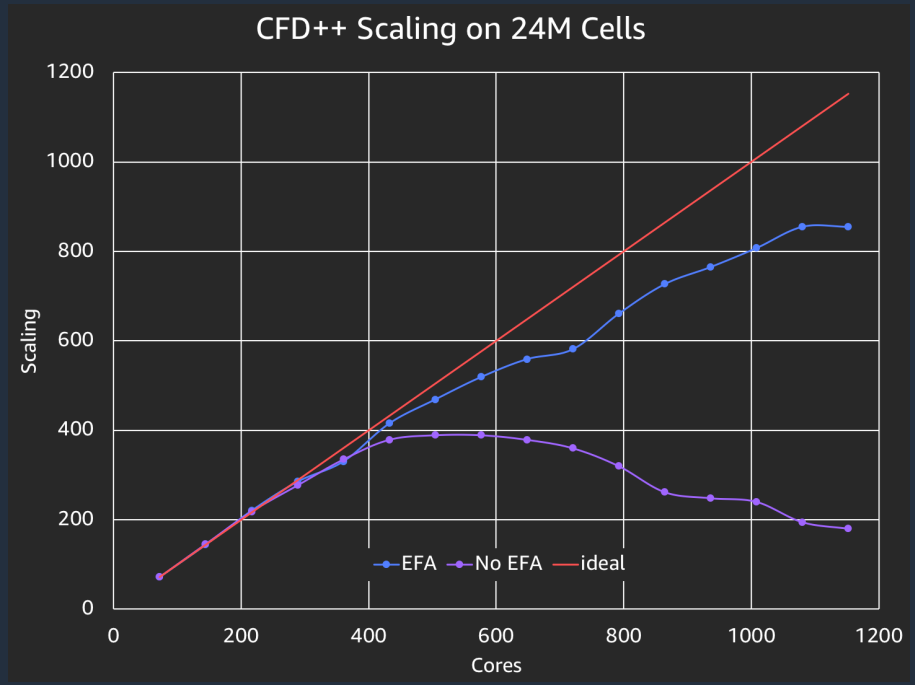
Thanks to Wikipedia and it's contributors for the pithy explanation and Peter Ashwood-Smith for the snappy animated GIF explaining the concept.

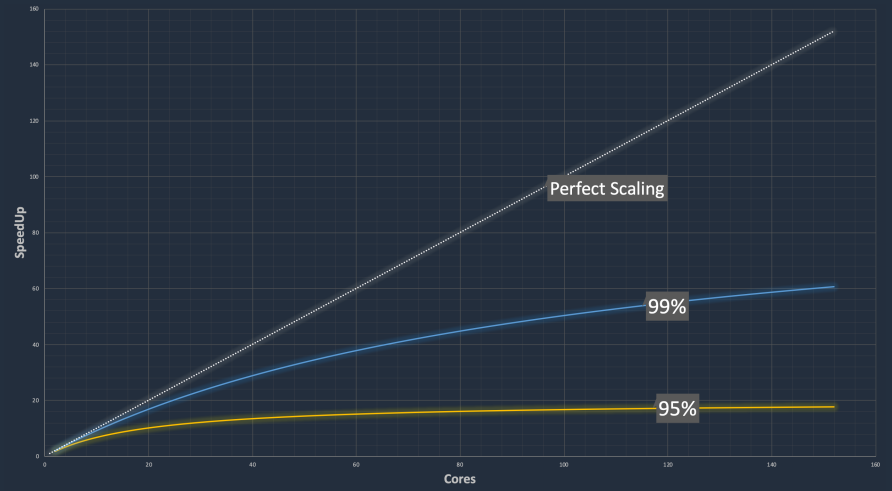
TCP vs Infiniband vs SRD

TCP	Infiniband	SRD
Stream	Messages	Messages
In-order	In-order	Out-of-order
Single path	Single (ish) path	ECMP spraying with load balancing
High limit on retransmit timeout (>50ms)	Static user-configured timeout (log scale)	Dynamically estimated timeout (µs resolution)
Loss-based congestion control	Semi-static rate limiting (limited set of supported rates)	Dynamic rate limiting
Inefficient software stack	Transport offload with scaling limitations	Scalable transport offload (same number of QPs regardless cluster size)

aws

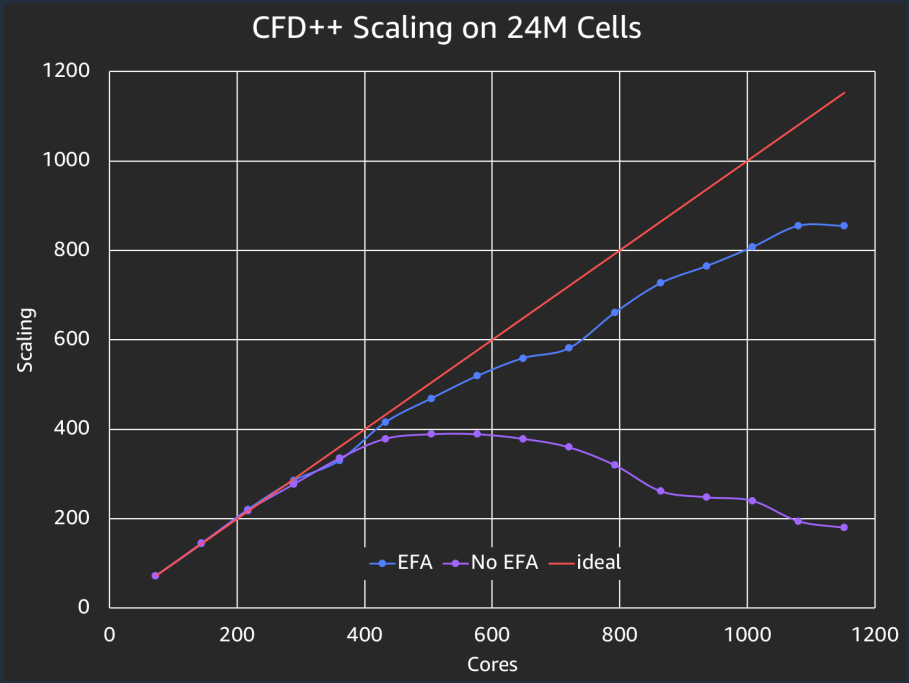
What can EFA do?

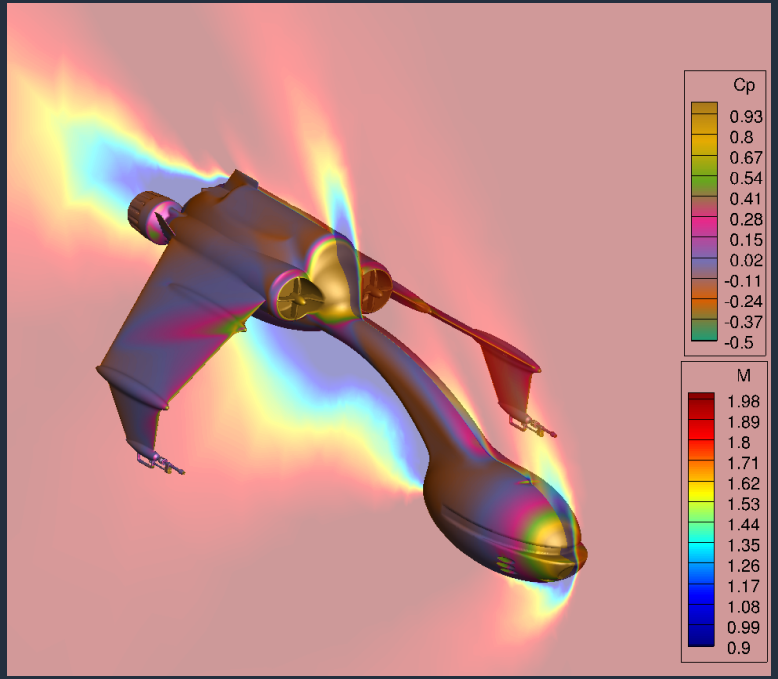




Amdahl's law (above) shows us how hard it is to scale an application even close to linearly.

What can EFA do?





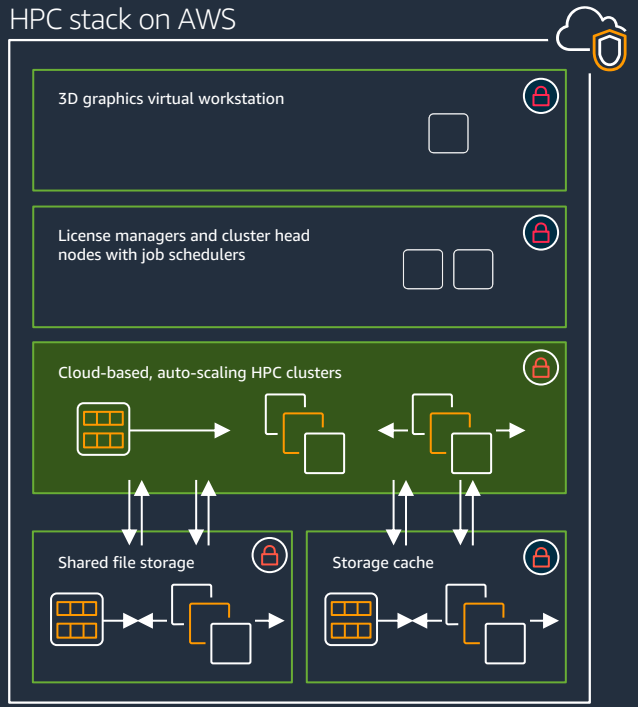
Thanks to Metacomp Technologies and the Klingon Empire. Garrrhhh.

High bandwidth compute instances: C5n

Massively scalable performance

- C5n Instances will offer up to 100 Gbps of network bandwidth
- Significant improvements in maximum bandwidth, packet per seconds, and packets processing
- Custom designed Nitro network cards
- Purpose-built to run network bound workloads including distributed cluster and database workloads, HPC, real-time communications and video streaming

Featuring Intel Xeon Scalable (Skylake) processor

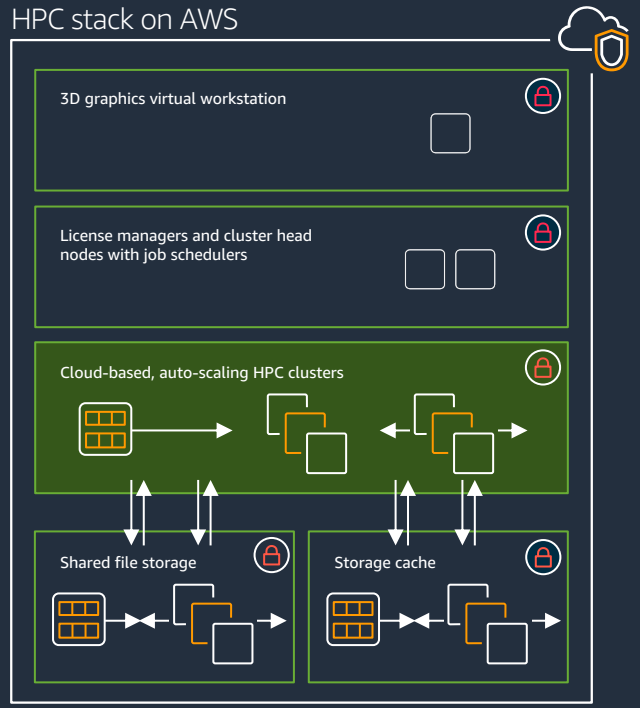


I/O Intensive Compute Instances: i3en

Dense SSD storage for data-intensive workloads

- One of the most I/O intensive instances available in the cloud
- 100 Gbps of networking throughput
- EFA enabled
- 96 vCPUs of Intel® Xeon® Scalable (Skylake) processors @ 3.1GHz
- 60 TB of total NVMe instance storage
- 768 GiB of memory

Featuring Intel Xeon Scalable (Skylake) processor



Where can you go from there ?

What if I want to test my HDF5 patches against a Lustre FS ?

```
{
    "AWSTemplateFormatVersion": "2010-09-09",
    "Resources": {
        "Type" : "AWS::FSx::FileSystem",
        "Properties" : {
            "FileSystemType" : "LUSTRE",
            "LustreConfiguration" : {
                "ExportPath" : "/scratch"
            },
            "SecurityGroupIds" : [ "sg-xxxxxxxx" ],
            "StorageCapacity" : 7200,
            "SubnetIds" : [ "subnet-xxxxxxxx" ]
        }
    }
}
```


aws

What about deployment ?

Let’s validate our template first

```
$ aws cloudformation validate-template --template-body file://fsx-lustre-template.json
{
    "Parameters": []
}
```

And now we can deploy it

```
$ aws cloudformation create-stack --stack-name MyTestFS --template-body file://fsx-lustre-template.json
{
    StackId: arn:aws:cloudformation:eu-west-1:xxxxxxxxxx:stack/MyTestFS/xxxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
}
```


aws

And now that I’m done with my tests, how do I get rid of it ?

```
$ aws cloudformation delete-stack --stack-name MyTestFS
```

If my test suite runs for an hour, how much will that cost?

0.14 x 7200 / (30 * 24) = 1.4 $

aws

What else can I do ?

- Test on different operating systems, different versions
- Test on different CPU/GPU architectures
- Test on different filesystems
- Automate my test infrastructure build system
- Test my code each time I do a commit

Thank you!

Arthur Petitpierre – Amazon Web Services
arthurpt@amazon.com – EMEA HPC Specialist Solutions Architect