# HPC, MPI, EFA, Oh My!

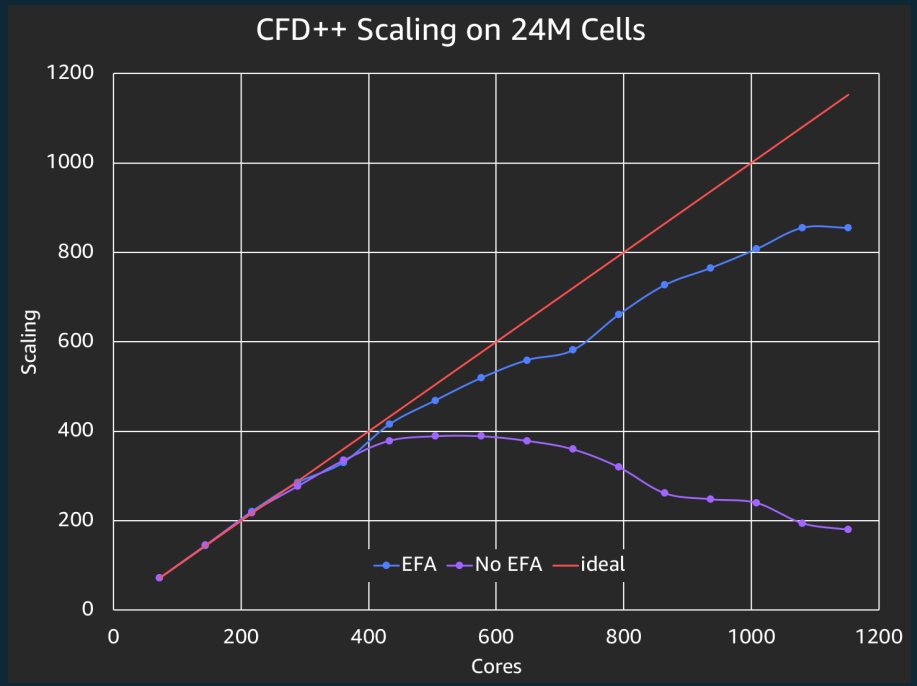
Brian Barrett, Principal Engineer, HPC

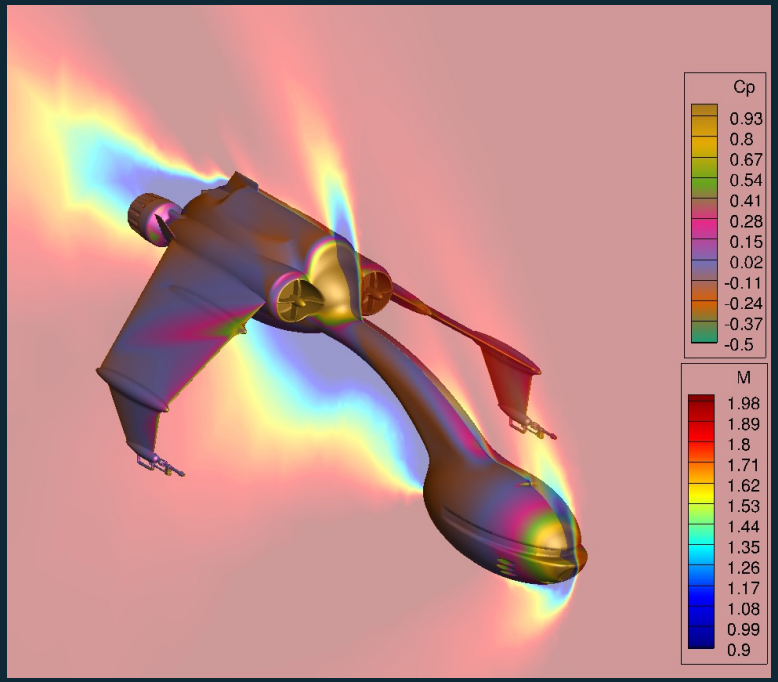


aws

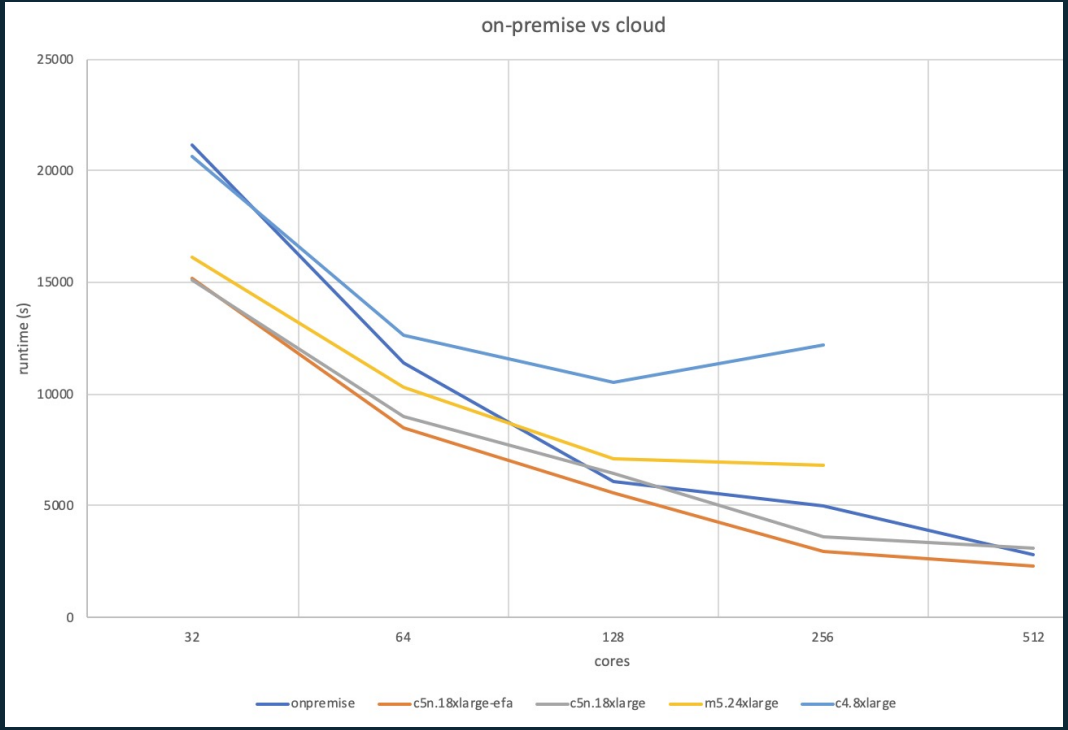
# What is High Performance Computing (HPC)?

# Metacomp CFD++



# LSTC LS-DYNA




Car2Car time to completion with C5n + EFA Vs On-Premise, C5n, M5, and C4

**At ~512 cores, C5n+EFA shows ~25% faster time to completion over C5n w/o EFA**



https://www.totalcae.com/learn/look-aws-efa-ls-dyna/

# That sounds easy?

# The Modern Platform Problem

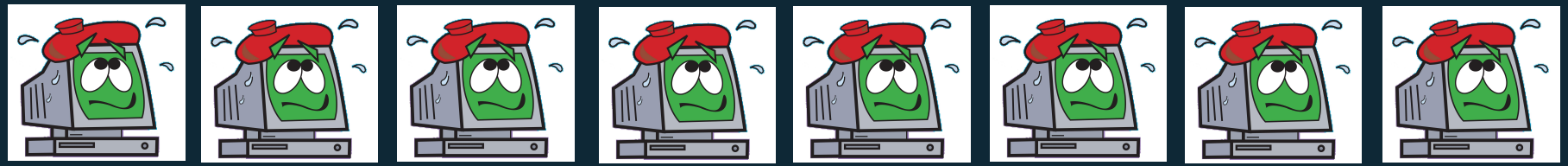
512 GiB

8,192 GiB



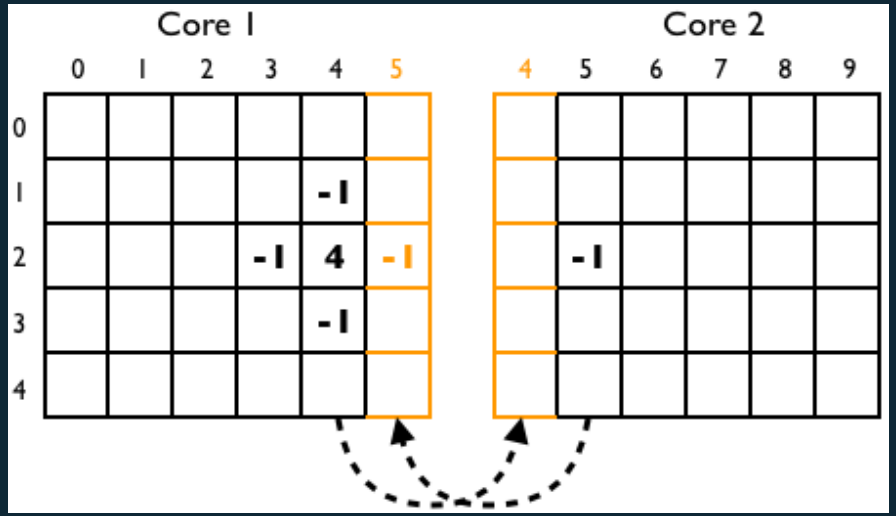
aws

# The Modern Platform Problem




8,192 GiB

# Ghost Cell Exchange

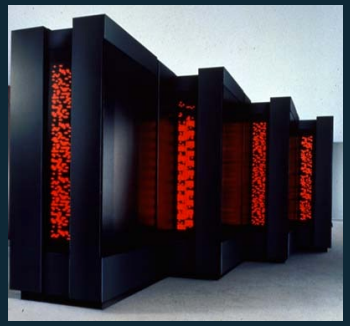


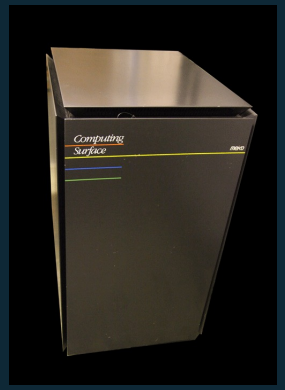

Everyone posts N-1 (Where N is the stencil count) Sends and Receives



aws

# In the 80s and 90s, many ways to do this

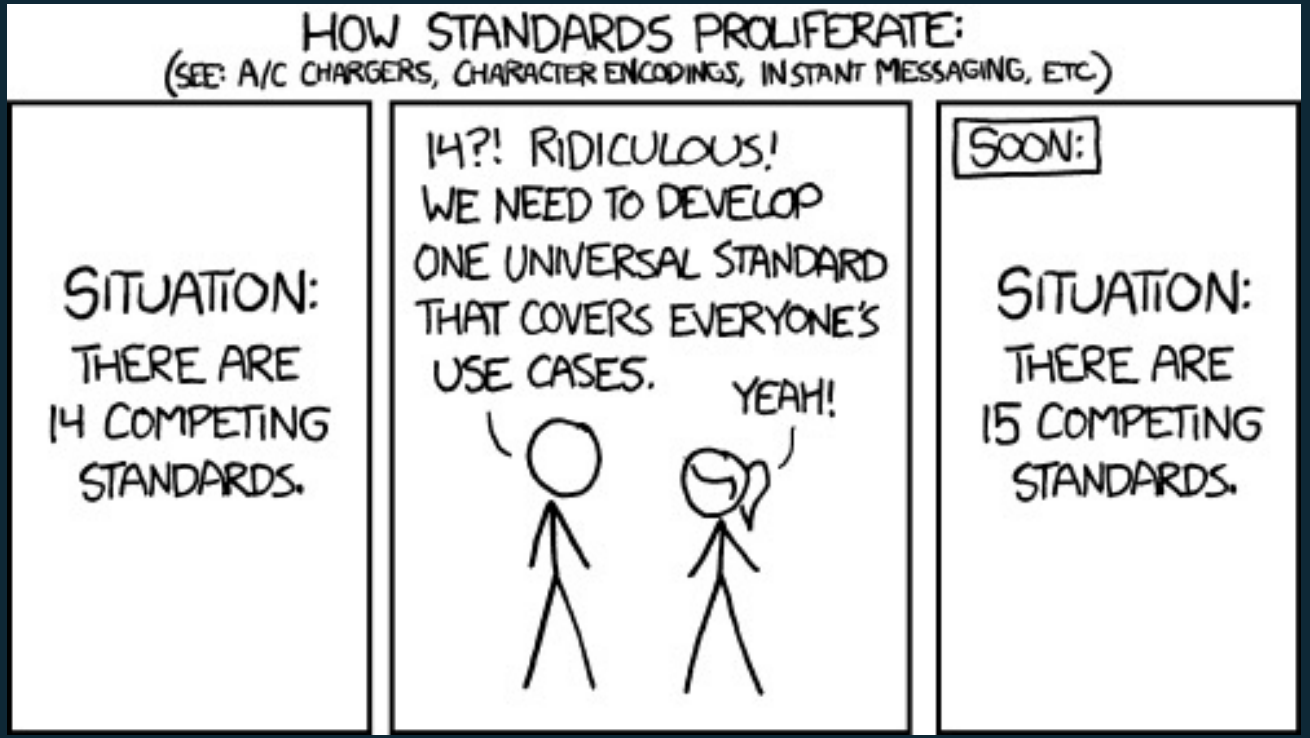





aws

# Once again, XKCD to the rescue…




Copyright XKCD - https://xkcd.com/927/

# MPI: The One Standard to Rule Them All[1]

- HPC communication interface
- Standard with multiple implementations
- Black box – application writers don't need to understand how it works, just that it does work
- Slowly evolving standard – new updates every 3-5 years



[1] As long as "them all" are HPC applications

# MPI Implementations









aws

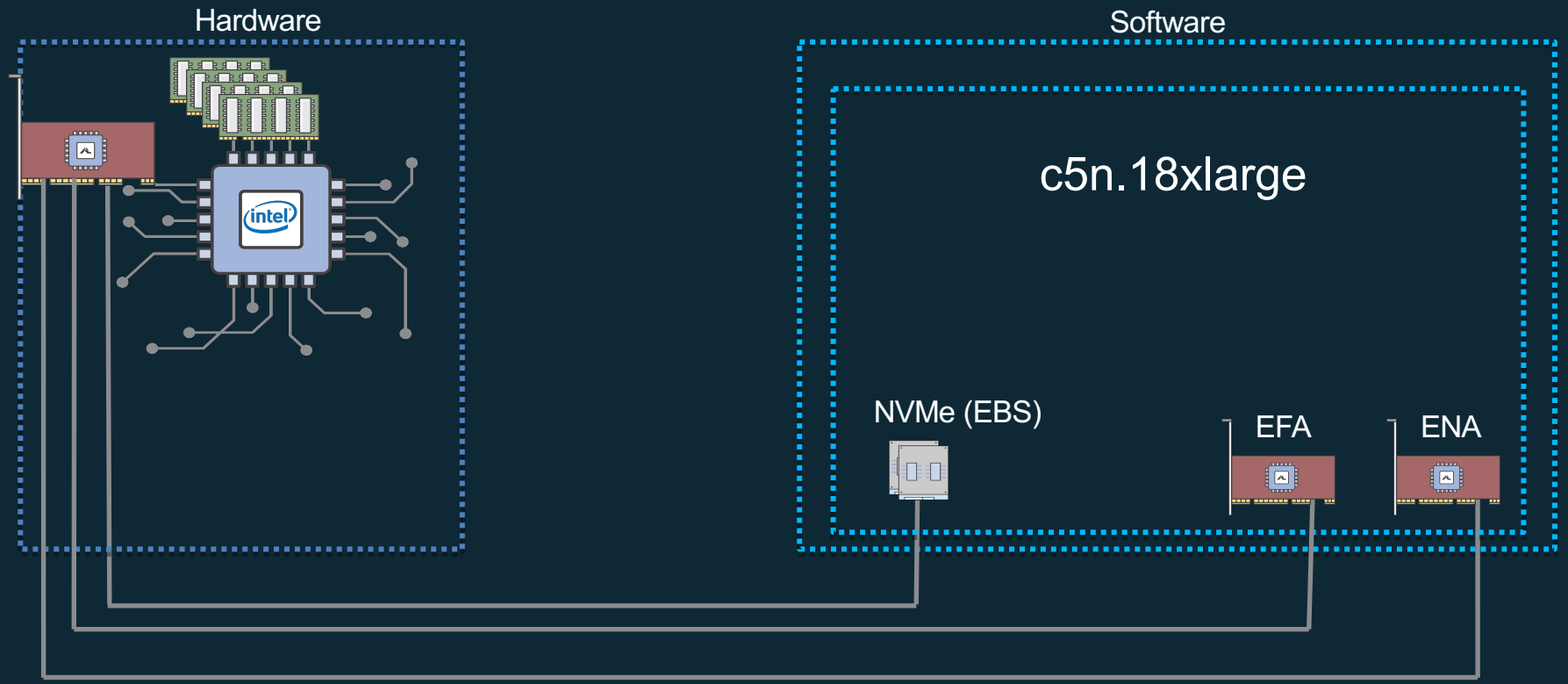
# Ok, Let's move some data! (example)

```
double the_buffer = 5.0;

MPI_Send(&the_buffer, 1,
   MPI_DOUBLE, 1, 1,
   MPI_COMM_WORLD);
```

```
Double the_buffer = 0.0;
MPI_Status status;

MPI_recv(&the_buffer, 1,
   MPI_DOUBLE, 0, 1,
   MPI_COMM_WORLD,
   &status);
```

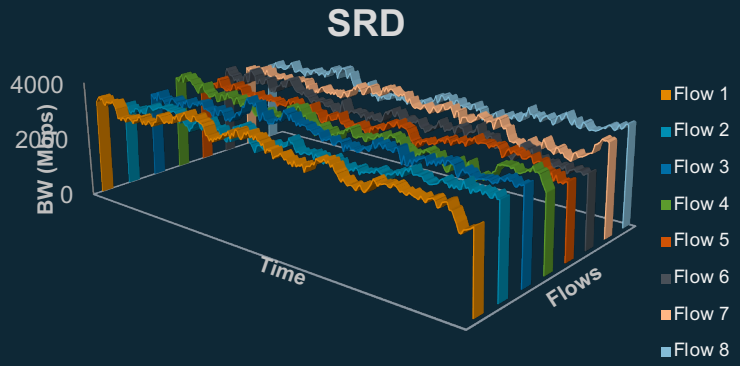
# How does all this work in AWS?

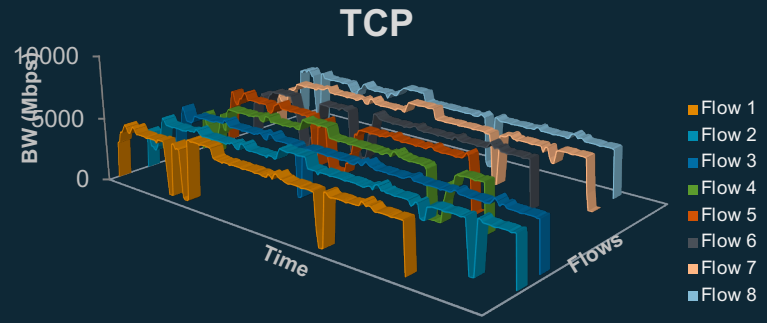
# EFA: HPC Networking in the Cloud



# Why build our own HPC network?

- Large Cloud-scale network an advantage
- Offer flexibility in instance choices
- Customize hardware to application needs

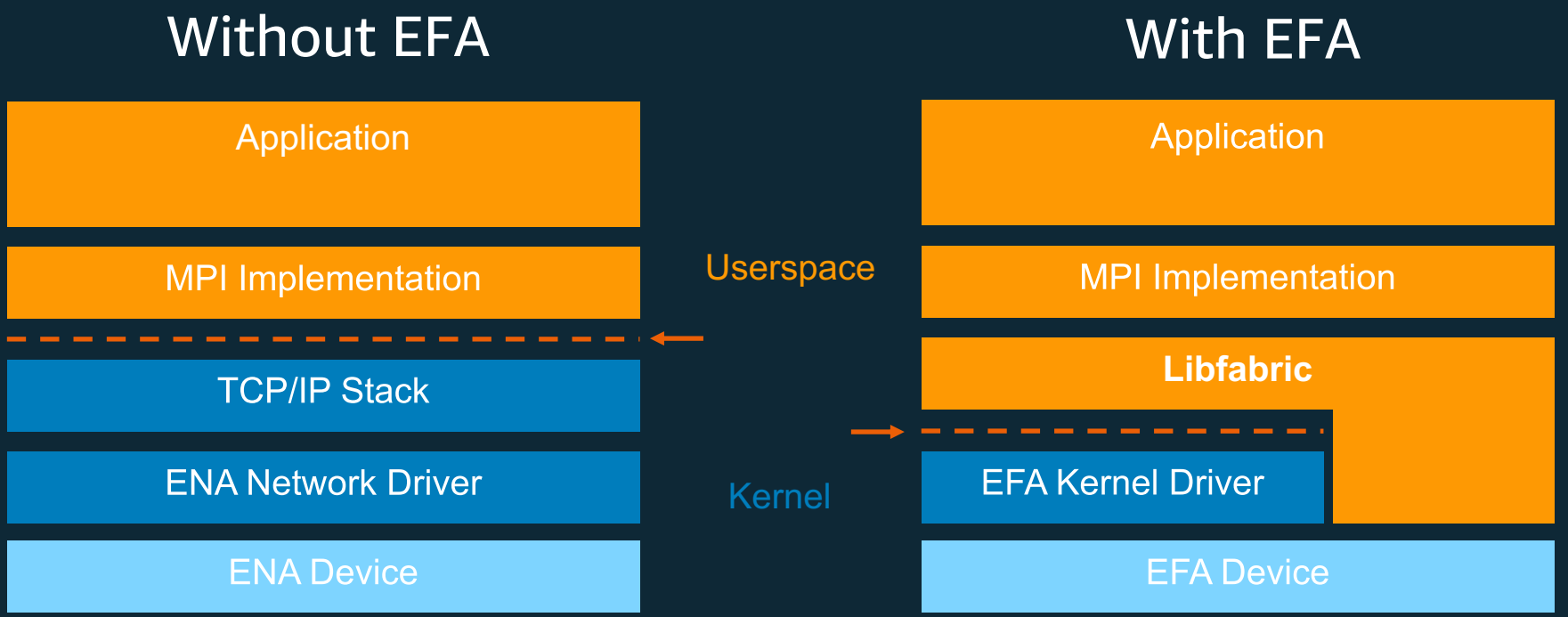





aws

# HPC software stack on Amazon EC2






aws

# MPI Implementations W/ EFA Support

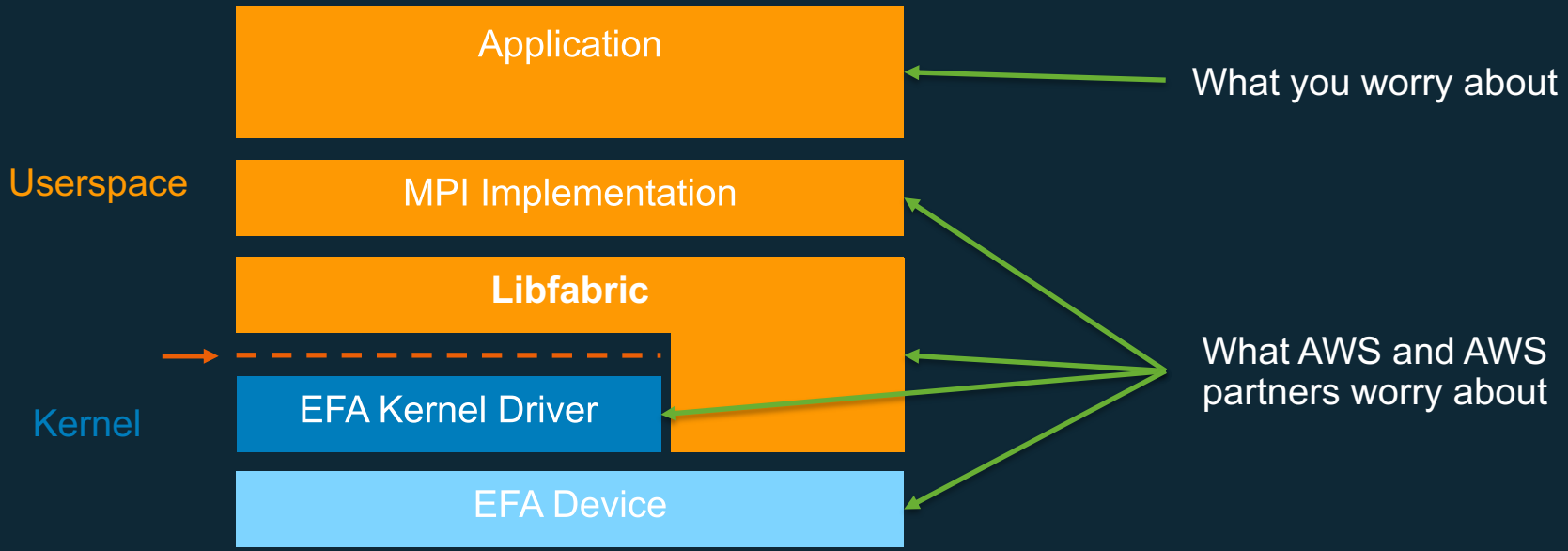






aws

# HPC software stack on Amazon EC2



# Performance

- EFA Instances: c5n.18xlarge, c6gn.16xlarge, p4d.24xlarge, p3dn.24xlarge, …
- Throughput: 100 Gbps
- Latency: 15 – 20 µs
- Message Rate: 10 Mmsg/second

- Today!



aws

# Getting Started

# Getting started

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/efa-start.html

- Amazon Linux 2 includes EFA stack in repositories, but slightly out of date
- AWS ParallelCluster includes recent EFA stack for all supported distributions
- EFA Installer Package to install:
    - EFA kernel module
    - Recent rdma-core
    - Libfabric
    - Open MPI



aws

# Common Tripping Points

- Security Groups
  - Need BOTH an ingress and egress rule that allows all traffic within the security group
  - Default 0.0.0.0/0 egress rule does NOT meet this requirement
  - ParallelCluster handles all this for you
- If building a cluster yourself, make sure using the same MPI build on every instance.



aws

# Using MPI with SLURM

- Open MPI and Intel MPI both have integration with SLURM, and SLURM handles all the hard work
- Two ways to use:
    - salloc –n 128
      mpirun ./a.out
    - srun –n 128 ./a.out
- SLURM will handle memory and process pinning for you. But if you're using threads, be sure to use the --cpus-per-task option



aws

# High Performance Computing on AWS

Innovate fast. Innovate securely. Innovate within budget.



aws