Leveraging the DL Software Stack for HPC: Vision and Planning

Peter Messmer
pmessmer@nvidia.com

SOS21, Davos, 3/22/2017

Vision: Data Science Drives Architecture

Microsoft
Olympus

Facebook
OCP Big Basin

Data science/Deep learning needs heavily influence architecture

Extensible NVLink, CPU as orchestrator, ..

System level: Dense nodes, high-performance intra-node communication

DGX-1/SATURNV, Big Basin, Minsky, Olympus, …

GPU level: Instruction set influenced by needs of data science

Half/mixed precision, int8, ..

Develop HPC applications for high-density nodes (also e.g. CORAL)

Leverage the DL hardware features for scientific computing

Vision: Data Science Drives Software-Stack

NVIDIA is the AI computing company, thinking lots about software!

Data science mission-critical to non-traditional HPC organizations

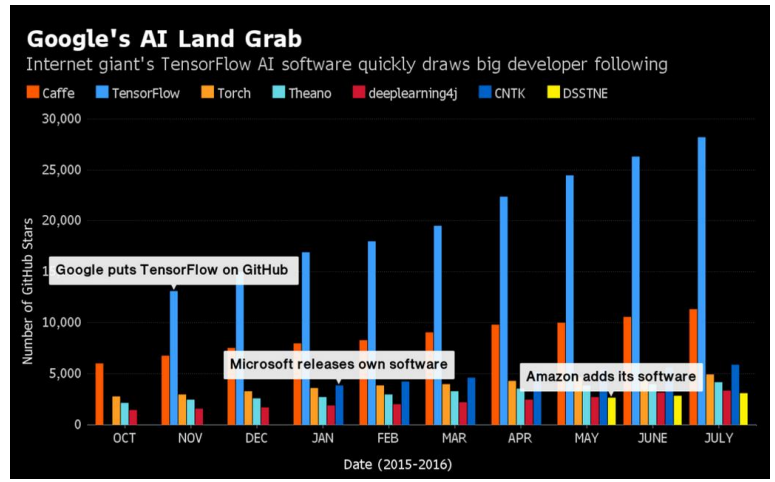
Deep learning, graph analytics, in-core databases,..

Sustainability and performance by scale

Frameworks supported by big corps, large communities

Big market, big support by all vendors

⇒ Economics drive performance portability and sustainability



Trilinos: 143 stars, Caffe: 16’679 stars

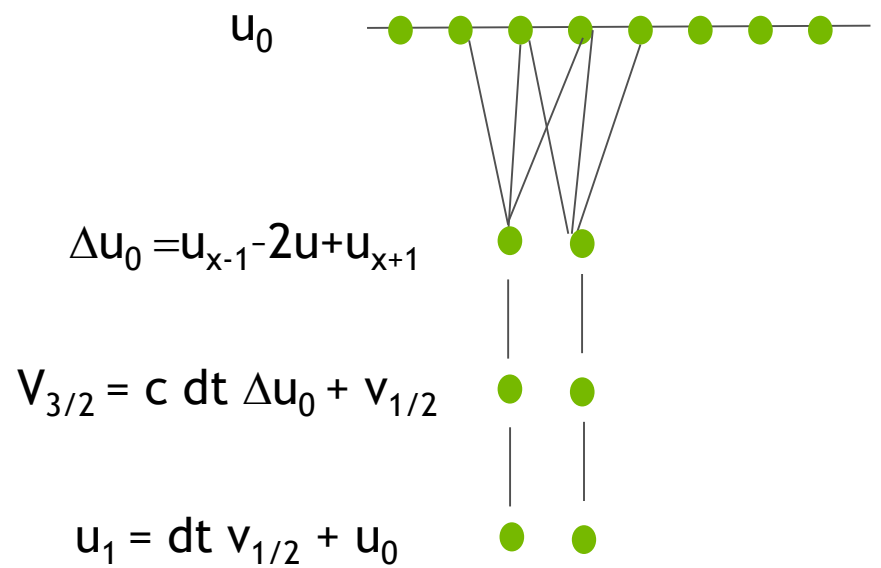
Cast HPC algorithms in DL terms

Example: Wave equation via Conv Neural Network

Applying stencils = Inference in Conv Network

$$\frac{dv}{dt} = c\,\Delta\,u$$

$$\frac{du}{dt} = v$$



Simplified 1D example cartoon
Approach works for higher dimensions

```
layer {
        name          : “laplace_u”
        type          : “convolution”
        bottom        : “u_0”
        top           : “laplace_u”
        convolution_param {
                num_output  : 1
                kernel_size : 3
                stride      : 1
                pad         : 1
..
}

layer {
        name          : “velocity_update”
        type          : “Eltwise”
        bottom        : “laplace_u”
        top           : “velocity_update”
}

layer  {
        name          : “position_update”
        type          : “Eltwise”
        bottom        : “velocity_update”
        top           : “position_update”
}
```

Other examples: Stencils, Spectral transforms, spectral elements, ..

Vision: HPC can Contribute to Emerging Data Science Needs

HPC solved lots of “new” problems in the past

DL will need distributed memory parallelism

New challenges for DL algorithms

HPC has probably hit those challenges in the past

Better implementation, better algorithms

https://www.hpcwire.com/2017/02/21/hpc-technique-benefits-deep-learning/

Collaborate with DL framework developers, contribute to DL frameworks

First step: speak a common language

Vision: Combined DL and HPC

Jointly solve new problems, better

Many HPC models have “inaccurate” components, eg parameterized sub-model

Often complex control flow

A trained network might result in higher performance, better accuracy

Possible examples: collisional cross-sections, chemical reaction chains,

Simplified if rest of application is already in DL friendly fashion

Revisit parameterized models

Example: Fog prediction at Zurich Airport

New data products via DL post-processing

Fog has operational impact

Not simulated by operational weather model

Require significantly higher resolution

Use deep learning to correlate fog conditions from weather simulation data

Historic weather forecast t+24hrs, correlation with observations

Promising initial results, ongoing research

Summary

DL, data science software stack backed by vendors, large corps

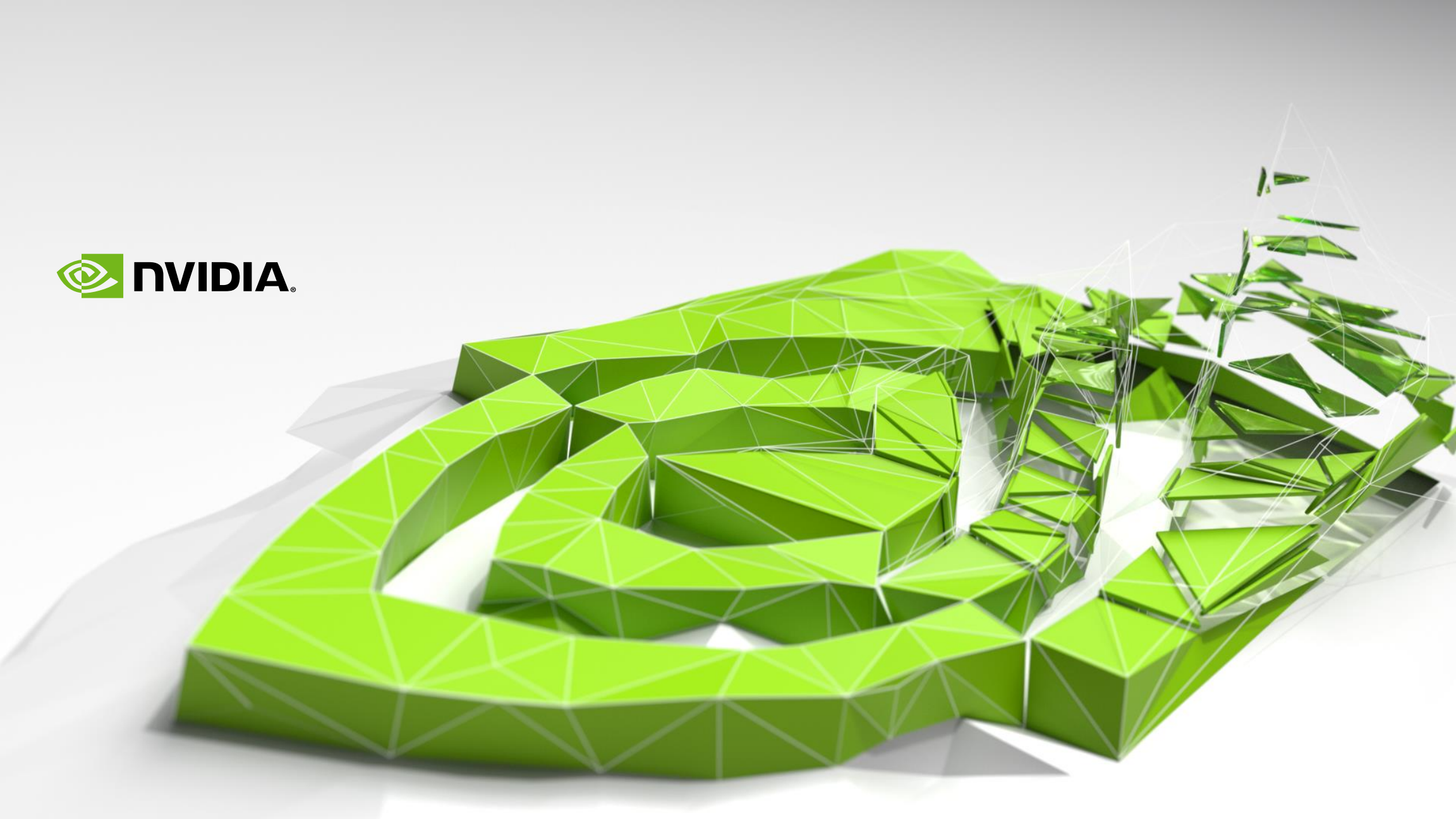
Large overlap with HPC applications

=> Leverage DL efforts for HPC

Need contributions: MPI parallel framework, support for FP64, use-cases, ..

Possible solution for HPC software sustainability, (performance) portability problem

Call to action: Look at your scientific problem through the eyes of data science


NVIDIA