High Performance Computing on AWS
For Design, Engineering, and Manufacturing

Dr. Karsten Gaier
HPC Sales Specialist, AWS
Munich, September 18th, 2018
Motivations – Why HPC in the Cloud

(and what is HPC, anyway?)
POPULAR HPC WORKLOADS ON AWS

- Genome Processing
- Monte Carlo Simulations
- Transcoding and Encoding
- Modeling and Simulation
- Government and Educational Research
- Computational Chemistry

...and many more
POPULAR HPC WORKLOADS
In Design, Engineering, Manufacturing

Genome Processing

Monte Carlo Simulations

Modeling and Simulation

Transcoding and Encoding

Government and Educational Research

Computational Chemistry
DEFINING HPC – EXAMPLE USE-CASES

Data Light
Minimal requirements for high performance storage

Fluid dynamics
Weather forecasting
Materials simulations
Crash simulations

Risk simulations
Molecular modeling
Contextual search
Logistics simulations

Clustered (Tightly Coupled)

Seismic processing
Metagenomics
Astrophysics
Deep learning

Data Heavy
Benefits from access to high performance storage

Animation and VFX
Semiconductor verification
Image processing
Genomics

Distributed/Grid (Loosely Coupled)
Business Drivers - HPC on AWS

- Faster design cycles - queue lengths too high, peak loads
- Agility – react quickly to market demands
- CAPEX/OPEX – avoid capital invest and binding
- HW refresh cycles – bring in new hardware takes too much time
- Capability – run a code on new hardware which is not available on-premises (GPU, FPGA, ...)
- Competitive advantage – be faster than the competition, realize the potential
- Efficiency - leverage expensive ISV licenses optimally on latest hw
IMPORTANT ENABLERS – HPC on AWS

- Compute performance – CPUs, GPUs, FPGAs
- Memory performance – high RAM requirements in many applications
- Network performance – throughput, latency, and consistency
- Storage performance – including shared filesystems
- Automation and cluster/job management
- Remote 3D graphics for pre- and post-processing
- ISV support – including license management

...and SCALE
FASTER THROUGHPUT WITH RAPID SCALING

Scale up when needed, then scale down

- In a traditional datacenter, the only certainty is that you always have the wrong number of servers – too few, or too many

- Every additional server launched in the cloud can improve speed of innovation – if there are no other constraints to scaling

- Over-night or over-weekend workloads - reduced to an hour or less

Think big: what if you could launch 1 million concurrent jobs?

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
Amazon EC2 Instances

- Select compute that best fits the workload profile; Match the architecture to the job, not vice versa
- Optimize price/performance of your HPC Workloads with widest range of compute instances
- Benefit from the AWS pace of innovation
FLUID DYNAMICS

- C4.8xlarge instance type
- 140M cell model
- F1 car CFD benchmark

![Graph showing external flow over a Formula-1 race car with AWS, C4.8XLarge instance type for computational fluid dynamics simulations.](attachment:image.png)
Transforming CFD with GPUs

Dr. Bastian Schnepf
Business Development Manager - CFD, Altair
Munich, September 18th, 2018
Altair transforms design and decision making by applying simulation, machine learning and optimization throughout product lifecycles.

This is really in our mission statement 😊
TRANSFORMATION IN THE AUTO INDUSTRY

Autonomous Driving
Connectivity
Mobility concepts
Not our topic for today

CO2 / Emissions
Efficiency
Reduce driving resistance
► Aerodynamic drag
Reduce losses in drivetrain
► Churning losses

Electrification
Thermal management
► Cool and lubricate high-rpm concepts
  (e-motors, e-drives)
EFFICIENCY THROUGH OPTIMIZED AERODYNAMICS

- For 70+ km/h aerodynamic drag becomes dominating driving resistance
- Target quantity drag coefficient
  \[ c_d = \frac{F_d}{0.5 A_{ref} \rho V^2} \]
- \(c_d\) values range 0.21 - 0.35 for modern passenger cars
  0.2 is the magical border
- Decreasing \(c_d\) by 0.01 saves \(~1.5\text{g CO}_2 / 100\text{km}\) in WLTP
- Modern cars’ aerodynamics are already optimized, differences are in the details
- To exploit even further potential aerodynamic analysis needs to happen 
  **early, fast** and **accurately** with the help of CFD.

Schütz, Aerodynamik des Automobils, 2013
EXPLOITING GPU COMPUTE POWER

• GPUs deliver tremendous speed for explicit, local algorithms:

  ultraFluidX: LBM (D3Q27)
  Streaming to next neighbors
  Collision purely local
  Explicit time discretization
  (weakly compressible)

  nanoFluidX: SPH (Navier-Stokes)
  Kernel interpolation within 3*dx (quintic)
  Explicit time-discretization
  (weakly compressible)

• Algorithms written from scratch for efficient parallel CUDA/C++ implementation on GPUs.

  • Overnight runs on single server node
  • Lower cost for performance (hardware & energy)
  • High pace of GPU development

1 GPU h corresponds to O(100s) CPU core h
TRANSFORMATION IN THE AUTO INDUSTRY

Autonomous Driving
Connectivity
Mobility concepts
Not our topic for today

CO2 / Emissions
Efficiency
Reduce driving resistance
► Aerodynamic drag
Reduce losses in drivetrain
► Churning losses

Electrification
Thermal management
► Cool and lubricate high-rpm concepts (e-motors, e-drives)
DRIVETRAIN LUBRICATION

• Drivetrain components need to be well lubricated to **avert damage**
• But too much oil degrades the **efficiency** of the drivetrain
• Experiments with transparent housings are **expensive** and **late** in the development process
• **Two-phase flow** field of oil and air (for high-speed e-concepts air influence is significant)
• Conventional CFD methods take **weeks** for pre-processing and analysis.

http://recongearbox.co.uk/

Schelbe (GKN), ATCx CFD, 2016
The GPU-based SPH approach of **nanoFluidX**

- **Decimates the pre-processing effort** and handles complex motion easily
- **Cuts down run time** for high-fidelity **multiphase** flow simulations to a few days vs weeks for conventional methods.
E-MOTOR COOLING

- Heat dissipation is one the major constraints for electric motor design
- Conventional cooling concepts rely on water jackets
- Modern electrical machines use direct oil cooling of windings (e.g. sprayed from the rotor)
- Challenging to simulate because of high rpm and two-phase flow (oil and air interaction)
E-MOTOR COOLING

Oil distribution

Streamlines
Two-phase flow simulation on GPUs using nanoFluidX

- 11 Mio. Particles
- 1.5 sec physical time at 4000 rpm
- Only 18h on AWS p3.16xlarge (8 Nvidia TESLA V100)
- Conventional CFD would take > 1 week
- Thermal analysis in AcuSolve using time-averaged flow field from nanoFluidX
- The multi-physics simulation covering electromagnetic losses, fluid flow and heat transfer can improve design decisions much earlier in the development process.
IF YOU WANT TO KNOW MORE…

featuring

SPH Simulation of an Oil Cooled Electrical Machine
Nicolas Brossardt, BMW Group

Developing the electric drivetrain of tomorrow - How numerical tools can support the design process and improve system understanding
Johannes Winklinger, Magna Powertrain ECS
THANK YOU!

Dr. Karsten Gaier, AWS
Dr. Bastian Schnepf, Altair
Munich, September 18th, 2018