AWS Graviton3:

The first cloud native SVE-enabled Arm based processor

Olly Perks – operks@amazon.com

Snr. Dev Advocate for HPC
AWS

Part 1: The Hardware

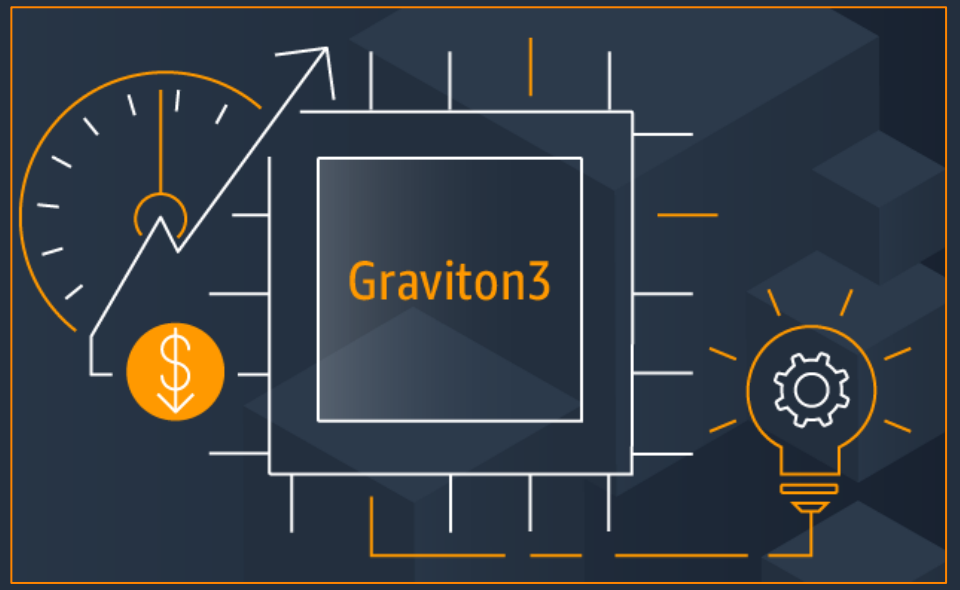
AWS Graviton 3: Available now

Announced at re:Invent ‘21

- With an early access program
- GA happened May ’22

Latest Arm base instances at AWS

- Compute optimized C7g
- Lots of cool new features



First, a history lesson

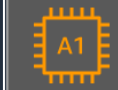
Graviton3 is the third generation of Arm at AWS

Annapurna Labs

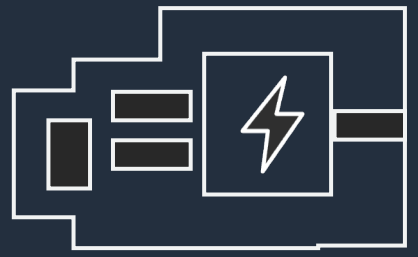
- Acquired by Amazon in 2015
- To build custom silicon

Each generation innovates more

- Extended reference Arm core

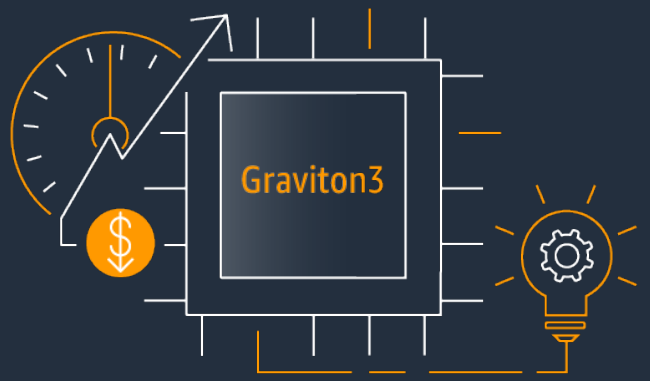
	Arm Core	Announced
 Graviton	Cortex-A72	2018
 Graviton 2	Neoverse-N1	2019
 Graviton 3	Neoverse-V1	2021

Silicon innovation at AWS



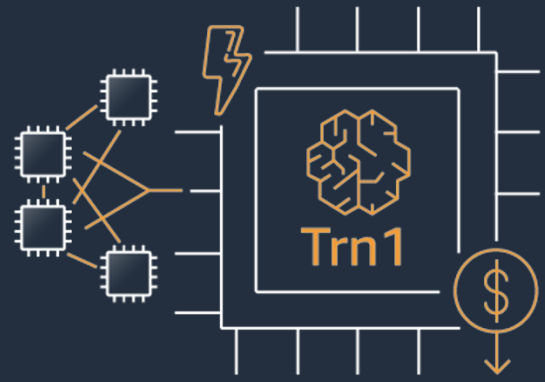
AWS Nitro System
AWS Nitro SSD

Hypervisor, network, storage, and security



AWS Graviton2
AWS Graviton3

Powerful and efficient, modern applications



AWS Inferentia
AWS Trainium

Machine learning, hardware and software at scale

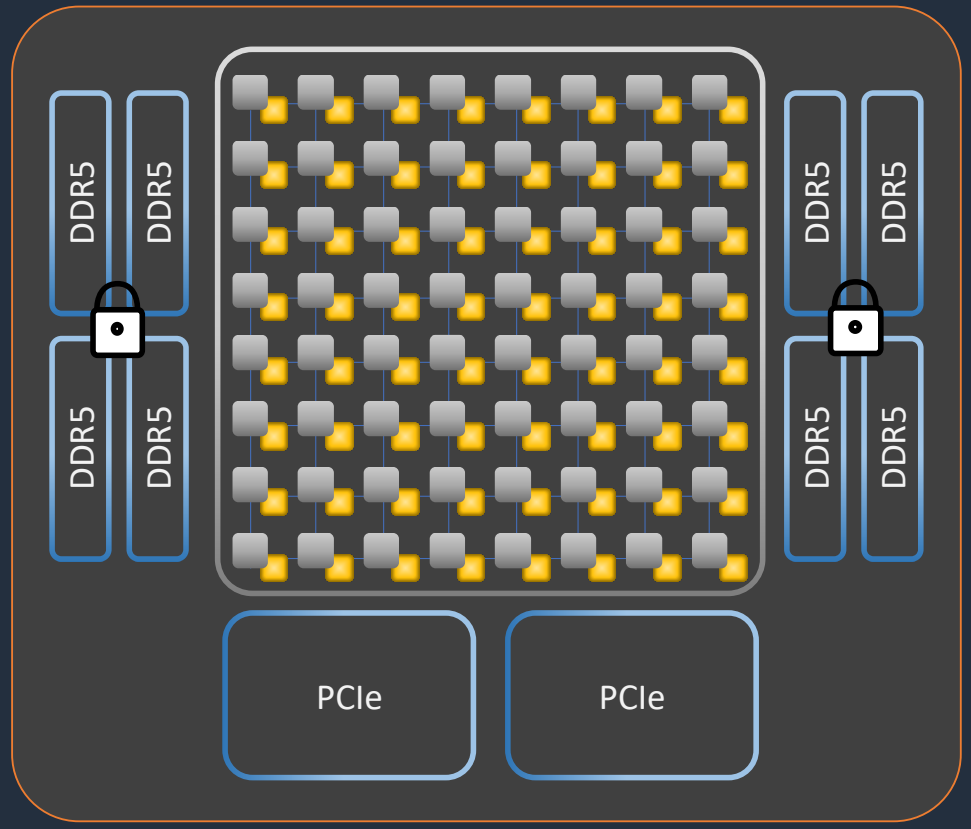
Graviton3: What's in the box?

SoC:

- 64 cores in a 2D mesh
- 300 GB/s of DDR5
- Single NUMA domain

Core:

- 2.6 GHz Neoverse-V1 based core
- 2x256 bit SVE
- BF16 support



Part 2: The Software

So what does that mean for HPC?

C7g in not an HPC specific instance

High usable memory bandwidth

- Great for traditionally memory bound codes (like CFD)

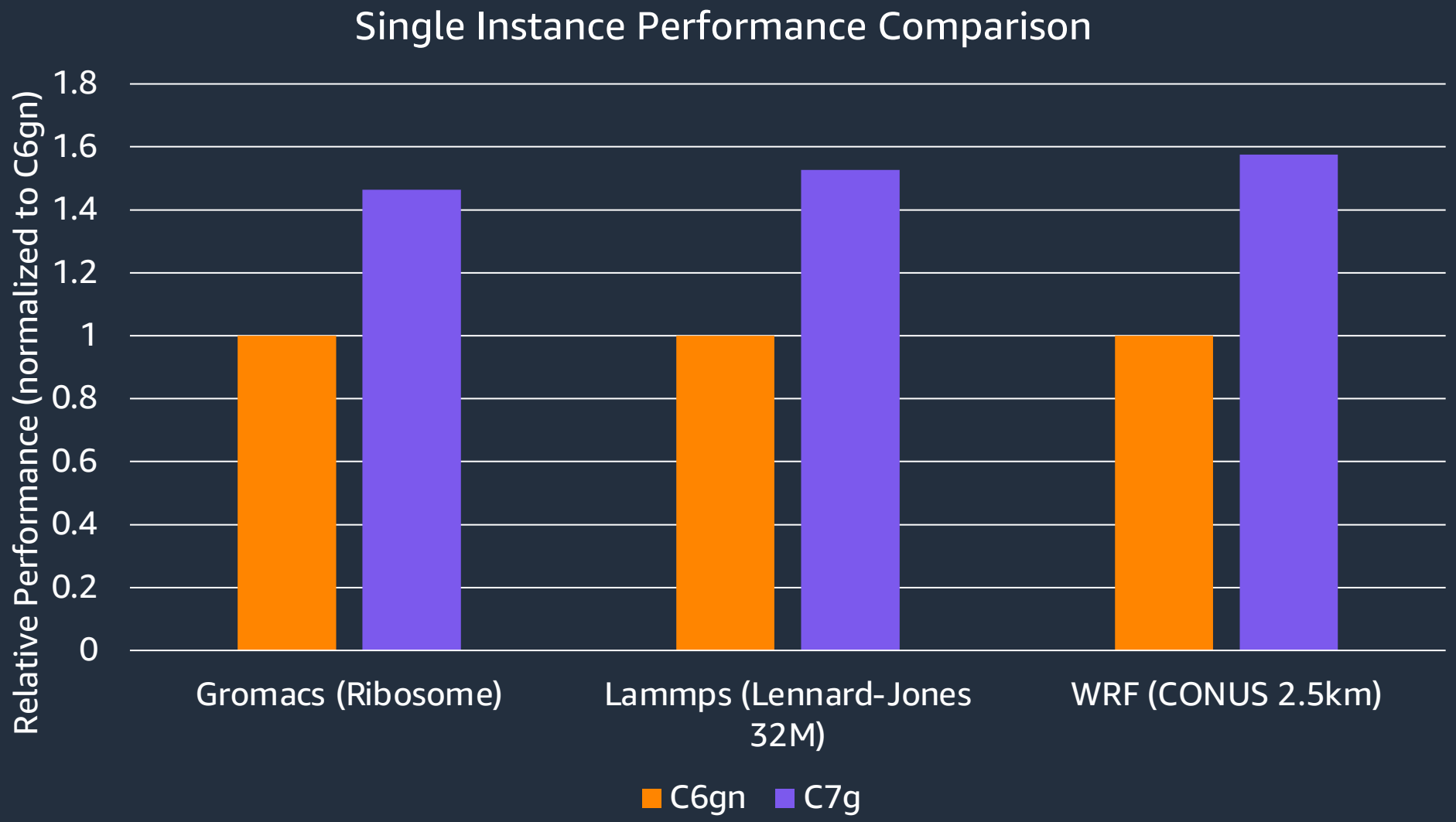
2x256bit SVE

- Makes SVE easily accessible

Currently only in a low network bandwidth configuration

- 30 Gbps (vs 100 of HPC specific instances)

Initial benchmarking



Graviton3 Support in Dev Tools

Work for A64FX and Fugaku cleared the path

SVE support across tools

- Like SVE register support in GDB

Support across the compiler community

- GCC, LLVM and ACfL support
- Pick up Graviton3 as a target
- Microarchitecture feature tuned

GCC 11.2.0
`-mcpu=native`

```
__ARM_FEATURE_AES
__ARM_FEATURE_ATOMICS
__ARM_FEATURE_BF16_SCALAR_ARITHMETIC
__ARM_FEATURE_BF16_VECTOR_ARITHMETIC
__ARM_FEATURE_CLZ
__ARM_FEATURE_COMPLEX
__ARM_FEATURE_CRC32
__ARM_FEATURE_CRYPTO
__ARM_FEATURE_FMA
__ARM_FEATURE_FP16_FML
__ARM_FEATURE_FP16_SCALAR_ARITHMETIC
__ARM_FEATURE_FP16_VECTOR_ARITHMETIC
__ARM_FEATURE_IDIV
__ARM_FEATURE_JCVT
__ARM_FEATURE_MATMUL_INT8
__ARM_FEATURE_NUMERIC_MAXMIN
__ARM_FEATURE_QRDMX
__ARM_FEATURE_RNG
__ARM_FEATURE_SHA2
__ARM_FEATURE_SHA3
__ARM_FEATURE_SHA512
__ARM_FEATURE_SM3
__ARM_FEATURE_SM4
__ARM_FEATURE_SVE
__ARM_FEATURE_SVE_MATMUL_INT8
__ARM_FEATURE_SVE_VECTOR_OPERATORS
__ARM_FEATURE_UNALIGNED
```

AWS ParallelCluster

Virtual HPC clusters on-demand

‘Spin-up’ a 40k core cluster in ~5mins

Brings together ~20 AWS services

- High performance networks
- Parallel file systems
- User management

Target different instance types (Graviton)

Software Stacks via Spack

Announced the Spack Binary Cache this week

Supports two targets

- Generic aarch64 and Graviton2
- ~315 unique packages (each)
- ~450 combinations (each)

Pipelines to support new targets (Graviton3)

Part 3: The People

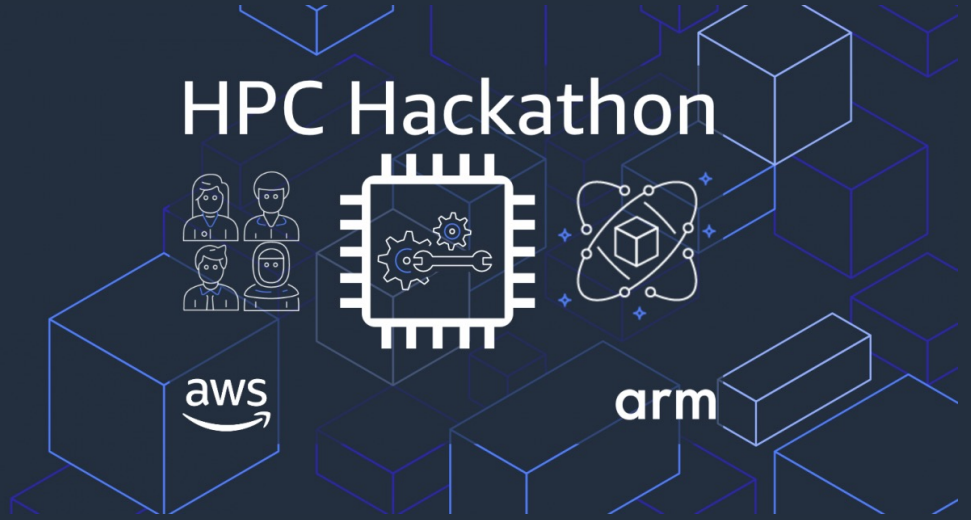
Raising the Community

Hardware isn’t sufficient!

Software isn’t sufficient!

We must bring the people along too:

- Exposure to environment
- Training on hardware
- Best practices
- Community!



AWS Community Builders

Working to further the community

Specific Arm group

- Champion HPC / Arm / AWS

Register interest:

- Waiting list
- Opens later this year

HPC Tech Shorts – hpc.news/techshorts

YouTube channel for discussing HPC at AWS

Putting out Graviton and Graviton3 content

- Intro to Graviton 3
- Developer environment on Arm
- Lots more content to come

Let us know what other content you want!

Staying in touch

We want to grow as part of the community

Our whole DevRel team is on Twitter

@boofla

@OllyPerksHPC

Find us on the A-HUG Slack