
RED HAT
SUMMIT
LEARN. NETWORK.
EXPERIENCE OPEN SOURCE.
June 11-14, 2013
Boston, MA
redhat.

Hyperscale Red Hat-powered ARM Server

Jon Masters
Chief ARM Architect, Red Hat A-Team
2013/06/13

Win an ARM system right now!

Raffle Draw (ticket on your wa

...sit through my talk first

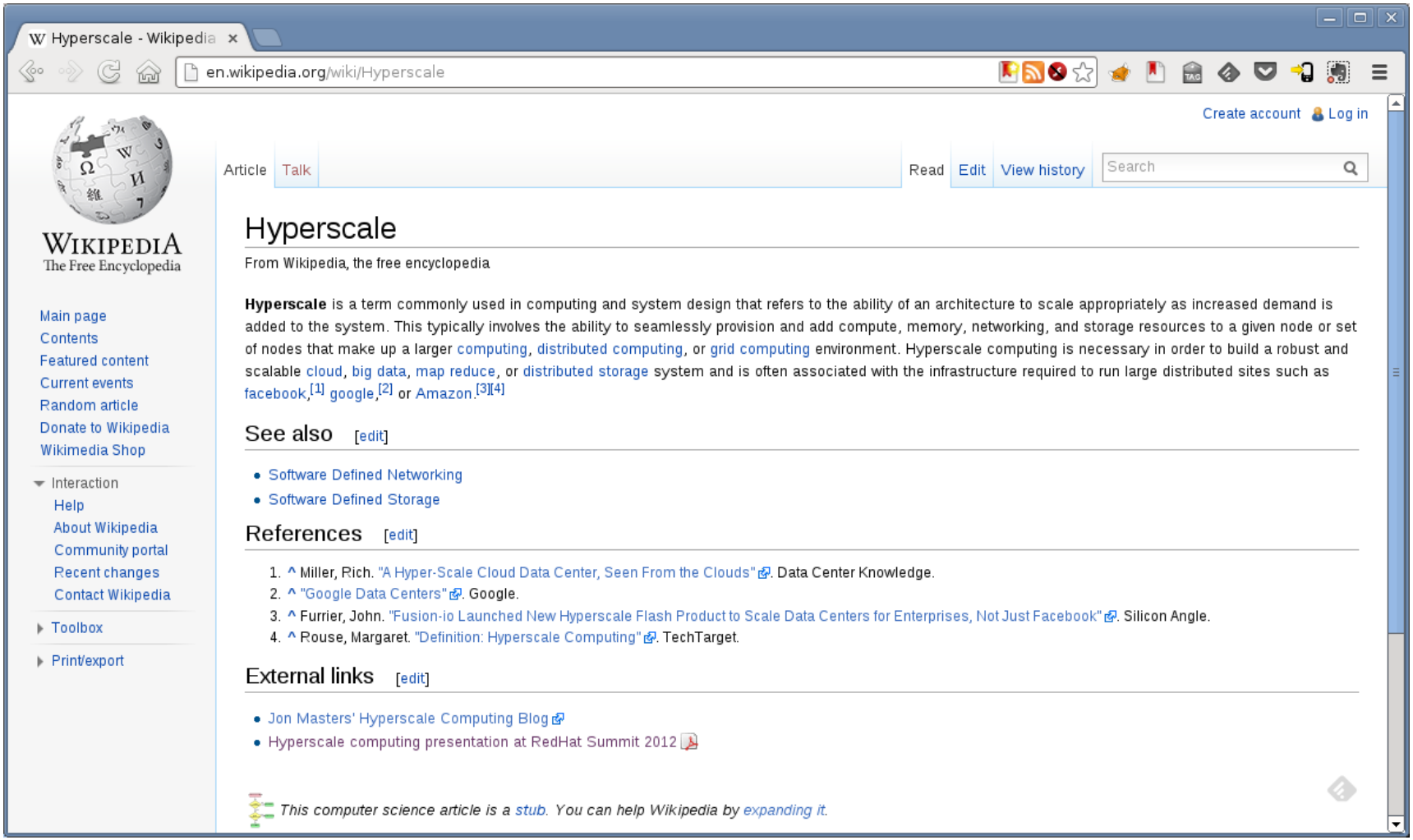
- Raspberry Pi Model-B (700Mhz SoC, 512MB RAM)
- Supports Pidora Fedora Remix (http://www.pidora.ca/)

Full competition rules available upon request

Overview

- One year ago...
  - Hyperscale Computing?
  - ARM Servers Became Reality
- Since last year...
  - Fedora 18 Released
  - Fedora ARM Build System migrated to ARM Servers
  - Red Hat co-founded Linaro Enterprise Group
  - Fedora AArch64 Bootstrap (64-bit ARM Architecture)

Hyperscale Computing



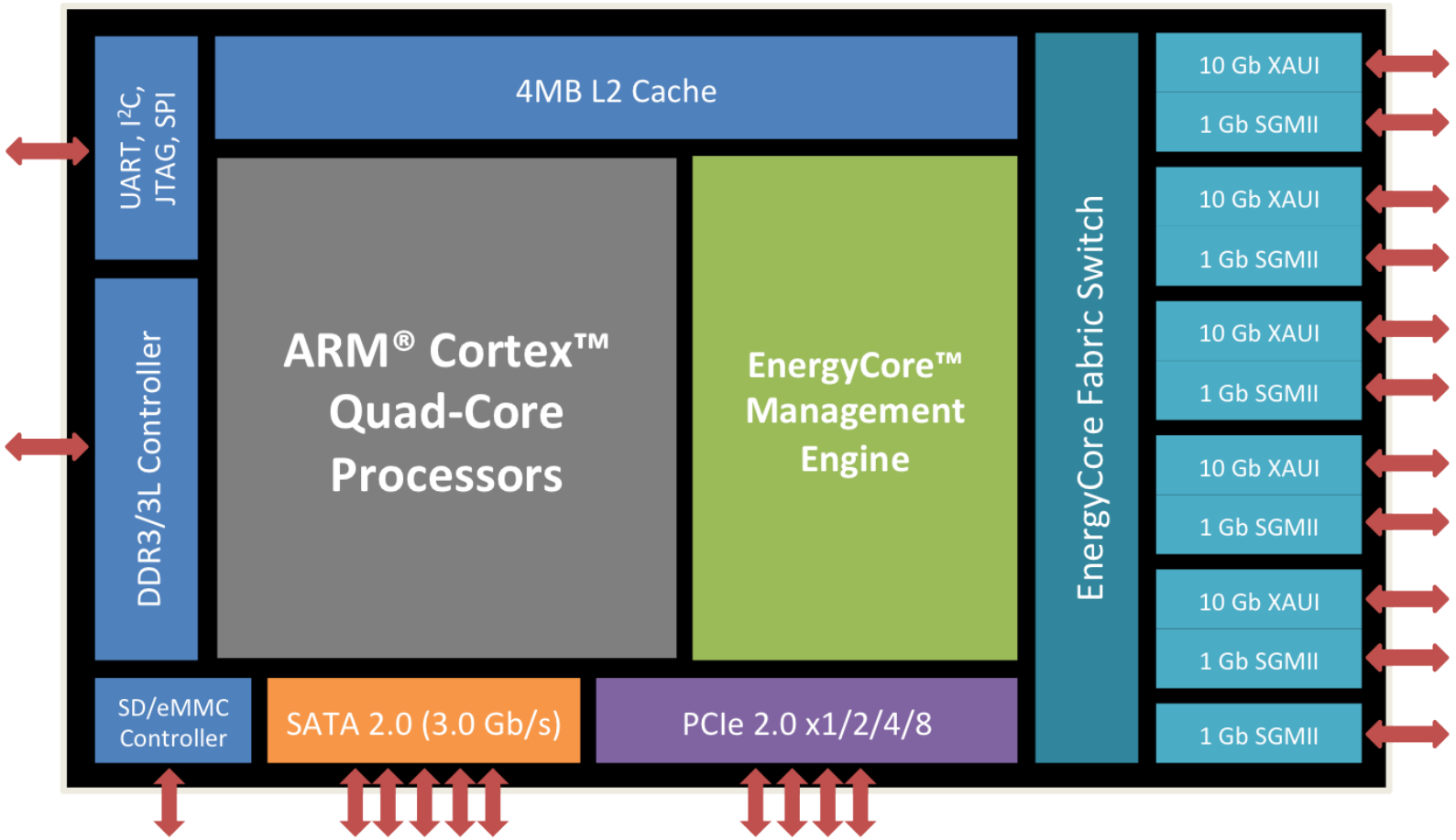
Hyperscale Computing

- Order of Magnitude Higher Density nodes per rack
- SoC Commodization of Server
- Fabric Interconnectivity
- Disaggregation of resources
- Ease of provisioning/management
- Failure-in-Place

Hyperscale Computing - Density

- 1,000-10,000 or more server nodes per rack
  - Tightly connected, whole rack-level granularity
- Multi-core won the linear GHz debate
  - Many simpler cores, lower energy
  - Target is aggregate performance
- Physicalization vs. Virtualization
  - Provision elastically, and reprovision often
  - Technical challenges include “Thundering Hurd”
  - Management controller allows image injection

Hyperscale Computing – SoC Commodization



Hyperscale Computing – SoC Commoditization

- System-on-Chip revolutionized mobile, now server
  - CPU, IO, devices integrated on-chip
  - Servers add offload engines, accelerators, management
- Very flexible designs, standardized (ARM) cores
  - Future servers will have more variety than today
- Fabric and management tight integration
  - IPMI, industry standard interfaces
- Eventually integrate memory, flash on-chip (PoP)

Hyperscale Computing – Fabric Interconnectivity

Hyperscale Computing – Fabric Interconnectivity

- Fabric powers future datacenters
  - Obviate the need for individual node/rack cabling
  - Obsolete the top-of-rack and inter-rack switching
- Virtualize network/other traffic
  - Multipath, redundant physical links to nodes/racks
  - Separate physical and virtual SoC/Fabric topology
  - Connects storage, PCIe, other protocols
- Enables true Software Defined Networking
  - Red Hat Founding member of OpenDaylight

Hyperscale Computing - Disaggregation

- Separate Compute from Memory, and Storage
  - Compute, memory key/value, storage nodes
  - High bandwidth, low-latency interconnectivity
  - 100GBit+ networking, silicon photonics emerging
- Depreciated resources on appropriate schedules
  - Leverage falling prices more effectively
  - Replacement on different 3-5 year timetables
- Facebook/OpenCompute blazing a trail

Hyperscale Computing - Failure-in-Place

- What happens to failed nodes?

Hyperscale Computing – Datacenter future

- Future datacenters are dark places nobody goes
  - Build near efficient power generation (PUE target)
  - Flood fill with SoCs from multiple vendors (3-5 years)
  - Fabric interconnectivity between nodes/racks
- Million node systems become commonplace
  - New opportunities in Big Data, Analytics
  - Ignore failed nodes completely

ARM Servers Became Reality

- ARM IP License Model applied to servers
- Calxeda EnergyCore HP Redstone (Moonshot)
  - System-on-Chip technology
  - Quad 32-bit ARM Processor Cores
  - Integrated (IPMI) Management
  - Integrated Fabric Interconnect
- More than just (ARM) compute

Bicycle Power Solves World Energy Shortage

Fedora 18 for ARM Systems

- Supports 32-bit ARM Servers out-of-the-box
  - Standard Red Hat technologies (kickstart, etc.)
- Supports many popular 32-bit ARM systems
  - BeagleBone, Trimslice, Chromebook
- Dropped support for older ARM Architectures
  - Now requires at least ARMv7 Application Profile
- Most complete, standardized release yet

Fedora ARM Build System – Before (Panda-on-a-stick)

Fedora ARM Build System – After (Boston Viridis)

Fedora ARM Build System

- Began life as a “Panda-on-a-stick”
- Embedded boards had reliability issues
  - Required special skills, images, etc.
- Migrated to industry standard ARM servers
- Provisioned using standard Red Hat technologies
  - Kickstart, puppet, etc.
- 100% Reliability record, with zero downtime

Linaro Enterprise Group

Linaro Enterprise Group

- Upstream-focused development effort
- Collaboration between Enterprise Vendors
- Ensures common implementation of standards
- Announced Nov '12 at Linaro Connect/ARM TechCon
- Red Hat represents LEG on Linaro TSC
- Red Hat assignees working on ACPI, validation (LAVA)

Fedora 19 shipping in July 2013

- Will release concurrently with x86, other architectures
- Will support many more 32-bit ARM systems
  - Includes support for emerging servers
  - Includes support for LPAE/Cortex-A15
- Most complete release of Fedora for ARM yet


AARCH
64

64-bit ARM Architecture (ARMv8, AArch64)

- Red Hat has collaborated with ARM for many years
- Engaged with every 64-bit ARM silicon vendor
  - Some will implement the ARM Architecture
  - Others will license from ARM (Cortex-A57/A53)
  - Red Hat will support both of these approaches
- Helped review elements of architecture design
- Assisted in upstream review and patches
- Porting Java (OpenJDK) to 64-bit ARM

Porting to a new Architecture 101

- Architecture design and definition
- Standardization (hardware, software)
- Hardware implementation stages
- Open Source Software porting
- Linux distribution bootstrap
- Putting it all together

Architecture design and definition

- Architecture design is iterative
  - Design parameters: low energy, reduced complexity
- Start with software models, port example code
  - Could use QEMU, or proprietary models
- Begin work on basic toolchain (binutils, gcc, etc.)
- Profile and feedback execution of example code
  - Which instructions, sequences are popular?
  - What legacy approaches no longer make sense?

Architecture design and definition (continued...)

- Implement reference architecture model
  - ARM FAST/Foundation Models
  - Software teams use these models for development
- Begin implementation of hardware (RTL)
  - Synthesis of hardware design in FPGA
  - Tape-out design into real silicon parts
  - Versatile Express/TC2 examples

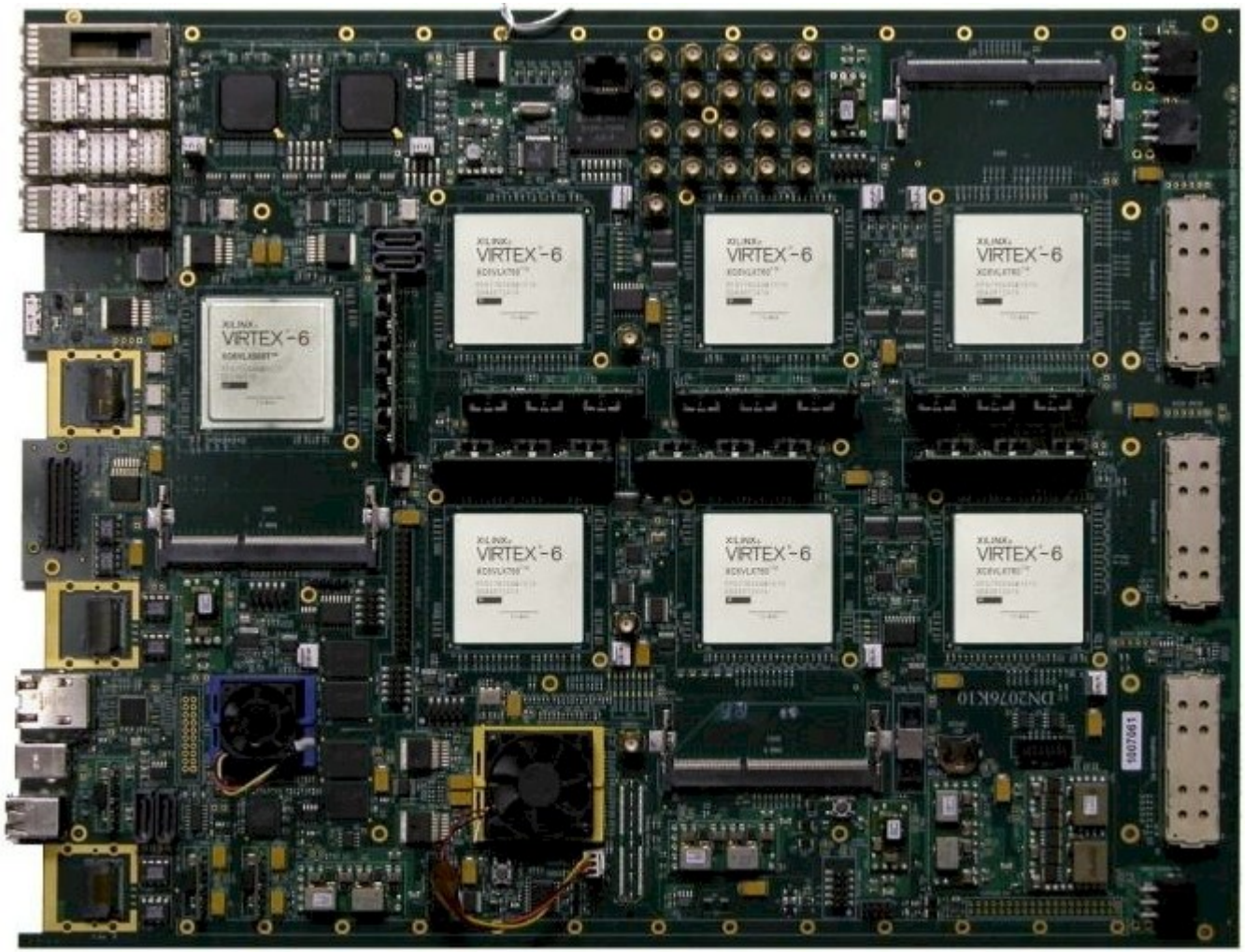


Image: Applied Micro X-Gene FPGA platform

Architectural Standardization

- My three favorite words are “shall”, “will”, and “must”
- Base architecture instruction encodings/behavior
- Required components of the core architecture
- Application Binary Interface (ABI, AAPCS)
- Standardize components of hardware platform
  - ARM cores, UARTs, device interfaces
- Standardize components of software platform
  - UEFI, ACPI, IPMI, SMBIOS, Boot Architecture

Open Source Software Porting

- Begin with binutils/gcc/gdb port to new architecture
- Port kernel in several stages (initial, iterative...)
- Port userspace libraries (glibc) to new architecture
  - gmp, mpfr, mpc, ppl, cloog, zlib...
- Patch other essential applications (autotools support)
- Co-ordinate with others to patch/optimize the rest
  - Provide examples, guide upstream developers

Porting the kernel

- Linux kernel is highly portable (31 architectures+)
- Basic kernel port has no SoC enablement/optimization
- Boot a ramdisk (AXF image) and show console output
- Iterate on kernel port to add architecture specifics
  - Optimizations (FPU lazy save/restore) come later
- Upstream review process followed by upstreaming
  - Initial support added in Linux 3.7
  - Support for SoC platforms added in 3.10

Porting Fedora - Overview

- New architecture bootstrap is very complex
- Prerequisites are a working toolchain, kernel, etc.
- Fedora is a “native built” (on hardware) distribution
  - Why is this the case?
- Package set has some complex/circular dependencies
- Leverage past experience to make life easier

Porting Fedora – Trial Run

- Practice for ARMv8 by bootstrapping ARMv7
- ARMv7 added a new optional ABI (hardfloat)
- Newer ABI not compatible with ARMv5 systems
- Perfect excuse for rebuilding the world
  - Immediate benefit: faster, cleaner, optimized builds
  - Longer term: preparation for ARMv8/AArch64

Porting Fedora - Stages

- Stage 1 - “cross build” bootstrap minimal deps
  - Disable optional docs, solve circular deps
- Stage 2 - “native build” bootstrap deps for RPM
  - “bootstrap” options in certain packages
  - Publish work in progress filesystem as a git repo
- Stage 3 - “mock build” bootstrap deps for mock
- Stage 4 - “Koji build” complete deps for Koji
- Stage 5 – Rebuild the world cleanly with Koji

Porting Fedora - AArch64

- Stage 1 – Using “cross build” on x86_64 systems
- Stage 2 – Internal using ARM FAST/Foundation Model
- Stage 3 – Public Using ARM FAST/Foundation Model
  - Documented on Fedora ARM AArch64 wiki
- Stage 4 – Public Using ARM FAST/Foundation Model
  - Freely available Quickstart images for download
- Stage 5 – Pending hardware availability

Putting it all together

- AArch64 Fedora systems will be entirely 64-bit clean
  - No 32-bit (AArch32) multi-lib, 64K page size, etc.
- AArch64 Fedora systems will boot using UEFI/ACPI
  - Standard UEFI/GRUB2 boot process adopted
- Reference model environment being built by Linaro
- Fedora 19 Remix will support early 64-bit systems
- Fedora 20/21 will support standard 64-bit systems

Image: Jon Masters speaking at Applied Micro announcement of world's first 64-bit ARM CPU

Red Hat and Applied Micro

- Applied Micro are the first 64-bit ARM silicon vendor
- Red Hat presented at unveiling of Applied/X-Gene
- Red Hat and Applied multi-year long-term partnership
  - Collaborated pre-silicon using FPGA platforms
  - Collaborated on reference board (XC-1 design)
  - Fedora will ship with X-Gene (out-of-the-box)

Demo

Image of X-C1 Board Removed From Online Presentation

64-bit ARM Demo

- 2 x Applied Micro 64-bit ARM servers with Fedora 19
- 2-3 GHz 8-core Server-on-a-Chip (Quad Issue, OoO)
- SLIMpro Lightweight Intelligent Management
  - Power Management, Secure Boot, Debug
- Fedora 19 (Bootstrap) for AArch64
- LAMP (Linux, Apache, MySQL, PHP)
- Wordpress serving content (streaming video)
- GlusterFS distributed multi-node cluster

The Future

- Fedora 20 for AArch64 systems (including X-Gene)
- 64-bit ARM servers will go into production
- Another even more awesome demo next year!