Progress Update on **PowerPC LE** and **BE**, and **CMSSW** with **Intel C++ Compiler**

David Abdurachmanov (FNAL)

2016-02-08

GCC 5.3.0 Migration

- Scheduled (?) for CMSSW_8_0_0

- Merge point of **ppc64le** and latest changes for **aarch64** into CMSDIST under a single branch, **IB/CMSSW_8_0_X/gcc530**:

  - **This brings:**
    slc6_amd64_gcc530, slc7_amd64_gcc530, slc7_aarch64_gcc530, fc22_ppc64le_gcc530

PowerPC LE

CMSSW IBs, QA, Releases

	Princeton	**CERN**
Architecture:	ppc64le	pc64le
Byte Order:	Little Endian	Little Endian
CPU(s):	160	32
On-line CPU(s) list:	0-159	0-31
Thread(s) per core:	8	8
Core(s) per socket:	10	4
Socket(s):	2	1
NUMA node(s):	4	1
Model:	IBM,8247-22L	TYAN,Palmetto
L1d cache:	64K	64K
L1i cache:	32K	32K
L2 cache:	512K	512K
L3 cache:	8192K	8192K
NUMA node0 CPU(s):	0-39	0-31
NUMA node1 CPU(s):	40-79	--
NUMA node2 CPU(s):	80-119	--
NUMA node3 CPU(s):	120-159	--
Distribution:	Fedora 22	~~Fedora 21~~

Fedora 22 now!

Issues Found

- T2_CH_CERN siteconf from CVMFS is "broken" as internally it points to AFS
- CVMFS server depends on not upstreamed "aufs" (**A**nother **U**nion **F**ile **S**ystem) module
  - CVMFS will use OverlayFS in the future, but it is only sane/stable in latest versions of kernel
- Running CMSSW validation requires re-mounting CVMFS volumes (TTL 15-60 minutes + 60 s. kernel cache drain)
- Complicated installation into cms-ib.cern.ch CVMFS repo

How to install release?

- Solution: static **proot** (5.1.0) + static **qemu-ppc64le** (2.5.0) + **fedora-22-ppc64le-rootfs** and all prepared for SLC6:

  -
    ```
    ./proot -R $PWD/fedora-22-ppc64le-rootfs -b /cvmfs:/cvmfs -b /
    build:/build -q "$PWD/qemu-ppc64le -cpu POWER8"
    ```
- Requires no root permissions (ptrace based)
- It's like a Linux container, but runs ppc64le binaries
- Use this environment to get releases installed via x86_64 machine

Quick Performance

```
                       PowerPC LE          Intel Xeon

Architecture:          ppc64le             x86_64
Byte Order:            Little Endian       Little Endian
CPU(s):                160                 72
On-line CPU(s) list:   0-159               0-71
Thread(s) per core:    8                   2
Core(s) per socket:    10                  18
Socket(s):             2                   2
NUMA node(s):          4                   4
Model:                 8247-22L @ 3.425GHz Xeon E5-2699 v3 @ 2.30GHz
L1d cache:             64K                 32K
L1i cache:             32K                 32K
L2 cache:              512K                256K
L3 cache:              8192K               23040K
NUMA node0 CPU(s):     0-39                0-8,36-44
NUMA node1 CPU(s):     40-79               9-17,45-53
NUMA node2 CPU(s):     80-119              18-26,54-62
NUMA node3 CPU(s):     120-159             27-35,63-71
Memory                 256GB CDIMM (DDR3?) 64GB DDR4
Distribution:          Fedora 22           SLC 6.7
```

```
                                      IBM PowerPC     Intel Xeon

# Physical core comparison (8 vs 2 threads)

Single thread (performance)           0.156907 ev/s   0.200261 ev/s
Multi threaded (performance)          0.155383 ev/s   0.198463 ev/s

Single thread (peak RSS)              15'190.5 MB     3'341.89 MB
Multi threaded (peak RSS)             3'145.62 MB     1'859.4 MB

# Full machine comparison (160 vs 72 threads)

Multi threaded (performance)          2.78965 ev/s    3.65784 ev/s

Multi threaded (peak RSS)             97'844 MB       38 824.2 MB
```

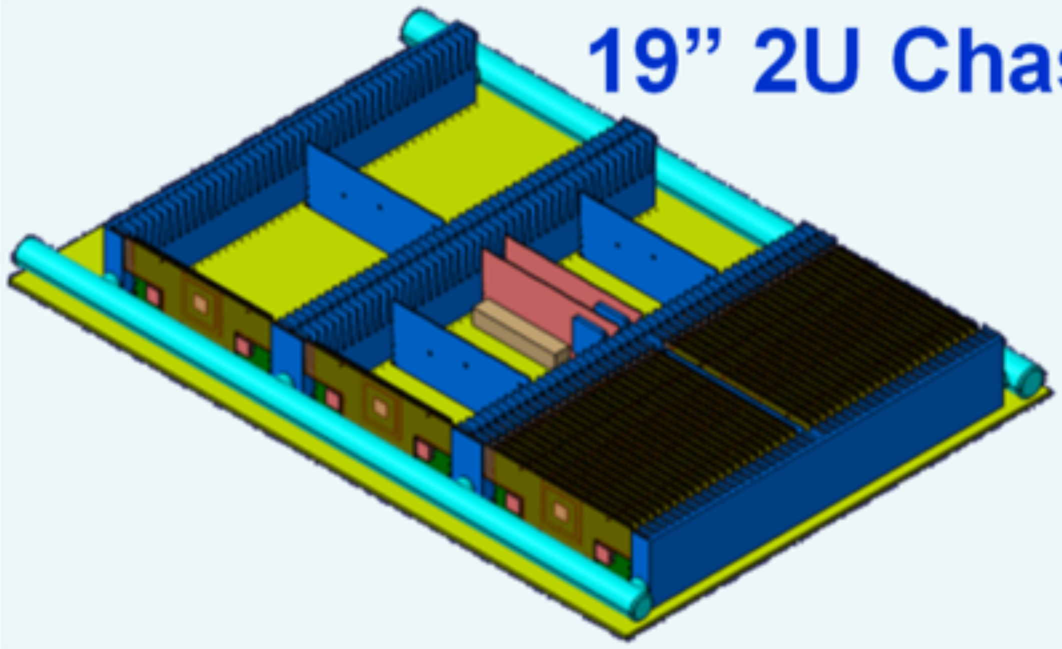
Intel Xeon (Haswell) machine provided **1.31x** more **ev/s**.
Single threaded job was consuming approx. **228MB** more of RSS on IBM PowerPC. Note, that PowerPC use **64K** pages instead of **4K**.

Benchmark was 25202.0 13TeV 25ns PU, step3 (RECO) without DQM and VALIDATION. Numbers from Framework Job Reports.

PowerPC BE

- First attempt to compile CMSSW on POWER7 BE was done
- ROOT and CMSSW ROOT based utilities segfaulted
  - Caused by a broken LLVM inside ROOT 6.06, reverting weak symbol support restored "sanity" and tools started compiling/executing
- Wrong class versions and checksums are reported due to endianness bugs in pyroot
  - It required rewiring pyroot internal type selection mechanism; patch WIP

Planned System: 2U rack unit



19” 2U Chassis w/ Combined Cooling & Power

128 compute node boards

1536 cores / 3072 Threads

6 TB DRAM

1.28Tbps Ethernet (@40Gbps)

→ Datacenter-in-a-box

- Expected 2U unit total power: ~ 6kW
- Integrated mains power converter to 12V distribution: 12V / 500A
- Each compute node has own 12V / 40W converter
- Common Power Converter boards for all other supplies
- High radix 10GbE / 40GbE switch boards (under construction)
- Connects to Mains, Rack level Water, 32x 40Gbps Ethernet
- **Hot-water cooled for efficiency and density**

Intel C++ Compiler

- CMS provides fully-compiling CMSSW_8_0_ICC_X IBs for some time already
  - Note, there are visible issues between ifort and gfortran (ABIs different)
- Recently we couldn't move forward to use new ICC releases due to breakage in compiler
  - On agreement Intel started to provide night ICC builds (incremental) for testing with CMSSW & {GCC 4.9.3, 5.3.0}
  - The exercise discovered another 3 bugs in ICC, thus we agreed it continue doing it and maybe even expand scope (ROOT, GEANT4)