SWT2, Google and ARM

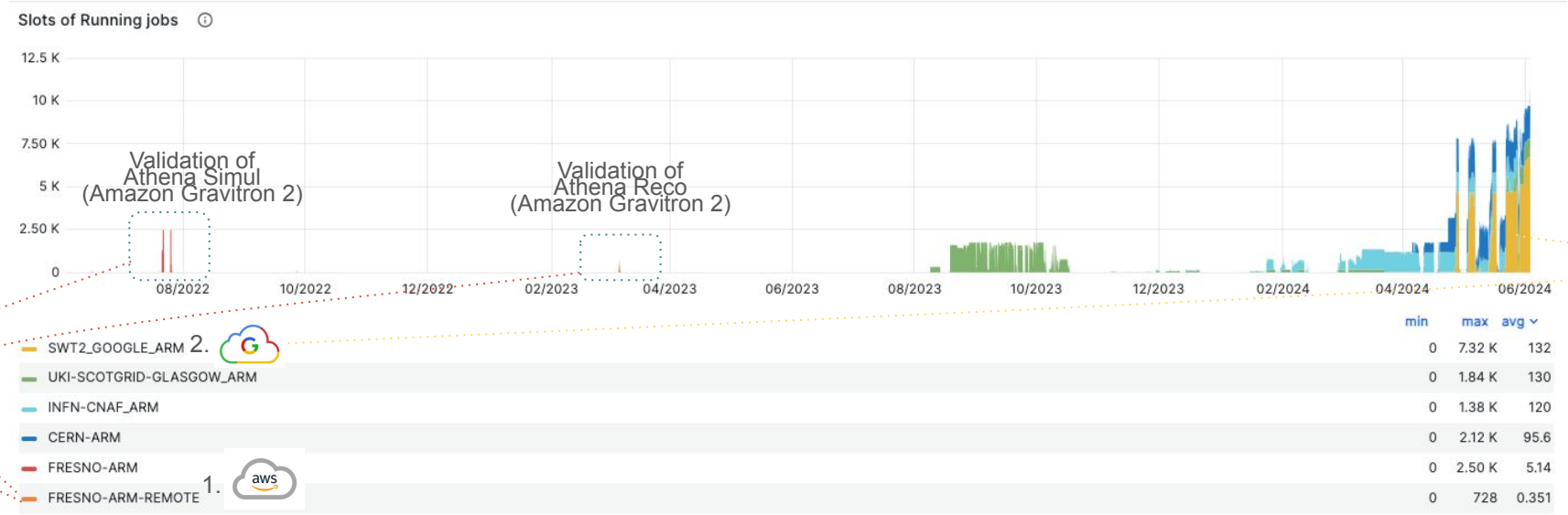
Fernando Barreiro Megino on behalf of SWT2

Grid Deployment Board, 12 June 2024

Recent (US) ATLAS Cloud projects

- US ATLAS led R&D projects on Cloud Computing (2019-2023)
  - **Phase I (~2019-2020)**: US ATLAS funded project with Google Cloud
    - Development of PanDA and Rucio integration models
  - **Phase II (~2021-2023)**: ATLAS funded project with Google Cloud
    - Extended viability of ATLAS T2 on Google during 15 months
    - Multiple R&D activities
  - **Phase II' (~2020-2023)**: Cal State (Fresno) and US ATLAS funded project on Amazon
- SWT2 phase III (2024):
    - (South West T2 = University of Texas at Arlington + University of Oklahoma)
    - Fraction of 2024 budget spent on Google Cloud subscription agreement
    - Activities in T2 scope, where the cloud can offer added value (e.g. VHIMEM, ARM)

Cloud contribution to ATLAS-ARM



1. Original collaboration to set up first ARM PanDA queue (FRESNO-ARM*) for physics validation
   a. Considerable effort to build middleware packages and extend `ALRB` configuration
2. Now contributing to ARM adoption/experience (SWT2_GOOGLE_ARM)

SWT2_GOOGLE_ARM origin

Thread exploring areas of interest for the SWT2 Google project

RW **Rodney Walker**
Random thought ... Can you get arm resources too? It could bolster the few we have to make it worthwhile assigning tasks. Another flexibility demo. Wed 4/24/2024 19:33

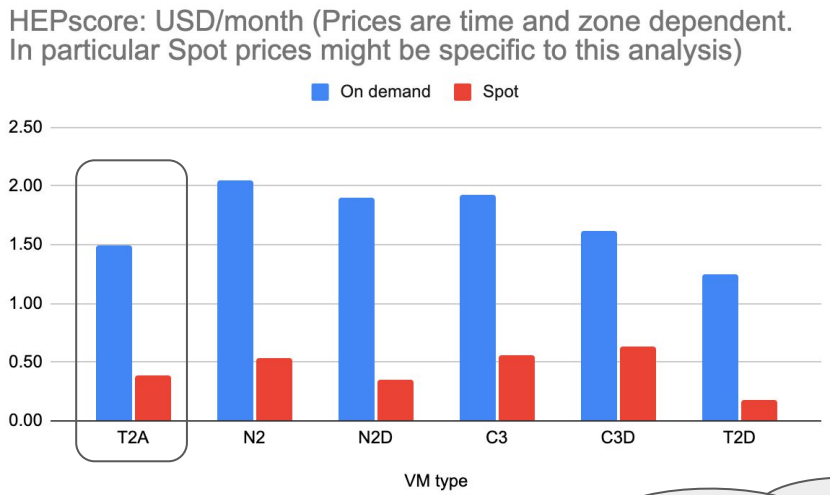
KD **Kaushik De**
ARM is a great idea. Fernando - how much work would it be to set up a new PanDA queue? Once the SE is available on Google, we can also try user analysis j... Wed 4/24/2024 19:50

F **Fernando Harald Barreiro Megino**
In principle we can get ARM resources. On Amazon it was not particularly complicated. I'm assuming on Google it will be similar, but I've never tried. I'll have a... Thu 4/25/2024 09:09

F **Fernando Harald Barreiro Megino**
Hi, (+Johannes) I created the ARM queue and it just finished the first HC job right now: https://bigpanda.cern.ch/job?pandaid=6188933235&timestamp=18:19... Thu 4/25/2024 18:22

Benchmarking VMs

- Possible to benchmark cloud VMs with HEPscore23
- GCP ARM VM families (T2A) offer good value for money



VM type	CPU type	Generation
T2A	Neoverse-N1	
N2	Intel(R) Xeon(R) CPU @ 2.60GHz	Ice Lake
N2D	AMD EPYC 7B13	Milan
C3	Intel(R) Xeon(R) Platinum 8481C CPU @ 2.70GHz	
C3D	AMD EPYC 9B14	Genoa
T2D	AMD EPYC 7B13	Milan

(not HT)

Google also plans to release their own ARM chips this year (Axion announcement)

Overall experience with ARM

- Cost effective option on Google
- Operating 4k to 8k vCPU cluster since end of April
  - Some operational hiccups due to a variety of reasons (related to ARM, cloud and site)
  - Spot instances show lower preemption rate than Intel VMs, but sometimes also funny behaviour
    - My interpretation is that the demand is not as high and the service a bit less mature
- Task assignment in ATLAS is still manual
  - Limited to a subset of activities (simulation and reconstruction)
  - Additional labor
  - Risk to have no tasks available
    - Auto-scaling shields us from sitting on an empty cluster
- ARM being more and more integrated in ATLAS
  - As of today 8 different nightly flavours for ARM
    - e.g. full Athena, simulation, analysis, new ROOT versions
  - Architecture-based brokerage available in PanDA

Conclusions

- Integration of ATLAS and ARM is advanced, but not complete
- Setting up the queue is not so complicated
  - Most work was abstracted by ATLAS SW and MW team
- Early adopters may face initial challenges
- Flexibility in cloud: freedom to choose/change architecture

Acknowledgements

- Kaushik De (UTA)
- Matt Doidge (Lancaster)
- Johannes Elmsheuser (BNL)
- Alessandro de Salvo (INFN)
- Asoka de Silva (TRIUMF)
- Steven Timm (FNAL)
- Rod Walker (LMU)