Science and Technology Facilities Council

Meerkat

A benchmarking framework

Motivation

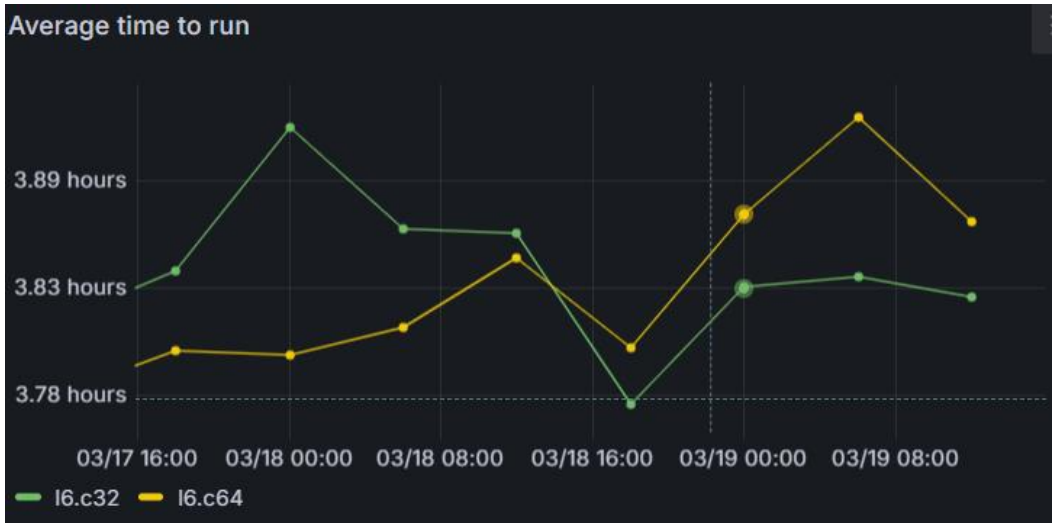
- Create a modular system which can benchmark different flavours of cloud VMs
  - I focussed on CPU benchmarks; Chris Green has been working in parallel on storage
- Called Meerkat because it pops up, does something, then goes back down
  - Spawns VM, runs benchmark then cleans up
- Provides an indicator of performance to users
  - Benchmarks are run on newly created VMs so as to replicate the typical user experience
  - Help them pick appropriate VM flavours
  - Could also be useful when setting quotas

Implementation

- Terraform to create VMs
  - Variables to set type of benchmark, flavours to run on and additional resources (e.g. storage volumes)
- Ansible to install and run benchmark
  - Each kind of benchmark (storage, CPU) has its own role
  - Designed to be easily expandable
- HEPScore benchmark used for CPU
  - Python script to pull and run containers and calculate a score
    - Score is a geometric mean of each benchmark in the suite
  - Suite of high energy physics workloads developed by CERN
  - Can define custom suites and use any containers you want
    - As long as you match the output syntax

Implementation (cont.)

- Once benchmark is run, various metrics are sent to a VictoriaMetrics instance:
  - Benchmark score
  - Time to run
  - CPU information
  - VM UUID
- Terraform deploy runs periodically with cron job
  - Currently once every 6 hours
  - Benchmarks take ~4 hours to run, but it varies even for the same flavour



Science and Technology Facilities Council

Visualisation

- Grafana!

Science and Technology Facilities Council

Extension

- Investigate custom benchmarks to better reflect cloud user communities
  - Could also aim to bring the runtime down to get finer-grained information
  - Default HEPScore remains useful for comparing with other systems
- Implement other kinds of benchmark using the Meerkat framework
  - GPU
    - I looked into this, but it was difficult to find relevant benchmarks (lots are very visualisation focussed) or I couldn’t get them to run
  - Networking
    - Potentially much more complex than CPU and storage

Science and Technology Facilities Council

Thank you

Facebook: Science and Technology Facilities Council

Twitter:@STFC_matters

YouTube: Science and Technology Facilities Council