In contrast to at EPCC (Section 4.1), the workload at ARCTUR HPC centre is characterized by a distinct "on-demand" utilisation pattern
In contrast to at EPCC (Section 4.1), the workload at ARCTUR HPC centre is characterized by a distinct "on-demand" utilisation pattern

Kronos

Benchmarking HPC Systems with Realistic Workloads

Antonino Bonanni, Simon D. Smart and Tiago Quintino

[1] European Centre for Medium-Range Weather Forecasts, Reading, UK

NEXTGenIO Project Partners: EPCC, ECMWF, ARCTUR, INTEL, Fujitsu, BSC, TU-Dresden

Introduction

Existing HPC benchmarking tools test single components of a system in isolation. Whilst this approach is valuable it fails to give a realistic evaluation of the system as a whole. A benchmarking tool that makes use of all of the components of an HPC system (including I/O, compute, network and scheduling) gives results more representative of operational conditions.

Kronos generates and executes workloads representative of the real-life computational workloads of HPC centres in a highly controlled and easily portable way. Furthermore, it aims to generate these workloads automatically, based on analysis of data collected from measured operational workloads.

Kronos is being developed at the European Centre for Medium-Range Weather Forecasts ECMWF as part of the NEXTGenIO project. The NEXTGenIO project is a 3-year EU-funded Horizon 2020 project started in October 2015. It is coordinated by the Edinburgh Supercomputing Centre (EPCC) and involves partners from several European countries. It aims to develop innovative solutions to the emerging I/O bottlenecks as high-performance approaches exascale.

HPC Workload Modelling Methodology

Kronos analyses profiling data collected from real workloads executing on HPC systems and generates synthetic model workloads through a modelling process (Figure 1). These synthetic workloads can then be expressed on a real HPC system using a set of simple, configurable and highly portable *"synthetic"* applications.

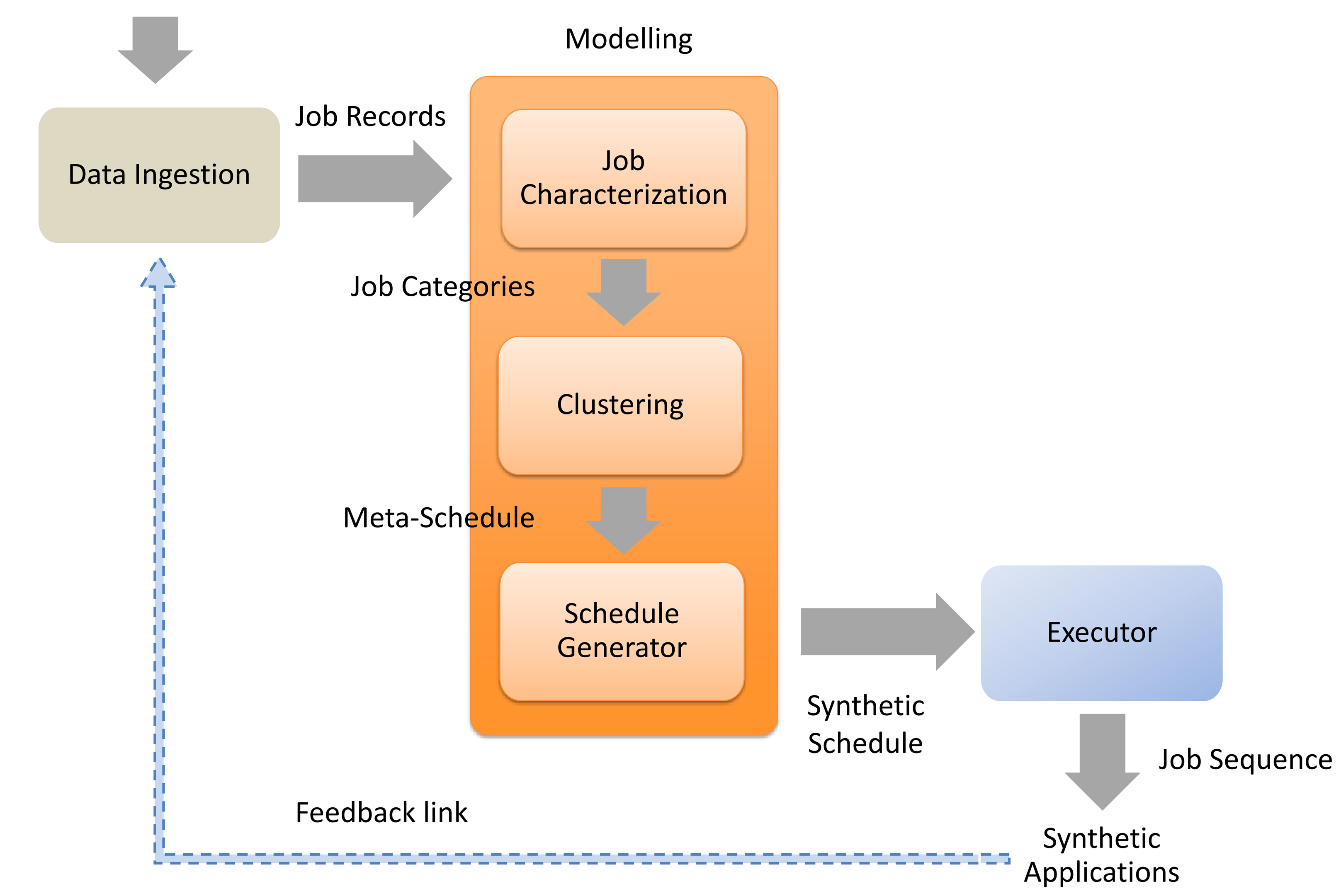
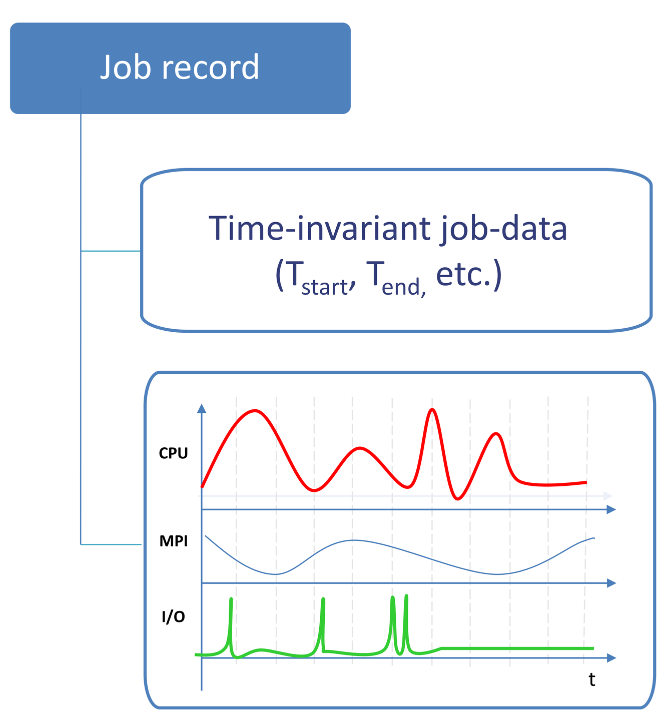


Figure 1: Kronos components and data pipeline, coloured according to the software tools; Ingester *(grey)*, Modeller *(orange)* and Executor *(blue)*

The modelling process can be described by the following steps

1. ***Data ingestion***. Workload profiling data is obtained and translated into the Kronos Profiling Format for use by Kronos. This data contains job records with metadata and time-dependent profiled values corresponding to a series of defined HPC metrics (Figure 2).
2. ***Job characterization***. Workload data is sparse; only a subset of metrics are available for each of the jobs and Machine Learning techniques are used to fill in the incomplete information.
3. ***Clustering***. Jobs are clustered according to their associated data such that jobs with similar characteristics form clusters from which a reduced set of prototypical jobs can be formed. These are then used to spawn the synthetic applications.
4. ***Schedule generation***. A schedule is built for abstract, idealised synthetic applications to reflect the profiled workload. This takes into account any modification or scaling factors passed in by the user.
5. ***Execution***. The abstract schedule is passed to the Kronos Executor. This submits concrete synthetic applications to the scheduling queue of a real HPC system.
6. **Feedback.** Profiling data collected when running the synthetic applications on the target HPC system can optionally be fed back into the modeller to fine-tune the target parameters.



- Time-invariant data
- Time-dependent job metrics defines job "fingerprints"
- Job "fingerprints" identify jobs according to their usage of the HPC resources: computing, network and I/O)
- Job fingerprints are then used to identify job "clusters" (jobs that behave similarly from the system perspective)

Figure 2: Each job record is described by both time-invariant and time-dependent data.

HPC Workload Data Collection

To provide Kronos with real-life input data, a workload profiling and data collection process was carried out in 2016 as part of the NEXTGenIO project

- HPC centres: **EPCC, ECMWF** and **ARCTUR.**
- A wide range of profiling tools were used, including both off-the shelf products and bespoke tools.
- **Allinea** and **TU-Dresden** provided extensive expertise in developing HPC profiling tools.
- Profiling strategies have been developed to cope with wildly different workloads and usage patterns at the three HPC centres

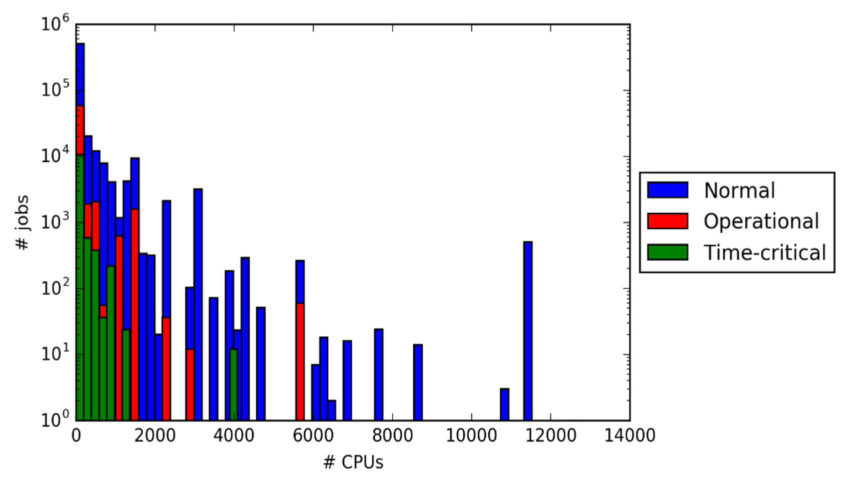


Figure 3: Histograms of jobs run at ECMWF centre categorized according to the type of workload.

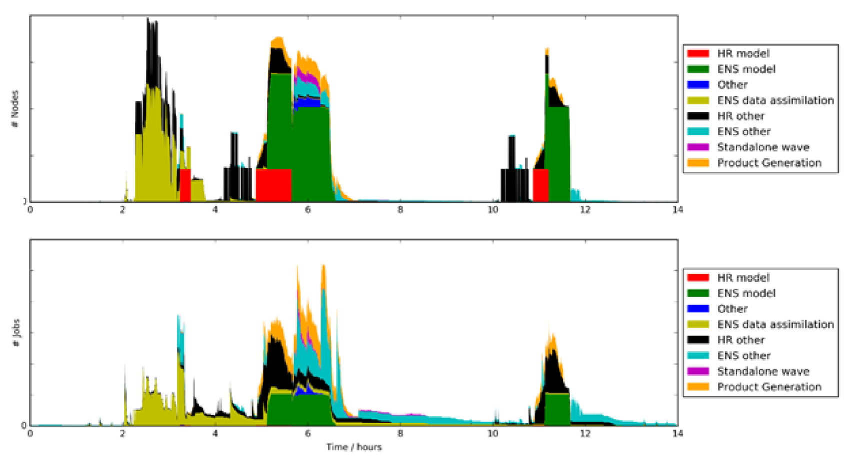
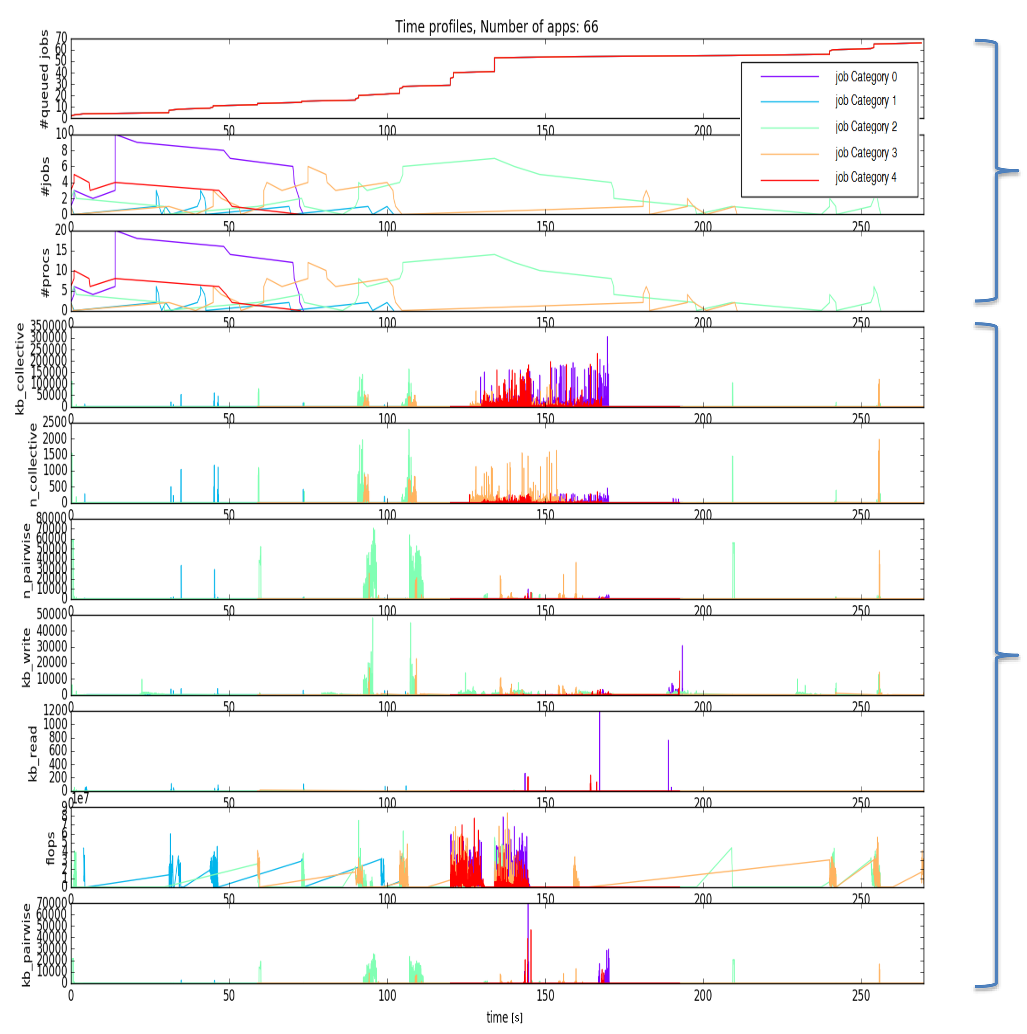


Figure 4: Example of timeline workload profiled collected at ECMWF (#jobs and #processors – sanitized values)

Preliminary Kronos Runs

Kronos is currently at an early stage of development, and therefore modelling uncertainty needs to be analysed and quantified. Figure 5 shows an example of collective time-dependent metrics for a small test run with a workload of **66 synthetic applications** generated from an operational workload at ECMWF. Testing, validation and comparison against profiled workload is on going.

Figure 5 shows the time-dependent metrics that have been collected from each synthetic application during the run of Kronos. These profiles are in the process of being compared against the original workload time profiles and similarity being defined and measured.



- Number of submitted jobs
- Concurrently running jobs
- Concurrently used processors

- Kronos time-dependent metrics
  - CPU metrics
  - MPI P2P and collective
  - I/O throughput and number of read/write

Figure 5: Example run of Kronos workload. Time dependent job-metrics of synthetic applications are collected and summed up in a time-line for analysis

Summary and Conclusions

- Workload profiling data have been collected at the HPC centres: EPCC, ECMWF and ARCTUR and used to develop a first working prototype of Kronos.
- ECMWF is also considering the usage of Kronos for overall system performance evaluation.
- Future work will focus on developing further modelling strategies and in particular on the process of scaling the generated workloads to model systems of different scales.
- Kronos will be used within NEXTGenIO to support the evaluation of the novel hardware platform that is being developed as part of the project.

Further Reading

1. NEXTGenIO project: *www.nextgenio.eu*
2. Allinea PR/MAP *http://www.allinea.com*
3. Score-P. from *http://www.vi-hps.org/projects/score-p*
4. Vampir. from *https://www.vampir.eu*