International Centre for Radio Astronomy Research

PAWSEY supercomputing centre

AUSSRC-DSP-2019-0002

AusSRC Prototyping Platform ***Strawman design***	
DESIGN STUDY PROGRAM	
Date:	**21.10.2019**
Work package:	**AusSRC Core**
Dissemination level:	**AusSRC MC**
Published at:	

Abstract

The document outlines a strawman architecture for AusSRC prototyping system that could be developed and tested with selected ASKAP and MWA survey projects.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	**Partner/WP**	**Date**
From	AusSRC DSP		
Author(s)	S.Kitaeff		
Reviewed by	A.Wicenec, J.C.Guzman		21-Oct-2019
Approved by			

III. DOCUMENT LOG

Issue	**Date**	**Comment**	**Who**
1	20-Sep-2019	Created	S.Kitaeff
2	8-Oct-2019	First draft	S.Kitaeff
3	21-Oct-2019	Second draft	S.Kitaeff

IV. GLOSSARY

PI Principal Investigator

AENEAS	Advanced European Network of E-infrastructures for Astronomy with the SKA
ALMA	Atacama Large Millimeter/submillimeter Array
API	Application Programming Interface
ARC	ALMA Regional Centre
ASKAP	Australian SKA Pathfinder
ASVO	All-Sky Virtual Observatory
ATNF	Australia Telescope National Facility
CASDA	CSIRO ASKAP Data Archive
CLI	Command Line Interface
CSP	Central Signal Processor
ESA	European Space Agency
ESDC	European SKA Data Centre
ESRC	European SKA Regional Centre
EVN	European VLBI Network
FFT	Fast Fourier Transform
FOV	Field of View
GSM	Global Sky Model
GUI	Graphical User Interface
HPC	High Power Computing
HPSO	High Priority Science Objective
HST	Hubble Space Telescope
HTC	High Tech Computing
IVOA	International Virtual Observatory Alliance
JIVE	Joint Institute for VLBI ERIC
LBA	Long Baseline Array
LOFAR	LOw Frequency ARray
LSM	Local Sky Model
MWA	Murchison Widefield Array
NASA	National Aeronautics and Space Administration
NVME	Non-Volatile Memory Express
QA	Quality Assessment
SDP	Science Data Processor
SED	Spectral Energy Distribution
SG	Science Gateway
SIAP	Simple Image Access Protocol
SKA	Square Kilometre Array
SRC	SKA Regional Centre
SSAP	Simple Spectral Access Protocol
SSD	Solid State Drive
TAP	Table Access Protocol
VLBA	Very Long Baseline Array
VLBI	Very Long Baseline Interferometry
VO	Virtual Observatory

V. LIST OF FIGURES

VI. LIST OF TABLES

VII. EXECUTIVE SUMMARY

It has been well recognised, including through DSP work, that classical HPC system architecture of compute and storage infrastructure in data commonly deployed in research centres, including Pawsey, are not optimal for data-intensive applications in general and radio astronomy in particular. The AusSRC DSP needs to investigate the options for hardware and system architectures that are optimally suited for the needs of SKA precursors and later for the SKA1 post-processing. This can best be achieved through prototyping the elements of such a system and testing them on ASKAP and MWA use-cases. The SDP CDR documentation already provides useful concepts, though the prototyping of software stacks on suitable hardware platforms for the SRC and post-processing needs is currently limited. Such prototyping is requires a considerable amount of systematic work and needs to be performed in collaboration with the international SRC partners. The ERIDANUS project initiated such work in 2017, by exploring and testing some essential concepts in two hackathon style workshops. ShAO, as the Chinese SRC lead organisation, has already made a significant investment in four different platforms for a similar prototyping work. Such work needs to begin in Australia to determine the future directions of development of the AusSRC, and to inform Pawsey on how the future systems should look like for the needs of the SKA precursor post-processing.

Currently, there is no system at Pawsey suitable for such prototyping. Some of the initial prototyping could be done in a public cloud such as Microsoft Azure, AWS or Google Cloud, in principle. However, the limited budget of the AusSRC DSP will not allow ASKAP and MWA science teams to use such prototypes for any large-scale processing or services. To make the prototyping work more practically useful, AusSRC DSP needs a hardware platform of sufficient size and capability to be able to post-process ASKAP/MWA scale data.

The white paper outlines a strawman architecture of a prototyping system that could be acquired and developed for the purposes described above.

Table of Context

1. Introduction

The Australian-hosted SKA-low will produce around 300 PB per year of data that science teams around the globe will need to access readily. Another 300 PB will be produced by the SKA-mid in South Africa. The SKA Observatory has no provisioning to store this data long term. Instead, SKA member states are forming a collaborative network of SKA Regional Centres (SRCs) to design, build, deliver, and operate end-to-end support for science data products, archives, and associated services.

Those SRCs will:

- Provide data flow and data dissemination solutions from the SKA to users;
- Store, publish and curate SKA data long-term;
- Post-process and analyse SKA data products;
- Provide SKA data and processing user support.

The Australian SRC Design Study Program is $4m program, funded by the Australian Government and CSIRO. It will define an AusSRC design and costing based on requirements and experiences gathered from the Australian and regional communities and the Australian SKA precursors ASKAP and MWA.

A significant body of work is required to develop the SRC concept. The program will use a top-down analysis of SRC requirements in global collaboration with other SRCs and the SKAO, and a bottom-up approach solving practical computational and data problems within the SKA precursor projects, leading to the design and prototyping of the architecture of the future AusSRC.

Part of what the program needs to achieve is to identify, assess, and test potential solutions for providing tools and services to post-process data from ASKAP, MWA and later SKA to extract the parameters leading to new scientific insights and publications.

2. Goals of this document

The document communicates some of the underlying requirements for the architecture of a system for data-intensive post-processing, and how such requirements could be met with the existing and developing software solutions. The document outlines a strawman architecture of a small prototyping system that if acquired could be used to verify the proposed architectural solutions using ASKAP and MWA survey data.

3. HPC and BDC

High Performance Computing (HPC) platforms (both hardware and software) have evolved over the last 2.5 decades into large and efficient computational clusters optimised for modelling and simulation workloads from science, industry, public decision making and commerce perspectives that require extreme amounts of computation and a class of highly tuned stacks that provide scalable compute and local communication capabilities that surpass hardware capabilities of a single server and can take on the aforementioned workloads. HPC system such as Pawsey's Magnus cluster are typically optimised for low latency process-to-process in distributed memory communications and load balancing as the computations are completed as fast as the slowest process in run-time. However, many radio astronomy applications in post-processing stage (source finding, cross matching, stacking etc) are not utilising these fast process-to-process communications due to the local nature of computations with respect to data, and data can be relatively easily partitioned in frequency, spatial or polarisation domains leading to almost embarrassingly parallel execution of post-processing pipelines[1].

Lately, Big Data Computing (BDC) has been recognised as a new trend in optimisation of large computing systems [RD1], thanks to the growing commercial sector of machine learning. Such systems are used to collect, organise and analyse large sets of data, often called Big Data. Big Data typically means datasets possessing data volume, velocity and variety characteristics, which are so large that it is difficult to process using traditional database and software techniques. Radio astronomy is an extreme case of Big Data with the datasets in post-processing ranging from hundreds of gigabytes to multiple TB for ASKAP, and upto multiple petabytes for the SKA-1. One of the typical characteristics of processing such datasets is that unlike in computer simulations and modeling HPC applications, the computations can't begin until a dataset is loaded into memory, leading to significant inefficiency of processing on a system that is optimised for simulations or modeling, and the wait time can be significant, leading to underutilisation of a system and resulting in long processing times. SSD, NVME and large memory per CPU-core available in each computational node of a cluster are necessary to address the I/O issue in radio astronomy applications. In addition, this data is not usually readily available for processing and needs to be staged, typically, from a tape storage to higher tier storage, which leads to significant latencies before the processing can even start and complicates the process for an unprepared end user - a scientist, post-doc, research student.

[1] Global FFT is commonly used in imaging algorithm, and requires intensive interprocess communications, however SRC post-processing is mostly concerned with Level 6 & 7 data, when global FFT is not normally required, with EoR perhaps being an exception.

The difference in requirements for HPC and BDC have been long recognised and lead to the development of a number of software tools, such as Hadoop, Spark, YARN, HIVE etc. that help manage the data and its analysis in new, better optimised ways. While these tools may offer significant improvements in data processing in astronomy, the adoption and adaptation of new technologies is slow due to the fact that the majority of research data centres, including Pawsey, are still providing mostly HPC infrastructure for modeling and simulations. It's important to note that the selection of software and hardware solutions need to be done with care and thorough testing as radio astronomy data is typically much larger and has a significantly different structure than commercial data, for which the tools had been developed. Some development of new tools that have an optimal fit for research is also happening (e.g. Singularity, PythonHub, etc).

AusSRC is placed ideally to collaborate with Pawsey and ASKAP/MWA teams on investigating new hardware and software solutions that will make post-processing of radio astronomy data significantly more efficient and optimised for data-centric processing.

4. Tools and environment

The Astronomy community has developed a wealth of software tools to inspect, manipulate and analyse astronomy data, including radio interferometric data. The vast majority of those tools are standalone software packages that can be downloaded and installed on a personal computer. The data then also needs to be downloaded and copied to local storage where the software can access it. Although these tools contain smart algorithms and are excellent in features that they offer, downloading data to a local computer for inspection, exploration and analysis is no longer a suitable framework because the size of the datasets well exceeds the capability of a single computer. This is already the case for ASKAP and MWA, although parts of data still can be downloaded to a local computer; however, such an approach is utterly unrealistic for the SKA data.

The SRCs will be the SKA specific data science platform[2] that will provide users with a complete software environment that contains all the tools, which can interact with data within the SRC system without the need to download tools or data. This would make automation of post-processing significantly more accessible, faster and overall more efficient as it'll minimise the data movement and the operations can be completed through a standard SRC API. The interaction with data can be made directly through and within the archive.

[2] https://whatis.techtarget.com/definition/data-science-platform

5. Strawman design

Figure 1 shows the proposed strawman system architecture design envisioned for AusSRC.

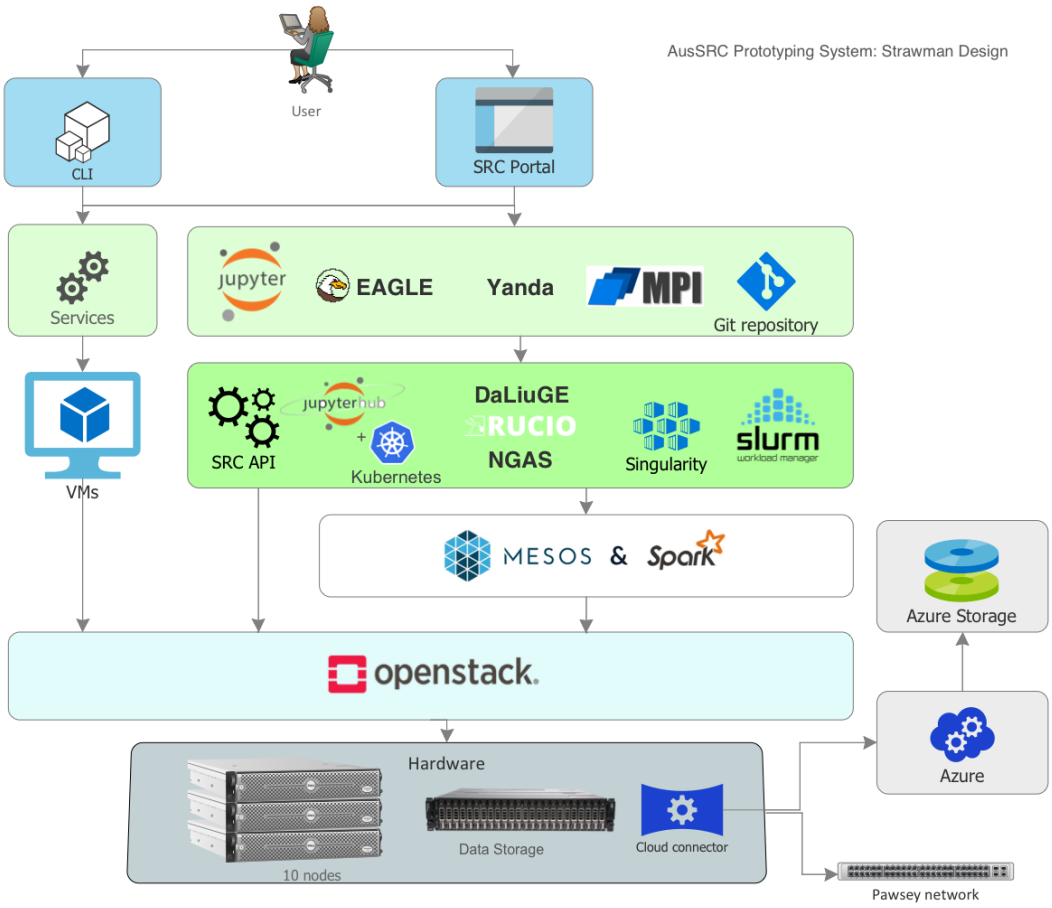

Figure 1. AusSRC strawman system architecture. The arrows should be generally interpreted as "uses". The layers of the stack are rather relative, because there are horizontal "uses" between the elements in each layer as well as vertical between the layers.

The critical attribute of the proposed architecture is the use of Openstack[3] software to abstract and virtualise the hardware to the rest of the software stack. Such virtualisation enables flexible use of the hardware in infrastructure-as-service fashion. Such a solution mirrors the current baseline design for the main SKA processing to which SRCs compute requirements are quite similar.

[3] https://www.openstack.org

Provided by the Openstack, Virtual Machines (VM) then used for various services such as data ingest, archive interface or specific survey user portals. Apache Mesos[4] & Spark[5] software suits manage the rest of the system resources.

Singularity[6], as a containerisation tool for high-performance computing applications minimises the effort to maintain the software (including legacy software) on the target system. Slurm and MPI can be made available as well, though Spark and DaLiuGE offer an entirely new level of managing large-scale post-processing in the AusSRC.

It is envisioned that the system needs to be able to offload tasks from the system in Pawsey to a cloud, on-demand. This way, the processing tasks that need immediate deployment or requires resources that are unavailable within the SRC can be deployed outside the SRC system. This kind of scalability in the form of a prototype has been discussed as a collaborative project with Microsoft Azure. A similar integration can also be explored with other cloud providers, such as AWS or Google.

Finally, the user should have two ways to interact with the system. One is through a command-line interface (CLI), which enables triggering the execution of processing on the system from a user computer, and the second is more interactive, through a web portal that provides a user with an instance of JupiterLab connected directly to the base of the system.

The SRC API will provide a unified way for all the elements of the system to interact with data, processes and each other.

6. Hardware Requirements for the Prototype

To prototype such an architecture, we need a hardware system with at least 10 very capable nodes and 1 PB disk storage with 10% SSD burst buffer, all connected to the Pawsey network, and preferably fully hosted by Pawsey.

Node specifications:

- Processors - 24 or 2 x 12 cores (possibly AMD Epyc)
- RAM - 2x384 GB (or more)
- NVME SSD - 6TB
- Interconnect - Infiniband or Ethernet

One of the nodes should host 2 or 4 NVIDIA V100 or P100 GPUs, as the budget allows.

[4] http://mesos.apache.org
[5] https://spark.apache.org
[6] https://sylabs.io/docs/

7. Use of the system

Initially, the system needs to undergo thorough benchmarking and optimisation work. However, the best improvement is going to come from the practical use of the system for the selected precursor projects which will receive AusSRC DSP assistance. The lessons learned, knowledge and experience developed in the process of the optimisation of such a system for the precursor projects will inform and benefit the future users and systems of the AusSRC, as well as the broader community at Pawsey.

8. Conclusions

The proposed strawman architecture aims at prototyping the transition from HPC to BDC for radio astronomy post-processing with the specific focus on ASKAP, MWA and future SKA post-processing within the SRCs. It's a flexible architectural solution for SRC computational and data needs. The solution will be tested and improved via practical use of the system by the AusSRC DSP assisted precursor survey projects.

VIII. REFERENCES

[RD1] M.D.Assunção et al, Big Data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing, v 79–80, 2015, p 3-15, https://doi.org/10.1016/j.jpdc.2014.08.003