Thursday, November 9

Week 45 of 2023

- Home
- Moon
- Work
- UA Holidays
- NOAO
- Siri Suggestions
- ADASS 2023 Calendar
- Telescopes

November 2023

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

December 2023

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

7AM

7:30AM MORNING COFFEE
Where
POSTER ROOM (Catalina/Tucson rooms)
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]

8AM

8:30AM C902: High performance visualization for Astronomy &
8:45AM C701: A brightening future for research software engineer-

9AM

9:00AM INVITED I701: From AI to L2 and beyond: A software engineer career
9:30AM PRIZE I901: Decades of Transformation: Evolution of the NASA

10AM

10:00AM COFFEE BUFFET
Where
POSTER ROOM (Catalina/Tucson rooms)
10:00AM Announce-
10:15AM FOCUS DEMO F802: Empowering SKA Data Challenges: A homogeneous

11AM

11:00AM C903: The VLA Sky Survey
11:15AM C904: Lessons learned from building LOFAR data pipelines
11:30AM C905: Towards automated structural analysis of galaxies in
11:45AM C802: Stimela 2, kubernauts, and dask-ms: radio interfer-

Noon

12:00PM LUNCH BREAK

1PM

1:30PM C205: Using The NEOfixer API for NEO Follow-Up and NEO
1:45PM C804: XMM-Newton Science Analysis System (SAS) on the

2PM

2:00PM INVITED I801: Open Source Science Initiative at NASA
Where
2:30PM C805: Taking TESSCut to the Cloud: Architecting for Avail-
2:45PM C806: A Good IDIA : Scientific Computing at Scale

3PM

3:00PM AFTERNOON BREAK
Where
POSTER ROOM (Catalina/Tucson rooms)
[https://union.arizona.edu/in-
3:15PM FOCUS DEMO F101: Exploring the Dark Side of the Uni-

4PM

4:00PM C807: NASA Archival Data in The Cloud: Service & Dis-
4:15PM C808: Processing All- Sky Images At Scale On The Amazon
4:30PM C809: Preparing a scientific data processing facility for Ru-
4:45PM C810: Rubin Science Platform: on cloud, on-prem, all of the

5PM

5:00PM IVOA Executive Committee Meeting

6PM

7PM

8PM

• Timed Events

7:30AM to 8:30AM MORNING COFFEE
Location: Where
POSTER ROOM (Catalina/Tucson rooms)
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

Coffee breaks will be held in the poster room.

Hot beverage service.

8:30AM to 8:45AM C902: High performance visualization for Astronomy & Cosmology: the VisIVO's pathway toward Exascale systems
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: OTHER CREATIVE TOPICS IN ASTRONOMICAL SOFTWARE
[https://adass2023.lpl.arizona.edu/themes-0#az-accordion--9]

pretalx [https://pretalx.com/adass2023/talk/WKXV7Y/]

Petabyte-scale data volumes are generated by observations and simulations in modern astronomy and astrophysics. Storage, access, and
data analysis are significantly hampered by such data volumes and are leading to the development of a new generation of software tools. The Visualization Interface for the Virtual Observatory (VisIVO) has been designed, developed and maintained by INAF since 2005 to perform multi-dimensional data analysis and knowledge discovery in multivariate astrophysical datasets. Utilizing containerization and virtualization technologies, VisIVO has already been used to exploit distributed computing infrastructures including the European Open Science Cloud (EOSC).

We intend to adapt VisIVO solutions for high performance visualization of data generated on the (pre-)Exascale systems by HPC applications in Astrophysics and Cosmology (A&C), including GADGET (GAlaxies with Dark
matter and Gas) and PLUTO simulations, thanks to the collaboration within the SPACE Center of Excellence, the H2020 EUPEX Project, and the ICSC National Research Centre. In this work, we outline the evolution's course as well as the execution strategies designed to achieve the following goals: enhance the portability of the VisIVO modular applications and their resource requirements; foster reproducibility and maintainability; take advantage of a more flexible resource exploitation over heterogeneous HPC facilities; and, finally, minimize data-movement overheads and improve I/O performances.

Acknowledgements: This work is funded by the European High Performance
Computing Joint Undertaking (JU) and Belgium, Czech Republic, France, Germany, Greece, Italy, Norway, and Spain under grant agreement No 101093441 and it is supported by the spoke "FutureHPC & BigData" of the ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing – and hosting entity, funded by European Union – NextGenerationEU"

8:45AM to 9:00AM C701: A brightening future for research software engineering
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: RESEARCH SOFTWARE ENGINEERING AS A CAREER PATH
[https://adass2023.lpl.arizona.edu/themes-0#az-accordion--7]

pretalx [https://pretalx.com/adass2023/talk/EKVAVU/]

It is understood that astronomy relies on research software and data engineering. From the collection of telescope proposals, the control of telescopes and their miriad instruments, to driving the archives, simulating and processing data, research software engineering underpins almost every process in the advancement of astronomy. And yet, for a large number of projects and institutes, in planning and funding conversations, the requirements of the discipline for producing the best results have at times been an afterthought, receiving little attendion or funding. Although our more enlightened institutes have always valued software engineering the community at large is slowly coming to realise that the discipline must be supported and career paths nurtured, so that the best science can be carried out.

In this talk I will discuss some of the joy and pain of pursuing a research software engineering career within astronomy, and the problems we must tackle if we wish to continue to attract excellent creative, engineering, and scientific minds to our field. Not just attract them but retain them, in an era where flexible working conditions are no longer a perk of academia, and salary disparity between our institutions and industry is larger than ever.
I will describe the AAO's Research Data & Software section's work to provide a stable career path for its research software and data engineers, and our aims to attract a portfolio of work which both satisfies the needs of the instrumentation and data projects of the community, and the needs of our team to have a challenging, creative, and fulfilling work life.

9:00AM to 9:30AM INVITED I701: From AI to L2 and beyond: A software engineer career
turned into a journey through fascinating territories, landscapes and,
of course, languages
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: RESEARCH SOFTWARE ENGINEERING AS A CAREER PATH
[https://adass2023.lpl.arizona.edu/themes-0#az-accordion--7]

pretalx [https://pretalx.com/adass2023/talk/CFW3DG/]

A talk about my career as a software engineer and how I like to see it as a journey through many different territories, landscapes and, of course, languages. Do not expect leadership or big money: I have been, and still I am, only a modest but curious traveler. My journey started more than thirty years ago in the land of AI with a thesis on an "Expert System", as rules-based AI was called at the time, and continued through remote observing, the birth of the web, telescope control systems, detector controllers, programming standards, virtualization, space applications (L2), MBSE, software quality assurance, and still continues even if the final station is not so far away. Different landscapes under the sky of Astronomical projects: Spectral classification, AI, Electronics, System Engineering, Control Software, Quality Assurance… different organizations and different countries. Also, different languages, some of them long forgotten: from Fortran to Occam, Ksh, Bash, C, C++, Java, Python etc. Not a journey that will make it to any travel guide but maybe it can give hints to other fellow travelers when the moment of picking the next destination arrives. I then also hope this travel experience can be of interest to the people visited by travelers like me in understanding what the spirit of a wandering software engineer is.

9:30AM to 10:00AM PRIZE I901: Decades of Transformation: Evolution of the NASA
Astrophysics Data Systems Infrastructure
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: OTHER CREATIVE TOPICS IN ASTRONOMICAL SOFTWARE
[https://adass2023.lpl.arizona.edu/themes-0#az-accordion--9]

pretalx [https://pretalx.com/adass2023/talk/LZ9FQV/]

The NASA Astrophysics Data System (ADS) is the primary Digital Library portal for researchers in astronomy and astrophysics. Over the past 30 years, the ADS has gone from being an astronomy-focused bibliographic
database to an open digital library system supporting research in space and (soon) earth sciences. In this talk I will describe the evolution of the ADS system, its capabilities, and the technological infrastructure underpinning it.

I will begin with an overview of the ADS's original architecture, constructed primarily around simple database models. This bespoke system allowed for the efficient indexing of metadata and citations, the digitization and archival of full-text articles, and the rapid development of discipline-specific capabilities running on commodity hardware. The move towards a cloud-based microservices architecture and an open-source search engine in the late 2010s marked a significant shift, bringing full-text search capabilities, a modern API, higher uptime, more reliable data retrieval, and integration of advanced visualizations and analytics.

Another crucial evolution came with the gradual and ongoing incorporation of Machine Learning and Natural Language Processing algorithms in our data pipelines. Originally used for information extraction and classification tasks, NLP and ML techniques are now being developed to improve metadata enrichment, search, notifications, and recommendations. I will describe how these computational techniques are being embedded into our software infrastructure, the challenges faced, and the benefits reaped.

Finally, I'll conclude by describing the future prospects of ADS and its ongoing expansion, discussing the challenges of managing an interdisciplinary information system in the era of AI and Open Science, where information is abundant, technology is transformative, but their trustworthiness can be elusive.

10:00AM to 10:10AM Announcement of ADASS XXXIV in 2024
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]

10:00AM to 11:00AM COFFEE BUFFET
Location: Where
POSTER ROOM (Catalina/Tucson rooms)
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

Coffee breaks will be held in the poster room.

Continental buffet.

10:15AM to 10:45AM FOCUS DEMO F802: Empowering SKA Data Challenges: A homogeneous
platform for enhanced collaboration and scalability fully aligned with
Open Science.
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

PREV
[https://adass2023.lpl.arizona.edu/event-categories/thursday-session-1]NEXT
[https://adass2023.lpl.arizona.edu/event-categories/thursday-session-2]

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS
[https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/MS7FEW/]

The Square Kilometre Array Observatory (SKAO) is an international collaborative effort focused on constructing and operating the world's most advanced radio telescope. The SKAO Science Data Challenges (SDCs)
are a series of competitions that are designed to help scientists and engineers develop new techniques for analysing the vast amounts of data that the SKAO will generate. These SDCs have traditionally been conceived to use computing resources kindly provided by scientific

institutions and facilities. The method of allocating computing resources for participants in the Data Challenges has varied among resource providers, resulting in a heterogeneous user experience where the users have access to Virtual Machines (VMs) with differing configurations, while others provide HPC-type resources. Providing an uniform platform for computing resources for SDC brings fairness, scalability, enhanced collaboration and consistency. Participants work with equal tools and streamlined collaboration. A standardised setup simplifies resource management, support, and evaluation, leading to enhanced efficiency and reliable results.

JupyterHub provides a platform for provisioning compute resources through a container orchestration service such as Kubernetes, in addition to providing user demand scaling, and enabling centrally managed authentication. The advantages of this approach include ease of deployment through Helm, homogenisation of the customisation for software and compute environment needed for the SDC, and horizontal scalability by allowing resources to be allocated to users by the Kubernetes cluster based on demand and availability.

With this contribution we want to present a highly portable, interactive and fully OpenScience-aligned analysis service for future participants in different Science Data Challenges to develop solutions on a horizontally scalable platform within the infrastructures of the SKA Regional Centres Network (SRCNet) and other IT facilities. In this context, we will show the process of configuring the Kubernetes cluster, the installation and preparation for BinderHub/JupyterHub, as well as a use case for a data analysis and workflow in radio astronomy, using Dask (a Python library for parallel and distributed computing) to take advantage of the capabilities of large distributed clusters in the cloud on Kubernetes. To ensure portability, two SRCNet cloud platforms such as ESPSRC (Spain) and CHSRC (Switzerland) have been used in addition to the infrastructure of a supercomputing centre (CESGA).

11:00AM to 11:15AM C903: The VLA Sky Survey
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: OTHER CREATIVE TOPICS IN ASTRONOMICAL SOFTWARE [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--9]

pretalx [https://pretalx.com/adass2023/talk/NUP7MD/]

The VLA Sky Survey (VLASS) is a multi-epoch radio survey of the whole sky visible to the Very Large Array. It has a frequency range of 2-4 GHz, with 2.5-arcsecond resolution, and is taken in "on the fly" (OTF) mode with the antennas rastered on the sky in sets of 10x4 deg tiles in three epochs. The combination of the high angular resolution of VLASS and the OTF observing mode produce significant challenges for data processing. Although “Quick Look” images are made within ~ 1 month of observing, we are exploring new algorithms involving GPUs to speed the gridding of the observed visibilities in order to make higher accuracy images for the final processing. The large computing resources needed for VLASS has led to us develop methods for processing on remote clusters in order to complete the survey imaging in a timely fashion. Finally, the scale of the survey also means that accessing and visualizing the 34,000 individual images per epoch is itself a challenge, both for the VLASS quality assessment (QA) team and for our users. In this talk I will discuss the data challenges associated with VLASS and the solutions we are adopting, including algorithmic and machine learning approaches to QA and VO services and
applications such as HIPS, SIA2 and SODA (via CADC) for data access and visualization.

11:15AM to 11:30AM C904: Lessons learned from building LOFAR data pipelines
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: OTHER CREATIVE TOPICS IN ASTRONOMICAL SOFTWARE [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--9]

pretalx [https://pretalx.com/adass2023/talk/K3FLZP/]

In this presentation I will show how automated data processing provides great opportunities in developing robust and efficient results in astronomical research. As research code and data processing pipelines grow ever more complex, it has become more important than ever that scientists have access to frameworks that facilitate the validation of their results, and ensure that those results are fully reproducible.

I will demonstrate the current state of pipeline development for the processing of data from the International LOFAR telescope, how this pipeline leverages familiarity of common software tools and community-supported frameworks, and how research software can be embedded into this pipeline to create complex but understandable and consistent processing steps that reliably produce science-ready results.

Another point I want to address is the importance of interdisciplinary communication and coding standards. The existence of such allows a larger part of the scientific community to collaborate on mutually shared goals — of which data processing is a prominent example — and allows us as developers to create and maintain tools that anticipate the needs of future scaling. I will show how the pipeline that I will present is partially a product of such collaboration.

Finally, during this talk I would like to reflect on the broader lessons I have learned during my time developing this pipeline as someone who had no prior experience as a scientific software developer. I hope that, by sharing my experiences, I can inspire others to build and improve on them, and that in turn I can learn from the experiences of others.

11:30AM to 11:45AM C905: Towards automated structural analysis of galaxies in large imaging surveys
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: OTHER CREATIVE TOPICS IN ASTRONOMICAL SOFTWARE [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--9]

pretalx [https://pretalx.com/adass2023/talk/EZ3THQ/]slides [https://pretalx.com/media/adass2023/submissions/EZ3THQ/resources/ADASS2023_C905_Casura_pbzNv7X.pdf]

I will present our pipeline for the surface brightness fitting of galaxies using optical and near infrared imaging data from large surveys, which we applied to ~13,000 nearby galaxies with z<0.08 from the Galaxy And Mass Assembly (GAMA) survey. We fit three models to each galaxy in each of our nine wavelength bands with a fully automated Markov-chain Monte Carlo analysis using the Bayesian two-dimensional profile fitting code ProFit. For the first time, we employ ProFits multi-frame fitting functionality, working with data at the pawprint level and fitting all exposures of the same galaxy in the same band simultaneously, thus avoiding point spread function (PSF) uncertainties due to stacking. All preparatory work, including image segmentation, background subtraction, PSF estimation, and obtaining initial guesses, is carried out using the complementary image analysis package ProFound; and we develop additional routines for post-processing, including model selection, extensive quality control and a detailed investigation into systematic uncertainties. The resulting catalogue of robust structural parameters for the stellar components of galaxies (bulges and disks) can be used to study a variety of properties of galaxies and their components such as colours, luminosity functions, mass-size relations and dust attenuation. At the same time, our work contributes to the advancement of image analysis, surface brightness fitting and post-processing routines for quality assurance in the context of automated large-scale bulge-disk decomposition studies. Such advancements are vital to fully exploit the high-quality data of current and upcoming large imaging surveys.

11:45AM to 12:00PM C802: Stimela 2, kubernauts, and dask-ms: radio interferometry data reduction in the cloud
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?

level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/BFKD9M/]

Radio interferometry has been slow in adopting cloud-based technologies, despite some of their apparent advantages. I argue that it has been difficult to make radio interferometry on the cloud cost-effective for a number of reasons, chief among them: (a) awkward legacy data formats ill-suited to object store, (b) complex and heterogeneous software stacks with a heavy reliance on legacy code, and (c) awkward and complicated "thick/thin" workflows with very different resource requirements at different stages of the pipeline.

Recent software developments, however, offer a way forward. I will showcase some of these, including the Stimela 2 workflow management and containerization framework, which streamlines the orchestration of complex workflows on a Kubernetes cluster, and the dask-ms library, which maps legacy data formats onto diverse storage backends, providing support for object store. A new generation of software packages leverages these technologies, providing cloud-efficient implementations of the basic processing steps, which are able to exploit the auto-scaling capabilities inherent to cloud architectures. I will demonstrate a full data reduction workflow running on AWS. I will also argue that cloud-compatible pipelines go a long way to providing fully reproducible workflows.

12:00PM to 1:30PM LUNCH BREAK
Notes: Event Description

There are many lunch options with walking distance or via the free streetcar or if you are in a hurry on the 2nd and 3rd floors of the Student Union.

1:30PM to 1:45PM C205: Using The NEOfixer API for NEO Follow-Up and NEO Queries
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: GROUND AND SPACE MISSION OPERATIONS SOFTWARE [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--2]

pretalx [https://pretalx.com/adass2023/talk/ZYWWSM/]

NEOfixer's primary goal is to provide NEO targeting recommendations that aid in coordinating follow-up efforts. To do that effectively it creates a unique database of all the known NEOs and NEO candidates. It monitors the obvious data sources at the MPC and JPL, but incorporates other information as well, such as lists of potential radar and mission targets. It calculates orbits, ephemerides, and a variety of custom scores for each NEO based on that information. Finally, a ranked list of target recommendations is generated, customized for each subscribing telescope. Much of this information is available on the website, and more still is available via the API. You do not need to have an account to use NEOfixer and its API.

The NEOfixer API allows users to participate in its primary mission using scripts and automation, but it can be used for more than that. I will demonstrate how to use the API for everyday NEO follow-up, how to obtain details about specific NEOs, and how to generate filtered lists of NEOs for other purposes. I will give examples of how Catalina Sky Survey incorporates NEOfixer API calls into its workflow.

Catalina Sky Survey would like to thank NASA's Planetary Defense Coordination Office for its continued support, including for NEOfixer, currently via grant 80NSSC21K0893-NEOO.

1:45PM to 2:00PM C804: XMM-Newton Science Analysis System (SAS) on the cloud
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/988UK9/]

The XMM-Newton satellite is one of the most successful missions ever built for ESA. It has been operating as an open X-ray observatory since the beginning of 2000, producing high quality scientific results since then.

The XMM-Newton Science Analysis Software (SAS) is the application used
for processing the data obtained with the scientific instruments on board XMM-Newton, an indispensable tool that has been helping scientists in the publication of nearly all refereed scientific papers published up to date. SAS is a robust software that has allowed users to produce good scientific results since the beginning of the mission. This has been possible given the SAS capability to evolve from a stand-alone to a SaaS (Software as a Service) application and adapt to the needs of the scientific community.

Today, the landscape of data analysis is evolving with the advent of cloud computing, offering new dimensions to enhance scalability and efficiency. Recently, XMM-Newton project developed a pilot prototype to migrate the current Remote Interface for Science Analysis (RISA), available through the XMM-Newton Science Archive (XSA) to Amazon Web
Services (AWS). This presentation explores the synergy between SAS and cloud processing, showcasing how this collaboration transforms the landscape of X-ray astronomy.

This presentation explores the collaborative potential between the XMM-Newton SAS, cloud processing and the European Space Agency's (ESA) DataLab initiative. In the future, we will explore as well other collaborative data-driven science platforms, like SciServer , which could form a synergy that revolutionizes X-ray astronomy analysis.

Furthermore, we address recent SAS developments, focused on Docker technologies, to prepare SAS for this new technology paradigm. In particular, new SAS python interfaces that will help users to run data processing threads based on ESA DataLab platform. Further developments, such as X-ray image interactivity is needed to exploit all SAS capabilities in these cloud environments.

In conclusion, the fusion of the XMM-Newton SAS, ESA DataLab, and cloud processing represents a significant leap forward in X-ray astronomy data analysis. This symbiotic relationship not only accelerates scientific discoveries but also paves the way for innovative research methodologies, empowering astronomers to explore
the depths of the universe with unprecedented efficiency and precision.

2:00PM to 2:30PM INVITED I801: Open Source Science Initiative at NASA
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/ENXMDW/]

The Open Source Science Initiative implements the ambitious, open science vision outlined in the NASA Science Mission Directorate's "Strategy for Data Management and Computing for Groundbreaking Science 2019-2024." OSSI includes the recently updated Scientific Information Policy (SPD-41a) that includes updated requirements, compliant with the recent OSTP memo on " Ensuring Free, Immediate, and Equitable Access to Federally Funded Research", for sharing data, publications, and software produced from SMD's research activities. The Initiative further aims to increase accessibility, inclusion, and reproducibility in Earth and Space Sciences through a range of activities including training in open science, development of Open Science technologies, and grants to support Open Science. NASA's Transform to Open Science, a program to train 20,000 scientists over the next 5 years in Open Science and broaden participation from historically excluded groups, kicked off with the

Year of Open Science in 2023. NASA had provided over $6 million to sustaining and scientific software as well as to support innovate open science projects. The latest developments include a range of new infrastructure to support open science. This includes the Science Discovery Engine providing cross divisional data search and the Science Explorer, an expansion of the Astrophysics Data Service to include the other divisions in SMD. With the release of the core services strategy, SMD is laying out a path to enable groundbreaking science through cloud and high performance computing access and services.

2:30PM to 2:45PM C805: Taking TESSCut to the Cloud: Architecting for Availability,
Performance and Cost
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/LCQ3XW/]

We present the challenges encountered and solutions we reached while converting the TESSCut application to run in the cloud. TESSCut is a web application that provides image cutouts of chronologically-stacked TESS full-frame images, without requiring the user to work with the image stacks themselves. While running inside our on-premises datacenter, the application ran on a large virtual machine: 32 cores, 64 GB of memory, and nearly 400 TB of high-performance local-storage for serving the image stack data. This single machine served terabytes of cutout request data to users each month. Replicating this specific environment in the cloud would have been prohibitively expensive and beyond our budget. Instead, our cloud architecture utilizes serverless tasks inside AWS ECS Fargate, performs cutouts from remote files on an open data S3 bucket, and relies heavily on autoscaling, to achieve our performance goals while keeping costs within budget. We hope that others can benefit from our experiences and lessons learned.

2:45PM to 3:00PM C806: A Good IDIA : Scientific Computing at Scale
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/KAWFFR/]

The high data rates from current and next generation radio interferometers (MeerKAT, JVLA; SKA, ngVLA) necessarily require the data to be processed via a highly parallelized architecture in order to complete processing at reasonable timescales.

In this talk, I will discuss the Institute for Data Intensive Astronomy (IDIA) facility in Cape Town, South Africa - a pathfinder SKA science regional data centre; the tools and systems developed and adapted to perform processing at scale with the aim of producing high-fidelity images from radio interferometers. I present the IDIA MeerKAT pipeline - an automated, parallel, scaleable full Stokes calibration and imaging pipeline for MeerKAT data designed to operate on the ilifu cluster using off-the-shelf software. Our setup uses the IDIA platform running on hardware provided by the ilifu national facility, taking advantage of cluster-level parallelism, resource management and software containers.

I also discuss use of the CARTA software to efficiently visualize terabyte scale image products remotely, and briefly discuss some of the algorithm developments in progress to produce high fidelity widefield polarimetric maps with MeerKAT and other interferometers.

3:00PM to 4:00PM AFTERNOON BREAK
Location: Where
POSTER ROOM (Catalina/Tucson rooms)
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

Coffee breaks will be held in the poster room.

Afternoon tea and snacks.

3:15PM to 3:45PM FOCUS DEMO F101: Exploring the Dark Side of the Universe: The EUCLID
Scientific Archive System
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

PREV
[https://adass2023.lpl.arizona.edu/event-categories/thursday-session-3]NEXT
[https://adass2023.lpl.arizona.edu/event-categories/thursday-session-4]

THEME: SCIENCE WITH DATA ARCHIVES: CHALLENGES IN MULTI-WAVELENGTH AND
TIME DOMAIN DATA ANALYSIS
[https://adass2023.lpl.arizona.edu/themes-0#az-accordion]

pretalx [https://pretalx.com/adass2023/talk/SN8T8Z/]

Euclid is the ESA mission to explore the dark universe in the next decade. Launched on the 1st of July this year, Euclid is orbiting around the Lagrange L2 point and will map the 3D distribution of billions of galaxies and dark matter associated with them. It will hence measure the large-scale structures of the Universe across 10 billion light years, revealing the history of its expansion and the growth of structures during the last three-quarters of its history. The Euclid Consortium (EC) is in charge of processing all the Euclid data, of which only the most scientifically valuable data will be released through the Euclid Science Archive System (ESAS) during 6 years of mission lifetime: images, various types of catalogues and spectra.

Regarding data release contents, it is planned to combine Euclid observations with ground-based images obtained from several telescopes, and a huge pixel data collection, catalogues and spectra. At the end of 2023, the first science ready data products of the Early Release Observations (EROs) shall be published in ESAS. At the same time, the first data of the EC pipeline will be made available in ESAS too but only to EC members. The first public release, Q1 is planned by the end of 2024.

In the meantime, the science archive already hosts simulated images, catalogues and spectra that were used to excercise the scientific exploitation. Thus, in order to demonstrate how to explore, visualize and analyze the first public data, within the next Focus Demo, we will show the latest functionalities of the archive and the tools available for the users, such as the ESA Euclid Astroquery and ESA Datalabs Science Platform among others.

4:00PM to 4:15PM C807: NASA Archival Data in The Cloud: Service & Discovery
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/SAX7BG/]

NASA data archives started serving data from the cloud for several Astrophysics space missions. Making this data findable and discoverable means that discovery services from the archives needs to be updated to include the cloud data. It also means clients used by scientists need to know how to process and interpret that cloud information. Here, I will present some of the work that the NASA archives have been doing in serving cloud data, and tools developed that allow users to find and access the cloud data seamlessly.

Thursday, November 9

Week 45 of 2023

4:15PM to 4:30PM C808: Processing All- Sky Images At Scale On The Amazon Cloud: A HiPS Example
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/79GV9H/]slides [https://pretalx.com/media/adass2023/submissions/79GV9H/resources/C808_upO0i7y.pdf]

We report here on a project that has has developed a practical approach to processing all-sky image collections on cloud platforms, using as an exemplar application the creation of 3-color Hierarchical Progressive Survey (HiPS) maps of the 2MASS data set with the Montage Image Mosaic Engine on Amazon Web Services. We will emphasize issues
that must be considered by scientists wishing to use cloud platforms to perform such parallel processing, so providing a guide for scientists wishing to take exploit cloud platforms for similar large-scale processing. A HiPS map is based on the HEALPix sky tiling scheme. Progressive zooming of a HiPS map reveals an image sampled at ever smaller or larger spatial scales that are defined by the HEALPix standard. Briefly, the approach used by Montage involves creating a base mosaic at the lowest required HEALPix level, usually chosen to match as closely as possible the spatial sampling of the input images, then cutting out the HiPS cells in PNG format from this mosaic. The process is repeated at successive HEALPix levels to create a nested collection of FITS files, from which are created PNG files that are shown in HiPS viewers. Stretching FITS files to produce PNGs is based on an image histogram. For composite regions (up and including the whole sky) the histograms for each tile can be combined to create a composite histogram for the region. Using this single histogram for each of the individual FITS files means all the PNGs are on the same brightness scale and displaying them side by side in a HiPS viewer produces a continuous uniform map across the entire sky.

All the processing just described can one readily performed in parallel on AWS instances. To create the HiPS maps on AWS, jobs were set up with a Docker container that contains the requisite data software components, including modules added to streamline processing
on cloud platforms, including adjusting for inter-image background variations and developing a global model for visualization stretches. Jobs are set up and run with the Amazon Web Services (AWS) Batch processing mode, which spins up server instances as needed, pulling from a pool of pre-defined job script. When a job is done it either the compute instance another job from the pool or shuts the instance down. This approach minimizes having idle instances which would still incur charges even when not processing. A set of script generators developed for this project create, by design, simple scripts that are handed to the instances to run jobs inside the containers. Processing the whole sky at three wavelengths requires about ten thousand such jobs. We will discuss processing times and costs.

4:30PM to 4:45PM C809: Preparing a scientific data processing facility for Rubin Observatory's LSST: the case of France's CC-IN2P3
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/BMDAP9/]

Located in Lyon, France, the IN2P3 / CNRS Computing Centre (CC-IN2P3)
has been preparing its contribution to produce the Legacy Survey of Space and Time in its role as the Rubin Observatory's France Data Facility.

An integral copy of the raw images will be imported and stored there for the duration of the 10 year-long survey and annual campaigns of reprocessing of 40% of the raw images recorded since the beginning of the survey will be performed on its premises. The data products of those campaigns will be sent back to the Observatory's archive center in the USA.

As a scientific data processing facility shared by several dozen international projects in high energy physics, nuclear physics and astroparticle physics, in recent years we have observed a significant increase in both the computing and storage capacity demand as well as in the complexity of the services required for supporting astroparticle physics projects. We expect their needs to continue increasing for the foreseeable future: major international projects like Rubin, Euclid, KM3NeT, Virgo/LIGO represent a sizeable fraction of the resources CC-IN2P3 provides to the science projects it supports, even if not yet at the level of the high energy physics projects.

In this contribution we will address how we have been preparing to perform bulk image processing for the needs of the Rubin Observatory annual data release processing campaigns for the duration of the survey. We will present the architecture of the system we deployed with focus on the storage, compute and data transfer components and how we have been testing the system at significant scale. We will highlight and motivate some of the solutions we adopted which have proven effective for our successful contribution to other large science projects like CERN's Large Hadron Collider. We will also cover our initial experience with components deployed for the specific needs of scientific exploitation of Rubin data such as the astronomical catalog database and the Rubin science platform.

4:45PM to 5:00PM C810: Rubin Science Platform: on cloud, on-prem, all of the above
Location: Where
BALLROOM
[https://union.arizona.edu/infodesk/maps/sumc_maps.php?level=level3]
Notes: Event Description

THEME: CLOUD INFRASTRUCTURES FOR ASTRONOMICAL DATA ANALYSIS [https://adass2023.lpl.arizona.edu/themes-0#az-accordion--8]

pretalx [https://pretalx.com/adass2023/talk/3CJEHG/]

The Rubin Science Platform is already in production before system first light and is approaching 1,000 registered early access users working with precursor data products on our outward-facing deployment
on Google Cloud . In this talk I describe the architecture that allows a small team to manage over a dozen separate deployments of the platform on cloud, on-premises (including the telescope summit) and in our hybrid model for operations, a mixture of both. I will also briefly address common mistakes made when evaluating cloud economics

5:00PM to 7:30PM IVOA Executive Committee Meeting
Notes: Event Description

https://wiki.ivoa.net/twiki/bin/view/IVOA/InterOpNov20