



Contribution ID: 106

Type: **Presentation**

Bridging the Gap: A Research-Ready AI/ML Infrastructure on the Nectar Research Cloud

Wednesday 15 October 2025 14:11 (11 minutes)

Introduction: Researchers and scientists are increasingly using programming languages for data processing, visualisation, and analysis. Advancement in machine learning (ML) and artificial intelligence (AI) helps accelerate the process of analysing complex research data and conducting experiments, leading to the discovery of hidden patterns in the data. However, installing and configuring the environment to begin researching using AI/ML or transitioning from local scripting to utilising cloud computing to accelerate research is often hindered by the need for extensive knowledge or expertise in the use of command line and back-end operating systems. In collaboration with the Australian Research Data Commons (ARDC), the Advanced Analytics and AI (3AI) Platform at Intersect has developed a solution to address this barrier by developing an application that provides a research-ready AI/ML infrastructure on the Nectar Research Cloud. The goal is to reduce the technical barrier of utilising cloud infrastructure for AI/ML-enabled research, empowering researchers to focus on science.

Methods: The application is featured with a modular design, offering an AI-ready environment that covers the foundational aspects of data-intensive research, including a pre-installed and configured Python environment, Data Operations (DataOps), Machine Learning Operations (MLOps), and real-time system monitoring. The pre-installed and configured Python environment is the core of the application with all the necessary libraries for users to begin their AI journey or to go further with optional additional layers focusing on specific AI topics including; Computer Vision, Large Language Models and Generative AI. This is enabled by the pre-installed and configured state-of-the-art open-source tools (i.e. Apache Airflow for DataOps, MLflow for MLOps, and Prometheus and Grafana for system monitoring). Additionally, interactive software like RStudio and JupyterLab are also included to offer user-friendly development environments. Researchers can customise their environment with specialised Python installations optimised for tasks like computer vision or generative AI. Guidance on hardware specifications (vCPUs, RAM, and storage) is provided. The comprehensive documentation, covering installation, setup, and usage, further supports researchers in utilising the platform effectively. For more advanced users, the infrastructure could be customised from ARDC's open-source image repository to create their unique platform to meet their needs.

Results: The initiative provides pre-configured, optimised environments that enable researchers to focus on their scientific inquiry rather than complex system administration. The infrastructure simplifies the entry to cloud computing for AI/ML, empowering researchers to quickly and efficiently begin their AI/ML journey. The platform offers essential tools for data analysis, workflow orchestration, and model development within a user-friendly framework. The guidance on hardware specifications also ensures optimal performance for research workloads.

Conclusion: This collaborative effort between Intersect and ARDC aims to empower researchers, regardless of their system administration expertise, to leverage cutting-edge AI/ML tools. By providing a pre-configured, documented, and user-friendly environment, the project removes a significant hurdle in the research workflow. The initiative aims to significantly accelerate AI/ML research on the Nectar Research Cloud, enabling researchers to focus on innovation and discovery and enhancing the Australian research landscape by providing accessible and powerful computational resources.

Primary author: CHARLTON, Glen (Intersect Australia)

Co-authors: Mr LE, Long (Intersect Australia); Dr FAN, Jiaxin (Intersect Australia); Dr PAPAIOANNOU,

Anastasios (University of Technology Sydney); Mr BOTTING, Andy (Australian Research Data Commons); Ms LOVELACE-TOZER, Meirian (Australian Research Data Commons); Mr CHIU, Ben (Australian Research Data Commons)

Presenter: CHARLTON, Glen (Intersect Australia)

Session Classification: Presentations Session 7: Open research through Interconnected, Interoperable, and Interdisciplinary Data

Track Classification: SciDataCon2025 Specific Themes: Infrastructures to Support Data-Intensive Research - Local to Global