



Australian Research Data Commons

AI-Ready Application

Glen Charlton

Lead Data Scientist

Andy Botting

Technical Lead, Nectar Cloud Services



:: Content

01

Setting the scene

02

About the application

03

Case Study

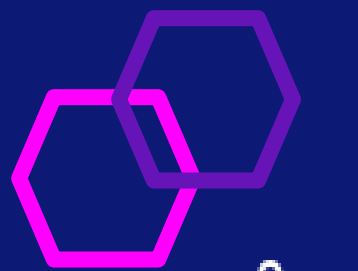
04

Getting Started

05

Technical Details

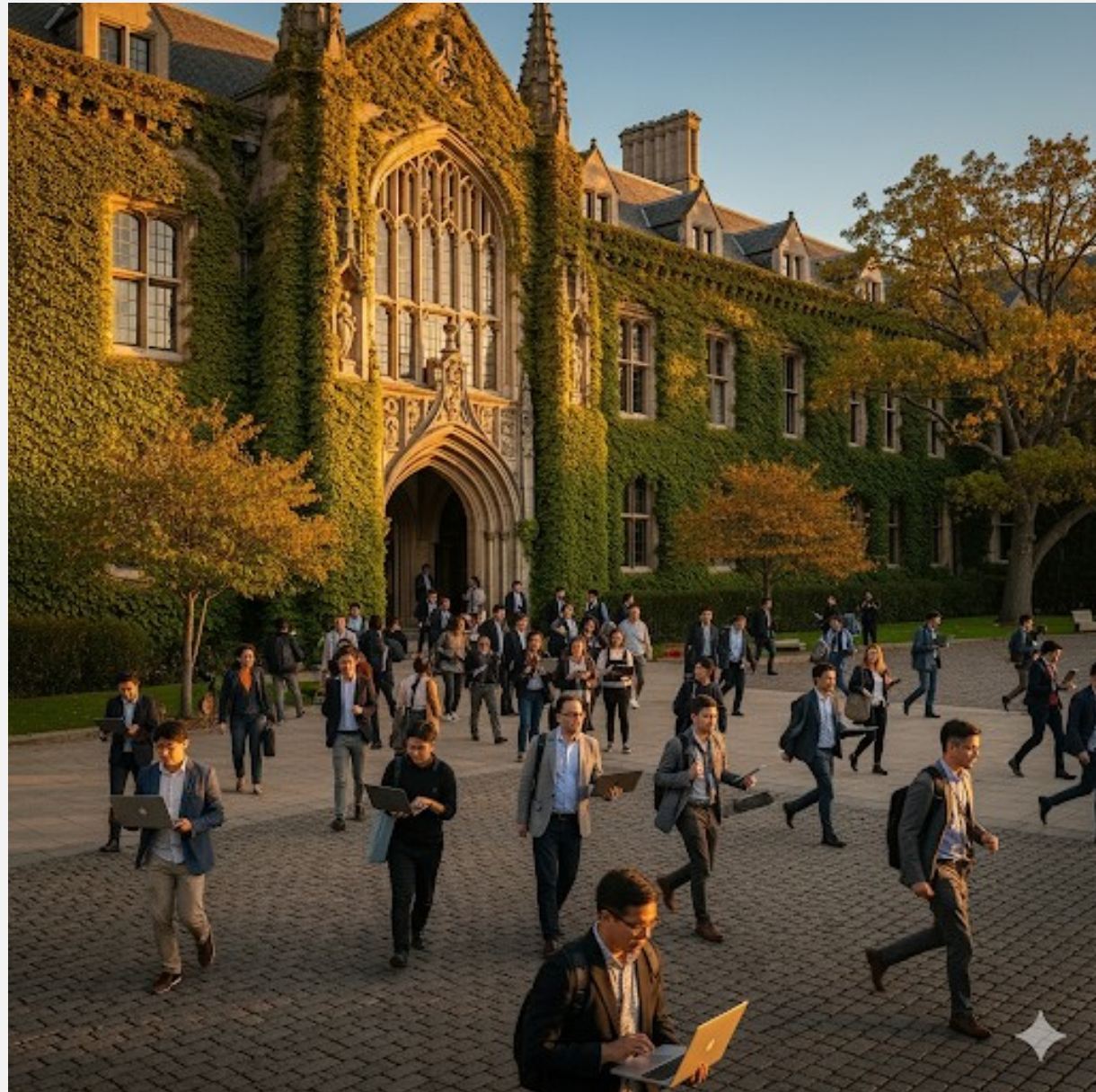




Setting the scene

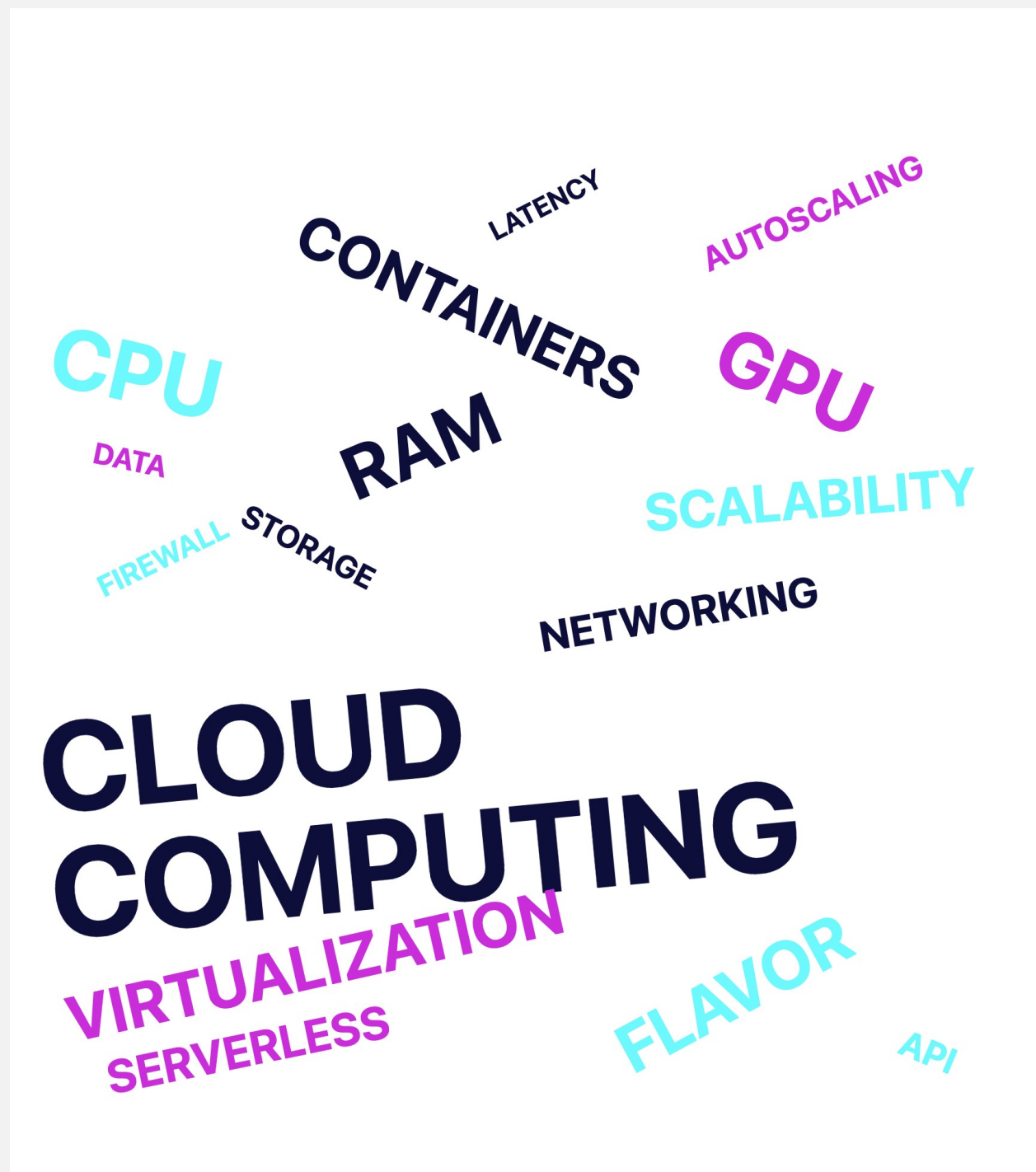


:: Why use the cloud?



- **More CPUs and RAM:** Cloud computing can offer flavours with more CPUs or RAM than most local computers can offer.
- **Access to high performing GPUs:** Cloud computing like Nectar offers access to high performance GPUs
- **Performance when you need it:** You probably don't need 32 cores and 128GB of RAM when you're not analysing your data.
- **You're in control:** We have all experienced that inconvenient forced update that stops our code.

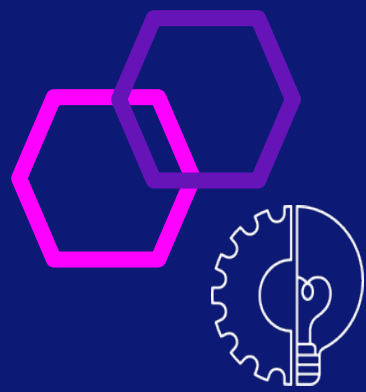
:: **Why** not?



- **All the lingo:** Computing has a lot of lingo I have already used in the presentation that most researchers don't need to know.
- **Cost and Complex Pricing:** Cloud computing (especially public cloud) can be confusing and have unknown costs making it difficult to budget for.
- **Data Privacy and Security Concerns:** It may not be appropriate for you to use cloud computing with your dataset or you may not know how to secure it correctly.
- **The knowledge gap:** Researchers are already experts in their specific field or domain, we can't also expect them to have both expertise in both cloud computing and machine learning.

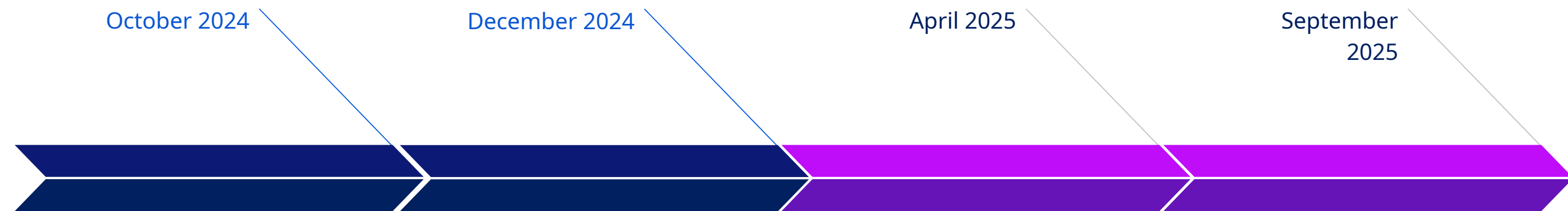
:: So **where** are we?





About the application

:: Ideation to end result



Ideation

3AI and Nectar team identified the gap for researchers and begun to develop project plan.

Conceptualised

The idea and concept was developed and the development of the application was started.

Development Complete

The development of the AI Ready Images was complete and testing initiated with a soft launch on the Nectar console.

Launched



In September, the AI Ready Application was officially launched and advertised by ARDC and Intersect.

:: How do we solve this?






AI Ready

Welcome to the ARDC Nectar AI Ready Application, offering a range of Python environments with AI and ML tools installed and ready to go.

This application was developed in partnership with [Intersect's Advanced Analytics and AI Platform](#).

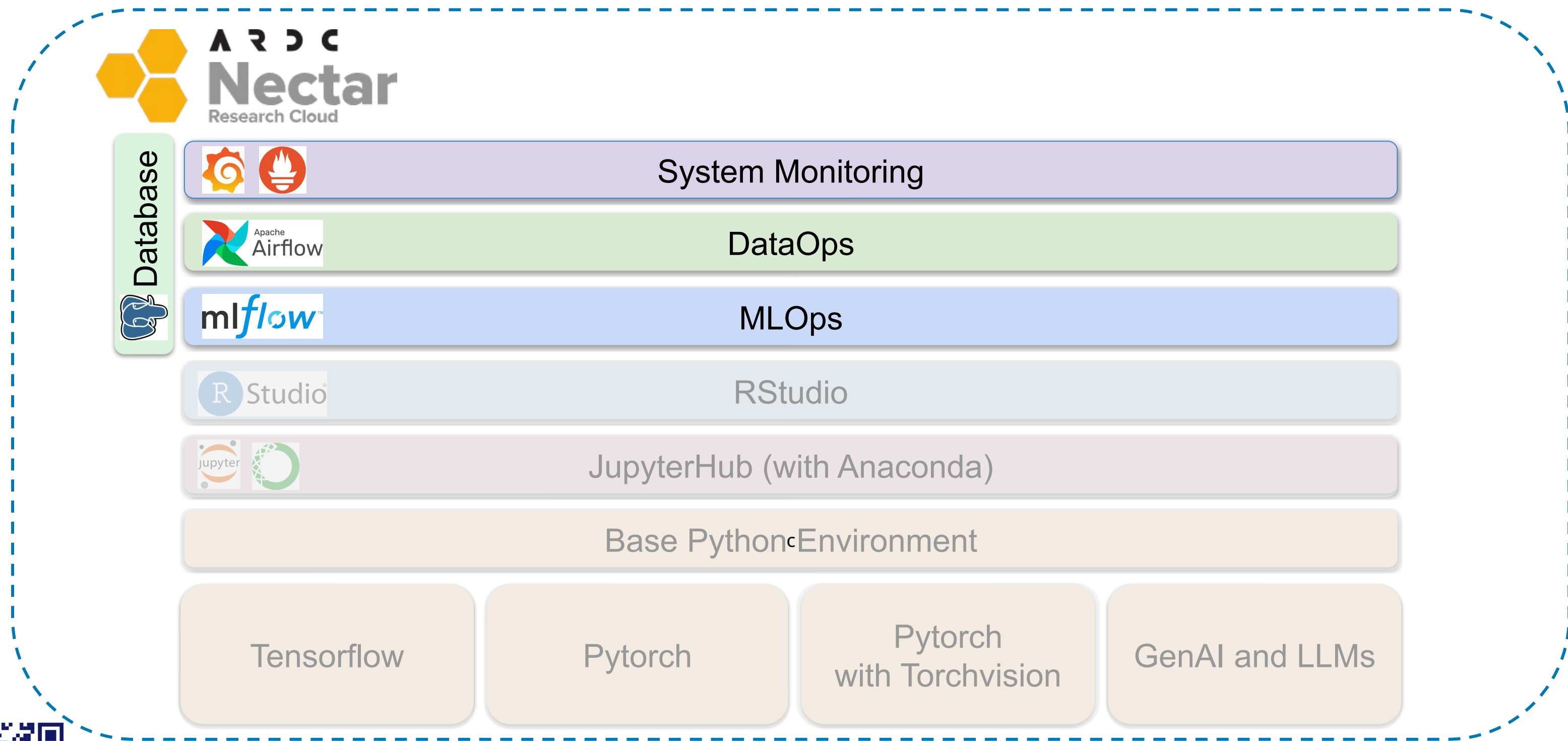
 

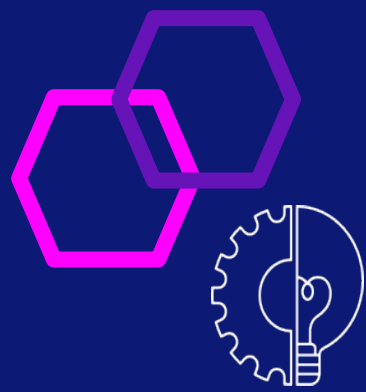
SELECT YOUR APPLICATION

-  R-Studio >
-  JupyterHub >
-  MLflow >
-  Apache Airflow >
-  Grafana >

Need help? Check the [AI Ready Application support article](#).

:: How do we solve this?





Case Study:
Megadetector Pipeline



:: Case Study: Using this tool to create a DataOps/MLOps

Pipeline

Import Data

- Researchers upload dataset into volume to SFTP.
- Researchers trigger Airflow DAG to start.

Find data

- Tasks to find data within input directory
- Tasks create matching output directories

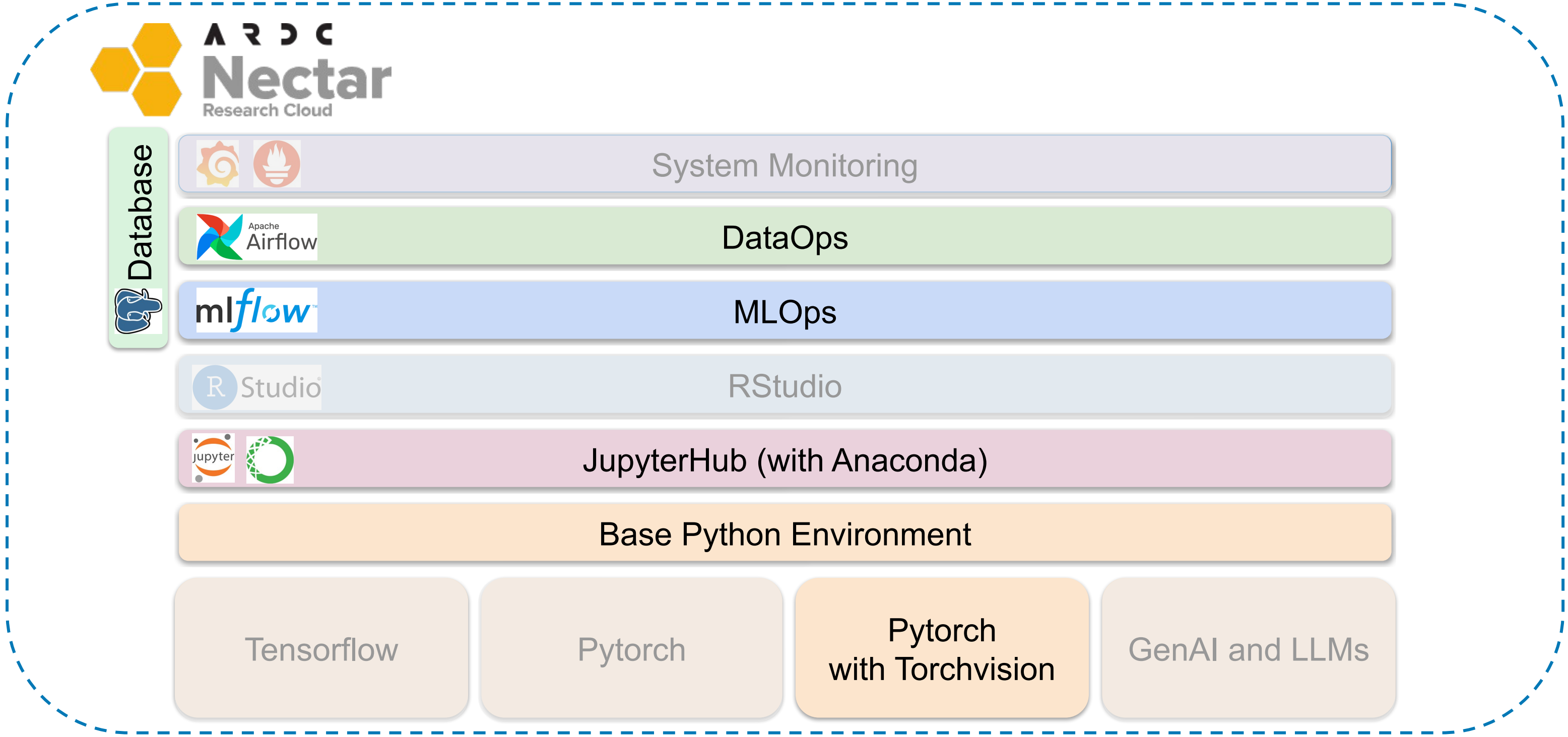
Annotate Data

- Task runs object detection algorithm (MegaDetector) to classify images.
- Image metadata is updated.
- Data is sorted into output directories by classification.

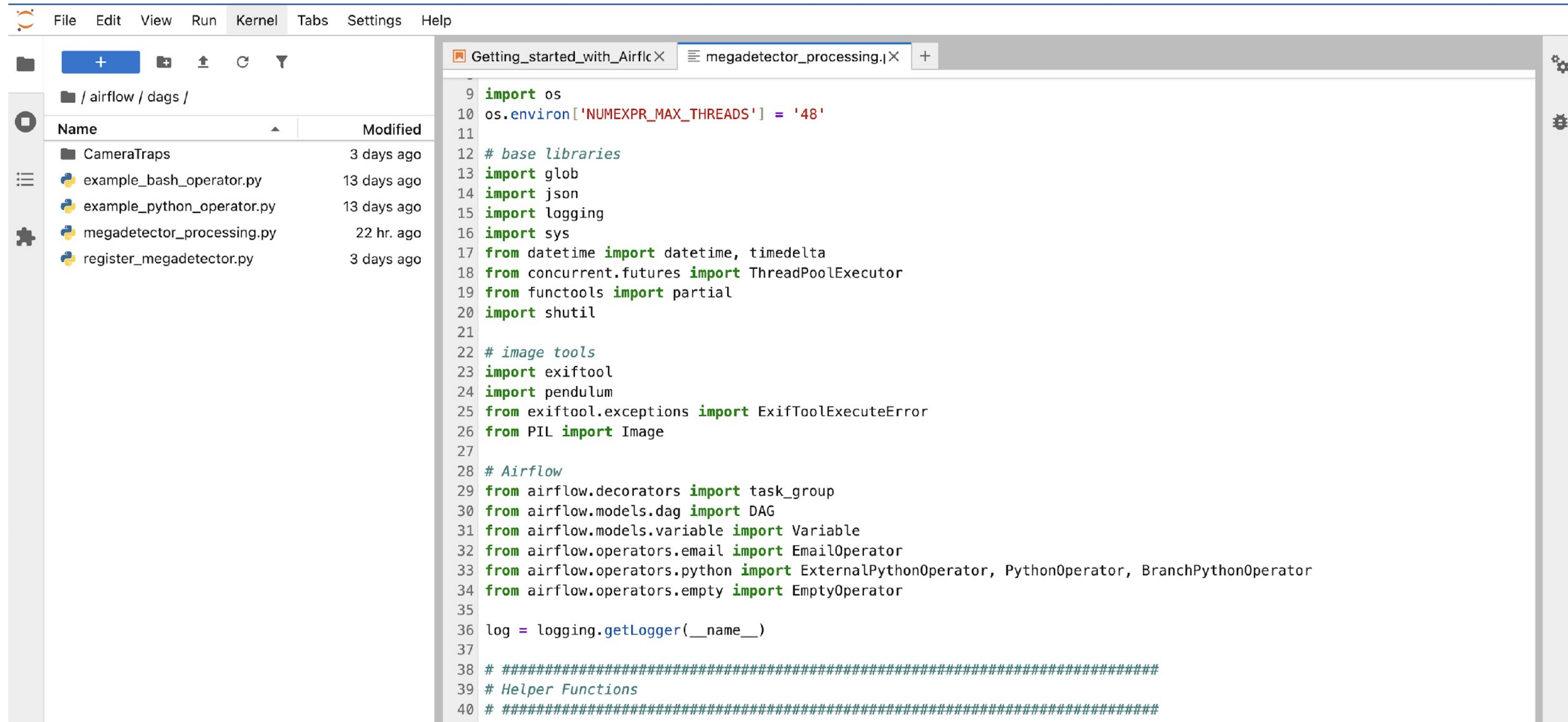
Export Data

- Researchers receive email when dataset is processed.
- Researchers download dataset to local system using SFTP.

:: Case Study: Using this tool to create a DataOps/MLOps Pipeline



:: Case Study: JupyterHub for Python Development



The screenshot displays the JupyterLab interface. On the left, a file browser shows the directory structure: /airflow/dags/. The files listed are: CameraTraps (3 days ago), example_bash_operator.py (13 days ago), example_python_operator.py (13 days ago), megadetector_processing.py (22 hr. ago), and register_megadetector.py (3 days ago). The main area shows a code editor with the following Python code:

```
9 import os
10 os.environ['NUMEXPR_MAX_THREADS'] = '48'
11
12 # base libraries
13 import glob
14 import json
15 import logging
16 import sys
17 from datetime import datetime, timedelta
18 from concurrent.futures import ThreadPoolExecutor
19 from functools import partial
20 import shutil
21
22 # image tools
23 import exiftool
24 import pendulum
25 from exiftool.exceptions import ExifToolExecuteError
26 from PIL import Image
27
28 # Airflow
29 from airflow.decorators import task_group
30 from airflow.models.dag import DAG
31 from airflow.models.variable import Variable
32 from airflow.operators.email import EmailOperator
33 from airflow.operators.python import ExternalPythonOperator, PythonOperator, BranchPythonOperator
34 from airflow.operators.empty import EmptyOperator
35
36 log = logging.getLogger(__name__)
37
38 # #####
39 # Helper Functions
40 # #####
```

:: Case Study: MLOps for model management

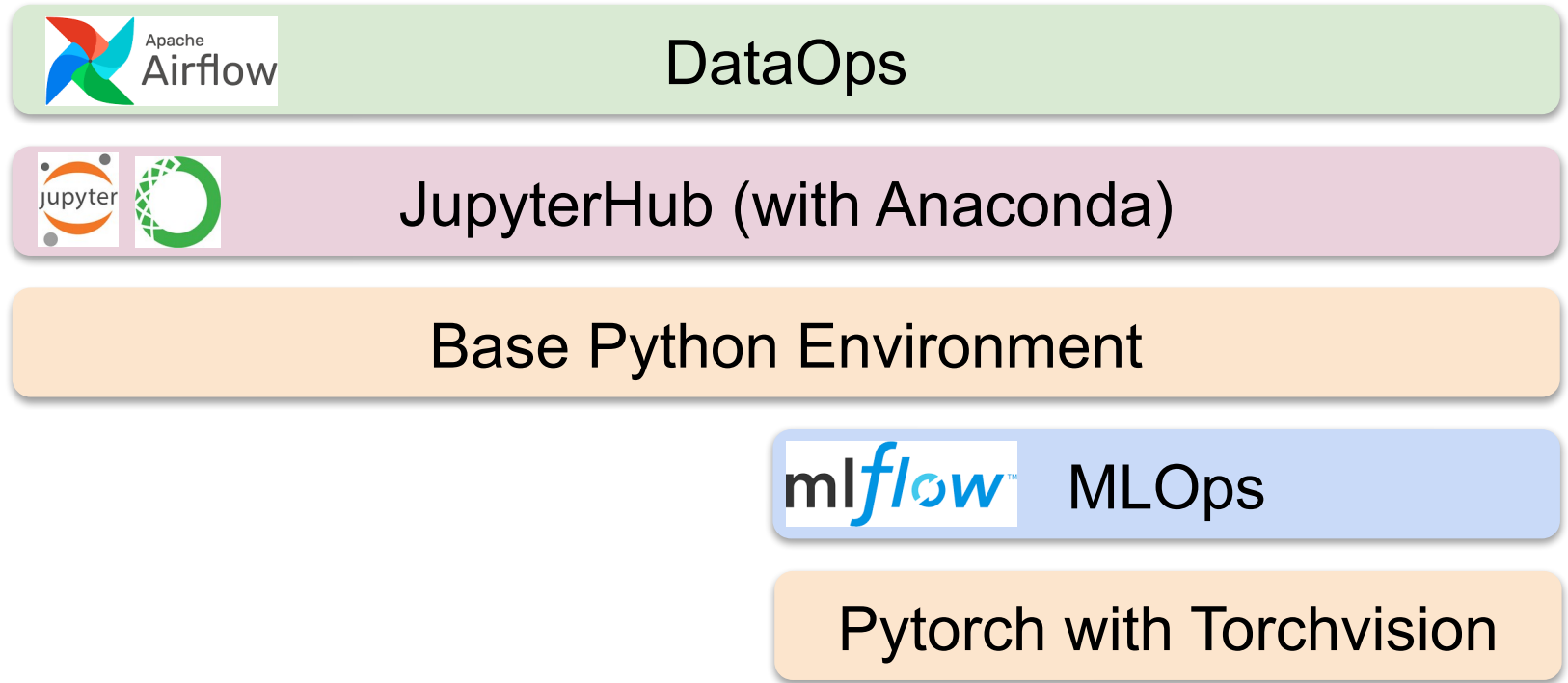
The screenshot displays the mlflow 3.3.2 web interface. The left sidebar contains navigation options: '+ New', 'Experiments', 'Models', and 'Prompts'. The main content area is titled 'MegaDetector Registration' and includes a 'Share' button. Below the title are tabs for 'Runs', 'Models', 'Experimental', 'Evaluation', and 'Traces', with 'Models' selected. A table lists registered models with columns for Model name, Status, Created, Logged from, Source run, and Registered models.

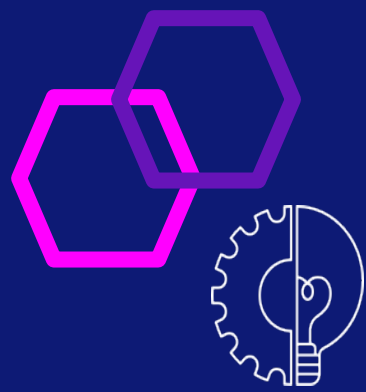
Model name	Status	Created	Logged from	Source run	Registered models
megadetector_model	Ready	38 minutes ago	airflow	handsome-tern-30	megadetector-v5 v7 +2
megadetector_model	Ready	38 minutes ago	airflow	salty-eel-592	megadetector-v5 v6 +1
megadetector_model	Ready	38 minutes ago	airflow	placid-vole-107	megadetector-v5 v5 +1
megadetector_model	Ready	38 minutes ago	airflow	unruly-newt-799	megadetector-v5 v4 +1
megadetector_model	Ready	38 minutes ago	airflow	skillful-hare-688	megadetector-v5 v2 +1
megadetector_model	Ready	39 minutes ago	airflow	bright-fawn-345	megadetector-v5 v3 +1
megadetector_model	Ready	3 days ago	airflow	upset-crow-519	megadetector-v5 v1 +1

:: Case Study: DataOps Pipeline on Airflow

The screenshot displays the Apache Airflow web interface for a DAG named 'megadetector_processing'. The interface includes a top navigation bar with links for DAGs, Cluster Activity, Datasets, Security, Browse, Admin, and Docs. The current time is 04:53 UTC, and there is a 'Log In' button. The DAG title is 'DAG: megadetector_processing' with a subtitle 'A dynamic DAG to process images with MegaDetectorV5'. The interface shows the DAG's execution history for the date 25/09/2025 at 03:57:53 am. A legend identifies various run states: deferred, failed, queued, removed, restarting, running, scheduled, shutdown, skipped, success, up_for_reschedule, up_for_retry, upstream_failed, and no_status. The current view is for a specific run of the DAG on 2025-09-24 at 03:52:31 UTC, specifically for the task 'tag_and_move_images' (ID 55). The interface provides options to view details, graph, gantt, code, audit log, logs, XCom, and task duration. A task duration bar chart on the left shows the execution time for various tasks, with 'start' at 1d20:45:48 and 'end' at 00:00:00. The main area shows a DAG graph with tasks: start, get_config, find_image_subdirectories, create_output_directories, process_directory [56] (containing run_megadetector_detection [1] and group_events [56]), tag_and_move_images [55], remove_processed_files [1], check_for_remaining_files, remove_processed_directories, send_files_remaining_email, send_completion_email, and end. A 'Layout: Left -> Right' dropdown is visible. A 'React Flow' logo is at the bottom right of the graph area.

:: **Case Study:** Using this tool to create a DataOps/MLOps Pipeline





Getting started



The ARDC is Australia's leading research data infrastructure facility.

- To provide Australian researchers with competitive advantage through data.
- To accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.



ARDC Nectar Research Cloud

- Australia's federated national research cloud providing the research community with on-demand computing infrastructure and software.



SELECT PROJECT

nectar-training ▾

[Learn more about allocations >](#)

RESOURCE BUNDLE ⓘ

Custom

PROJECT USAGE ⓘ

2941.29 / 10000 SU

[Usage Summary](#) | [Limit Summary](#)

Home

Project >

Identity >

Applications >

Allocations ▾

New Request

My Requests

Allocations / None


New Allocation Request: About the Project

All My Requests

This form allows you to request a project specific allocation on the Nectar Research Cloud. Allocations take up to 3 weeks to process. If you have any questions please [create a support ticket](#).

NOTE If you would like to request an extension or amendment of an existing allocation, please use the *Amend/Extend allocation* action on the existing allocation request on the [My Requests](#) page.

✓ ORCID Connected

 [0000-0002-5765-6292](#)

Is this ORCID incorrect? Sign out at [orcid.org](#) and then go to the [My Details](#) page to disconnect.

INFO: Required fields are marked with an * sign.

Before you begin...

Before requesting an allocation, please ensure that you:

- meet the [Eligibility Requirements](#) for a national or local allocation
- understand and follow Nectar's [Sensitive Data Guidelines](#)
- have an ORCID and it is connected to your Nectar account
- have the Project Information ready
- know the Chief Investigator name and contact details
- can provide the Research Grant information
- know the FoR code to describe the research

STEP 1: About the project

STEP 2: Cloud resources

1. Project Details

Project allocation title * ?

Contact e-mail ?

Project Identifier * ?

Generated project DNS zone name ?

Project identifier cannot be changed after allocation is approved.

Estimated project duration * ?

Convert trial project? ?

Step 1 out of 9

Next >

2. Usage Information

3. Fields of Research (FOR)

4. Supported Institutions

5. Investigator

6. Project Support

7. Research Grant Information

SELECT PROJECT

nectar-training

Learn more about allocations

RESOURCE BUNDLE

Custom

PROJECT USAGE

2941.29 / 10000 SU

Usage Summary | Limit Summary

Home

Project

Identity

Applications

Browse

Browse Local

Manage

Applications

Allocations

Applications / Browse / Browse

Recent Activity

No recent activity to report at this time.

App Category: All

Environment: quick-env-10

Filter

Filter

AI Ready
Python environment with essential libraries for data analysis, machine learning, and AI development. Ideal for users starting the...
[Details »](#)
Add to Env Quick Deploy

Apache HTTP Server
The Apache HTTP Server Project is an effort to develop and maintain an open-source HTTP server for modern operating systems incl...
[Details »](#)
Add to Env Quick Deploy

Docker Container
The application provides ability to create Docker container to deploy one of the 13,000+ apps available on Docker Hub at https://...
[Details »](#)
Add to Env Quick Deploy

Docker HTTPd
The Apache HTTP Server, colloquially called Apache, is a Web server application notable for playing a key role in the initial gro...
[Details »](#)
Add to Env Quick Deploy

Docker Standalone Host
Docker is an open platform for developers and sysadmins to build, ship, and run distributed applications. Consisting of Docker En...
[Details »](#)
Add to Env Quick Deploy

Duply Backups
Simple cloud backups using a Swift container and duply, a frontend for the duplicity incremental archiving tool. Package author e...
[Details »](#)
Add to Env Quick Deploy

Next Page



The Australian Research Data Commons is enabled by NCRIS.

ARDC CONNECT NEWSLETTER

First Name Last Name
Email A Researcher

Subscribe

QUICK LINKS

- Dashboard Home
- Nectar Terms of Service
- Nectar Support
- About Nectar
- ARDC Services for Research
- Contact ARDC

We acknowledge and celebrate the First Australians on whose traditional lands we live and work, and we pay our respects to Elders past, present and emerging.

Configure Application: AI Ready

Environment *

Base Python Environment

Base Python Environment

PyTorch

PyTorch and TorchVision

TensorFlow

Generative AI and Large Language Models (LLMs)

environment:

Select the Python environment that best suits your activity. For details, see <https://support.ehelp.edu.au/support/solutions/articles/6000280865>

Host name ?

my-ai-app

Host name:

Specify a string, that will be used in instance hostname and possible DNS entry. Just letters, numbers and hyphens are allowed.

DNS zone

nectar-training.cloud.edu.au

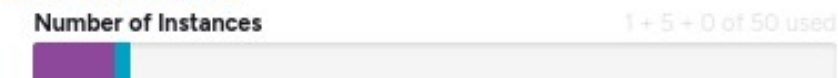
DNS zone:

Select DNS zone if you'd like a DNS entry created for you. A HTTPS certificate will be provisioned for you automatically if you choose a zone.

Instance flavor *

m3.xlarge

Instance flavor:



Key pair

andy



Availability zone *

intersect



Disk size (GB) *

100

Select your desired flavor. Consider that application performance will be affected by your choice.

Key pair:

Select the keypair to control access to instances. You can login to the instance using SSH with this keypair after the deployment has completed.

Availability zone:

Select availability zone where the application would be created. If you plan to use a volume for your data, this will need match the Availability Zone...

[Show more](#)

Disk size (GB):

Specify disk size in GB which is going to be created from image. The minimum size we allow is 60GB, and can be increased depending on your workload.

Username *

myusername

Username:

Provide a username for the new user account to be created. The username should be lowercase letters and numbers only, and must start with a letter.

Password * ?

.....

Password:

Password must be at least 7 characters and contain both lower and uppercase letters, numbers and special characters, except single quote (').

Confirm password * ?

.....

SELECT PROJECT

ncs-demo ▾

[Learn more about allocations](#) >

RESOURCE BUNDLE ⓘ

Powerful

PROJECT USAGE ⓘ

1349.82 / 8000 SU

[Usage Summary](#) | [Limit Summary](#)

Home

Project >

Identity >

Applications ▾

Browse >

Manage >

Applications ▾

Environments

Allocations >

Applications / Applications / [Environments](#) / quick-env-1


Components **Topology** Deployment History Latest Deployment Log

Application Components

App category All ▾

Find

<




AI Ready

shared



AI Ready




Apache HTTP Server

shared




Docker Container

shared



Docker HTTPd

shared



Docker Standalone Host

shared

>

📍 Drop Components here

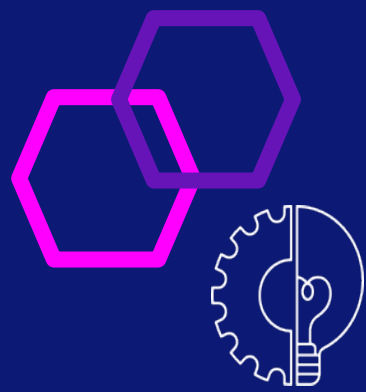
[+ Add Component](#)

Displaying 1 item

Name	Type	Status	Last operation	Time updated	Actions
AIReady	AI Ready	Ready	Application is available at https://andydemo1.ncs-demo.cloud.edu.au	Sept. 16, 2025, 4:41 p.m.	Delete Component ▾

Displaying 1 item

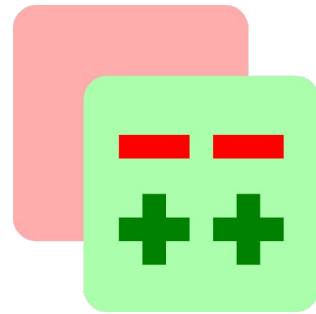
[? Help](#)



Technical Details

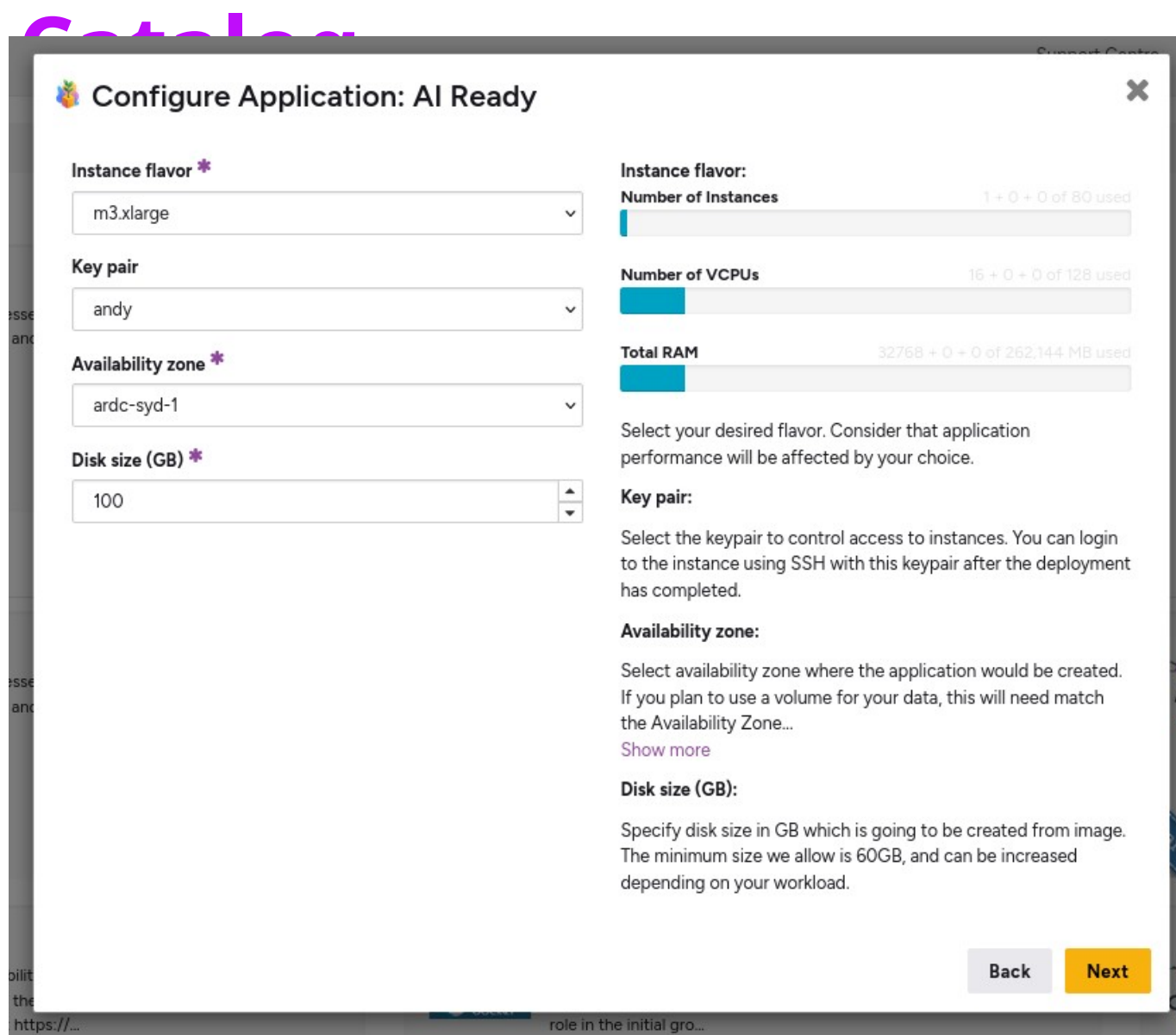


:: How images are built



- **Changes** proposed to Nectar Code Review service
- Code merged into the **Nectar Images** git repository
- **Jenkins** starts the image pipeline workflow
- **Packer** is used for the image creation process
- **Ansible** is called by Packer to apply the configuration changes including installing packages, deploying configuration files and testing scripts
- Image is uploaded to the Research Cloud (**OpenStack**)
- **Test** scripts are executed to validate the image
- Image is made **public** for general access

:: Nectar Application



- Cloud Service for deploying applications
- Provides simple interface for user to choose the properties of their environment
- Generates orchestration stack that provisions cloud resources (Virtual machine, Volume storage)
- Boot-time provisioning scripts to handle user provisioning and HTTPS Let's Encrypt certificate



Australian Research Data Commons

A special thank you to those who contributed to this work:

Anastasios Papaioannou, Andy Botting, Ben Chiu, Bernard Meade,

Dylan McCulloch, Long Le, Jiaxin Fan, Meirian Lovelace-Tozer,

Paul Coddington, Sonia Ramza

