

Australian Ocean Data Network utilises Amazon Web Service Batch processing for gridded data

¹*Cameron Moloney*

Cameron Moloney¹, Sebastien Mancini², Philip Bohm³, Leigh Gordon⁴, Craig Jones⁵, Alex McKeown⁶,
George Sattler⁷, Angus Scheibner⁸, Kate Reid⁹

¹UTAS/AODN/IMOS, Hobart, Australia, Cameron.moloney@utas.edu.au

²UTAS/AODN/IMOS, Hobart, Australia, sebastien.mancini@utas.edu.au

³UTAS/AODN/IMOS, Hobart, Australia, philip.bohm@utas.edu.au

⁴UTAS/AODN/IMOS, Hobart, Australia, leigh.gordon@utas.edu.au

⁵UTAS/AODN/IMOS, Hobart, Australia, craig.jones@utas.edu.au

⁶UTAS/AODN/IMOS, Hobart, Australia, a.mckeown@utas.edu.au

⁷UTAS/AODN/IMOS, Hobart, Australia, George.sattler@utas.edu.au

⁸UTAS/AODN/IMOS, Hobart, Australia, Angus.Scheibner@utas.edu.au

⁹UTAS/AODN/IMOS, Hobart, Australia, Kate.Reid@utas.edu.au

EXTENDED ABSTRACT

The Australian Ocean Data Network (AODN) is an interoperable online network of marine and climate data resources. It is a collaboration between six major Australian Commonwealth agencies, the Integrated Marine Observing System (IMOS), and a growing list of universities, state government offices and others organisations in Australia, New-Zealand and the Pacific.

AODN data collections cover a large geographic area (from coast to open ocean, from equator to Antarctica), a wide range of observed parameters (physical, chemical, biological) and are obtained using a variety of platforms and technologies (e.g. ships, autonomous floats, gliders, moorings, satellites, animal tags, coastal radar. The end users include researchers, students, managers, policy makers, consultants, sailors and fishers. All data products contributed to the AODN are made freely and openly available to the public via the AODN Portal (<https://portal.aodn.org.au/>).

The AODN Portal provides access to the following different types of gridded dataset collections:

- Sea Surface Temperature (SST) products.
- Ocean Colour products.
- Satellite altimetry products.
- Coastal radar products.
- Climatology.
- Bathymetry.

Many of the gridded data collections are very large and therefore tools have been developed over the years to enable data products to be visualised through the AODN Portal as well as to be partitioned temporally and spatially to then be aggregated into a single file. Tools like netCDF Operator (NCO) or the netCDF Java library are widely used by the scientific community in order to subset and aggregate gridded data. The main issue for the AODN has been the management of

jobs submitted by many different users through the AODN Portal. A single queue, no indication of the position of a particular job in the queue, loss of jobs during a restart of services and restriction on output files were some of the problems encountered over the past few years.

In 2016, the AODN relocated its services to the commercial cloud service Amazon Web Services (AWS). In addition to the generic advantages of cloud computing, AWS offers a growing array of complimentary services. These services include data durability, the automated deployment of auto scaling and load balancing architecture, which increase system reliability. One of the most recent service to be implemented in the AODN infrastructure is AWS Batch, a fully managed batch processing service.

AWS Batch dynamically provisions the optimal quantity and type of compute resources (e.g. CPU or memory optimised instances) based on the volume and specific resource requirements of the batch jobs submitted. AWS Batch offers many benefits like reduced operational complexities, time saving and reduced costs, the possibility to configure multiple queues and increased download size options.

An Open Geospatial Consortium (OGC) Web Processing Service (WPS) allows standardised requests and responses to the asynchronous processing service used to subset and aggregate gridded data into a single netCDF file for a variety of dataset collections.

AODN have also implemented customised dashboards in order to efficiently track jobs in real-time and to ensure rapid responses to queries that come to the AODN helpdesk. The content of data reports used by the AODN management team have been extended to include information, such as the most popular dataset collection, response time to provide download output, a confidential list of users and their affiliation.