NeSI New Zealand eScience Infrastructure

RSE RSE-AUNZ

Science New Zealand

New Zealand
Research Software Engineering Conference 2020
Virtual conference | 9-11 September 2020

Parallel Computing with Dask

Wolfgang Hayek, Maxime Rio, Chris Scott
NeSI
wolfgang.hayek@nesi.org.nz, maxime.rio@nesi.org.nz, chris.scott@nesi.org.nz

ABSTRACT / INTRODUCTION (Up to 200 words)

Parallel computing has become a necessity for a wide range of modern scientific computing problems, including data-oriented computing at large scale to achieve reasonable processing times. Implementing parallel computing can be challenging and time-consuming - APIs such as the Message Passing Interface (MPI) are powerful but can be hard to learn and implement.

Dask is a popular toolkit for the Python programming language that addresses this issue. While requiring very little programming effort, it offers a variety of parallelisation paradigms, including work sharing via parallel function evaluation, task graphs, and direct integration with packages such as NumPy, Pandas, and Scikit-Learn. Dask can be used interactively and as a batch processing tool. The Dask-MPI package adds MPI as a parallelisation backend, enabling scalability and high throughput on high-performance computing (HPC) systems.

In this presentation, I will introduce Dask, discuss some of its parallelisation mechanisms, and demonstrate how to use the MPI backend for batch processing.

[Note: This presentation should precede Maxime Rio's demo of using Dask with SciKit-learn in Jupyter notebooks as it will cover off the basics of Dask.]

ABOUT THE AUTHOR(S)

Wolfgang Hayek is a research software engineer at NeSI and NIWA, and group manager of NIWA's scientific programming group, with many years of experience in scientific computing and HPC.

Maxime is a data scientist at NeSI and NIWA. He enjoys helping researchers to analyse their data, from visualisation to probabilistic modelling.

Chris Scott is a Research Software Engineer at NeSI with a background in scientific computing and HPC.