DELIVERING ASTRONOMICAL IMPACT FOR SCIENCE, INDUSTRY AND THE COMMUNITY
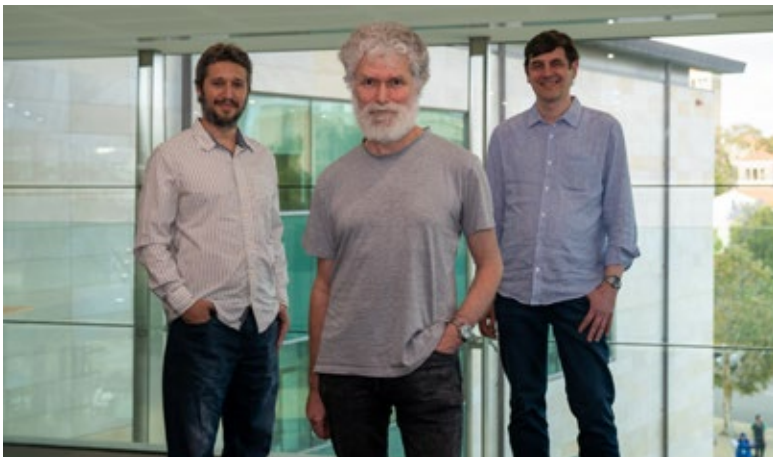
ICRAR

SUMMIT - SUPERCOMPUTER

How the world's most powerful supercomputer was used to test the 'brain' of the SKA.

On 3 October 2019, Professor Andreas Wicenec was nervous.

The head of data intensive astronomy at ICRAR was about to test a program he and his team had been working on for more than three years.

If it worked, the software could become the 'brain' of the world largest radio telescope—the Square Kilometre Array (SKA).

And it was about to be put to its biggest test yet.

Professor Wicenec's team was about to process more data than ever been tried in radio astronomy, on the world's most powerful supercomputer.

"We had no idea if we could take an algorithm designed for processing data coming from today's radio telescopes and apply it to something a thousand times bigger," he said.

"The test was going to tell us if we're able to deal with the data deluge from the SKA when it comes online in the next decade."

The billion-dollar SKA is one of the world's largest science projects.

Hour after hour, day after day, the telescope will peer deep into the Universe, looking back more than 13 billion years in time to the dawn of the cosmos.

The low frequency part of the SKA—which is set to be built in Western Australia—is expected to generate about 550 gigabytes of data every second.

It's so much data, the observations would instantly overwhelm most of today's supercomputing facilities.

So how do you test data pipelines for a telescope that hasn't been built and a supercomputer that doesn't exist yet?

You simulate it on the smartest, most powerful supercomputer in the world.

In 2019, that was Summit, at the US Department of Energy's Oak Ridge National Laboratory in Tennessee.

The facility is a million times more powerful than a high-end laptop, with a peak performance of 200,000 trillion calculations per second.

To simulate data processing for the SKA, Professor Wicenec's team in Perth worked with

researchers at Oak Ridge National Laboratory and Shanghai Astronomical Observatory.

Oak Ridge software engineer and researcher Dr Ruonan Wang, a former ICRAR PhD student, said the huge volume of data used for the SKA test run meant the data had to be generated on the supercomputer itself.

"We used a sophisticated software simulator written by scientists at the University of Oxford, and gave it a cosmological model and the array configuration of the telescope so it could generate data as it would come from the telescope observing the sky," he said.

"Usually this simulator runs on just a single computer, generating only a tiny fraction of what the SKA would produce.

"So we used another piece of software written by ICRAR, called the Data Activated Flow Graph Engine (DALiuGE), to distribute one of these simulators to 27,360 of the 27,648 graphics processing units that make up Summit.

"We also used the Adaptable IO System (ADIOS), developed at the Oak Ridge National Laboratory, to resolve a bottleneck caused by trying to read and write so much data at the same time."

The simulated observations mimicked a time in the early Universe known as the Epoch of Reionisation—when the first stars and galaxies were beginning to form.

Professor Tao An, from Shanghai Astronomical Observatory, said the data was first averaged down to a size 36 times smaller.

"The averaged data was then used to produce an image cube of a kind that can be analysed by astronomers," he said.

"Finally, the image cube was sent to Perth, simulating the complete data flow from the telescope to the end-users."

Professor Wicenec needn't have been worried—the test was a success.

"It worked really, really well," he said. "We were able to average 400 GB of data processed a second on average, with a maximum of 925 GB a second."

"And the effective throughput of the complete simulation was about a factor of two better than what is required by the SKA."

In 2020, the work saw the team shortlisted for the prestigious Gordon Bell Prize for outstanding achievement in high-performance computing.

The award is commonly referred to as the 'Nobel Prize of supercomputing', and rewards innovation in applying high-performance computing to science, engineering, and large-scale data analytics.

"The fact that we needed about 99% of the world's biggest supercomputer to run this test successfully shows the SKA's needs exist at the very edge of what today's supercomputers are capable of delivering," Professor Wicenec said.

TESTING DALiuGE ON SUMMIT

400 GIGABYTES
Average data processed per second, equivalent to 1600 YouTube videos

925 GIGABTYES
Maximum data processed per second, equivalent to 3700 YouTube videos

THREE HOURS
Time the simulation ran on Summit

64,900 TRILLION
Average mathematical operations per second

4560 COMPUTERS
Used during simulation