



Cambridge University reaches for the stars with Kognitio WX₂ and HP FCS

“Gathering, classifying and cross-referencing very large quantities of data is a highly demanding process. Kognitio WX₂ has added a new dimension to our data analysis by dramatically accelerating the process.” *Jim Lewis, senior research associate, CASU*

Introduction

The discovery of planets, a sun-centered solar system, Halley's Comet, general relativity. These are some of astronomy's biggest discoveries and they have spanned more than 500 years. Technology has continued to enhance astronomical discovery. Today the Cambridge Astronomical Survey Unit (CASU), part of the Institute of Astronomy at Cambridge University, is implementing technology that could accelerate some of astronomy's biggest future discoveries.

Modern astronomers no longer peer through telescopes at the dead of night. Instead, digital detectors do the peering, with specialist software analyzing the resulting images and producing a catalogue entry for each object that appears in every frame. Each catalogue entry contains information about the object's location in the sky and a vast array of information about its light output in different wavebands. A typical frame can contain around 10,000 objects, each having approximately 100 separate measured properties, and each night astronomers might take a few thousand images. As a result, gathering such a large amount of information equates to a data processing problem that is far from trivial.

Most catalogues are stored in databases and astronomers use SQL queries rather than optical devices to search the skies. But because these catalogues can contain billions of objects and because the nature of the searching is quite ad hoc and unpredictable, the process of searching can be very slow and tedious when using a conventional database.

Challenge

The biggest challenge for Cambridge Astronomical Survey Unit was the time it took to make the telescope imaging data available, data that the unit collects on the stars every night and which is forecast to grow daily volumes in excess of 500GB per night. It was taking far too long for the data to be loaded into CASU's database and then made available for analysis and research purposes. This was because each data set had to be indexed in multiple ways so that CASU's database could then store the data appropriately and make it searchable for research queries around the world.

Images had to be grouped into catalogues which then had to be uploaded, cross-referenced and heavily indexed to enable research astronomers to ask useful and complex data mining queries. As each catalogue

contains billions of objects, this process used to take months when carried out using traditional technologies. Moreover, the worldwide trend to make all astronomical databases available on the Internet to allow researchers to mine high-quality data, was placing additional pressure on the catalogue and archiving facilities. And it is this data that is the foundation for thousands of astronomy and science discoveries around the world. Such discoveries can only happen as fast as this research is made available.

Rising to the challenge

CASU turned to Kognitio and HP Flexible Computing Services (FCS) for a technology solution that could enable the survey unit to analyze and catalogue this research faster, so it could be used for important astronomy and science discovery worldwide. In this field of large and complex data analytics, where CASU was faced with the challenge of how they could quickly analyze highly granular data, they needed a solution that was high performance but not high cost.

Running unpredictable ad-hoc queries on large volumes of data is exactly what Kognitio WX₂ does best, so CASU took the opportunity to use the high performance analytical database software solution using hardware provided by HP FCS in their datacenter in Houston, Texas. CASU first used a catalogue that contained 500 million objects, each of which equated to a database row. The catalogue took about one and a half hours to load into WX₂ and the database itself provided Cambridge scientists with some remarkable performance. For instance, CASU could scan every object for a match to some given criteria in less than a second. Complex queries that search for objects which differ from the norm ran in just a few seconds.

Today, CASU is continuing to use Kognitio WX₂ to help them in their aim to understand as much about space objects as possible.

Benefits

Kognitio WX₂ has enabled CASU to get answers to detailed queries from large amounts of granular data in just seconds. In addition, it has dramatically cut set up costs by eliminating costly proprietary hardware thanks to its ability to run on standard hardware. Being able to tap into the HP Flexible Computing Services power pool has also enabled CASU to lower costs while maintaining performance.

The Challenge

Leading research unit seeks to process and analyze gigabytes of data generated every night from the observation of space objects in a timely manner.

The Solution

CASU uses Kognitio WX₂ and HP FCS which allows them to benefit from highly flexible yet affordable state-of-the-art data analytics software and on-demand computing power.

The ROI

Start-up and running costs have been kept to a minimum and performance has increased dramatically with operations that took days now only take minutes. As such, more complex queries can be asked of the data and more discovery projects can be initiated.

