Future Generation Computer Systems 76 (2017) 92–97

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Forward Observer system for radar data workflows: Big data management in the field

Richard Knepper *, Matthew Standish

Indiana University, Research Technologies, 2709 E Tenth St., Bloomington, IN, 47408, United States

HIGHLIGHTS

- We created an in flight data management and analysis system for NASA's Operation Ice Bridge.
- In-flight system allows for protection of the data as soon as it is collected through out its lifetime.
- Real-time processing allows researchers to make course corrections based on the results.
- Including an in-flight system permits the reduction of staff on field missions and saves costs.

ARTICLE INFO

Article history:
Received 9 December 2013
Received in revised form 29 April 2017
Accepted 21 May 2017
Available online 29 May 2017

Keywords:
Microcomputers
Information storage
Physical sciences and engineering

ABSTRACT

There are unique challenges in managing data collection and management from instruments in the field in general. These issues become extreme when "in the field" means "in a plane over the Antarctic". In this paper we present the design and function of the Forward Observer a computer cluster and data analysis system that flies in a plane in the Arctic and Antarctic to collect, analyze in real time, and store Synthetic Aperture Radar (SAR) data. SAR is used to analyze the thickness and structure of polar ice sheets. We also discuss the processing of data once it is returned to the continental US and made available via data grids. The needs for in-flight data analysis and storage in the Antarctic and Arctic are highly unusual, and we have developed a novel system to meet those needs.

We describe the constraints and requirements that led to the creation of this system and the general functionality which it applies to any instrument. We discuss the main means for handling replication and creating checksum information to ensure that data collected in polar regions are returned safely to mainland US for analysis. So far, not a single byte of data collected in the field has failed to make it home to the US for analysis (although many particular data storage devices have failed or been damaged due to the challenges of the extreme environments in which this system is used).

While the Forward Observer system is developed for the extreme situation of data management in the field in the Antarctic, the technology and solutions we have developed are applicable and potentially usable in many situations where researchers wish to do real time data management in the field in areas that are constrained in terms of electrical supply.

1. Introduction

There are unique challenges in managing data collection and management from instruments in the field in general. These issues become extreme when "in the field" means "in a plane over the Antarctic". In this paper we present the design and function of the Forward Observer a computer cluster and data analysis system that flies in a plane in the Arctic and Antarctic to collect, analyze in real time, and store Synthetic Aperture Radar data used to analyze the thickness and structure of polar ice sheets. We also discuss the processing of data once it is returned to the continental US and made available via data grids. The needs for in-flight data analysis and storage in the Antarctic and Arctic are highly unusual, and we have developed a novel system to meet those needs.

Ailamaki et al. [1] describe a number of purpose-built systems for the capture, analysis, and management of scientific data by high-performance systems. "Big science" projects such as analysis of data from the Large Hadron Collider (LHC), earthquake simulations, and astronomy often involve very big storage and compute facilities. In some cases, such as analysis of LHC data, computational scientists have the benefit of years to prepare massive international storage and computational grids to analyze data. When

* Corresponding author.
E-mail addresses: rknepper@iu.edu (R. Knepper), mstandis@iu.edu (M. Standish).

http://dx.doi.org/10.1016/j.future.2017.05.031
0167-739X/

Table 1
Data rates for Ice Bridge field missions.

Campaign	TB per day	Total data (TB)
2009 Antarctica	2.8	33
2011 Greenland	3.38	105.57
2011 Antarctica	2.86	56.77
2012 Greenland	4.02	161.50
2012 Antarctica	2.47	39.55
2013 Greenland	1.98	46.68
2013 Antarctica	1.69	11.88
2014 Greenland	1.81	86.81
2014 Antarctica	1.69	43
2015 Greenland	2.13	70.29
2016 Greenland	1.81	29
2016 Antarctica	2.13	60

Note: 2015 Antarctica mission was conducted by a minimal team and did not utilize the system.

doing research in the field there are very often constraints on the size, power consumption, and durability of computing equipment used in the field to do data collection [2]. A brief survey of literature on radar data processing systems indicates that due to the nature of SAR data analysis, data is a concern, as it produces holographic or volumetric data, and that analyses of data can be difficult, however, work on instrument data has not previously been a concern [3–7].

Starting in 2007, Indiana University (IU) began working with the University of Kansas' Center for the Remote Sensing of Ice Sheets (CReSIS) in order to support management of data collected by CReSIS during arctic field missions. This began with the PolarGrid project [8] to create cyberinfrastructure in support of remote sensing systems, particularly Synthetic Aperture Radar (SAR) imaging of polar ice sheets. Research into the ice sheets at our poles is vital information to understand the effects of climate change, of glacier interactions, and to monitor the thickness and condition of ice sheets over time. While there is still uncertainty regarding the rate of change of the global climate and the causes of that change, what we know certainly is this: there can be no confidence that data collected next year will provide data identical to data collected this year. That means that there is a tremendous premium for researchers working in polar regions during any year to collect as much data as possible and to ensure that the data make it back to the laboratory for analysis. Fig. 1 shows the general workflow for gathering ice sheet data, to be processed back at IU and CReSIS.

Continuing this relationship, IU provides data collection and management for SAR data gathered in flight by CReSIS staff participating in NASA Operation Ice Bridge. NASA Operation Ice Bridge will last six years "the largest airborne survey of Earth's polar ice ever flown. It will yield an unprecedented three-dimensional view of Arctic and Antarctic ice sheets, ice shelves and sea ice. These flights will provide a yearly, multi-instrument look at the behavior of the rapidly changing features of the Greenland and Antarctic ice" [9]. Ice Bridge flights involve a SAR radar system and computer equipment taken over a predetermined path to image polar ice sheets in a plane (either the NASA Operation Ice Bridge DC-8 or P-3 aircraft) [10]. The NASA Operation Ice Bridge radar systems [11–13] are enhanced and improved on an ongoing basis and as a result the rate of data collection goes up every field data collection season. Table 1 below shows the rate of data production over time and the expansion of the amount of data collected per flight day and season.

Prior to 2012 SAR data collected on NASA Operation Ice Bridge (hereafter OIB) flights were written directly to disk, and analyzed in as close to real time as possible. In 2011 this near real time analysis proved absolutely essential to the scientific mission when vibrations from the plane interfered with the spinning disks used by SAR radar equipment, causing data to be only partly usable for analysis. Other environmental conditions such as the physical shock of shipping and handling in difficult conditions, thermal changes due to differences in environment and weather, and issues due to unconditioned power in the hotel rooms and offices used by the team for processing contributed to multiple failures of drives that were not destructive to radar data, because the team had already made duplicate copies of the data. These incidents and the risks they exposed convinced the Indiana University team that it was essential to develop true real time data analysis facilities so that the SAR data could be observed by scientists in the plane as data were being collected. In addition to ensuring data quality, this creates the possibility that the planned route for a flight could be changed in response to detection of some particularly interesting feature in the ice. IU thus developed a new and novel computing system called the Forward Observer.

The Forward Observer system (hereafter FO) represents a considerable improvement over earlier in-flight data management capabilities. FO provides the ability to capture, process, and display data in real time while radar instruments are recording data, whereas systems that had been previously used as part of NASA OIB data collection only captured radar data files. Metadata associated with the flight and GPS coordinates are associated with the data in real time. The FO interface provides information analysis of data in real time and in addition a real time interface to the data management process. This process includes copying data to multiple different physical devices (with verification), so that even if a particular storage device is lost or damaged, one of the copies of each byte of data collected makes it back to the US for analysis. Working in extreme environments sometimes means things do not go perfectly, and the data copying interface allows for things such as monitoring the speed at which data are written and checked.

2. The Forward Observer system overview

The FO was designed first and foremost to provide better data replication and performance, improving on the initial idea of a simple RAID array installed in the plane. A single RAID array would provide a minimal amount of data protection in the event of one or two drives failing, and depending on the architecture would provide sufficient speed for writing the data to drives. A RAID array alone would not fulfill the requirements of a system that provided checksum verification of the data written to the storage, allowed separate paths for data to return to the processing facilities, or provided a read-only workspace for real-time analysis.

2.1. System constraints and requirements

IU developed the FO in order to meet several and in some cases stringent constraints:

- Power on the NASA plane is limited to two 110 V circuits that must support all of the running systems, including the radar, airborne topography mapper, gravimeter, and lidar instruments and systems.
- Space is limited to an aircraft-rated machine rack that fits within the cabin enclosure which is 54″ tall × 24″ deep.
- The system must also be as light as possible so that the equipment and passengers together fit into the weight limits for the plane.
- All of the equipment must be shock-mounted and vibration resistant; the ride in a plane over Greenland or the Antarctic can be very bumpy. All of the equipment that is mounted on the plane experiences flight conditions well in excess of what is allowable for commercial flights and must be rated for safety in the event of high turbulence or other unsafe conditions. Vibration occurring during flights is particularly difficult to manage for mechanical (rotating) disk storage.