NISAR Real Time Data Processing – A Simple and Futuristic View

Manavalan(✉)

Centre for Development of Advanced Computing (C-DAC), Bangalore, India
manavalan@cdac.in, saimanaindia@gmail.com

Abstract. NASA – ISRO Synthetic Aperture Radar (NISAR) is an Interferometric synthetic aperture radar mission will be launched during 2021. NISAR will be the first radar imaging satellite to use dual frequency NASA L-band SAR and ISRO's S-band SAR payloads. In a day NISAR generates near to 85 TB of raw data and during its life time it generates more than 140 PB of data. Processing of any such Big Data is a tough task for standalone computers and servers. Moreover, SAR based application models should support in making available the critical processed results in real time or near real time mode. In line to this, a HPC based NISAR processing architecture that can meets the data processing and analyzing needs of NISAR like Big Data generating space missions is conceptualized and proposed. The proposed model is unique due to its support to the simultaneous processing of different frequency, polarization channels of SAR data as well as in virtually visualizing the end results. All together the proposed HPC-Big Data system will make available the simulation results within shortest possible time.

Keywords: NISAR · HPC · Big Data · System architecture

1 Introduction

Synthetic Aperture Radar (SAR) based earth imaging technology is becoming a most sought after earth imaging technology due to its strength in capturing the terrain information during all weather and round the clock periods as well as in capturing the earth surface both in high resolution spot mode and course resolution wide swath mode [1, 2]. Many countries are shifting considerable part of their space budget towards SAR based space imaging system due to its consistent reliability at critical occasions. With multiple advances happening in various fronts of Synthetic Aperture Radar (SAR) technology, as on date processing of any such high resolution multi frequency, multi polarization SAR signals and corresponding pre and post image processing operations requires, the support of using advanced HPC systems and related Big Data technologies as this can only make available the end results in near real time or real time mode [3, 4].

To be specific the NISAR (NASA-ISRO Synthetic Aperture Radar) mission which is a joint space mission of NASA and ISRO will be launched in the year 2021 is an dual frequency L- and S-band polarimetric SAR with 12-day interferometric orbit that will support systematic global coverage including cryosphere [5, 6]. As on date, projections

R. Patgiri et al. (Eds.): BigDML 2019, CCIS 1317, pp. 95–101, 2020.
https://doi.org/10.1007/978-3-030-62625-9_9

of NISAR satellite data acquisitions exhibits, that in a single day pass it can captures more than 85 Terabytes (TB) of earth observation data and during its life span of time this mission alone likely to generate 140 Petabytes (PB) of SAR raw data [7, 8]. Calibration of raw data and subsequent pre and post processing processes further generates TB to PB of SAR images. Similar situations likely to happen with many other future space borne missions as the respective SAR technology is already moving from space borne to air borne UAVSAR (Uninhabited Aerial Vehicle Synthetic Aperture Radar) mode which is already in use in many defense sectors of the world [9]. Overall, in future many such SAR systems generates TB to PB of data due to its multi frequency, multi-polarization capturing modes as well as due to the preference of generating high resolution spot mode images. Hence, processing of NISAR like Big Data will certainly become a challenge to many HPC cluster environment. The primary challenges includes:

- Storage of PBs of SAR raw data of each pass of satellite missions
- Temporal data indexing of routine satellite pass periods and in supporting specific data extraction of smaller terrain
- In supporting, the data download process of user specific data frames covering the region of interest or specific study area
- Computation involved in calibrating the source/raw radar signals to generate first level SAR data (Level 0B data)
- Pre and post processing of calibrated data as per the need and requirements of application specific research models
- Real Time or Near real time data processing of TBs of SAR data frames of critical locations is still a major challenge

To overcome many such challenges the data utilization of NISAR like dedicated SAR mission's has to be supported through distributed HPC storage and processing clusters. Any such setup that supports the real time/near real time processing requirements of SAR specific application models helps in delivering the critical information within shortest possible time. The only thing expected from the application researcher group is that they have to fine tune and port their application models into such dedicated HPC clusters which is specially configured as per the processing requirements of NISAR like SAR system. Any model build over such dedicated or distributed cluster environment will be subsequently hardcoded as a routine service of respective application, mainly to support the temporal data processing of future requirements. In general, during regional scale disaster modelling when the amount of input data increases, the advantages of distributed HPC processing environment is quickly engrossed due to its scalable nature of providing the computing resources with reference to the input data size and other modelling requirements which altogether reduces the processing time. Such SAR processing systems on distributed environment is already well setup and supported by European Space Agency (ESA) through its Grid Processing on Demand (G-POD) environment where user move their specific simulations code or algorithms to the dedicated cluster of G-POD and then relate the source data that resides in respective G-POD clusters [10, 11]. Similar system supporting SAR data processing was also prototyped with India's first Grid Computing environment Grid GARUDA [12] which is currently non-operational.

In line to above detailed distributed SAR data processing systems, this article mainly discusses the core design perspectives of a HPC based system architecture that supports both real time as well as offline data processing of NISAR like SAR missions. Emphasis has been given in evolving a Big Data based HPC system which can handle TB of input SAR data processing of same geographical region of different frequency, polarization and disseminates the processed end results through remote visualization techniques. Any such operational setup supports the researchers as well as end users to visualize the geographical or terrain signature content of SAR images of his area of interest simultaneously at different frequencies, polarization channels [13]. Due to this simultaneous processing of different frequencies, different polarization channels of SAR system this article differs from existing HPC based Grid/cluster environment. The proposed architecture of the NISAR data processing system is detailed in the following section.

2 Architecture

The architecture of proposed SAR SLC data processing system which support simultaneous processing of NISAR data is shown in Fig. 1. The efforts involved in RADAR signal processing and subsequent SAR raw data to SLC data (Single Look Complex) processing of the NISAR data is out of the scope of this article as emphasis has been given in evolving an HPC architecture that can suit for the simultaneous processing of NISAR SLC data of different frequencies and polarizations. Hence, as shown in Fig. 1, it has been assumed that SAR raw data and corresponding SLC data of same geographical region resides in Storage Cluster-I which shares input SLC data for the next level processing.

From storage cluster-I, the SLC data of L- and S-band SAR of same geographical regions is transferred to two different computing clusters where the first one will be processing the L-band SLC data and the second one supports processing of S-band SLC data. Depending upon the multi-look calibration factor (1:4; 2:8 and 3:16), which is mainly decided by the SAR data processing scientist with reference to the end user application specific requirements, the master node of individual L-and S- band computing cluster allocates the no of slave nodes as shown in Fig. 1. For example when 1:4 and 2:8 multi-look parameter is hardcoded as part of the programming paradigm only two slave processing nodes for each L- and S-band clusters will be activated. When the user further interested to compare the output of 3:16 multi-look image additional slave processing node will be activated. Subsequent to this, the respective SLC data preferably has to be copied to each slave node or communication between the slave node and master node has to be maintained all the time by which the slave nodes can read the source SLC data that has been placed in master node. In the first option, the data is processed within short period of duration as the input SLC data is copied to each computing slave node whereas the second option consumes considerable processing time as communication between master and slave node has to be maintained to support constant read and write process which becomes a memory intensive model. Hence, with reference to meeting the real time requirement, simultaneous transfer of source SLC data to compute node i.e. slave nodes has to be automated through a workflow environment. At the end of this process

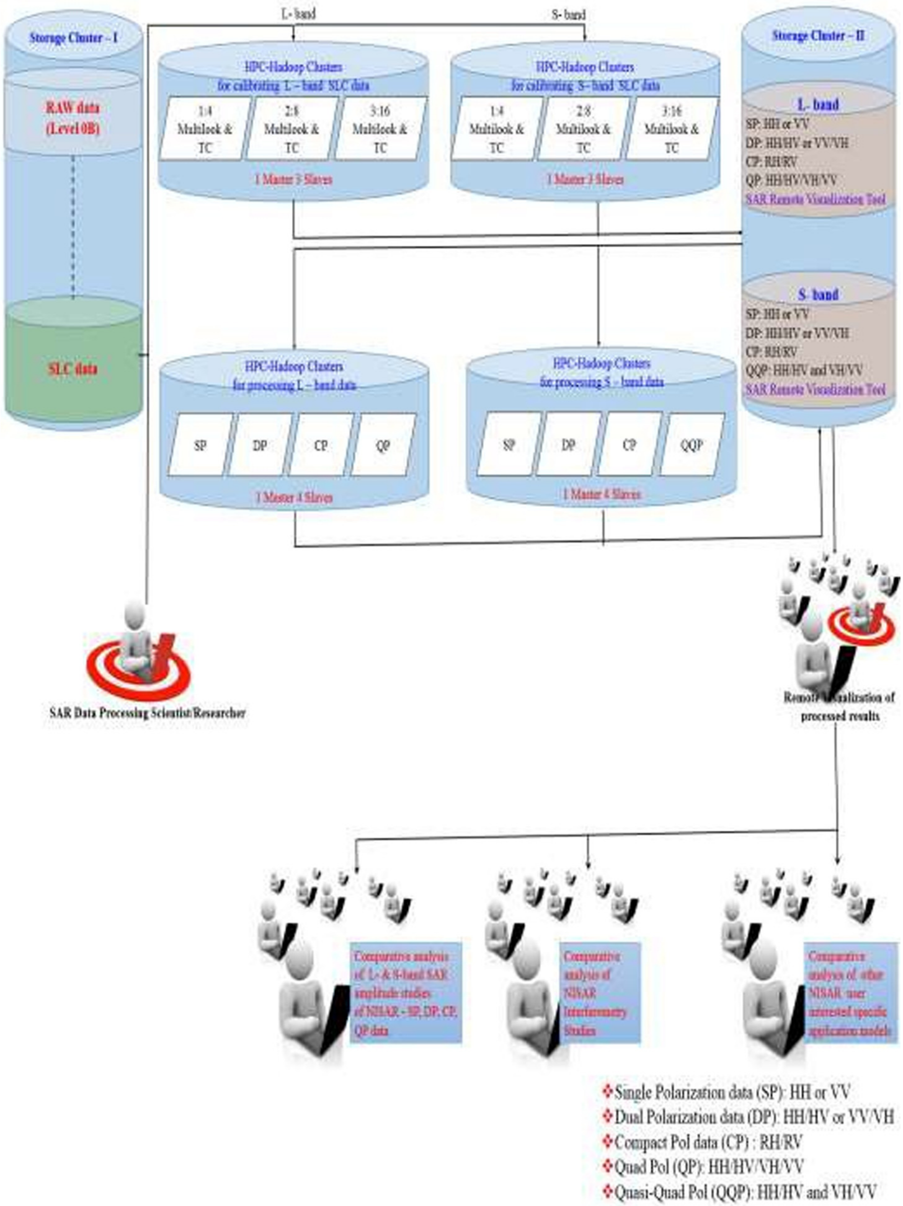

Fig. 1. Architecture of NISAR SLC SAR data processing system

multi-looked terrain corrected SAR images are generated and same is transferred to the Storage Cluster-II where L- and S-band SAR images will be indexed in corresponding storage nodes. Following such intermediate processes helps the user to retrieve the processed SAR image of different multi-looked data straight from the Storage Cluster-II without performing any SLC data calibration process which is a time consuming task. To support terrain correction process it is mandatory the user has to relate the accurate DEM of the respective region or need to use open source DEM [14].

To continue the next level processing, as shown in Fig. 1 the different polarization SAR images (SP, DP, CP and QP) of L- and S-band frequency has to be shared to two different computing clusters as one supports the L-band data processing and the other one supports S-band data processing. By this, the L-band Cluster will be supporting Single Polarization (SP) - HH or VV; Dual Polarization (DP) - HH/HV or VV/VH; Compact Pol (CP) - RH/RV and Quad Pol (QP) - HH/HV/VH/VV data processing whereas in the case of S-band Quasi-Quad Pol data (QQP) - HH/HV and VH/VV replaces the Quad Pol data of its counterpart. In such case, each individual computing cluster transfers one single polarization image from an individual computing node by which four slave nodes has to be allotted to process all the polarization channels of L- and S-band SAR images. In parallel to this, it is expected the application model which has to extract the end user specific information from each individual polarization SAR channels will be compiled in each slave nodes and respective simulation is executed immediately after receiving the multi-looked image from Storage Cluster-II. To meet this purpose another automated workflow environment which can copy the multi-look data to the different slave nodes as well as monitor the respective data transfer and compute process has to be developed and deployed at master node level. In addition to above data processing methods, it is also possible to allocate the individual slave nodes to process the individual polarization image such as HH, VV, HH/HV, VV/VH, RH/RV, HH/HV, VH/VV along with the application model that has to be ported to each slave node. This is not addressed in this article as well as reflected in Fig. 1 mainly to maintain the simplicity of the proposed architecture.

Once the application simulation is completed at slave computing nodes the results are transferred back to the Storage Cluster-II and indexed against corresponding L- and S-band storage nodes. A SAR visualization tool which supports remote visualization of individual SAR images and corresponding results has to be installed in each L- and S-band storage nodes. To support remote visualization the X-11 forwarding has to be suitably configured at Storage Cluster-II nodes as well as to the remote system of the end-user with a stipulated time frame. This will support the end user in simultaneously viewing and comparing the results of his application that has been derived using different frequency, polarization SAR channels. At the end of the process completion, simultaneous remote visualization and its analysis, end users will certainly bring out many significant insights of the application as the data of same geographical region of different frequencies, polarizations has been derived, compared and analyzed in the same time [14].

3 Conclusion

As on date there is no single distributed SAR processing environment is proposed that can support the simultaneous processing of different frequency, polarization data where the end user will be able to visualize and compare the processed results of their area of interest of same geographical region. Setting up the proposed NISAR processing system can meet such expectations by which it will become a significant HPC-Big Data platform for worldwide researchers who can build specific application models in the fields of - Natural Ecosystem studies such as River linking, Agriculture Biomass estimation, Crop monitoring, Forest mapping and Biomass estimation, Mangroves-Wetlands mapping; Land Surface Deformation studies such as Inter and Co-seismic deformation studies, Land Subsidence, Landslides and Volcanic deformation studies; Cryosphere studies covering Polar Ice Shelf monitoring & estimation, Sea Ice dynamics, Glacier dynamics of Himalayan region, Mountain Snow-Glacier relationship studies; Coastal Studies & Oceanography which includes Coastal erosion, shoreline dynamics, High and Low Tide lines mapping, Bathymetry studies, Ocean surface wind dynamics, Ocean wave spectra, Ship detection; Disaster Response studies covering Floods, Forest Fire, Oil Spill, Earthquakes as well as analyzing the impact of any other man induced disasters; Geological applications such as Structural & Lithological mapping, Lineament study, Paleo-Channel study, Geomorphological mapping, etc.

To be specific NISAR mission during its lifespan likely to generate 140 PB of raw data which in turn will generates same amount of additional data during its pre, post processing levels as well as during various application specific simulations as mentioned above. For any application, even for a small scale terrain data of such nature cannot be handled as well as processed by high end standalone computers or servers. In such case setting up the proposed HPC-Big Data SAR data processing platform will become a timely solution. To completely absorb the advantages of such distributed processing environment the application scientists also need to find tune and enable their application models as part of the proposed system as this meets the expectations of real time simulations where the end results from different frequency and polarization SAR channels can be derived within shortest possible time and compared.

References

1. Manavalan, R.: SAR image analysis techniques for flood area mapping - literature survey. Earth Sci. Inf. **10**(1), 1–14 (2016). https://doi.org/10.1007/s12145-016-0274-2
2. Manavalan, R.: Review article, Review of synthetic aperture radar frequency, polarization, and incidence angle data for mapping the inundated regions. Int. J. Appl. Remote Sensing **12**(2), 021501 (2018). https://doi.org/10.1117/1.JRS.12.021501
3. Lanari, R (2015) From ERS-1 TO SENTINEL-1, A big data challenge for 25 years of DInSAR observations. Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Milan, pp. 1460–1463. https://doi.org/10.1109/igarss.2015.7326054
4. Elefante, S., Zinno, I., De Luca, C., Manunta, M., Lanari, R., Casu, F.: Big DInSAR data processing through the P-SBAS algorithm. In: IEEE International Geoscience and Remote Sensing Symposium, Milan, pp. 2696–2698 (2015). https://doi.org/10.1109/igarss.2015.7326369
5. Rosen, P.A, Hensley, S., Shaffer, S., Veilleux, L., RajuSagi, V., Satish, R.: The NASA-ISRO SAR mission - An international space partnership for science and societal benefit. In: IEEE Radar Conference, Arlington, VA, pp. 1610–1613 (2015). https://doi.org/10.1109/radar.2015.7131255
6. Xaypraseuth, P., Satish, R., Chatterjee, A.: NISAR spacecraft concept overview: design challenges for a proposed flagship dual-frequency SAR mission. In: IEEE Aerospace Conference, Big Sky, MT, pp. 1–11 (2015). https://doi.org/10.1109/aero.2015.7118935

7. Earth Science Data Systems (ESDS) Program. https://earthdata.nasa.gov/getting-ready-for nisar
8. Blumenfeld, J.: Getting Ready for NISAR – and for Managing Big Data using the Commercial Cloud, 17 April 2019. Accessed on 12 Oct 2019
9. Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR), NASA-JPL. 22 January 2019. https://uavsar.jpl.nasa.gov. Accessed 12 Oct 2019
10. ESA Grid Processing on Demand. https://gpod.eo.esa.int/. Accessed 12 Oct 2019
11. De Luca, C., et al.: Unsupervised on-demand web service for DInSAR processing: The P-SBAS implementation within the ESA G-POD environment. In: IEEE International Geoscience and Remote Sensing Symposium, Milan, pp 2692–2695 (2015). https://doi.org/10.1109/igarss.2015.7326368
12. Grid Computing, Centre for Development of Advanced Computing. https://www.cdac.in/index.aspx?id=grid_comp. Accessed 12 Oct 2019
13. Manavalan, R., Rao, Y.S., Krishna Mohan, B.: Comparative flood area analysis of C- band VH, VV and L-band HH polarizations SAR data. Int. J. Remote Sens. **38**(16), 4645–4654 (2017). https://doi.org/10.1080/01431161.2017.1325534
14. Manavalan, R., Rao, Y.S.: DEM and SAR image based flood feature extraction tech- niques to map the deep and shallow flood inundated regions of known as well as remote disaster regions. Geocarto Int. **29**(7), 745–757 (2013). https://doi.org/10.1080/10106049.2013.838310