1. Abstract

RFI data is stored in a variety of different format at MeerKAT scattered in several files i.e. tables, arrays, images, PDFs, etc., which makes it difficult for many potential users with varying needs to efficient store, manage and quickly retrieve this data. We are investigating if we can store all associated forms of RFI data in a unified storage while providing the underlying database functionalities such as querying, indexing, etc. The proposed framework (polystores) integrates data from different sources to support such functionalities required by the users e.g. cross-systems querying.

2. Current RFI storage challenge

- **Table (a)**: RFI metadata - mainly static self describing data, e.g. frequency, source, etc.
- **Table (b)**: Known RFI transmitters (ICASA) - frequencies of transmitters.
- **Array (c)**: Two-dimension array (3600 X 14700) data - sequence data in HDF5 files.
- **Image (d)**: RFI spectral image illustrating each sequence point in array dataset.

RFI data scattered files in tables, arrays, images, PDFs etc. (a) Table: RFI metadata - mainly static self describing data, (b) Table: ICASA - Known frequencies of transmitters, (c) Two-dimension Array: (3600 X 14700) data - sequence data in HDF5 files, and (d) Image: RFI spectral image illustrating each sequence point in array dataset.

3. Research aim

Supply a model of the RFI database that will assist astronomers and engineers to store, manage and quickly retrieval RFI data.

4. Proposed framework - Polystores

- **A)** Figure A) illustrates an integrated framework of heterogeneous data stores that exposed through a middleware that enforces polyglot persistence –permits cross-system querying, indexing etc. Figure B) is a basic prototype - Postgres (metadata/tables), SciDB (array data), accumulo (RFI text/images) and s-store (RFI real-time data).

5. Preliminary investigation

1) **Figure 1)** -- over 50% users stored RFI data mainly in CSV and HDF5 files. While **Figure 2)** -- 75% users ranked RFI storage and classification as high priority. Although for classification to be sustainable, storage should be successful.

2) **Figure 1)** -- over 50% users stored RFI data mainly in CSV and HDF5 files. While **Figure 2)** -- 75% users ranked RFI storage and classification as high priority. Although for classification to be sustainable, storage should be successful.

6. Evaluation

- Speed of execution (single & mixed queries)
- Scalability – in terms of volume & applications

7. Project outcomes

- Project will forms a basis of RFI identification
- Provision of reliable RFI storage at MeerKAT
- Solution can be extended to others Observatories e.g. MRO, LOFAR, NRAO etc.