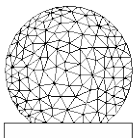


Data to Correlators Challenges / Expectations

Chet Ruszczyk



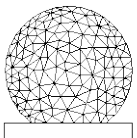
MIT
HAYSTACK
OBSERVATORY

VGOS Observational Data

- Two stations types for how observational data stored:
 - Disk modules
 - Network attached storage NAS
- VGOS Specification (VLBI 2010) states three type of stations network access:
 - 4Gbps / 1Gbps / 100 Mbps or less (none)
 - None – based upon importance of station location
 - Network costs may be prohibitive based on location, but from science standpoint critical
- Disk module based stations
 - e-transfer, or
 - Ship disk modules to correlation center
- e-transfer only stations only one option

VGOS Correlation Center (CC) Requirements

- To process VGOS 24 hour sessions CCs must be able to handle
 - Receive / send modules
 - e-transfer
- Present typical 24 hour session
 - Recording 8Gbps
 - due to only observing 512 MHz bands
 - 30 seconds on source, 30 second slewing
 - allow storage to a single module for a session
 - reducing shipping costs
 - Max storage is 48TB
 - Min storage 24TB

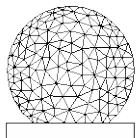


VGOS CC Requirements (con't)

- If e-transfers from 4 stations for VT session
 - 160TB of Network Attached Storage (NAS) required
- For efficient e-transfer of data
 - 5 Gbps
 - Based upon data arriving to correlator within 5 days for processing
 - This excludes other S/X sessions that must be e-transferred to correlators
- Recording rate for VGOS to increase to 16Gbps
 - Full 1024 MHz band recording
 - 15 seconds on source, 30 seconds for slew
 - 33% more data / session
- What are the capabilities today?

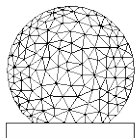
Network Last Mile (or km)

Station	Maximum	Using
GGAO	1Gbps	1Gbps
Ishioka	10Gbps	1Gbps
KPGO	1Gbps	100Mbps
Onsala	20Gbps	9Gbps
Wettzell	1Gbps	800Mbps
Westford	1Gbps	1Gbps



Network Last Mile (or km)

Correlation Center	Maximum	Utilization	Storage
BONN	1Gbps	1 Gbps	1400 TB
GSI	10Gbps	?	513 TB
Haystack	10Gbps	10Gbps	172 TB
Shanghai	4Gbps	2Gbps	800 TB
USNO	1Gbps	950Mbps	278 TB
Vienna	10Gbps	?	230 TB

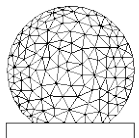


Challenges for IVS

- e-transfer stations are pushing the problem to correlation centers
 - In the past all institutions contributed to media pool, no longer true
 - Not providing storage resources
 - Due to funding agency requirements
 - Different correlation centers have different requirements (security)
 - Even if you could provide storage, requirements on NAS
 - How to address challenge

VGOS Data Format

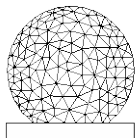
- Problem has been presented to IVS DB, with a solution recommended to have to come from top down.
 - I will be contacting sites to investigate cloud storage options
 - Cost for storing
 - Cost to retrieve data
 - Network data rate costs
 - Already know writing to is much cheaper than retrieving (at least in USA)
- Recommend the CC publish the data format they can process
 - Specify the data format you are cable of correlating (e-transfer)
 - Threads / band





Data Format (cont)

- Impact on e-transfer only sites
 - Specify the final version of data to be correlated (threads / band)
 - Have the stations convert the data to specified format
 - not push it to correlator if it cannot handle it
 - or if interested working with DiFX community, can investigate it.
- Ramifications
 - Data cannot be correlated
 - Time wasted
 - with e-transfer
 - with CC having to perform conversion (if conversion routines are not native to correlation software process)



Data Format (con't)

- MHO correlator has e-transfer only sites convert from N threads / 4 bands to 1 thread for 4 bands
 - Efficiency (correlation time) to process data
- Please note:
 - These are recommendations
 - What you do is your choice
 - Trying to make the process easier on CC's.
 - Based on our experience over the last 4 years.

Questions?

Thank you