LOFAR: Online and Offline Pipeline Processing

Anastasia Alexov
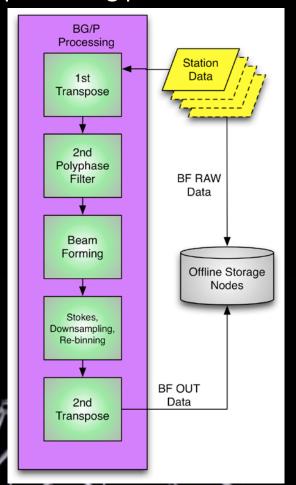
On behalf of the LOFAR BF Pipeline Group

IDIA 2011

LOFAR "Online" Data Processing

- IBM Blue Gene/P supercomputer, Groningen, NL:
  - 10,880 processor cores, 45.0 TFLOPS peak processing power
  - Data Correlation
  - Tied-Array Beam Pipeline
  - Online Coherent Dedispersion
  - Streams data to offline storage
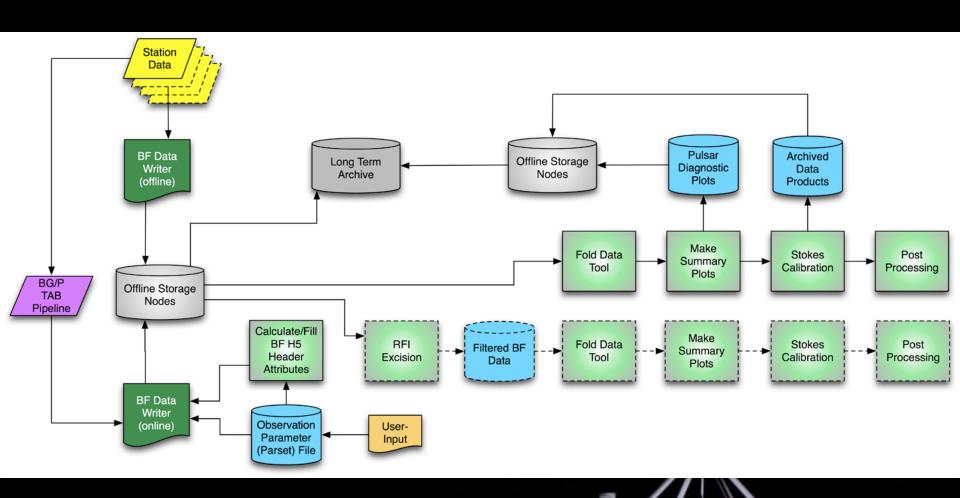- "Real-time processing"

LOFAR "Offline" Data Processing
(Bring compute power to the data)

- April 2011 @ 3rd generation of "offline" processing hardware with 100 nodes:
  - 24 CPU's (6 quad cores @2.1GHz)
  - 64GB RAM
  - 21TB diskspace, XFS filesystem (~2PB)
  - 50 Gflops
  - Ubuntu 10.04 LTS (long time support till April 2015) server edition (a.k.a. lucid)
  - Requirement: minimum write capacity 3.1 Gbps/node
  - 40 Gbps Infiniband network interface

	Read (MB/s)	Write (MB/s)	Read/Write (MB/s)	
HP P800	228	259	118 / 278	<== previous
Areca ARC1880	1052	919	418 / 670	<== new

- Science Pipelines (Imaging, Pulsar, Transient) are run on this system via Pipeline Framework

LOFAR "Offline" (BF) Data Processing

LOFAR Data Examples

Source:
  Pulsar B0329+54
Frequency (HBA):
  115-165 MHz
Sample size:
  (5.12 micro-sec)
  1.3 milli-sec
Observation time:
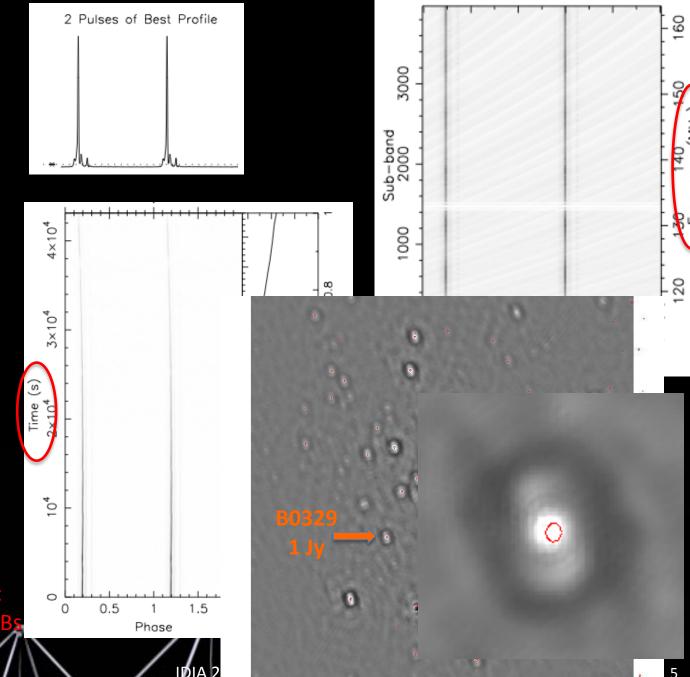 12 hours
Total # of samples:
  32,899,072
Total raw data size:
  1TB (23 Stations)

Simultaneous Mode:
  Imaging + BF

LOFAR Time-Resolution:
5 nano-seconds TTBs

B0329
1 Jy

LOFAR Long Term Archive (LTA)

- Long Term Archive has: 2.2PB disk, 5PB tape
- Currently 150TB LTA space is in use
- Archive uses Oracle DB
- Catalog: AstroWise technology adapted for LOFAR
- Need to process where you store your data; access to 22,600 cores via BigGrid and JUROPA:
  - Established a "Lightpath" between LOFAR Offline cluster and Long Term Archive to speed data transfer
  - In process of porting LOFAR Pipelines and Framework to GRID (Pulsar Search Pipeline working)
  - Change of methodology to think in terms of parallelizing software tasks (difficult using 30+ year-old off-the-shelf tools)