JDEM SOC Database Tests

Igor Mandrichenko,
Vladimir Podstavkov

August 4, 2010

Performance Tests

- Galaxy coordinate reconstruction

  https://cdcvs.fnal.gov/redmine/wiki/jdemsoc-private/DatabaseTests

- Catalog matching
  - For each detection found in a frame (0.5x1.0 degrees, ~80K objects), find matching galaxy in the "base catalog" with 500 million objects
  - closest object within ~0.5 arc sec circle

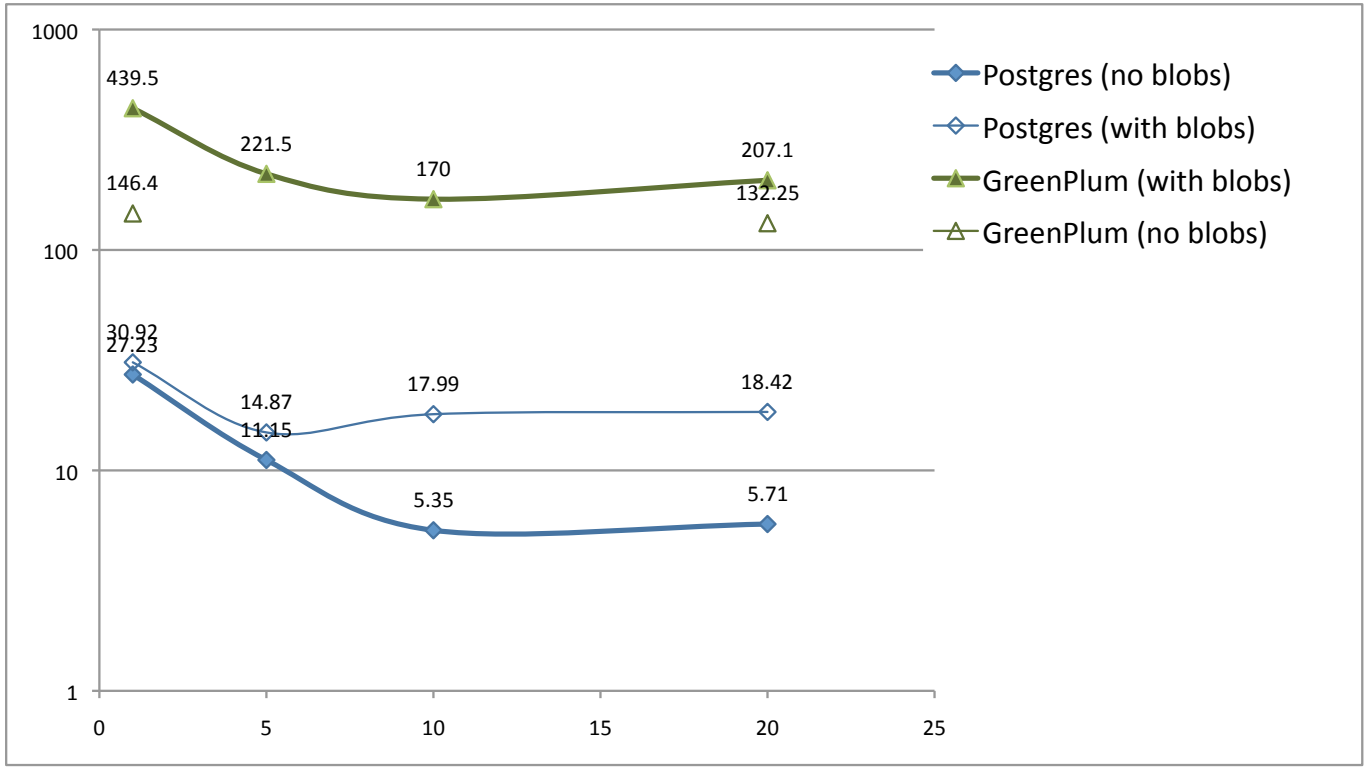
Galaxy Coordinate Reconstruction

Time to process 10,000 detections in seconds

N Processes	Postgres, no blobs	Posrgres, blobs	GreenPlum, no blobs	GreenPlum, blobs
1	27.23	30.92	146.4	439.5
5	11.15	14.87		221.5
10	5.35	17.99		170.0
20	5.71	18.42	132.25	207.1

Galaxy Coordinate Reconstruction

Reconstruction time in seconds per 10,000 detections vs. number of concurrent processes



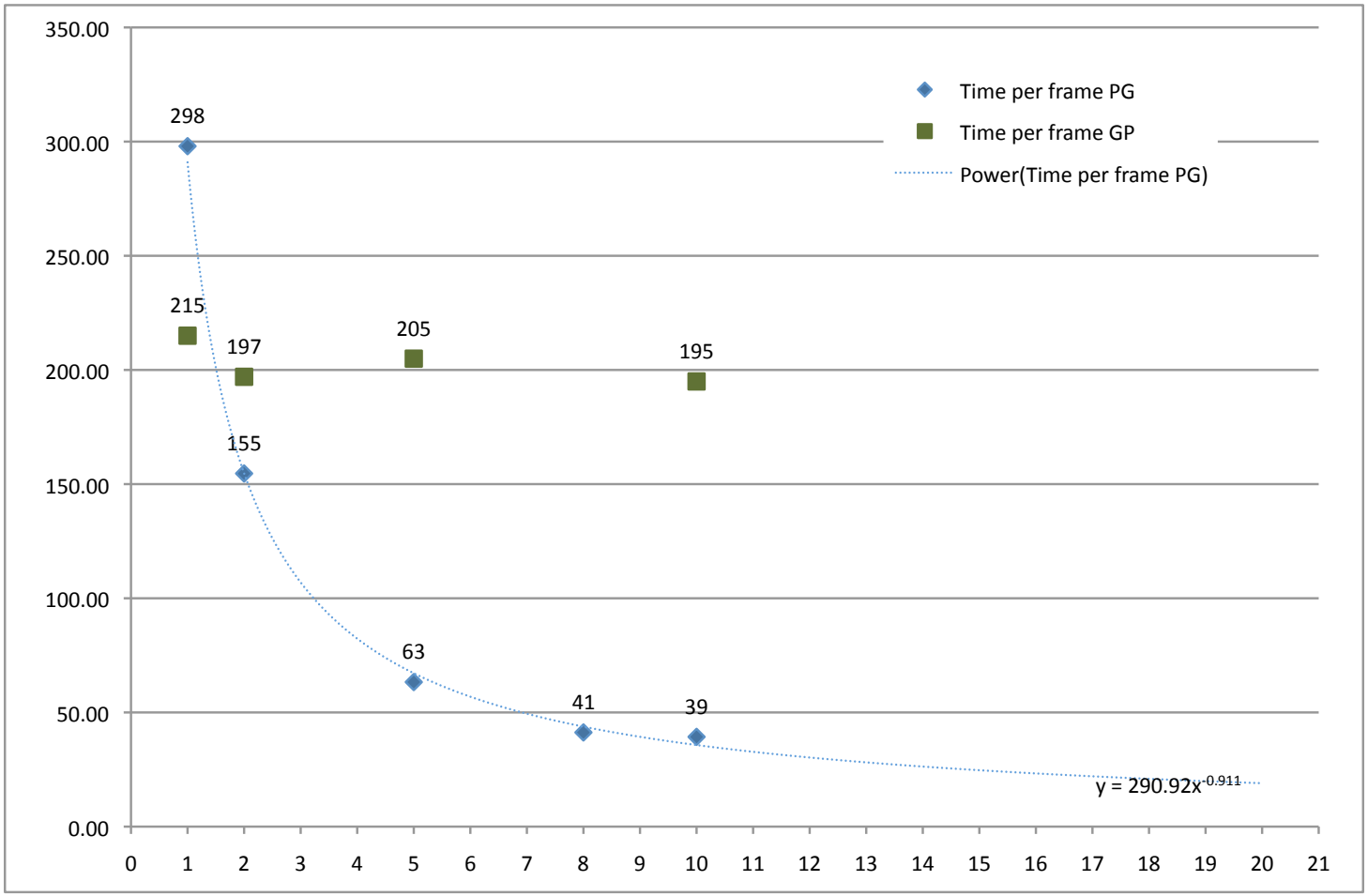
Catalog Matching

Time to match all objects found in one frame (~80K objects) vs. number of concurrent processes, in seconds

	1 process	2	5	8	10	20
Postgres	298	155	63	39	41	55
Greenplum	215	197	205		195	

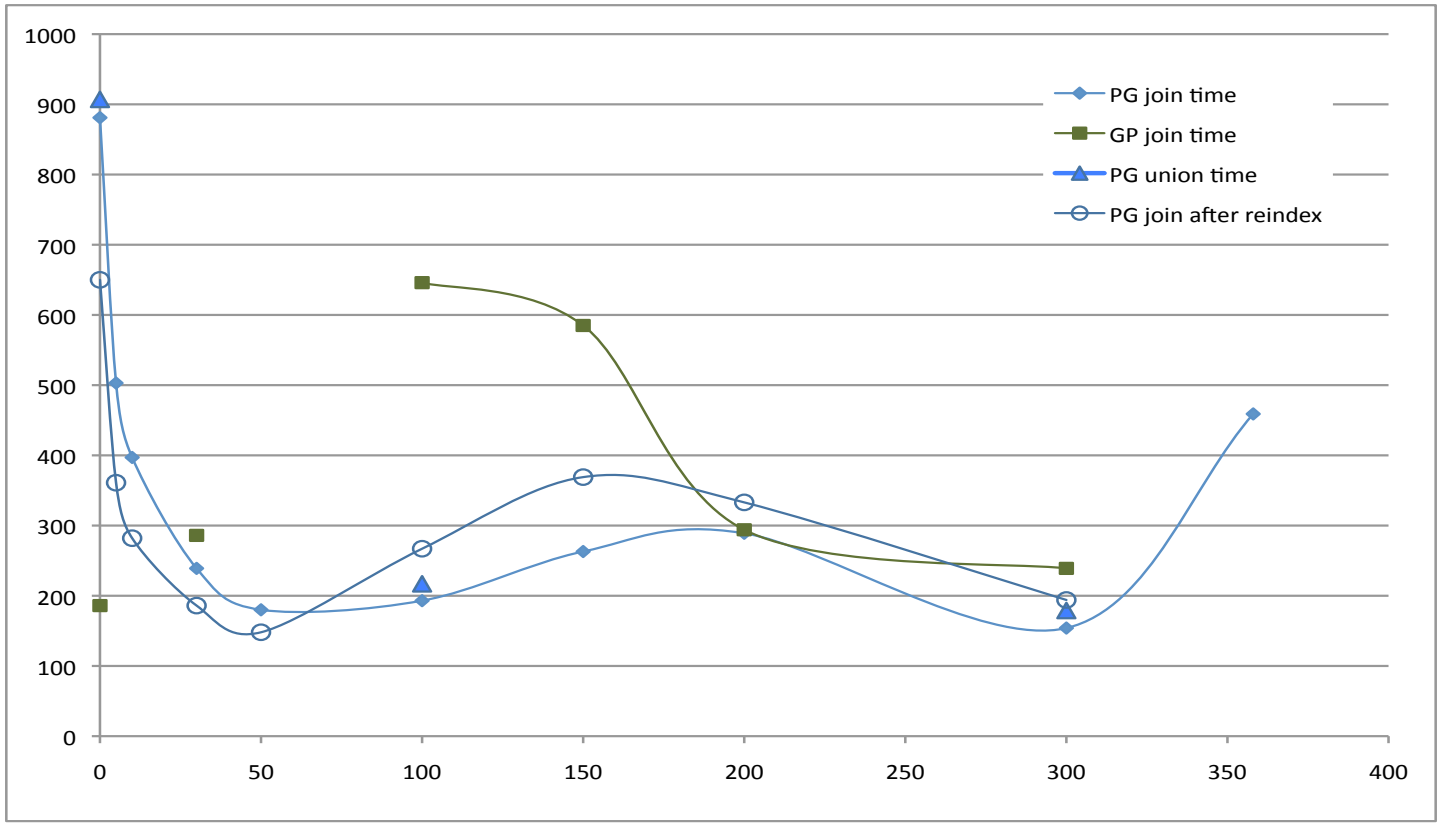
Target is < 350 seconds (246 exposures per day)

Catalog Matching



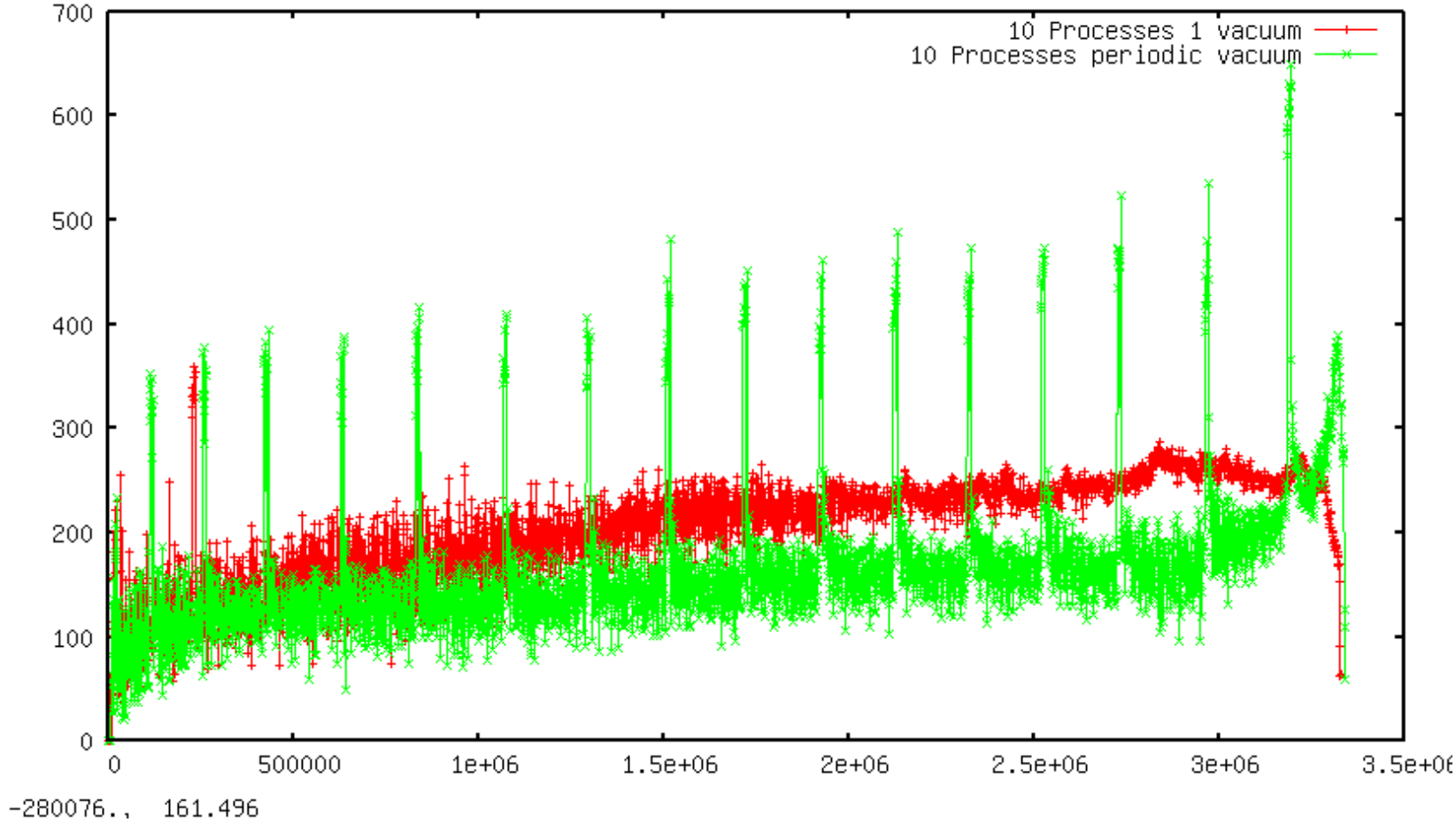
Observations: index performance

Catalog matching algorithm performance as a function of frame number
Frame is a rectangle (0.5*frame,0) - (0.5*(frame+1),1) degrees



Observations: vaccuming

- Postgres (and GreenPlum) do need vacuuming



Observations: single-node GreenPlum vs. Postgres

- Postgres:
  - 1 user process creates 1 Postgres process
- GreenPlum:
  - 1 user process creates 9 GreenPlum processes
    - 1 for the client-master connection
    - 8 for master-segment connection
- In single-node configuration, GreenPlum may not scale as easily

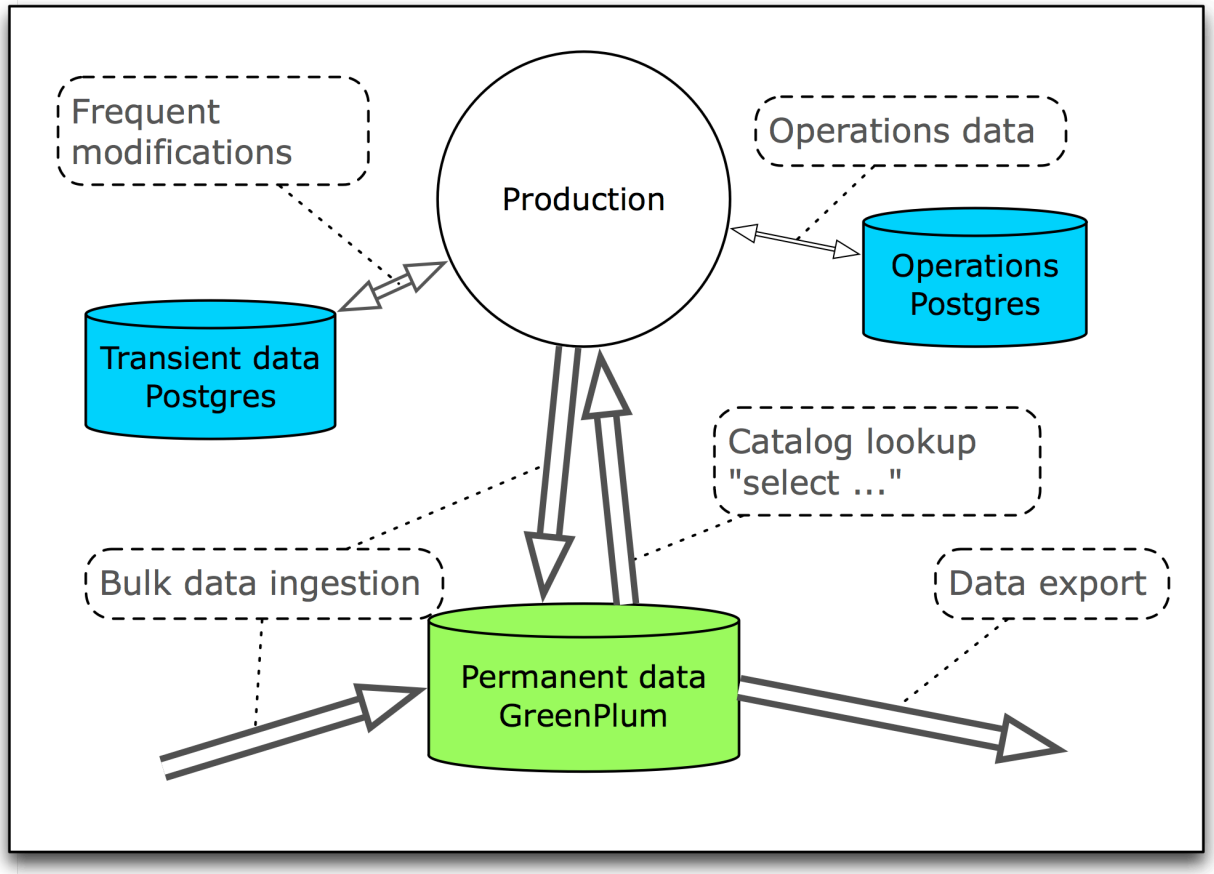
Observations: blobs/arrays

- Blobs/arrays are fine, but *if possible*, split them off into separate, less frequently updated table – *if possible*

GreenPlum vs. Postgres

Feature	Winner
Individual row operations (inserts, updates)	Postgres
Bulk data ingestion	GreenPlum
Utilities (vaccum, reindexing, etc.)	GreenPlum
Lookup	GreenPlum
Concurrency	Postgres
Scalability	?

JDEM/SOC Database Architecture



Next Steps

- Bulk Slitless Spectroscopy algorithm

  https://cdcvs.fnal.gov/redmine/wiki/jdemsoc-private/Mock_Slitless_Spectroscopy_Algorithm

  - Bulk lookup
  - Update of spectra arrays
  - File (CSV ,FITS) / DB interface
  - Compare GreenPlum and Postgres
  - Measure scalability
- Consult with GreenPlum
- Data backup/recovery procedures
- Update database architecture recommendations