ZTF Science Data System: status update

Frank Masci & the IPAC-ZTF Team

Board Meeting, April 7, 2016

Accomplishments (since Jan 2016)

- Software Definitions, Deliverables & Services document written and distributed to collaboration.
  - ➤ Much feedback received on pipeline flow plans presented therein
  - ➤ Some services envisaged to be a challenge and need discussion; more later

- 2016 PTF/iPTF lightcurve data release preparation (MSIP deliverable):
  - ➤ Lightcurves and metadata spanning 03/01/2009 – 01/28/2015 generated; many issues resolved
  - ➤ Delivery will contain ~600 million lightcurves (transients and variables); 30 billion detections
  - ➤ Loading of the above lightcurves and metadata into database is in progress

- Data volumes, source counts, DB sizes scoped for 3 year survey: used for hardware planning
- Recent hardware orders:
  - ➤ Two Oracle servers for pipeline databases with large RAM and solid-state disks;
  - ➤ Six pipeline fileservers: 4 to hold temporary sandbox/pipeline content; 2 for archive; storage arrays; fiber channel switches for fileservers; 10Gbit network switches with copper ports

- Raw camera image packaging and metadata specs in review; to be finalized end of this week
- Raw image compression study; analysis document written. Goal: need to lossy-compress raw-image data at telescope by 3x to account for bandwidth limitations; can preserve 4 noise-bits per pixel, leading to ~ 0.3% increase in pixel noise across different environments (used PTF data).

Accomplishments contd.

- ZTF images simulated according to recent camera specs: to assist with pipeline development

- Quality Assurance metrics across ZTF processing system identified; used to design DB schema

- Design and development of new ZTF Transients Database Schema (ongoing)

- Build and deployment of PTF codebase on new ZTF 32-node cluster to exercise new pipeline job manager, filesystems, and databases in new environment; testing in progress

- Strategy for testing data latency / bandwidth from summit to IPAC devised; in progress

- End-to-end data flow model design / data partitioning / system engineering plan in progress given better knowledge of deliverables and hardware requirements
  - ➢ see schematic on next page (from ZSDS Data Flow & Processing document)

Data Flow / Processing Model (preview)
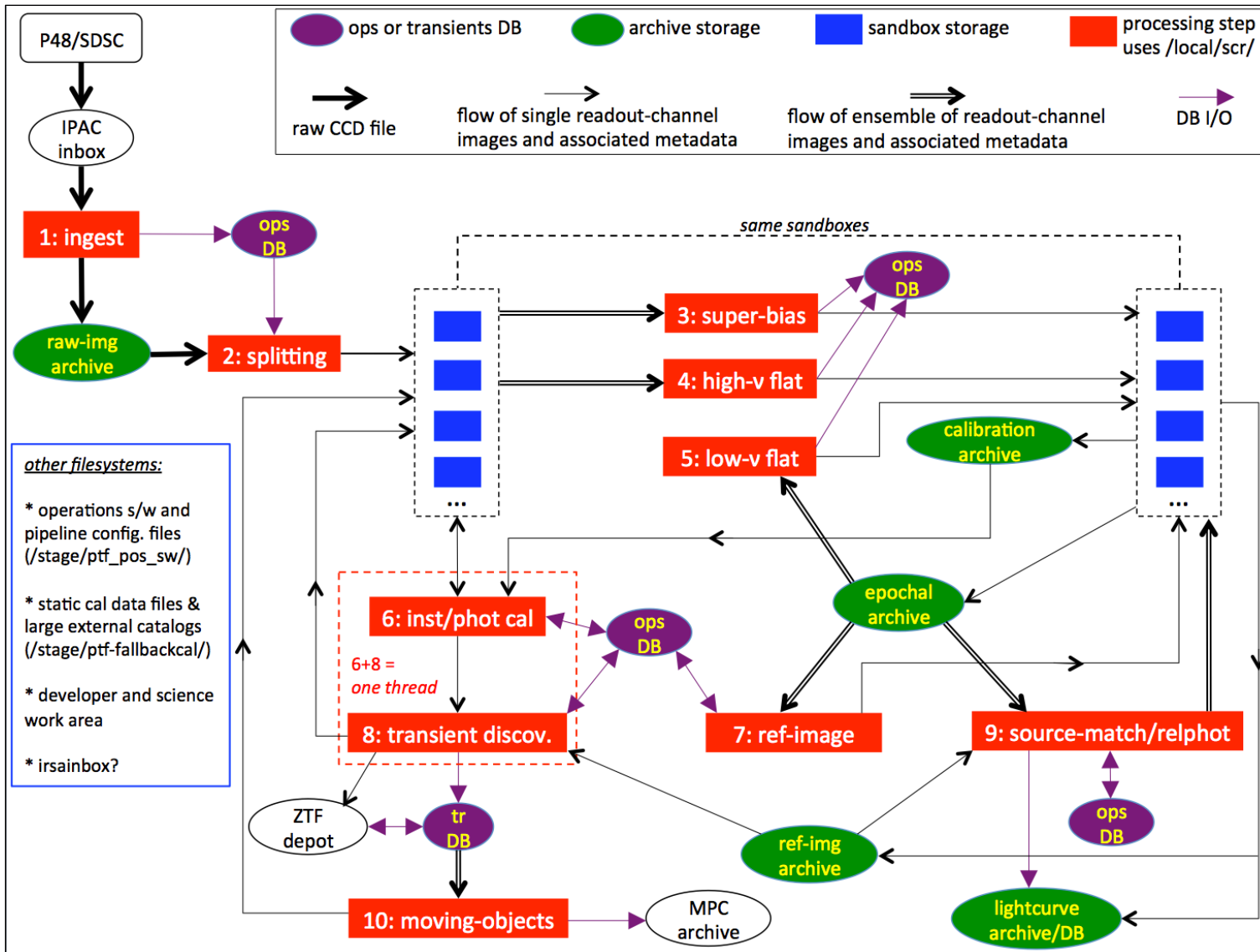
Near-term schedule (expected completion dates)

- **April 20, 2016:** complete testing / configuration of new ZTF cluster environment; w/ filesystems
- **April 30, 3016:** new ingest / camera-splitting pipeline interfacing with new operations DB
- **April 30, 2016:** prototype instrumental calibration pipeline (with basic astrom./phot. calibration)
- **May 15, 2016:** complete initial data-latency / bandwidth study from P48 to IPAC; identify needs
- **May 20, 2016:** prototype image-subtraction transient-discovery pipeline
- **May 31, 2016:** astrometric (& distortion) calibration infrastructure and software module in place
- **June 15, 2016:** prototype reference-image (co-addition) pipeline
- **June 15, 2016:** prototype source-matching pipeline to support lightcurve generation
- **August 15, 2016:** user-interface for PTF 2016 lightcurve release; simple version is planned for iPTF Summer School in July.

Issues / discussion points

- From sizing model estimates for ZTF products and services, hardware budget cannot accommodate lightcurve database based on <u>single-epoch extractions</u>.
  - ➤ plan is to use PSF-fit extracted sources with no filtering (get ~ 920 million >5σ extractions per night => ~ 700 billion row database after 3 years)
  - ➤ will be difficult to load a single night's data fast enough; will also be difficult to query
  - ➤ can support ~ 300 million extractions per night
  - ➤ **Note:** above does not affect processing of realtime transient stream and archiving
- Also, baseline budget cannot support storage of difference-image products and related metadata (6TB/night => ~ 4.5PB/3years): valuable for archival research later, e.g., forced-photometry

- Flat-fielding plan: whether low-spatial frequency responsivity correction maps are needed to achieve best photometric precision in a relative sense; currently a placeholder in ZTF pipeline

- External catalogs to support astrometric and photometric calibration in pipeline:
  - ➤ Gaia first release is expected in September 2016
  - ➤ Pan-STARRS1 images expected in August 2016 timeframe; catalogs expected end of year(?)
  - ➤ **Note:** we can still proceed with pipeline development and testing using existing catalogs
  - ➤ slight concern is PS1 availability, primarily to allow enough time for R&D on catalog filtering for optimal ZTF calibration