

# LSST Data Management Brief Status Update

**Mario Juric**

*University of Washington*

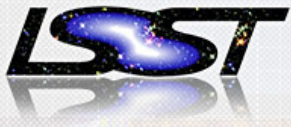
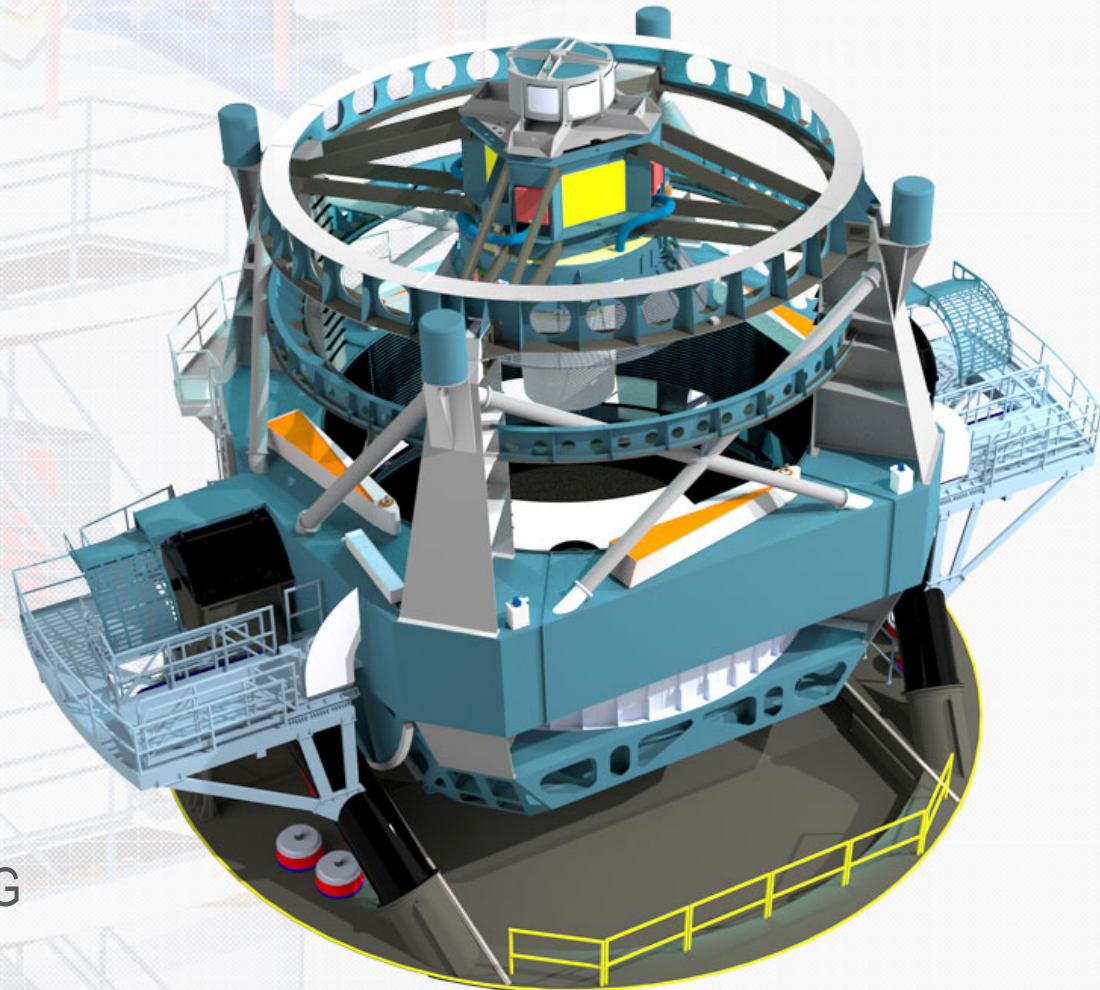
*LSST Data Mgmt. Subsystem Lead*

**Robert Lupton**

*Princeton University*

*LSST DM Science Pipelines Lead*

*with John Swinbank, Simon Krughoff,  
Jacek Becla, K-T Lim, Jim Bosch  
and the LSST DM Team*



LSST DESC MEETING

March 6, 2016



- Jim Bosch gave an excellent summary <http://ls.st/4o4>; this talk is an update with some highlights since last October.

# Data Management Status

Jim Bosch  
Princeton University



- Status in ~October 2015:
  - We were porting over science pipelines changes from HSC
  - We were improving the long-term development plan
  - We were in the midst hiring



# A year of growth

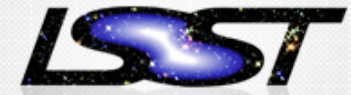


- Staff has quadrupled since start of MREFC, will be at **~80 people** by ~October this year (including contractors and fractional staff)
- Some are fractional; **~54 FTE hired/assigned**

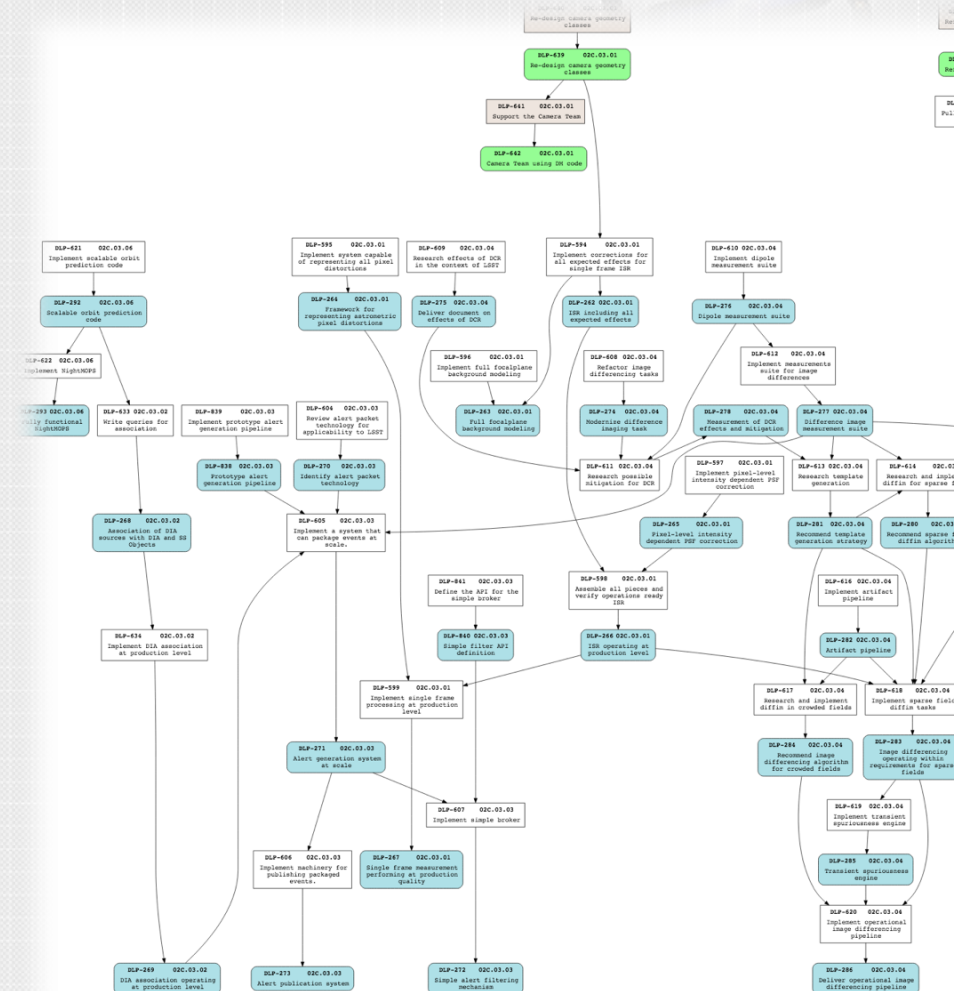




# Status of the development plan



- <https://jira.lsstcorp.org/projects/DLP>
- A long-term plan for science pipeline development and related development.
- A significant update since construction start, but more work still remains
- Next refresh will come later this year



Plan visualizations: <http://ls.st/zmf> <http://ls.st/ms1>



# DM Leadership Changes



- We've streamlined the leadership and structure of the DM subsystem.
- **Used to be three-in-a-box**
  - DM Project Manager: Jeff Kantor
  - DM Project Engineer: K-T Lim
  - DM Project Scientist: Mario Juric
- **Leadership structure reorganized**
  - Subsystem Lead: **Mario Juric** (interim)
    - Project Manager: **Jacek Becla** (interim)
    - Project Scientist: **Zeljko Ivezić** (interim)
    - System Architect: **K-T Lim**
  - **Andy Connolly** named interim Level 1 processing science lead
    - Search underway for permanent lead
- Announced two weeks ago at the LSST Joint Technical Meeting. Transition still in process (not all documents & websites updated to reflect the new structure).
- We will be conducting a world-wide search for the Project Manager position





- Established working groups to make a push to the next level of DM architecture fidelity. Two of special interest to DESC:
  - **Science Pipelines Definition Workgroup (scipi-wg)**
    - Co-Chairs: Zeljko Ivezic & Robert Lupton
    - Refine and improve the high-level architecture of the DM Science Pipelines to support further refinement into detailed design, the creation and execution of well planned and resourced construction activities, as well as provide science pipeline related inputs necessary for the work of dps-wg. This includes **enumerating all pipelines required to deliver products in the DPDD, defining overall processing flows, and specifying algorithms (as best understood at this time)**. For the purposes of this WG, science pipelines include SDQA and Calibration.
  - **Data Processing System Architecture Workgroup (dps-wg)**
    - Co-Chairs: K-T Lim and Don Petravick
    - **Refine and improve the overall architecture of the data processing system and its coarse-grain concept of operations**, to support further refinement into detailed design, the creation and execution of adequately planned and resourced construction activities, as well as project-wide operations planning.

<https://confluence.lsstcorp.org/display/DM/DMLT+Working+Groups>



# Significant New Science Pipeline Features



- ‘Multi-band’ coadd processing.
  - Provides consistent outputs across filters.
- ‘Safe’ clipping in coadd construction.
  - Much more robust and reliable rejection of outliers.
- Prototype ‘Brighter-Fatter’ correction.
- System for removing false detections around bright objects.
- Integrated external PSFEx package for state-of-the-art PSF estimation.
- Joint astrometry/photometry calibration (impl. by Pierre Astier’s group)
- ctrl\_pool task parallelization framework
  - High level solution for MPI-based task distribution across a cluster.
  - Not the long-term solution to LSST’s needs, but good enough for now.
- Ability to inject ‘fake’ sources into processing.
- Rework of ProcessCcdTask

*IMPORTANT: All of these are currently on master, and will appear in the the next release (~3 months). Intermediate builds may be available sooner.*

*More from Robert in just a minute!*



# DMTN-006: False Positive Rates in the LSST Image Differencing Pipeline

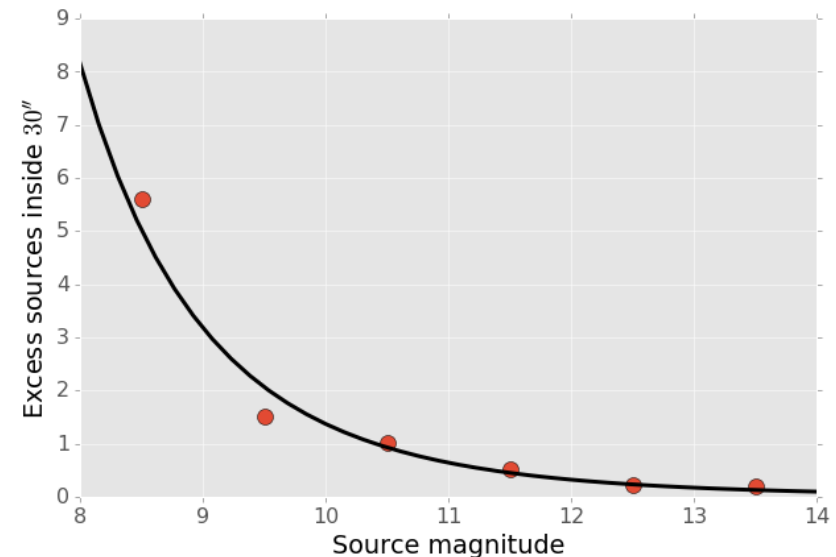
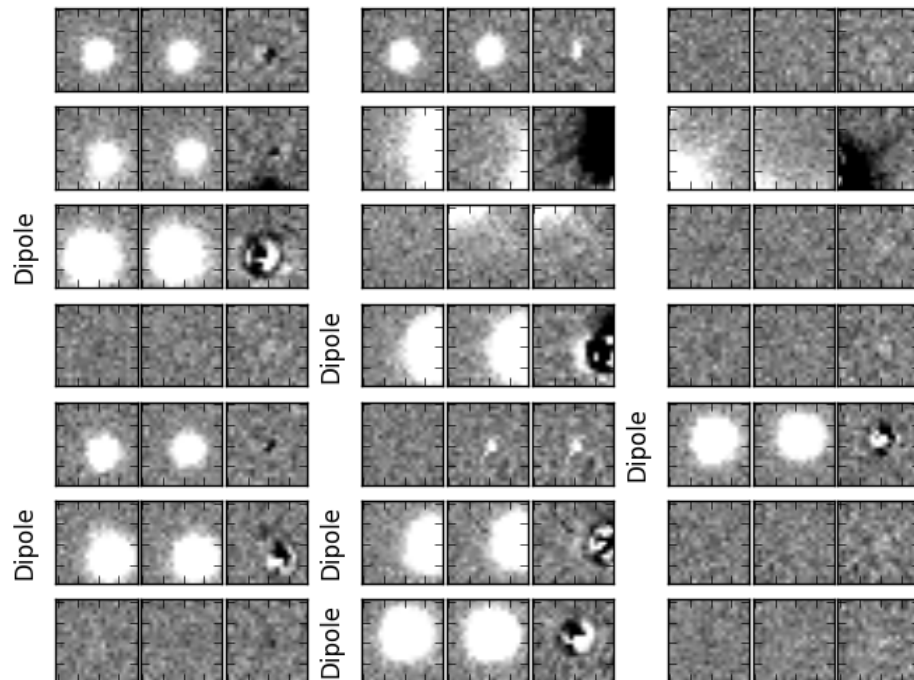


Colin Slater, Mario Jurić, Željko Ivezić and Lynne Jones

Latest Revision: [2016-03-04](#)

Significant work done on understanding of current image differencing pipeline performance. With DECam data, we're capable of producing a clean sample of difference image detections at roughly the 200-400 per square degree level (early results)

- Raw counts, before real-bogus filtering (only a simple dipole detector applied)
- A significant number of these is likely due to real variability + asteroids
- This is image-to-image (not image-to-template) differencing



Left: A sample of (false) detections  
Top: Detection excess near bright sources



# Documentation: DM Technical Notes



**Listing of available DM Technotes**

**Data Management Technote**

**jsick** · Jonathan Sick · LSST Data Management · 5 Jan 21 · [Announcing the DM Technical Note Series](#)

In November 2015 we [started a new technote series](#) in Data Management. Although the platform is still an MVP, we're seeing great early adoption.

One of the next steps for the platform is a highly usable index/landing page for technotes and other DM documents. I have a [ticket](#) to do this. In the mean time, this post will be updated to list all available technotes. I've pinned the post to the DM category; **if you wish you can personally un-pin it by clicking on the pin icon.**

**Data Management Technotes (DMTN)**

- [DMTN-001: Porting the stack to OS X El Capitan](#) by Tim Jenness [[GitHub](#)] [[Zenodo](#)]
- [DMTN-002: SuperTask and Activator Notes](#) by Mattias Carrasco Kind [[GitHub](#)]
- [DMTN-003: Description of v1.0 of the Alert Production Simulator](#) by Stephen Pietrowicz [[GitHub](#)] [[Zenodo](#)]
- [DMTN-004: Debugging in Docker Containers](#) by Vaikunth Thukral [[GitHub](#)]
- [DMTN-005: Current LSST stack WCS usage](#) by John Parejko [[GitHub](#)]
- [DMTN-006: False Positive Rates in the LSST Image Differencing Pipeline](#) by Colin Slater [[GitHub](#)]
- [DMTN-007: Dipole characterization for image differencing](#) by David Reiss [[GitHub](#)]
- [DMTN-008: Introducing validate\\_drp: Validate an output repository against SRD Key Performance Metrics](#) by Michael Wood-Vasey [[GitHub](#)]
- [DMTN-009: Vertical-partition Join Performance in MySQL](#) by Vaikunth Thukral [[GitHub](#)]

All DMTN repos, including drafts: <https://github.com/lsst-dm?utf8=✓&query=dmtn>

**SQuaRE Technotes (SQR)**

- [SQR-000: The LSST DM Technical Note Publishing Platform](#) by Jonathan Sick [[GitHub](#)]
- [SQR-001: Git LFS Architecture Note](#) by Frossie Economou, J Matt Peterson and Joshua Hoblitt [[GitHub](#)] [[Zenodo](#)]
- [SQR-002: Binary Science Pipeline Software Distribution](#) by Joshua Hoblitt [[GitHub](#)]
- [SQR-003: Reserved](#)
- [SQR-004: Reserved](#)
- [SQR-005: Publication Board JIRA Project - User Note](#) by Frossie Economou and Željko Ivezić [[GitHub](#)]
- [SQR-006: Documentation Deployment Service for LSST's Eups-based Software](#) by Jonathan Sick [[GitHub](#)]
- [SQR-007: Reserved](#)
- [SQR-008: SQuaRE QA Database](#) by Angelo Fausti [[GitHub](#)]
- [SQR-009: SQUASH Dashboard Prototype](#) by Angelo Fausti [[GitHub](#)]

1 / 2

A system to make it easy to publish, maintain, and cross-reference technical notes (credit: Jonathan Sick and the SQuaRE team).

Current list of tech notes is on [community.lsst.org](https://community.lsst.org) (link below).

Plan to have it indexed by ADS.

Same mechanism may be usable for DESC purposes.

<https://community.lsst.org/t/listing-of-available-dm-technotes/496>



community.lsst.org

all categories ▾ all tags ▾ Latest Categories Bookmarks Unread (14) My Posts + New Topic

Topic	Category	Users	Replies	Views	Activity
Welcome to community.lsst.org community.lsst.org is a place for the astronomy community to discuss the Large Synoptic Survey Telescope's ongoing development and get help with using LSST's software today. What's here Community members can read, pos... <a href="#">read more</a>	Meta		2	341	Aug '15
Proposal for Butler DM-4168 "Data repository selection based on version" <a href="#">new</a> butler	Data Management		2	23	1m
How do we update EUPS? <a href="#">i</a> eups	Support		17	163	10m
Pipeline execution in HSC execution	Data Management		9	120	8h
What to include in DM status presentation to DESC next week?	Data Management		6	65	10h
Given a config field name as a string, how to get the value? <a href="#">new</a> dm-dev	Data Management		2	34	20h
Recent Butler changes: Repository class and 'cfg' configuration <a href="#">new</a> butler	Data Management		0	13	21h
Poll: Timeslot for current-infrastructure meeting <a href="#">new</a> dm-infrastructure	Data Management		2	34	22h
4 March meeting <a href="#">new</a>	Camera-DM Visualization		0	27	1d
Notes from DRP High-Level Data Flow Discussion <a href="#">new</a> planning	Data Management		5	66	4d
User interface design	Camera-DM Visualization		3	69	4d

Display a menu

A forum for both internal and external communication and interaction.  
Note: More formal mechanisms for Project-Collaboration interactions are also being put in place.



# User and Developer Guides cleaned up, updated



**Credit:** Jonathan Sick and the SQuaRE team.

The image displays two web browser windows. The left window shows the 'LSST Software User Guide' on the Confluence site, featuring a sidebar with navigation links like 'Pages', 'Blog', 'Questions', and 'SPACE SHORTCUTS'. The main content area includes 'About this document', 'Document Downloads', and a 'Release History' table.

Tag	Version	Date	Notes
v11_0	v11.0	2015-October-02	Summer2015 release <ul style="list-style-type: none"><li>Release Notes,</li><li>Characterisation Measurements,</li><li>Known Issues,</li><li>ADASS XXV poster paper describing release</li></ul>
v10_1	v10.1	2015-May-12	Winter2015 release <ul style="list-style-type: none"><li>Release Notes.</li></ul>

The right window shows the 'LSST DM Developer Guide' on the developer.lsst.io site, featuring a sidebar with navigation links like 'GETTING STARTED', 'PROCESSES', and 'CODING GUIDES'. The main content area includes 'About this document', 'Document Downloads', and a 'Release History' table.

A significantly improved software user guide (<http://ls.st/ug>) and the developer guide (<http://developer.lsst.io>)

These will be interesting for DESC to mine for standards, best practices, etc.

## Upcoming work (next ~6-9 months)



- We're inserting a short ~three-month development cycle
  - The "X16" (for "eXtra") cycle
  - Having releases coincide with project meetings didn't work well
  - Next cycle will be "Fall 2016"
- Expect to do significant work on the overall system architecture (working group process).
- **Upcoming science pipelines work (some highlights):**
  - MultiFit (initial version) and deblender improvements
  - Astrometry
  - QA Pipelines
- **Architectural and middleware work (some highlights):**
  - Looking at better integration with AstroPy
  - Work on data access components (see next slide)
- **More work on documentation and usability**

*Tracking X16 planning and work:*  
<https://jira.lsstcorp.org/issues/?filter=14115>





The screenshot shows a web browser window displaying a ticket page for 'LDM-463: Data Access Design'. The browser's address bar shows the URL 'ldm-463.lsst.io/en/tickets-dm-5091/#'. The page has a blue header with the title 'LDM-463: Data Access Design' and the ticket ID 'tickets-DM-5091'. A dark sidebar on the left contains a table of contents with three items: '1 Introduction', '2 Butler', and '3 Change Record'. The main content area has a large heading 'LDM-463: Data Access Design', followed by the author 'Jacek Becla' and the latest revision date 'February 5, 2016'. The content is organized into sections: '1 Introduction', '2 Butler', and '2.1 Overview - What is the Data Butler?'. The '2.1 Overview' section contains two paragraphs of text. At the bottom of the sidebar, there is a link 'Read the Docs' and a dropdown menu showing 'v: tickets-DM-5091'.

LDM-463: Data Access Design  
tickets-DM-5091

1 Introduction  
2 Butler  
3 Change Record

Read the Docs v: tickets-DM-5091

## LDM-463: Data Access Design

Jacek Becla

Latest Revision: February 5, 2016

### 1 Introduction

### 2 Butler

#### 2.1 Overview - What is the Data Butler?

The Butler is a framework for generic I/O and data management. It isolates application code from the underlying data-access implementation in terms of storage formats, physical locations, data staging, database mapping, etc. Butler is configured by a Policy that provides configuration details, as well as by parameters provided at initialization time.

This general concept makes the Data Butler potentially able to serve as a data *router* (or hub or switch). Data can be published and sent to multiple locations based on the Butler configuration. Those locations may include “persistent storages” that are actually dynamic (like displays or subscription streams) rather than truly persistent storage.

<http://ldm-463.lsst.io/en/tickets-dm-5091/#>

# Usability: universal binary installs with conda



```

mistune: 0.7.1-py27_0
nbconvert: 4.0.0-py27_0
nbformat: 4.0.1-py27_0
notebook: 4.0.6-py27_0
numpy: 1.9.0-py27_0
openblas: 0.2.14-3
pcr: 8.31-0
pixman: 0.26.2-0
ptyprocess: 0.5-py27_0
py2cairo: 1.10.0-py27_2
pygments: 2.0.2-py27_0
pyparsing: 2.0.3-py27_0
pyqt: 4.11.4-py27_0
python-dateutil: 2.4.2-py27_0
pytz: 2015.7-py27_0
pyzmq: 14.7.0-py27_1
qt: 4.8.7-1
scipy: 0.16.0-np19py27_1
scons: 2.3.0-py27_0
setuptools-git: 1.1-0
sip: 4.16.9-py27_0
six: 1.10.0-py27_0
sqlalchemy: 1.0.9-py27_0
ssl_match_hostname: 3.4.0.2-py27_0
stsci.distutils: 0.3.7-0
swig: 3.0.2-0
terminado: 0.5-py27_1
tornado: 4.2.1-py27_1
zeromq: 4.1.3-0

Proceed ([y]/n)? y

Fetching packages ...
libgcc-4.8.5-1 100% |#####| Time: 0:00:00 3.92 MB/s
libgfortran-1. 100% |#####|
libpng-1.6.17- 100% |#####|
libsodium-1.0. 100% |#####|
pcr-8.31-0.ta 100% |#####|
pixman-0.26.2- 100% |#####|
freetype-2.5.5 100% |#####|
libxml2-2.9.2- 100% |#####|
openblas-0.2.1 100% |#####|
swig-3.0.2-0.t 100% |#####|
zeromq-4.1.3-0 100% |#####|
python-0.23.4- 100% |#####|
 cups-1.5.9_6-p 100% |#####|
fontconfig-2.1 100% |#####|
sconschema-2.4 100% |#####|
markupsafe-0.2 100% |#####|
mistune-0.7.1- 100% |#####|
```

We're working to make it easier to install the LSST stack, and provide binary distributions:

- Internally already building OpenStack images for our Nebula cluster.
- Containers, native packages, CERN-VMFS builds
- **Working on conda packaging of the LSST stack.**

## Installing and using

No root privileges are required to install or use this code.

If you're using [Anaconda](#) (or [Miniconda](#)) Python distribution, you should be able to install the LSST stack by doing the following:

```
conda update conda # to ensure you're running latest conda

conda config --add channels http://research.majuric.org/conda/stable

conda install lsst-distrib
```

An experimental conda build of Summer 2015 is at <http://ls.st/k2i>. More recent builds are also available (ask on [community.lsst.org](http://community.lsst.org)).





- We've staffed up (with a bit more to go)
- Visible progress in a number of areas, especially in science pipelines and documentation. Much more underneath the surface.
- Key challenge for the next year: keeping the development coordinated (both internally and with the community)
- A push to strengthen the architecture and project management
- Continue the conversation on <http://community.lsst.org>

# Questions?

