

# LSST



Un défi et une opportunité  
pour le CC-IN2P3

Dominique Boutigny

Journées Prospectives 23 – 24 mai 2013

23 - 24 mai 2013

1



# Large Synoptic Survey Telescope

LSST est un télescope automatique capable de tout voir (ou presque) dans son champ de vue (20 000 degrés<sup>2</sup>)

- Phénomènes transitoires
  - ✓ Supernovae
  - ✓ Astéroïdes géocroiseurs
  - ✓ ...
- Objets extrêmement lointains
  - ✓ Galaxies
  - ✓ Quasars

- Grandes structures de l'univers
  - ✓ Amas de galaxie
  - ✓ Super-amas
  - ✓ Distribution des galaxies dans l'univers visible
- Phénomènes cosmologiques
  - ✓ Matière noire
  - ✓ Énergie noire
  - ✓ ...

Capacité à observer les objets très lointains (très peu lumineux)

Capacité à mesurer le décalage vers le rouge des objets observés (mesure de distance)



# Supernovæ de Type 1A

Accrétion de matière dans un système binaire :

naine blanche + étoile

→ Masse critique

→ Explosion cataclysmique

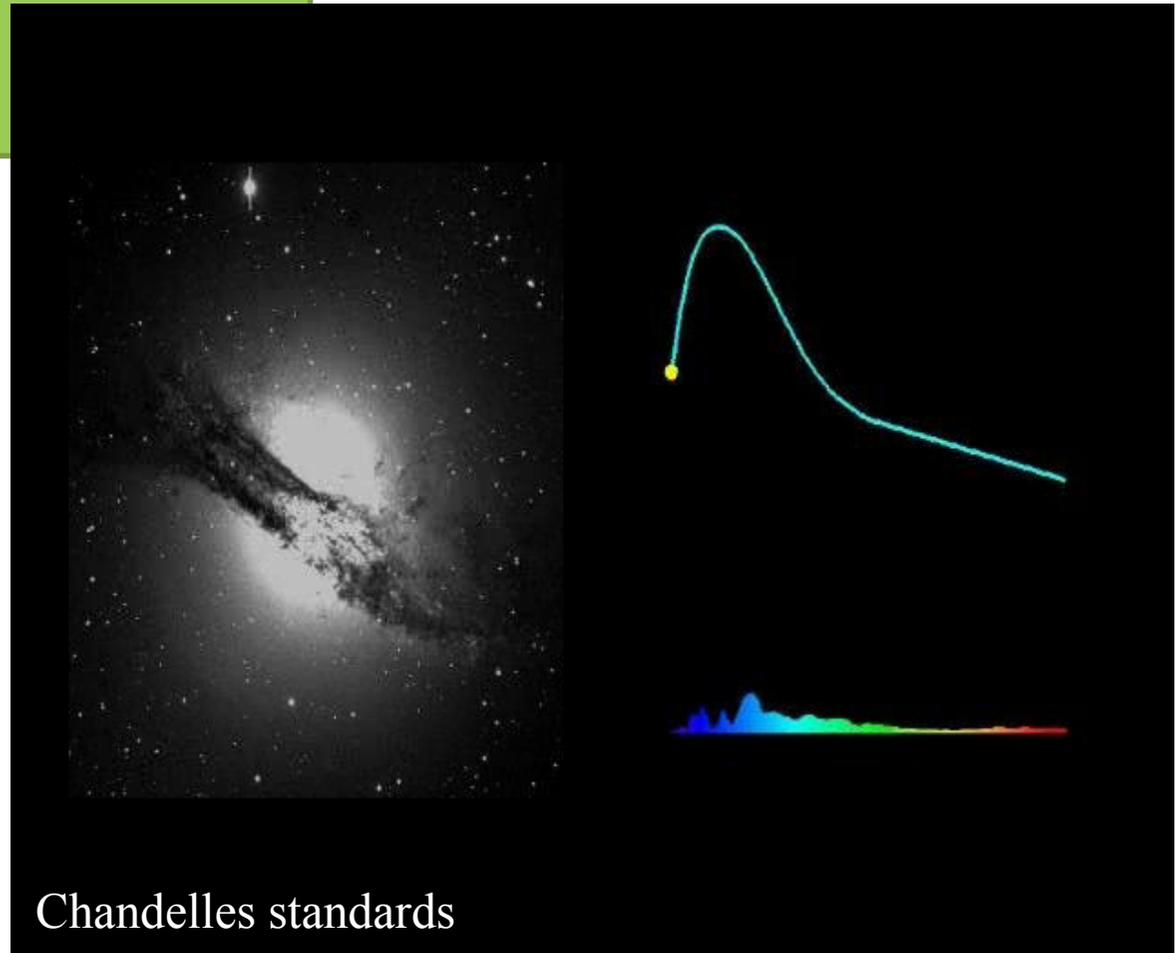
Courbe de lumière  
caractéristique

Phénomène supposé connu et  
suffisamment compris

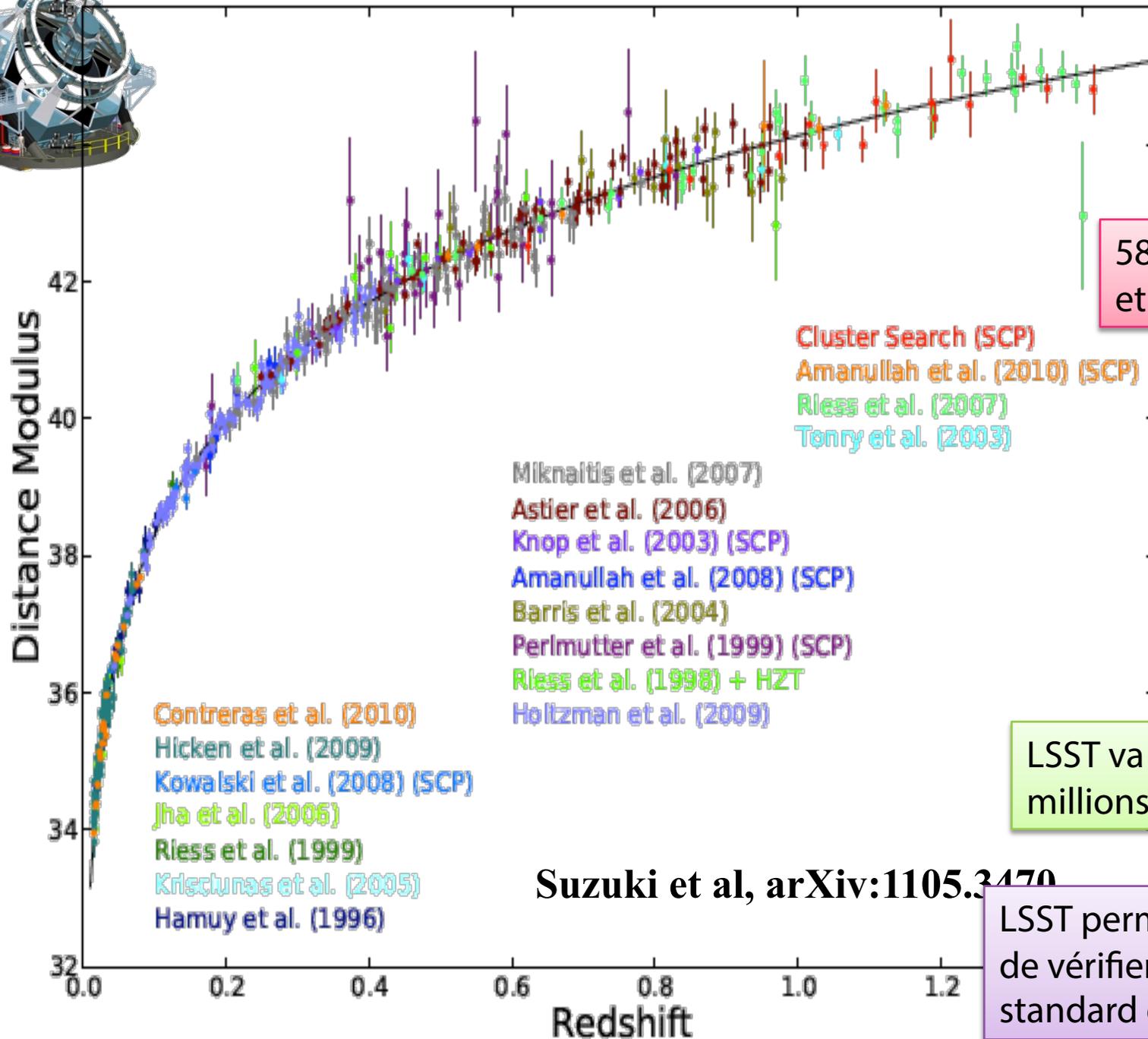
→ Connaissance de la luminosité  
absolue de la Supernovæ

Comparaison entre le décalage  
vers le rouge ( $z$ ) et la luminosité  
absolue

→ Information sur l'évolution  
de l'univers



Chandelles standards



580 SN1A mesurées et utilisables

Cluster Search (SCP)  
Amanullah et al. (2010) (SCP)  
Riess et al. (2007)  
Tonry et al. (2003)

Miknaitis et al. (2007)  
Astier et al. (2006)  
Knop et al. (2003) (SCP)  
Amanullah et al. (2008) (SCP)  
Barris et al. (2004)  
Perlmutter et al. (1999) (SCP)  
Riess et al. (1998) + HZT  
Holtzman et al. (2009)

Contreras et al. (2010)  
Hicken et al. (2009)  
Kowalski et al. (2008) (SCP)  
Jha et al. (2006)  
Riess et al. (1999)  
Kristunas et al. (2005)  
Hamuy et al. (1996)

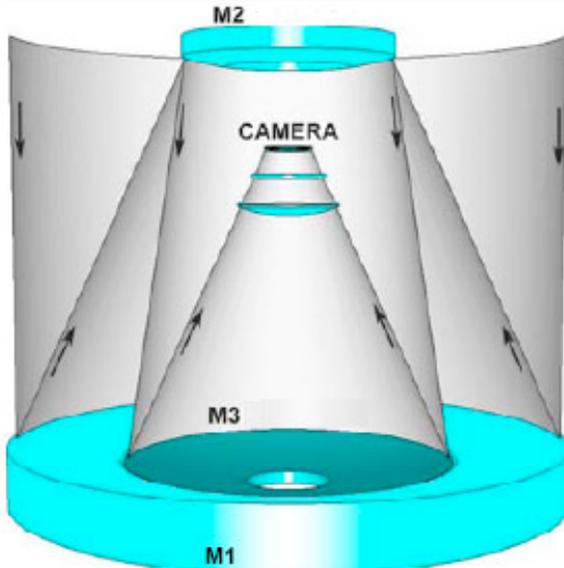
Suzuki et al, arXiv:1105.3470

LSST va en détecter 10 millions

LSST permettra également de vérifier le caractère standard ou non des SN1A



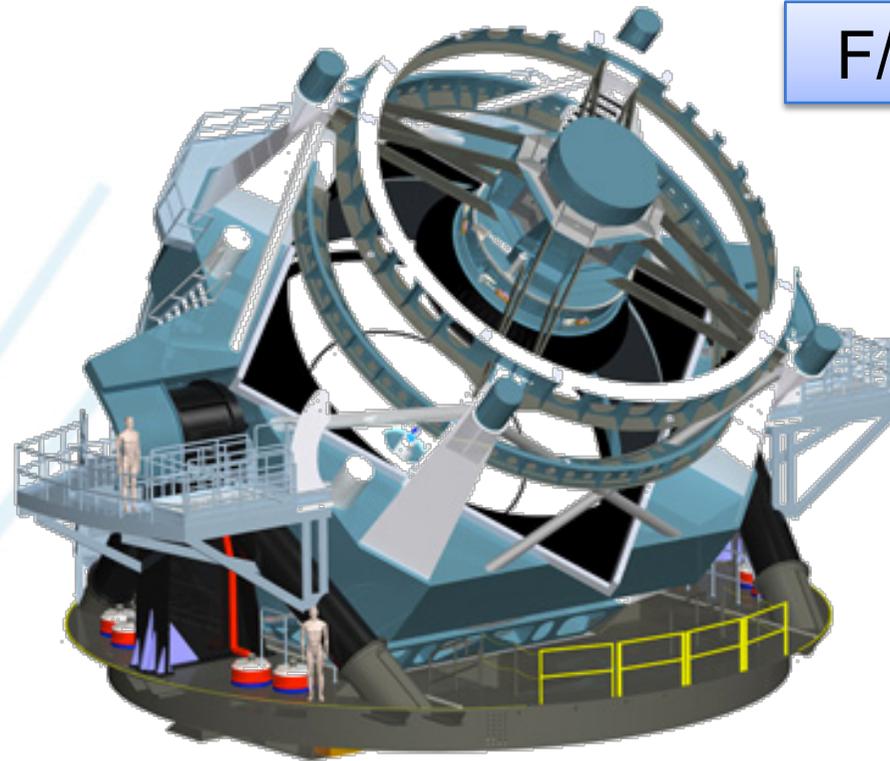
# Le télescope



Diamètre 8.4 m  
Formule optique à 3 miroirs "Paul Baker"

$F/D = 1.23$

M1 : 8.4 m f/1.14  
M2 : 3.4 m f/1.05  
M3 : 3.5 m f/1.19



Structure mobile de 350 tonnes (dont 60 tonnes pour le système optique seul)

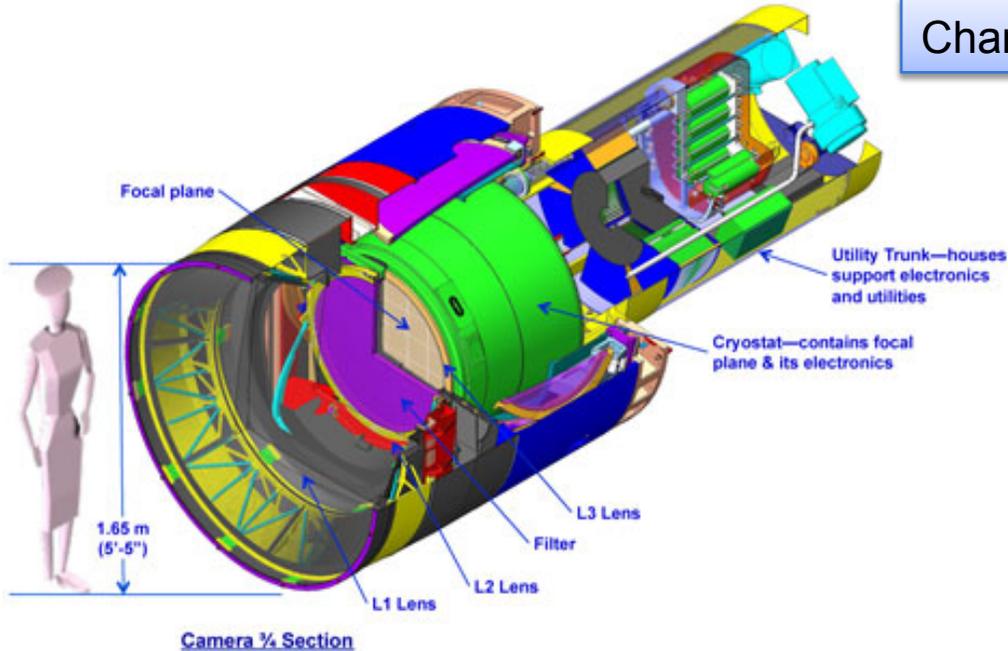


# La caméra



Champ de vue : 9.6 deg<sup>2</sup>

3.2 milliards de pixels  
6 filtres : UV (320 nm) → IR (1070 nm)  
résolution : 0.2 arcsec / pixel



- Pose de 15 s
- Positionnement de l'obturateur durant 1 s
- Lecture des CCD durant 2 s
- Nouvelle pose de 15 s
- Positionnement de l'obturateur durant 1 s
- Lecture des CCD durant 2 s
- Positionnement du télescope sur une nouvelle zone (proche) durant 5 s

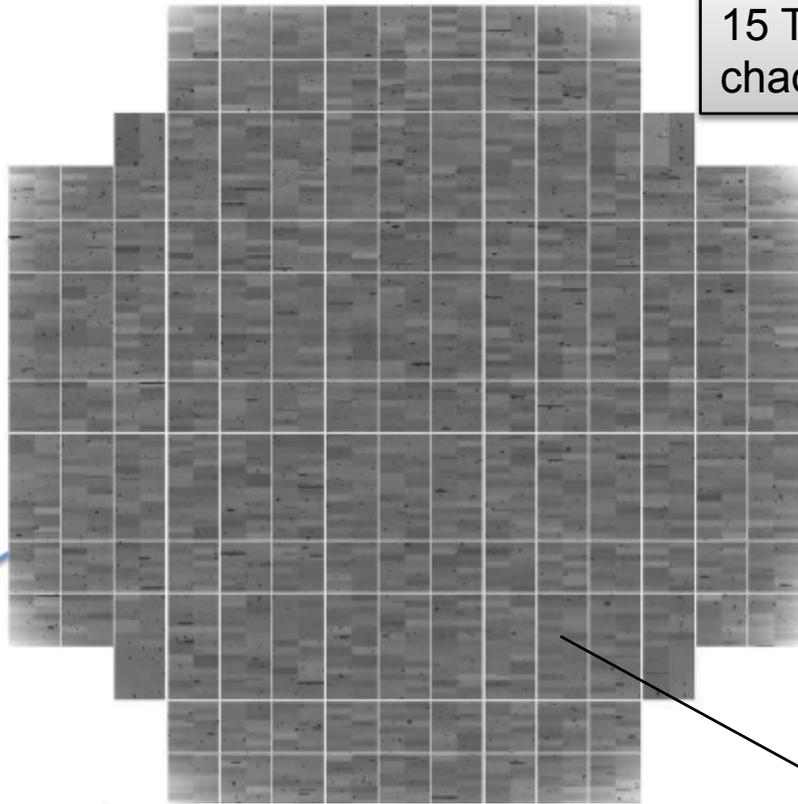


189 plans de CCD (4k x 4k)  
assemblés dans 21 modules  
de 3x3 CCD



# Les images

Image simulée 16 Mpixels  
1/189<sup>ème</sup> d'une pose complète de  
LSST



15 To de données  
chaque nuit



Chaque zone du ciel est photographiée  
~1000 fois en 10 ans → On additionne  
les images ce qui permet de gagner en  
profondeur au cours du temps  
(magnitude 24 → 27)

→ 50 Po d'images brutes

Pré-traitement quasi en ligne des  
données  
Alertes générées en moins de 60s  
>1 000 000 d'alertes / nuit





# Calendrier / Collaboration



Démarrage prévu en 2018 pour 2 ans de validations ("commissioning") de l'instrument

10 ans de prise de données : 2020 - 2029

Le calendrier est susceptible de glisser un peu en raison de la situation économique mondiale ...

LSST est une expérience prioritaire du DOE et de la NSF

Le rôle pionnier de l'IN2P3 dans LSST lui vaut d'être considéré comme un contributeur américain (seul dans ce cas) → 51<sup>ème</sup> état 😊

## Collaboration française :

LPNHE – LAL – APC – LPC – CPPM –  
LPSC – LMA et **CC-IN2P3**

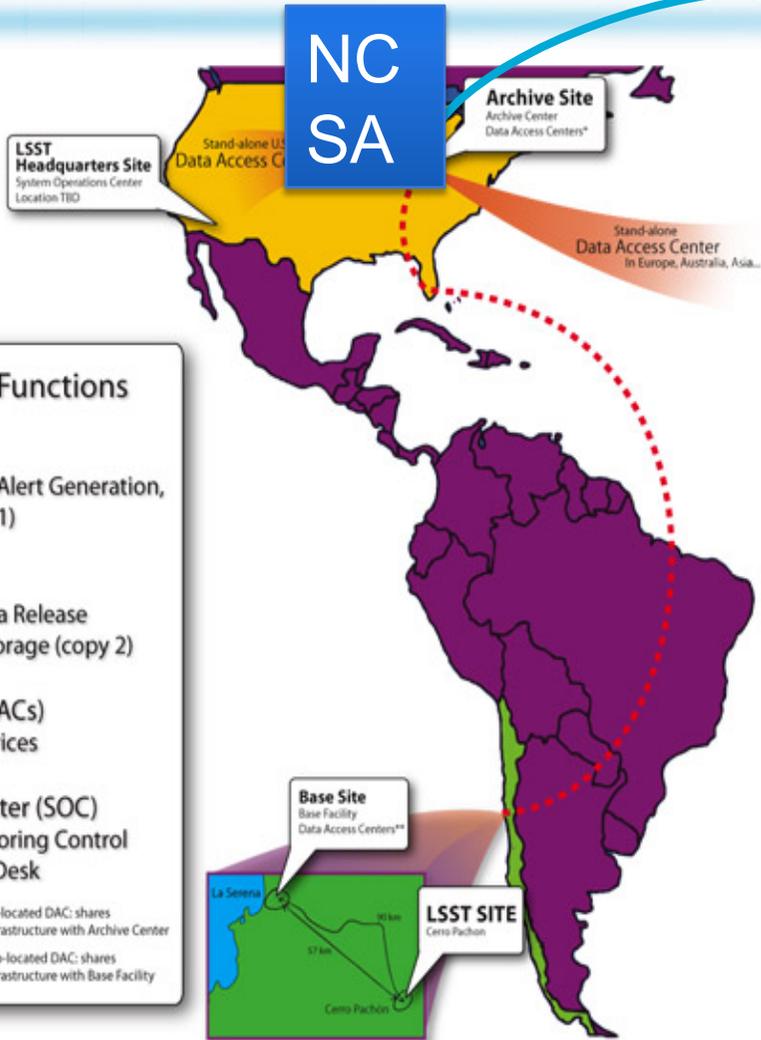
Les données seront publiques !

LSST est un grand instrument (comme le LHC) :  
La collaboration LSST met à disposition des catalogues d'objets

Des collaborations se forment pour exploiter ces données : IN2P3 → **Dark Energy Science Collaboration (DESC)**



# CC-IN2P3



- ### Site Roles and their Functions
- Base Facility  
Real-time Processing and Alert Generation,  
Long-term storage (copy 1)
  - Archive Center  
Nightly Reprocessing, Data Release  
Processing, Long-term Storage (copy 2)
  - Data Access Centers (DACs)  
Data Access and User Services
  - System Operations Center (SOC)  
System Supervisory Monitoring Control  
& End User Support/Help Desk
- \* Co-located DAC: shares infrastructure with Archive Center  
\*\* Co-located DAC: shares infrastructure with Base Facility



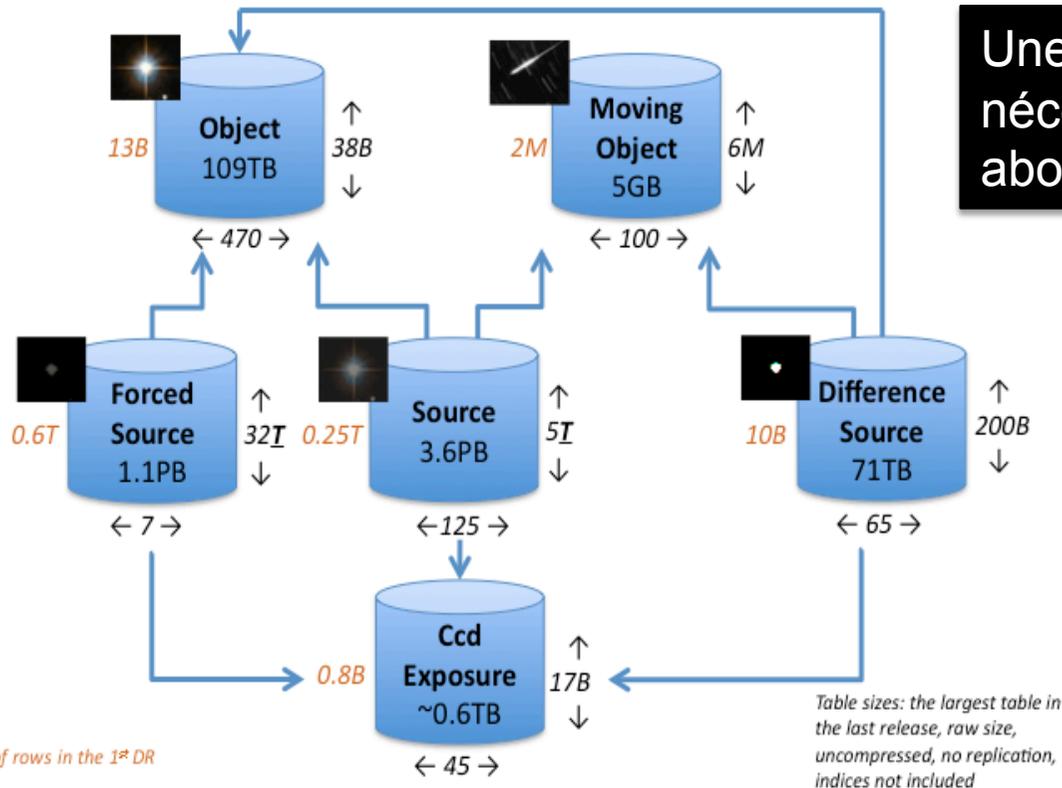
Le NCSA et le CC se partageront le traitement des données – 50% - 50%

Les données seront ensuite ré-échangées de façon à ce que chaque site dispose d'un lot complet

# Le système de base de données de LSST



Après le traitement des images des algorithmes permettent d'identifier les objets astrophysiques présents (étoiles, galaxies, quasars, candidats SN, astéroïdes, etc.)



Une requête dans une telle DB peut nécessiter un temps quasi infini pour aboutir → Mieux vaut être astucieux !

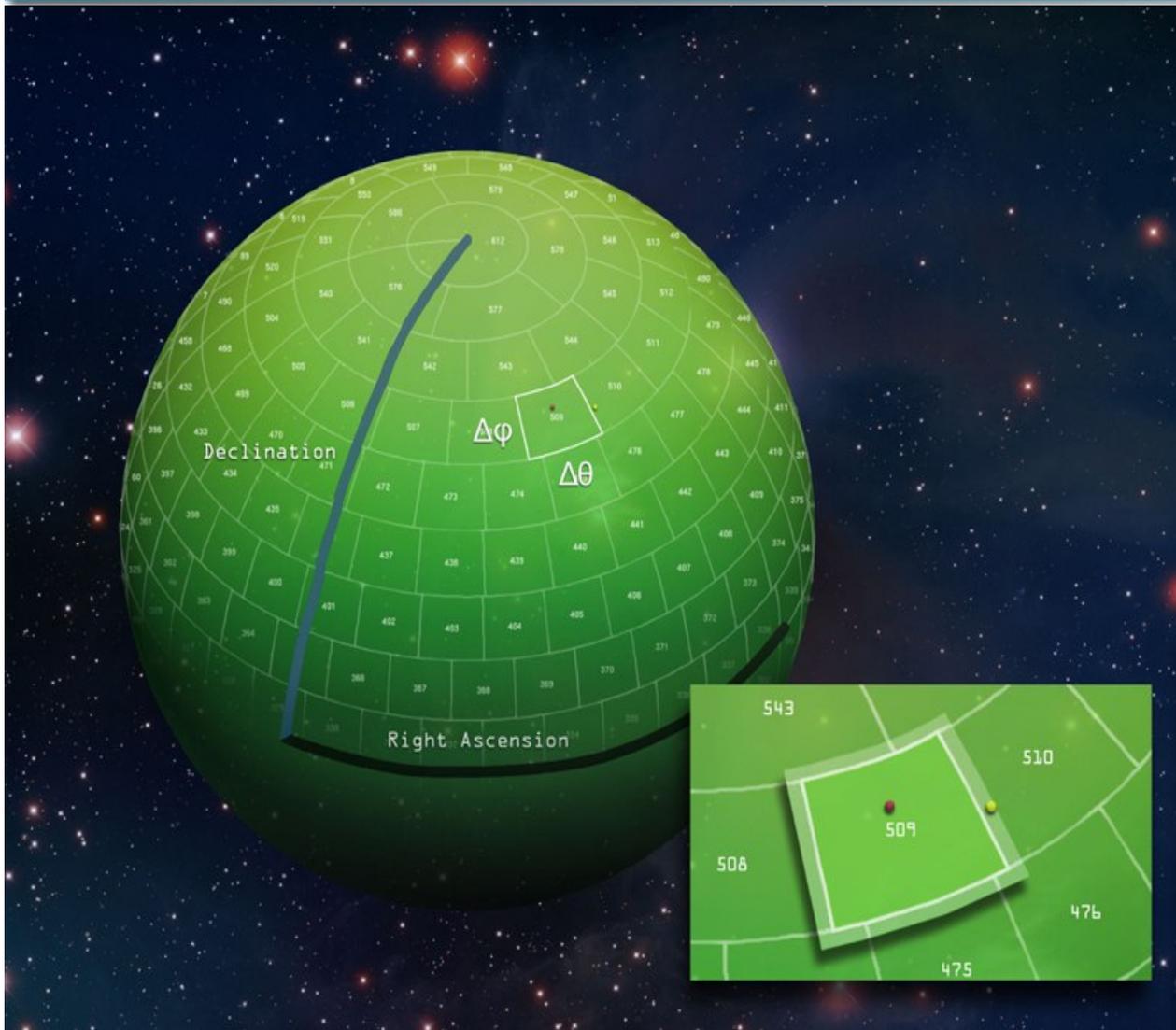
A terme : 20-30 Po de données dans la base de données

Pour DESC il sera sans doute nécessaire de remonter à l'information "pixels"

La table des "sources" contiendra 250 milliards de lignes après la première année et croitra jusqu'à 5000 milliards au bout de 10 ans



# Bases de données



Les tables dans la DB correspondent à un pavage géométrique du ciel observable

Optimisé pour la plupart des recherches astro

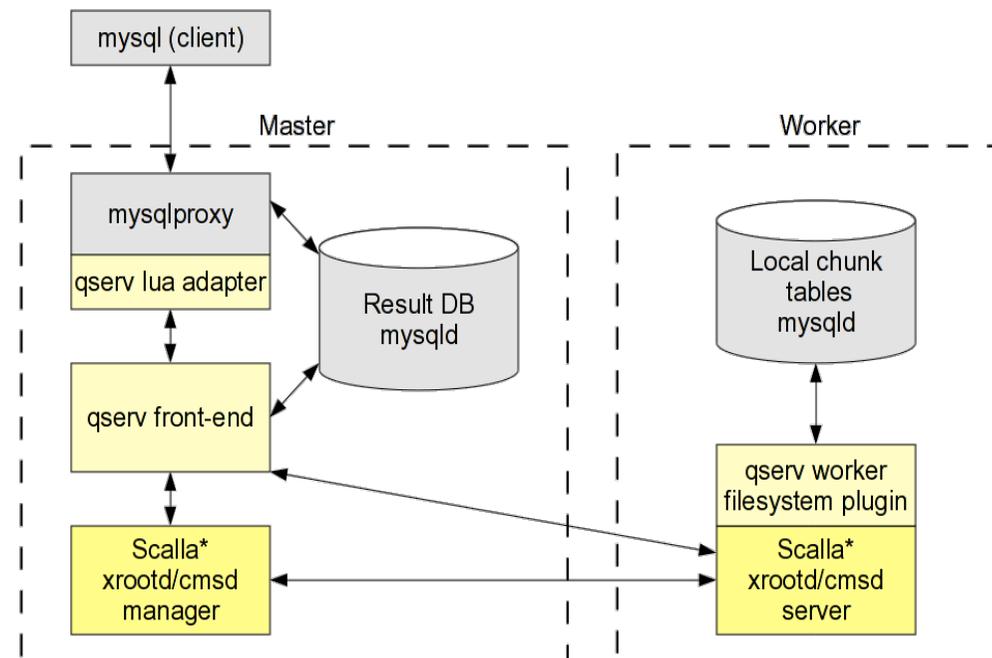
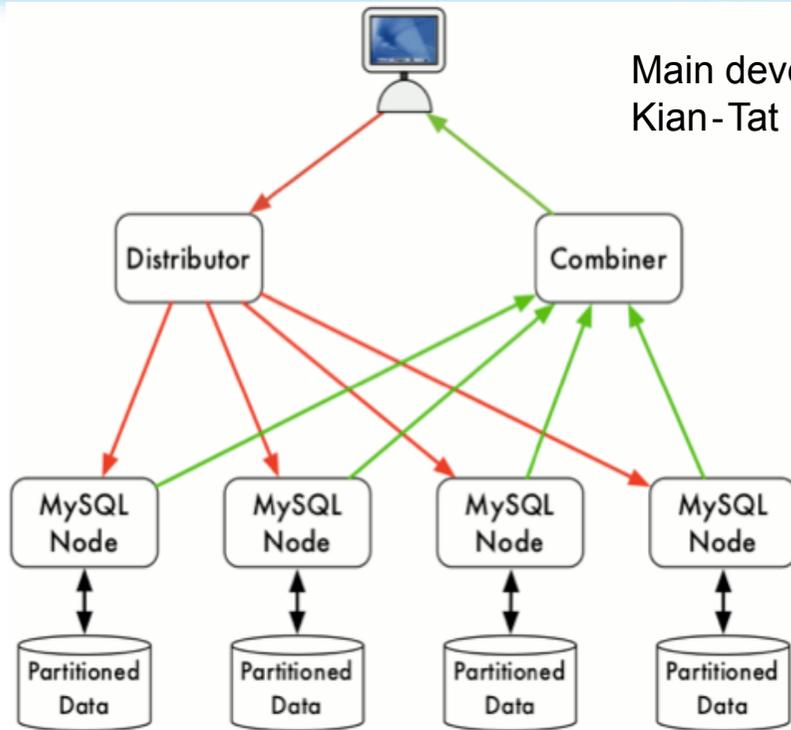
Certaines requêtes ne peuvent pas aboutir en un temps raisonnable

→ "give me the 5 most brightest objects of the sky" is an impossible query !

# Qserv: une base de donnée distribuée "shared nothing"



Main developers : Daniel L. Wang (SLAC), Serge M. Monkewitz (IPAC), Kian-Tat Lim (SLAC), Jacek Becla (SLAC), Douglas Smith (SLAC)



\* Includes a custom set of patches which have not yet part of the official Scalla

MySQL n'est pas une techno obligatoire





# Simulation des images

LSST dispose d'un simulateur d'images très complet

Intéressant pour la mise au point des algorithmes d'analyse d'images

→ On compare ce que l'algorithme détecte avec ce qu'on a voulu simuler

- Sources : étoiles, galaxies, quasars, etc...
- Effet de l'atmosphère (turbulences, propagation dans les différentes couches...)
- Effet de la pollution lumineuse (si, si, il y en a aussi dans les montagnes du Chili...)
- Effet de la couverture nuageuse
- Prise en compte de la Lune
- Simulation de l'optique du télescope
- Réponse des capteurs CCD

Permet également de mettre au point la stratégie d'observation

- Pointage
- Filtre
- Lune
- etc.

Pour optimiser le cycle utile du télescope

Peut être couplé avec un programme de simulation cosmologique

La simulation complète d'un couple d'images sur la totalité du plan focal nécessite **80 à 100 h** de calcul sur 20 processeurs en parallèle  
Plusieurs centaines de milliards de photons tracés individuellement



# Rôle du CC dans Qserv



Actuellement Qserv est installé sur 4 nœuds iDataplex au CC (1 master + 3 nœuds)

File System de 5 To monté via NFS sur le master pour le chargement des données

Le CC fournira un banc de test de 50 nœuds iDataplex dès que la procédure de déploiement automatique des nœuds sera disponible

- Tests systématiques
- Optimisation des paramètres (taille des tables, ...)
- Tests des versions successives de Qserv

En vue de la Final Design Review de LSST (fin de l'année) nous allons étendre le banc de test à 300 nœuds pour réaliser un test à grande échelle

- Mise à disposition des 300 nœuds pour quelques semaines

Apport majeur pour la collaboration !



# Data Challenge



Le CC va être fortement impliqué dans le prochain data challenge (été)

L'un des buts principaux est de démontrer la capacité de LSST à traiter les données sur 2 sites

- 1 – 2 MHS06
- 100 To de stockage (3 X4540 – iRODS)
- Liaison transatlantique

Utilisation d'iRODS pour gérer les échanges de données

Revu à la baisse : concentration sur le traitement des données de SDSS avec le code LSST

Les "pipeline" de traitement des images sont actuellement en cours de test au CC  
→ Complexe mais excellente interaction avec les experts américains

En plus du Data Challenge, il y aura probablement un traitement spécifique de données du CFHT-LS avec le code LSST → possibilité d'utiliser des données pour de "vraies" études scientifiques



# Bilan après 6 mois dans LSST



Portage de la simulation sur le cloud OpenStack

- Des milliers d'heures de calcul utilisées
- Très grande souplesse (VM taillées sur mesure, jobs multi-cores, pas de limitation sur le temps de résidence des jobs en machine...)
- Des tests systématiques sur les performances restent à faire (overhead ?)

- Bonne collaboration avec les développeurs
- Communauté LSST-Calcul française active

L'architecture du soft est moderne et bien faite

- Réécriture de 0
- Python + C++

L'architecture matérielle et la partie opérationnelle au NCSA ne passent pas à l'échelle (à mon avis)

- Le CC peut apporter beaucoup à ce niveau
- Le data challenge et la plateforme qserv constituent une excellente opportunité pour mettre en avant le CC



# Bilan après 6 mois dans LSST



Beaucoup plus grande autonomie du CC dans LSST comparé à LHC

- Pas de contrainte au niveau des outils
- Un vrai rôle à jouer

Nous allons certainement rapidement devoir faire face à des problèmes de performance en terme d'I/O

- Traitement d'image sur une multitude de petits fichiers
- Besoin d'expertise

Se poser la question d'une participation du CC-IN2P3 au-delà de l'architecture de traitement de données

- Base de données distribuées
- Parallélisation massive
  - Interaction avec l'équipe recherche
  - Plateforme Dell