Computing model for Large Cosmological Surveys and Simulations

PAU and MICE

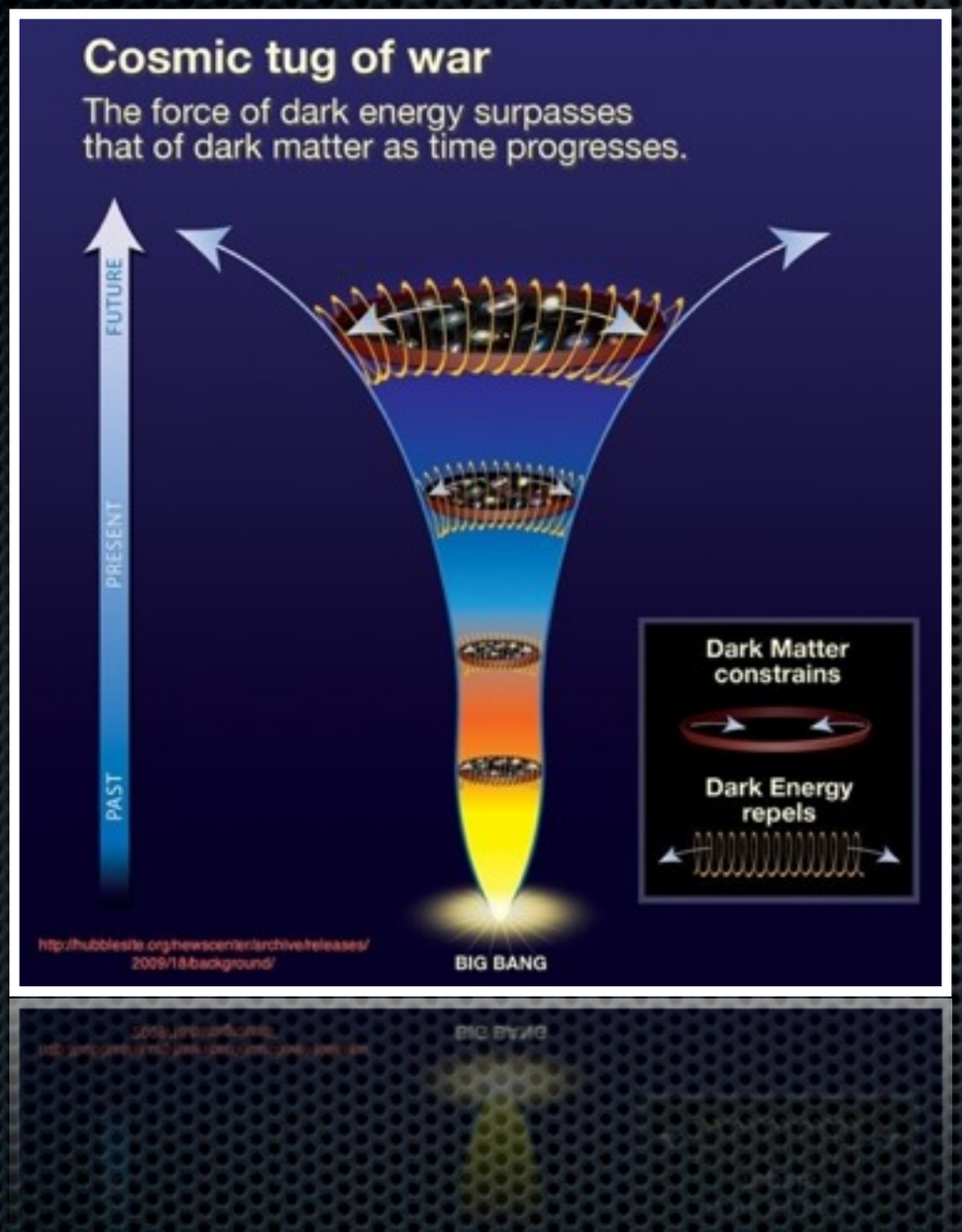
Nadia Tonello
Santiago Serrano

Outline

- PAU survey
  - Science
  - Camera
  - Data Management
    - Nightly pipeline
    - Multi Epoch pipeline
    - Pixel Simulation pipeline
    - Analysis pipeline
    - Data Base model
- MICE simulations
- Conclusions
- Team

Physics of the Accelerating Universe

Survey

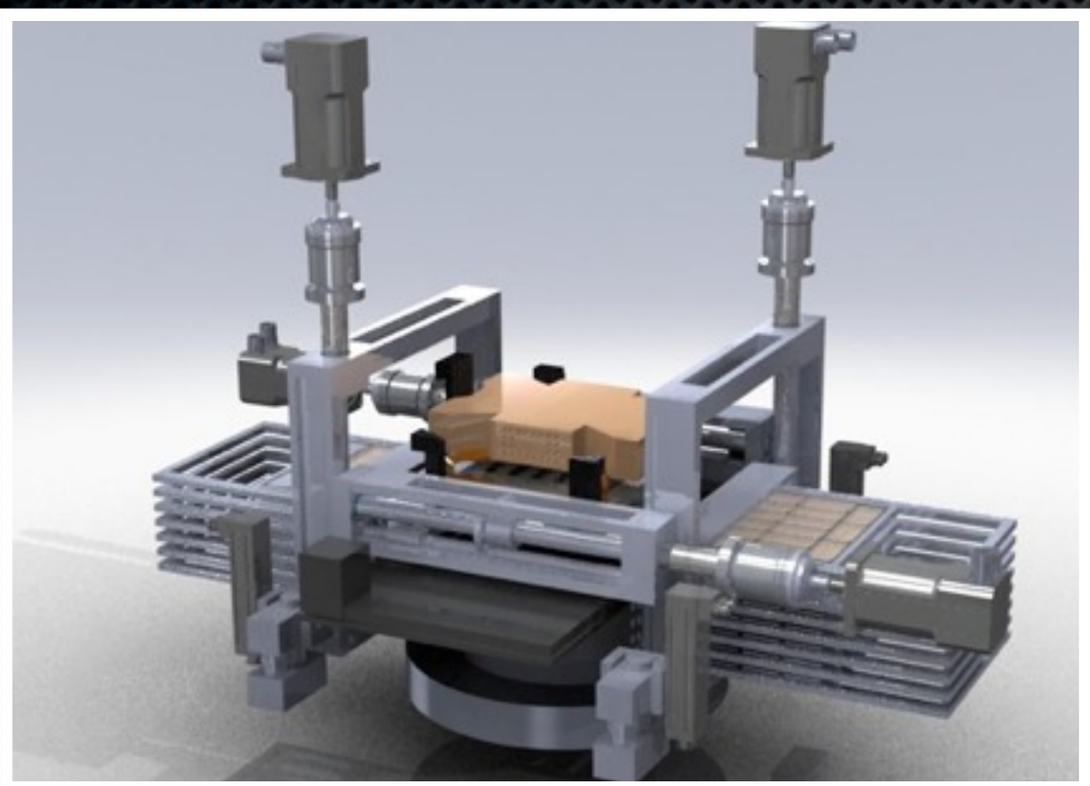


- Study dark energy and cosmic acceleration
- Measure ~10^6 galaxies up to redshift 0.9
- Photo-z error under 0.3%
- 200 deg^2 survey area

http://www.pausurvey.org

Physics of the Accelerating Universe

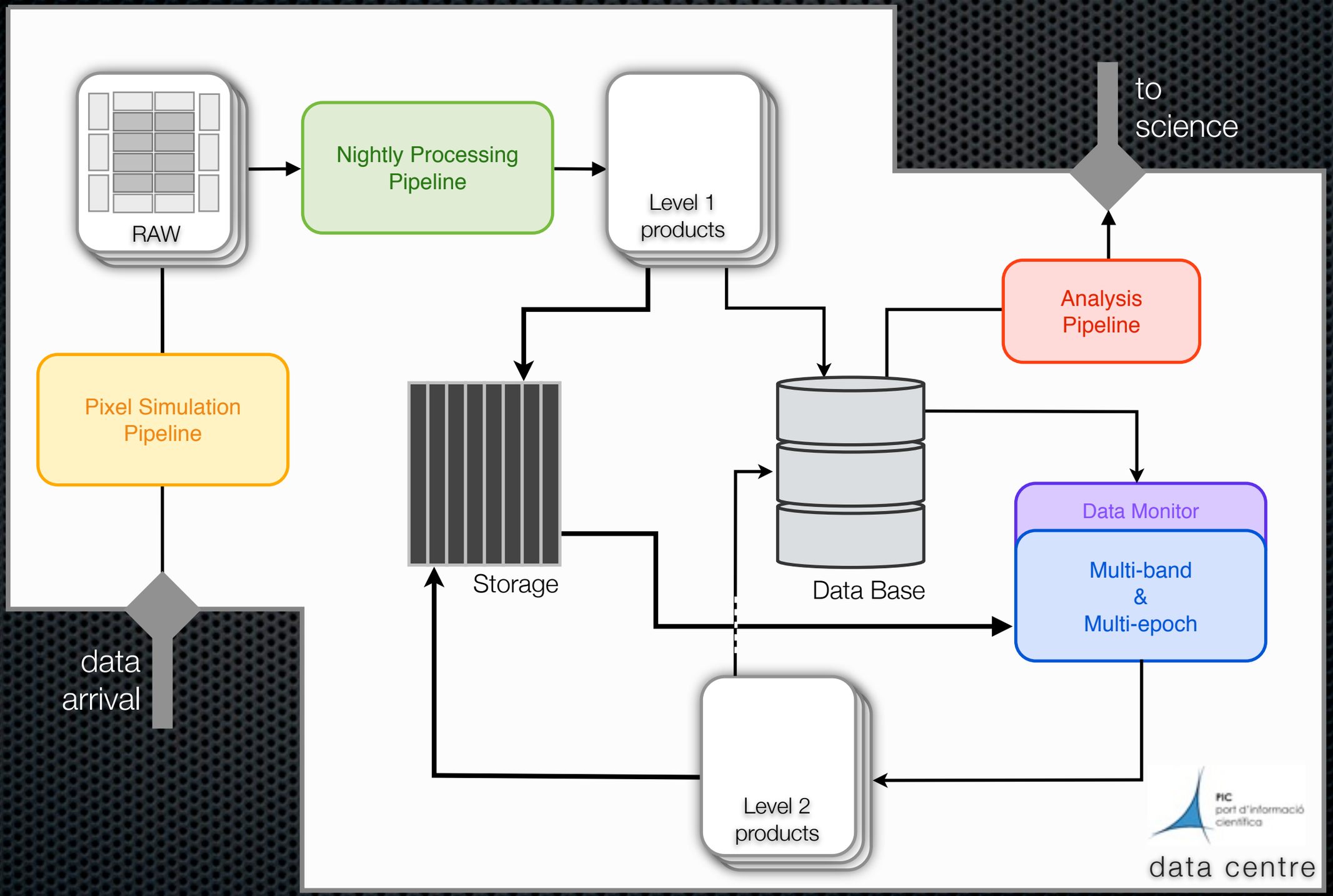
Camera



- To be installed in primary focus at 4.2m WHT
- Mosaic with 18 4k×2k HPK CCDs
- FOV 40’ unvignetted
- 6 Broad Band filters (SDSS)
- 40 Narrow band filters
- 300MB / mosaic frame
- 200GB / night

data management

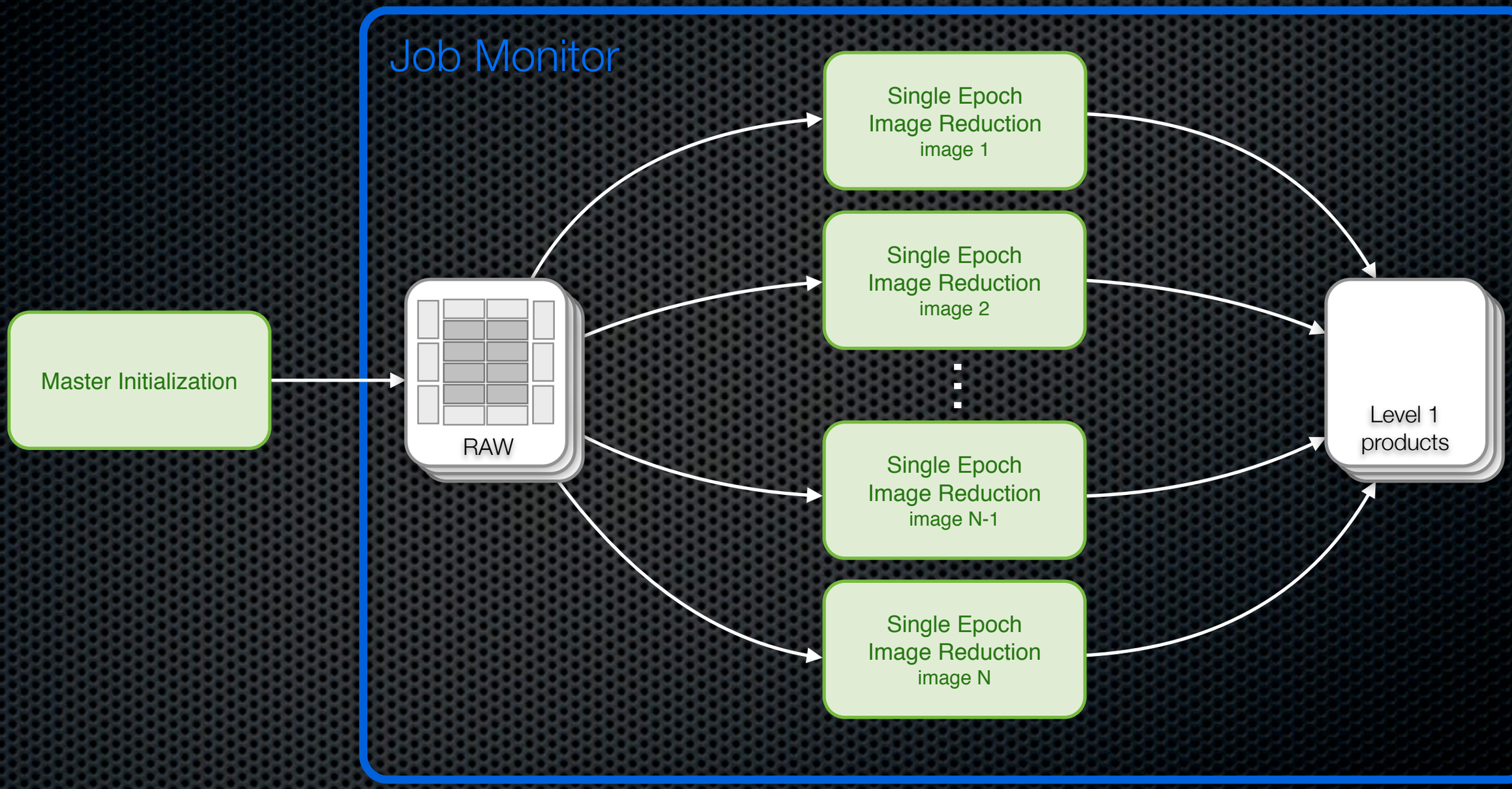
Physics of the Accelerating Universe
Data Management



data management

Data Management

Nightly pipeline



12-hour constraint ➙ Parallelize jobs (100/200)

Data Management

Multi-Epoch and Multi-Band pipeline

data management

New Event

New Remap Tile

Boundaries

New Remap Tile

Central

New Coadd Tile

Reference

New Detection Objects

Remaps

Ref. Tile complete?

No

Exit

Yes

Tile complete?

No

Exit

Remaps Waiting queue +1

Yes

Coadds or Remaps waiting?

No

Exit

Coadds

Remaining overlapping scans?

Yes

Exit

Detections Waiting queue +1

No

Logic

Astro Refinement

PSF Homogen.

Image Stacking

Other

New Event!

Reference Band

Ref. Tile complete?

No

Exit

Coadds Waiting queue +1

Yes

Image Processing

Dual-Image Source Extraction

Direct Source Extraction

Dual-Image Source Extraction

Source Extraction

Inter-Tile Photometric Check

Global Match + ZP_{global} calculation

Photo Calib.

Depth Mask

Mask

Catalogue-level Combination

Catalogue Ingestion

Global Object Ingestion

Catalog

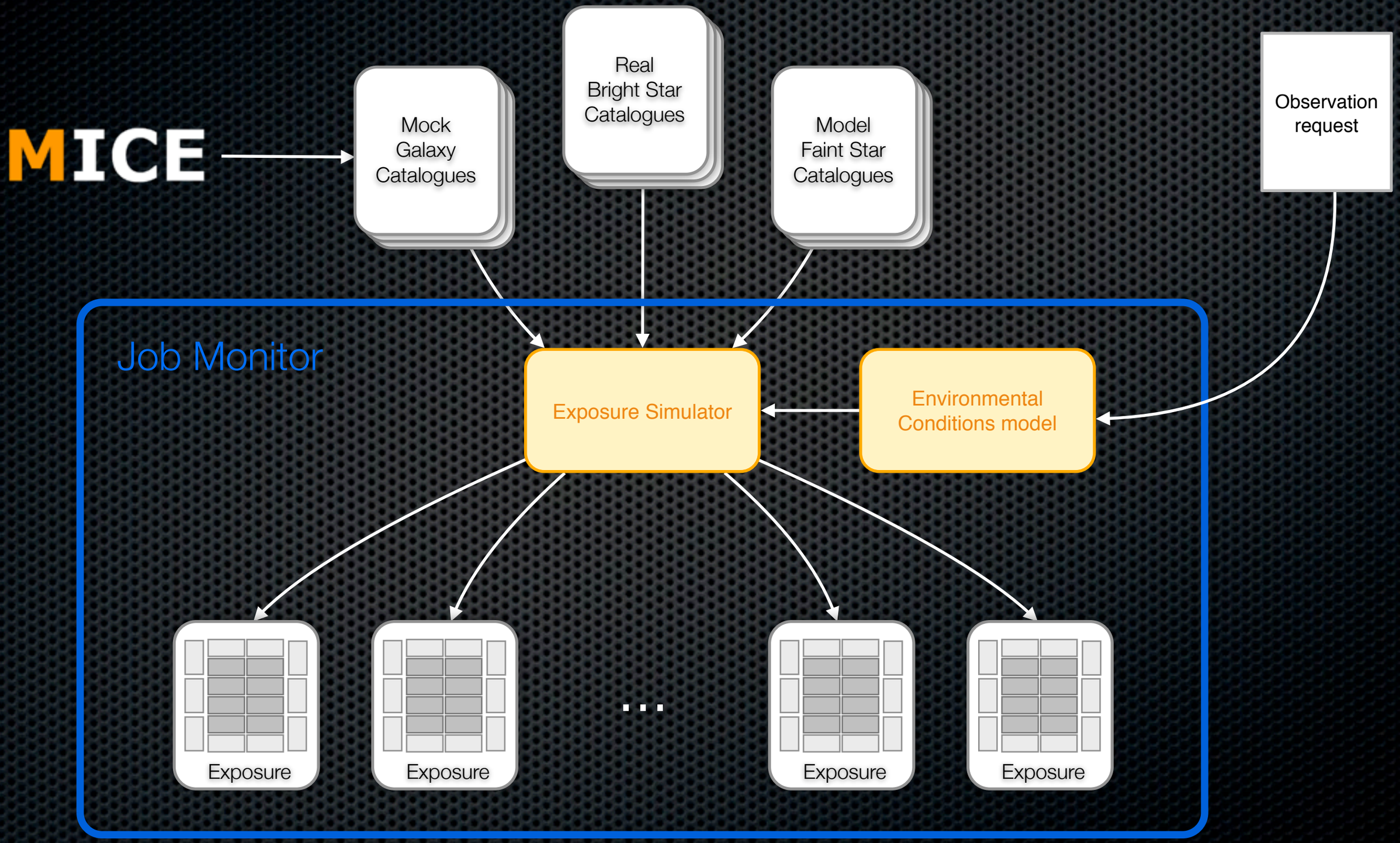
New Event!

Remap CCD

data management

Data Management

Pixel Simulation Pipeline

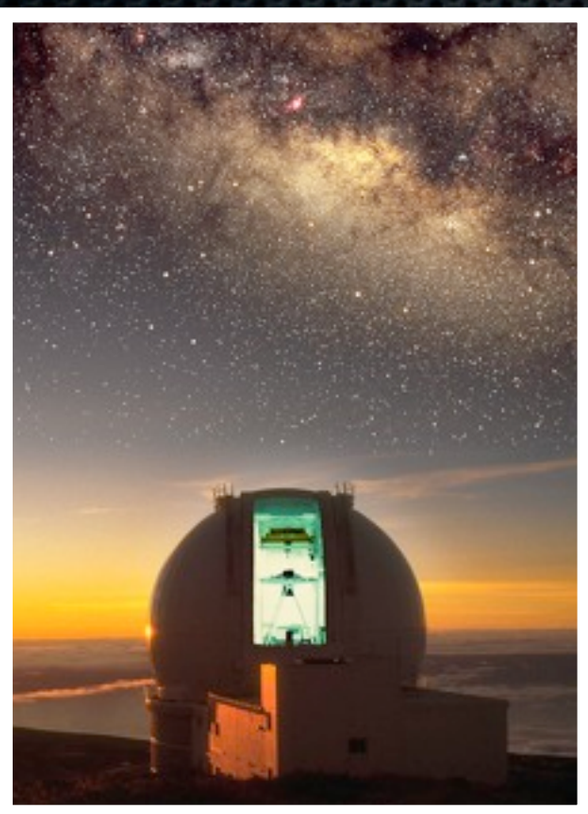


data management

Data Management

Analysis Pipeline

Real Universe + Telescope



PAUdm
Processing Pipelines

Output catalogues

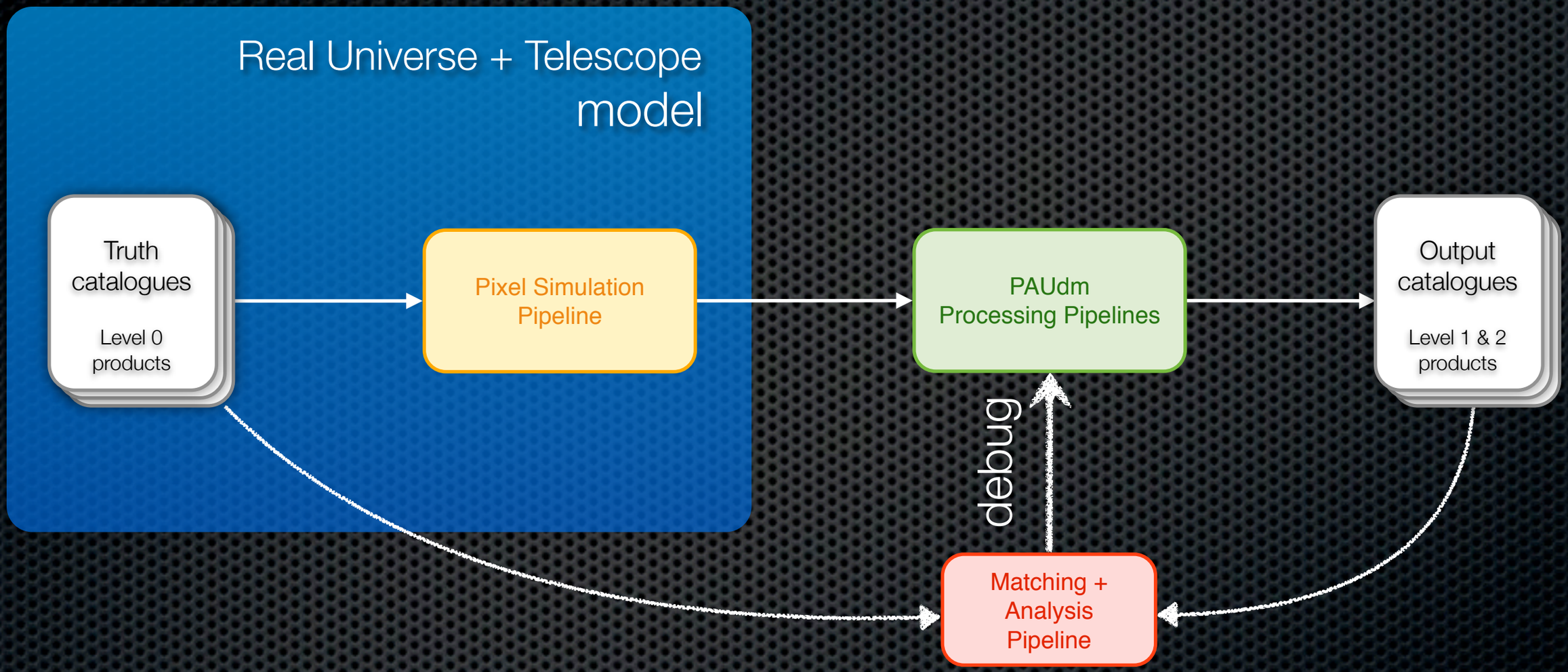
Level 1 & 2 products

Pipeline check ?

data management

Data Management

Analysis Pipeline

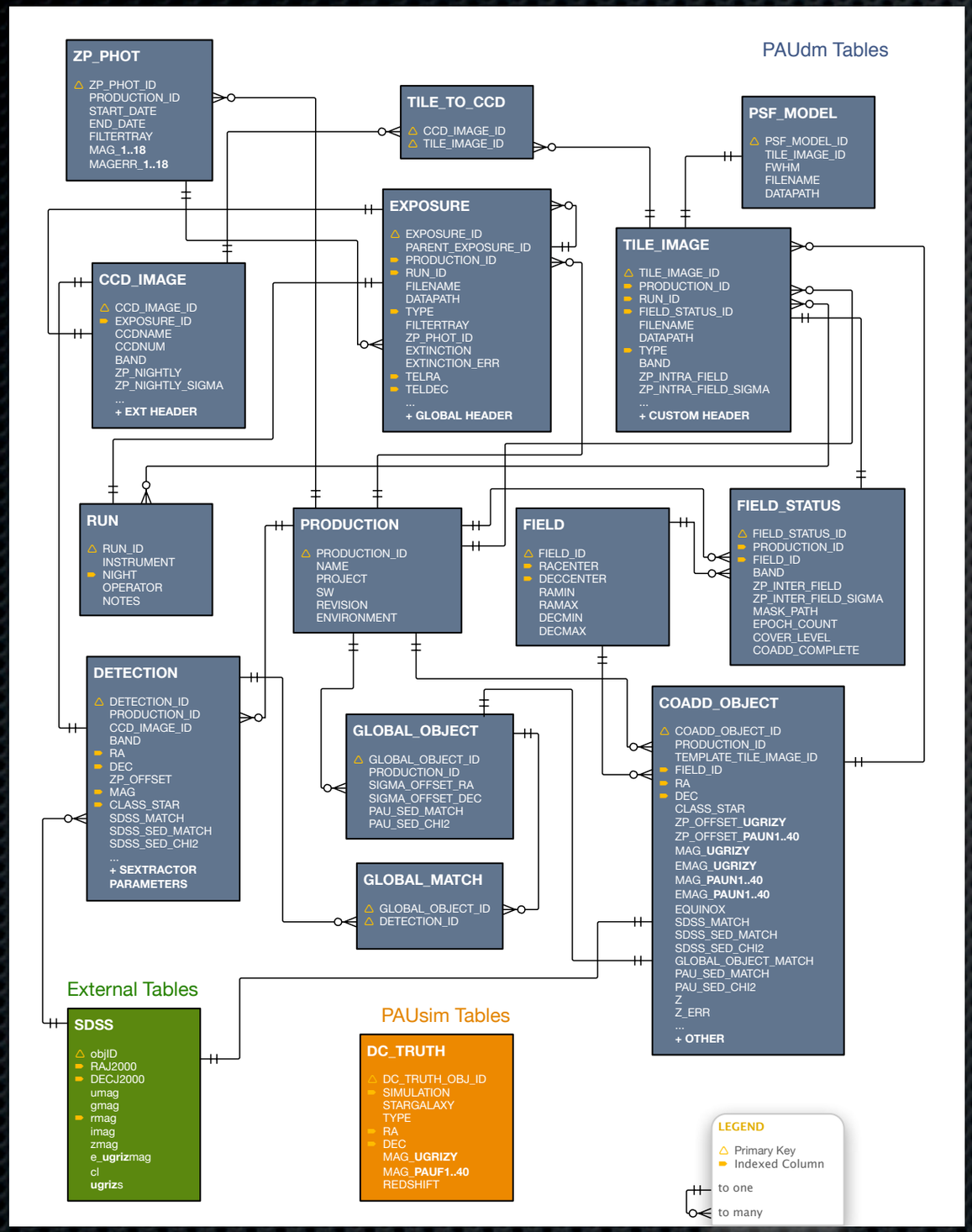


Pipeline check ➔ Matching + Analysis *Truth input* vs *output*

data management

Data Management

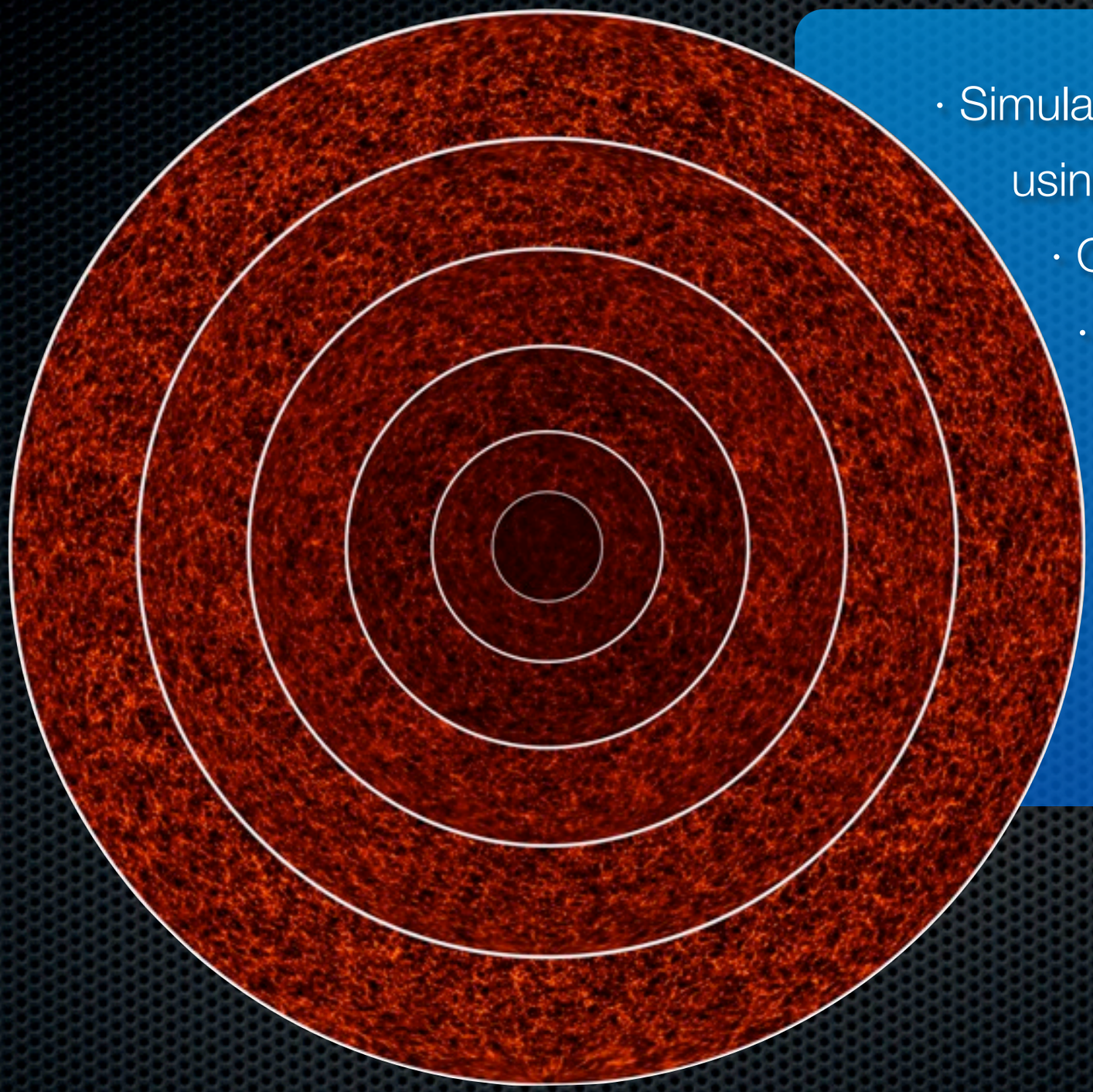
Data Base model



- Integrated in pipelines
- Interface between storage and processing
- Dual operation mode:
  - PostgreSQL in GRID
  - SQLite in PC
- Large data
  - > 2.000.000.000 entries
  - Contains 500 GB of data
  - Links ~1 PB of data

MICE

Marenostrum Institut de Ciències de l'Espai
Simulations



- Simulations processed @ Marenostrum using 4000 processors in parallel
- Grand Challenge 4096^3 DM particles
- Transfered and stored @ PIC (100TB)
- Used to produce mock galaxy surveys (input for pixel simulation)
- Used to develop the scientific case of large astronomical surveys (PAU, DES, EUCLID, PLANCK,...)

www.ice.cat/mice

MICE

Marenostrum Institut de Ciències de l'Espai
Simulations

Data Transfer from BSC

- Issue: BSC security measures are very restrictive, thus data transfer methods are limited.
- Solution: Mount a remote file system over ssh at BSC and install srm server _to serve_ it.

Conclusions

· We have developed processing pipelines to process high volumes of data.

· Parallelization model of the jobs in Grid has been an effective solution to deal with the time constraint.

· The whole PAUdm system is expected to be fully implemented and tested before fall 2012, when the PAU camera will start to produce data.

· MICE simulations have been successfully transferred and stored at PIC and being used to develop science case and data processing pipelines of major astronomical surveys.

The Team

PAU data management

MICE

IEEC

Anne Bauer
Christopher Bonnet
Ricard Casas
Jorge Carretero
Francisco Castander
Martin Crocce
Samuel Farrens
Pablo Fosalba
Enrique Gaztañaga
Santiago Serrano

Ciemat

Eusebio Sanchez
Nacho Sevilla

PIC port d'informació científica

Christian Neissner
Pau Tallada
Nadia Tonello

IFAE

Marino Maiorino
Pol Martí
Ramón Miquel
Carlos Sanchez