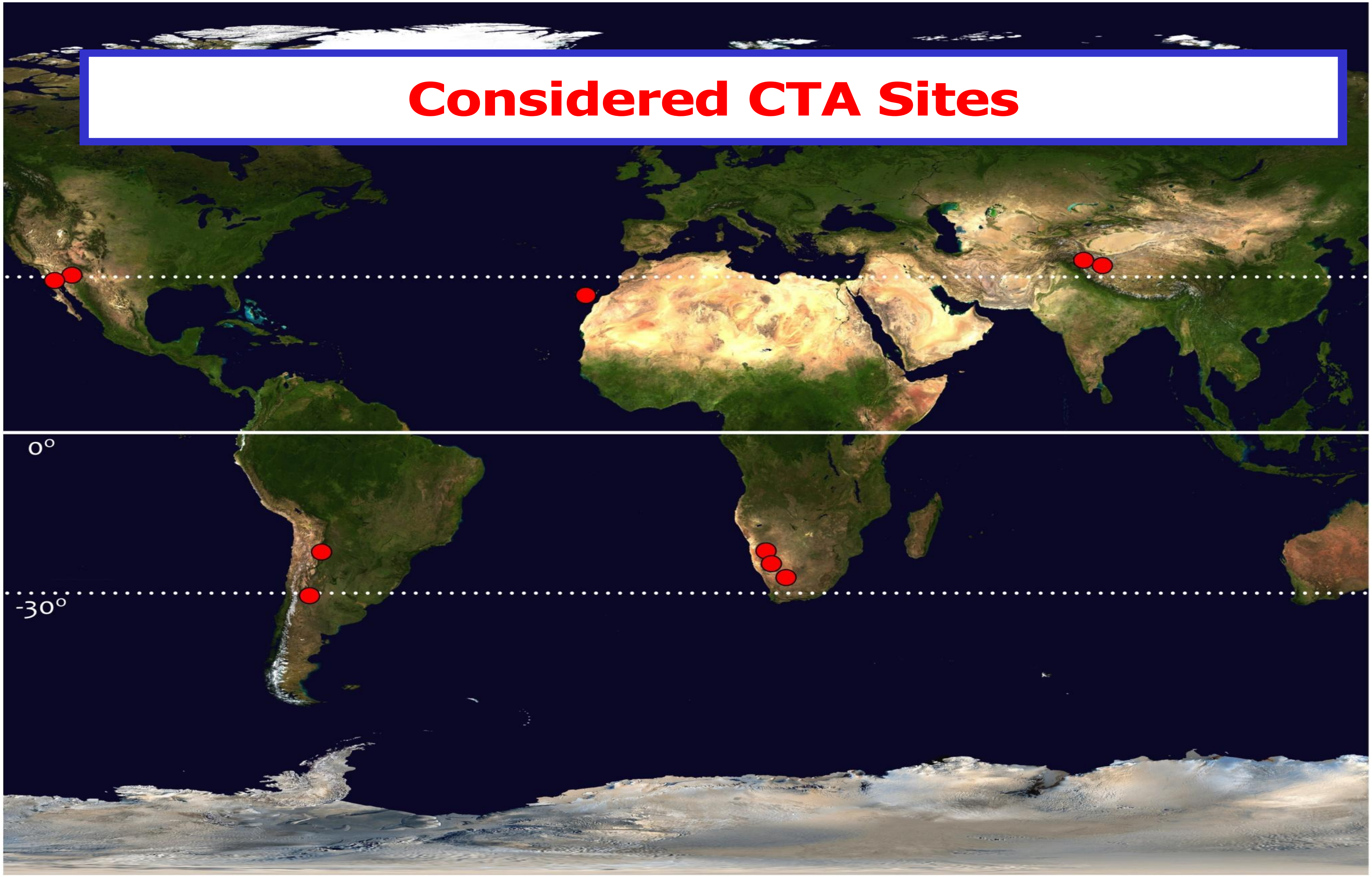
C T A
Cherenkov Telescope Array

Au nom des groupes CTA :
"ACTL (Array ConTRol)" et "Data Management"
G. Lamana(LAPP), N. Neyroud(LAPP)
E. Lyard (ISDC)
U. Schwanke(DESY)
Et al.

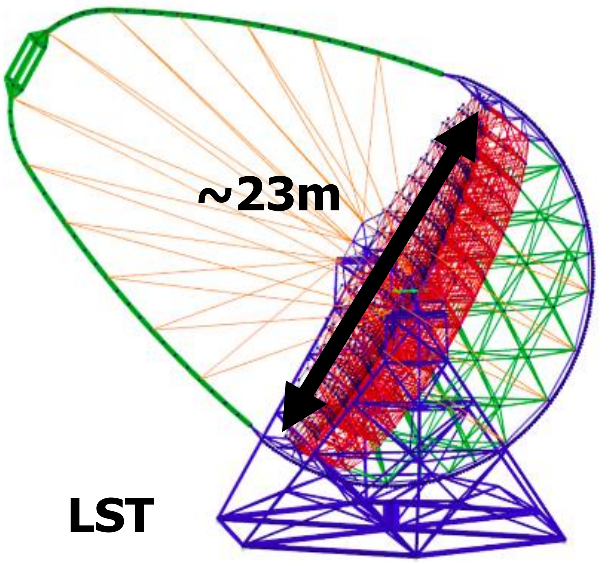
- CTA is the next-generation facility for the detection of VHE* gamma-rays
- A northern and a southern array for full sky coverage
- An instrument with:
  - better flux sensitivity
  - better angular resolution (~0.05°)
  - wider energy coverage (some 10 GeV – 100s TeV)
- Main science goals:
  - Understanding the origin of cosmic rays and their role in the Universe
  - Understanding the nature and variety of particle acceleration around black holes
  - Searching for the ultimate nature of matter and physics beyond the standard model

*VHE = very high energy, E>some 10 GeV

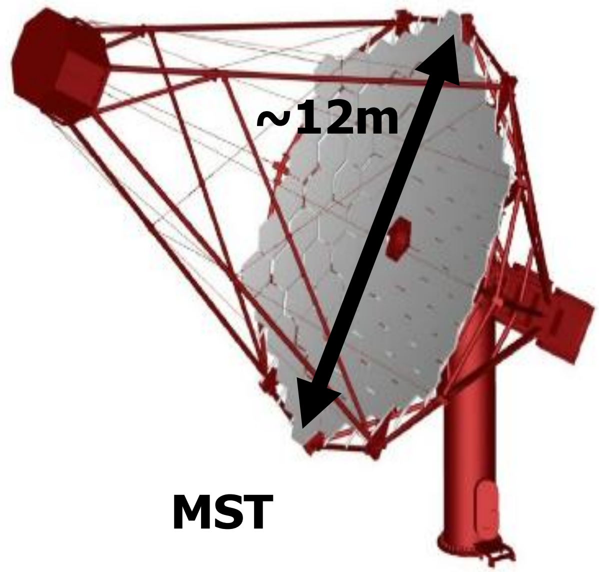
Considered CTA Sites



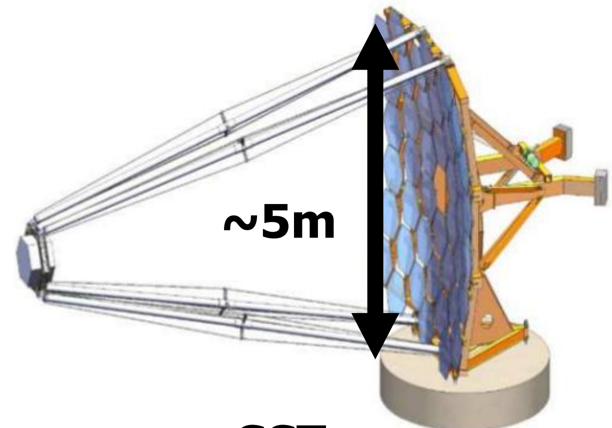
3 sizes of telescopes



LST



MST



SST

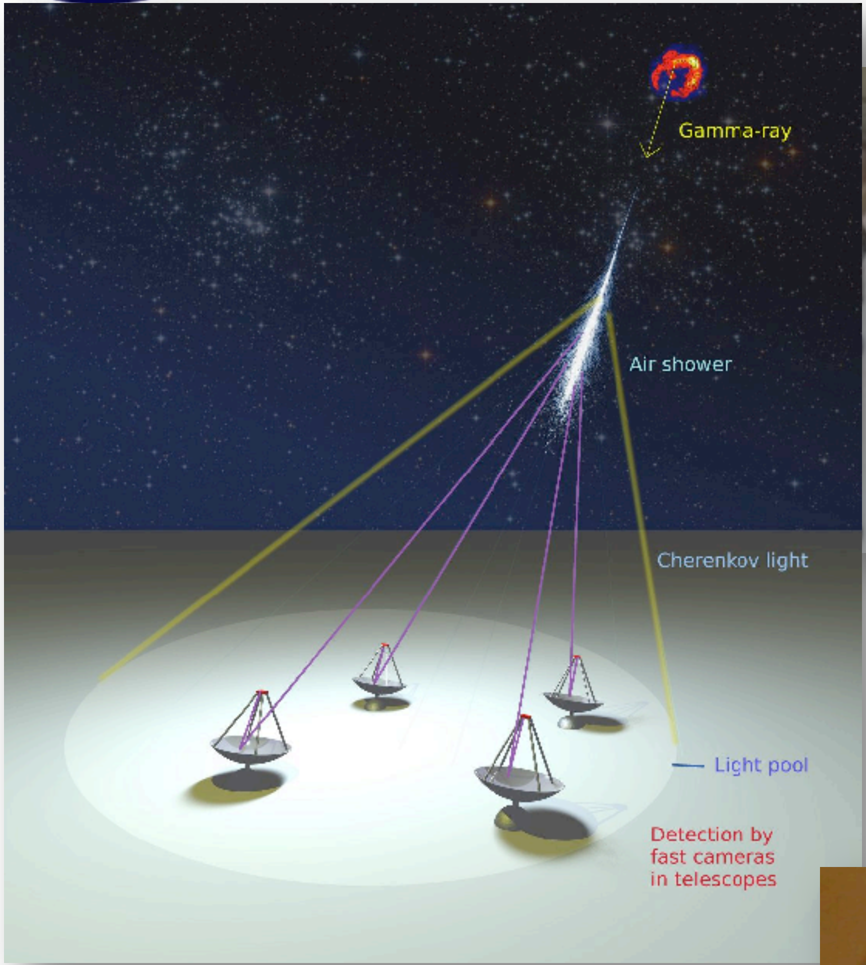
- **Large size telescope**
- **Access to E<100 GeV**
- **Parabolic mirror**
- **FoV* diameter: 5°**
- **~4 LSTs**

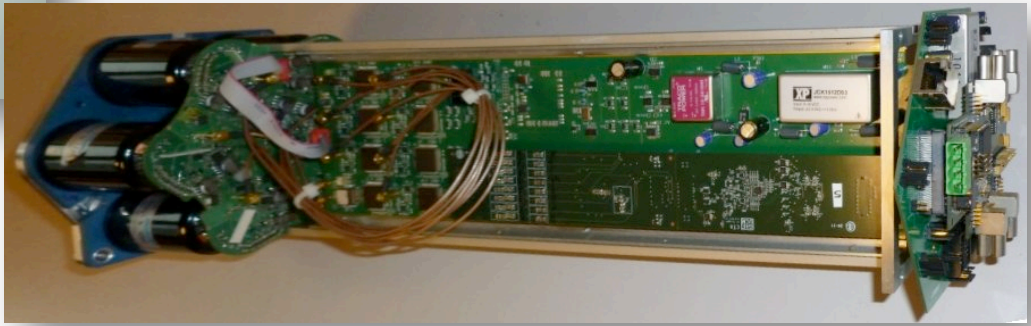
***FoV = field of view**

- **Medium size telescope**
- **100 GeV<E<10 TeV**
- **1 or 2 mirrors**
- **FoV: 7-8°**
- **~25 MSTs**

- **Small size telescope**
- **E>10 TeV**
- **1 or 2 mirrors**
- **FoV: 10°**
- **~30 MSTs**

Camera

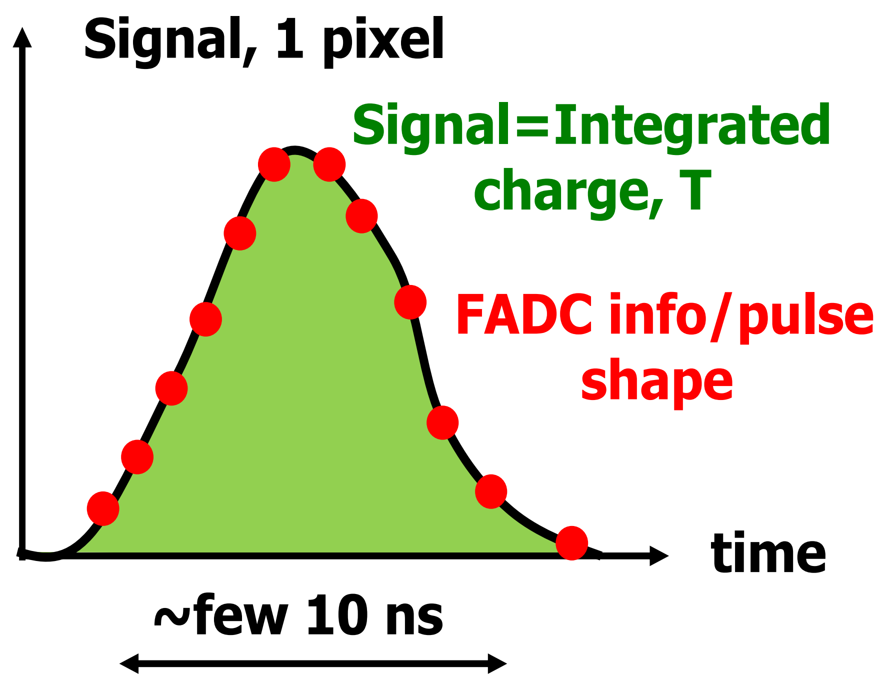

Gamma-ray
Air shower
Cherenkov light
Light pool
Detection by
fast cameras
in telescopes



Data Rates and Data Sources

Camera	Telescope	Array
Camera Electronics High Voltage System Array Trigger Interface Camera Lid Local Clock	Drive Systems Safety System Camera Lock/Shelters CCD camera(s) Radiometer Calibration Devices (LED pulsers, Lasers) Active Mirror Control	Array Trigger/GPS clock LIDAR(s) Weather Station(s) Optical Telescope(s)

- **Data rate is dominated by the Cherenkov cameras**
- **Signal integration vs. pulse shape recoding has not been decided yet**
- **There are numerous telescope-specific and central devices that must be controlled and read-out**



Data Types

- **Camera Events Data**
  - Camera Events data (CE data) consisting of photo-detectors-registered signals and triggered by the camera system.
- **Camera Slow Control Data**
  - Camera Slow Control data (CSC data) consisting of ancillary information related to calibrations, alignment, functioning, safety and general environmental/status conditions of the camera system.
- **Telescope pointing data**
  - Telescope drive system data related to pointing (TEL data).
- **Instrument Auxiliary Data**
  - Any other Instrumental Auxiliary slow control data (IA data) from any other auxiliary system (e.g. AMC, GPS, Lidar, etc…) but the camera.
- **Observatory Raw Events Data**
  - Observation Raw Events data (ORE data) consisting of the outcome of the event building processes which matches CE data from cameras involved in an observation and related TEL data.
- **Level 0 data**
  - Raw data which are defined as those event-based data which are the original input of the Reconstruction Pipelines are called “Level 0 data”.

Raw Data Rate

Data reduction options must be studied

- Integrate FADC samples before DL0 .
- Suppression of pixels with low signals.
- Compression ?
- Software trigger to reject hadrons.

-> Mixed data formats to keep flexibility:

- Should be standard since the first prototype

… then format should evolve to reduce data size.

Ref. Document: DATA-ICT/130110
(provided to the review committee)

Studies based on MC
(K. Bernlohr et al. 2009, CTA, U. Schwanke 2011)

CTA south array:
Type E, 69 telescopes: 4 LST , 23MST and 32 SST
Energy threshold: 34 GeV
Gamma-rate: 12Hz
Proton-rate: 13kHz
Mean trigger tel. mult.: 1.7(LST), 2.3(MST), 1.9(SST)
Array triggered event data size: 300kB

CTA north array:
Type NA, 21 telescopes: 4 LST and 17 MST
Energy threshold: 34 GeV
Gamma-rate: 8Hz
Proton-rate: 7kHz
Mean trigger tel. mult.: 1.9(LST), 2.0(MST)
Array triggered event data size: 200kB

	Data rate (GB/s)	Maximum data volume per day (TB)	Data volume per month (TB)	Data volume per year (PB)	Total data volume in 30 years (PB)
CTA South (array E & scenario 1)	0.3	13	120	1.42	42
CTA South (array E & scenario 2)	0.4	17	160	1.9	56
CTA South (array E & scenario 3)	8	346	3000	38	1110
CTA North (array NA & scenario 1)	0.12	5.2	50	0.57	17
CTA North (array NA & scenario 2)	0.15	6.5	60	0.7	21
CTA North (array NA & scenario 3)	1.8	78	700	8	250
CTA (scenario 1)	0.42	18	170	1.99	59
CTA (scenario 2)	0.55	24	220	2.6	77
CTA (scenario 3)	9.8	423	3700	46	1360

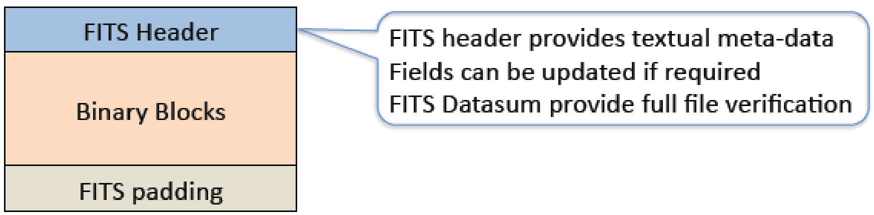
Scenario 1: Readout without waveform sampling (i.e. $\delta = 0$). Only the time-integrated Cherenkov signal is stored in 2 bytes per pixel.
Scenario 2: Readout with waveform sampling for selected pixels. In this case, it was assumed that the full waveform will be kept for the few important pixels that make up the shower image, while all other pixels will be treated like in Scenario 1 (i.e. $\delta = 1$ to 9%).
Scenario 3: Readout with waveform sampling for all camera pixels (i.e. $\delta = 1$)

CTA Raw Data (L0) is the bulk data coming from the cameras (actual events, EVT0)

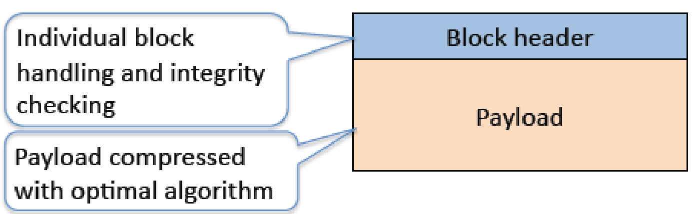
- Two possible approaches emerge

1) Block structure: sequential records of incoming data

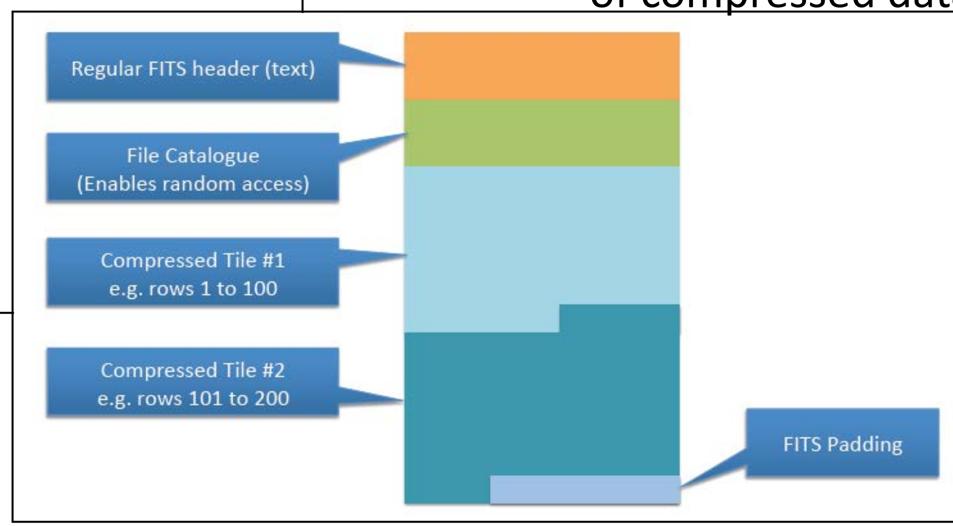
- Files -> FITS compliant



- Binary blocks



2) Compressed FITS: structured organization of compressed data



PROTOTYPES and TESTS
under evaluation (E. Lyard)

Raw data flow → Computing model

- **First step – evaluation of data acquisition rates & CPU needs**
- **Second step - related Hardware resources needs**
  - **Data storage volume**
  - **Average and peak processing power**
  - **Network bandwidth**
- **Third step - Rough costs evaluation**
- **Fourth step - Computing models study**
  - **Centralized versus distributed**
  - **Distributed datacenters, Grid, Cloud, …**

First step – evaluation of data acquisition rates & CPU needs

Based on MC simulations (trigger-task), three data acquisition rates have been evaluated:

- **Minimum : no waveform sampling (2 bytes per pixel) – Cumulated data rate 0.42GB/s (0.3+0.12)**
- **Close to minimum: Waveform sampling for selected pixels – 0.55GB/s (0.4+0.15)**
- **Maximum : Waveform sampling for all camera pixels – 9.8GB/s (8+1.8)**

For IC infrastructure evaluation, three scenarios have been studied:

- **Close to minimum: Waveform sampling for selected pixels -0.55GB/s**
- **Intermediate: During a commissioning phase of 3 years from construction start a raw data rate ~10 times larger than the baseline data rate – 5.5GB/s during commissioning phase**
- **Maximum: Full wavelength – 9.8GB/s**

Based on HESS CPU computing needs to be able <u>to process one hour of raw data</u>, the CTA computing needs are extrapolated:

- **3 Million HS06.sec for HESS 1,**

=> 43 Million HS06.sec for CTA

Note: HS06 definition : HEP-wide benchmark for measuring CPU performance

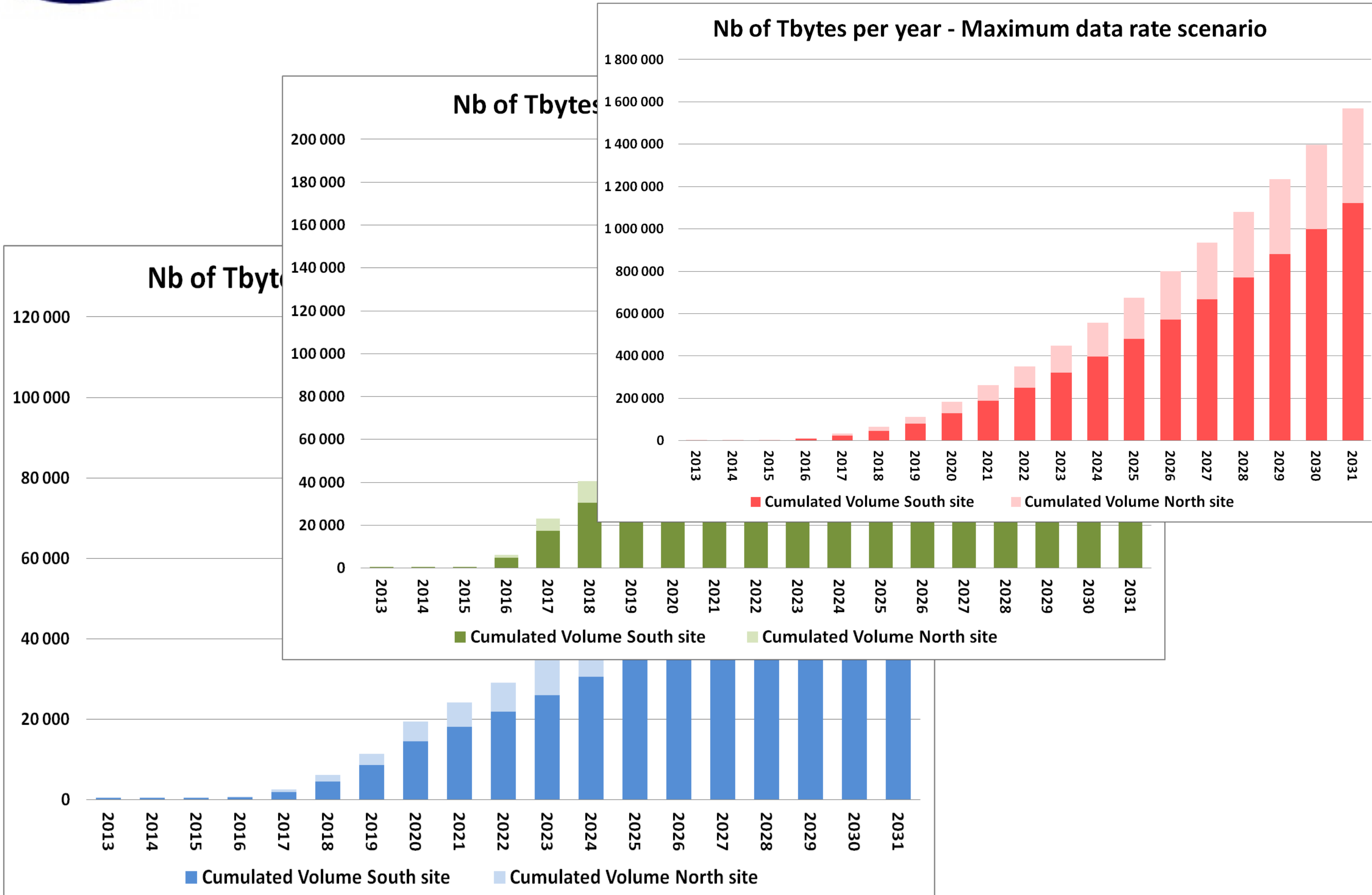
Computing model and datacenter (Step 2)

Second step - related Hardware resources needs

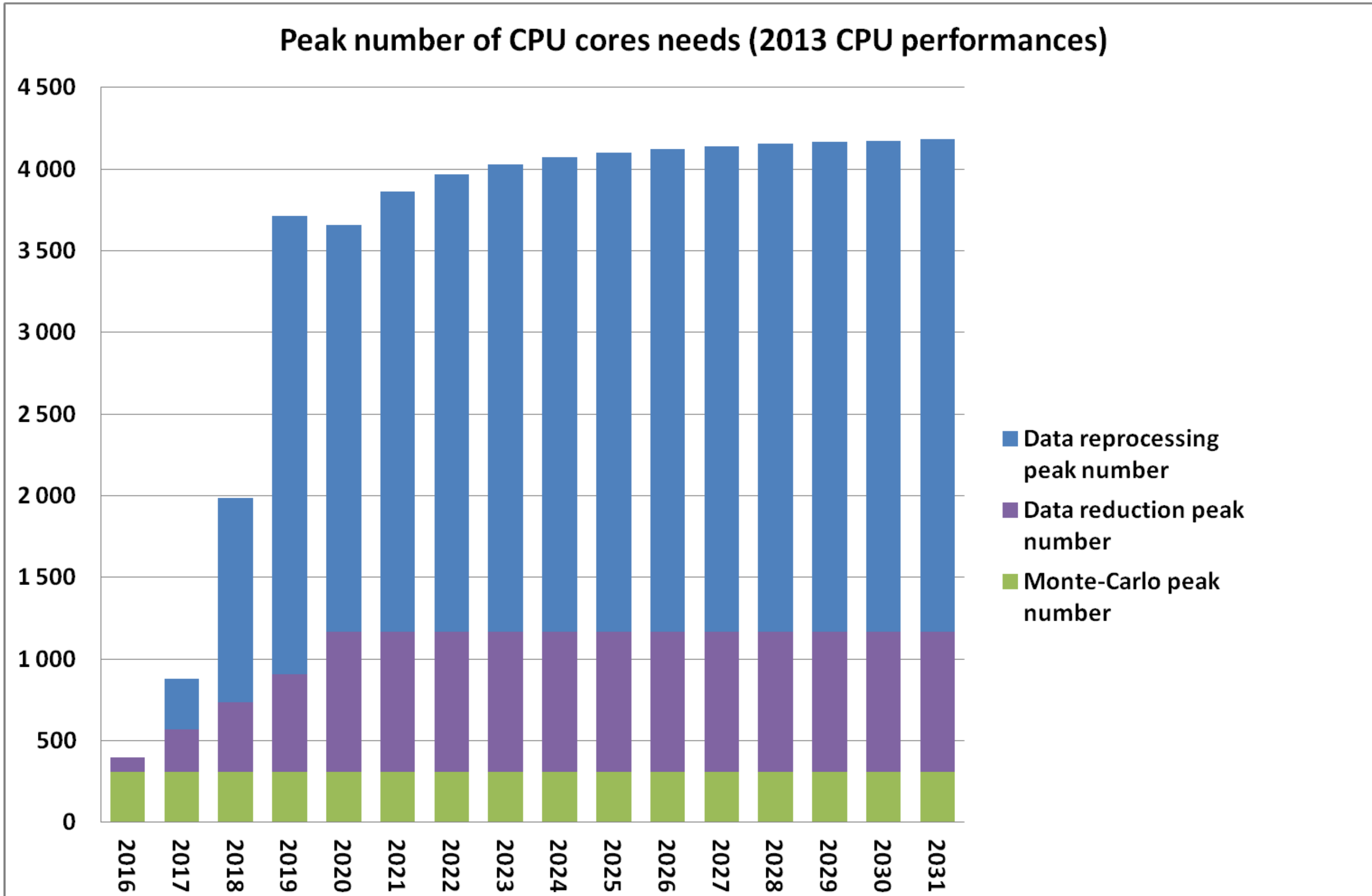
Assumptions/Requirements:

- **1314 hours** of data taking per year, maximum **12 hours per day**
- **Monte-Carlo data volume = yearly cumulated data volume** (minimum 500TB & maximum 10 000TB)
- **Reconstructed data /Raw data volume = 20%**
- All data must be **archived during the CTA lifetime**
- **Annual Monte-Carlo computing need = 10% of annual raw data computing need**
- **Reprocess 1 year of raw data in one month** & **one reprocessing per year**
- **Transfer of daily maximum raw data in less than 10 days** (B-DATA-0960)

Computing model and datacenter(Step 2)

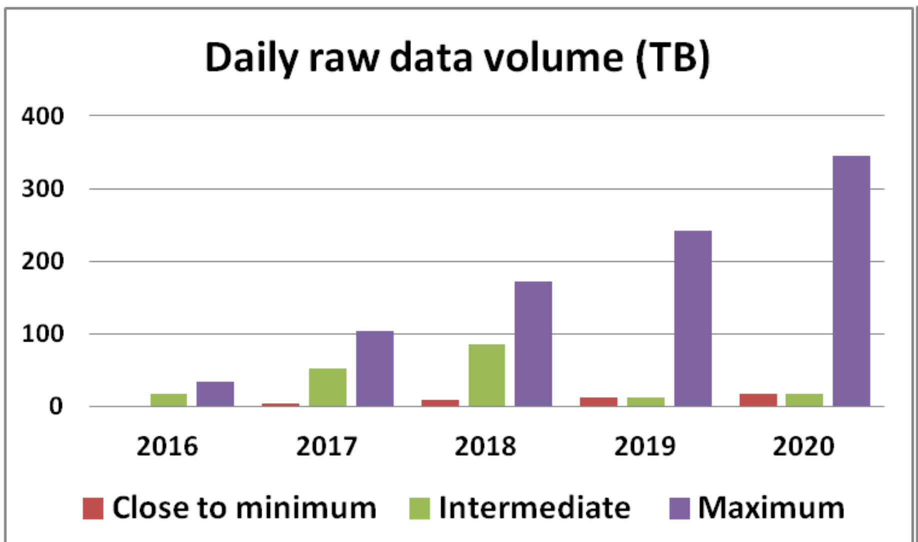


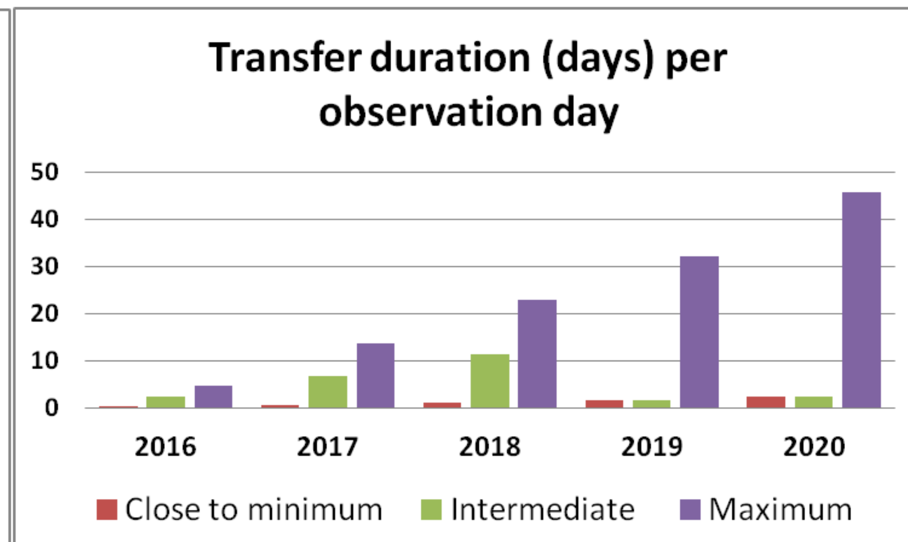
Computing model and datacenter (Step 2)

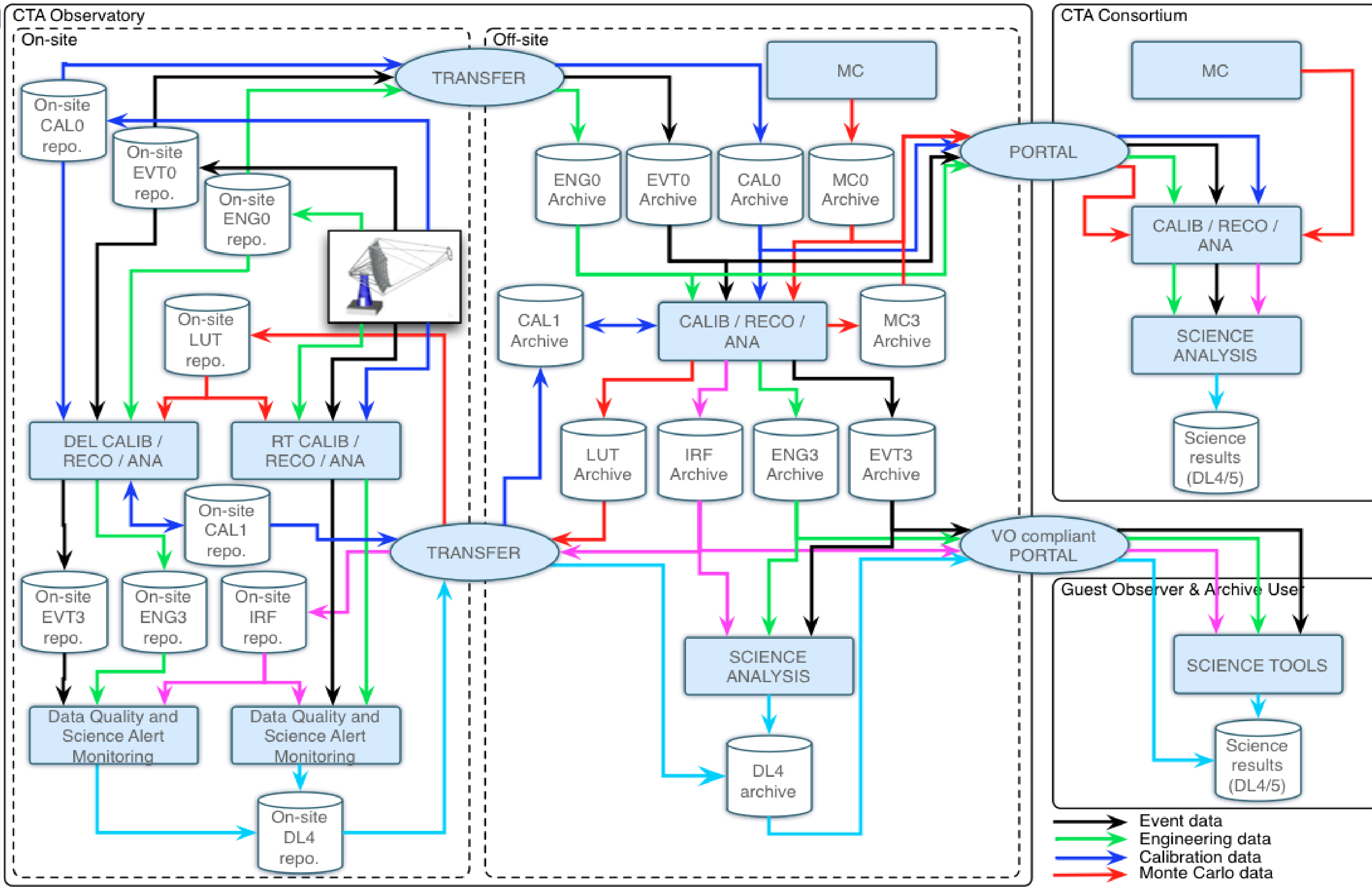


Impact on delay for transferring one day of raw data with a 1Gb/s Network bandwidth:

- For the **close to minimum** scenario: **48 hours** for the South site
- For the **Intermediate** scenario: **48 hours for the South site from 2020** but almost **12 days the third year** of commissioning
- For the **maximum** data rate scenario is more than **48 days**.







A preliminary functional model of the CTA data flow

CTA and Big Data ?

- Which part of data to used :
  - Full set
  - Meta data
  - Engineering data
  - Slow Control data