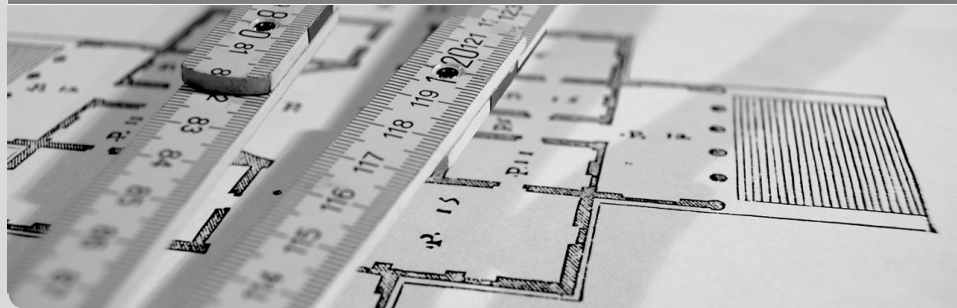


Current status of data center for cosmic rays based on KCDC

GRID-2018, Dubna

Dmitriy Kostunin, Victoria Tokareva | September 12, 2018

INSTITUTE FOR NUCLEAR PHYSICS (IKP)



- Вводная:
 - astroparticle physics и как это всё дофига важно
 - тренды: больше станций или общие данные?
 - итог: совместная российско-немецкая инициатива

The secret life of the astroparticle data

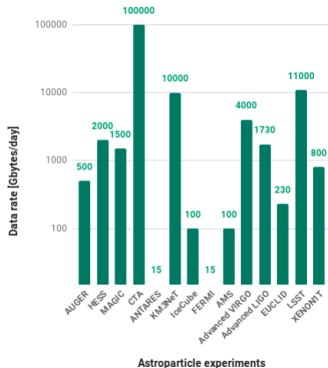
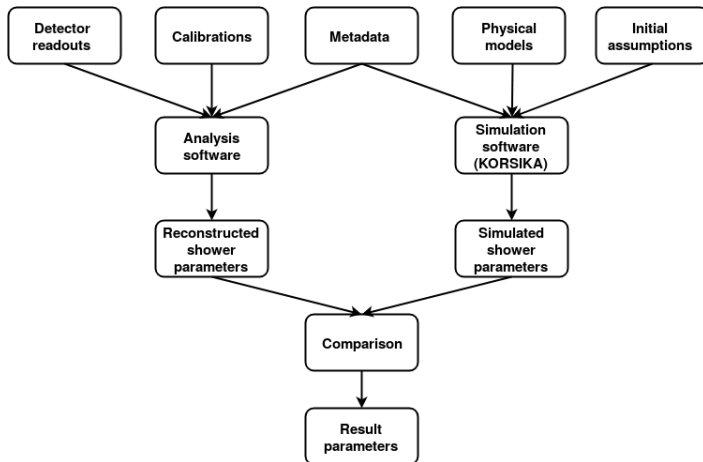


Figure: Modern astroparticle experiments data rate [Gbytes/day], source: APPEC brochure on Computing, 2016

- Wide range of experiments;
- Looking at the same sky with different eyes: different detectors, different reactions under the study;
- Common data rate for astrophysical experiments all together is a few PBytes/yearly, which is comparable to the current LHC output ([ССЫЛКА на брошюру!!!](#))
- Deep learning is coming...
- Need for collaboration

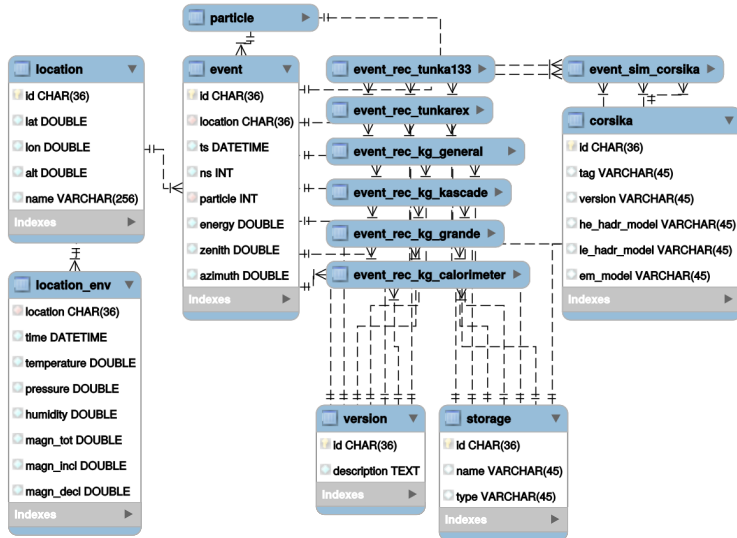
Схема - проверить!)



- Storage
 - Different data formats (depends on available detectors)
 - Common metadata format (e.g. time, location, atmospheric conditions)
- Analysis
 - Special software and system environment for each experiment
 - Virtualization could be useful
- Mapping
 - Calibrations provided by experiments
 - Various models
 - Implementations are stored
- Access
 - Benefits of open access
 - Education and outreach
 - Which infrastructure to use?

- Each experiment has its own detectors and provides its own format of raw events
- Events share common metadata that is stored in the database
- Some experiments are situated at remote locations with no or limited internet access, data are transferred periodically to servers on tapes or disks
- With growing data sizes, distributed storage could be useful

Data storage II



- Some software is unique for each experiment (e.g. detector calibrations, reconstruction based on raw detector readouts) and may even require dedicated system environment
 - Virtualization could be useful
- Some software is common for all EAS experiments, or is planned to be (e.g. software for shower simulation like KORSIKA)
 - Distributed computing could be useful

Содрять слайд 8 у Костюнина

Mapping of air-shower observables



Observables

- "Direct" D (e , μ , γ , \hat{C} , radio, etc)
- "Reconstructed" R (E , X_{\max} , N_{μ}/N_e , etc)
- "Virtual" V (EM component, hadr.component, $\ln A$, etc)

Mapping

- Calibration (hardware response) $D \rightarrow R$: provided by experiments, but can be change on the flight by the *end-user*
- Models $V \Leftarrow R$: provided by experiments (reconstruction methods), theorists (proposed methods), *end-users*
- Mapping can be implemented, indexed with DOI and cited in papers
- *Implementation is stored and accessible using proposed framework*

Proof-of-principle: $E_{\text{KG}} \Rightarrow V_{\text{EM}} \Rightarrow D_{\text{radio}} \Rightarrow V_{\text{EM}} \Rightarrow E_{\text{Tunka}}$

Navigation icons: back, forward, search, etc.

- Benefits of open access
- Education and outreach
- Which infrastructure to use?

TODO

Что ещё не хватает?

KRAD: Karlsruhe-Russian Astroparticle Data Life Cycle

Как можно догадаться по названию, являет совместным
Российско-немецким проектом, включающим в себя с российской
стороны такие институты как ... (перечислить, какие) Суть:
создание единого центра обработки астрофизических данных.

- сравнение KASCADE и Tunka (почему должно быть интересно объединить данные, проблемы и решения):
 - физика: в чем разница собираемых данных?
 - разница по метаданным
 - организация хранения: у KASCADE есть KCDC, и всё крайне аккуратно организовано (рассказать, как), у Tunka все в процессе (оказывается, есть крутой слайд у Костюнина про это (5/13))
 - обработка? общая она или отдельная? в чем разница? как могла бы выглядеть общая схема обработки?
 - как можно было бы организовать совместный быстрый доступ пользователей к данным/инструментам анализа?
 - почему мы считаем, что эксперименты можно определить, и почему именно Tunka()?
- Как мы думаем, это можно было бы сделать:

- MWS как идея
- какие-нибудь схематические догадки о том, как это все касается нас
- job workflow
- какую систему будем юзать?
- А что уже сделано к наст моменту?
 - общая схема KCDC
 - astroparticle.online
 - описание метаданных
- Conclusion: мы собрались делать большое дело, у нас есть богатая история, много планов и даже чуть-чуть из них уже сделано. Следующим шагом проекта станет...