

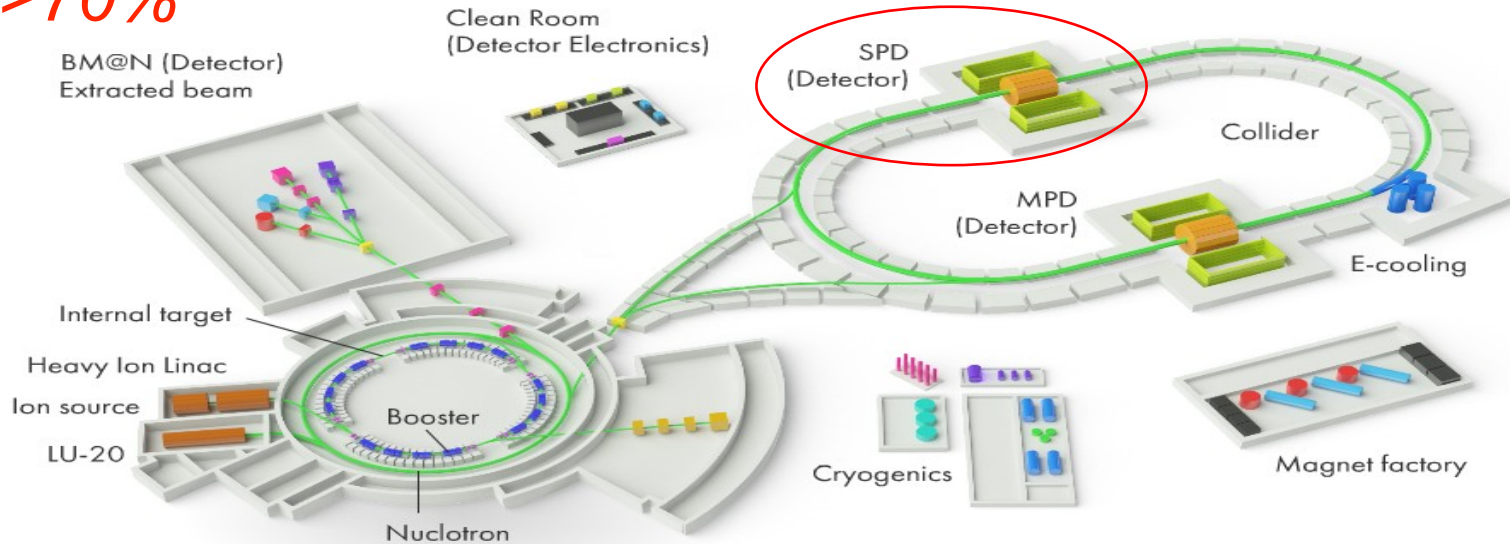
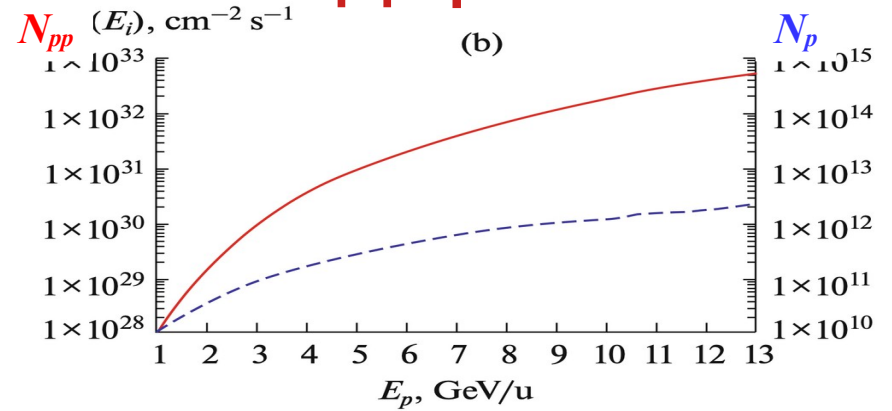
Программное обеспечение и компьютеринг эксперимента SPD

Жемчугов Алексей Сергеевич
ОИЯИ, г.Дубна

Эксперимент SPD на коллайдере NICA

$p \uparrow p \uparrow : \sqrt{s} \leq 27 \text{ GeV}$
 $d \uparrow d \uparrow : \sqrt{s} \leq 13.5 \text{ GeV}$
 $d \uparrow p \uparrow : \sqrt{s} \leq 19 \text{ GeV}$

U, L, T
 $|P| > 70\%$



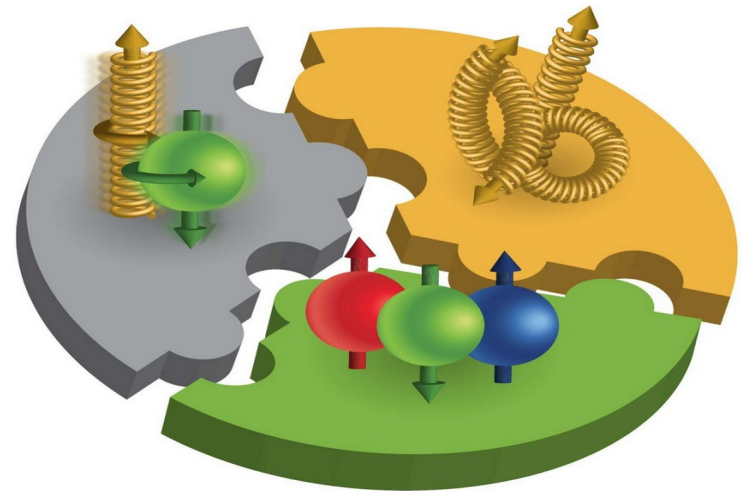
Программа исследований

- SPD — универсальная установка для всестороннего изучения глюонной составляющей протона и нейтрона при больших x
 - Прямые фотоны
 - Чармоний
 - Состояния с открытым чармом
- Другие явления, связанные со спином
- Другие физические задачи

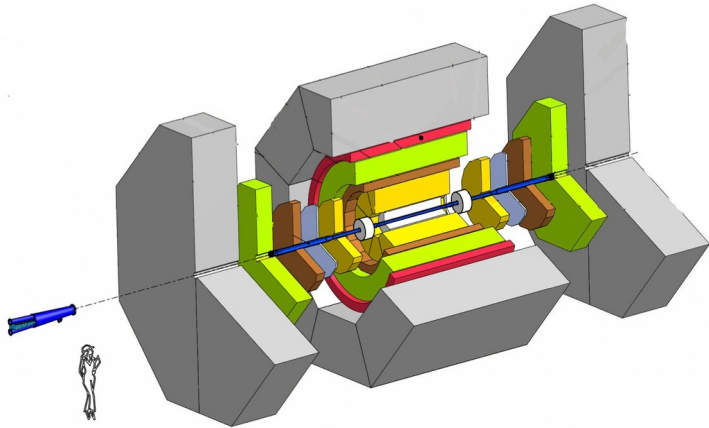
Подробнее:

Prog.Part.Nucl.Phys. 119 (2021) 103858

arXiv:2011.15005



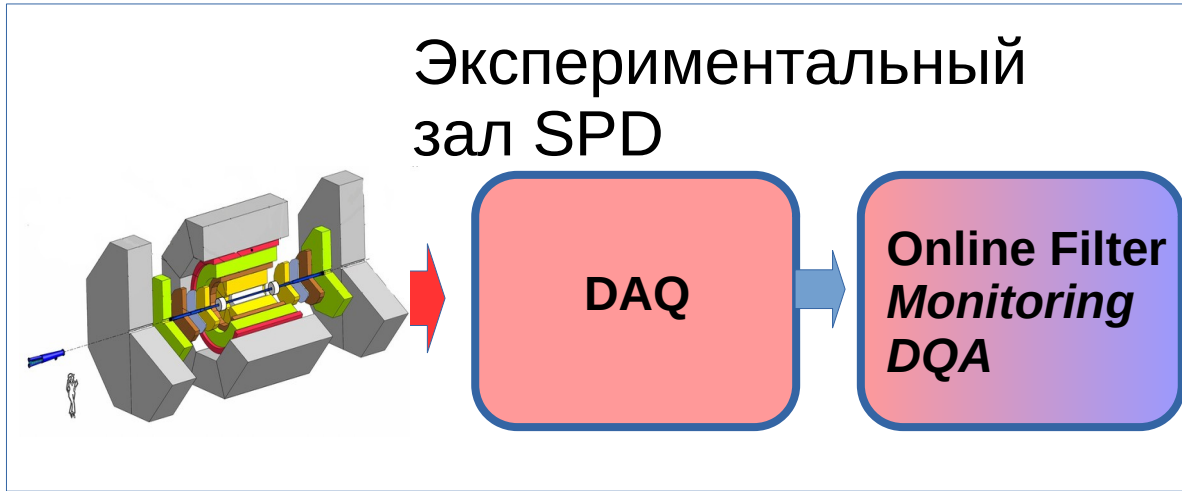
Установка SPD как источник данных



- Пересечение сгустков каждые 76.3 нс = 13 MHz
- Частота событий ~ 3 MHz (при проектной светимости 10^{32} $\text{cm}^{-2}\text{s}^{-1}$)
- 20 GB/s (или 200 PB/год (сырые данные), $3 \cdot 10^{13}$ событий/год)
- Отбор полезных событий требует восстановления импульса и вершины \rightarrow **простой триггер невозможен**

Установка SPD, имея средние размеры, оказывается источником больших данных!

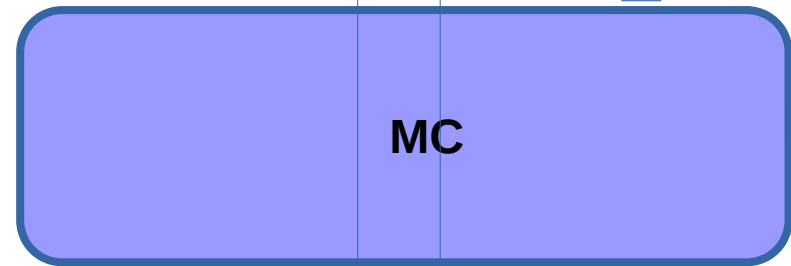
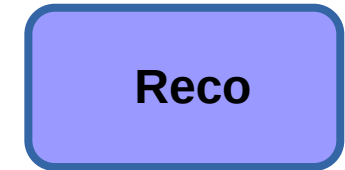
Конвейер обработки данных



Tier-0: ВЦ ОИЯИ



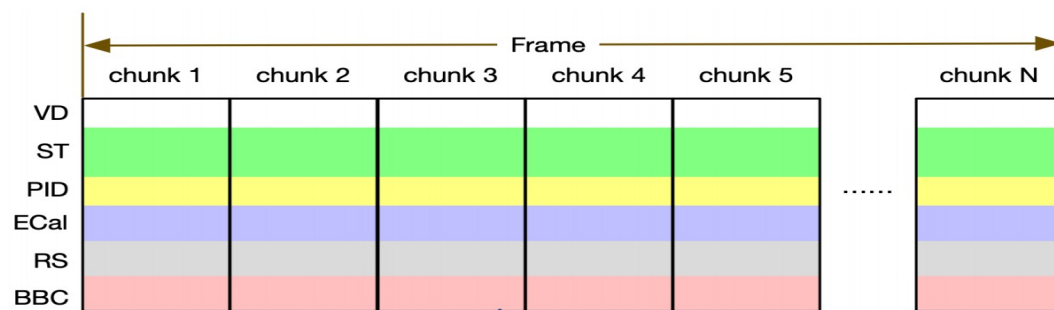
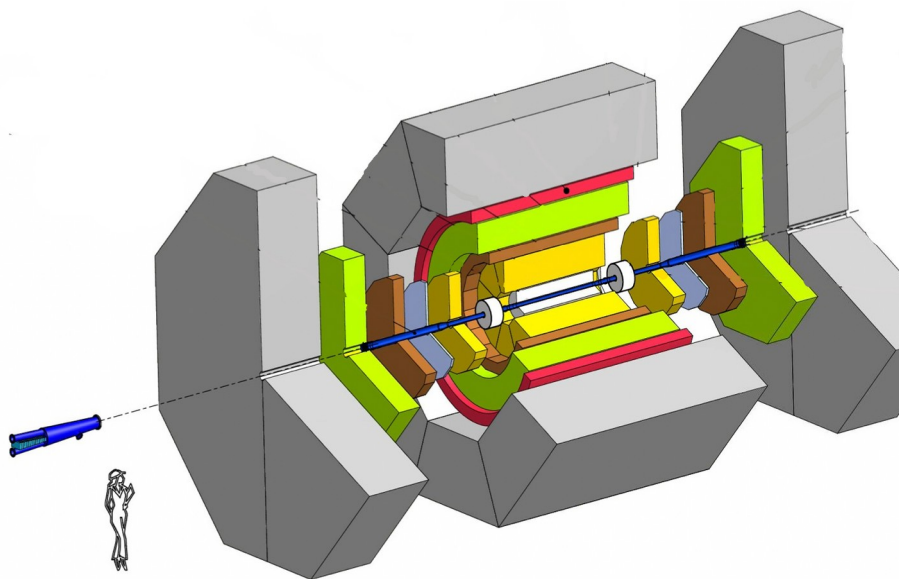
Всемирный Грид



Постоянное уменьшение количества данных:

- DAQ: подавление шума
- Онлайн-фильтр: частичная реконструкция и отбор полезных событий
- Оффлайн: полная реконструкция, моделирование, долговременное хранение и анализ данных

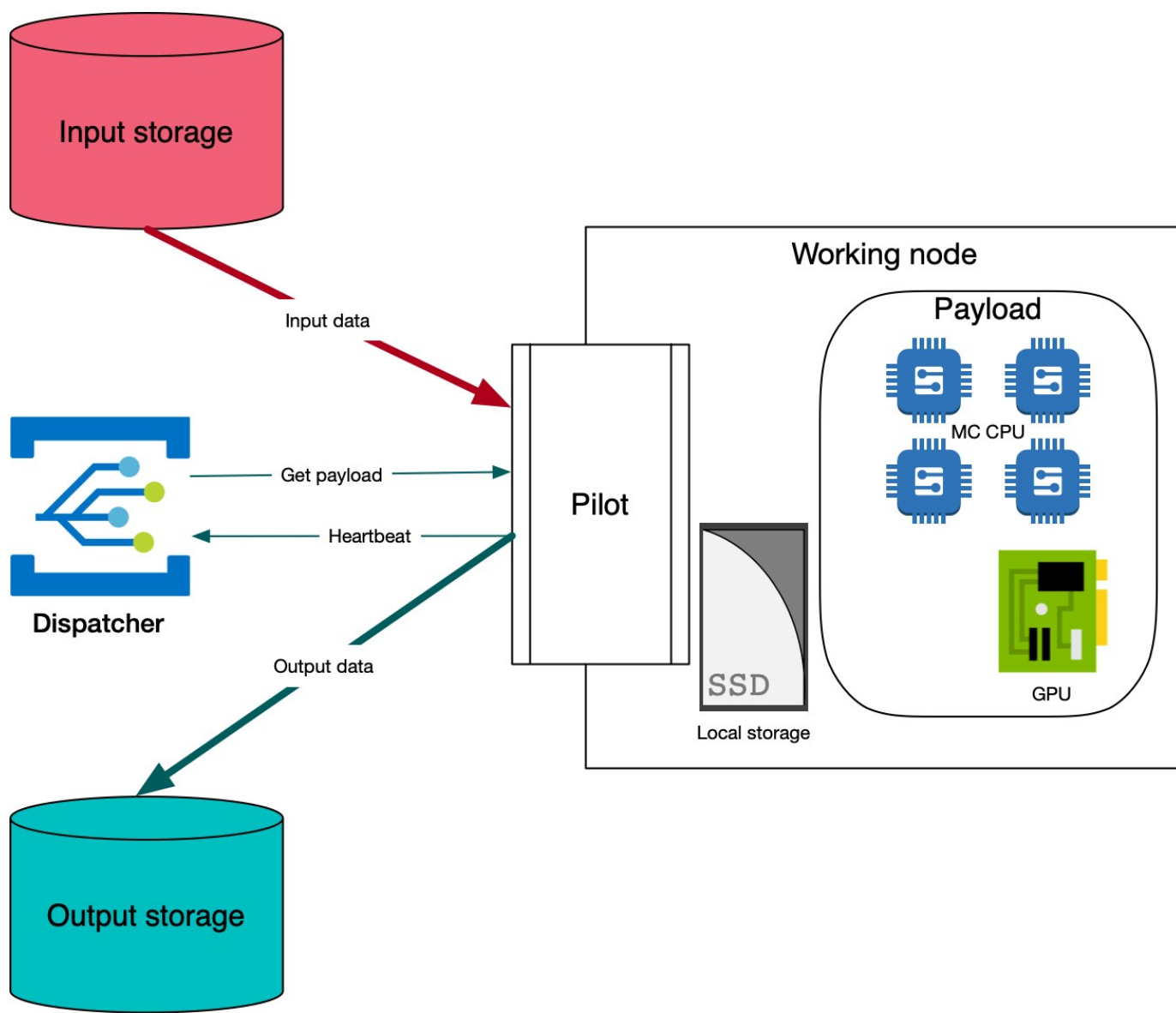
Бестриггерная система DAQ



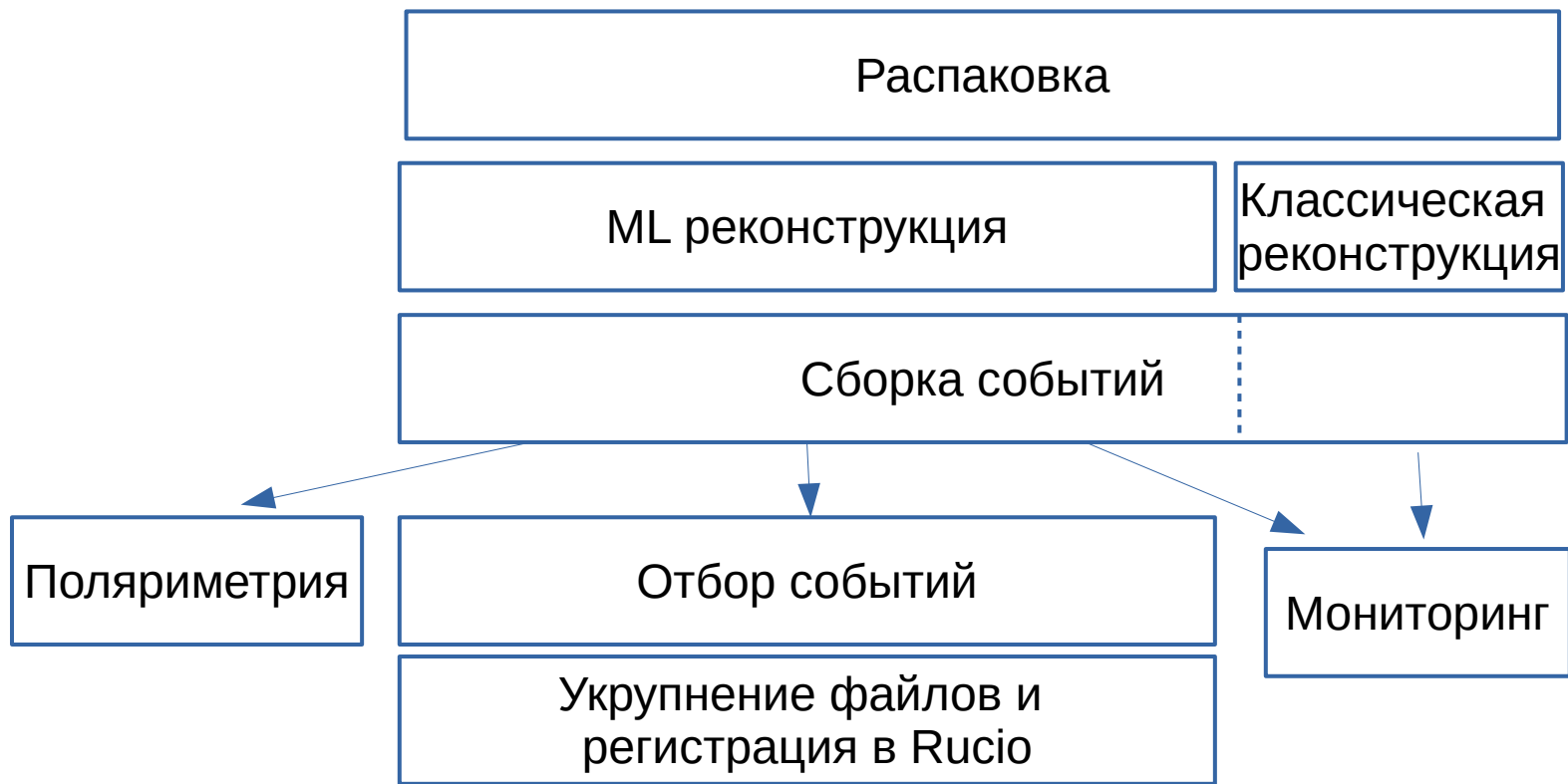
Онлайн-фильтр данных

Высокопроизводительный гетерогенный кластер

- Частичная реконструкция
 - Восстановление треков и вершин
 - Поиск кластеров в калориметре
 - Восстановление событий
 - Отбор
 - несколько потоков
- Машинное обучение — ключевой подход
- Как быть с систематической погрешностью?
- Мониторинг установки и контроль качества данных
 - Измерение поляризации пучка



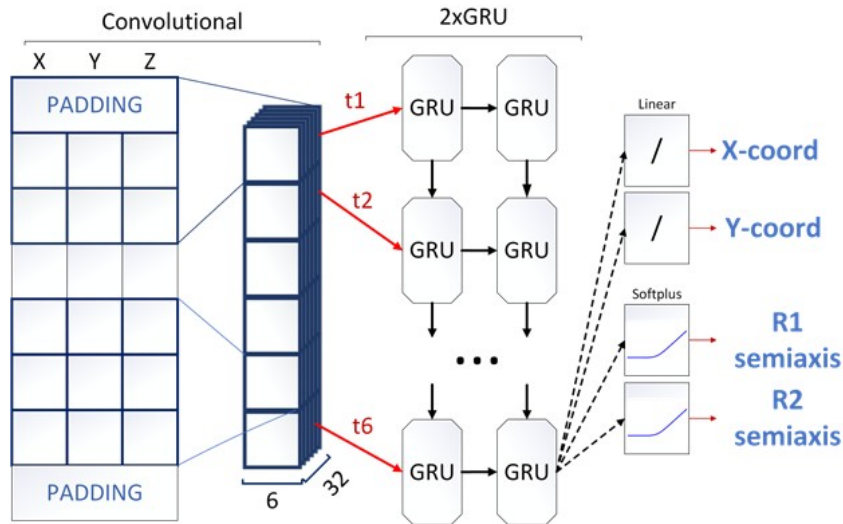
Приложение



Пример: TrackNETv3 для поиска треков

<https://arxiv.org/abs/2210.00599>

D. Rusov et al, talk at PCT'2023



- Сеть предсказывает область, в которой следует искать продолжение трека
- Если точка найдена, она добавляется в трек-кандидат и процедура повторяется
- Фактически воспроизводится фильтр Калмана: параметры трека предсказываются сетью после соотв. обучения
- **Обобщение? Устойчивость?**

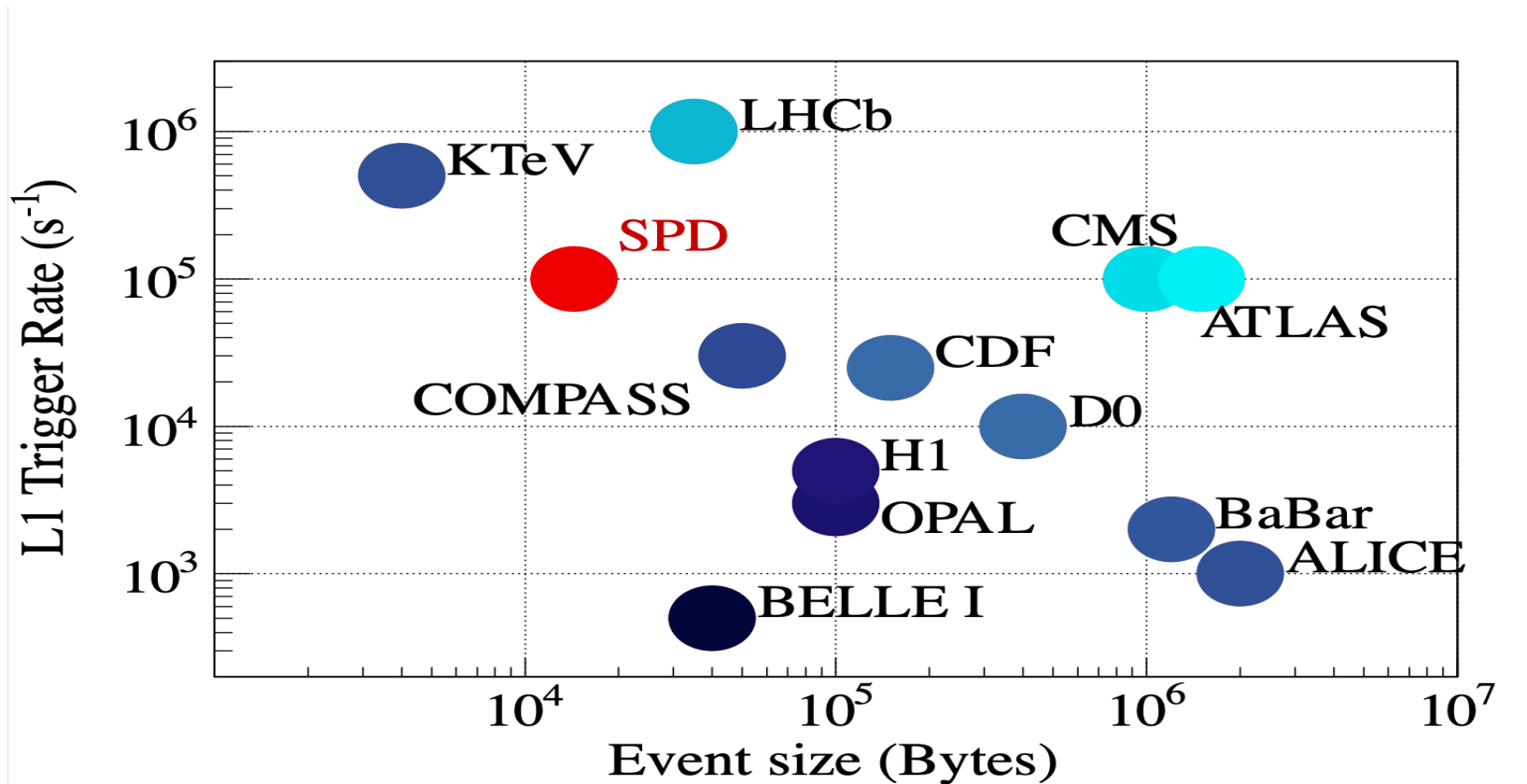
Неэффективность детектора?

Срезы по 40 событий

Эффективность (recall) (%)	96,54
Чистота (precision) (%)	94.75
Производительность (slices / sec)	63.74 (* 40 = 2549.6)

Предварительный
результат

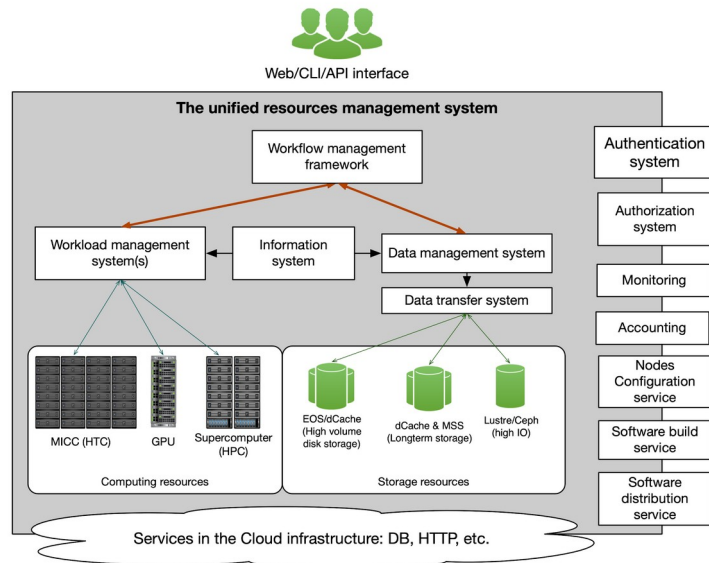
После фильтрации данных



Распределенные вычисления

К 2030 году:

- до 30 PВ хранилища
- до 1.5 Pflops вычислительной мощности



Все основные компоненты уже созданы в рамках компьютеринга LHC:

- *Информационная система: CRIC*
- *Управление задачами: PANDA*
- *Управление данными: RUCIO и FTS*
- *Распространение ПО: CVMFS*

Требуется адаптация к модели обработки данных SPD

Автономное ПО

SpdRoot — основной инструмент для моделирования и подготовки физической программы в ближайшие годы.

Однако, для обработки экспериментальных данных его нужно очень серьезно дорабатывать.

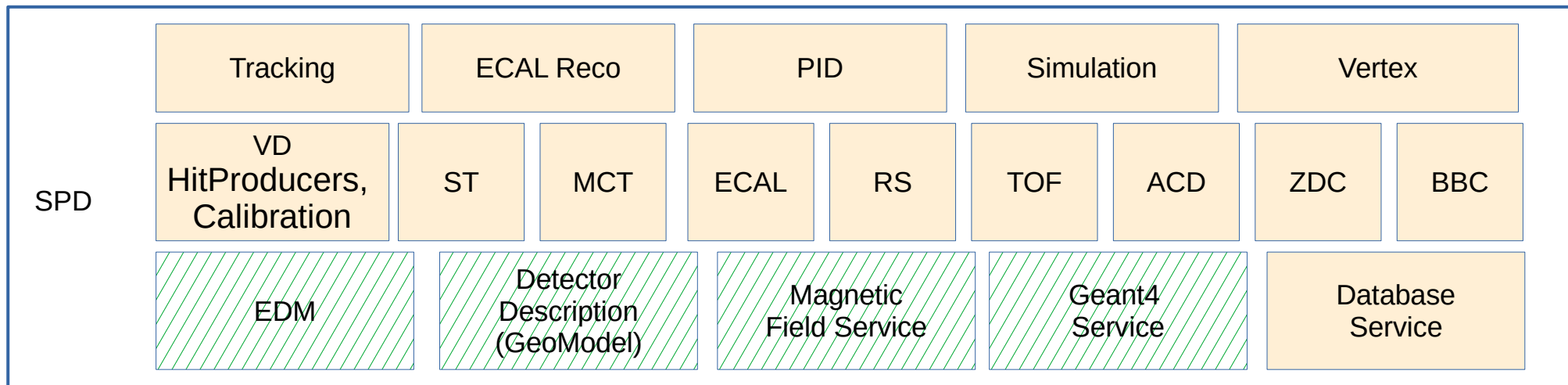
Или искать альтернативу

В настоящее время в SPD активно развивается новое ПО, основанное на фреймворке Gaudi

- **Gaudi-v36r9**
- **CentOS7, gcc-11.3.0, python-3.9.6**
- **Распространение в виде контейнера docker**

Новое ПО

User's code



Gaudi

External libs (18 for Gaudi + cmake/gcc/python + Geant4 + OTHER)

Итог

- Онлайн-фильтр, основанный на машинном обучении, является ключевым элементом эксперимента SPD
- Большие потоки (и объемы) данных требуют широкого применения параллельных вычислений
 - многоядерные CPU, GPU, FPGA ...
- Большое число интересных задач при разработке автономного ПО, основанного на фреймворке Gaudi
- Большое число интересных задач при создании системы распределенных вычислений
 - Управление задачами
 - Управление данными
 - Информационная система