



Национальный исследовательский ядерный университет «МИФИ» Институт
ядерной физики и технологий
Кафедра физики элементарных частиц №40

Отчет о преддипломной практике на тему:

СИСТЕМА УПРАВЛЕНИЯ ПРОЦЕССАМИ ОБРАБОТКИ В ГЕОГРАФИЧЕСКИ РАСПРЕДЕЛЕННОЙ СИСТЕМЕ ОБРАБОТКИ ДАННЫХ SPD

Студент

Н.Г. Монаков

Научный руководитель Е. Ю. Солдатов

Научный консультант А.Ш. Петросян

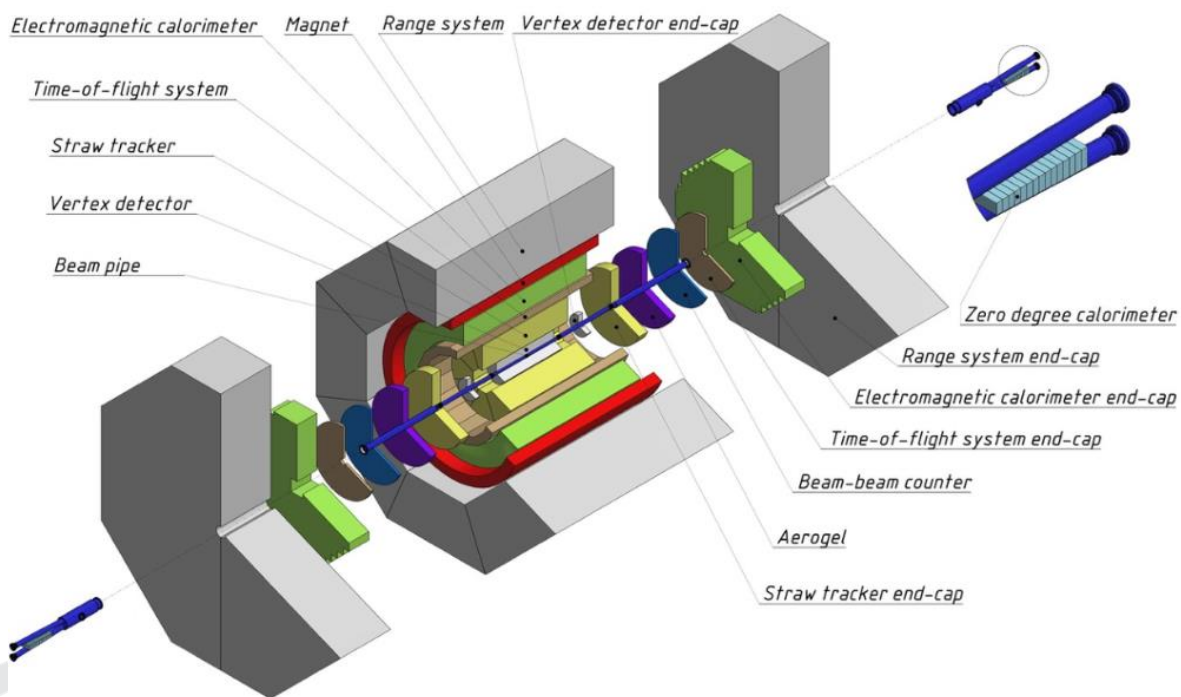
Москва 2025



Эксперимент SPD

Установка SPD, размещаемая в одной из двух точек пересечения пучков коллайдера NICA, предназначена для всестороннего изучения **спиновой структуры протона и дейтрона**.

- $\sqrt{s}$ до 27 ГэВ
 - L до $10^{32} \text{ см}^{-2} \text{ с}^{-1}$
- $\Rightarrow 10^{13} \text{ событий} \times 10\text{-}15 \text{ КБ}$
 $\approx 10^2 \text{ ПБ/год}$



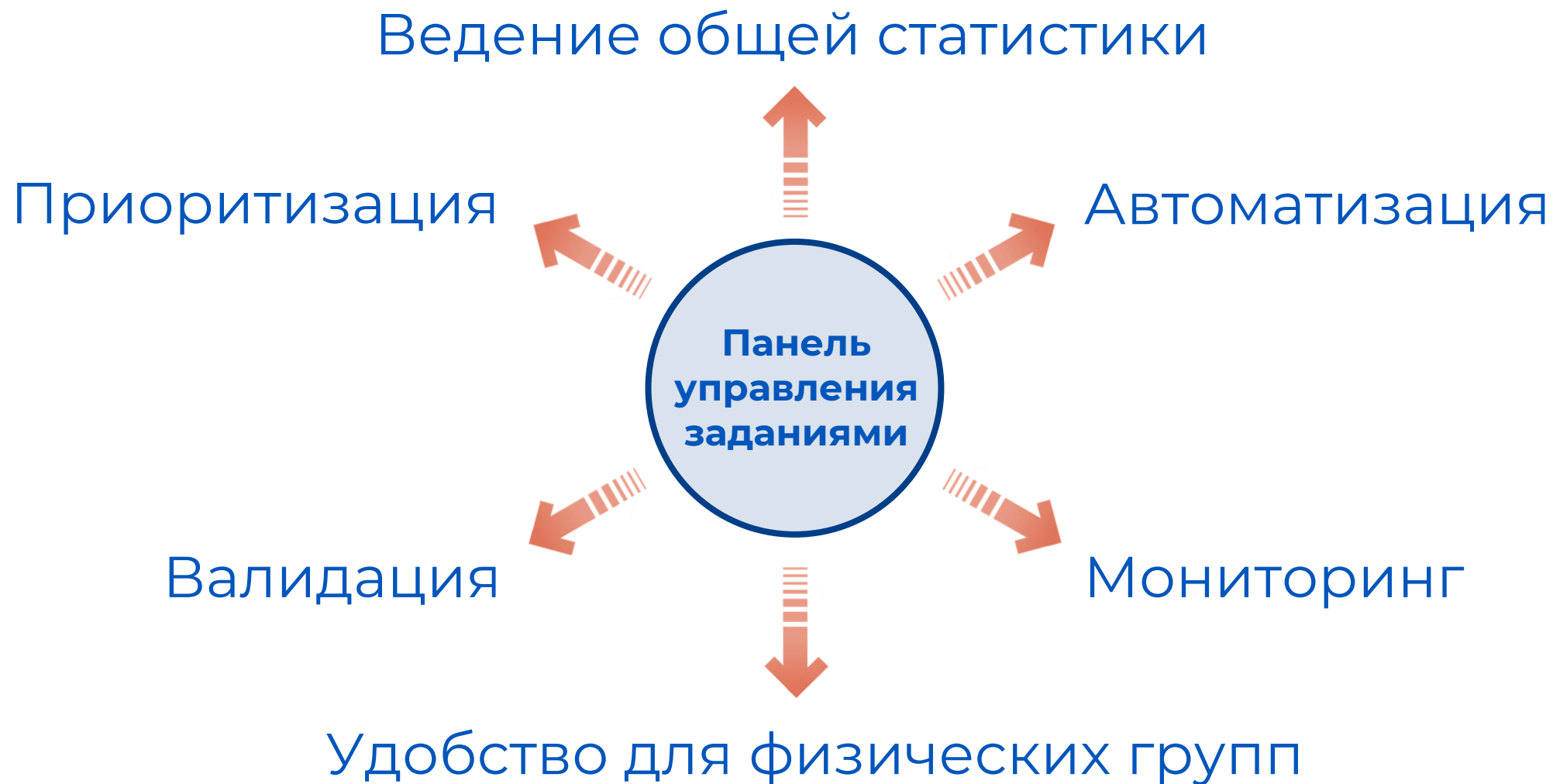
Основное внимание будет уделено изучению

- поляризованной глюонной компоненты в реакциях инклюзивного рождения чармониев
- открытого чарма
- прямых фотонов
- прочих спин-зависимых явлений в p-p / d-d столкновениях

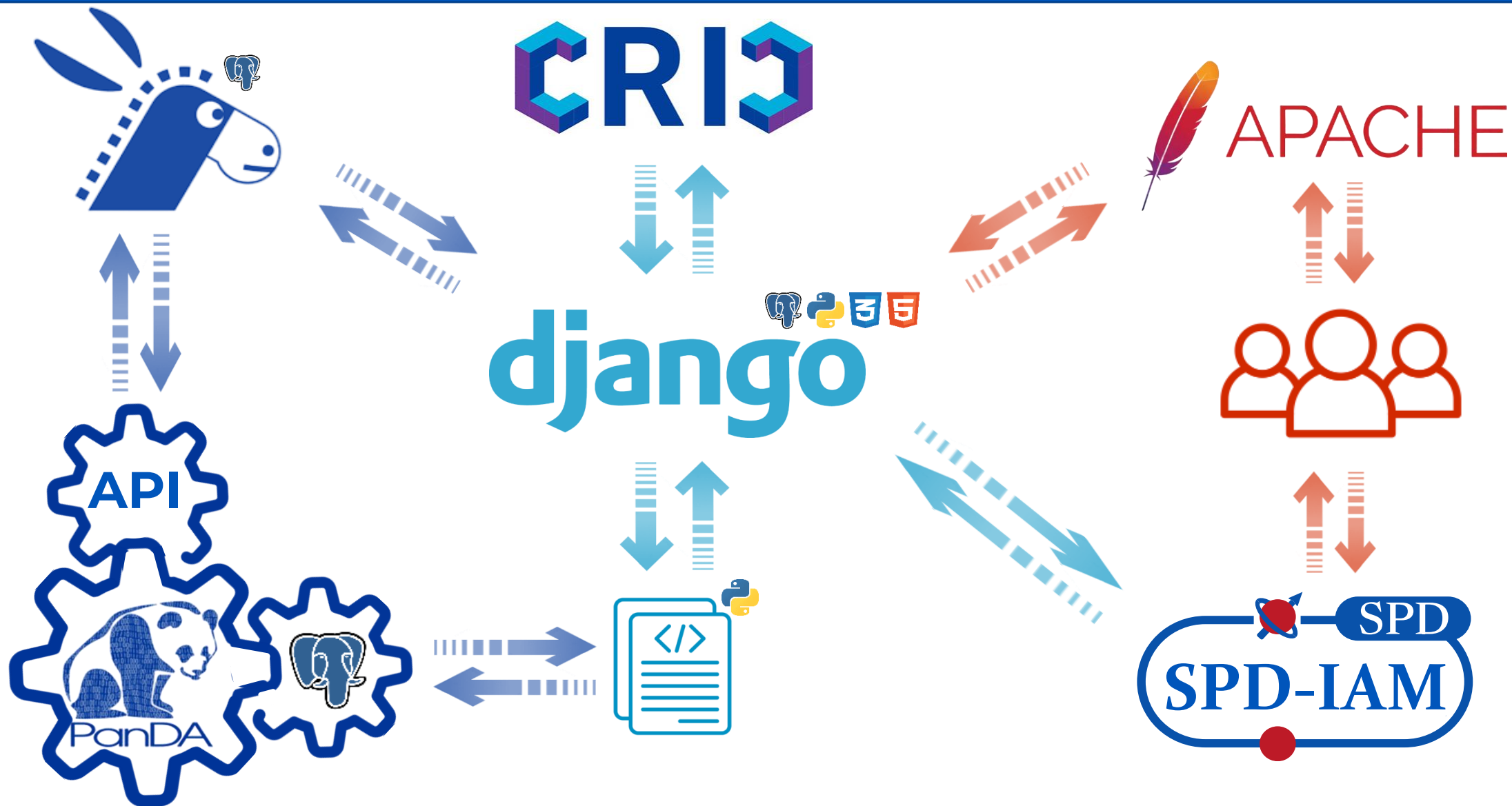
Мотивация создания системы



Текущая концепция сервиса



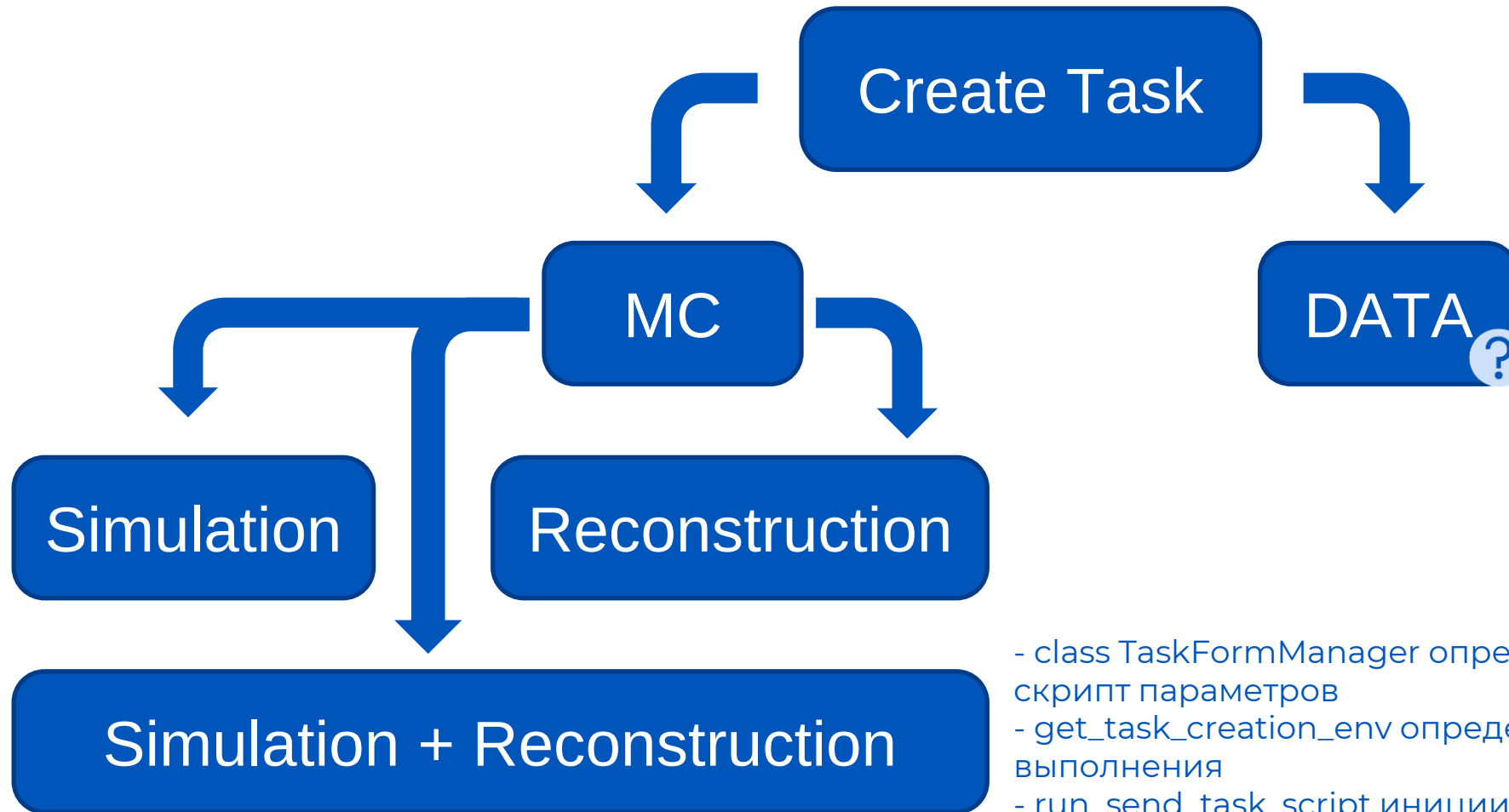
Взаимосвязи компонентов и систем



Проделанная работа

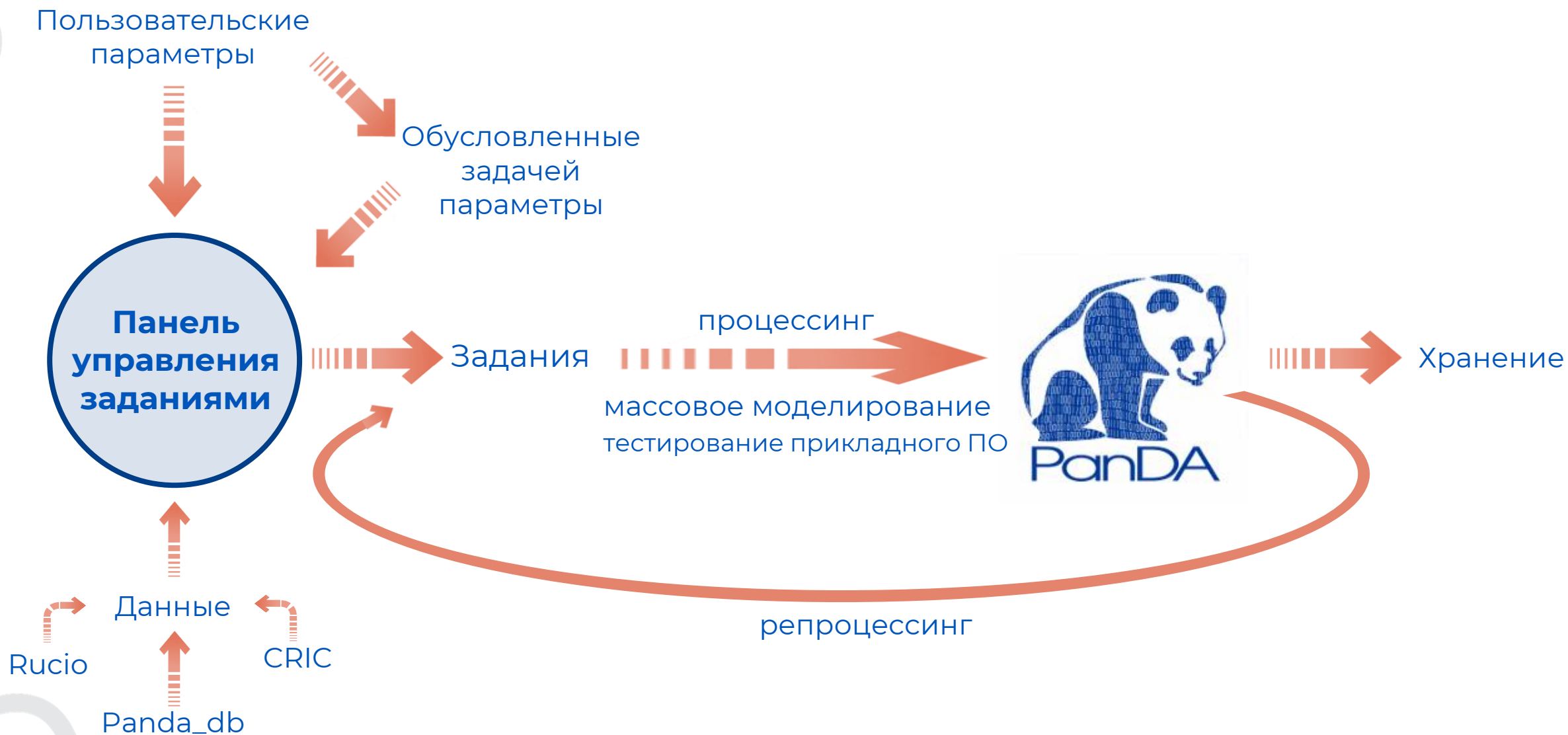
- Создан прототип веб-приложения
 - аутентификации пользователя с использованием `spd-iam`
 - создание заданий в соответствии с принятыми конвенциями
 - отслеживание состояния заданий с возможностью фильтрации и сортировки
 - настроены декораторы и `middleware`
- Начата реализация автоматизирующих скриптов
- Налажено взаимодействие с PanDA WMS с использованием JWT
- Логирование внутренних и внешних действий и событий

Создание заданий 1/3



- class TaskFormManager определяет набор передаваемых в скрипт параметров
- get_task_creation_env определяет переменные среды для выполнения
- run_send_task_script инициирует выполнение скрипта, ведет логгирование, передает данные, необходимые для цепочек

Создание заданий 2/3



Создание заданий 3/3

Simulation

Task Creation

Task name:

Output dataset name:
[Naming convention here](#)

Total events:

Events per job:

Cloud:

Data disk:

Skip scout: ☐

Offset:

Path to execution files:
smth like -> /cvmfs/spd.jinr.ru/production/MC/minbias-P8-spdroot417-dev.10GeV.V01

Path to container:
smth like -> /cvmfs/spd.jinr.ru/images/spdroot-dev-4.1.7.sif

Create task

Reconstruction

Task Creation

Task name:

Input dataset name:
[Naming convention here](#), note that no extension expected

Output dataset name:
[Naming convention here](#)

Files per job:

Cloud:

Data disk:

Skip scout: ☐

Offset:

Path to execution files:
smth like -> /cvmfs/spd.jinr.ru/production/MC/minbias-P8-spdroot417-dev.10GeV.V01

Path to container:
smth like -> /cvmfs/spd.jinr.ru/images/spdroot-dev-4.1.7.sif

Create task

Мониторинг

Select field

Enter value

Filter

Task ID	Task name ↑ ↓	Parent ID	Creator	Status	Done jobs	Default/Current priority	Total events	Submit time ↑ ↓	Start time ↑ ↓	End time ↑ ↓
370	PROD2025-017.SIM	370	Elena Zemlyanichkina	running	2	900/900	0	17:03, 20 May 2025	17:11, 20 May 2025	None
369	PROD2025-016.RECO	369	Elena Zemlyanichkina	finished	9999	900/900	0	03:31, 10 May 2025	08:33, 10 May 2025	22:12, 13 May 2025
368	PROD2025-016.SIM	368	Elena Zemlyanichkina	done	10000	900/900	0	15:08, 08 May 2025	15:16, 08 May 2025	22:07, 09 May 2025
367	PROD2025-015.RECO	367	Elena Zemlyanichkina	finished	9992	900/900	0	13:16, 07 May 2025	13:24, 07 May 2025	18:58, 09 May 2025
366	PROD2025-015.SIM	366	Elena Zemlyanichkina	done	10000	900/900	0	04:26, 06 May 2025	07:01, 06 May 2025	12:52, 07 May 2025
365	PROD2025-014.RECO	365	Elena Zemlyanichkina	finished	9991	900/900	0	23:13, 04 May 2025	23:21, 04 May 2025	20:37, 06 May 2025
364	PROD2025-014.SIM	364	Elena Zemlyanichkina	done	10000	900/900	0	14:21, 03 May 2025	14:28, 03 May 2025	21:35, 04 May 2025
363	PROD2025-013.RECO	363	Elena Zemlyanichkina	finished	9992	900/900	0	10:03, 02 May 2025	10:11, 02 May 2025	00:51, 04 May 2025
362	PROD2025-013.SIM.2	362	Elena Zemlyanichkina	done	10000	900/900	0	23:35, 30 Apr 2025	23:43, 30 Apr 2025	08:57, 02 May 2025
361	PROD2025-013.SIM	361	Elena Zemlyanichkina	aborted	869	900/900	0	10:59, 30 Apr 2025	11:10, 30 Apr 2025	15:13, 30 Apr 2025

Page: 1 2 3 ... 27 28 29

← Текущие задания

Успешные задания



Select field

Enter value

Filter

Task ID	Task name ↑ ↓	Status	Start date	End date	Walltime	Total events	Events per job	Total jobs	Out DS size, GB	Out Log size, GB
369	PROD2025-016.RECO	finished	10 May 2025	13 May 2025	30	None	None	9999	18557.43	5.07
368	PROD2025-016.SIM	done	08 May 2025	09 May 2025	22931	40000000	4000	10000	18365.39	1.92
367	PROD2025-015.RECO	finished	07 May 2025	09 May 2025	23	None	None	9992	18543.30	5.05
366	PROD2025-015.SIM	done	06 May 2025	07 May 2025	24486	40000000	4000	10000	18362.72	1.97
365	PROD2025-014.RECO	finished	04 May 2025	06 May 2025	24	None	None	9991	18540.20	5.05
364	PROD2025-014.SIM	done	03 May 2025	04 May 2025	24612	40000000	4000	10000	18358.86	1.96
363	PROD2025-013.RECO	finished	02 May 2025	04 May 2025	20	None	None	9992	18536.20	5.06
362	PROD2025-013.SIM.2	done	30 Apr 2025	02 May 2025	24899	40000000	4000	10000	18357.95	1.93
359	PROD2025-012.RECO	finished	28 Apr 2025	29 Apr 2025	24	None	None	9993	18546.74	5.08
358	PROD2025-012.SIM	done	25 Apr 2025	26 Apr 2025	23316	40000000	4000	10000	18360.80	1.89

Page: 1 2 3 4 5 6 7

Export csv

Содержание заданий

Моделирование и реконструкция выполняются для p-p столкновений с использованием SpdRoot.

Фреймворк включает:

- геометрическое описание SPD-детектора
- распространение частиц
- упрощенное моделирование отклика детектора
- алгоритмы реконструкции

Во время моделирования и реконструкции были включен:

- вершинный детектор
- straw – трекер
- система определения времени полета
- электромагнитный калориметр
- система определения дальности полета

На текущей стадии разработки сохраняется избыточная информация о взаимодействии частиц внутри детектора, которая является предметом будущей оптимизации размера события, а также на этапах моделирования и реконструкции



Этап	S1 & S2
Энергия	4 - 27 GeV
Поляризация	UU
Количество событий	5 – 40 М

Событий на файл	4000
Initial seed	1-1250 1-5000 5001-10000 1-10000 10001-20000 20001-30000 30001-40000 40001-50000

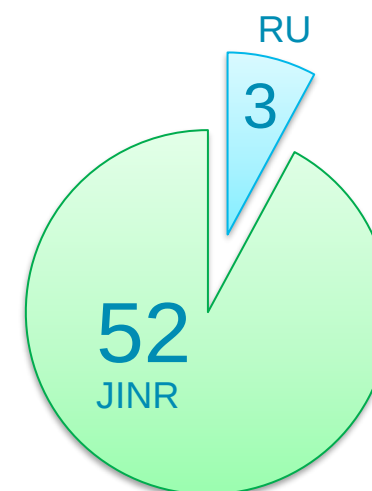
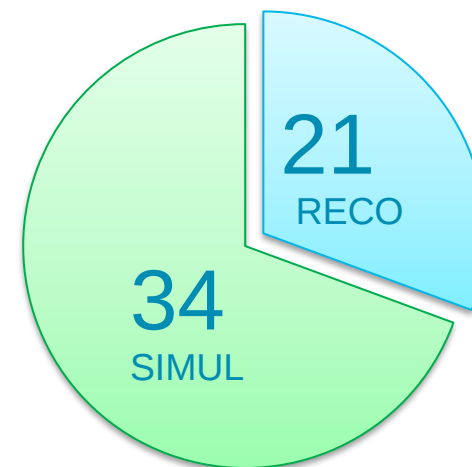
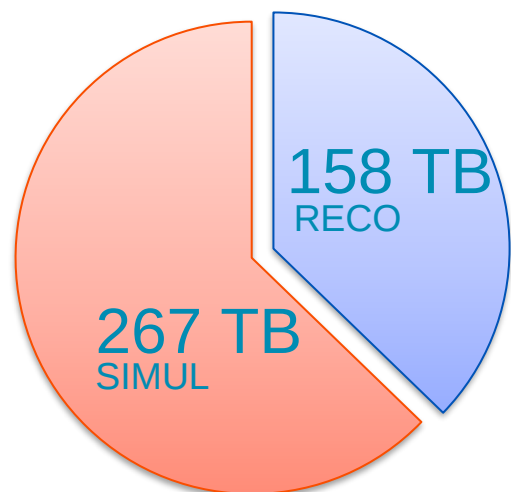
Шаблон наименования датасетов

(DATA SOURCE)(YEAR)_(STAGE):(SHORT DESCRIPTION).(ENERGY)-(POLARIZATION).(PRODUCTION NAME).(TASK TYPE).(VERSION).(EXTENSION)

https://docs.google.com/spreadsheets/d/10uG1p6wPQ_GWe7wkZ_t4RPFLmsHuHxnth5MqM49uOYQ/

Статистика

- Успешно обработано
около 300 тысяч задач 
в рамках 55 заданий 



- Суммарный объем выходных
датасетов – около 425 ТБ 

Текущие задачи и дальнейшие планы

- Дальнейшая автоматизация создания и валидации заданий
- Расширение мониторинга заданий
- Внедрение дополнительного уровня коллекций для больших заданий
- Создание документации
- Развертывание дополнительного сервера для разработки и тестирования



Благодарю за внимание

