

# Прогнозирование сбоев работы систем хранения данных методами машинного обучения

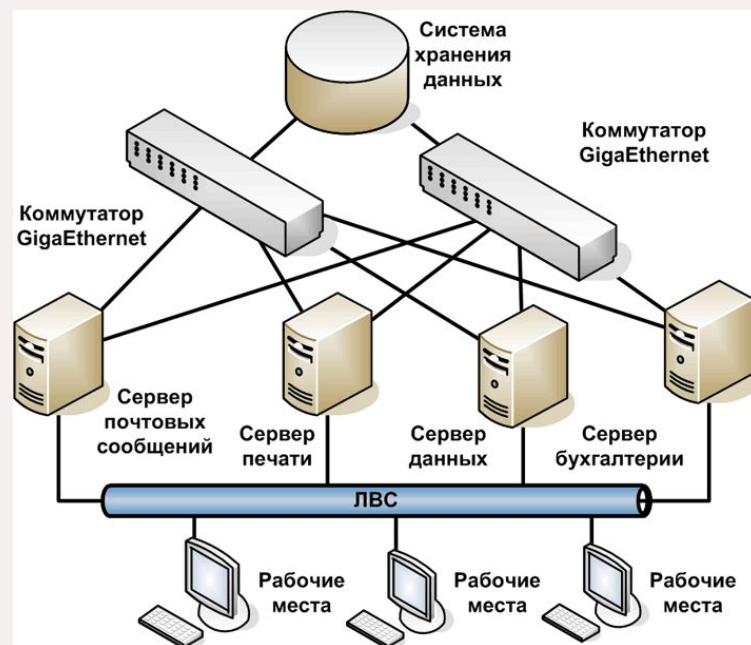
Андрей Сапронов  
LAMBDA ВШЭ, СПбПУ НИЛ «АСПОД»

# Содержание

- Система хранения данных - общие сведения
- Мотивация, цели и задачи
- СХД Tatlin компании Yadro
- Система мониторинга и диагностики
- Моделирование работы СХД
- Алгоритмы диагностики и предсказания сбоев
- Планы и заключение

# Система хранения данных

СХД – это программно-аппаратное решение для надежного хранения, управления и обмена данными.



# Мотивация

Ввиду того, что доступность данных является критически важной для функционирования бизнеса, отказоустойчивость систем хранения данных имеет высокое значение.

К отказам СХД можно отнести:

- Поломки элементов системы
- Деградация их свойств
- Потеря/повреждение данных
- Задержки в отклике на команды записи и чтения данных
- Ошибки служебного программного обеспечения
- Другие нештатные ситуации

Разрабатываемое решение - специализированное программное обеспечение для:

- Диагностики текущего состояния
- Предсказания вероятности тех или иных сбоев

# Постановка задачи



**Цель:** система диагностики и предсказания сбоев СХД

**Задачи:**

- разработка инфраструктуры для диагностики и предсказания сбоев
- создание симулятора СХД
- получение реальных и модельных тренировочных данных о состоянии СХД
- создание и настройка алгоритмов диагностики и предсказания состояний СХД на основе подходов машинного обучения
- внедрение диагностического комплекса в аппаратно-программное обеспечение СХД

**В решении мы используем data-oriented подход: для построения диагностических и предсказательных инструментов не описываем все причинно-следственные связи, а используем только собранные данные**

# СХД Tatlin компании Yadro



Комплекс прогнозирования сбоев разрабатывается для применения на серийном продукте СХД Tatlin, созданном компанией Yadro.

СХД предоставляется в пяти различных конфигурациях и обладает следующими характеристиками:

- Различные типы доступа - блочный, объектный или файловый
- До четырех контроллеров хранения
- Гибкая политика защиты данных на основе кодов Рида-Соломона с минимальной избыточностью
- Возможность комбинирования типов носителей: NVMe/SAS SSD и SAS/SATA HDD

# СХД: виды сбоев

Ниже приведены возможные сбои СХД:

- Выход из строя контроллеров хранения
- Сбои сетевых контроллеров и/или разрыв сетевых соединений
- Дисковые сбои различного характера (выход из строя, деградация производительности, повреждение данных)
- Сбои SAS-адаптеров
- Ошибки служебного программного обеспечения

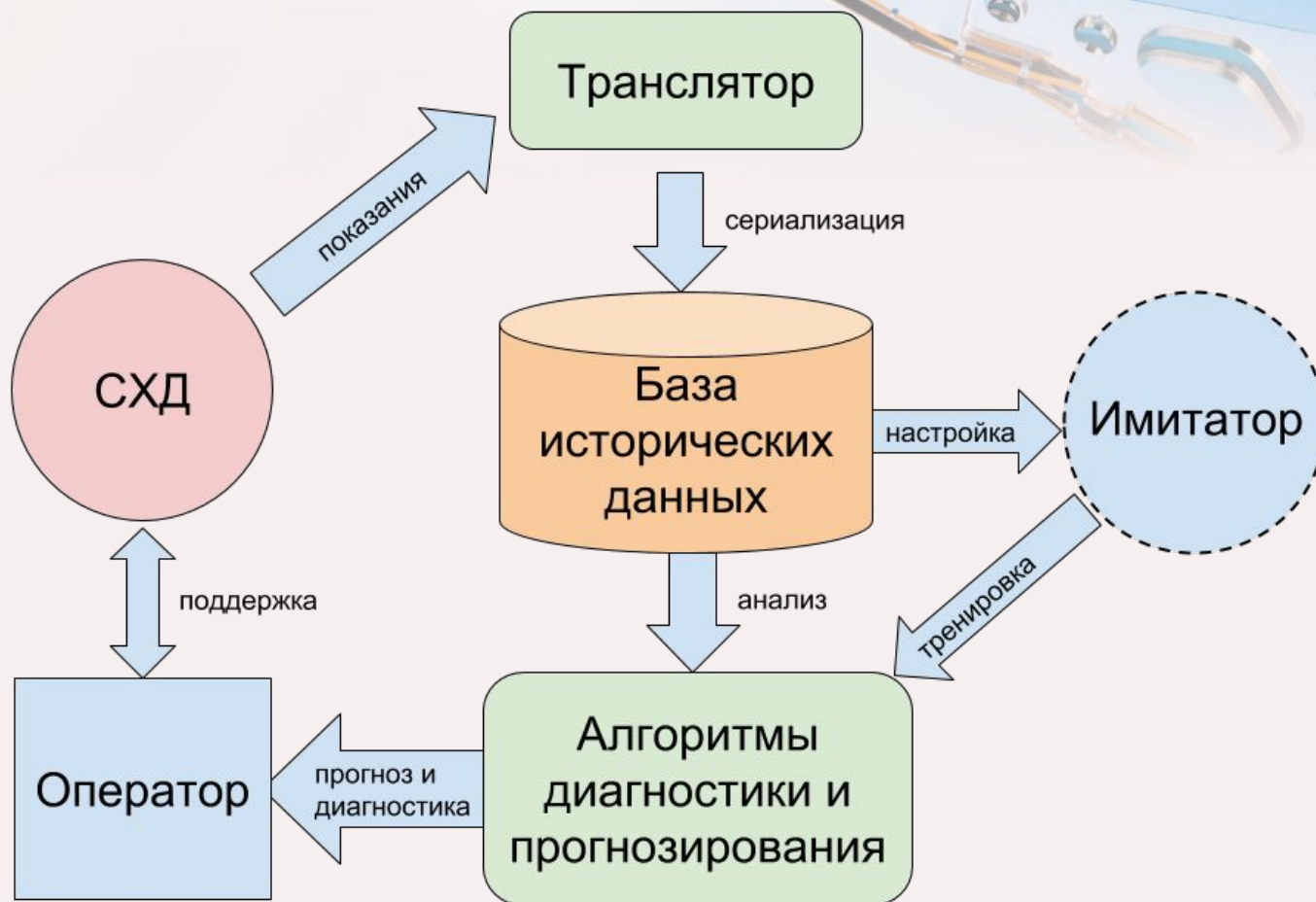
---

Условно состояния СХД можно классифицировать как:

- Работоспособное состояние
- Предотказное состояние
- Сбой




# Система мониторинга и диагностики





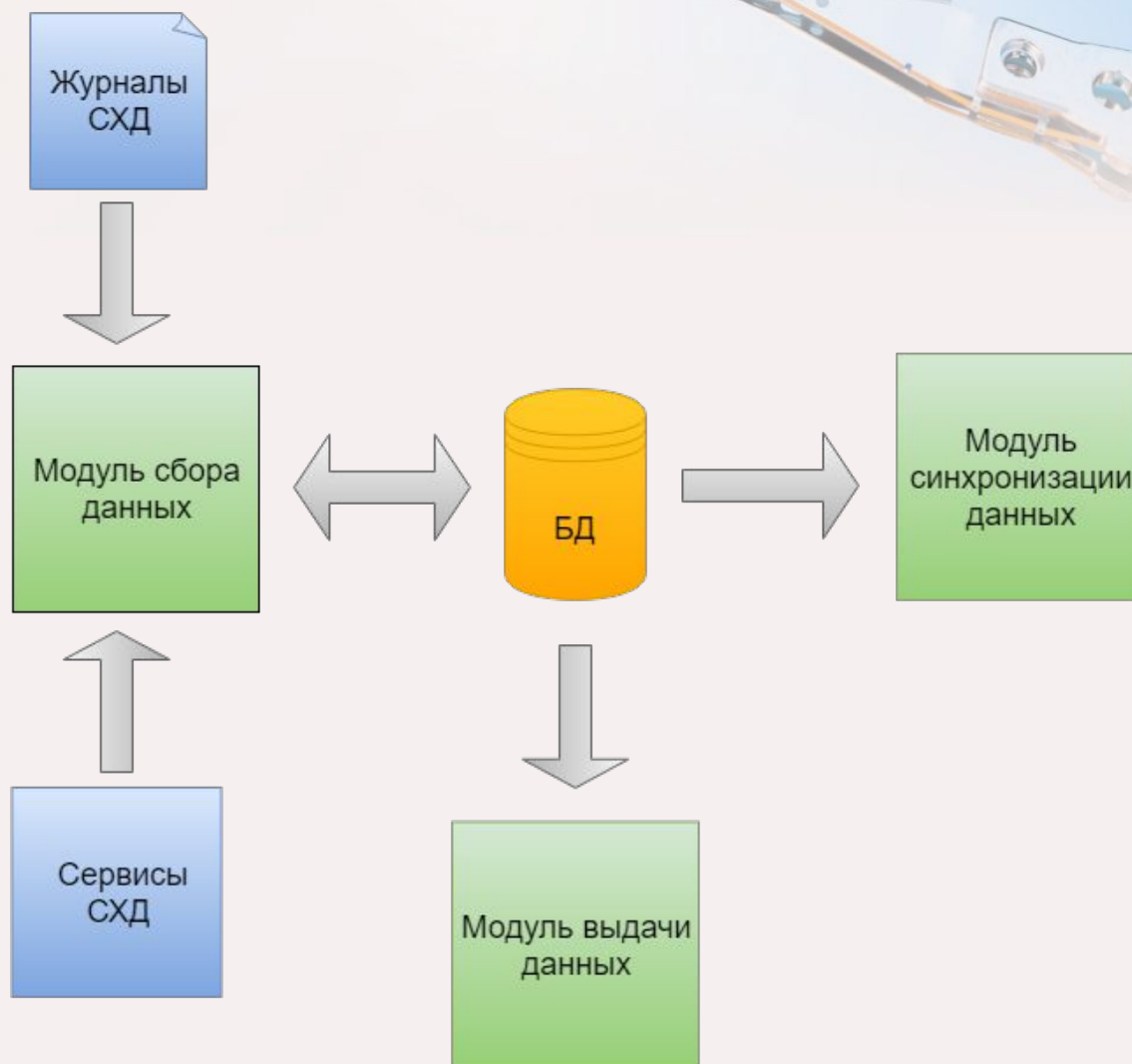
# Инфраструктура для диагностики



Решает следующие задачи:

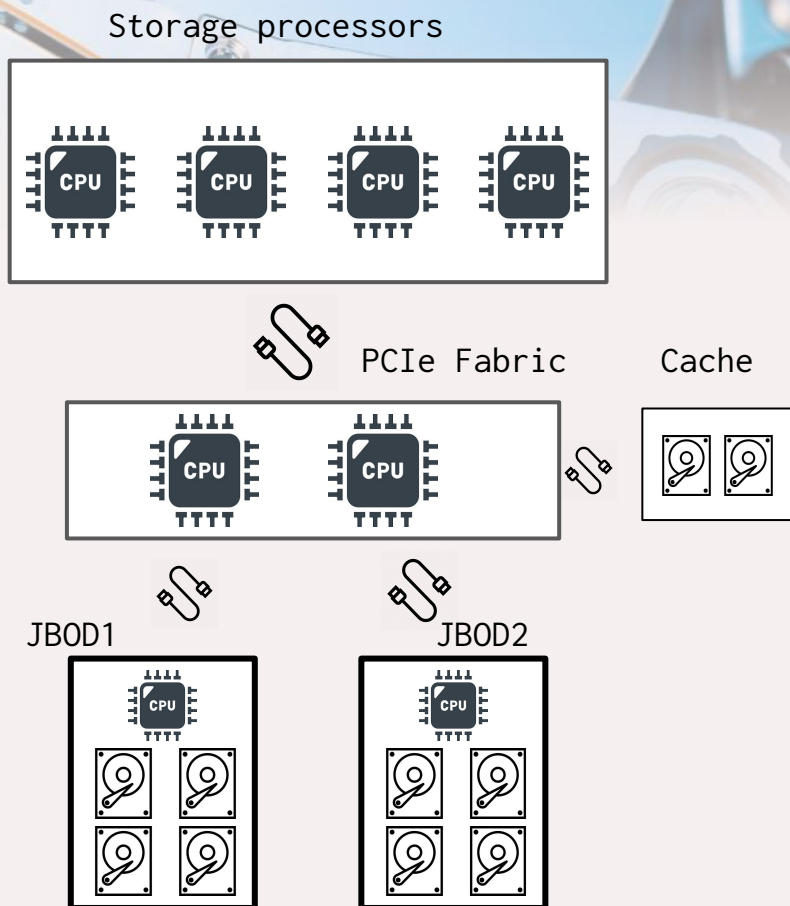
- Мониторинг значений параметров СХД, необходимых для диагностики и прогнозирования сбоев и отказов
- Хранение значений параметров СХД за определенный временной интервал
- Синхронизация накопленных данных между контроллерами хранения
- Взаимодействие с программными сервисами СХД (получение уведомлений о возникновении сбоев и отказов)

# Инфраструктура для диагностики: схема



# Моделирование СХД: симулятор GoTatlin

- Детальное описание компонентов СХД и их взаимодействия
- Моделирование динамики показателей и свойств элементов СХД
- Реализация нештатных ситуаций и процесса деградации оборудования
- Оптимизация параметров моделирования производится с помощью Deep Controller
- **Набор большого количества тренировочных данных для обучения алгоритмов диагностики и предсказания сбоев**



# Моделирование СХД: Deep Controller

**Основная задача:** повышение реалистичности моделирования или, другими словами, тонкая настройка симулятора на реальных данных

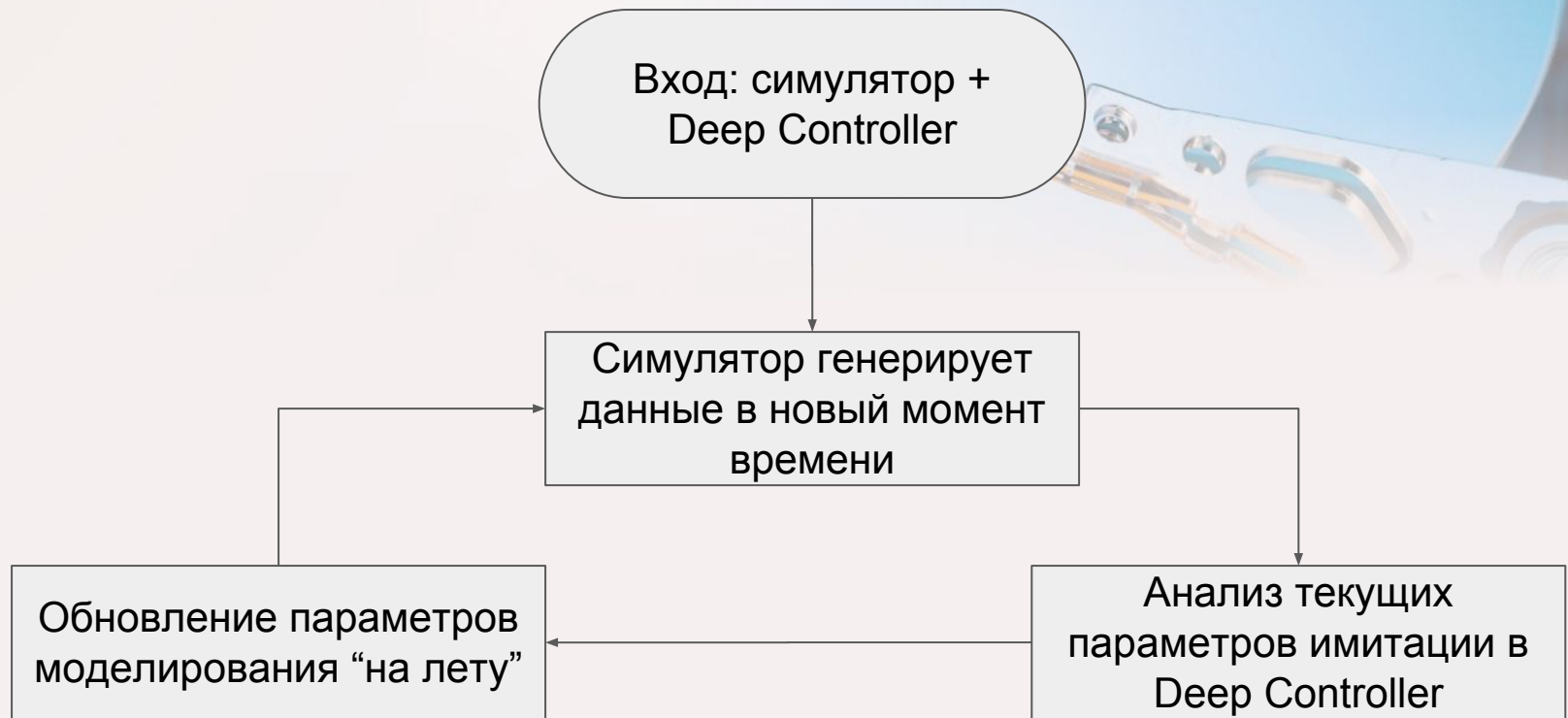
## **Изначальные условия:**

- GoTatlin моделирует работу СХД с некоторой точностью
- количество реальных данных ограничено

## **Основные преимущества решения:**

- основано на глубинных нейронных сетях
- использует алгоритмы обучения с подкреплением
- оценивает улучшение качества моделирования методами спектрального анализа, вейвлет-анализа
- постоянная подстройка параметров моделирования во время работы имитатора (оптимальное управление)

# Deep Controller: применение



- Симулятор гарантирует интерпретируемость результатов моделирования ⇨ больше **доверия** модели благодаря экспертным знаниям;
- Нейронные сети вносят коррекцию в поведение симулятора на основе реальных данных ⇨ **улучшаем** качество моделирования и компенсируем те экспертные знания, которыми мы не обладаем.

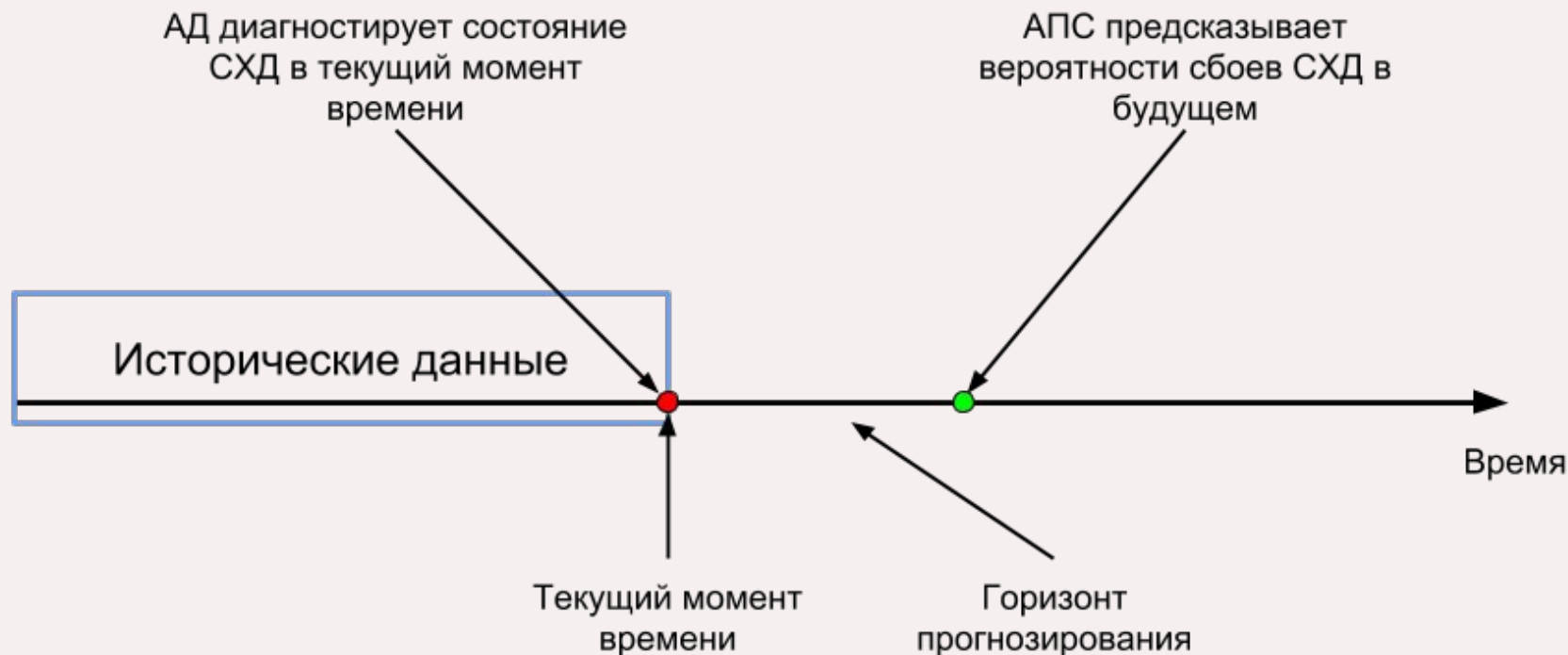
# Deep Controller: обучение



# АД и АПС

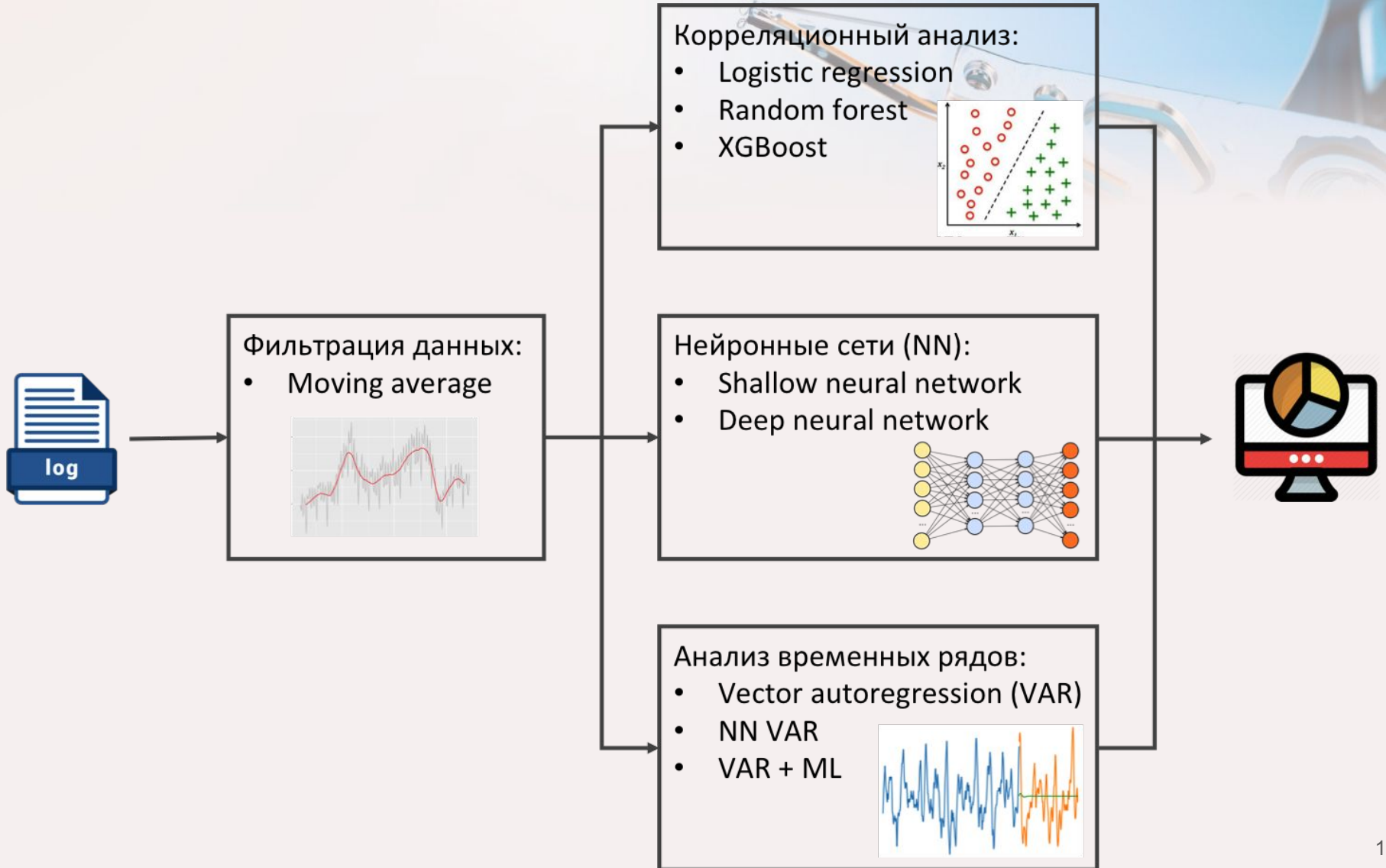
Алгоритмы диагностирования (АД) и прогнозирования сбоев (АПС) являются ключевыми элементами системы мониторинга.

Они обеспечивают возможность ранней диагностики проблемы и дают время на принятия превентивных мер.



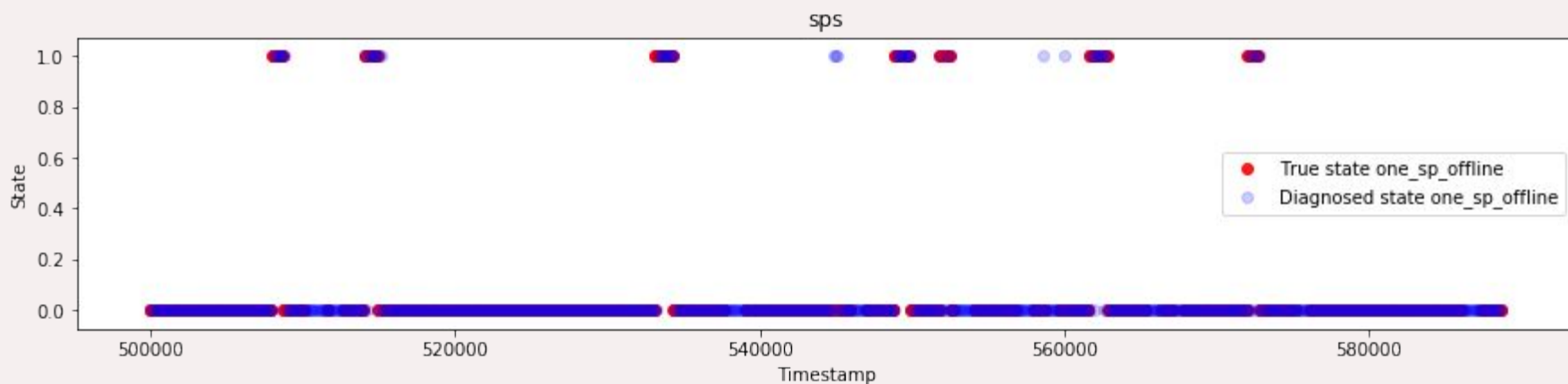
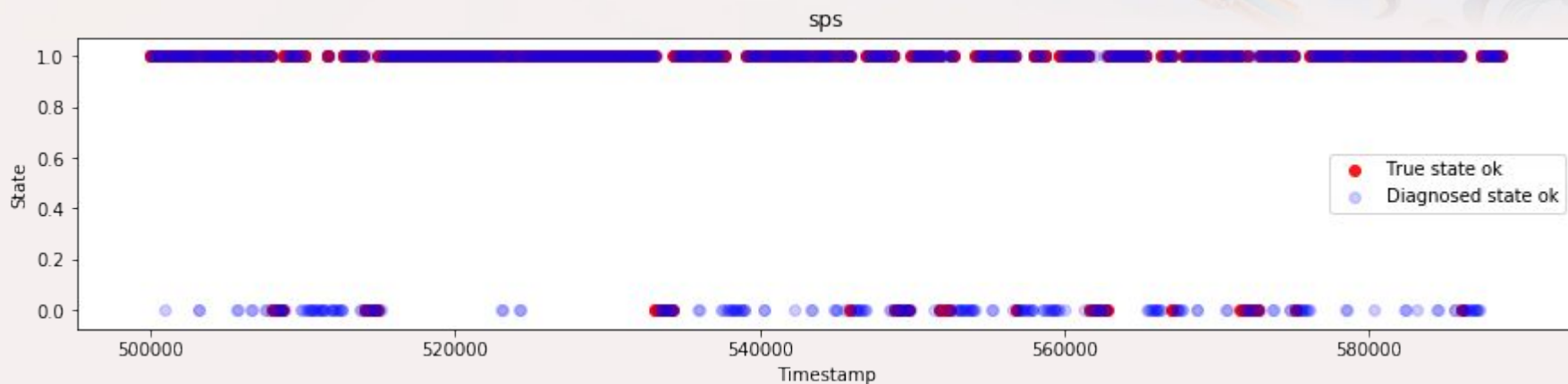


# АД и АПС: применяемые методы



# АД и АПС: применение

На примере искусственных данных, полученных в виртуальной среде, алгоритмы позволяют с точностью 90% получать текущее состояние элементов СХД по ее показателям:



# Следующие шаги

- Сбор тренировочных данных о функционировании СХД Tatlin
- Обучение симулятора для реалистичного моделирования СХД и сбор искусственных данных
- Поиск эффективной схемы смешивания реальных и имитационных данных
- Обучение алгоритмов диагностики и предсказания сбоев
- Внедрение программного комплекса алгоритмов и инфраструктуры в программное обеспечение СХД Tatlin

# Заключение

- Разработана методика повышения надежности системы хранения данных путем предсказания возможных сбоев
- Система диагностики и предсказаний сбоев основана на алгоритмах машинного обучения. Согласно нашему подходу мы не закладываем типичных сценариев сбоев, а извлекаем их из имеющихся данных
- Для тренировки алгоритмов ML используются реальные и искусственные данные о состоянии “здоровья” СХД
- Искусственные данные получаются со специализированного симулятора СХД. Точность существенно увеличивается за счет применения интеллектуальной оптимизации параметров СХД
- В настоящее время доказана перспективность подхода и ведутся разработки промышленного решения для прогнозирования сбоев СХД.