



St Petersburg
University
www.spbu.ru

О КОНЦЕПЦИИ ПОСТРОЕНИЯ ЭКОСИСТЕМЫ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ В РАМКАХ ОТРАСЛЕВОЙ ПЛАТФОРМЫ ДЛЯ РАЗРАБОТКИ IT-РЕШЕНИЙ

St-Petersburg State
University,
St. Petersburg, Russia, 2025

Автомона Татьяна Владимировна
Дик Геннадий Давидович
Богданов Александр Владимирович
Дик Александр Геннадьевич
Юров Виктор Иванович

Все большее значение при разработке решений в области IT технологий приобретает возможность организации требуемого доступа к данным при поддержании мощных вычислительных ресурсов для их обработки

Ограничения временного ресурса

Предоставление постоянного доступа

Обеспечение уровня безопасности

Решение вопроса организации требуемого доступа к данным находится в списке приоритетных направлений технологического развития РФ, где при внедрении технологий ИИ разработчики ПО должны использовать достаточно сложные локальные системы для работы с Big Data

Основная IT-задача отраслевой платформы-повышения качества и ускорения разработок в сфере IT технологий

Отраслевые платформы (ОТПП или IP — Industry Platforms) — это адаптируемые программные бизнес-решения, которые отвечают потребностям конкретной отрасли. Такие платформы предоставляют набор интегрированных сервисов и услуг, включая управление данными и аналитику, управление бизнес-процессами и инструменты разработки интеллектуальных приложений



Роль БД в отраслевых платформах реализуется через системы управления базами данных (СУБД)

Базы данных (БД) играют важную роль в отраслевых платформах — они обеспечивают хранение, управление и анализ больших объёмов информации, что способствует принятию обоснованных решений и оптимизации процессов

функции

Структурированное хранение данных

Работа с данными: создание, обновление, удаление и выборки

Безопасность и контроль доступа, консистентность и отказоустойчивость

Консистентность и отказоустойчивость, резервное копирование и восстановление

Системы, объединяющую функции организации хранения данных по принципу «озер» и хранилищ данных, которые осуществляют сбор и накопление данных из разных источников, хранение «сырых» и обработанных данных с целью предоставления разработчикам возможности работать с этими данными, принято называть экосистемой данных (ЭСД)

Место ПО в рамках ЭСД и платформы



Компоненты ЭСД

Для решения задач сбора и хранения информации (данных, методов, алгоритмов) целесообразно использовать

комплексирование СУБД разных типов. Сочетание реляционных баз данных с базами NoSQL и файловой системой позволяет хранить информацию любого типа для решения задач, связанных с данными

Реляционные БД
(PostgreSQL,
MySQL)

NoSQL БД
(MongoDB (Mg),
Redis (Rd))

**Каталоги
файловой
системы**

**Встроенные
расширения
(Timescale DB,
pgvector, FDW,
patroni)**

Логическая схема ЭСД

Логическую схему ЭСД можно представить следующим образом:



Основные функции блоков

Работа блоков ЭСД, как единой системы, управляется программным слоем платформы

Взаимодействие с внешними (относительно ЭСД) источниками информации

Организация сбора внешних данных с использованием различных технологий

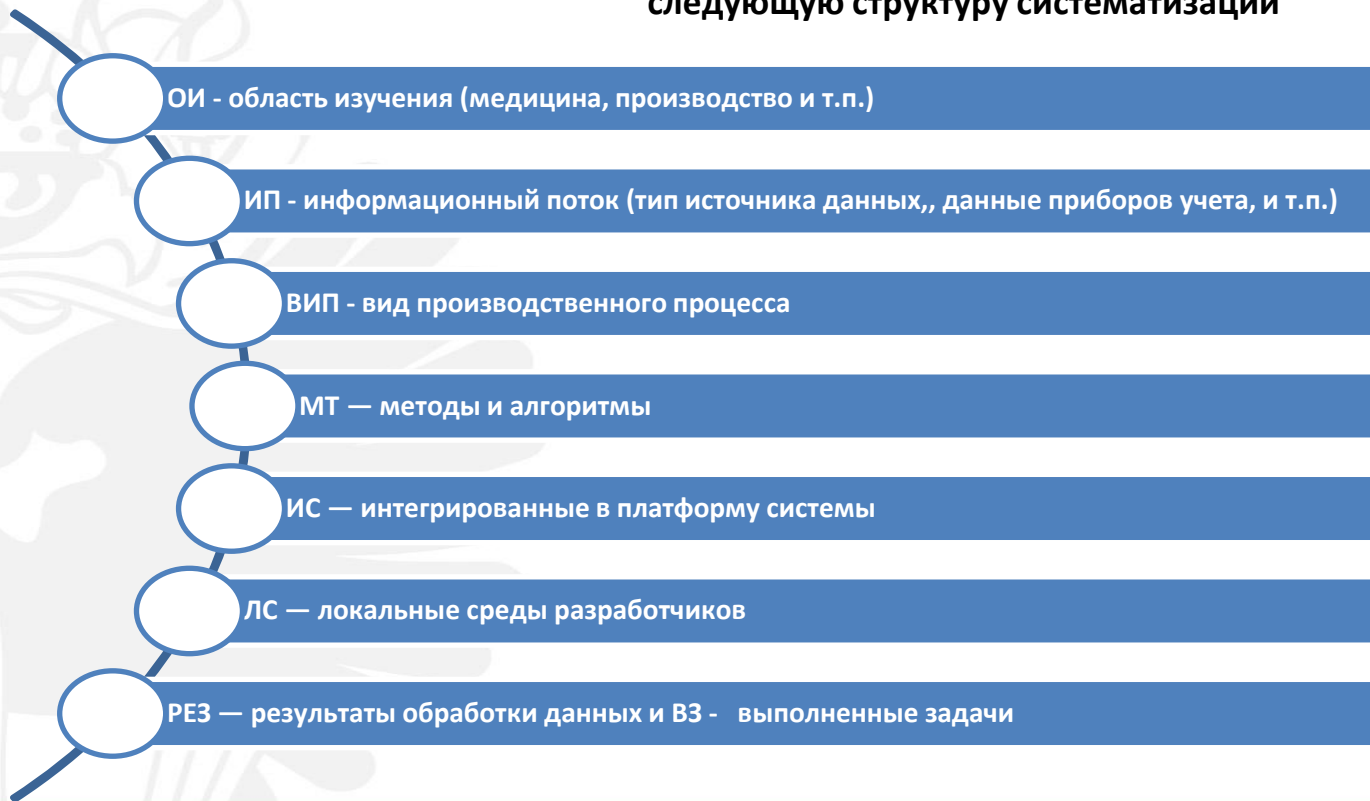
Предоставления платформой возможности для разработчика создавать свою среду данных как из имеющихся в ЭСД, так и дополнительно подгружаемых для обработки и анализа выбранных алгоритмов и методами

Организация хранения информации в базах данных MainDB и в каталогах файловой системы для обработанных данных, справочных данных, методов и алгоритмов работы с данными.

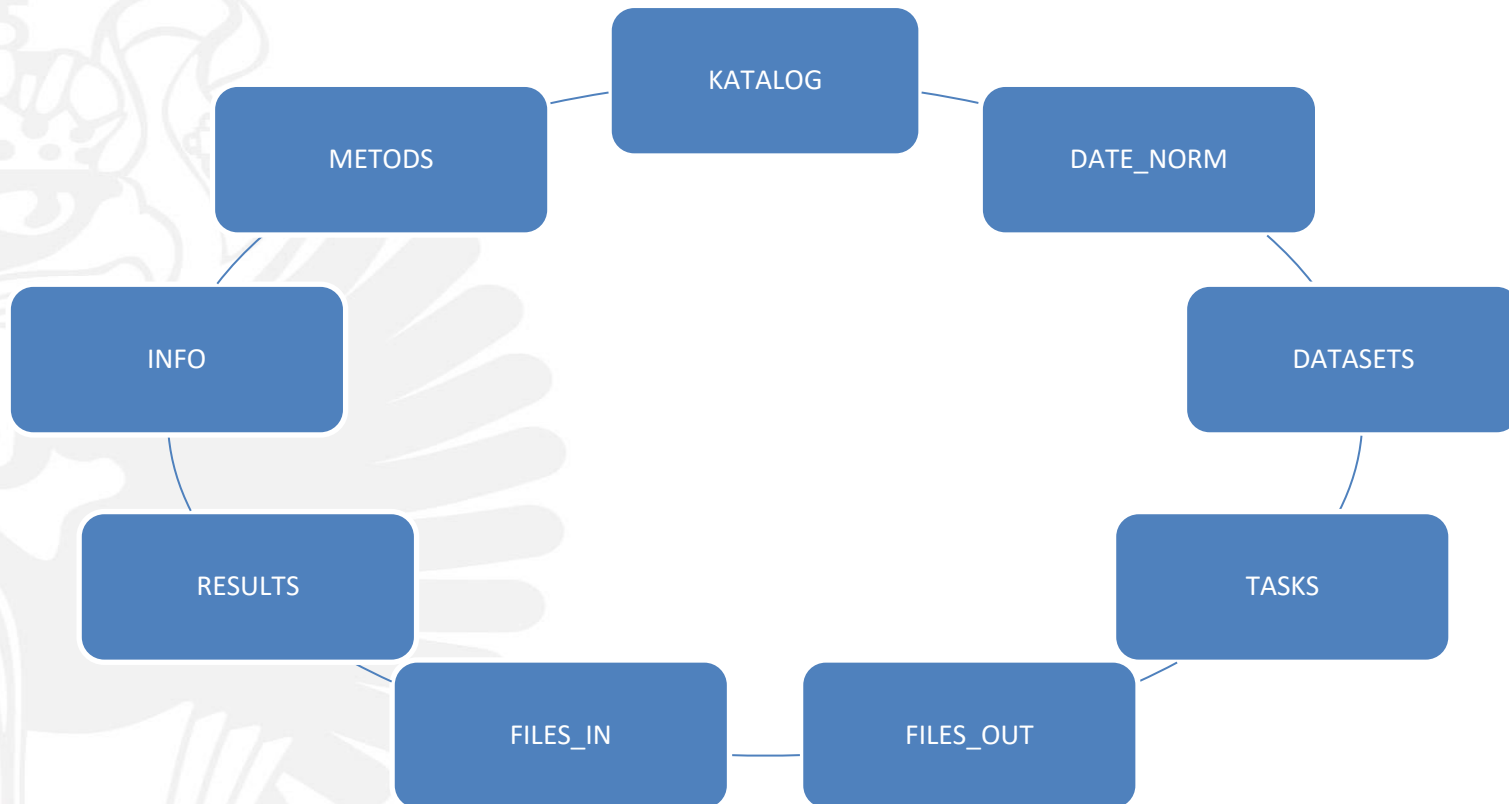
Организация хранения информации в интеграционные базы данных (IntegrBD) и каталоги файловой системы для сырых данных

Организация хранения очень старых данных, если это необходимо. Определяется при разработке архитектуры платформы

Для организации верхнего уровня поиска можно применить примерно следующую структуру систематизации



MainDB может содержать информационные блоки



Выводы

Сбор большого по объему количества разнородных данных в рамках одной платформы позволит разработчикам не тратить ресурсы на их поиск и подготовку. Создав локальную среду разработки в рамках платформы и развернув в ней при необходимости разрабатываемое ПО, разработчики смогут полноценно работать с данными, например, получить доступ на чтение напрямую к потокам данных (Kafka), подключиться к сырым и подготовленным данным, подходящим для решения поставленной задачи и хранящимся с ЭСД, найти готовое решение и попробовать применить его на собственных данных, создать из имеющихся в системе собственные наборы данных, применить свои или интегрированные в платформу системы анализа данных (BI-системы). При этом не нужно тратить время и средства на создание и обслуживание сложных вычислительных структур. Все собрано воедино в рамках платформы. Особенно это актуально при создании технологий ИИ, где для обучения и тестирования НС нужны большие сырые и подготовленные данные. Платформы может использоваться и для оказания услуг в формате - данные как услуга и функция как услуга.

Таким образом можно ожидать экономии и повышения эффективности ИТ разработок при использовании подобной отраслевой платформы



Спасибо за внимание!

St Petersburg University
spbu.ru