



КХД 2.0 Росатом и Форсайт: от концепции к успешной реализации

Конференция TechEd

Студенцов Павел Вячеславович

Тригубчак Евгений Петрович

Тамбовский Александр Анатольевич

Центр отраслевой архитектуры и интеграции

Задачи ЦОАИ сегодня



Выбор Форсайт. Аналитическая Платформа для проекта ИАС КХД 2.0



Осуществлен выбор решения:
Форсайт. Аналитическая Платформа
в варианте архитектуры с продуктами
– Arenadata DB и Arenadata QM



3
этап

- Подготовлены материалы для вынесения на архитектурный совет материалов
- Осуществлен выбор решения

2
этап

- Настройка вендорами прототипов
- Проведение вендорами демонстраций прототипов
- Подготовка и проведение нагрузочного тестирования (НТ) прототипов
- Оценка результатов НТ
- Расчет совокупной стоимости владения (ТСО)



Определены **3** варианта архитектуры

Проведено НТ по **3** вариантам архитектуры

Проведено прототипирование по **9** сценариям

1
этап

- Формирование Long List платформ-участников
- Разработка критериев оценки платформ
- Разработка требований на прототип
- Анализ рынка и оценка аналитических платформ по разработанной системе критериев
- Определение лидеров



8 платформ-участников

~300 критериев оценки

Определены **3** лидера из участников

Основные технические критерии: **~50 Тб** данных, **~2000** одновременных пользователей

Выбор инфраструктуры

Динамика разработки и уточнения сайзинга



Первичный сайзинг проводился с привлечением вендоров перед стартом проекта



Уточнение первичного сайзинга производилось подрядчиком на первом этапе проекта



Текущий сайзинг получен по результатам проведенных нагрузочных тестирований

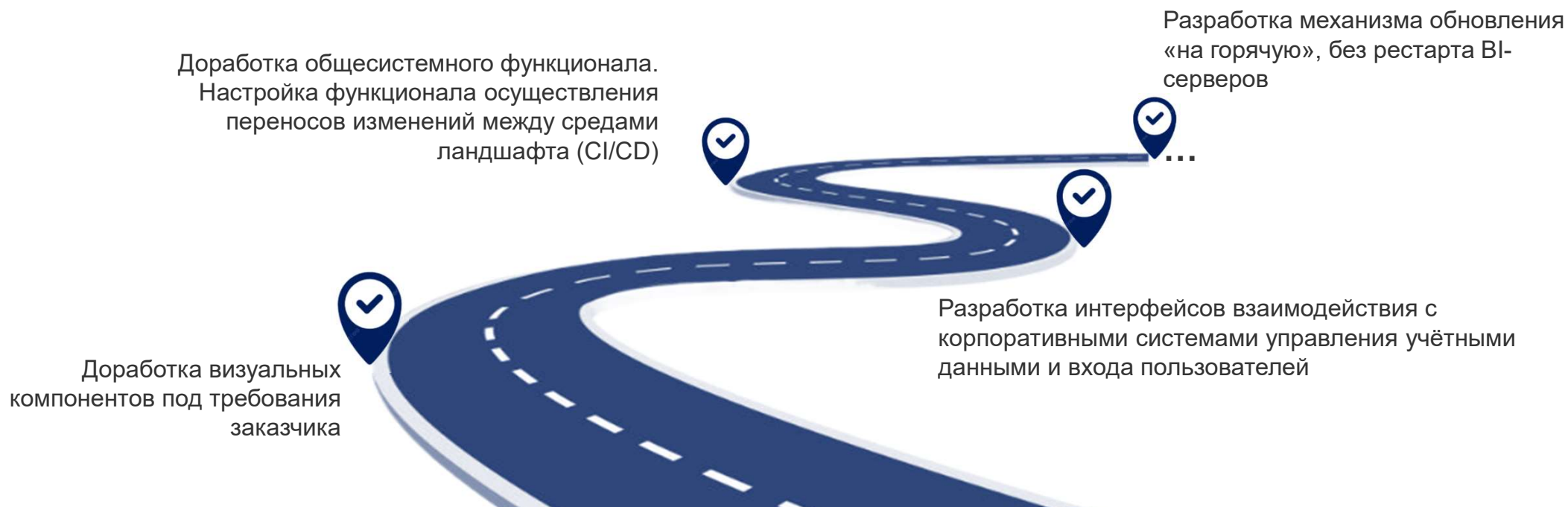
Ключевые метрики и рекомендации:

- Необходимо учитывать параметры виртуализации корпоративного облака. Коэффициент переподписки CPU, RAM, ёмкость гипервизоров, соседства с высоконагруженными ИС
- При разработке сайзинга необходимо учитывать рекомендации – 4-5 пользователей на ядро и 1,5-2 Гб RAM на одного пользователя (имеется в виду одновременно работающих пользователей)
- Учитывая, что слой Data Marts (DM) был реализован на ADQM, важно тщательно подбирать конфигурацию серверов DM, исходя из правила 1000 одновременных пользователей на 1 ноду



Важно помнить, что увеличение инфраструктуры – последний шаг в борьбе за производительность. Оптимизация кода, тюнинг параметров прикладного ПО, применение исключений для антивируса – первые шаги для оптимизации ИС и обеспечения необходимых требований по производительности.

Развитие Форсайт. Аналитическая Платформа в проекте



Движение навстречу. Проект улучшает продукт, продукт улучшает проект.
За время проекта было выпущено не менее трех новых сборок продукта Форсайт. Аналитическая Платформа, доработанных под требования.

Хранилище данных

Цели и задачи



- 01** Развитие экспертизы на новом технологическом стеке
- 02** Создание новых решений и продуктов
- 03** Централизованная поддержка и развитие систем
- 04** Реализация проектов собственными силами
- 05** Осуществление контроля выполняемых разработок
- 06** Разработка и поддержание единого унифицированного подхода к ведению разработок.

Создаваемые и поддерживаемые продукты



ЦИП
цифровая
инвестпрограмма



ИСУБ
система управления
бюджетированием



ВДНМ 2.0
витрина налогового
мониторинга

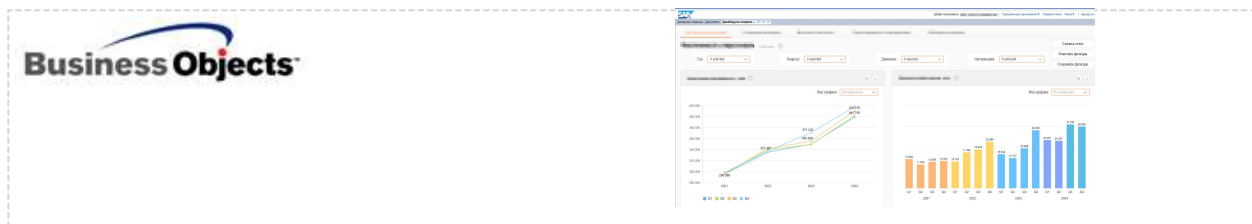


КХД 2.0
корпоративное хранилище
данных

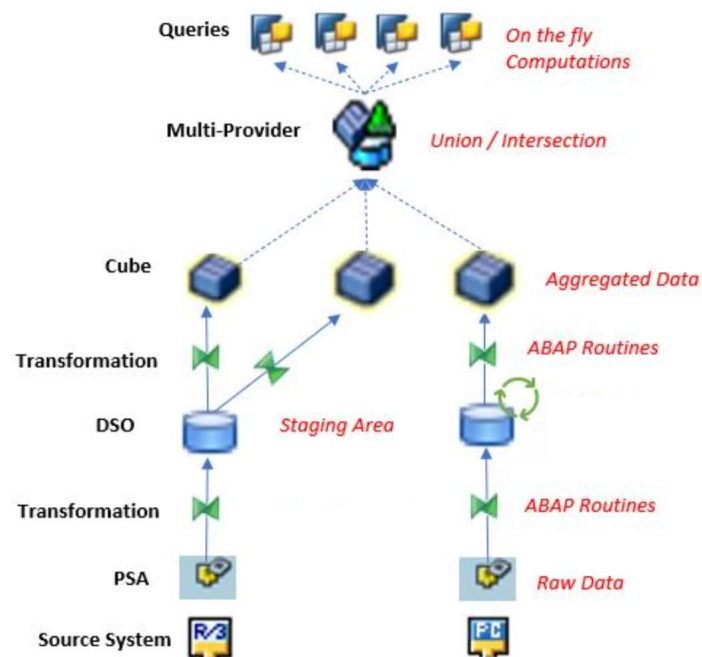


**АСУСС-
Прогноз**
Система управления
стоимостью сооружения

Архитектура КХД 1.0



BW/HANA



20 Тб

Объем информации
в хранилище
данных

60 тыс.

Объектов базы
данных: таблиц,
ракурсов

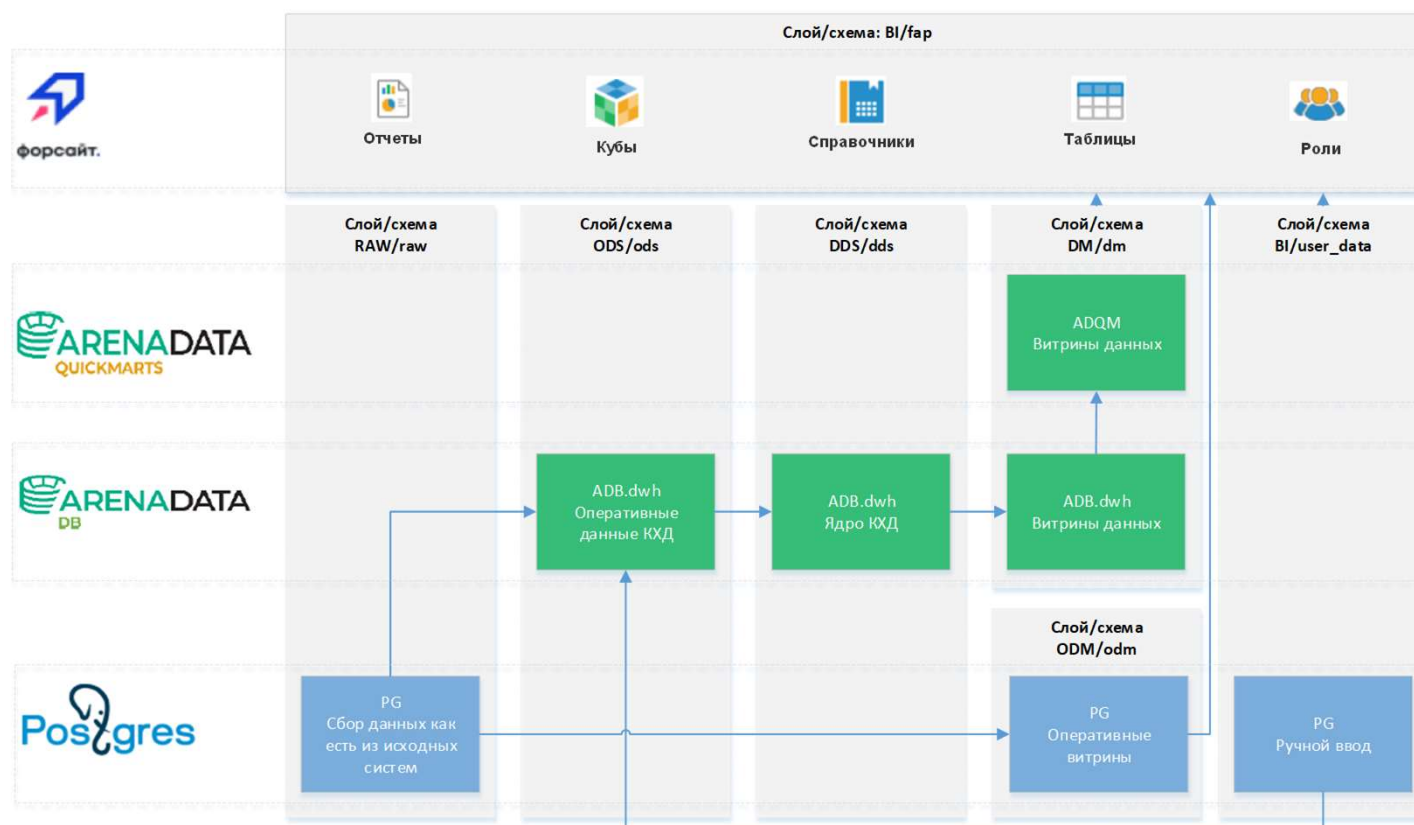
500+

Одновременных
пользователей
системы

200+

Интеграционных
поток

Архитектура КХД 2.0 – целевое решение



50 Тб

Объем информации в хранилище данных

8000+

Объектов базы данных: таблиц, ракурсов

2000+

Одновременных пользователей системы

400+

Интеграционных потоков

Процесс проектирования и разработки КХД 2.0



Проектирование

Сквозное документирование проектируемых объектов



Регламент разработки

Включает описание стандарта ведения разработки, доставки изменений и типовых решений



Быстрый старт

Реализация упрощенных витрин на Postgres без ядра хранилища, команды BI и DWH идут встречными потоками



Интеграционный стандарт

Данные должны передаваться безопасно, с достаточной скоростью, гарантировано, с прозрачным процессом согласования взаимодействия



Автоматизация разработки

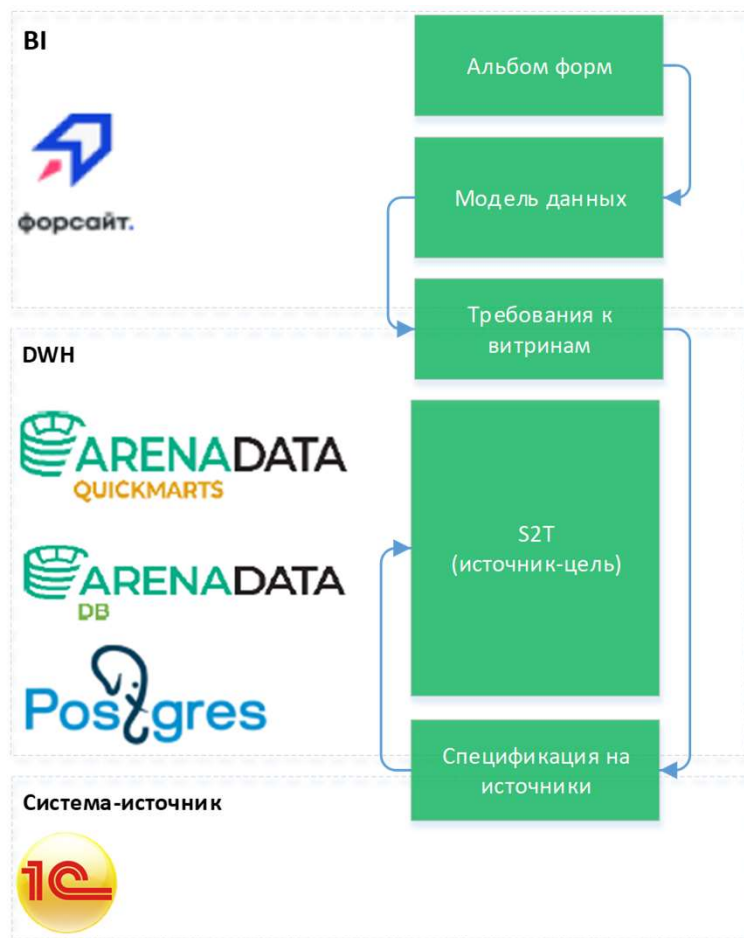
Разработка хранилища данных с использованием средств автоматизации создания и поддержания в актуальном состоянии



Проверка результата

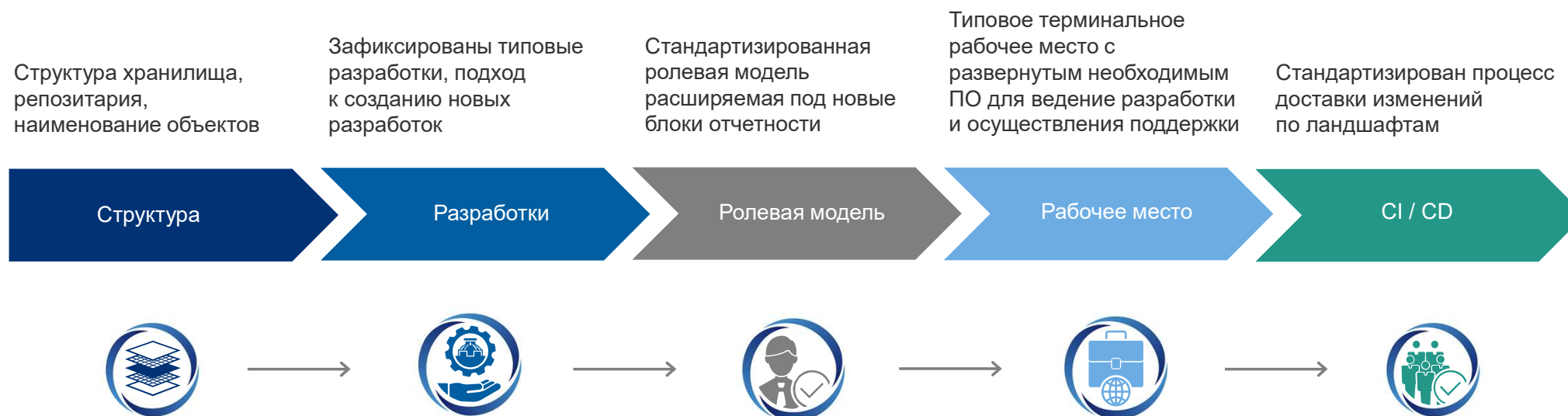
Сквозное тестирование данных от источника до отчета

Проектирование КХД 2.0

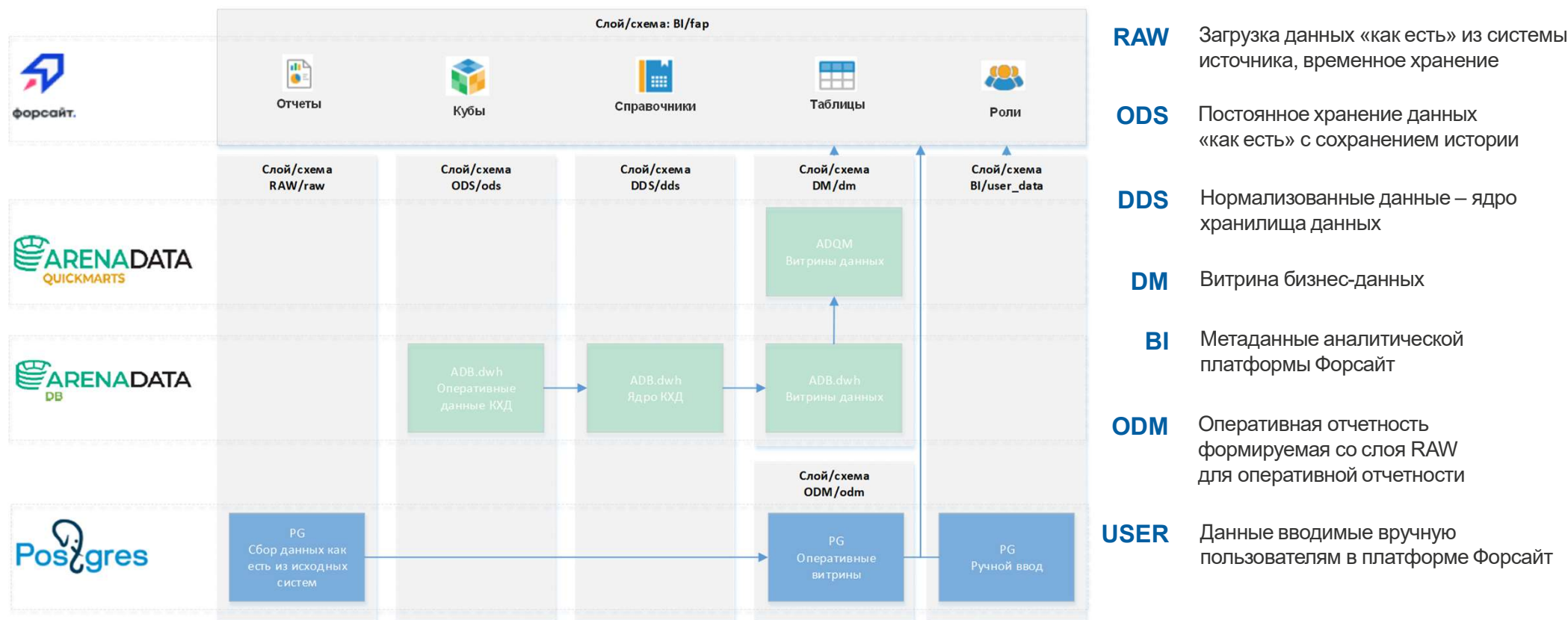


- **Альбом форм**
Визуализация, привязка к измерениям и показателям
Модели данных, алгоритмы расчетов
- **Модель данных**
Измерения, показатели, витрины данных
- **Требования к витринам**
Бизнес-описание витрин данных
- **Спецификация на источники**
Описание алгоритмов и структур в нотации системы-источника для получения необходимых данных
- **S2T (источник-цель)**
Описание всех шагов преобразования данных в хранилище от слоя «сырых данных» до слоя витрин

Регламент разработки КХД 2.0



Архитектура КХД 2.0 – быстрый старт



Интеграционный стандарт



Конфиденциальность данных

Допустимость потоков данных и взаимодействий между ИС в зависимости от уровня конфиденциальности информации интегрируемых систем

Зоны в безопасности

Допустимость потоков данных и взаимодействий между ИС в зависимости от зоны безопасности ЦОД

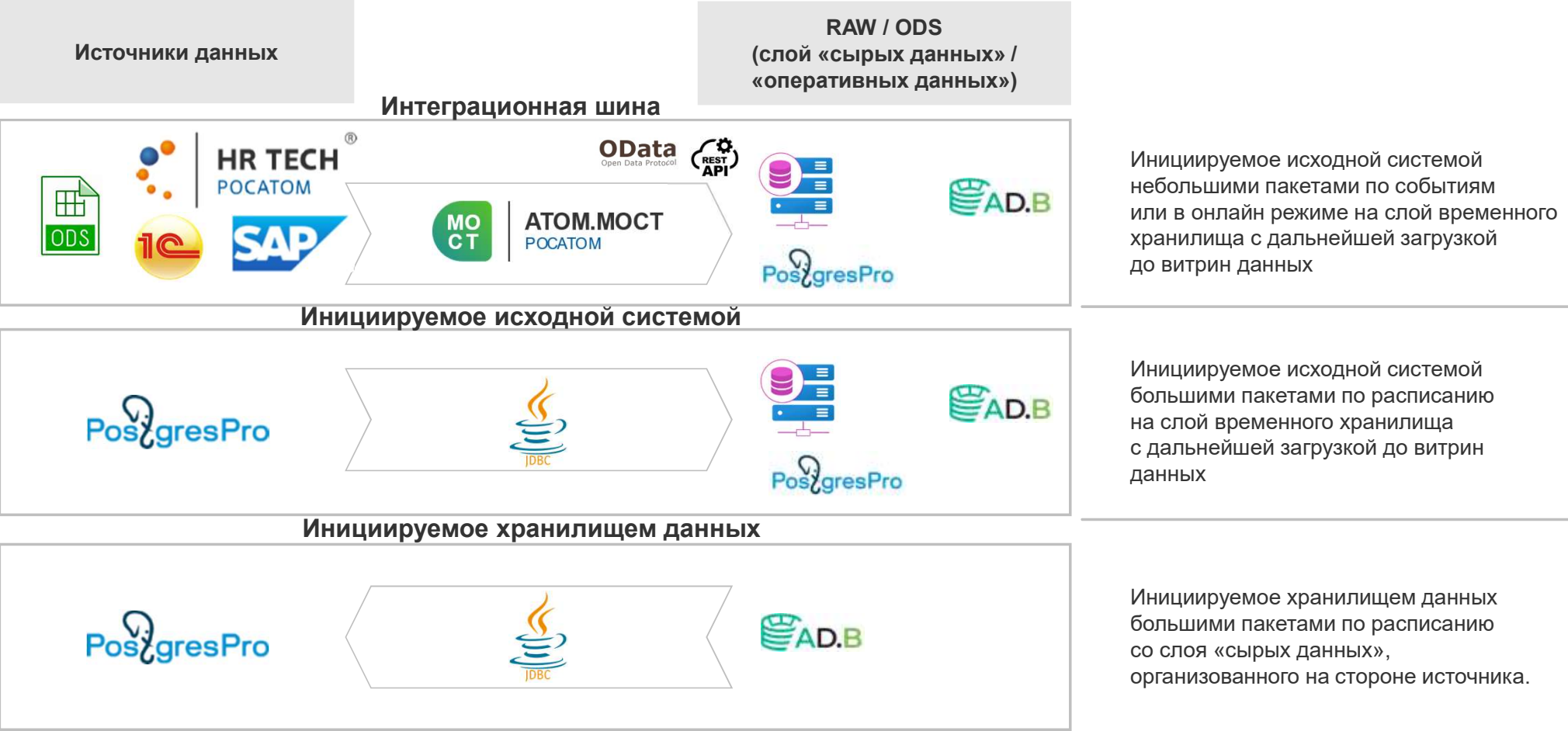
Стандарт интеграции

Корпоративный стандарт – Интеграционная шина (ИШ)

Альтернативы

Критерии, позволяющие выбрать альтернативный способ – прямого взаимодействия (ПВ) систем

Интеграционный стандарт КХД 2.0



Автоматизация разработки КХД 2.0



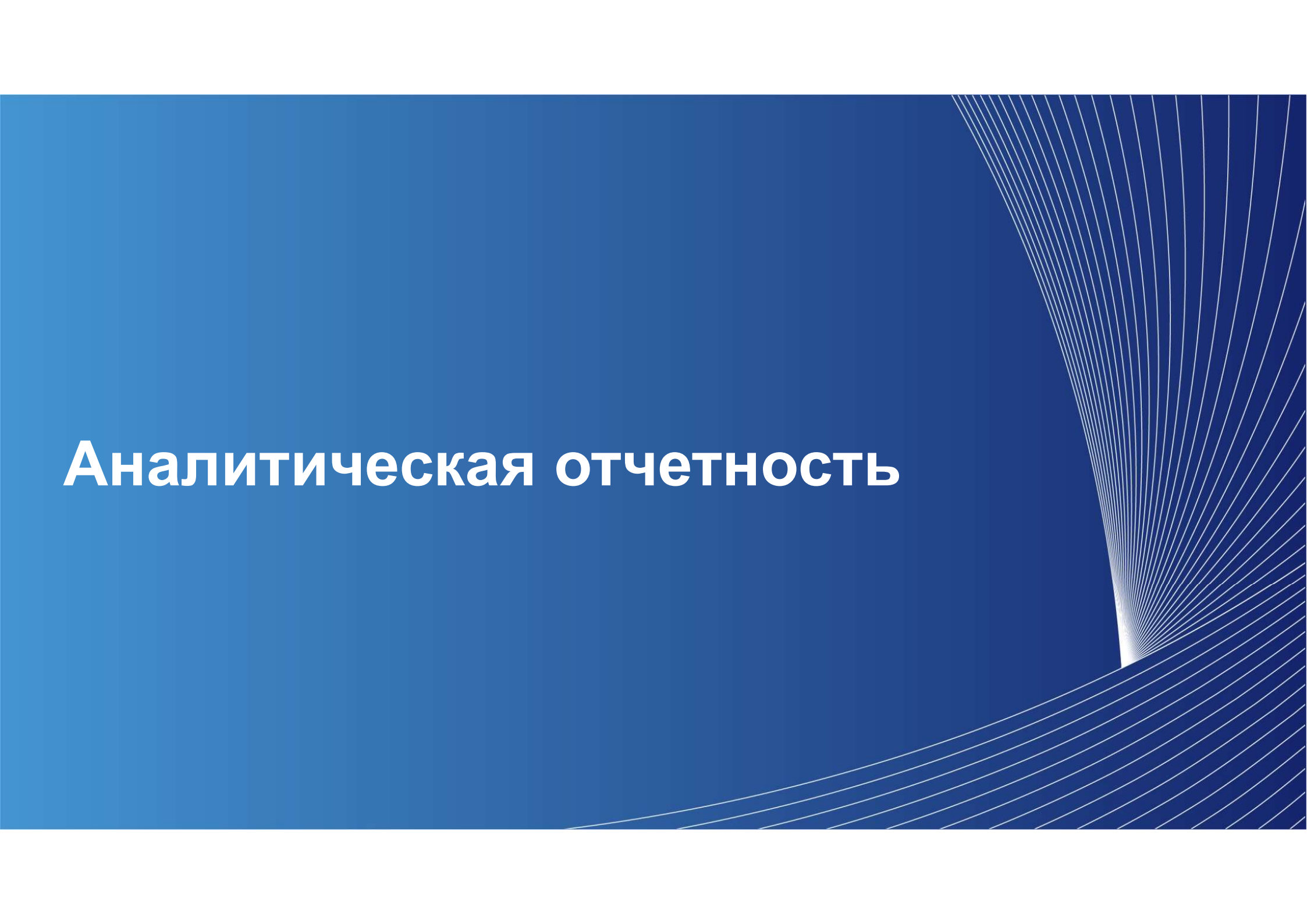
Generate_flow –
универсальная функция
автоматизированное создание таблиц

Grid	flow	attr_name	attr_type	attr_desc	ods	link	hub
1	imp	priority	varchar	Идентификатор приоритета обучения	[v]	[v]	[]
2	ZGF	SUBTY	text	Подтип	[v]	[]	[]
3	cou	title	varchar	Название приоритета обучения	[v]	[]	[]
4	imp	title	varchar	Наименование действия	[v]	[v]	[]
5	use	min_login_date	text	Дата первого входа в Rosatom Life	[v]	[v]	[]
6	TES	attr4	text	attr4	[v]	[v]	[v]
7	ZGF	UNAME	text	Имя исполнителя, изменившего объект	[v]	[]	[]
8	cou	kind	varchar	Ссылка на Справочник Виды обучения	[v]	[]	[]
9	OEN	BSGRD	decimal(5, 2)	Степень занятости	[v]	[]	[]
10	OEN	PERSK	text	Категория сотрудников	[v]	[]	[]
11	OEN	SLREG	text	Область для зарплаты	[v]	[]	[]
12	OEN	TRFGR	text	Тарифная группа	[v]	[]	[]
13	OEN	TRFST	text	Уровень тарифа	[v]	[]	[]

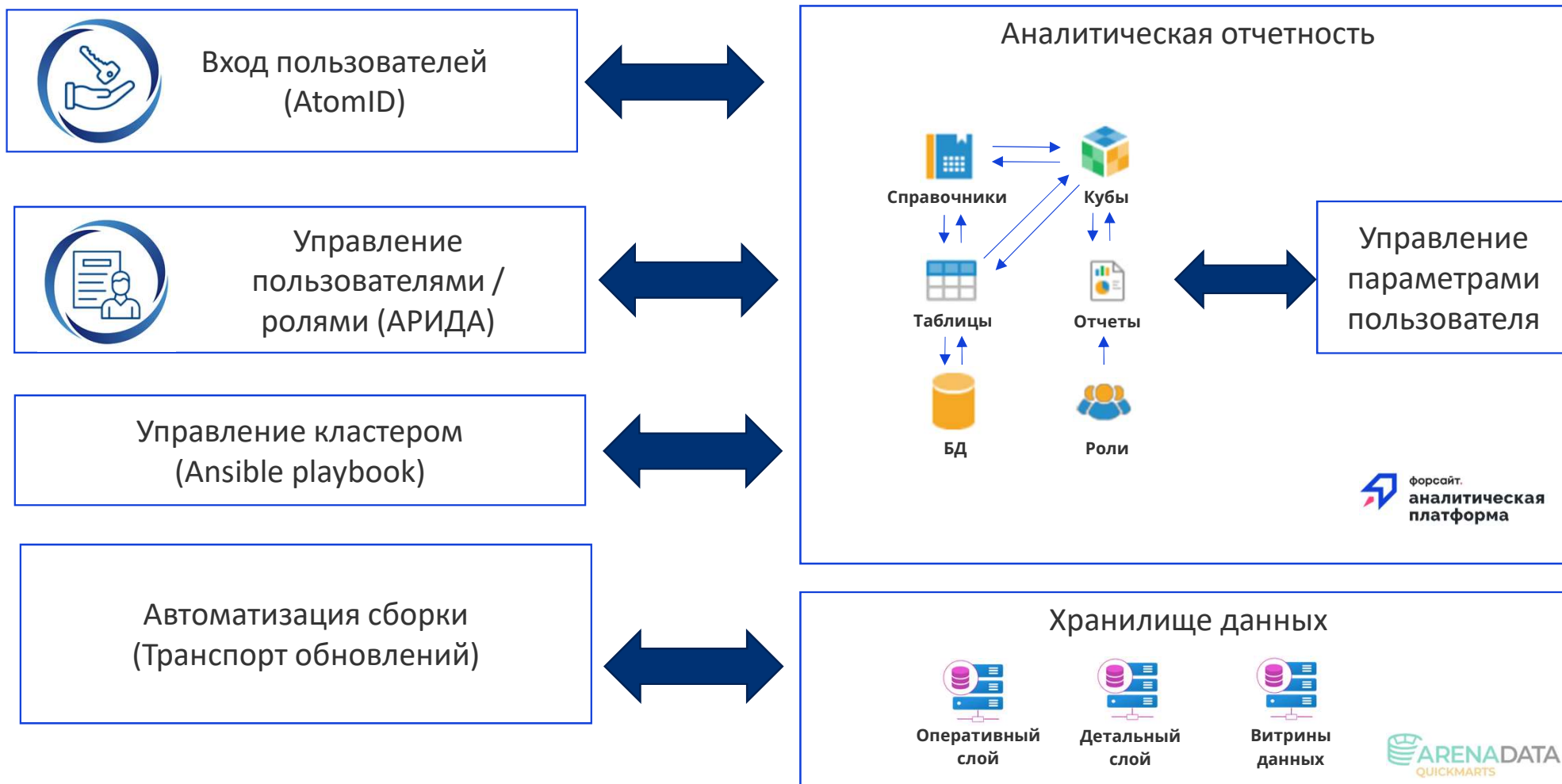
Load_general – универсальная функция загрузки между от RAW до DDS

Grid	transform_id	func_name	source	target	load_type	fields_mapping
1	dds.hub_iasup_hr	logger.dwm_load_general	ods.iasup	dds.hub_iasup	insert	{}
2	dds.sat_iasup_h	logger.dwm_load_general	ods.iasup	dds.sat_iasup	td_attr	{"hdiff": "hdiff"}
3	dds.hub_iasup	logger.dwm_load_general	ods.iasup	dds.hub_iasup	insert	{}
4	dds.sat_iasup	logger.dwm_load_general	ods.iasup	dds.sat_iasup	insert	{}
5	dds.hub_iasup	logger.dwm_load_general	ods.iasup	dds.hub_iasup	insert	{}
6	ods.iasup	logger.dwm_load_general	dwh.ods.iasup	ods.iasup	ods	{"hkey": "hkey", "
7	dds.sat_iasup	logger.dwm_load_general	ods.iasup	dds.sat_iasup	td_attr	{"dateto": "endda
8	dds.sat_record	logger.dwm_load_general	ods.record	dds.sat_record	insert	{}
9	ods.record_m	logger.dwm_load_general	dwh.ods.record	ods.record	full	{"hkey": "[bkey]",
10	dds.hub_iasup	logger.dwm_load_general	ods.iasup	dds.hub	insert	{}

Аналитическая отчетность

The background is a solid blue gradient. On the right side, there is a series of thin, white, curved lines that fan out from a point near the bottom right corner, creating a sense of motion or data flow.

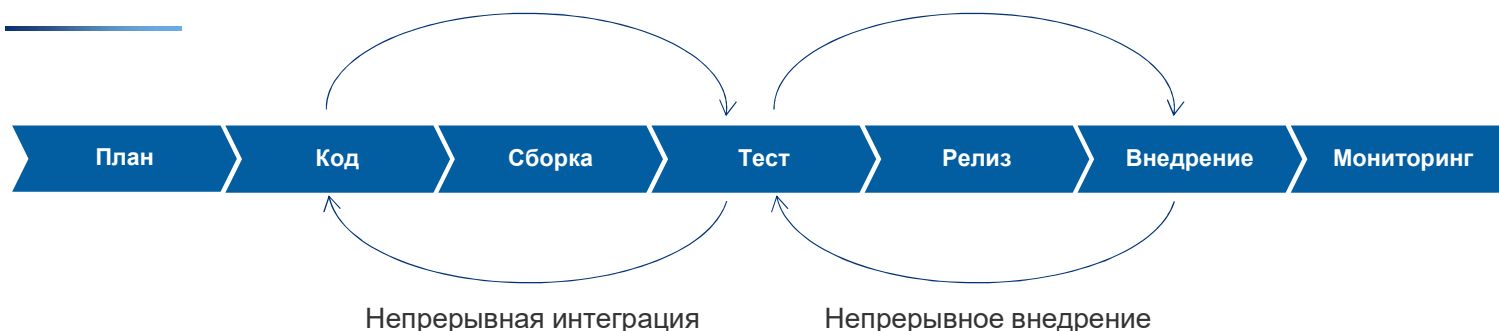
Архитектура модуля аналитической отчетности КХД 2.0



Транспорт переноса обновлений объектов репозитория Форсайт. АП КХД 2.0



Требования



Результат

- Унифицирован подход к формированию пакетов пепфов;
- Упрощена диагностика и минимизированы ошибки при переносе;
- Охвачены два обособленных контура (ЗКО и МБ);
- Сокращено непосредственное участие специалистов в процессе переноса обновлений;
- Обеспечена легкая адаптация процесса для новых блоков.

>5

организация работы
обособленных
команд разработки

~100

технических специалистов,
использующих функционал

Обновление объектов репозитория Форсайт. АП «на горячую», без рестарта ВІ-серверов КХД 2.0



01 Поставить на выполнение по расписанию ansible-скрипт, выполняющего очистку кэшей



02 После выполнения указанных действий при следующей авторизации пользователю будет доступен обновленный функционал

Механизм сохранения наборов параметров для каждого пользователя КХД 2.0



Задача

Предоставить возможность формирования предустановленных наборов фильтров для каждого пользователя



Быстро получать разные выборки данных под разные цели (статусы, встречи по результатам, напоминания)



Сократить количество отчетов, отличающихся только отметкой параметров

Основные требования:

- Создание раздела параметров для отдельной формы или для группы форм;
- Добавление / удаление параметров;
- Изменение наименований параметров;
- Сохранение / удаление набора выбранных пользователем параметров;
- Очистка всех сохраненных наборов параметров для конкретного раздела параметров.



Для настройки выборки данных, на основании которых формируются отчеты, предусмотрена настройка фильтрации через дополнительную форму «Управление параметрами».

Интеграция КХД 2.0 с корпоративными системами управления учётными данными и входа пользователей



AtomID –

централизованный сервис идентификации, аутентификации и авторизации работников



AtomID

- Идентификация и аутентификация
- Единый вход (Single Sign On)
- Контроль доступа к приложениям и сервисам (OAuth 2.0, API Gateway)

АРИДА –

система централизованного управления доступом пользователей



АРИДА

- Управление учетными записями пользователей
- Делегированное администрирование
- Одобрение и назначение доступов пользователям
- Синхронизация учетных записей



Active Directory

- Доменные учетные записи пользователей
- Атрибуты учетных записей
- Пароли пользователей
- Полномочия и права доступа

Результаты нагрузочного тестирования КХД 2.0 по блокам УК и ЕОСДО



Проводилось тестирование базовой нагрузки с помощью программы Apache Jmeter

> Блок УК

Профиль нагрузки:

1 000 конкурентных
пользователей

Продолжительность
тестирования:

8 часов

> Блок ЕОСДО

Профиль нагрузки:

1 000 конкурентных
пользователей

Продолжительность
тестирования:

16 часов формирования фоновых
отчетов

Итоги тестирования:

- Среднее время открытия отчётов – **10 сек** (соответствует целевым показателям).
- Блокировки в базе данных во время тестирования.
Решение: Включение кэша сборок на локальном диске в толстом клиенте. Включение DebugMode на планировщике.
- Отсутствие идентификатора сессии в логах BI, вызванное некорректной работой Apache Jmeter.
Решение: Увеличение времени жизни потока в Apache Jmeter. Увеличение ОЗУ перед стартом тестирования на 1 Гбайт.

- Потеря сессий нагрузочными пользователями.
Решение: Ошибка устранена в сборке ФАП 10.4.1087.
- Ошибки и задержки при выходе пользователей из системы.
Решение: Скорректированы права для нагрузочных пользователей.
- Повышенная загрузка CPU в ADQM до и после тестирования.
Решение: Внедрён новый алгоритм для стабилизации нагрузки.
- Большая очередь фоновых отчетов после завершения тестирования.
Решение: Оптимизация механизма освобождения памяти на сервере с планировщиком.
- Ошибки при выборе значений параметров для фонового формирования отчета.
Решение: Скорректирован файл с параметрами скрипта.

Подходы обновления Форсайт. Аналитическая платформа с минимальным влиянием на разработку



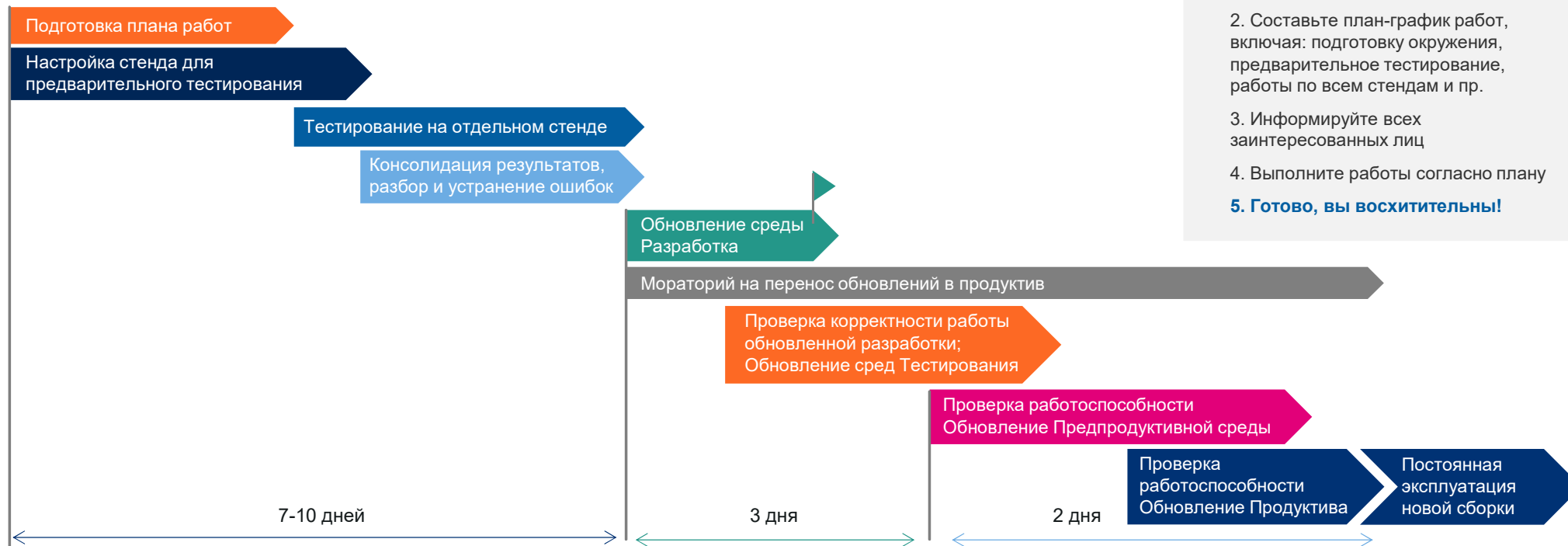
Новая сборка – это не страшно! Всего четыре шага для обновления

Предпосылки перехода:

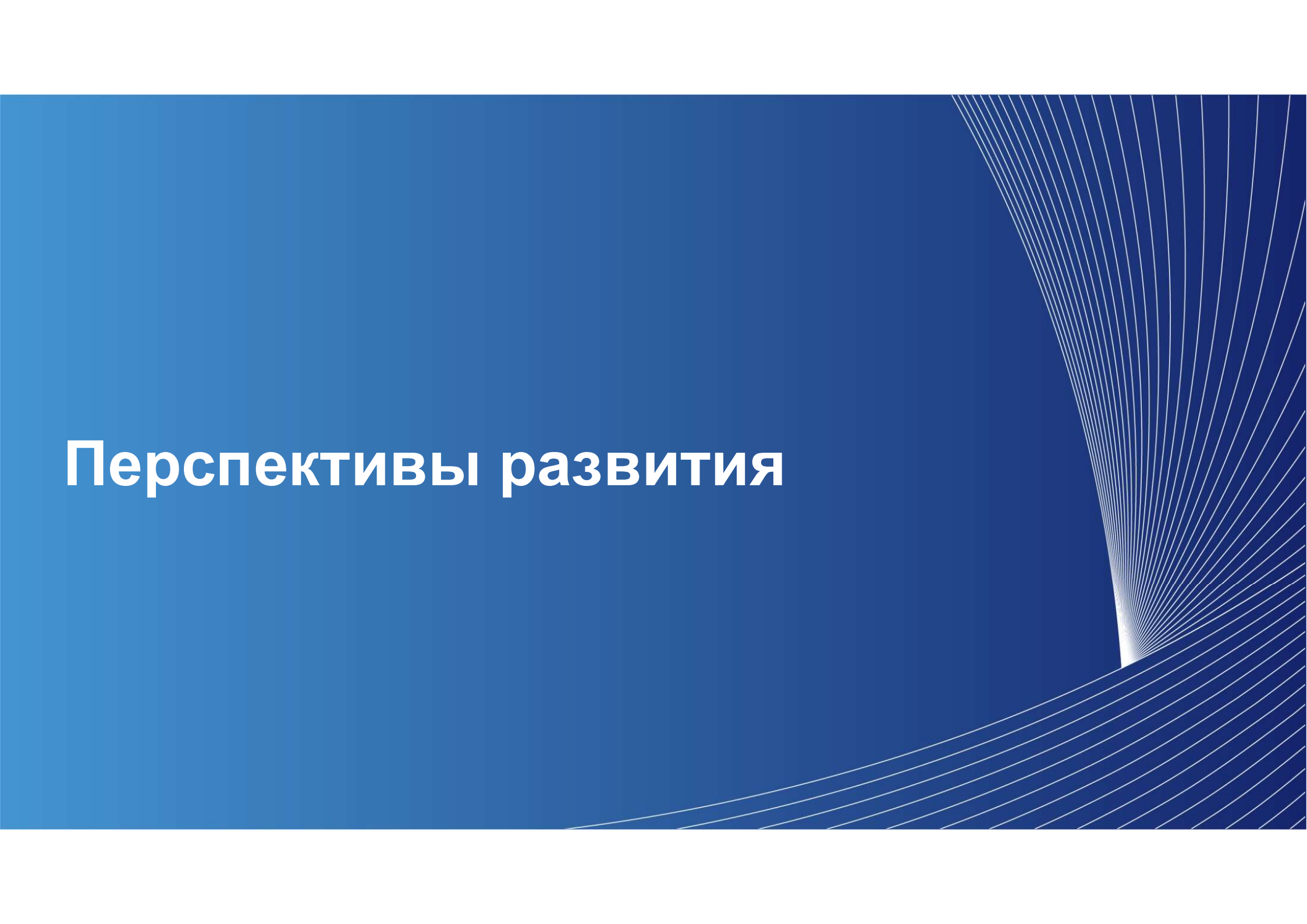
- Исправление ошибок текущей версии ФАП
- Реализация доработок и нового функционала в рамках развития инструмента
- Выпуск новой сертифицированной ФСТЭК версии
- **Снятие версии с поддержки (этого ждать не надо!)**

Чек-лист

1. Определитесь с целевой версией – в зависимости от итоговых ожиданий
2. Составьте план-график работ, включая: подготовку окружения, предварительное тестирование, работы по всем стендам и пр.
3. Информировать всех заинтересованных лиц
4. Выполните работы согласно плану
5. **Готово, вы восхитительны!**



Перспективы развития



Развитие ИАС КХД 2.0 в части аналитических инструментов



Управление данными

Интеграция инструмента управления Моделью данных ИАС КХД 2.0. Создание каталога данных. Обеспечение датацентричности



Замещение Redis

Замещение in-memory БД Redis для серверов состояний. Выбор целевого ПО, проведение стендирования и переход



Развитие Self Service BI

Определение объемов потребности в SSBI. Разворачивание инфраструктуры и разработка витрин для использования инструмента Форсайт FlyBI



Интеграция Hadoop

Интеграция в ИАС КХД 2.0 хранилища неструктурированных данных, проработка возможности поднимать данные в отчеты с холодного хранения

Спасибо за внимание!

Студенцов Павел Вячеславович
Тригубчак Евгений Петрович
Тамбовский Александр Анатольевич

