Общие сведения

Система бизнес-анализа BI (СБА)

10-01-2023

Перечень сокращений

В данном разделе представлен перечень сокращений, используемых в настоящем документе (см. Таблица 1).

Таблица 1 – Перечень сокращений

Сокращение	Расшифровка
API	Application Programming Interface, это WCF - сервис, который развернут на веб-сервере IIS
ГОСТ	Государственный стандарт
ЕХД	Единое хранилище данных
ОС	Операционная система
ПО	Программное обеспечение
СБА	Система бизнес-анализа BI - система программных средств многомерного анализа данных с использованием интегрированных хранилищ данных
НСИ	Система нормативно-справочной информации
СУБД	Системы управления базами данных
ЦСОД	Централизованная система обработки данных

Термины и определения

В данном разделе представлен перечень терминов и определений, используемых в настоящем документе (см. Таблица 2).

Таблица 2 – Перечень терминов и определений

Термин	Определение
Динамическая отчетность	Аналитическая отчетность, реализуемая Системой в виде интерактивных аналитических панелей (дашбордов), обеспечивающая пользователю возможность взаимодействия с анализируемыми данными: фильтрация, полнотекстовый поиск, сортировка, детализация и агрегирование, изменение перечня отображаемых показателей, изменение используемых аналитических разрезов и т.п.
Показатель	Является количественной характеристикой, при помощи которой предоставляются итоговые данные о единицах наблюдения (например, количество лиц или количество хозяйств). Показатель характеризуется набором признаков (ед. измерения, периодичность и др.) и справочников

Термин	Определение
Разрезность (показателя)	Разрезность показателя определяет количество признаков (справочников, измерений), в соответствии с которыми происходит хранение, сбор и отображения показателя
Таблица показателя	Сущность в ЕХД, которая содержит данные показателя
Витрина данных	Набор структурированных узконаправленных данных доступных для пользователей ЕХД.

Общие сведения

1.1. Полное наименование Системы

Система бизнес-анализа BI (Далее – Система).

1.2. Назначение Системы

Система предназначена для реализации аналитической функции Заказчика. Основным назначением Системы является автоматизация информационно-аналитической деятельности в бизнес-процессах Заказчика. Система предназначена для автоматизированной поддержки решения задач мониторинга и анализа показателей производственной деятельности.

Осуществление автоматизации информационно-аналитической деятельности в следующих бизнес-процессах:

- анализ финансово-хозяйственной деятельности;
- информационная поддержка процессов бюджетирования;
- анализ производственной деятельности;
- анализ системы мотивации и стимулирования труда;
- анализ бизнес-процессов организации.

В результате создания Системы данных улучшены значения следующих показателей:

- время сбора и первичной обработки исходной информации;
- количество информационных систем, используемых для подготовки аналитической отчетности;
- время, затрачиваемое на информационно-аналитическую деятельность;
- предоставление пользователям Системы возможностей создания интерактивной аналитики;
- повышение доступности аналитических данных для конечных пользователей;
- сокращение сроков подготовки аналитической отчетности за счет интеграции с действующими информационными системами Заказчика;
- снижение трудоемкости подготовки аналитических материалов для руководства и по внешним запросам по ключевым показателям за счет предоставления функциональных средств конструирования аналитических отчетов и интерактивных панелей.

2. Характеристика Системы

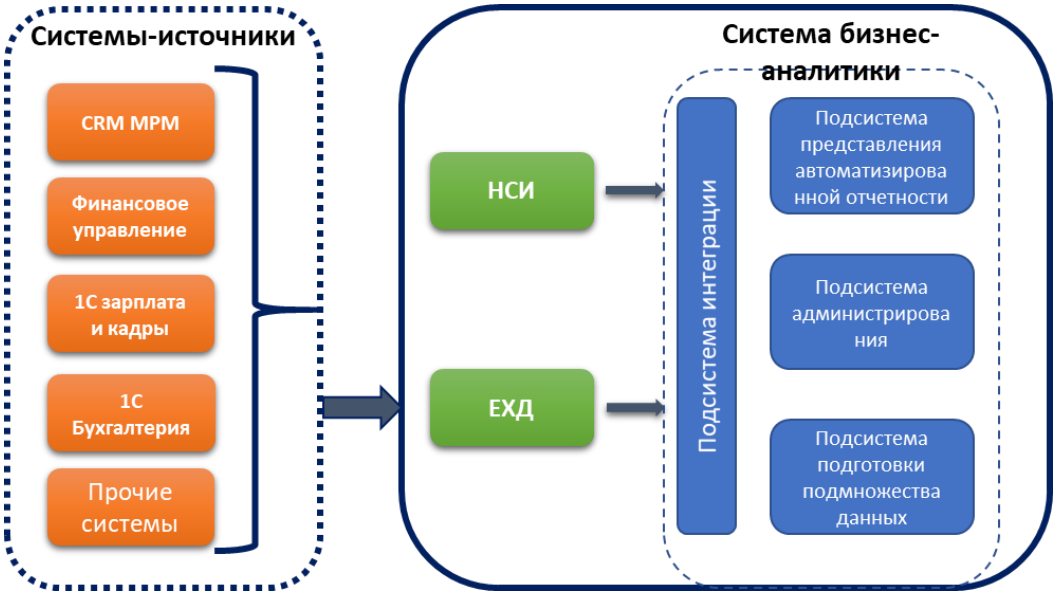
Система включает в себя следующие подсистемы:

- подсистему интеграции;
- подсистему администрирования
- подсистему подготовки подмножества данных
- подсистему представления автоматизированной отчетности
- подсистему единого хранилища данных.

2.1. Описание архитектуры и структуры Системы

Система развернута на предоставленных серверах облачного провайдера услуг и сервисов Sbercloud.

Структура Системы бизнес-анализа BI



Комплектность

Программное обеспечение Системы (см. Таблицу 1).

Таблица 1. Перечень программного обеспечения

№ п/п	Функциональная роль	Программное обеспечение и характеристики
1	Сервер СУБД	• операционная система: Ubuntu 20.04.1 Server 64(bit) • прикладное программное обеспечение: ○ postgreSQL 13.2 on x86_64-pc-linux-gnu 2-4-40-100 (2 ядра, 4ГБ оперативка, 40ГБ под ОС, 100ГБ жесткий диск раб.)
2	Сервер приложений Apache Airflow	• операционная система: Ubuntu 20.04.1 Server 64(bit) • прикладное программное обеспечение: ○ Apache Airflow 2.3.4 2-4-40-100 (2 ядра, 4ГБ оперативка, 40ГБ под ОС, 100ГБ жесткий диск раб.)
5	Сервер приложений Apache Superset	• операционная система: Ubuntu 20.04.1 Server 64(bit) • прикладное программное обеспечение: ○ Apache Superset 2-8-40-100 (2 ядра, 8ГБ оперативка, 40ГБ под ОС, 100ГБ жесткий диск раб.)
7	АРМ пользователей и разработчиков	• операционная система: MS Windows (7, 8, 8.1, 10, 11); • прикладное программное обеспечение: ○ Браузер Google Chrome, Яндекс и др.

Для реализации функций Системы использовано следующее общесистемное программное обеспечение:

– Ubuntu 20.04.1 Server 64(bit) (10GB);

– Apache Airflow 2.3.4 (подсистема подготовки подмножества данных)

– Apache Superset (подсистема представления автоматизированной отчетности)

– PostgreSQL 13.2 on x86_64-pc-linux-gnu, compiled by gcc (GCC) 7.3.0, 64-bit (для реализации модуля хранения данных);

– PostgreSQL сервис SberCloud (для реализации механизма формирования резервной копии целевой БД).

Подсистема интеграции

Подсистема обеспечивает следующие функций:

– интеграцию Системы со смежными подсистемами.

Интеграционная подсистема реализована на ETL-платформе.

Подсистема администрирования

Подсистема обеспечивает следующие функций:

– ведения списка пользователей, управления правами доступа.

Подсистема подготовки подмножества данных

Подсистема обеспечивает следующие функций:

– автоматизацию процессов загрузки и трансформации данных в подсистему единого хранилища данных;

– идентификацию и приведение к единому формату текстовых и числовых данных систем-источников;

– расчет и передачу показателей в подсистему единого хранилища данных в разрезах;

– предоставление данных во внешние системы;

– формирование отчетов в соответствии с требованиями.

Подсистема представления автоматизированной отчетности

Подсистема обеспечивает следующие функции:

– представления автоматизированной отчетности с возможностью формирования, настройки и построения графиков и диаграмм в веб-интерфейсе для бизнес-пользователей. ПС ПД реализована на платформе Apache Superset, работающей на операционной системе Linux. BI-сервер, отвечающий за работу с данными. Язык запросов – SQL.

Подсистема единого хранилища данных

Подсистема обеспечивает следующие функции:

– хранение информации:

- хранение данных в виде таблиц показателей, служебных таблиц и таблиц справочников в специальном взаимосвязанном виде, оптимизированном под извлечение и анализ данных;
- извлечение информации, содержащейся в хранилище данных, для анализа, консолидации и свода данных для дальнейшего предоставления в подсистему представления автоматизированной отчетности.

– сжатие сегментов – сжатие, как данных, так и индексов для увеличения быстродействия.

– ускорение работы SQL-запросов – ускорение SQL-запросов за счет использования возможностей индексирования и партиционирования.

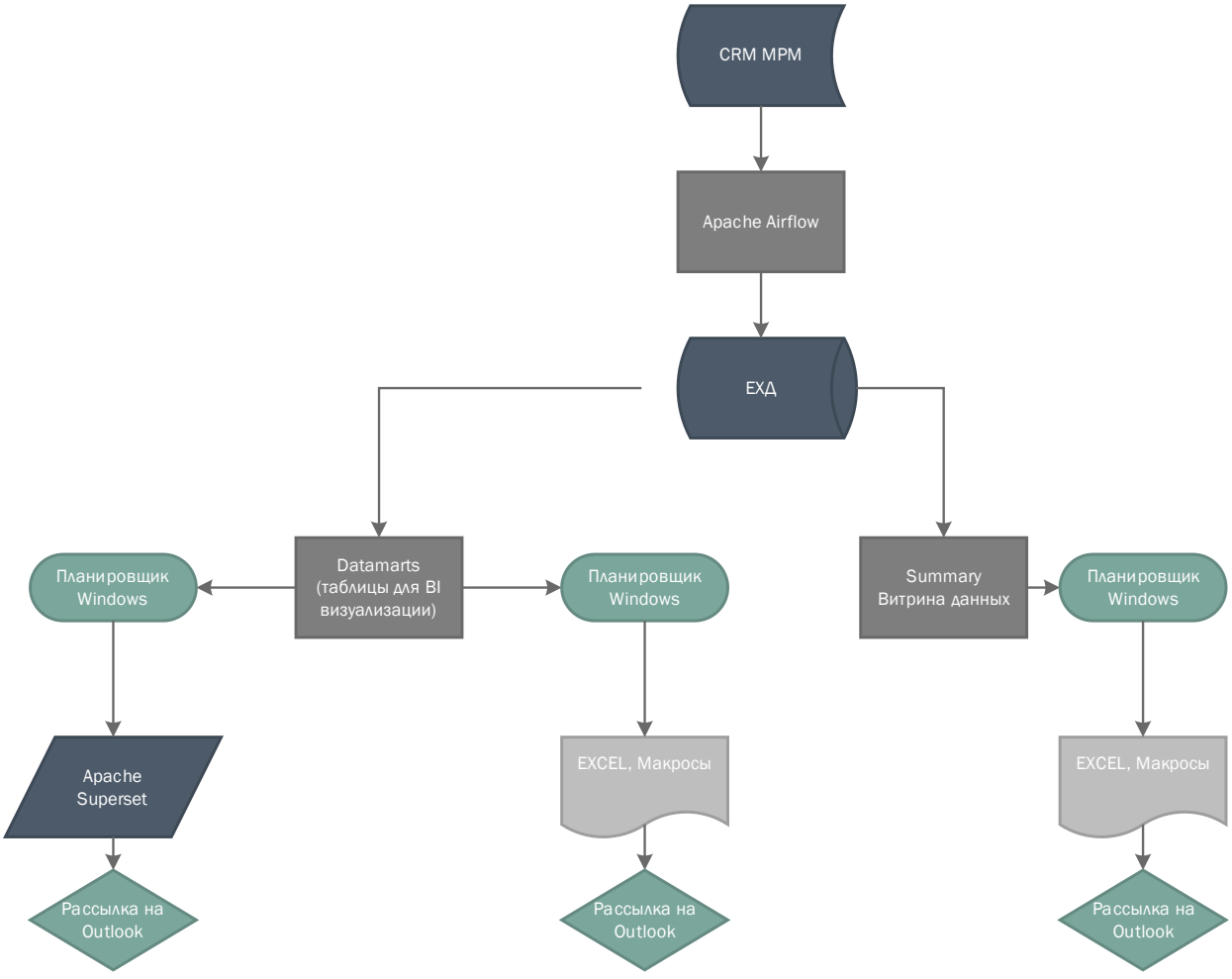
– разграничение прав доступа к данным – настройка и ограничение прав доступа пользователей к данным подсистемы единого хранилища данных, как по таблицам данных, так и по времени доступа.

– объединение серверов в кластер – возможность повышать производительность Системы в целом путем добавления в кластер нового оборудования, без замены оборудования на более производительное; также повышать отказоустойчивость СУБД: при выходе из строя или плановом выключении одного из серверов, входящих в кластер, СУБД полностью сохранит свою работоспособность.

– создание резервного сервера – возможность создания резервного сервера основной базы данных с применением всех изменений, сделанных на основном сервере, с автоматическим переключением пользователей на резервный сервер в случае выхода из строя или плановой остановки основного сервера.

– хранение исторической информации с минимальной нагрузкой на базу данных – возможность разгрузки базы данных от информации устаревшей, но требующей хранения по требованиям рабочих процессов или контролирующих органов с сохранением обычного доступа к этой информации. Использование данной функции позволяет разгрузить таблицы базы данных от накопленной и редко использующейся информации, но сохранить возможность выполнить запрос и получить такую информацию на любой момент в прошлом, то есть не требует внесения изменений в существующее приложение, работающее с базой.

Схема формирования отчетности.



2.1.1. Функциональные требования к модулю хранения данных

Модуль хранения данных обеспечивает выполнение следующих функций:

– Хранение информации:

- хранение данных в виде таблиц показателей, служебных таблиц и таблиц справочников в специальном взаимосвязанном виде, оптимизированном под извлечение и анализ данных;
- извлечение информации, содержащейся в хранилище данных, для анализа, консолидации и свода данных для дальнейшего предоставления в систему аналитики данных.

– Сжатие сегментов – сжатие, как данных, так и индексов для увеличения быстродействия.

– Ускорение работы SQL-запросов – ускорение SQL-запросов за счет использования возможностей индексирования и партиционирования.

– Разграничение прав доступа к данным – настройка и ограничение прав доступа пользователей к данным модуля хранения данных, как по таблицам данных, так и по времени доступа.

– Объединение серверов в кластер – возможность повышать производительность СБА в целом путем добавления в кластер нового оборудования, без замены оборудования на более производительное; также повышать отказоустойчивость СУБД: при выходе из строя или плановом выключении одного из серверов, входящих в кластер, СУБД полностью сохранит свою работоспособность.

– Создание резервного сервера – возможность создания резервного сервера основной базы данных с применением всех изменений, сделанных на основном сервере, с автоматическим переключением пользователей на резервный сервер в случае выхода из строя или плановой остановки основного сервера.

– Хранение исторической информации с минимальной нагрузкой на базу данных – возможность разгрузки базы данных от информации устаревшей, но требующей хранения по требованиям рабочих процессов или контролирующих органов с сохранением обычного доступа к этой информации. Использование данной функции позволяет разгрузить таблицы базы данных от накопленной и редко использующейся информации, но сохранить возможность выполнить запрос и получить такую информацию на любой момент в прошлом, то есть не требует внесения изменений в существующее приложение, работающее с базой.

Система имеет круглосуточный режим работы (24х7). Допускаются технологические перерывы для выполнения резервирования и проведения регламентных работ (копирование БД, установка новых версий ППО и т.п.), при этом технологический перерыв не подразумевает полной недоступности системы для бизнес-пользователей, но допускает снижение производительности Системы.

Система имеет техническую возможность создания резервных копий данных встроенными средствами резервного копирования на базе Sbercloud. Максимальное время восстановления Системы после сбоя составляет не более 8 часов. Максимальное время восстановления – промежуток времени от обнаружения инцидента, который привел к

отказу в функционировании Системы, до полного восстановления функционирования Системы.

Система развернута на имеющихся у Заказчика ресурсах. Выделяемые для Системы ресурсы соответствуют требованиям, указанным в эксплуатационной документации производителя Apache Airflow, с возможностью модернизации.

Процесс сбора, обработки, передачи информации в Системе реализованы в автоматическом режиме. Данные из систем-источников выгружаются ежедневно средствами Apache Airflow.

В качестве системы резервного копирования информации для серверов ИС ЕХД используется сервис, предоставляемый облачным провайдером SberCloud.

Для серверов ИС ЕХД настроен следующий план резервного копирования:

- полный Backup – 1 раз в день в 00:50 по МСК;
- инкрементальный – каждые 5 минут, в зависимости от наличия инкремента.

ЕХД состоит из нескольких слоев хранения данных и модулей, каждый из которых имеет определенную направленность.

Stage - данные извлекаются из всех подключенных источников, очищаются от излишней информации, обогащаются необходимыми полями;

Facts – очищенные и структурированные данные, на данном слое происходит контроль качества загруженных данных. Слой предоставляет аналитику быстрый доступ к необходимой информации и построение AD HOC отчетности;

Summary – слой, содержащий предрассчитанные агрегации данных (например: таблица сгруппированная по клиентам с помесячными данными по звонкам/встречам/лидам);

DATAMARTS – слой подготовленных таблиц для использования их в настройке визуализации отчетности в инструменте BI.