

beeline **cloud**

**Облако или on-premise?
Честное сравнение подходов
обработки данных**

Agenda

1 Основные трудности при работе с данными

2 Что такое платформа данных и зачем это бизнесу?

3 Строим сами или идем в облака?

4 СУБД, BI, S3 – простые инструменты для построения КХД?

5 Расчет TCO на примере эксплуатируемых решений

6 Вопросы и ответы

Проблемы с данными

Много информационных систем

- Рассогласованность данных между базами данных систем (1С, CRM, ERP и т.п.)
- Привязанность к конкретному вендору
- Нет совместимости данных друг с другом

Ручная работа

- Нет непрерывного бизнес-процесса
- Недостаток компетенций
- Много ошибок в Excel

Скорость отчетности

- Долго строятся отчеты
- Долгая загрузка данных в хранилища построения отчетности

Развитие ИИ

- Долго загружается информация на сайте, на корп. портале

Нет единого окна

- Нет единого окна со всей корпоративной отчетностью в компании
- Необходимы отчеты в реальном или околореальном времени

Много потребителей данных

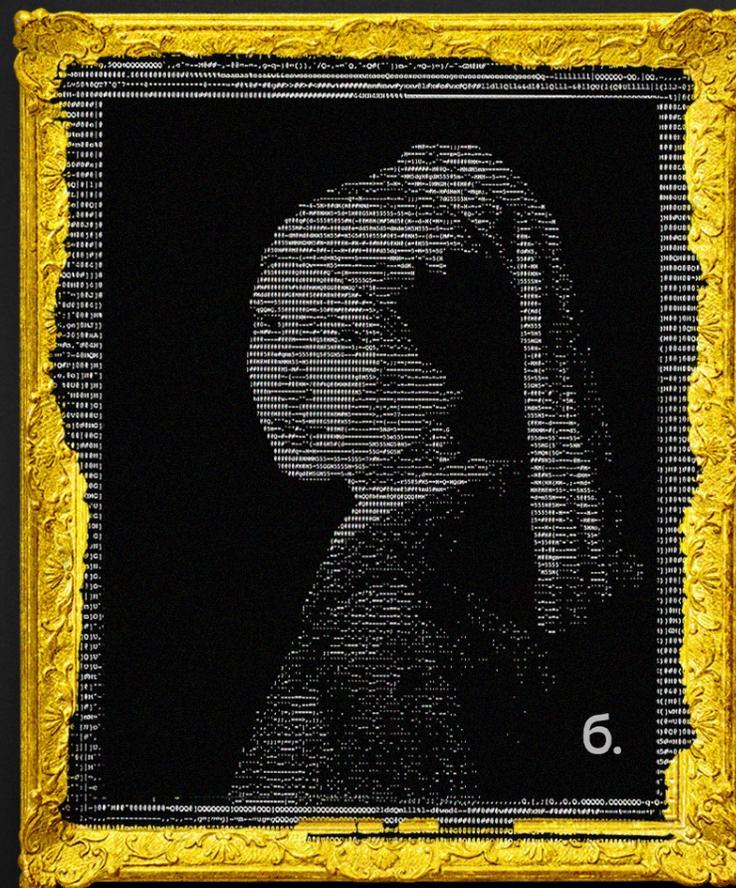
- BI-витрина не справляется с пользовательской нагрузкой, есть риск нарушения непрерывности бизнеса

Данные нужны в различных представлениях

- Кратный рост стоимости трансфера данных, отсутствие прозрачности передачи данных
- Большой объем сырых данных, которые представляют ценность, но их негде хранить

beeline **cloud**

Что такое
платформа
данных и зачем
это бизнесу?



Information Systems
Management
Information Systems
Management



Инструменты для работы с данными

Корпоративное хранилище данных

- OLAP и аналитические запросы
- Построение единого КХД
- СУБД для систем клиентской аналитики (CRM)
- СУБД для систем управления маркетинговыми кампаниями
- Реляционная облачная MPP-СУБД Greenplum

Система визуализации и аналитики данных

- BI-система для анализа и визуализации данных конечными пользователями (финдиректора, коммерческая функция, руководители)

Сбор данных

- Шина данных для сбора на базе Apache Kafka

Озеро данных

- Построение озера данных на базе Hadoop для хранения слабоструктурированных данных

Кеширование данных

- Использование Redis как дешевого кеша или построение Data Grid

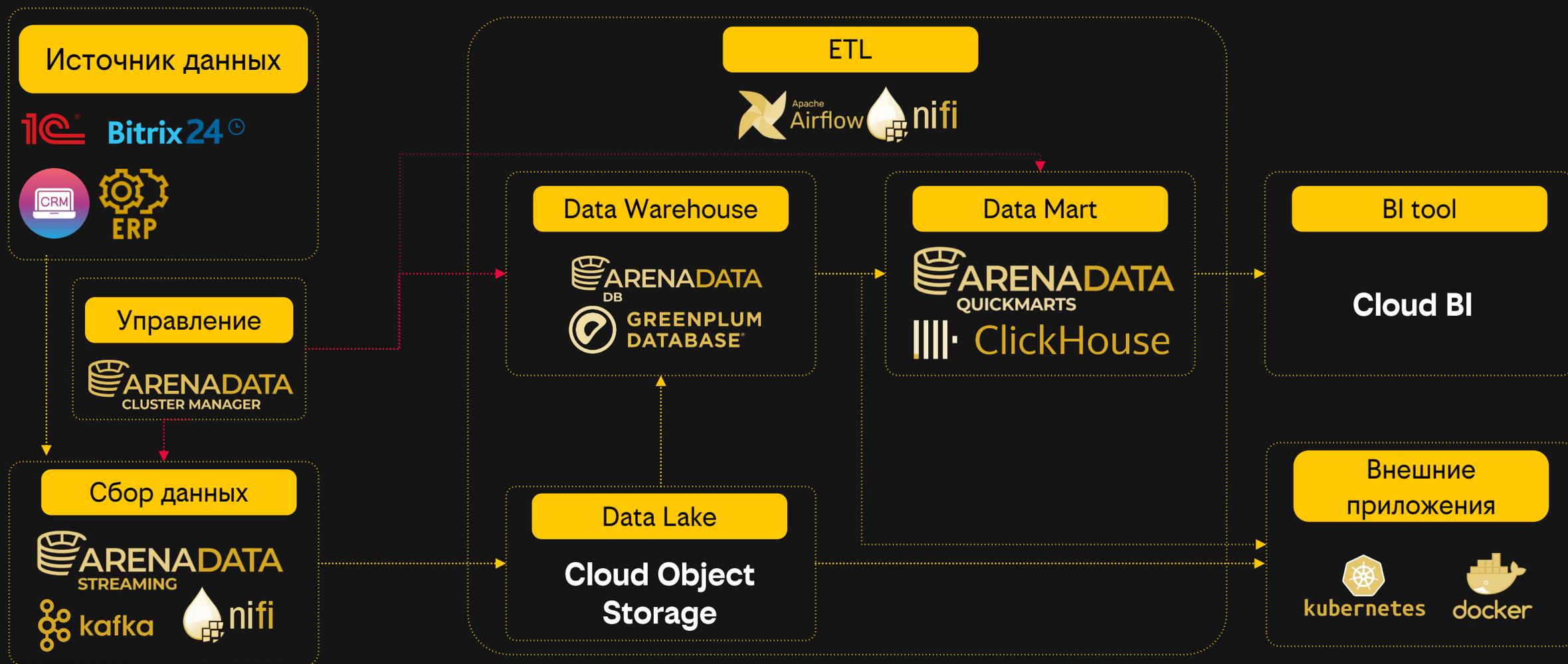
Витрины данных

- OLAP-СУБД Clickhouse для «горячего» доступа к данным для подготовки аналитических отчетов

Преобразование данных

- Инструменты для реализации ETL (extract, transform, load) процессов, трансформации и перегонки данных в разные хранилища на базе Apache Nifi

Платформа данных



Выгода для бизнеса

1

Скорость

Использование платформы данных в процессах компании положительно влияет на их скорость

- Быстрая синхронизация систем
- Быстрое построение отчетности
- Быстрый трансфер данных между системами до конечных потребителей

2

Стандартизация

Внедрение подхода управления данными на базе платформы данных обеспечивает точность, чистоту и стандартизацию данных

- Данные из всех систем унифицируются
- Данные из всех систем согласованы в едином месте
- Качество данных соблюдается согласно стратегии управления данными

3

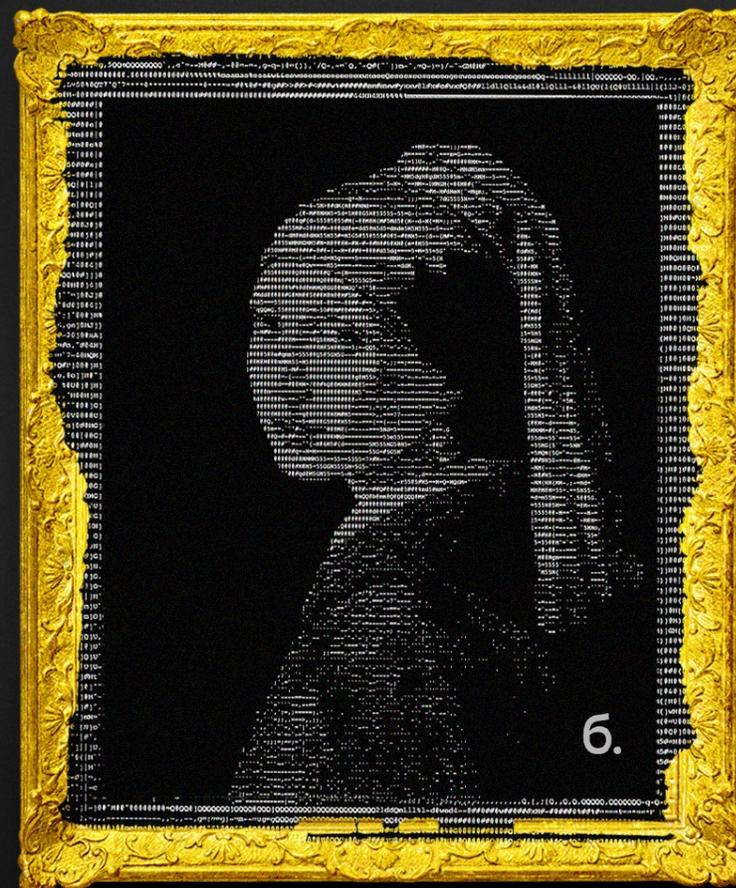
Новые сценарии

Управляя данными в компании можно развивать новые сценарии их использования

- Искусственный интеллект
- Предиктивная аналитика
- Обучение и инференс нейронных сетей

beeline cloud

Строим сами или
идем в облако?



Содержание
1. Введение
2. Основные понятия
3. Структура программы
4. Алгоритмы
5. Примеры
6. Заключение

Отличие On-Premise и Cloud

	On-Premise	Cloud
Затраты	CAPEX	OPEX
Устойчивость	Ограничение собственным железом	Гибкое масштабирование
Контроль данных	Полностью на стороне клиента	Провайдер обеспечивает доступность решения для работы с данными
Уровень и скорость настройки	Руками, начиная с инсталляции серверов	IaaS, PaaS, SaaS в зависимости от стоимости и задач
Доступность	В рамках собственной сети	Интернет везде и всегда

Ключевые преимущества обработки данных в облаке



Оптимизация TCO

Оптимизация расходов на эксплуатацию решения в обычном режиме и во время нештатных ситуаций

- Снижение затрат на закупку железа (CAPEX -> OPEX)
- Сокращение TCO на обслуживание внутренних бизнес-процессов
- Оптимизация поддержки при возникновении нештатных ситуаций



Масштабирование и гибкая тарификация

- Помесячная оплата за использование ресурсов и лицензий
- Возможность расширяться и сужаться в зависимости от нагрузки и объема данных
- PoC за несколько недель вместо долгой и сложной закупки «железа»

Пример сравнения инфраструктур для ADB в крупном ритейлере

Аппаратная спецификация Arenadata ADB



Arenadata Database Production 2024 (115TB Compressed data), ядер к лицензированию: 864

Тип узла кластера	Конфигурация аппаратного обеспечения узла	Количество узлов	Ядер к лицензированию на узел	Bare-metal Virtual	Примечание
Arenadata DB: Master Server	CPU: 36/48 cores (2 x 18 или 2 x 24, 2.4+ GHz) HT enabled, min Intel Cascade Lake (R) Gold RAM: не менее 512 GB (*) Storage: 2 x 480 GB HDD / SSD RAID1 for OS + 2 x 3.84TB SAS12G SSD MixedUse RAID1 (PCIe Gen3 x4+) for Data (1100+ mb/sec sequential read/write recommended) Network: 2 x 10/25 Gbps (PCIe Gen3 x4) + 2 x 50 Gbps (PCIe Gen3 x8)	2	0	Bare Metal	Коммутатор сети Interconnect должен поддерживать размер MTU равный 8300.
Arenadata DB: Segment Server	CPU: 48 cores (2 x 24, 2.4+ GHz) HT enabled, only Intel, min Cascade Lake (R) Gold RAM: не менее 768 GB (*) Storage: 2 x 480 GB HDD / SSD RAID1 for OS + 24 x 1.6 TB SAS12G SSD MixedUse RAID10 (3 controllers, each PCIe Gen3 x8) for Data (10000+ mb/sec sequential read/write recommended) Network: 2 x 10/25 Gbps (PCIe Gen3 x4) + 2 x 50 Gbps (PCIe Gen3 x8)	18	48	Bare Metal	Каждый узел хранит и обрабатывает 6.4 TB сжатых данных. Коммутатор сети Interconnect должен поддерживать размер MTU равный 8300.

Arenadata Database DEV 2024, ядер к лицензированию: 8

Тип узла кластера	Конфигурация аппаратного обеспечения узла	Количество узлов	Ядер к лицензированию на узел	Bare-metal Virtual	Примечание
Arenadata DB: Master Server	CPU: 4-6 vCores (1 pCore with HT enabled = 2 vCore), Intel Cascade Lake (R) 62xx RAM: 48 GB SDS / Storage: 250 GiB HDD for OS + 1.2 TiB SSD for Data Network: 1 x 10 Gbps	2	0	Virtual	Рекомендуется обеспечить в сетевой инфраструктуре поддержку IP-фреймов, равную 8300 байт.
Arenadata DB: Segment Server	CPU: 8 vCores (1 pCore with HT enabled = 2 vCore), Intel Cascade Lake (R) 62xx RAM: 64 GB SDS Storage: 250 GiB HDD for OS + 4 TiB SSD/NVME for Data Network: 1 x 10 Gbps	2	4	Virtual	Рекомендуется обеспечить в сетевой инфраструктуре поддержку IP-фреймов, равную 8300 байт.

Локальная ИТ под кластер ADB



Solution Name: arndt					
Solution Category: General					
Solution List Price: Курс 1\$ =91,64 руб		181 030 403 P			
Group Name	Product Name	Product Qty	Unit Price (List)	Extended Price (List)	
Group 1	master_PowerEdge R650 - Full Configuration - [EMEA_R650]	2	\$41755,21	\$83510,42	
Group 1	segment_PowerEdge R750 - Full Configuration - [EMEA_R750]	18	\$102099,61	\$1837792,98	
Group 1	virt-dev_PowerEdge R650 - Full Configuration - [EMEA_R650]	1	\$29985,21	\$29985,21	
Group 1	adcc_PowerEdge R650 - Full Configuration - [EMEA_R650]	1	\$24163,19	\$24163,19	
				\$1 975451,80	

Product short	Version	Type	Program	Metrics	Period	Цена с НДС, руб.	Наименование	Кол-во ядер	Сумма без НДС
ADB	Enterprise	Prod	PL	1 core	Perpetual	345 000 P	Неисключительное право на использование ПО Arenadata Analytical DB (ADB) для продуктивной среды.	864	298 080 000 P
ADB	Enterprise	Prod	Supp PL 24x7	1 core	3 years	248 400 P	Сопровождение ПО Arenadata Analytical DB (ADB) для продуктивной среды, пакет "Расширенный", 1 год	864	214 617 600 P
ADB	Enterprise	Dev	PL	1 core	Perpetual	172 500 P	Неисключительное право на использование ПО Arenadata Analytical DB (ADB) для среды разработки.	8	1 380 000 P
ADB	Enterprise	Dev	Supp PL 8x5	1 core	3 years	62 100 P	Сопровождение ПО Arenadata Analytical DB (ADB) для среды разработки, пакет "Базовый", 1 год	8	496 800 P
									514 574 400 P

Облачная ИТ под ADB кластер

Greenplum

Нефункциональные требования		Цена в месяц без НДС				
Целевая доступность сервиса на запись	99,50%					
Объем сжатых данных для хранения, Тб	126					
Хранение резервных копий в распределенном катастрофоустойчивом хранилище	300Тб	135000				
Point-in-Time Recovery (PITR) с точкой консистентности (RPO)	10 минут					
Предложение по целевой конфигурации						
		vCPU	SSD, Тб	RAM, Гб	Количество	
Прод. - выделенные сервера	Мастер	48	3	428	2	
	Сегмент-нода	104	17,5	624	18	
Дев. Виртуальные машины	Мастер	8	0,5	64	2	
	Сегмент-нода	12	4	96	2	
	Итого:	172	25	1212	24	
	Цена за ед. в месяц без НДС	287,00	5595,00	230,00	Итого инсталляция в месяц без НДС	Итого инсталляция в год с НДС
	Итого в месяц без НДС	577000	1846647	2853840	5412487	77 939 813 Р

Product short	Product	Version	Type	Program	Metrics	Period	Наименование	Цена без НДС, руб.	Кол-во ядер	Сумма с НДС в год, руб.
ADB	Arenadata Analytical DB	Enterprise	Prod	Temporary license with warranty support 24/7	License 1 Core	1 month	Неисключительное право на использование ПО Arenadata Analytical DB (ADB).	5 716 Р	864	71 114 941 Р
ADB	Arenadata Analytical DB	Enterprise	Dev	Temporary license with warranty support 8/5	License 1 Core	1 month	Неисключительное право на использование ПО Arenadata Analytical DB (ADB).	2 287 Р	8	263 451 Р
										71 378 392 Р

Сравнение затрат в перспективе лет

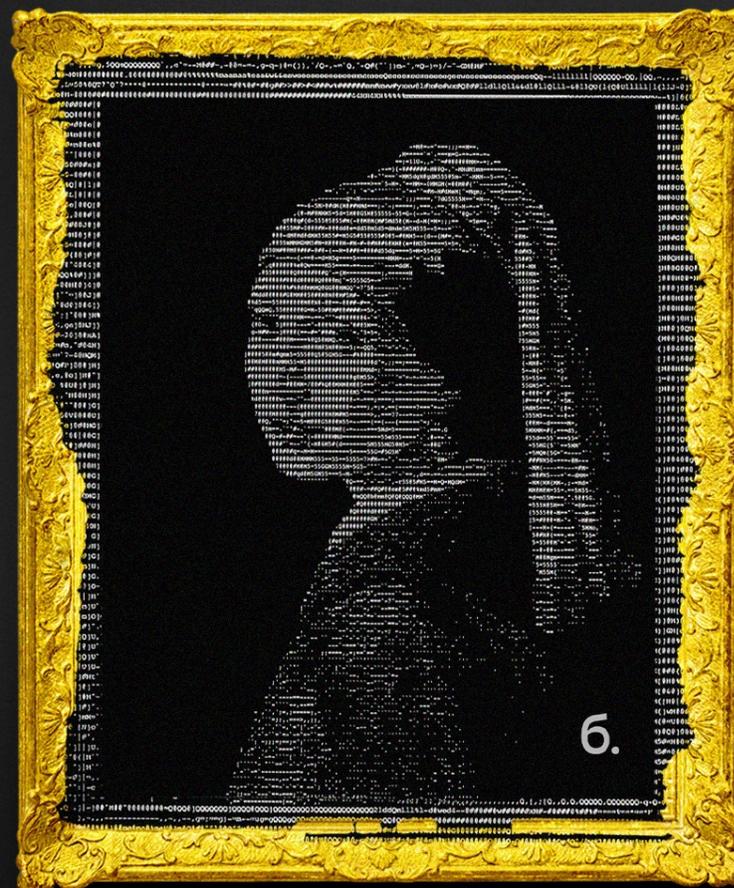
Инфраструктура на земле	
Hardware	181 030 403 Р
Software	514 574 400 Р
Итого	695 604 803 Р
	4,66

Инфраструктура в облаке в год с НДС	Более 4 лет 7 мес (55 месяцев)
Hardware	77 939 813
Software	71 378 392
Итого	149 318 205

beeline cloud

+8

причин к выгоде
сервисов над
локальной ИТ



Informational text block, likely a description or label for the artwork.



1. Безопасность данных и ресурсов в облаке

- Соответствие требованиям: соответствие стандартам, ISO, УЗ-1, PCI DSS
- Конфиденциальность: вы владеете собственными данными и контролируете их
- Прозрачность: вы знаете, как хранятся ваши данные, кто получает к ним доступ и каким образом обеспечивается их безопасность
- Защита от DDOS и хакерских атак
- Обеспечение безопасности сетей
- Управление угрозами

Совокупная стоимость может превышать 100+млн руб в год

Примерно 8 месяцев потребления облачного сервиса

2. Экономия оборотных средств

Лок. ИТ (железо + лицензии равна N месяцев использования облака)

$$\sum_N a \cdot (y - N \cdot x)$$

a – коэффициент оборачиваемости средств в год («сколько вы зарабатываете с 1 рубля»)

y – Стоимость лок. ИТ

x – Стоимость облака в месяц

N – кол-во месяцев, когда расходы облака сравняются с лок. ИТ

Оборачиваемость средств клиента ~20% (0,2)

1 год = 109 257 320 ₽

2 год = 79 393 679 ₽

3 год = 49 530 038 ₽

4 год = 19 666 397 ₽

Итого 257 847 432 ₽

Равноценно потреблению облачного сервиса ~ 21 месяц (1 год 9 мес)

3. «Деньги сегодня дороже, чем завтра»

Инфляция 7-15 % (за 4 года ~40%)

Удорожание товаров и услуг происходит сильнее инфляции

1 год	54 628 660 ₽
2 год	39 696 839 ₽
3 год	24 765 019 ₽
4 год	9 833 198 ₽
Итого	128 923 716 ₽

Равносильно 10 месяцам потребления облачного сервиса

4. «Инфраструктура не должна работать в холостую»

При графике работы с 9-18,

понедельник-пятница

итого 22 дня,

всего 196 рабочих часов

в месяц

Это всего лишь 26%

рабочего времени в месяц

5. Стоимость отказоустойчивости лок. ИТ

Нужна система резервного копирования
Нужен резервный сервер с репликацией
и многое другое (стоимость лок. ИТ умножить на 3)

$181\ 030\ 403 * 3 = 543\ 091\ 209\ \text{₽}$

Удорожание: $543\ 091\ 209\ \text{₽} - 181\ 030\ 403 = 362\ 060\ 806\ \text{₽}$

Что равносильно использованию облачного сервиса 29 месяцев (2 года 5 мес)

6. «Сколько стоит 1 день простоя вашей компании?»

Потери от простоя за 1 день = Сумма з/п всех сотрудников за 1 день +
Однодневный оборот компании + % штрафа от неоплаченных счетов, налогов + прочее

15 000+ работников средняя ЗП в ритейлере `30 тыс руб

Тогда ФОТ = 450 млн. руб, не может быть больше 0,3 всей прибыли

Прибыль = 450 млн. руб/0,3 = 1,5 млрд. руб

Маржинальность 20%

Выручка = 1,5 млрд. руб/0,2 = 7,5 млрд. руб

30 рабочих дней

Выручка в день = 250 млн. руб — цена простоя

7. Предсказуемость расходов

Снижение
замороженного
резерва на черный
день



8. Прочие расходы (ФОТ админа, аренда доп. помещения, система вентиляции, электричество и проч.)

Итого

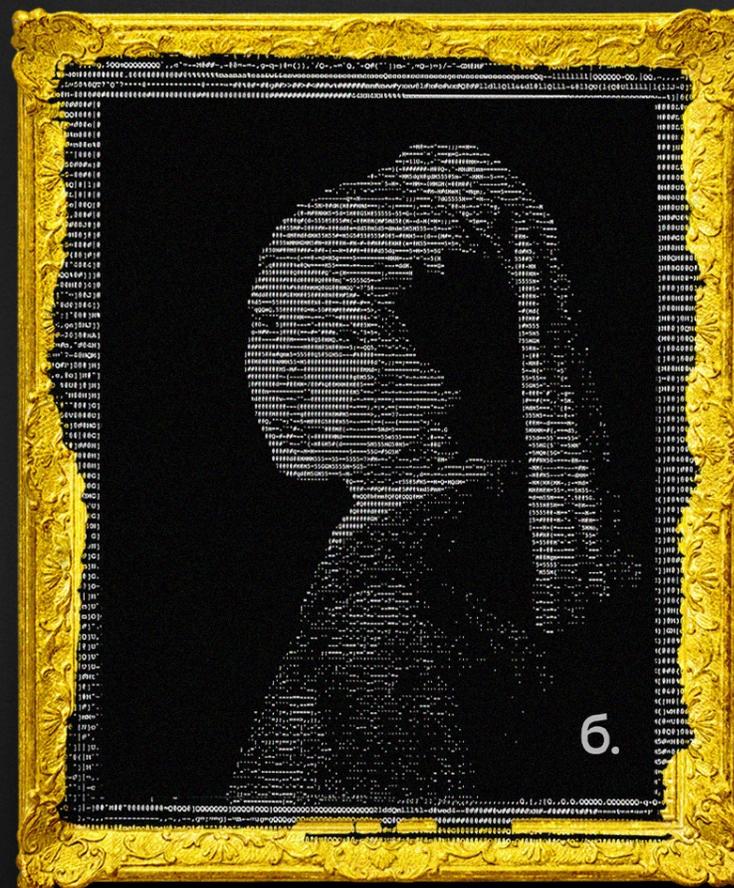
$55+8+21+10+29=123$

месяца (более 10 лет)

Средняя окупаемость не должна быть более 5 лет

beeline **cloud**

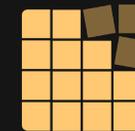
**СУБД, ВІ, S3 —
простые
инструменты
для построения
КХД**



Содержание
1. Введение
2. Описание системы
3. Архитектура
4. Описание компонентов
5. Описание работы системы
6. Заключение



Cloud Object Storage



Описание

Облачное S3 хранилище на базе российской разработки для хранения резервных копий, архивов, больших данных с мгновенным доступом и IAM для управления идентификацией

- Отказоустойчивость и репликация
- Репликация по бакетам
- Защита от DDoS/WAF на уровне сервиса
- Полная совместимость с AWS S3
- Object Lock – WORM model
- Нетарифицируемые GET и POST запросы и нет разделения на «горячее» и «холодное»

Решает задачи

Хранение бэкапов

Хранение документов

Хранение сырых данных в качестве озера данных

Типовой клиент

20-30 ТБ данных

Топ-10 компаний в своей отрасли

Банки ритейл телеком, производство

Вид предоставления

Paas

От 100 ГБ

Стандартный интернет-канал
150 мбит\с

Возможность расширения
канала на 100, 200, 300, 500,
700, 900 1000 мбит\с

Описание

Аналитическая распределенная массово-параллельная СУБД, построенная на базе проекта с открытым исходным кодом Greenplum.

- Двусторонние коннекторы
- ADB Control (мониторинг на уровне запросов)
- ADBM – управление бинарными бэкапами
- Мониторинг и алертинг
- Управление через ADCM
- Включена техподдержка
- Типовые конфигурации на 3, 5 и 10 ТБ данных

Решает задачи

Реализация аналитических хранилищ
Сложные аналитические запросы
Построение КХД
Эффективное соединение таблиц

Типовой клиент

3+ ТБ данных и тенденция к росту, наличие аналитических запросов, готовность идти в облако, отрасли: Финтех, Ритейл, промышленность, гейминг, ИТ-компания

Вид предоставления

PaaS на базе нашего IaaS.

- Инфраструктурные ресурсы + лицензия на облачное ПО
- Лицензия считается по количеству ядер на сегмент-серверах (там где хранятся данные)

ADB Enterprise Light

- Dev+Test+Prod среда по единой цене в месяц
- До 10 несжатых данных клиента (сертифицировано вендором)





Описание

Сервис визуальной аналитики данных для построения корпоративной отчетности (дашбордов), анализа данных, управления метриками и мониторингом бизнес-процессов в компании.

- Более 40 графиков визуализации
- Продвинутой SQL-движок для анализа данных в одном окне
- Возможность Drill-down и подключения всех возможных СУБД и хранилищ данных

Решает задачи

Корпоративная отчетность, отслеживание тенденций бизнеса, замещение PowerBI, мониторинг внутренних процессов, сегментация клиентской базы, анализ рынка

Типовой клиент

Enterprise (финтех, ритейл, промышленность, медтех)

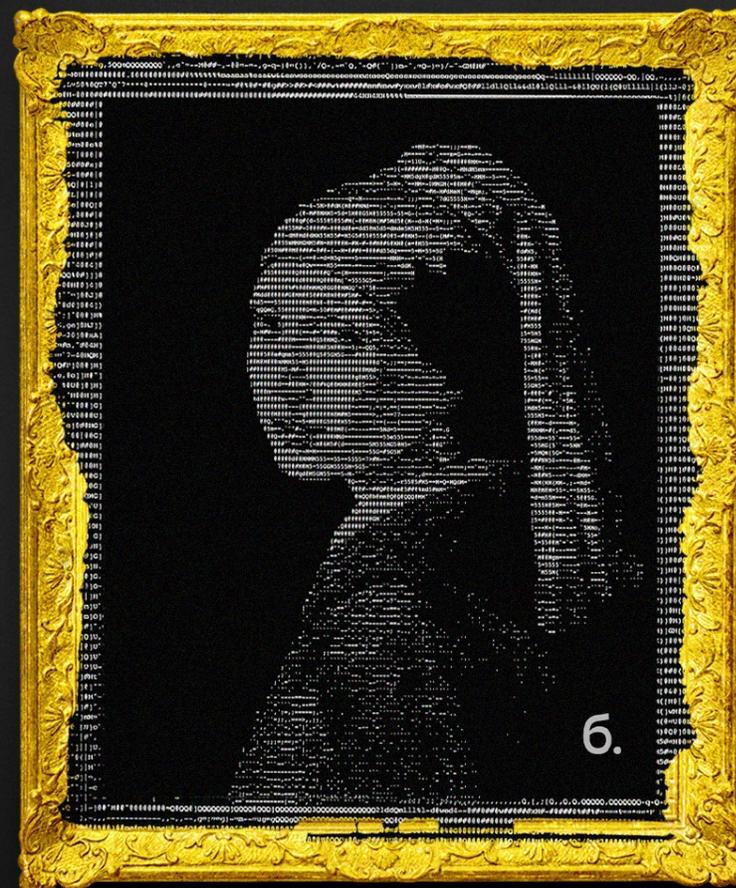
Вид предоставления

Модель предоставления сервиса «ПО как услуга» по подписке.

Тарификация по количеству пользователей (пакет + свыше пакета)

beeline cloud

Вопросы и ответы



Informational text block, possibly a list of questions or answers, located to the right of the main artwork.