



Informatika



AI и Copilot: Революция в управлении данными

DIS Group Kazakhstan

Болат Таймагамбетов

Сентябрь, 2024 v. 1.0

ОФИЦИАЛЬНЫЙ ДИСТРИБЬЮТОР В СНГ

DIS Group предлагает решения в областях:

- Управления данными
- Бизнес-аналитики
- Управления знаниями
- Защиты информации на базе технологий признанных международных лидеров



14+ лет

30+ партнеров

250+ реализованных проектов

200+ клиентов компаний-лидеров

ПРАКТИКА ВНЕДРЕНИЯ И ПРОДУКТИВНАЯ ЭКСПЛУАТАЦИЯ В КАЗАХСТАНЕ

БОЛЕЕ 15 ЛЕТ ПРОДУКТИВНОЙ ЭКСПЛУАТАЦИИ РЕШЕНИЙ INFORMATICA

Решение/Заказчик	Ключевые Заказчики								
ETL/ELT	Халык Банк	Банк Центр Кредит	Форте Банк	Береке Банк	Отбасы Банк	Евразийский банк	Билайн КЗ	Magnum	Jusan Банк
Бизнес-гlossарий	Халык Банк		Банк Центр Кредит		Казахтелеком		КазМунайГаз		
Каталог	Халык Банк		Банк Центр Кредит		Казахтелеком		КазМунайГаз		
Data Quality	Халык Банк		Банк Центр Кредит		Казахтелеком		КазМунайГаз		
MDM (НСИ)	КазМунайГаз					ERG			

Решения Informatica эксплуатируются в продуктиве с 2008 год по настоящее время

Куда все движется?

100 миллионов пользователей

используют ChatGPT каждую неделю!

Все **ожидают**
похожего опыта

взаимодействия в организациях

К 2026 году **95% работников**
будут использовать в своей работе AI для
выполнения задач

Source: Gartner, 2024



Source: ChatGPT 4o

Самый популярный интерфейс на 2024 год

CLAIRE[™] GPT

Need help getting started? Try one of these sample prompts.

Discover

Show me the assets which are profiled in last 30 days

Find catalog sources of type Snowflake, Databricks, and Oracle

Show me the assets controlled by **Bolat Taimagambetov**

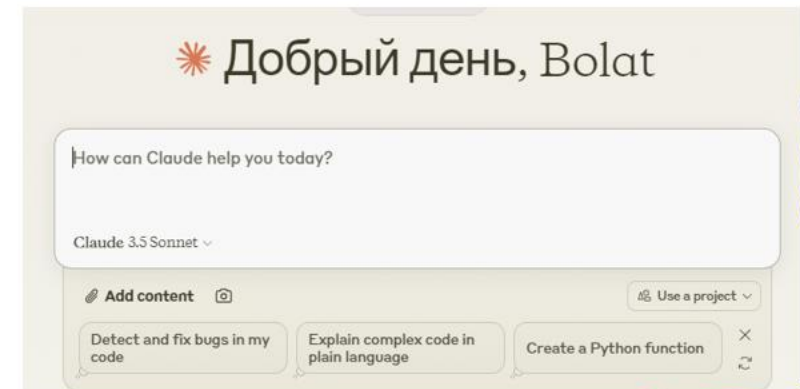
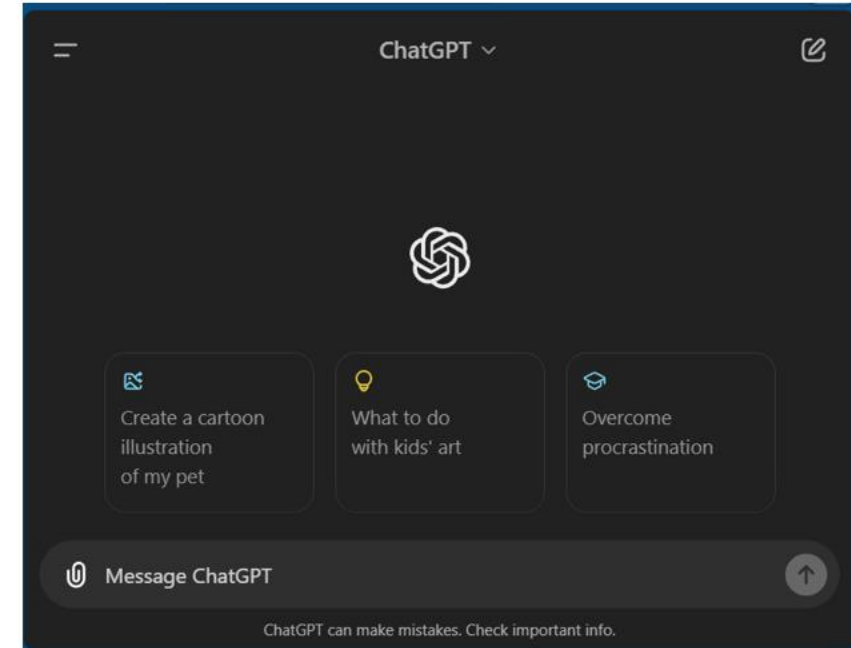
Ask CLAIRE...

Ask

What API can be used to get the details of an asset in CDGC?

How do I set up a bulk data import task?


Can mass ingestion cause any restrictions when using SQL Server?



GenAI: Данные, говорящие на вашем языке

Q: Покажи мне список отчетов с персональными данными.

Q: Какой объем продаж в филиале в Алматы?

Когда GenAI  встречает данные, Естественный Язык становится Языком Данных

Q: Почему судник так долго формируется?

Q: Удали дубликаты в моем отчете

Сегодня самый популярный «язык программирования» - Английский

Как выглядит будущее Управления данными?



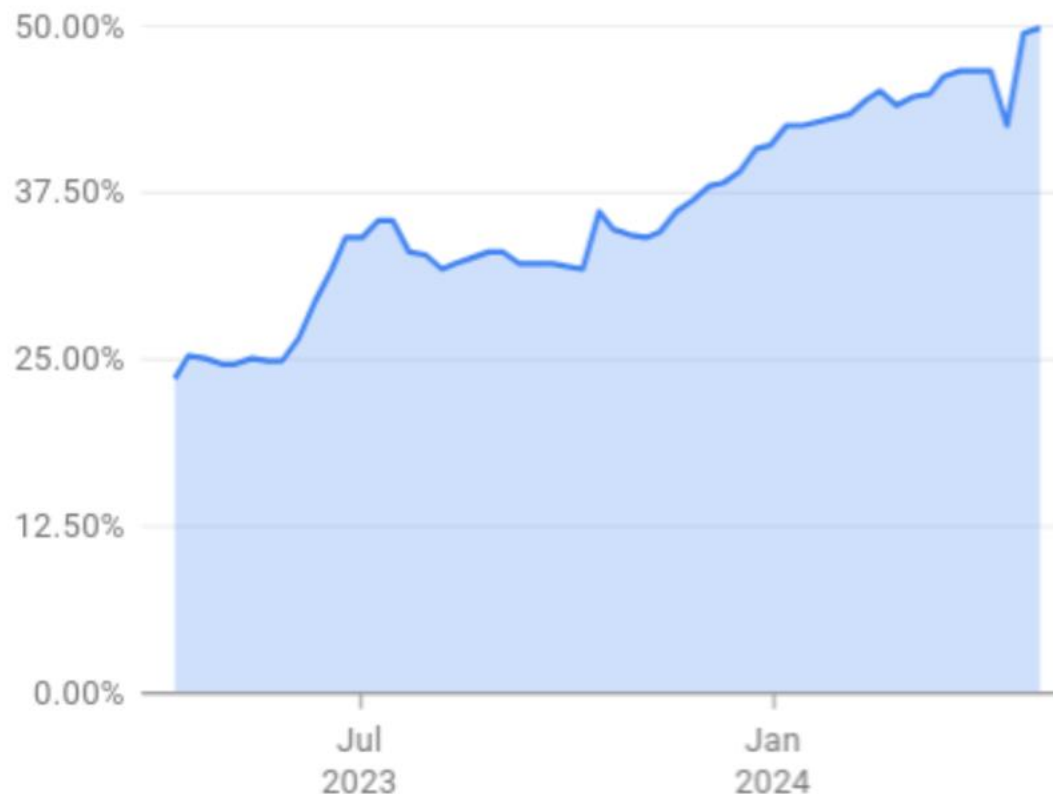
Выгоды

- Упрощение пользовательского опыта
- Быстрый и гибкий доступ
- Централизованное управление метаданными и Data Governance
- Сокращение затрат за счет упрощения работы с данными с помощью AI

Влияние AI на ИТ-персонал неуклонно растет

Google. Рост использования AI-инструментов автодополнения кода, который выражается в увеличении доли символов, принятых из AI-подсказок, по сравнению с вручную введенными символами

Amazon. Руководитель облачного подразделения Amazon заявил, что в ближайшем будущем ИИ может заменить большую часть работы разработчиков, делая их участие в кодировании менее необходимым



BUSINESS INSIDER

AI

In a leaked recording, Amazon cloud chief tells employees that most developers could stop coding soon as AI takes over

[Eugene Kim](#)

От нас не убежишь!



AI in software engineering at Google: Progress and the path ahead

Основные проблемы при внедрении генеративного AI

Точность, достоверность
и прозрачность данных



46% Качество данных



Защита, безопасность данных,
регулирующие требования

Направления развития

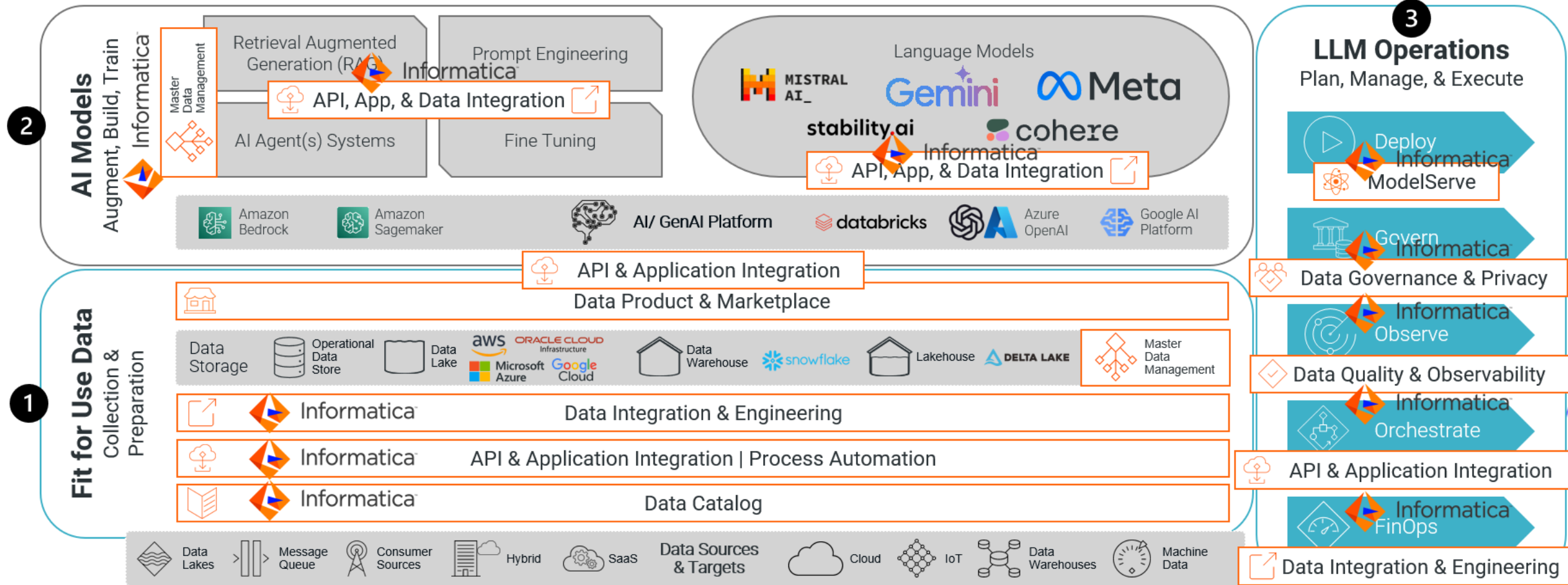
Навыки и умения

Интеграция и внедрение

Деньги и ценность

AI Референсная Архитектура

Попробуй собрать сам 😊



Управление данными с поддержкой AI для ваших AI инициатив



Всё готово для AI

Ваши данные готовы к AI?

- **CLAIRE**® расширяет возможности управления данными, помогая клиентам достигать результатов за пределами человеческих возможностей.
- Informatica **Intelligent Data Management Cloud**™ предоставляет **управляемые, доверенные и демократизированные данные**, необходимые клиентам для успешного внедрения **собственных решений на основе AI**



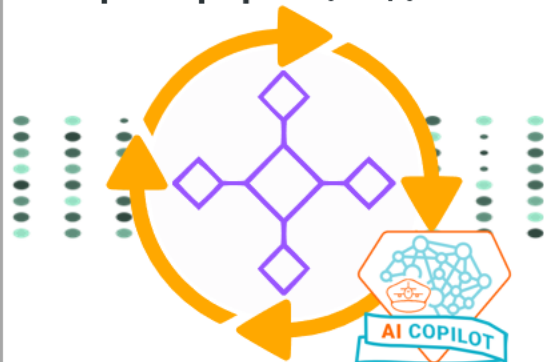
INFORMATICA. Единая облачная платформа сквозного управления данными на основе AI



Informatica

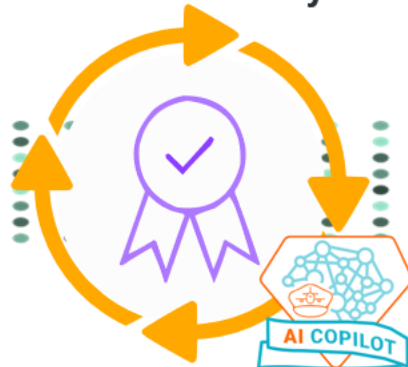
Intelligent Data Management Cloud

Интеграция & Трансформация данных



- Интеллектуальный анализ структур данных
- Рекомендации по Трансформации

Качество данных & Observability



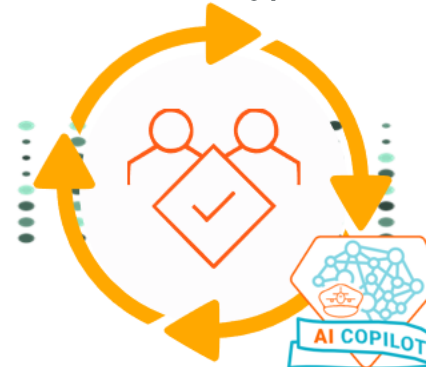
- Создание правил Качества данных
- Рекомендации и Инсайты

НСИ & Приложения 360



- Сопоставление сущностей
- Выводы по связям сущностей

Data Governance & Каталог данных



- Выводы по связям данных
- Определение связей в бизнес-гlossарии

CLAIRE[®]

AI-Powered Metadata Intelligence & Automation

Connectivity

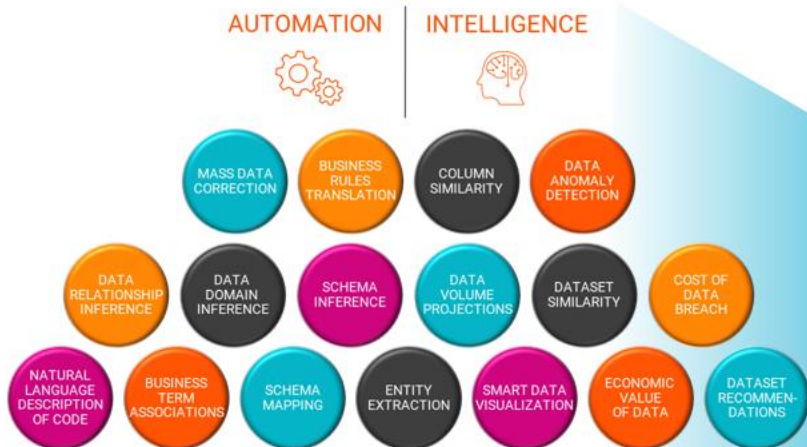
Metadata System of Record

CLAIRE® Управление данными на основе ИИ



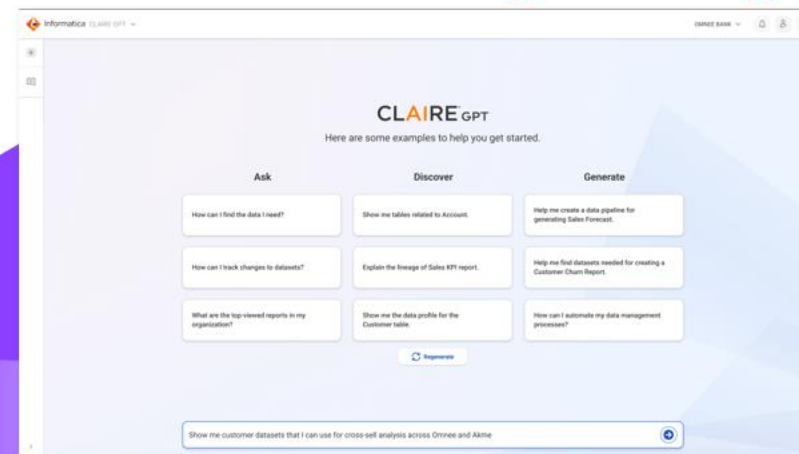
CLAIRE Copilot

АВТОМАТИЗАЦИЯ МЕТАДАННЫХ НА ОСНОВЕ ИИ



CLAIRE GPT

ИНТЕРФЕЙС ЕСТЕСТВЕННОГО ЯЗЫКА ДЛЯ РАБОТЫ С ДАННЫМИ



Автоматизация управления данными

Активные метаданные для обеспечения автоматизированных возможностей управления данными		
Сокращение ручной работы и повышение операционной эффективности в задачах управления данными		
Проактивные рекомендации и оповещения о потенциальных проблемах		
ИНТЕЛЛЕКТУАЛЬНЫЙ ГЛОССАРИЙ	УМНЫЕ ДОМЕНЫ	СХОДСТВО СТОЛБЦОВ
СОПОСТАВЛЕНИЕ СУЩНОСТЕЙ	СОПОСТАВЛЕНИЕ СХЕМ	ИНСАЙТЫ ПО КАЧЕСТВУ ДАННЫХ
РЕКОМЕНДАЦИИ ПО ОБЪЕДИНЕНИЮ	РЕКОМЕНДАЦИИ ПО ТРАНСФОРМАЦИИ ДАННЫХ	РЕКОМЕНДАЦИИ ПО НАБОРАМ ДАННЫХ

Управление данными на основе генеративного ИИ

Повышение эффективности за счет автоматизации работы специалистов с данными	Расширение доступа к данным: понятный и простой доступ к данным для рядовых пользователей	Снижение затрат на управление данными	
NL TO ETL/ELT	NL TO DQ	NL TO PREP	NL TO GOVERNANCE
NL-ОБНАРУЖЕНИЕ	NL-ИССЛЕДОВАНИЕ	NL-ТЕСТИРОВАНИЕ И ДОКУМЕНТИРОВАНИЕ	АВТОМАТИЗИРОВАННЫЙ FINOPS
КРОСС-ПРОДУКТОВЫЙ ОПЫТ	CO-PILOT	КОНТЕКСТНЫЙ АНАЛИЗ ДАННЫХ	OPEN, POLYGLOT

CLAIRE® GPT ДЛЯ КАЖДОГО ДАННЫЕ, ГОВОРЯЩИЕ НА ВАШЕМ ЯЗЫКЕ



Бизнес-пользователи

Кто наши самые ценные клиенты с точки зрения покупок за последний год?

Какие продукты или услуги клиенты часто покупают вместе?

Каковы тенденции в показателе удержания клиентов за последние 4 квартала этого года?

Аналитик

Какие СИ используют для отчета по показателям удержания клиентов?

Объясни происхождение данных для показателя удержания клиентов

Есть ли какие-либо аномалии или отклонения в данных удержания клиентов, которые необходимо исследовать?

Data-Инженер

Где я могу найти наборы данных, необходимые для расчета стоимости привлечения клиента (CAC)?

Все ли наборы данных, необходимые для расчета CAC, подключены и доступны?

Создай пайплайн, который рассчитывает CAC для различных маркетинговых каналов

DT Overview of @Sales KPIs

Sure, providing an overview of Sales KPIs

Sales KPIs Overview

The Sales KPIs is a Dashboard type asset, housed in a Tableau environment. Unfortunately, it has not undergone profiling to assess its quality and structure comprehensively.

Key Components Overview

Stakeholdership

- No specific stakeholdership information is available: It's unclear who owns or uses this asset, which could impact its management and strategic planning.

Data Characteristics

- Resource Type: Tableau
- Profiled: No
 - This indicates that the asset has not undergone a comprehensive assessment of its structure and quality.

Data Quality

- No data quality metrics are available: This suggests that the asset has not undergone a data quality process, which could affect its reliability and usefulness.

JD show me the lineage of @DIM_CUSTOMER

Absolutely, showing the lineage of DIM_CUSTOMER

- The lineage of DIM_CUSTOMER involves a total of 5 catalog sources.
- The data flow begins from two sources: iics_retail_dataFlow and snowflake_retail_dwh, which collectively include 4 unique assets such as mt_m_onlineSales_02_s3_landing_to_enrichment and DIM_CURRENCY.
- This data then flows into DIM_CUSTOMER and proceeds towards two target resources: Tableau_sales_kpis_Tableau and snowflake_retail_dwh, covering another 8 assets including Product List and V_TEMP_FACTSTORESALES_DETAILED_EUROPE.

Lineage of DIM_CUSTOMER Table

ST Identify the staff members who managed stores located in 'London' or 'Santa Cruz'

Sure, identifying the staff members who managed stores located in 'London' or 'Santa Cruz' using STAFFS

TABLE | STAFFS

SnowflakeScan_Copy_New_Sales / PLATFORMQA / SALES Columns: 2

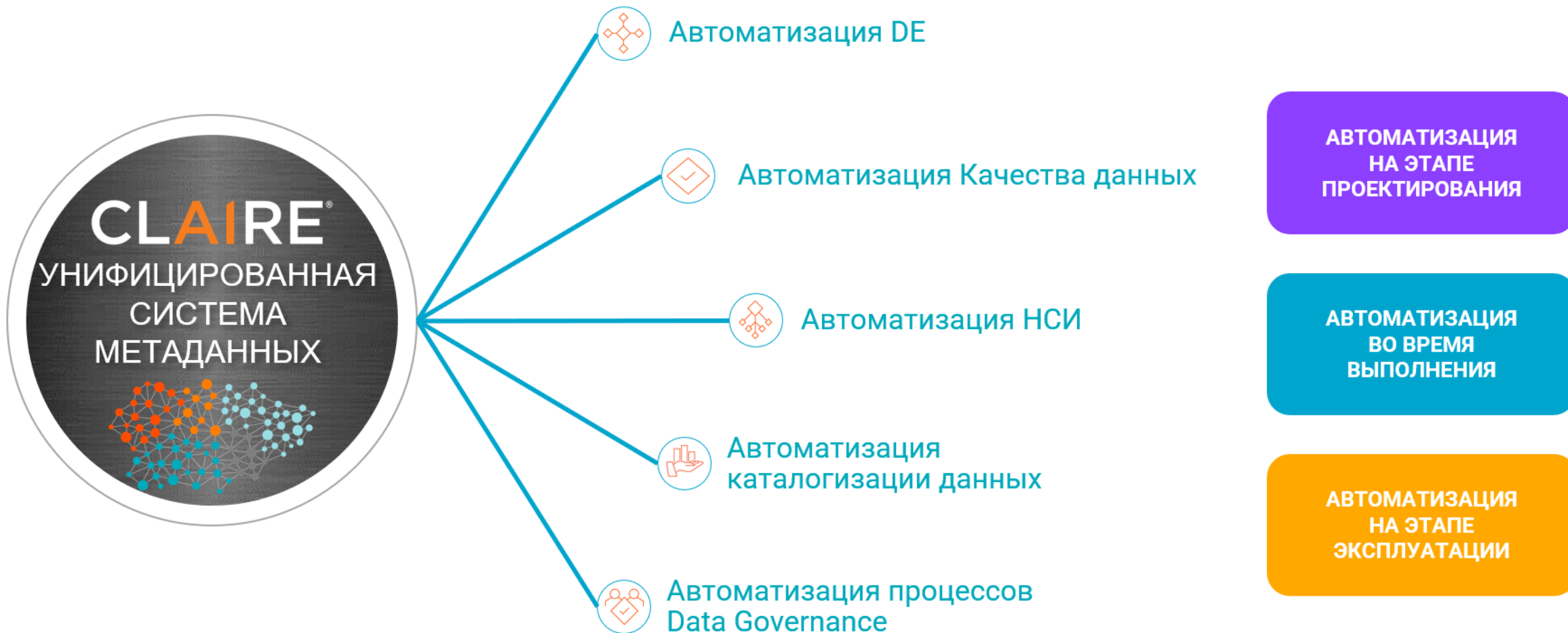
Privacy guidelines prevent us from storing sample data, but you can download the data for your reference.

Бизнес-ценность CLAIRe GPT

Новый разговорный интерфейс для управления данными для всех потребителей данных

Обнаружение данных	Исследование данных	Трансформация данных	Помощь по продукту
<ul style="list-style-type: none">▪ Демократизация доступа к данным▪ Разговорный интерфейс для поиска данных▪ Обеспечение доступности данных▪ Повышение скорости принятия решений▪ Повышение грамотности в области данных в организации	<ul style="list-style-type: none">▪ Единое место для понимания данных, готовых к использованию▪ Запрос данных на естественном языке▪ Мгновенный ответ на вопросы о данных▪ Просмотр данных▪ Понимание профиля данных▪ Исследование данных▪ Постановка вопросов о данных	<ul style="list-style-type: none">▪ Преобразование данных между источником и потребителем▪ Быстрый старт для создания конвейеров данных▪ Использование естественного языка как языка преобразования▪ Создание черновиков ELT-конвейеров данных▪ Редактирование и модификация в интеграции данных	<ul style="list-style-type: none">▪ Повышение грамотности использования продуктов INFA для управления данными▪ Ответы на вопросы документации▪ Помощь по продуктам▪ Рекомендации по использованию и лучшим практикам

Ключевые возможности CLAIRES[®] AI Copilot



Расширенное, E2E управление данными

Преимущества управления данными с поддержкой **CLAIRE** на этапах проектирования, выполнения и эксплуатации



CLAIRE®

Преимущества работы с ИИ

Этап проектирования

Повышенная эффективность разработчиков
*(Прирост производительности на 28 процентов)

Ускоренная интеграция данных и экономия времени
*(90% экономии усилий на ETL)

Этап выполнения

Максимизированное соотношение цена-производительность
(Итеративная оптимизация на основе приоритетов задач и SLA)

Оптимизированное использование ресурсов *(65% экономии затрат благодаря автонастройке, умному управлению кластерами, оптимальной параллельности)

Этап эксплуатации

Улучшенная производительность *(80% ускорения благодаря автоматическому выделению ресурсов)

Увеличенная экономия средств *(В среднем \$1.54 млн ежегодной экономии за счет снижения зависимости от персонала)

CLAIRE® AI Copilot: Истории успеха клиентов



Повышение эффективности, ясности и понимания



Автоматизация каталогизации данных

Классификация элементов данных, классификация сущностей, сопоставление схем

Автоматизированная классификация данных может сократить время и усилия на ручную классификацию данных до 50%, что приводит к значительной экономии времени и средств.

Anthem экономит 1 млн долларов ежемесячно, используя возможности автоматизированной организации и обнаружения с помощью Data Catalog от Informatica.



Автоматизация управления данными

Интеллектуальные ассоциации глоссария, автоматическое применение правил качества данных, панель активности CLAIRE®

Автоматическое добавление бизнес-контекста и оценок качества данных позволяет достичь стократного ускорения в процессах их обнаружения и повышения качества по сравнению с ручными методами

Ведущая компания в сфере доставки и логистики достигла впечатляющей 95% точности за считанные минуты, автоматизировав связывание бизнес-терминов с 970 тыс. из 1,04 млн записей, что в противном случае потребовало бы многих месяцев ручной работы.



Автоматизация качества данных

Аналитика качества данных, преобразование естественного языка в правила качества данных

Автоматизированные процессы контроля качества данных значительно экономят время, так как проблемы с качеством данных выявляются путем глубокого анализа статистики профилирования.



Автоматизация инжиниринга данных

Рекомендации по следующей трансформации, интеллектуальное обнаружение структуры, рекомендации по UDF/Mapplet

Автоматизированное завершение маппингов и обнаружение структуры сложных файлов может значительно ускорить разработку конвейеров данных и повысить продуктивность инженеров данных.

90% клиентов Informatica Cloud Data Integration ежедневно используют рекомендации по следующей трансформации при разработке конвейеров данных, что приводит к повышению производительности более чем на 20%.



Автоматизация управления НСИ

Сопоставление идентичности, классификация продуктов

Точность автоматизированной классификации и сопоставления улучшается с увеличением количества примеров, уменьшая потребность в ручном вмешательстве, минимизируя неоднозначность и устраняя несоответствия.

JDRF оптимизировала записи своих сторонников с 10 миллионов до 3 миллионов, включая 1 миллион неактивных сторонников, мгновенно увеличив свою потенциальную базу постоянных доноров на 50%.

ВЫВОДЫ

- Мы еще в начале пути внедрения генеративного ИИ, есть возможность сделать правильные инвестиции и проекты
- Начните с пилотных проектов применения ИИ в управлении данными на основе облачных платформ
- Ценность качественных данных возросла еще больше. Продолжайте заниматься DG и управлением данными
- ИИ будет драйвером для оцифровки организации
- ИИ оптимизирует процессы, штат и создает ценность для организации



Спасибо за внимание

УЗНАЙТЕ БОЛЬШЕ НА НОВОМ
САЙТЕ

WWW.DIS-GROUP.WORLD