Хранилища данных: современные подходы и архитектуры

Анонимный опрос о текущих архитектурах DWH, инструментах и планах

Роль в команде *
Data Engineer
Analytics Engineer
Oata Architect
○ BI/DA Analyst
ML Engineer
Руководитель
Студент/стажёр
Другое:
Опыт работы c DWH *
O 1-3
○ 3-5
○ 5-10
O 10+

Размер компании *
O 1-50
O 51-200
O 201-1000
O 1001-5000
○ 5000÷
Не трудоустроен / Учебный проект
Отрасль*
○ IT/Telecom
○ E-commerce
○ Finance
Manufacturing
○ Healthcare
O Public/Government
○ Education

Текущая архитектура хранилища *							
○ Он-прем DWH							
Облачный DWH							
O Data Lake + DWH (Lakehouse)							
○ Гибрид (On-prem + Cloud)							
○ Нет единого DWH							
Учебный проект (нет DWH)							
Основная платформа/СУБД *							
○ Snowflake							
BigQuery							
○ Redshift							
○ ClickHouse							
○ Vertica							
○ Greenplum							
○ MS SQL Server							
○ Oracle							
○ PostgreSQL							
O Databricks/Delta Lake							
○ Локальная БД (PostgreSQL/MySQL)							
СlickHouse (локально)							
Другое:							

Подход интеграции *
○ ETL
○ ELT
<u>Смешанный</u>
Нет ответа
Латентность обновления данных *
O Раз в неделю+
Раз в день
ОПочасово
○ Каждые 15-60 мин
○ Почти real-time (<15 мин)
○ Real-time (стриминг)
○ Нет ответа

Инструменты интеграции/оркестрации *
Apache Airflow
dbt
Fivetran/Stitch
Spark/Dataproc
Apache NiFi
☐ Talend
Pentaho
☐ Prefect
Luigi
Скрипты (bash/python)
Пока не использую
Другое:
Практики Data Governance *
Data Catalog
☐ Data Lineage
Мониторинг качества данных (DQ)
Ролевые доступы/АВАС
Нет формализованных практик
Не применимо (учебный проект)

Оцените текущий DWH по шкале 1–5 *									
	1	2	3	4	5	Нет ответа			
Производит	\circ	0	0	\circ	\circ	0			
Масштабир	0	0	0	\circ	\circ	0			
Стоимость	0	\circ	0	\circ	\circ	\circ			
Простота с	0	0	0	\circ	\circ	0			
Планы на 12 месяцев * Миграция в облако Переход к Lakehouse Внедрение ELT/dbt Внедрение Data Catalog/Lineage Оптимизация стоимости Увеличение покрытия real-time Ничего из перечисленного									