

Feedback Labo 4

Data Lineage en Metadata Management

Data Foundations

Metadata Management

- **Hardgecodeerde metadata:** Hoewel het geen groot probleem is om een kleine set metadata te *hardcoden*, wordt steeds aangeraden om grotere sets metadata dynamisch te genereren.
 - **Aanbeveling:** voeg een functie toe om metadata automatisch te genereren op basis van datasetkolommen.
- **Beperkte zoekfunctionaliteit:** filteropties beperken zich tot eenvoudige zoekopdrachten.
 - **Aanbeveling:** breid zoekfunctionaliteit uit met ondersteuning voor samengestelde voorwaarden (zoals AND/OR-logica).

- **Geen validatie van metadata:** metadata werd vaak opgeslagen zonder controle op overeenstemming met de datasetkolommen.
 - **Aanbeveling:** controleer of alle metadata-items corresponderen met datasetkolommen.

Data Lineage Transformaties

- **Geen validatie bij transformaties:**
 - **Probleem:** bij het toevoegen van `total_sales` werd vaak niet gecontroleerd op missende waarden of verkeerde datatypes.
 - **Aanbeveling:** voeg validatiestappen toe voor invoergegevens voordat transformaties worden uitgevoerd.
- **Beperkte logging:**
 - **Probleem:** het logboek moet voldoende details bevatten, op zijn minst de gewijzigde rijen.
 - **Aanbeveling:** integreer gedetailleerdere logging, inclusief specifieke kolomwijzigingen.

Data Lineage Visualisatie

- **Geen exportopties:**
 - **Probleem:** lineage-grafieken worden niet opgeslagen.
 - **Aanbeveling:** voeg opties toe om grafieken als afbeelding op te slaan. Zo kan je ze makkelijk mee indienen op Digitap.
- **Basisdata lineage:**
 - **Probleem:** de lineage-grafiek toont vaak geen annotaties over betrokken rijen of kolommen.
 - **Aanbeveling:** voeg labels en context toe over de impact van elke transformatie.

Aanbevolen verbeteringen

- Integreer foutafhandelingsmechanismen bij datasetverwerking en lineage-generatie.
- Zorg voor consistentie tussen metadata, logging, en visualisaties.

Conclusie

- De analyses tonen dat jullie de basisconcepten van data lineage begrijpen, maar dat er meer focus moet zijn op robuustheid.
- Implementatie van aanbevelingen zal leiden tot betere resultaten en een meer professionele workflow.