

## LLISTA DE CANVIS ENTRE INFORMES

### INFORME INICIAL – INFORME DE SEGUIMENT 1

En la taula 1, es mostra un petit resum de l'evolució dels objectius entre aquests 2 informes on es pot destacar la modificació del segon objectiu, a banda d'haver completat el primer dels reptes.

Objectiu	Informe Inicial	Informe 1
<b>1. Llinatge (camins de dependències)</b>	Identificar els camins de dependència de les taules "mart".	<b>Objectiu completat.</b> Ja s'han obtingut els camins i emmagatzemat en un "dataframe" de Pandas preparat per visualització.
<b>2. Optimització de fluxos</b>	Optimitzar els fluxos de càrrega i garantir dades actualitzades a temps.	<b>Modificació:</b> Ara es concreta que es vol crear un algorisme que defineixi grups de taules a "intermediate" que es puguin executar abans i en paral·lel. També es vol identificar les taules "staging" que retarden la cadena.
<b>3. Dashboard amb PowerBI</b>	Visualitzar resultats obtinguts.	<b>Es manté igual.</b> Es confirma la seva utilitat per plasmar el llinatge i la proposta d'optimització.

Taula 1. Evolució objectius 1.

Seguidament, en quan a progrés o desenvolupament s'ha avançat en la realització de varies tasques:

- **Accés al repositori de models SQL via GitHub API** amb "requests" i autenticació amb token.
- **Extracció de dependències amb expressions regulars (re)** en les clàusules FROM.
- Generació del **"dataframe" estructurat amb informació de dependències** ("source", "destination", nivell).
- Inici del **anàlisi de complexitat del llinatge**, mostrant:
  - Les taules amb més dependències.
  - La distribució de nombre de dependències per taula.
- **Creació d'un prototip** per a testejar l'algorisme a més petita escala i afavorir la detecció d'errors

## INFORME DE SEGUIMENT 1 – INFORME DE SEGUIMENT 2

Com en el primer informe, en el segon també hi ha hagut una evolució amb els objectius plantejats i hi ha hagut en aquest cas una modificació en el darrer objectiu. En la taula 2 podem veure els canvis.

Objectiu	Informe 1	Informe 2
<b>1. Llinatge (camins de dependències)</b>	Identificar els camins de dependència de les taules "mart".	<b>Objectiu completat.</b> Ja s'han obtingut els camins i emmagatzemat en un "dataframe" de Pandas preparat per visualització.
<b>2. Optimització de fluxos</b>	Optimitzar els fluxos de càrrega i garantir dades actualitzades a temps.	<b>Objectiu Completat.</b> Ara es concreta que es vol crear un algorisme que defineixi grups de taules a "intermediate" que es puguin executar abans i en paral·lel. També es vol identificar les taules "staging" que retarden la cadena.
<b>3. Aplicació Streamlit</b>	Visualitzar resultats obtinguts.	<b>Modificació.</b> Finalment en comptes de fer un "dashboard" amb PowerBI es realitza una aplicació amb Streamlit.

Taula 2. Evolució objectius 2.

Seguidament, en quan a progrés o desenvolupament s'ha avançat en la realització de varies tasques:

- **Obtenció de les dades reals de temps**, accedint al repositori de Github de l'empresa i a la taula que emmagatzema la informació de Prefect.
- **Realització de l'algorisme de optimització.**
- **Disseny i creació de l'aplicació amb Streamlit**, és a dir codificació a través de Python per a poder generar un entorn visual on veure els resultats i fer ús del simulador.

## INFORME DE SEGUIMENT 2 – MEMÒRIA FINAL

De l'últim informe de seguiment a la memòria final simplement s'ha condensat tota la documentació generada en els tres informes previs i s'ha redactat en format d'article.

També s'ha afegit algunes seccions com el resum o les paraules clau i s'ha referenciat en el text les fonts usades durant la realització del projecte.