

# DataHub

## ▼ CoreDB

- Descrizione: Non è un modulo operativo, ma rappresenta il database centrale che contiene tutte le tabelle strutturali importate (es. emesso, prodotti, diritti). È il cuore del sistema, popolato da DataIngestion e utilizzato da altri moduli come AutoLink per le integrazioni temporanee.

## ▼ QueryFlex

- Descrizione: Modulo non ancora attivo, ma proposto per eseguire query analitiche sui dati di CoreDB. Permette di estrarre informazioni specifiche per analisi o report, sfruttando i dati strutturali e (potenzialmente) quelli temporanei generati da AutoLink. Esempio di utilizzo: Testare query come "Film trasmessi su Canale 5 nel 2023 con audience > 1M".

## ▼ AutoLink

- Descrizione: Macroblocco dedicato all'integrazione temporanea di file extra (provenienti da Mediaset o fonti esterne) con i dati già presenti in CoreDB. Non modifica le tabelle strutturali, ma crea tabelle temporanee per combinare i dati extra con quelli di CoreDB (es. join con emesso per analisi a breve termine). È utilizzato per necessità limitate nel tempo, come report o analisi temporanee, e non aggiunge dati permanentemente a CoreDB.

## ▼ DataIngestion

- Descrizione: Macroblocco responsabile del caricamento dei dati strutturali in CoreDB. Gestisce gli script di caricamento forniti dalla Direzione Sistemi (settimanali per tabelle come emesso, giornalieri per altre tabelle) per popolare le tabelle principali del database, come emesso, prodotti, e diritti. Questo modulo si occupa di tutti i dati permanenti e strutturali, garantendo che CoreDB sia sempre aggiornato con i dati di base. Se c'è bisogno di aggiungere dati permanentemente a CoreDB, questo viene gestito integrando i nuovi dati in DataIngestion.

## ▼ DataQualityGate

- Descrizione: Componente integrato in DataIngestion che analizza e normalizza i dati prima dell'inserimento in CoreDB. Esegue la validazione dei file CSV (verifica formati, strutture, valori anomali), standardizza i dati in ingresso (uniforma nomi, formatti valori, converte tipi) e genera report di qualità dopo ogni importazione. Non modifica direttamente il database, ma agisce come "filtro" preventivo garantendo che solo dati conformi e normalizzati raggiungano le tabelle strutturali. Opera automaticamente ad ogni ciclo di aggiornamento delle tabelle.