

**Sistemi Informativi**  
**Progetto di Laboratorio a.a. 2024/2025**

**Salvatore Rudisi**  
**944705**

<b>Esercizio 1. Progettazione di una DW .....</b>	<b>2</b>
<b>1.1 Modello di Dominio.....</b>	<b>3</b>
<b>Schema E-R .....</b>	<b>4</b>
<b>Schema Relazionale.....</b>	<b>5</b>
<b>1.2 Modello BPMN.....</b>	<b>6</b>
<b>1.3 Data Warehouse in DFM .....</b>	<b>7</b>
<b>1.4 Star Schema .....</b>	<b>8</b>
<b>Esercizio 3. Data Mining.....</b>	<b>9</b>

## Esercizio 1. Progettazione di una DW

Un'etichetta discografica vuole tenere traccia di tutti i contratti stipulati con le band, i dischi che produce e delle vendite di questi dischi. Al momento, l'etichetta conserva solo i dati relativi ai contratti in corso e non conserva alcuna informazione storica. Si decide quindi di costruire un data warehouse per raccogliere e archiviare le informazioni storiche. Con il data warehouse l'azienda vuole analizzare le proprie vendite.

I dirigenti dell'azienda vi raccontano che quando si decide di finanziare un nuovo disco di una band viene stipulato un contratto. Ogni disco ha il suo contratto, ma lo stesso contratto può applicarsi a più dischi. Per redigere il contratto, la band viene iscritta presso l'etichetta discografica: ogni band ha un nome, una categoria musicale principale e uno o più artisti. Eventualmente viene anche specificato il manager del gruppo, ma non è obbligatorio averlo. In questo caso, le band che non hanno un manager indicano uno degli artisti come rappresentante principale.

Ogni artista è specialista di uno strumento principale, che può essere anche la sua voce.

È importante notare che, nel corso del tempo, le band possono dividersi, attrarre nuovi membri, o sostituire i membri del gruppo. Questi cambiamenti saranno eventualmente catturati prima della sottoscrizione di un nuovo contratto.

Le band registrano dischi fissando una o più sessioni di registrazione. I dischi sono poi prodotti e distribuiti dall'etichetta discografica. I dischi, compresi quelli non ancora terminati, hanno un titolo e sono identificati da uno speciale codice internazionale.

Per ogni disco vengono mantenuti il costo di produzione e la data di uscita.

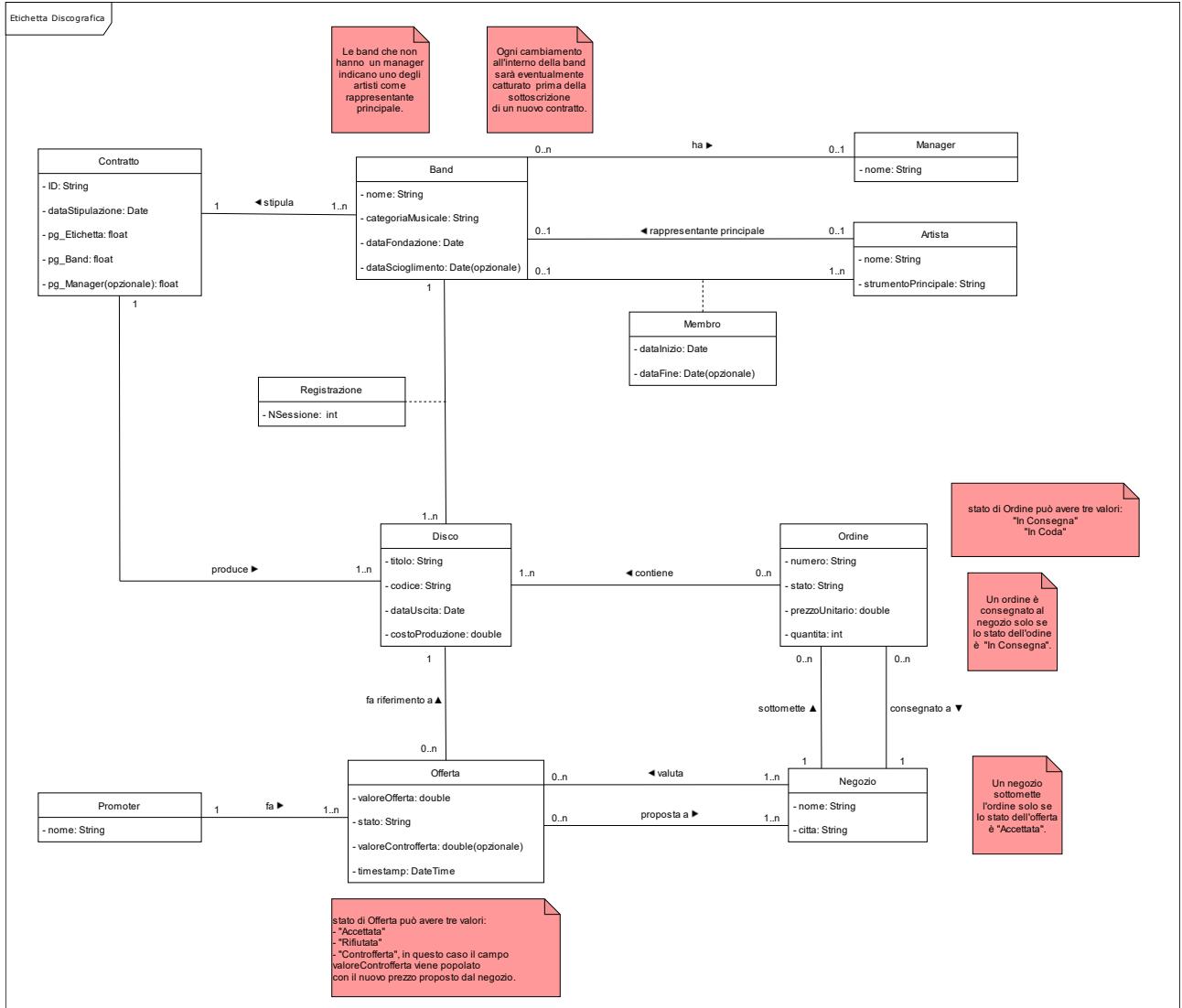
Il contratto stipulato tra l'etichetta e la band stabilisce anche quale percentuale di guadagno deve essere assegnata alla band, al manager (se presente) e all'etichetta.

Una volta prodotto, il disco viene messo in vendita.

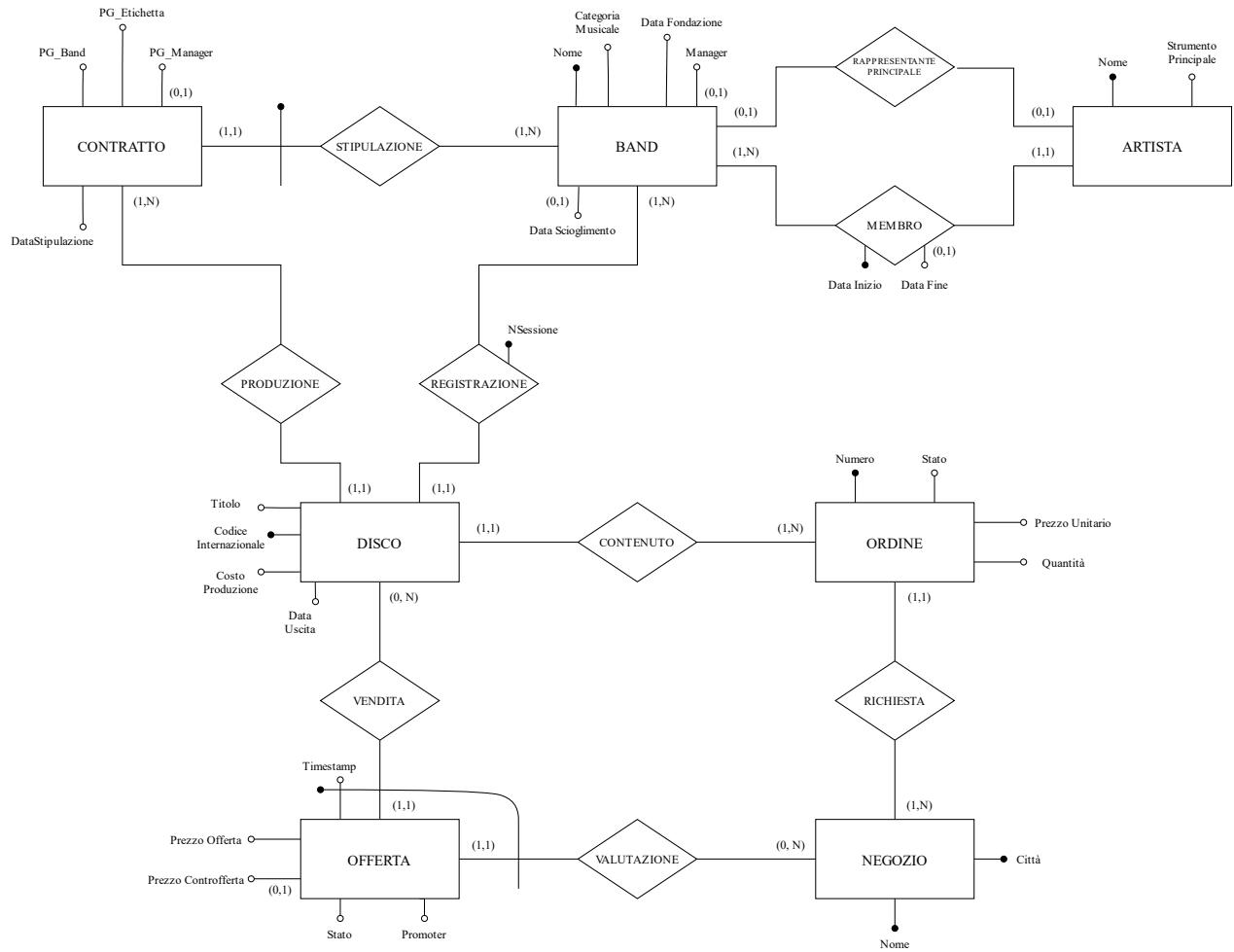
Negozi specializzati vengono contattati da un promoter dell'etichetta discografica che offre il nuovo disco per la vendita, il negozio valuta l'offerta che può accettare, rifiutare o può fare una controfferta. Una volta stabilito il prezzo di unitario di acquisto del disco da parte del negozio, il negozio può sottomettere uno o più ordini per un preciso numero di pezzi. L'etichetta può soddisfare gli ordini consegnando il materiale richiesto o, in caso di carenza, l'ordine viene messo in coda. Ovviamente, lo stesso negozio può emettere più ordini per uno stesso disco e ogni volta il prezzo unitario può essere diverso.

## 1.1 Modello di Dominio

Progettare un modello di dominio in UML che sia in grado di catturare tutti i concetti (e le relazioni tra questi) rilevanti introdotti nella descrizione.



## Schema E-R



## BUSINESS RULES:

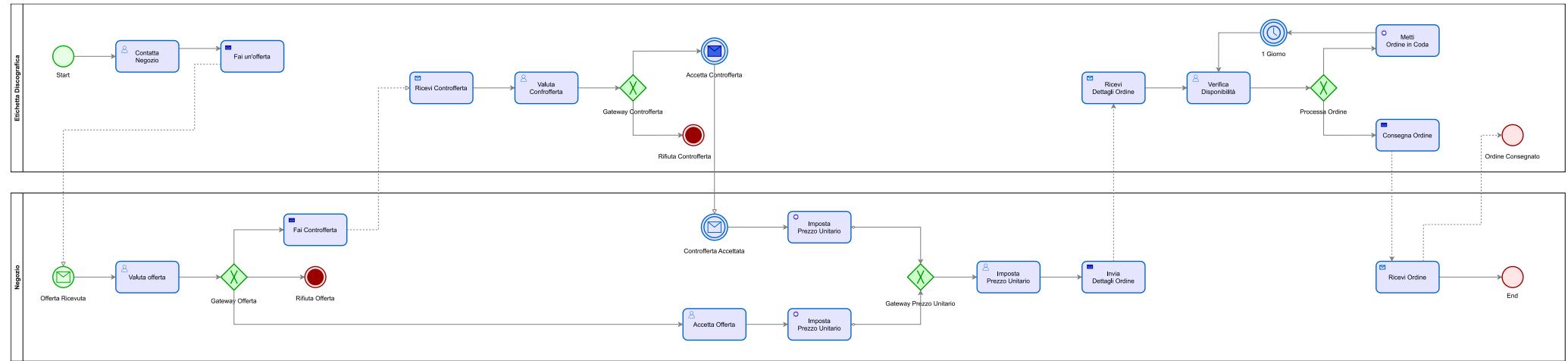
- Ogni cambiamento all'interno della band sarà eventualmente catturato prima della sottoscrizione di un nuovo contratto.
- Le *Band* che non hanno un *Manager* indicano un *Artista* come *Rappresentante Principale*.
- *Stato* di *Offerta* può avere tre valori:
  - "Accettata"
  - "Rifiutata"
  - "Controfferta", in questo caso il campo *Prezzo Controfferta* viene popolato con il nuovo prezzo proposto dal negozio
- *Stato* di *Ordine* può avere due valori:
  - "Consegnato"
  - "In Coda"

## Schema Relazionale

- Band (Nome, CategoriaMusicale, DataFondazione, DataScioglimento\*, Manager\*)
- Artista (Nome, StrumentoPrincipale)
- Contratto (Band, DataStipulazione, PG\_Band, PG\_Etichetta, PG\_Manager\*)
  - Contratto (Band) referenzia la chiave di Band
- Disco (CodiceInternazionale, Titolo, CostoProduzione, DataUscita)
- Ordine (Numero, Stato, PrezzoUnitario, Quantità)
- Negozio (Nome, Città)
- Offerta (NomeNegozio, CittàNegozio Timestamp, Stato, Promoter, PrezzoOfferta, PrezzoControfferta\*)
  - Offerta (NomeNegozio) referenzia la chiave Nome di Negozio
  - Offerta (CittàNegozio) referenzia la chiave Città di Negozio
- Membro (Band, Artista, DataInizio, DataFine\*)
  - Membro (Band) referenzia la chiave di Band
  - Membro (Artista) referenzia la chiave di Artista
- Registrazione (Band, Disco, NSessione)
  - Registrazione (Band) referenzia la chiave di Band
  - Registrazione (Disco) referenzia la chiave di Disco

## 1.2 Modello BPMN

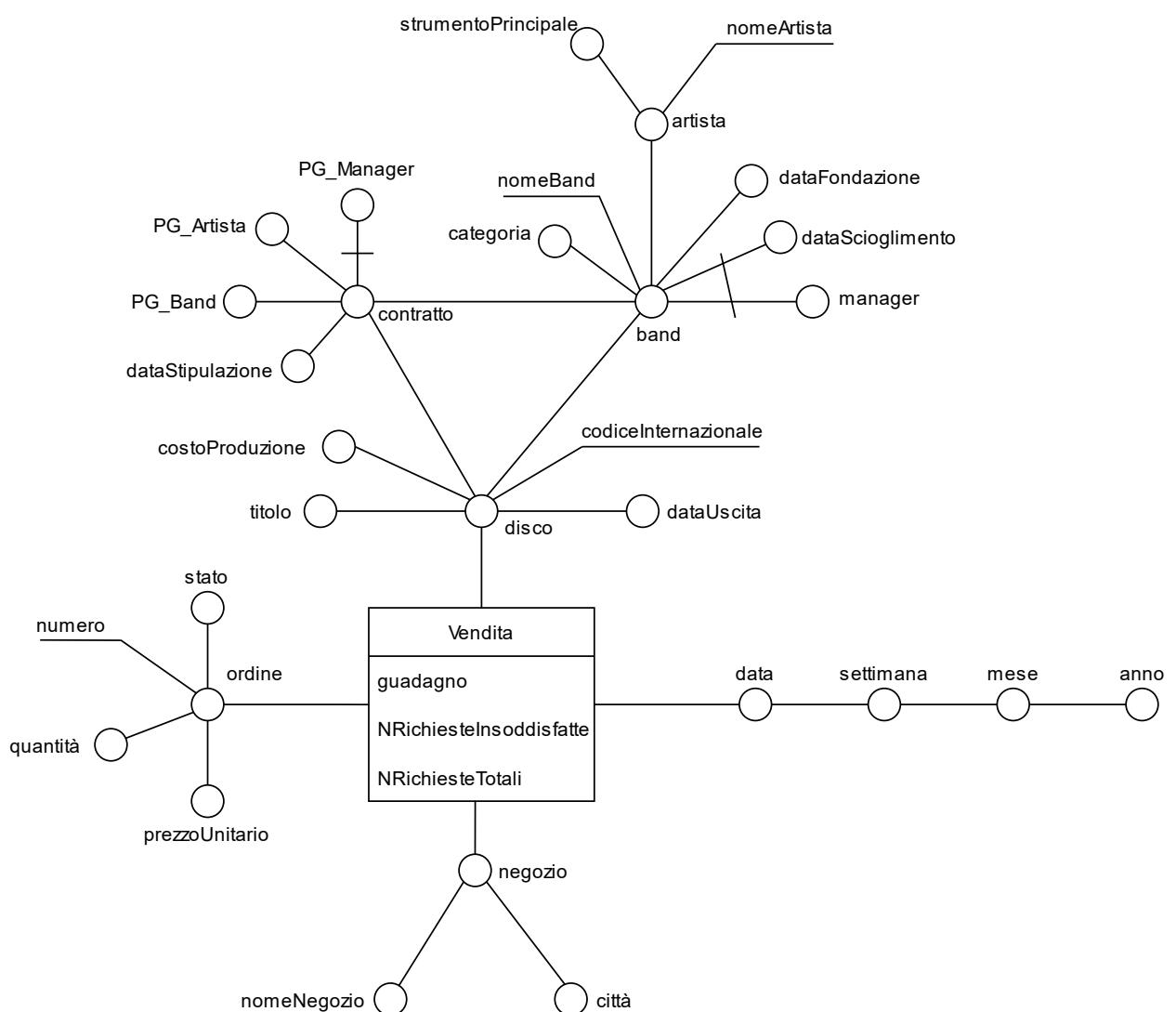
Progettare in BPMN il processo di acquisto di un disco da parte di un negozio modellando sia le attività dell'etichetta sia quelle del negozio.



### 1.3 Data Warehouse in DFM

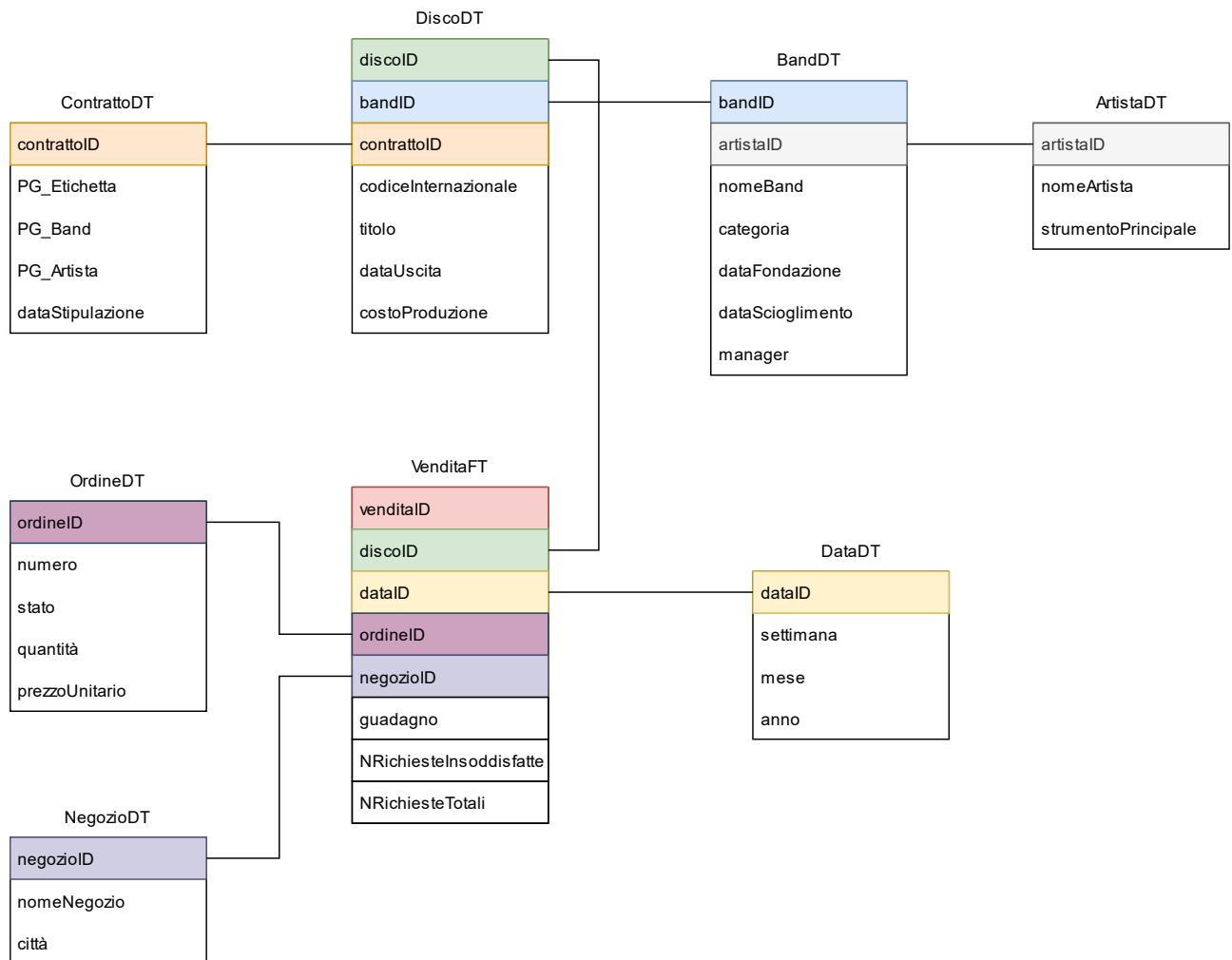
Progettare il DFM di una DW che risponda (almeno) alle seguenti query analitiche:

- Qual è il disco/gruppo/artista che ha dato all'etichetta discografica il più alto/basso guadagno/rendimento sull'investimento per settimana/mese/anno.
- Quale negozio/città/paese ha il maggior numero di richieste che non è stato possibile soddisfare?
- Quale disco/gruppo/artista ha il maggior numero di richieste?



## 1.4 Star Schema

Definire lo SS della DW progettata al passo precedente.



### Esercizio 3. Data Mining

Usando Knime, applicare un algoritmo di data mining a scelta ai dati del database Northwind, spiegando perché lo avete scelto (cosa vi permette di scoprire).

Ho scelto K-Means per il clustering dei clienti in quanto è un algoritmo efficace per la segmentazione basata su variabili numeriche.

Questo metodo consente di individuare gruppi di clienti con caratteristiche simili.

In questo studio, ho applicato l'algoritmo di clustering K-Means ai clienti del database Northwind, suddividendoli in gruppi basati su:

- Numero totale di ordini effettuati (**TotalOrders**)
- Totale speso (**SUM\_Gain**)
- Valore medio per ordine (**AVG\_Gain**)

Questa segmentazione permette di identificare diversi profili di clienti e ottimizzare le strategie di marketing e vendita.

Dopo l'analisi, i clienti sono stati suddivisi in gruppi distinti, che potrebbero essere interpretati come:

- **Cluster\_0**: Clienti occasionali con pochi ordini e valore medio basso.
- **Cluster\_1**: Clienti abituali con spese moderate e acquisti regolari.
- **Cluster\_2**: Grandi acquirenti con ordini frequenti e valore medio alto.