



UNIVERSITÀ DI PISA

Progetto di Basi di Dati
A. A. 2022 - 2023
Corso di Laurea in Ingegneria Informatica

Paolo Walsh, Pietro Balestri

Indice

1	Introduzione	7
2	Progettazione Concettuale	9
2.1	Area Contenuti	9
2.2	Area Formati	12
2.3	Area Clienti	14
2.4	Area Streaming	17
2.5	Chiarimenti su Entità e Relationship	18
3	Ristrutturazione del diagramma E-R	23
3.1	Eliminazione Attributi Composti	23
3.2	Eliminazione Generalizzazioni	23
4	Tavola dei Volumi	25
4.1	Area Contenuti	25
4.2	Area Formati	27
4.3	Area Clienti	28
4.4	Area Streaming	31
4.5	Chiarimenti Volume Film	31
5	Individuazione e Analisi delle Operazioni	33
5.1	Calcolo del Rating di un Film	33
5.1.1	Tavola degli Accessi in Assenza di Ridondanze	35
5.1.2	Tavola degli Accessi in Presenza di Ridondanza <i>Rating</i>	35
5.2	Raccomandazione dei Contenuti	37
5.2.1	Tavola degli Accessi in Assenza di Ridondanze	38
5.2.2	Tavola degli Accessi in Presenza di Ridondanza Raccoman- dazioni	40
5.3	Connessione Utente-Server	43
5.3.1	Tavola degli Accessi in Assenza di Ridondanze	45
5.3.2	Tavola degli Accessi in Presenza di Ridondanza <i>Banda Dispo-</i> <i>nibile</i>	48
5.4	Acquisto Piano Abbonamento	53
5.4.1	Tavola degli Accessi	54

5.5	Classifiche	55
5.5.1	Tavola degli Accessi in Assenza di Ridondanze	56
5.5.2	Tavola degli Accessi in Presenza di Ridondanza Classifiche	58
5.6	Inserimento Film, Anteprime e Esclusive	61
5.6.1	Tavola degli Accessi	62
5.7	Caching	64
5.7.1	Tavola degli Accessi	65
5.8	Bilanciamento Del Carico	68
5.8.1	Tavola degli Accessi della Fase 1	69
5.8.2	Tavola degli Accessi della Fase 2	70
5.9	Custom Analytics - Analisi Visualizzazioni	71
5.9.1	Chiarimenti sulla Ritenzione	72
5.9.2	Tavola degli Accessi Fase 1	72
5.9.3	Tavola degli Accessi Fase 2	73
6	Progettazione Logica	75
6.1	Traduzione dello Schema Concettuale nel Modello Logico Relazionale	75
6.1.1	Tabelle dell'Area Contenuti	75
6.1.2	Tabelle dell'Area Formati	76
6.1.3	Tabelle dell'Area Clienti	76
6.1.4	Tabelle dell'Area Streaming	76
6.2	Vincoli	77
6.2.1	Vincoli di Integrità Referenziale	77
6.2.2	Vincoli di Tupla	79
7	Analisi delle Dipendenze Funzionali e Normalizzazione	83
7.1	Verifica BCNF dell'Area Contenuti	83
7.2	Verifica BCNF dell'Area Formati	84
7.3	Verifica BCNF dell'Area Clienti	84
7.4	Verifica BCNF dell'Area Streaming	85
8	Implementazione su DBMS Oracle MySQL	87
8.1	Organizzazione dei File	87
8.2	Popolamento	87
8.3	Ridondanze	88
8.4	Vincoli	88
8.4.1	Vincoli Integrità Referenziale	88
8.4.2	Vincoli Di Tupla	88
	Bibliografia	89

Elenco delle tabelle

2.1	Dizionario delle entità dell'Area Contenuti	10
2.2	Dizionario delle relationship dell'Area Contenuti	12
2.3	Dizionario delle entità dell'Area Formati	13
2.4	Dizionario delle relationship dell'Area Formati	14
2.5	Dizionario delle entità dell'Area Clienti	16
2.6	Dizionario delle relationship dell'Area Clienti	17
2.7	Dizionario delle entità dell'Area Streaming	18
2.8	Dizionario delle relationship dell'Area Streaming	18
4.1	Tavola dei Volumi Area Contenuti	27
4.2	Tavola dei Volumi Area Formati	28
4.3	Tavola dei Volumi Area Clienti	31
4.4	Tavola dei Volumi Area Streaming	31
5.1	Tavola dei Volumi per il Calcolo del Rating	35
5.2	Tavola degli accessi in assenza di ridondanze per il Calcolo del Rating	35
5.3	Tavola degli accessi in presenza di ridondanza <i>Rating</i> per il Calcolo del Rating	36
5.4	Tavola dei Volumi per il calcolo della Raccomandazione dei Contenuti	38
5.5	Tavola degli accessi in assenza di ridondanze per Raccomandazione dei Contenuti	40
5.6	Schema della materialized view Raccomandazioni	41
5.7	Tavola degli accessi in presenza di ridondanze per Raccomandazione dei Contenuti	41
5.8	Tavola degli accessi per l'aggiornamento della ridondanza Racco- mandazioni	43
5.9	Tavola dei Volumi per il calcolo della Connessione Utente-Server . . .	45
5.10	Tavola degli accessi in assenza di ridondanze: c , rapporto tra la banda utilizzata e la banda libera	48
5.11	Tavola degli accessi in assenza di ridondanze: d , distanza tra Utente e Server CDN	48
5.12	Tavola degli accessi in presenza di ridondanza <i>Banda Disponibile</i> : c , rapporto tra la banda utilizzata e la banda libera	51

5.13	Tavola degli accessi in presenza di ridondanza <i>Banda Disponibile: d</i> , distanza tra Utente e Server CDN	51
5.14	Tavola degli accessi per aggiornamento <i>Banda Disponibile</i> all'inizio della Visualizzazione	52
5.15	Tavola degli accessi per aggiornamento <i>Banda Disponibile</i> al termine della Visualizzazione	52
5.16	Tavola dei Volumi per calcolo Acquisto Piano Abbonamento	53
5.17	Tavola degli accessi per Acquisto Piano Abbonamento	55
5.18	Tavola dei Volumi per calcolo Classifiche	56
5.19	Tavola degli accessi in assenza di ridondanze per Classifiche	58
5.20	Descrizione delle materialized view Classifiche	59
5.21	Tavola degli accessi in presenza di ridondanze per Classifiche	59
5.22	Tavola degli accessi per aggiornamento Classifiche	60
5.23	Tavola dei Volumi per Inserimento Film	62
5.24	Tavola dei accessi per Inserimento Film	64
5.25	Tavola dei Volumi per operazione Caching	65
5.26	Tavola degli accessi per Caching	67
5.27	Tavola dei Volumi per Bilanciamento del Carico	69
5.28	Tavola degli accessi per Bilanciamento del Carico Fase 1	70
5.29	Tavola degli accessi per Bilanciamento del Carico Fase 2	71
5.30	Tavola dei Volumi per Analisi delle Visualizzazioni	72
5.31	Tavola degli accessi per Analisi delle Visualizzazioni Fase 1	73
5.32	Tavola degli accessi per Analisi delle Visualizzazioni Fase 2	73

Capitolo 1

Introduzione

Benvenuti nella documentazione del progetto svolto per il corso di Basi di Dati dell'A. A. 2022-2023, Corso di Laurea in Ingegneria Informatica, presso l'Università di Pisa. Questo documento accompagna il lavoro svolto, rappresentando una panoramica completa delle attività di progettazione e implementazione del sistema di basi di dati per il servizio di streaming online Film Sphere.

La documentazione qui presente è strutturata per guidare il lettore attraverso le diverse fasi del progetto, offrendo una visione dettagliata delle scelte progettuali, delle metodologie implementate e delle soluzioni adottate per affrontare le sfide incontrate. Questa documentazione costituisce una risorsa informativa completa e sottolinea il percorso evolutivo del progetto. All'interno della documentazione abbiamo evidenziato in **grassetto** i termini che si riferiscono a entità o relationship della base di dati e in *corsivo* i termini che si riferiscono agli attributi di queste. Speriamo che questo documento fornisca una visione approfondita del nostro progetto. Buona lettura.

Capitolo 2

Progettazione Concettuale

In questo capitolo sono descritte le scelte che sono state effettuate in fase di progettazione concettuale. Viene introdotto il dizionario dei termini che descrive le entità e le relationship della base di dati. Indicheremo con * le entità, le relationship e gli attributi eliminati in fase di ristrutturazione del diagramma E-R. Indicheremo invece con ** le entità, le relationship e gli attributi aggiunti in fase di ristrutturazione del diagramma E-R. Infine indicheremo con *** le entità, le relationship e gli attributi aggiunti come ridondanze.

2.1 Area Contenuti

Riportiamo la tabella che descrive tutte le entità dell'Area Contenuti.

Entità	Attributi	Chiave	Descrizione
Film	CodiceFilm, Titolo, Paese- Produzione, Durata, Anno, Anteprima, Escusiva, Rating***	CodiceFilm	Rappresenta i contenuti presenti in Film Sphere. <i>Anteprima</i> e <i>Esclusiva</i> sono due attributi booleani.
Attore	CodiceAttore, Nome, Cognome	CodiceAttore	Figura professionale che interpreta un personaggio in almeno un Film che è o è stato presente nella base di dati. Il codice identifica univocamente ogni Attore presente nella base di dati.

Entità	Attributi	Chiave	Descrizione
Regista	CodiceRegista, Nome, Cognome	CodiceRegista	Figura professionale che dirige almeno un Film che è o è stato presente nella base di dati. Il codice identifica univocamente ogni Regista presente nella base di dati.
Genere	NomeGenere	NomeGenere	Raggruppa una data tipologia di Film della base di dati.
Premio	NomePremio, Categoria	NomePremio, Categoria	Un Premio è un riconoscimento al valore di un Film , un Attore o un Regista .
Vincita	CodiceVincita, Data	CodiceVincita	Rappresenta l'effettiva vincita di un Premio . Utile per risalire al Film per il quale un dato Attore o Regista è stato premiato.
Recensione	Upvote, Downvote, Valutazione, Descrizione	CodiceUtente (esterno), CodiceFilm (esterno)	Le Recensioni sono scritte dai vari Utenti . Ogni Utente può inoltre giudicare le recensioni scritte da altri Utenti con <i>Upvote</i> o <i>Downvote</i> . Ogni Recensione ha inoltre una <i>Valutazione</i> in decimi, e una <i>Descrizione</i> che giustifica la <i>Valutazione</i> . La chiave di una Recensione è composta dall' Utente che la scrive e dal Film a cui è riferita.

Tabella 2.1: Dizionario delle entità dell'Area Contenuti

Nella tabella seguente sono riportate le relationship dell'Area Contenuti.

Relationship	Attributi	Entità Coinvolte	Descrizione
Partecipazione	CodiceFilm, CodiceAttore	Film (1, N), Attore (0, N)	La cardinalità è stata inserita (1, N) perchè benché esistano Film senza Attori , come Film d'animazione, ci sono sempre i doppiatori che vengono considerati come Attori . La cardinalità (0, N) è presente perché nella base di dati possono essere memorizzati Attori che non hanno partecipato ad alcun Film attualmente disponibile.
Regia	CodiceRegista, CodiceFilm	Film (1, 1), Regista (0, N)	Ogni Film ha un e un solo Regista . Un Regista può invece dirigere più Film .
Classificazione	CodiceFilm, NomeGenere	Film (1, N), Genere (0, N)	Ogni Film presente nella base di dati appartiene almeno ad un Genere .
Vincita Film	NomePremio, CodiceFilm	Film (0, N), Vincita (1, 1)	Ogni Film può aver vinto più Premi . Ogni Vincita fa però riferimento ad un Film , per esempio il Film per cui è stato premiato un dato Attore o Regista .
Vincita Regista	CodiceRegista, NomePremio	Regista (0, N), Vincita (0, 1)	-
Vincita Attore	CodiceAttore, NomePremio	Attore (0, N), Vincita (0, 1)	-
Autore	CodiceUtente, CodiceFilm ¹	Utente (0, N), Recensione (1, 1)	Descrive la relazione tra l' Utente che scrive una Recensione e la Recensione stessa. Ogni Utente può scrivere più Recensione ma ovviamente ogni Recensione può essere scritta da un unico Utente .

¹In quanto fa parte della chiave di **Recensione**

Relationship	Attributi	Entità Coinvolte	Descrizione
Giudizio	CodiceUtente ² , CodiceFilm	Film (0, N), Recensione (1, 1)	Descrive la relazione tra il Film a cui fa riferimento una Recensione e la Recensione stessa. Ogni Film può essere recensito più volte ma ovviamente ogni Recensione fa riferimento ad un unico Film .

Tabella 2.2: Dizionario delle relationship dell'Area Contenuti

2.2 Area Formati

Riportiamo a seguire la tabella che descrive le entità dell'Area Formati.

Entità	Attributi	Chiave	Descrizione
File Film	CodiceFile, Dimensione	CodiceFile	Rappresenta gli effettivi file dei vari Film che vengono visualizzati dagli Utenti .
Contentitore	CodiceContentitore, Bitrate	CodiceContentitore	I Contentitori sono wrapper che uniscono un Formato Audio , e un Formato Video .
Formato Audio	CodecAudio, QualitàAudio	CodecAudio	Un Formato Audio è un tipo di compressione audio.
Formato Video	CodecVideo, QualitàVideo, Risoluzione	CodecVideo	Ogni Formato Video corrisponde a un tipo di compressione video.
Lingua	NomeLingua	NomeLingua	Sono le Lingue in cui ogni Film può essere doppiato o sottotitolato.
Stato	NomeStato, Latitudine, Longitudine	NomeStato	Rappresenta uno Stato geografico. Le coordinate fanno riferimento al centro geografico dello Stato .

²In quanto fa parte della chiave di **Recensione**

Entità	Attributi	Chiave	Descrizione
--------	-----------	--------	-------------

Tabella 2.3: Dizionario delle entità dell'Area Formati

Nella tabella seguente sono riportate le relationship dell'Area Formati.

Relationship	Attributi	Entità Coinvolte	Descrizione
Distribuzione	CodiceFilm, FileFilm	Film (1, N), File Film (1, 1)	Ogni Film avrà memorizzati nella base di dati diversi File Film .
Supporto	CodiceContenitore, CodiceFile	File Film (1, 1), Contenitore (0, N)	Ogni File Film ha un Contenitore .
Disponibilità Audio	CodiceFile, NomeLingua	File Film (0, N), Lingua (1, N)	Ogni File Film comprime più flussi audio e quindi può contenere più Lingue . È anche possibile che non ne contenga alcuna nel caso di un Film muto.
Disponibilità Sottotitoli	CodiceFile, NomeLingua	File Film (0, N), Lingua (1, N)	Ogni File Film contiene le informazioni per i sottotitoli.
Indisponibilità Formato	NomeStato, CodiceContenitore	Contenitore (0, N), Stato (0, N)	I Contenitori che compaiono nella relationship non sono disponibili nel rispettivo Stato .
Prima Lingua	NomeLingua, NomeStato	Lingua (0, N), Stato (1, 1)	Ogni Stato ha lingua principale. Una Lingua può essere parlata in più Stati . Se la Lingua principale di uno Stato non è presente tra quelle memorizzate nella base di dati verrà considerata come Prima Lingua la lingua inglese.
Fruizione	CodiceFile, TimeStampV, TimeStampC, CodiceDispositivo ³	File Film (0, N), Visualizzazione (1, 1)	Ogni Visualizzazione usufruisce di un File Film . Un File Film può essere visto da più Visualizzazioni .

³TimeStampC e CodiceDispositivo sono la chiave di **Connessione**

Relationship	Attributi	Entità Coinvolte	Descrizione
--------------	-----------	------------------	-------------

Tabella 2.4: Dizionario delle relationship dell'Area Formati

2.3 Area Clienti

Riportiamo a seguire la tabella che descrive le entità dell'Area Clienti.

Entità	Attributi	Chiave	Descrizione
Utente	CodiceUtente, Nome, Cognome, Password, Email	CodiceUtente	Un Utente è un individuo che usufruisce del servizio Film Sphere, esso deve aver acquistato almeno una volta un Piano Abbonamento .
Connessione	TimeStampC, TempoFineC, Ip, Latitudine, Longitudine	CodiceDispositivo (esterno), TimeStampC,	La Connessione rappresenta un Utente che sta attualmente usufruendo del servizio Film Sphere. Una Connessione inizia quando un Utente accede al servizio ed è nella fase di scelta di un contenuto. È identificata dal <i>CodiceDispositivo</i> ⁴ e da <i>TimeStampC</i> che indica l'istante di connessione al Server CDN .
Visualizzazione	TimeStampV, TempoFineV	TimeStampV, TimeStampC (esterno), CodiceDispositivo (esterno) ⁵	La Visualizzazione è fruizione di un File Film da parte di una Connessione .

⁴Attributo esterno e chiave di **Dispositivo**

⁵*TimeStampC* e *CodiceDispositivo* sono la chiave di **Connessione**

Entità	Attributi	Chiave	Descrizione
Dispositivo	CodiceDispositivo, Categoria, Risoluzione, RapportoAseptto	CodiceDispositivo	Il <i>CodiceDispositivo</i> identifica univocamente tutti i Dispositivi tramite cui gli Utenti hanno usufruito di Film Sphere. La <i>Categoria</i> fa riferimento al tipo del dispositivo ad esempio smartphone, computer, tablet etc.
Piano Abbonamento	NomeAbbonamento, Prezzo, Definizione, Anteprime, Esclusive	NomeAbbonamento	Film Sphere offre 5 tipi di abbonamenti differenti: Basic, Premium, Pro, Deluxe e Ultimate. Gli attributi Anteprime, Esclusive sono valori booleani e indicano se è possibile o no accedere a quelle funzionalità con l'abbonamento attualmente attivo. La durata di ogni abbonamento è mensile.
Fattura*	CodiceFattura, Importo	CodiceFattura	Rappresenta le informazioni di tutte le Fatture che sono state già pagate o che dovranno essere pagate nel prossimo mese dagli Utenti di Film Sphere.
Pagamento Effettuato	Data, Importo**	Data, CodiceUtente (esterno)	Rappresenta tutti i pagamenti effettuati dagli Utenti di Film Sphere.
Pagamento Atteso	DataAttesa, Importo**	DataAttesa, CodiceUtente (esterno)	Rappresenta le informazioni dei pagamenti attesi nel prossimo mese. Si suppone che un Utente voglia rinnovare il suo abbonamento attualmente attivo alla fine del mese.

Entità	Attributi	Chiave	Descrizione
Carta Di Credito	NumeroCarta, MeseScadenza, AnnoScadenza, CVV, Circuito, Tipo	NumeroCarta	Comprende tutte le Carte Di Credito , o meglio le Carte Elettroniche, che hanno effettuato pagamenti per usufruire dei servizi di Film Sphere. L'attributo <i>Tipo</i> indica se la Carta Elettronica sia un Bancomat, Carta Credito etc.

Tabella 2.5: Dizionario delle entità dell'Area Clienti

Nella tabella seguente sono riportate le relationship dell'Area Clienti.

Relationship	Attributi	Entità Coinvolte	Descrizione
Legame	Dispositivo, TimeStampC	Connessione (1, 1), Dispositivo (1, N)	Ogni Connessione di un Utente avviene mediante un Dispositivo , il quale può connettersi molteplici volte.
Sessione	Utente, TimeStampC, CodiceDispositivo	Utente (1, N), Connessione (1, 1)	Ogni Connessione ha un solo Utente , mentre un Utente può avere più Connessioni .
Azione	TimeStampV, TimeStampC, CodiceDispositivo	Visualizzazione (1, 1), Connessione (1, N)	Ogni Connessione deve avere almeno una Visualizzazione , ma può averne anche più di una, mentre ogni Visualizzazione è effettuata da una sola Connessione .
Pagamento Predefinito	CodiceUtente, CodiceCarta	Utente (1, 1), Carta Di Credito (0, N)	L'ultimo metodo di pagamento effettuato dall' Utente viene memorizzato nella base di dati e si considera come metodo di Pagamento Predefinito per il dato Utente . Utenti diversi possono avere lo stesso metodo di Pagamento Predefinito .

Relationship	Attributi	Entità Coinvolte	Descrizione
Addebito	CodiceFattura*, Data**, CodiceUtente**, NumeroCarta	Pagamento Effettuato (1, 1), Carta di Credito (0, N)	Ogni Carta Di Credito può effettuare più Pagamenti ma ogni Pagamento è effettuato da una e una sola Carta Di Credito .
Destinatario*	CodiceFattura, CodiceUtente	Fattura (1, 1), Utente (1, N)	Ogni Utente può effettuare il pagamento di più Fatture ma ogni Fattura è pagata da uno e un solo Utente .
Saldamento*	Nome-Abbonamento, CodiceFattura	Fattura (1, 1), PianoAbbonamento (0, N)	Ogni Piano Abbonamento può essere acquistato da più Fattura ma ogni Fattura acquista uno e un solo Piano Abbonamento .
Destinatario Effettuato**	Data, CodiceUtente	Pagamento Effettuato (1, 1), Utente (1, N)	Ogni Utente può effettuare più Pagamenti ma ogni Pagamento è effettuato da uno e un solo Utente .
Saldamento Effettuato**	Nome-Abbonamento, Data, CodiceUtente	Pagamento Effettuato (1, 1), PianoAbbonamento (0, N)	Ogni Piano Abbonamento può essere acquistato da più Pagamenti ma ogni Pagamento acquista uno e un solo Piano Abbonamento .
Destinatario Atteso**	DataAttesa, CodiceUtente	Pagamento Atteso (1, 1), Utente (1, 1)	Lega un Pagamento Atteso all' Utente che dovrà pagarlo.
Saldamento Atteso**	Nome-Abbonamento, DataAttesa, CodiceUtente	Pagamento Atteso (1, 1), PianoAbbonamento (0, N)	Lega un Pagamento Atteso al Piano Abbonamento che verrà acquistato.

Tabella 2.6: Dizionario delle relationship dell'Area Clienti

2.4 Area Streaming

Riportiamo qui una tabella che descrive le entità dell'Area Streaming.

Entità	Attributi	Chiave	Descrizione
Server CDN	CodiceServer, LarghezzaBanda, CapacitàMaxTrasmissione, BandaDisponibile***	CodiceServer	I Server CDN sono gli effettivi server a cui si collegano gli Utenti per usufruire dei servizi di Film Sphere. Film Sphere utilizza una rete composta da 13 Server CDN che sono distribuiti in diversi Stati . Ogni Server CDN è identificato da un <i>CodiceServer</i> univoco.

Tabella 2.7: Dizionario delle entità dell'Area Streaming

Nella tabella seguente sono riportate le relationship dell'Area Streaming.

Relationship	Attributi	Entità Coinvolte	Descrizione
Disponibilità Immediata	CodiceServer, CodiceFile	Server CDN (0, N), File Film (1, N)	Ogni File Film viene necessariamente caricato in uno o più Server CDN a cui si connettono gli Utenti per visualizzare il contenuto. Ogni Server CDN può contenere 0 o più File Film .
Appartenenza Geografica	CodiceServer, NomeStato	Server CDN (1, 1), Stato (0, N)	Ogni Stato contiene un Server CDN tranne gli Stati Uniti che ne contiene 3.
Presso	CodiceServer, TimeStampV, TimeStampC, CodiceDispositivo	Server CDN (0, N), Visualizzazione (1, N)	Ogni Visualizzazione può utilizzare più Server CDN e ogni Server CDN può avere più Visualizzazioni .

Tabella 2.8: Dizionario delle relationship dell'Area Streaming

2.5 Chiarimenti su Entità e Relationship

Piano Abbonamento

La tabella seguente descrive tutti i tipi di **Piano Abbonamento** disponibili e le loro caratteristiche.

Nome	Prezzo	Benefici
Basic	5€	Questo abbonamento permette lo streaming dei contenuti in SD
Premium	10€	Con l'abbonamento Premium si possono visionare contenuti in Full HD
Pro	12,5€	Con questo abbonamento si possono visionare i contenuti in 4k
Deluxe	15€	Con questo piano, oltre a poter visionare i contenuti in 4k , si possono anche visionare i Film in <i>Anteprima</i>
Ultimate	20€	Con questo piano, si possono visionare i contenuti in 4k , si possono inoltre guardare i Film in <i>Anteprima</i> e i Film Esclusivi

Connessione

Una **Connessione** rappresenta una sessione di utilizzo del servizio Film Sphere da parte di un **Utente** con un dato **Dispositivo**. La **Connessione** inizia quando un **Utente** è in fase di scelta di un contenuto, per questo abbiamo ipotizzato che fosse necessario mostrare all'**Utente** tutte le informazioni, prodotte da diverse operazioni che illustreremo nel Capitolo 5, che permettono di fare una scelta migliore.

Pagamento Effettuato

Il **Pagamento Effettuato** di un **Piano Abbonamento** da parte di un **Utente** permette all'**Utente** di usufruire del servizio per 1 mese. Infatti ogni abbonamento ha una durata mensile.

Pagamento Atteso

Si suppone che nel mese successivo un **Utente** voglia rinnovare il **Pagamento Effettuato** nel mese precedente, quindi acquistando lo stesso **Piano Abbonamento**.

Stato

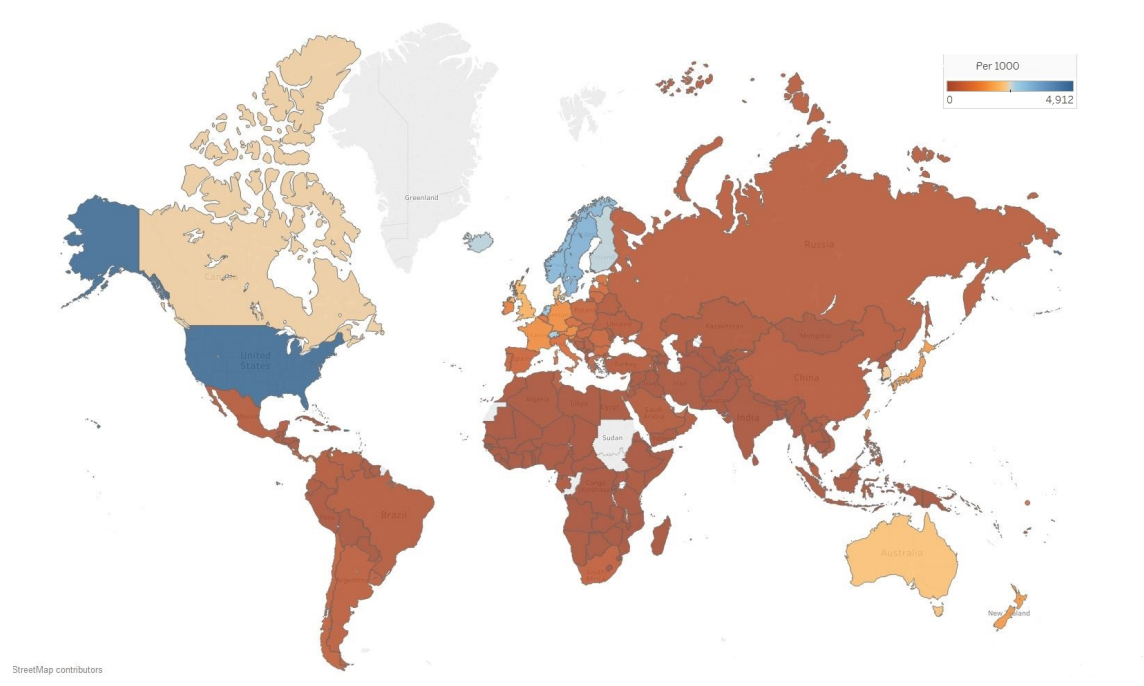
Nella base di dati sono memorizzate le informazioni riguardo a tutti gli **Stati** del mondo anche se il servizio è disponibile solo in 30 **Stati** differenti. Questa scelta è stata effettuata così da poter espandere il servizio in nuovi **Stati** senza dover modificare la base di dati.

Server CDN

Abbiamo deciso di dotare Film Sphere di una rete composta da 13 **Server CDN** in modo da coprire nella maniera più uniforme possibile gli **Stati** da cui provengono le **Connessioni**. I **Server CDN** sono stati collocati come segue:



Per la scelta della posizione dei **Server CDN** abbiamo utilizzato anche il seguente grafico che illustra la quantità di indirizzi *Ip* assegnati ad ogni **Stato**. Considerando un indirizzo *Ip* come un possibile **Utente** otteniamo un dato importante per la scelta della posizione dei **Server CDN**.



Capitolo 3

Ristrutturazione del diagramma E-R

In questo capitolo sono discusse le scelte che sono state effettuate in fase di ristrutturazione del diagramma E-R.

3.1 Eliminazione Attributi Composti

L'attributo composto *CoordinateCentro* dell'entità **Stato** è stato eliminato e sostituito da gli attributi *Latitudine* e *Longitudine* che proprio come *CoordinateCentro* indicato il centro geografico di uno **Stato**.

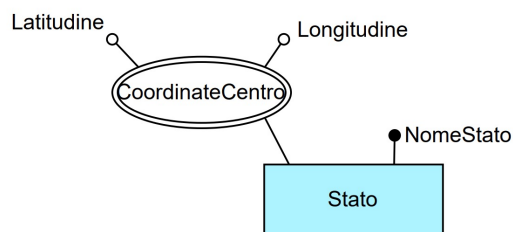


Figura 3.1: E-R non Ristrutturato

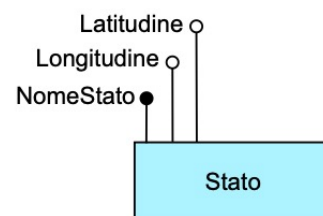
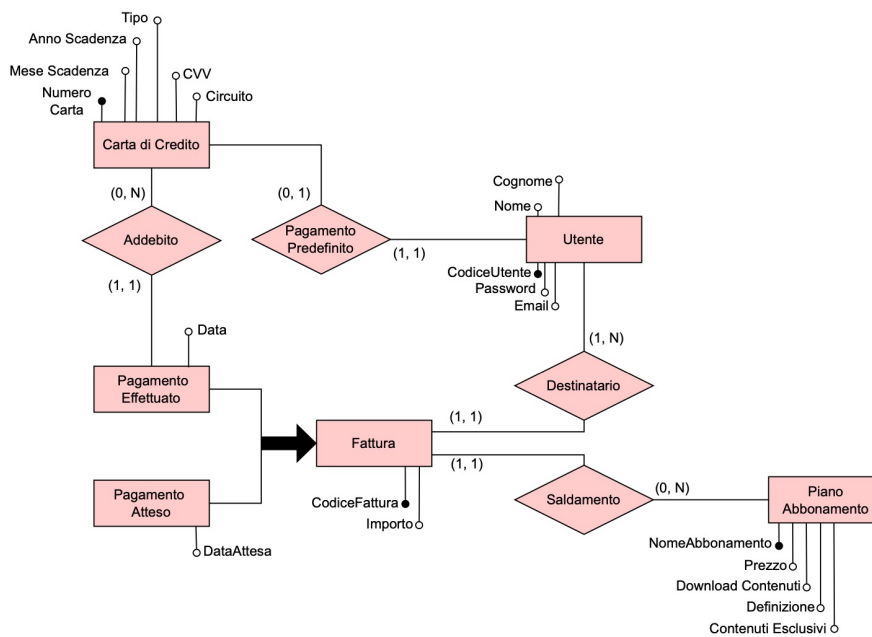


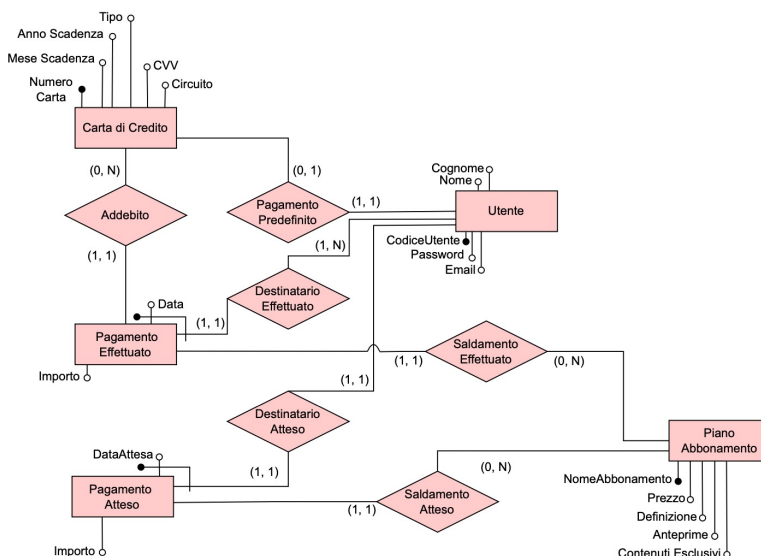
Figura 3.2: E-R Ristrutturato

3.2 Eliminazione Generalizzazioni

Nello schema E-R non ristrutturato compariva la generalizzazione totale **Fattura** di **Pagamento Effettuato** e **Pagamento Atteso**, come si vede nella seguente figura.



In fase di ristrutturazione è stato deciso di accorpare il padre alle figlie in quanto è necessario spesso accedere alle due associazioni in tempi e operazioni distinte. L'accorpamento delle figlie al padre sarebbe stato quindi sconveniente. Nello schema ristrutturato sparisce l'entità **Fattura** e gli attributi *CodiceFattura*¹ e *Importo* passano ad entrambe le figlie. Si raddoppiano anche le relationship del padre. In particolare **Destinatario** si divide in **Destinatario Effettuato** e **Destinatario Atteso**, e **Saldamento** si divide in **Saldamento Effettuato** e **Saldamento Atteso**.



¹In realtà l'attributo *CodiceFattura* che era chiave primaria di **Fattura** sparisce in entrambe le figlie in quanto si sceglie una chiave esterna composta da *Data* e *CodiceUtente*.

Capitolo 4

Tavola dei Volumi

Riportiamo a seguire la Tavola dei Volumi. Per il calcolo dei volumi abbiamo ipotizzato che Film Sphere sia attivo da 1 anno. Per ogni concetto del diagramma E-R sono specificate nella colonna Descrizione le ulteriori assunzioni utilizzate per il calcolo del relativo volume. All'interno della tabella si troveranno prima le entità di una data Area della base di dati, a seguire le rispettive relationship. Abbiamo deciso di presentare le aree nel seguente ordine:

1. Area Contenuti
2. Area Formati
3. Area Clienti
4. Area Streaming

4.1 Area Contenuti

Concetto	Tipo	Volume	Descrizione
Film	E	200	Abbiamo assunto che un servizio di streaming con una vita di un anno possa realisticamente contare 200 contenuti. Si rimanda alla sezione 4.5 per maggiori chiarimenti su questa scelta.
Recensione	E	2'000	Ogni Film ha in media 10 Recensioni .
Genere	E	14	Abbiamo individuato 14 Generi significativi a cui possono appartenere i Film di Film Sphere.

Concetto	Tipo	Volume	Descrizione
Attore	E	1'000	Abbiamo assunto che ogni Film abbia in media 15 Attori significativi (escluse comparse) e che ogni Attore presente nella base di dati abbia partecipato ad almeno 3 Film . $(15 * 200/3) = 1'000$.
Regista	E	100	Si assume che ogni Regista abbia diretto in media 2 Film presenti nel database.
Premio	E	110	Abbiamo individuato e scelto gli 11 Premi più prestigiosi da memorizzare all'interno della base di dati. Per ogni Premio abbiamo individuato 10 <i>Categorie</i> .
Vincita	E	75	Ogni Film ha vinto in media 0,1 Premi . Ogni Attore ha vinto in media 0,05 Premi . Ogni Regista ha vinto in media 0,05 Premi . $(200 * 0,1 + 1'000 * 0,05 + 100 * 0,05) = 75$.
Partecipazione	R	3'000	Ogni Film ha in media 15 Attori significativi.
Vincita Premio	R	75	Ogni Vincita ha un Premio associato.
Vincita Film	R	75	Ogni Film ha vinto in media 0,1 Premi , ma se consideriamo le Vincite di un Premio da parte di un Attore o Regista del Film come Vincita del Film allora le Vincite Film ammontano a $(200*0,1 + 100*0,05 + 1'000*0,05) = 75$.
Vincita Attore	R	50	Ogni Attore ha vinto in media 0,05 Premi .
Vincita Regista	R	5	Ogni Regista ha vinto in media 0,05 Premi .
Regia	R	200	Ogni Film ha un Regista .

Concetto	Tipo	Volume	Descrizione
Giudizio	R	2'000	Ogni Recensione fa riferimento ad un Film . Il volume di Giudizio deve essere uguale a quello di Recensione .
Autore	R	2'000	Ogni Recensione è scritta da un Utente . Il volume di Autore deve essere uguale a quello di Recensione .
Classificazione	R	600	Si stima che ogni Film appartenga in media a 3 Generi .

Tabella 4.1: Tavola dei Volumi Area Contenuti

4.2 Area Formati

Concetto	Tipo	Volume	Descrizione
Stato	E	193	La base di dati memorizza informazioni su tutti gli Stati del pianeta anche se il servizio è attivo solo in 30 Stati differenti.
Lingua	E	10	Film Sphere ha Film in 10 Lingue diverse.
File Film	E	3'000	Si hanno 15 File Film per ogni Film ($200 * 15$). Per ogni Film si ha un File Film codificato in ogni Contenitore .
Contenitore	E	15	Abbiamo ipotizzato che ogni Formato Video può essere accoppiato con ogni Formato Audio per formare un Contenitore .
Formato Video	E	5	Abbiamo considerato i 5 Formati Video più comuni da memorizzare nella base di dati.

Concetto	Tipo	Volume	Descrizione
Formato Audio	E	3	Abbiamo memorizzato, nella base di dati, i 3 Formati Audio più comuni.
Fruizione	R	2'190'000	Ogni Visualizzazione usufruisce di un solo File Film e dunque il volume di Fruizione deve essere uguale a quello di Visualizzazione .
Distribuzione	R	3'000	Ogni File Film è un file codificato di un unico Film presente nella base di dati. Il volume di Distribuzione deve essere uguale a quello di File Film .
Supporto	R	3'000	Ogni File Film ha un Contentitore e dunque il volume di Supporto deve essere uguale a quello di File Film .
Disponibilità Audio	R	9'000	Ipotizziamo ogni File Film racchiude in media 3 Lingue per la traccia audio.
Disponibilità Sottotitoli	R	9'000	Ogni File Film comprende almeno 3 Lingue per i sottotitoli.
Indisponibilità Formato	R	150	Si stima che in ogni Stato in cui sia disponibile Film Sphere, che ricordiamo essere 30, non sia possibile visionare File Film con 5 Contentitori differenti ($30 * 5 = 150$).
Contenimento Audio	R	15	Ogni Contentitore ha un Formato Audio .
Contenimento Video	R	15	Ogni Contentitore ha un Formato Video .
Prima Lingua	R	193	Ogni Stato ha una Prima Lingua .

Tabella 4.2: Tavola dei Volumi Area Formati

4.3 Area Clienti

Concetto	Tipo	Volume	Descrizione
Utente ¹	E	50'000	Abbiamo stimato che un servizio di streaming dopo 1 anno di attività possa avere 50'000 Utenti .
Dispositivo	E	100'000	Ogni Utente utilizza in media 2 Dispositivi .
Piano Abbonamento	E	5	Abbiamo scelto di introdurre 5 Piani Abbonamento differenti come spiegato nella sezione 2.5.
Connessione	E	1'825'000	Un Utente si connette in media 1 volta ogni due giorni per 365 giorni. Supponiamo che in media ci siano sempre almeno 10'000 Utenti con abbonamento attivo. $(10'000 * 0,5 * 365 = 1'825'000)$.
Visualizzazione	E	2'190'000	Si stima che una Connessione effettui in media 1,2 Visualizzazioni .
Carte di Credito	E	60'000	Un Utente utilizza in media 1,2 Carte di Credito .
Pagamento Effettuato	E	120'000	Si stima che ogni mese ci siano 10'000 Utenti attivi che rinnovano .
Pagamento Atteso	E	10'000	Se supponiamo che gli Utenti con un abbonamento attivo sono in media 10'000 per ogni mese, allora si hanno anche 10'000 Pagamenti Attesi alla fine del mese.
Pagamento Predefinito	R	50'000	Ogni Utente ha un metodo di Pagamento Predefinito .
Destinatario Atteso	R	10'000	Ogni Pagamento Atteso ha un destinatario e dunque il volume di Destinatario Atteso deve essere uguale a quello di Pagamento Atteso .

¹Questi sono gli **Utenti** che hanno avuto o hanno un **Piano Abbonamento** attivo, si stima che ogni mese ci siano circa 10'000 **Utenti** attualmente attivi.

Concetto	Tipo	Volume	Descrizione
Saldamento Atteso	R	10'000	Ogni Pagamento Atteso ha un solo Piano Abbonamento acquistato e dunque il volume di Saldamento Atteso deve essere uguale a quello di Pagamento Atteso .
Destinatario Effettuato	R	120'000	Ogni Pagamento Effettuato ha un solo destinatario e dunque il volume di Destinatario Effettuato deve essere uguale a quello di Pagamento Effettuato .
Saldamento Effettuato	R	120'000	Ogni Pagamento Effettuato ha un solo Piano Abbonamento acquistato e dunque il volume di Saldamento Effettuato deve essere uguale a quello di Pagamento Effettuato .
Addebito	R	120'000	Ogni Pagamento Effettuato ha una sola Carta Di Credito utilizzata e quindi il volume di Addebito deve essere uguale a quello di Pagamento Effettuato .
Sessione	R	1'825'000	Ogni Connessione è effettuata da un solo Utente e dunque il volume di Sessione deve essere uguale a quello di Connessione .
Azione	R	2'190'000	Ogni Visualizzazione è effettuata da una sola Connessione e dunque il volume di Azione deve essere uguale a quello di Visualizzazione .
Legame	R	1'825'000	Ogni Connessione è effettuata da un solo Dispositivo e dunque il volume di Legame deve essere uguale a quello di Connessione .

Concetto	Tipo	Volume	Descrizione
----------	------	--------	-------------

Tabella 4.3: Tavola dei Volumi Area Clienti

4.4 Area Streaming

Concetto	Tipo	Volume	Descrizione
Server CDN	E	13	Abbiamo scelto 13 Server CDN distribuiti su tutto il pianeta.
Disponibilità Immediata	R	9'000	Ogni File Film è memorizzato contemporaneamente in 3 Server CDN .
Presso	R	2'628'000	Si stima che ogni Visualizzazione utilizzi in media 1,2 Server CDN per visualizzare File Film .
Appartenenza Geografica	R	13	Ogni Server CDN è localizzato in uno Stato .

Tabella 4.4: Tavola dei Volumi Area Streaming

4.5 Chiarimenti Volume Film

Il numero di **Film** che Film Sphere offre, ovvero 200, può sembrare piccolo. Utilizziamo però dati reali per contestualizzare la scelta. Netflix, uno dei più grandi servizi di streaming online, conta 247'150'000 **Utenti** e 17'000 **Film**. Possiamo calcolare dunque il rapporto:

$$R = \frac{\text{Utenti}}{\text{Film}}$$

Vediamo che per Netflix questo rapporto è di 14'538,24. Mantendo lo stesso rapporto in Film Sphere, con 50'000 **Utenti** avremmo avuto solo $(50'000/14'538,24) = 3,44$ **Film**.

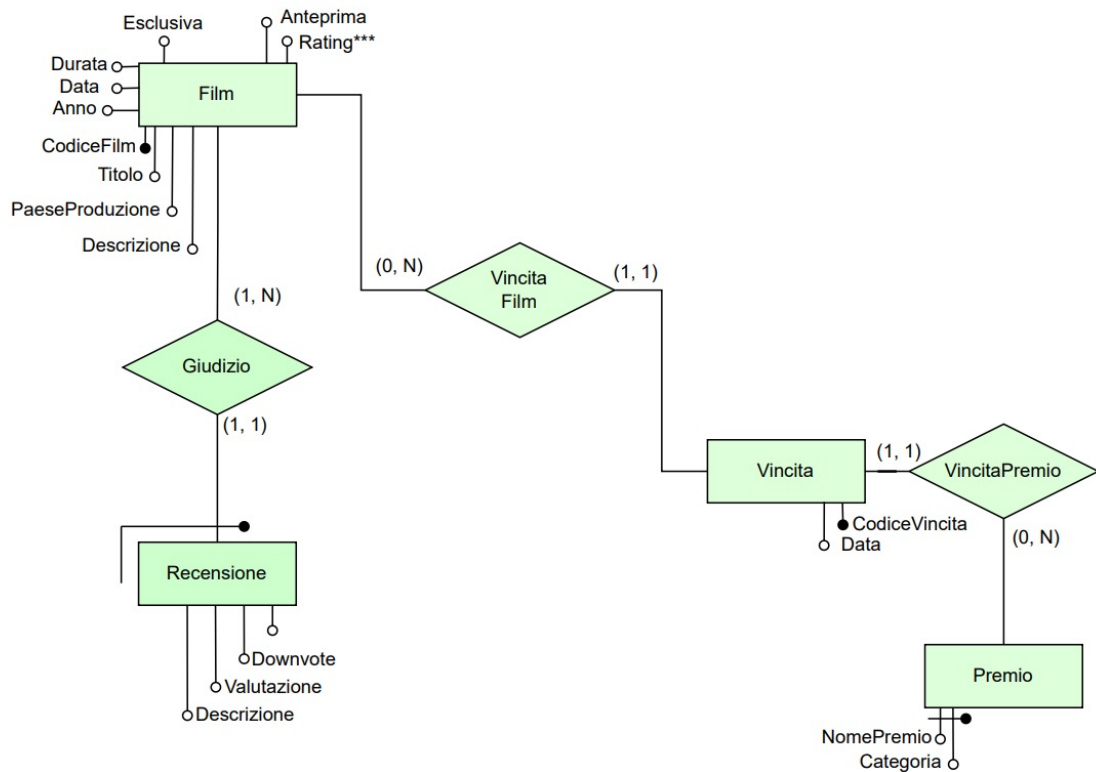
Abbiamo però ipotizzato che il rapporto tra **Utenti** e **Film** non scali linearmente con il diminuire degli **Utenti** e siamo dunque arrivati ad un numero ragionevole di 200 **Film**.

Capitolo 5

Individuazione e Analisi delle Operazioni

In questo capitolo andremo a descrivere e analizzare le operazioni che sono state individuate sulla base di dati.

5.1 Calcolo del Rating di un Film



Operazione che prende in input un **Film** e produce in output un *Rating*.

L'operazione viene eseguita ogni volta che **Utente** visualizza un **Film** ovvero 6'000 volte al giorno in quanto si vuole mostrare il *Rating* del **Film** all'**Utente** che lo visualizza. Per calcolo del *Rating* R viene utilizzata una media pesata delle **Recensioni**, che conta 9/10 del *Rating*, a cui viene sommata una costante V^p , che viene calcolata in base ai **Premi** vinti dal **Film**.

Introduciamo le varie quantità che entrano nel calcolo del *Rating* R .

Sia r_i la valutazione in decimi della i -esima recensione di un determinato **Film**.¹

Sia p_i il peso della suddetta **Recensione** definito come:

$$p_i = \begin{cases} u_i - d_i & u_i - d_i \geq 0 \\ \frac{1}{d_i - u_i} & u_i - d_i < 0 \end{cases}$$

dove u_i sono i *Upvote* ricevuti dalla **Recensione** i , mentre d_i sono i *Downvote* ricevuti dalla **Recensione** i . Con questa scelta vengono pesate maggiormente le **Recensioni** che hanno riscontrato feedback positivo dagli **Utenti**.

Inoltre si considerano nel modo seguente i **Premi** vinti dal **Film**. Supponiamo che il **Film** abbia vinto m **Premi** e che ogni **Premio** abbia un valore v_i^p . Si sommano i valori dei **Premi** vinti fino ad un massimo di 1 punto definendo nel modo seguente la quantità V^p somma del valore totale dei **Premi** vinti da un **Film**:

$$V^p = \begin{cases} \sum_{i=1}^m v_i^p & \sum_{i=1}^m v_i^p \leq 1 \\ 1 & \sum_{i=1}^m v_i^p > 1 \end{cases}$$

Il *Rating* R di ogni **Film** che abbia ottenuto n **Recensioni**, ognuna con *Valutazione* r_i e peso p_i , viene infine calcolato secondo la seguente formula²:

$$R = \frac{\sum_{i=1}^n p_i r_i}{\sum_{i=1}^n p_i} \cdot \frac{9}{10} + V_p$$

I volumi coinvolti nel Calcolo del *Rating* di conseguenza sono:

Concetto	Tipo	Volume
Giudizio	R	2'000

¹Ricordiamo che ogni **Recensione** ha una valutazione in decimi, un numero di *Upvote*, un numero di *Downvote* e una *Descrizione*.

²La media delle recensioni pesate è normalizzata a 9 in modo tale da ottenere R minore o uguale a 10.

Concetto	Tipo	Volume
Recensione	E	2'000
Vincita Film	R	75
Premio	E	110

Tabella 5.1: Tavola dei Volumi per il Calcolo del Rating

Analizziamo gli accessi necessari per il calcolo del *Rating* sia in assenza che in presenza di ridondanze per poi trarre le nostre conclusioni sull'utilità di queste ultime.

5.1.1 Tavola degli Accessi in Assenza di Ridondanze

Scriviamo gli accessi per il calcolo del *Rating*:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Giudizio	R	2'000	R	Data la chiave di Film è necessario scorrere Giudizio per trovare le Recensioni del Film in esame
2	Recensione	E	10	R	Trovate le 10 Recensioni , vengono letti i loro valori e i valori degli <i>Upvote</i> e <i>Downvote</i>
3	Vincita Film	R	75	R	Scansione necessaria al fine di trovare gli eventuali Premi vinti dal Film
4	Premio	E	0,38	R	Lettura del Premio vinto dal Film . Si stima che ogni Film abbia vinto in media 0,38 Premi . $75/200 = 0,38$.

Tabella 5.2: Tavola degli accessi in assenza di ridondanze per il Calcolo del Rating

Di conseguenza il totale di accessi giornalieri risultano: $(2'000 + 10 + 25 + 0,38) * 6'000 = \mathbf{12'512'280}$ accessi.

5.1.2 Tavola degli Accessi in Presenza di Ridondanza *Rating*

Valutiamo a seguire la tavola degli accessi per il calcolo del rating in presenza della ridondanza *Rating* in **Film**, aggiornata con frequenza giornaliera. Tavola degli accessi per il calcolo del rating:

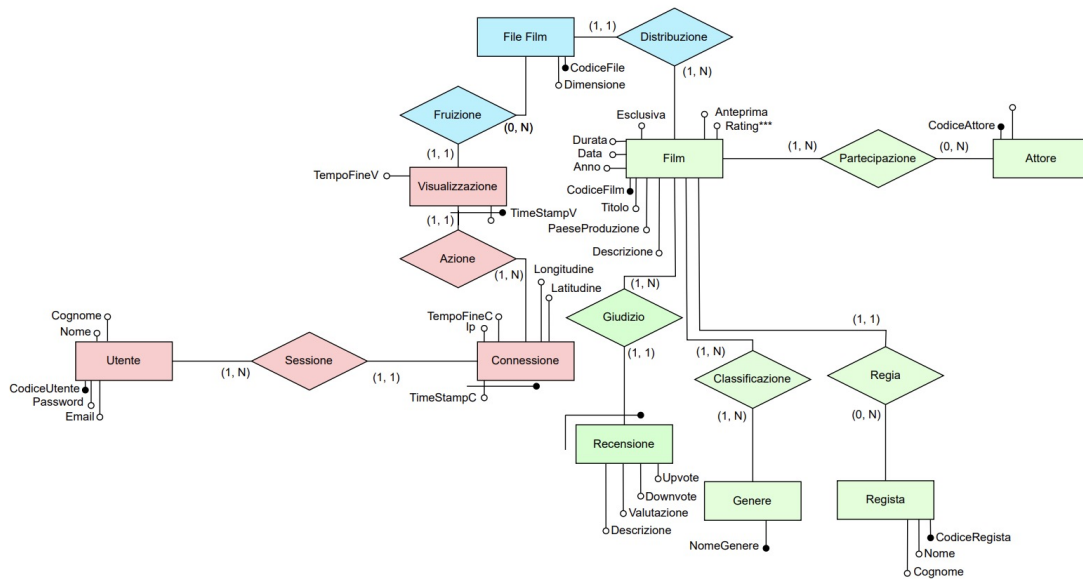
Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Giudizio	R	2'000	R	Data la chiave di Film è necessario scorrere Giudizio per trovare le Recensioni del Film in esame
2	Recensione	E	10	R	Trovate le 10 Recensioni , vengono letti i loro valori e i valori degli <i>Upvote</i> e <i>Downvote</i>
3	Vincita Film	R	75	R	Scansione necessaria al fine di trovare gli eventuali Premi vinti dal Film
4	Premio	E	0,38	R	Lettura del premio vinto dal film ³

Tabella 5.3: Tavola degli accessi in presenza di ridondanza *Rating* per il Calcolo del Rating

In questo caso gli accessi totali giornalieri ammontano a $(2'000 + 10 + 75 + 0,38) * 200 = 407'076$. In presenza di *Rating* come ridondanza cambia la frequenza dell'operazione in quanto il *Rating* non si deve calcolare ogni volta che un **Utente** visualizza un contenuto, ma si calcola un'unica volta per ogni **Film** della piattaforma. A questi accessi si sommano i singoli accessi di lettura fatti al **Film** per mostrare il *Rating* all'**Utente**, che sono 6'000 al giorno. In conclusione, gli accessi totali giornalieri in presenza di ridondanza ammontano a **412'076**. Possiamo affermare che la ridondanza *Rating* alleggerisce considerevolmente il carico di questa operazione.

³Si stima che ogni **Film** abbia vinto in media 0,38 **Premi**.

5.2 Raccomandazione dei Contenuti



L'operazione ha come input un *Codice Utente*, e come output i 5 **Film** che hanno punteggio P più alto. L'operazione viene effettuata ogni volta che si vogliono mostrare ad un **Utente** i **Film** che il sistema gli raccomanda, ovvero ogni volta che l'**Utente** si connette a Film Sphere, 5'000 volte al giorno. I dati utilizzati per il calcolo del punteggio P sono quelli relativi ai 30 giorni antecedenti al primo giorno della settimana in corso⁴.

Il punteggio P di ogni **Film** è calcolato sommando i seguenti contributi:

a) Genere

- se il **Film** appartiene al primo **Genere** più visto si somma 100 al punteggio P
- se il **Film** appartiene al secondo **Genere** più visto si somma 85 al punteggio P
- se il **Film** appartiene al terzo **Genere** più visto si somma 70 al punteggio P

b) Rating

Si somma al punteggio P il punteggio P_r relativo al *Rating* R calcolato come segue:

$$P_r = R * 50$$

Questa scelta è stata messa in atto per consigliare **Film** meglio recensiti a parità di compatibilità con l'**Utente**

⁴Esempio: se l'operazione viene richiesta Mercoledì 22 Novembre, i dati utilizzati saranno quelli raccolti tra il 20 Ottobre e Lunedì 20 Novembre.

c) **Attore**

Se l'**Attore** compare tra i 5 **Attori** più apprezzati dall'**Utente** nell'ultimo mese ⁵ si somma 25 al punteggio P

d) **Regista**

Se il **Regista** compare tra i 3 **Registi** più apprezzati ⁶ si somma 35 al punteggio P

I volumi coinvolti di conseguenza sono:

Concetto	Tipo	Volume
Sessione	R	1'825'000
Azione	R	2'190'000
Fruizione	R	2'190'000
Distribuzione	R	3'000
Classificazione	R	3'000
Genere	E	14
Partecipazione	R	3'000
Regia	E	200
Attore	E	1'000
Regista	E	100
Giudizio	R	2'000
Recensione	E	2'000

Tabella 5.4: Tavola dei Volumi per il calcolo della Raccomandazione dei Contenuti

Analizziamo gli accessi necessari per il calcolo delle Raccomandazioni dei Contenuti sia in assenza che in presenza di ridondanze per poi trarre le nostre conclusioni sull'utilità di queste ultime.

5.2.1 Tavola degli Accessi in Assenza di Ridondanze

A seguire ripostiamo la tavola degli accessi:

⁵Si considerano come **Attori** più apprezzati gli **Attori** che compaiono il maggior numero di volte nei **Film** visti.

⁶Si considerano come **Registi** più apprezzati i **Registi** che compaiono il maggior numero di volte nei **Film** visti.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Sessione	R	1'825'000	R	Dato il <i>Codice Utente</i> si scorre la relation-ship Sessione per trovare le Connessioni che l' Utente ha effettuato nell'ultimo mese
2	Connessione	R	45 ⁷	R	Si leggono le Connessioni effettuate dall' Utente dall'inizio della sua attività per poi individuare quelle fatte nell'ultimo mese che si stimano essere 15
3	Azione	R	2'190'000	R	Date le Connessioni effettuate nell'ultimo mese si cercano le chiavi di Visualizzazione , che si stimano essere 18 ($15 \cdot 1,2 = 18$), necessarie per trovare i Film visti dall' Utente
4	Fruizione	R	18	R	Date le Visualizzazioni effettuate nell'ultimo mese si cercano le chiavi dei File Film visualizzati
5	Distribuzione	R	18	R	Dati i File Film visualizzati dall' Utente si trovano gli effettivi 18 Film
6	Classificazione	R	600	R	Dati i Film si calcolano i 3 generi più visti dall' Utente nell' ultimo mese
7	Regia	R	18	R	Dati i Film si calcolano i 3 Registi più apprezzati
8	Partecipazione	R	3'000	R	Dati i Film si calcolano i 5 Attori più apprezzati

⁷È stato stimato che un **Utente** abbia in media un **Piano Abbonamento** attivo per 3 mesi. Facendo 1 **Connessione** ogni due giorni, si conatterà 45 volte.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
9	Film	R	200	R	Si scorre la tabella Film per trovare i Film non ancora visti dall' Utente nell'ultimo mese
10	Classificazione	R	600	R	Per ogni Film non ancora visto dall' Utente si somma il punteggio in base al Genere
11	Regia	R	182	R	Per ogni Film non ancora visto dall' Utente si sommano i punteggi in base al Regista
12	Partecipazione	R	3'000	R	Per ogni Film non ancora visto dall' Utente si sommano i punteggi in base agli Attori
13	Giudizio	R	2'000	R	Si trovano le Recensioni relative al Film
14	Recensione	R	10	R	Si legge il numero di <i>Up-vote</i> e <i>Downvote</i> delle Recensioni del Film

Tabella 5.5: Tavola degli accessi in assenza di ridondanze per Raccomandazione dei Contenuti

Di conseguenza gli accessi totali per il calcolo delle raccomandazioni per un **Utente** sono **4'032'201**. Dato che l'operazione è ripetuta ogni volta che un **Utente** si connette, ovvero 5'000 volte al giorno, gli accessi giornalieri ammontano a **20'161'005'000**.

5.2.2 Tavola degli Accessi in Presenza di Ridondanza Raccomandazioni

Andiamo ad introdurre la materialized view **Raccomandazioni** che contiene i **Film** raccomandati dal sistema ad ogni **Utente**. La materialized view viene aggiornata con modalità **scheduled** settimanalmente, utilizzando la stessa politica per la raccolta dati dell'operazione in assenza di ridondanza ovvero: i dati utilizzati per il calcolo del punteggio P sono quelli relativi ai 30 giorni antecedenti al primo giorno della settimana in corso.

Andiamo a vedere la struttura di **Raccomandazione**

Concetto	Attributi	Chiave	Descrizione
Raccomandazioni	CodiceUtente, Film, Punteggio	CodiceUtente, Film	La materialized view contiene per ogni Utente i 5 Film che gli sono stati raccomandati nell'ultima settimana e il relativo <i>Punteggio</i> P

Tabella 5.6: Schema della materialized view **Raccomandazioni**

Vediamo come cambiano gli accessi per mostrare ad un **Utente** le sue **Raccomandazioni** della settimana:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Raccomandazioni	MV	50'000 ⁸	R	Si leggono i Film raccomandati all' Utente

Tabella 5.7: Tavola degli accessi in presenza di ridondanze per Raccomandazione dei Contenuti

Gli accessi giornalieri diventano dunque **250'000'000**.

Vediamo gli accessi per tenere aggiornata **Raccomandazioni**:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Sessione	R	1'825'000	R	Dato il <i>Codice Utente</i> si scorre la relationship Sessione per trovare le Connessioni che l' Utente ha effettuato nell'ultimo mese
2	Connessione	R	45 ⁹	R	Si leggono le Connessioni effettuate dall' Utente per poi individuare quelle fatte nell'ultimo mese che si stimano essere 15

⁸Dato che la chiave di **Raccomandazioni** è *CodiceFilm*, *CodiceUtente* non si può fare accesso diretto avendo solo il *CodiceUtente*.

⁹È stato stimato che un **Utente** abbia in media un **Piano Abbonamento** attivo per 3 mesi. Facendo 1 **Connessione** ogni due giorni, si conatterà 45 volte.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
3	Azione	R	2'190'000	R	Date le Connessioni effettuate nell'ultimo mese si cercano le chiavi di Visualizzazione , che si stimano essere 18, necessarie per trovare i Film visti dall' Utente
4	Fruizione	R	18	R	Date le Visualizzazioni effettuate nell'ultimo mese si cercano le chiavi dei File Film visualizzati
5	Distribuzione	R	18	R	Dati i File Film visualizzati dall' Utente si trovano gli effettivi Film visti, si stimano essere 18
6	Classificazione	R	600	R	Dati i Film si calcolano i 3 generi più visti dall' Utente nell' ultimo mese
7	Regia	R	18	R	Dati i Film si calcolano i 3 Registi più apprezzati
8	Partecipazione	R	3'000	R	Dati i Film si calcolano i 5 Attori più apprezzati
9	Film	R	200	R	Si scorre la tabella Film per trovare i Film non ancora visti dall' Utente nell'ultimo mese
10	Classificazione	R	600	R	Per ogni Film non ancora visto dall' Utente si somma il punteggio in base al Genere
11	Regia	R	182	R	Per ogni Film non ancora visto dall' Utente si sommano i punteggi in base al Regista
12	Partecipazione	R	3'000	R	Per ogni Film non ancora visto dall' Utente si sommano i punteggi in base agli Attori
13	Giudizio	R	2'000	R	Si trovano le Recensioni relative al Film

operazione è di **6'000** volte al giorno, viene effettuata ogni volta che si ha una nuova **Visualizzazione**. Utilizziamo la seguente formula per il calcolo del **Server CDN** più conveniente. Il calcolo viene effettuato solo per i **Server CDN** che hanno il **File Film** del **Film** che è richiesto dalla **Connessione**. Per la scelta del **File Film** addatto si deve tener conto della posizione della **Connessione** per determinare la **Lingua** in cui si vuole visionare il **File Film**. Ricordiamo anche che la *Risoluzione* massima dei **File Film** che un **Utente** può visualizzare è determinata dal **Piano Abbonamento** che l'**Utente** ha attivo. Ricordiamo infine che alcuni **Contentitori** non sono disponibili in diversi **Stati**.

Definiamo c come il rapporto tra la banda utilizzata e la banda libera di un **Server CDN**, e d come la distanza tra **Utente** e **Server CDN**, al minimo pari a 1, se il **Server CDN** e l'**Utente** si trovano nello stesso **Stato**. Al momento della **Connessione** di un **Utente** ad un **Server CDN** verrà associato a quest'ultimo un indice di compatibilità I_c , calcolato come segue:

$$I_c = \frac{1 - c}{d^2}$$

Più alto è l'indice I_c , migliore è considerata la **Connessione** tra l'**Utente** e il **Server CDN**, di conseguenza per un **Utente** verrà scelto il **Server CDN** che genera indice più alto.

I volumi coinvolti sono:

Concetto	Tipo	Volume
Connessione	E	1'825'000
Film	E	200
Distribuzione	R	3'000
File Film	E	3'000
Supporto	R	3'000
Contentitore	E	50
Disponibilità Immediata	R	9'000
Server CDN	E	13
Presso	R	2'628'000
Sessione	R	1'825'000
Visualizzazione	R	2'190'000
Fruizione	R	2'190'000
Azione	R	2'190'000
Stato	E	193
Indisponibilità Formato	R	150
Appartenenza Geografica	R	13
Sessione	R	1'825'000
Destinatario Effettuato	R	120'000
Saldamento Effettuato	R	120'000
Piano Abbonamento	E	5

Concetto	Tipo	Volume
Prima Lingua	R	193
Disponibilità Audio	R	9'000
Formato Video	E	5
Formato Audio	E	3

Tabella 5.9: Tavola dei Volumi per il calcolo della Connessione Utente-Server

Andiamo ora ad analizzare gli accessi necessari per il calcolo di I_c . Andremmo a vedere gli accessi in assenza e in presenza di ridondanze per poi trarre le nostre conclusioni sull'utilità di quest'ultime.

5.3.1 Tavola degli Accessi in Assenza di Ridondanze

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Distribuzione	R	3'000	R	Data la chiave di Film dobbiamo scorrere tutta la relationship per trovare i 15 File Film corrispondenti
2	Destinatario Effettuato	R	120'000	R	Trovo la chiave dei Pagamenti Effettuati dall' Utente che stimiamo essere 3
3	Pagamento Effettuato	E	3	R	Leggo la <i>Data</i> per trovare la chiave dell'ultimo Pagamento Effettuato
4	Saldamento Effettuato	R	1	R	Trovo la chiave di Piano Abbonamento
5	Piano Abbonamento	E	1	R	Si legge l'attributo <i>Definizione</i> , così da scegliere il File Film utilizzabile dall' Utente
6	Stato	E	1	R	Si fa un accesso diretto allo stato da cui proviene la Connessione ¹¹
7	Indisponibilità Formato	R	150	R	Si trovano i Contenitori non disponibili nello Stato da cui proviene la Connessione

¹¹Ciò è possibile perché di ogni **Connessione** abbiamo memorizzate le coordinate geografiche che sono anche superchiave di **Stato**.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
8	Prima Lingua	R	1	R	Si trova la chiave della Prima Lingua per lo Stato
9	Disponibilità Audio	R	9'000	R	Si trovano i File Film del Film richiesto che contengono la Lingua che si stimano essere 8
10	Supporto	R	5	R	Trovo le chiavi dei Contentitori dei File Film con la Lingua richiesta
11	Contenimento Video	R	2	R	Trovo le chiavi dei 2 rispettivi Formati Video . Si stima che 5 Contentitori distinti abbiano 2 Formati Video distinti
12	Formato Video	R	2	R	Leggo l'attributo <i>Risoluzione</i> per verificare che rispetti il vincolo del Piano Abbonamento ¹²
13	Disponibilità Immediata	R	9'000	R	Si scorre per trovare i Server CDN che contengono i 8 File Film che rispettano i vincoli di Lingua e di <i>Definizione</i> . Si stima che i File Film siano equamente distribuiti in 3 Server CDN
14	Presso	R	2'628'000	R	Tramite Presso si ottengo le chiavi delle altre Visualizzazioni che utilizzano i 3 Server CDN in esame. In uno anno si stima di trovare 505'384 Visualizzazioni , (Totale Visualizzazioni)/13 * 3.

¹²Ho quindi trovato le chiavi dei **Contentitori** che rispettano tutti i criteri e posso trovare i **File Film**. Supponiamo siano 8 distribuiti in 3 **Server CDN**.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
15	Visualizzazioni	E	505'384	R	Si legge l'attributo <i>TempoFineV</i> per trovare le Visualizzazioni attive che si stimano essere 58, considerando che una Visualizzazione dura in media 1 ora e che le 6'000 Visualizzazioni giornaliere sono equidistribuite sui 13 Server CDN .
16	Fruizione	R	58	R	Si trovano le chiavi dei File Film che le Visualizzazioni stanno visionando, che si stimano essere 40
17	Supporto	R	40	R	Tramite Supporto si trovano le chiavi dei Contentori dei vari File Film .
18	Contentore	R	10	R	Si legge il <i>Bitrate</i> ¹³ dei 10 Contentori trovati in precedenza ¹⁴
19	Connessione	E	1'825'000	R	Si controlla se la Connessione , che fa la Visualizzazione , è già presente nella base di dati
20	Connessione	E	1	W	Se non presente si scrive la Connessione
21	Visualizzazione	E	1	W	Si scrive la nuova Visualizzazione
22	Azione	R	1	W	Si scrive la coppia Connessione - Visualizzazione
23	Fruizione	R	1	W	Si scrive la coppia Visualizzazione - File Film

¹³Si suppone che un File Film in SD abbia un bitrate di 1 Mbps, uno in 720p 3 Mbps, uno in 1080p 5 Mbps, uno in 4k 15 Mbps.

¹⁴A questo punto sono disponibili tutti i dati per calcolare I_c , viene quindi scelto il **Server CDN**. Si devono ora aggiungere alla base di dati tutte le informazioni relative alla nuova **Visualizzazione**.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
24	Sessione	R	1	W	Si scrive la coppia Utente - Connessione
25	Presso	R	1	W	Si scrive la coppia Visualizzazione - Server CDN

Tabella 5.10: Tavola degli accessi in assenza di ridondanze: c , rapporto tra la banda utilizzata e la banda libera

Gli accessi per il calcolo di c sono dunque **5'090'569**. Per un totale di $5'090'569 * 6'000 = \mathbf{30'543'414'000}$ accessi giornalieri.

Vediamo gli accessi per il calcolo di d distanza tra **Utente** e **Server CDN**:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Appartenenza Geografica	R	13	R	Si trovano le chiavi degli Stati dei 3 Server CDN in esame
2	Stato	E	3	R	Si vanno a leggere i valori delle coordinate degli Stati

Tabella 5.11: Tavola degli accessi in assenza di ridondanze: d , distanza tra **Utente** e **Server CDN**

Gli accessi per il calcolo di d sono dunque **16** e gli accessi giornalieri sono $16 * 6'000 = \mathbf{96'000}$.

Abbiamo dunque trovato che gli accessi totali per il calcolo di I_c per una **Visualizzazione** sono **Accessi** = $5'090'569 + 16 = \mathbf{5'090'585}$. Per un totale di $5'090'585 * 6'000 = \mathbf{30'543'510'000}$ accessi giornalieri.

5.3.2 Tavola degli Accessi in Presenza di Ridondanza *Banda Disponibile*

Andiamo ad introdurre l'attributo ridondante e analizziamo come cambiano gli accessi. L'attributo che andiamo ad introdurre è *Banda Disponibile* per l'entità **Server**.

Vediamo come cambiano gli accessi per il calcolo di c

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Distribuzione	R	3'000	R	Data la chiave di Film dobbiamo scorrere tutta la relationship per trovare i 15 File Film corrispondenti
2	Destinatario Effettuato	R	120'000	R	Trovo la chiave dei Pagamenti Effettuati dall' Utente che stimiamo essere 3
3	Pagamento Effettuato	E	3	R	Leggo la Data per trovare la chiave dell'ultimo Pagamento Effettuato
4	Saldamento Effettuato	R	1	R	Trovo la chiave di Piano Abbonamento
5	Piano Abbonamento	E	1	R	Si legge l'attributo <i>Definizione</i> , così da scegliere il File Film utilizzabile dall' Utente
6	Stato	E	1	R	Si fa un accesso diretto allo stato da cui proviene la Connessione ¹⁵
7	Indisponibilità Formato	R	150	R	Si trovano i Contenitori non disponibili nello Stato da cui proviene la Connessione
8	Prima Lingua	R	1	R	Si trova la chiave della Prima Lingua per lo Stato
9	Disponibilità Audio	R	9'000	R	Si trovano i File Film del Film richiesto che contengono la Lingua che si stimano essere 15
10	Supporto	R	5	R	Trovo le chiavi dei Contenitori File Film con la Lingua richiesta

¹⁵Ciò è possibile perché di ogni **Connessione** abbiamo memorizzate le coordinate geografiche che sono anche superchiave di **Stato**.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
11	Contenimento Video	R	2	R	Trovo le chiavi dei 2 rispettivi Formati Video . Si stima che 5 Contenitori distinti abbiano 2 Formati Video distinti
12	Formato Video	R	2	R	Leggo l'attributo <i>Risoluzione</i> per verificare che rispetti il vincolo del Piano Abbonamento ¹⁶
13	Disponibilità Immediata	R	9'000	R	Si scorre per trovare i Server CDN che contengono i 8 File Film che rispettano i vincoli di Lingua e di <i>Definizione</i> . Si stima che i File Film siano equamente distribuiti in 3 Server CDN
14	Server CDN	E	8	R	Si legge l'attributo <i>BandaDisponibile</i> dei 8 Server CDN trovati
15	Connessione	E	1'825'000	R	Si controlla se la Connessione , che fa la Visualizzazione , è già presente nella base di dati
16	Connessione	E	1	W	Se non presente si scrive la Connessione
17	Visualizzazione	E	1	W	Si scrive la nuova Visualizzazione
18	Azione	R	1	W	Si scrive la coppia Connessione - Visualizzazione
20	Fruizione	R	1	W	Si scrive la coppia Visualizzazione - File Film
21	Sessione	R	1	W	Si scrive la coppia Utente - Connessione

¹⁶Ho quindi trovato le chiavi dei **Contenitori** che rispetto tutti i criteri e posso trovare i **File Film** che rispettano i criteri. Supponiamo siano 8 distribuiti in 3 **Server CDN**.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
22	Presso	R	1	W	Si scrive la coppia Visualizzazione - Server

Tabella 5.12: Tavola degli accessi in presenza di ridondanza *Banda Disponibile: c*, rapporto tra la banda utilizzata e la banda libera

Gli accessi per il calcolo di c diventano **1'957'218**, per un totale di $1'957'218 * 6'000 = \mathbf{11'743'308'000}$ accessi giornalieri.

Vediamo gli accessi per il calcolo di d :

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Appartenenza Geografica	R	13	R	Si trovano le chiavi degli Stati dei 3 Server CDN in esame
2	Stato	E	3	R	Si vanno a leggere i valori delle coordinate degli Stati

Tabella 5.13: Tavola degli accessi in presenza di ridondanza *Banda Disponibile: d*, distanza tra **Utente** e **Server CDN**

Gli accessi per il calcolo di d rimangono invariati a **16**, per un totale di **96'000**.

Aggiornamento *Banda Disponibile*

Vediamo gli accessi per l'aggiornamento di *Banda Disponibile* che varrà fatto in modalità **Immediate** ogni volta che si ha una nuova **Visualizzazione**, ovvero 6'000 volte al giorno. Per aggiornare la ridondanza si deve modificare *Banda Disponibile* quando una **Visualizzazione** inizia e quando termina.

L'aggiornamento di *Banda Disponibile* quando si ha una nuova **Visualizzazione** si può fare direttamente durante l'esecuzione dell'operazione con i seguenti accessi:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Supporto	R	1	R	Per accedere alla chiave di Contentore del File Film di cui andrò ad usufruire
2	Contentore	E	1	R	Si legge l'attributo <i>Bitrate</i> di Contentore

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
3	ServerCDN	E	1	W	Si va a modificare l'attributo <i>Banda Disponibile</i> del Server con la quale viene instaurata la Connessione

Tabella 5.14: Tavola degli accessi per aggiornamento *Banda Disponibile* all'inizio della **Visualizzazione**

Per aggiornare la *Banda Disponibile* quando termina una **Visualizzazione** servirà un trigger che parte quando si modifica l'attributo *Tempo Fine V* della **Visualizzazione**. Gli accessi che si dovranno fare sono i seguenti:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Fruizione	R	1	R	Trovo il File Film che era utilizzato dalla Visualizzazione
2	Supporto	R	1	R	Per accedere alla chiave di Contenitore del File Film di cui stavo usufruendo
3	Contenitore	E	1	R	Si legge l'attributo <i>bitrate</i> di Contenitore
4	Server CDN	E	1	W	Si va a modificare l'attributo <i>Banda Disponibile</i> del Server con la quale ho instaurata la Connessione

Tabella 5.15: Tavola degli accessi per aggiornamento *Banda Disponibile* al termine della **Visualizzazione**

Gli accessi per mantenere aggiornata la *Banda Disponibile* sono **9**, per un totale di **54'000** accessi giornalieri.

Otteniamo quindi che in presenza di ridondanza gli accessi sono $1'957'218 + 16 + 9 = 1'957'243$. Gli **accessi giornalieri** $= 1'957'243 * 6'000 = 11'743'458'000$. Vediamo dunque che gli accessi sono diminuiti drasticamente e che l'inserimento della ridondanza *Banda Disponibile* è molto vantaggioso.

5.4.1 Tavola degli Accessi

Andiamo a scrivere la tavola degli accessi.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Utente	E	50'000	R	Si controlla se l' Utente è già memorizzato nella base di dati
2	Utente	E	1	W	Si aggiunge in caso non fosse presente
3	Pagamento Effettuato	E	1	W	Si fa una scrittura con le informazioni dell'input
4	Destinatario Effettuato	R	1	W	Si aggiunge la coppia Utente - Pagamento Effettuato
5	Saldamento Effettuato	R	1	W	Si aggiunge la coppia Pagamento Effettuato - Piano Abbonamento
6	Carta di Credito	E	60'000	R	Si controlla se la Carta di Credito appena utilizzata è già presente nella base di dati
7	Carta di Credito	E	1	W	Se non presente si aggiunge
8	Pagamento Predifinito	E	50'000	R	Si controlla se la Carta di Credito appena utilizzata è già memorizzata come Pagamento Predifinito per il dato Utente
9	Pagamento Predifinito	R	1	W	Si aggiorna il Pagamento Predifinito dell' Utente .
10	Addebito	R	1	W	Si scrive la coppia Pagamento Effettuato - Carta Di Credito
11	Pagamento Atteso	E	1	W	Si scrive in Pagamento Atteso le informazioni pertinenti ¹⁷
12	Destinatario Atteso	R	1	W	Si aggiunge la coppia Utente - Pagamento Atteso

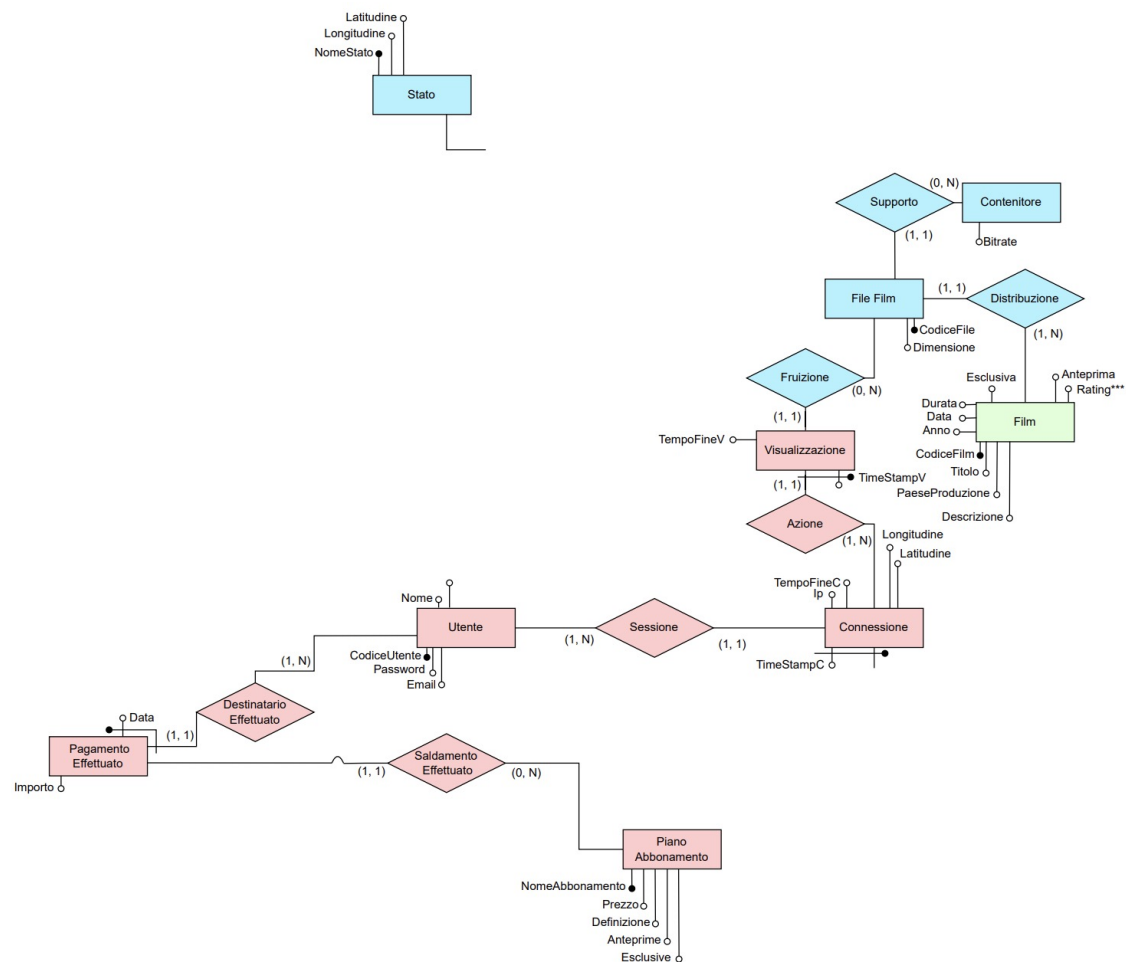
¹⁷Si assume che un **Utente** di default voglia rinnovare il **Piano Abbonamento** attuale.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
13	Saldamento Atteso	R	1	W	Si aggiunge la coppia Pa- gamento Effettuato - Piano Abbonamento

Tabella 5.17: Tavola degli accessi per Acquisto **Piano Abbonamento**

Gli accessi sono quindi $20 + 160'000 = \mathbf{160'020}$, per un totale di $160'020 * 10'000/30 = \mathbf{53'340'000}$ accessi giornalieri.

5.5 Classifiche



Operazione che non ha input che e come output produce delle Classifiche in base ai seguenti criteri:

- Classifica Rating: Si ordinano in maniera decrescente i 10 **Film** in base al loro *Rating*
- Classifica **Film** più visti per **Stato**: Si ordinano in maniera decrescente i 10 **Film** più visti nell'ultima settimana per ogni **Stato**

- Classifica per *Paese Produzione* di un **Film**: Si ordinano in maniera decrescente i 10 **Film** più visti nell'ultima settimana per ogni *Paese di Produzione*¹⁸
- Classifica per **Piano Abbonamento**: Si stila una classifica dei 10 **Film** più visti da **Utenti** con un dato **Piano Abbonamento**
- Classifica dei **Contenitori** più visti per **Film**: Si fa una classifica dei 10 **Contenitori** dei **File Film** più visualizzati per ogni **Film**

Le classifiche sono ricalcolate con cadenza settimanale. Sono però mostrate ogni volta che un **Utente** effettua una **Connessione** e quindi hanno una frequenza di calcolo di 5'000 volte al giorno.

Vediamo i volumi coinvolti:

Concetto	Tipo	Volume
Visualizzazione	E	2'190'000
Azione	R	2'190'000
Fruizione	R	2'190'000
File Film	E	3'000
Supporto	R	3'000
Contenitore	E	15
Distribuzione	R	3'000
Stato	E	193
Connessione	E	1'825'000
Sessione	R	1'825'000
Utente	E	50'000
Destinatario Effettuato	R	120'000
Saldamento Effettuato	R	120'000
Piano Abbonamento	E	5

Tabella 5.18: Tavola dei Volumi per calcolo Classifiche

5.5.1 Tavola degli Accessi in Assenza di Ridondanze

Andiamo ora a scrivere la tavola degli accessi:

¹⁸Si stima che i diversi *Paesi Produzione* siano 20.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Visualizzazione	E	2'190'000	R	Si trovano le chiavi delle Visualizzazioni dell'ultima settimana, che si stimano essere $42'000 = 6'000 * 7$.
2	Azione	R	35'000	R	Si trovano le Connessioni che hanno effettuato le Visualizzazioni , si stimano essere 35'000
3	Sessione	R	35'000	R	Si trovano gli Utenti che hanno fatto le Connessioni che si stimano essere 10'000 (ovvero tutti gli Utenti attivi)
4	Destinatario Effettuato	R	120'000	R	Si trovano i Pagamenti Effettuati che si stimano essere 3 per ogni Utente 30'000
5	Pagamento Effettuato	E	30'000	R	Si legge l'attributo <i>Data</i> per trovare l'ultimo Pagamento Effettuato
6	Saldamento Effettuato	R	10'000	R	Si trovano i Piani Abbonamento degli Utenti attivi
7	Stato	E	30	R	Si leggono gli Stati da cui provengono le Connessioni
8	Fruizione	R	3'000	R	Si trovano i File Film che sono visualizzati, si stimano essere 3'000
9	Supporto	R	15	R	Si trovano le chiavi dei Contenitori dei File Film che sono 15
10	Distribuzione	R	200	R	Si trovano i Film i cui File Film sono stati visualizzati. Si stimano essere 200
11	Film	E	200	R	Si leggono le informazioni relative al Film come il <i>Rating</i> e il <i>Paese Produzione</i>

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
--------	----------	------	---------	-------	-------------

Tabella 5.19: Tavola degli accessi in assenza di ridondanze per Classifiche

Gli accessi sono dunque **2'393'641** per un totale di $2'393'641 * 5'000 =$
11'968'205'000 accessi giornalieri.

5.5.2 Tavola degli Accessi in Presenza di Ridondanza Classifiche

Andiamo ad introdurre le materialized view classifiche. Saranno 5 materialized view chiamate:

- **Classifica Rating**
- **Classifica Stato**
- **Classifica Paese Produzione**
- **Classifica Piano Abbonamento**
- **Classifica Contenitore**

Le materialized view contengono le informazioni relative alle classifiche. Sono aggiornate con modalità **scheduled** con frequenza settimanale. Vediamone la struttura:

Concetto	Attributi	Chiave	Descrizione
Classifica Rating	CodiceFilm, Posizione, Rating	CodiceFilm	Contiene informazioni sulla classifica relativa ai <i>Rating</i>
Classifica Stato	CodiceFilm, Stato, Posizione, Visualizzazioni	CodiceFilm, Stato	Contiene informazioni sulla classifica relativa ai Film più visti per Stato
Classifica Paese Produzione	CodiceFilm, PaeseProduzione, Posizione, Visualizzazioni	CodiceFilm, PaeseProduzione	Contiene informazioni sulla classifica relativa ai Film più visti per <i>PaeseProduzione</i>
Classifica Piano Abbonamento	CodiceFilm, PianoAbbonamento, Posizione, Visualizzazioni	CodiceFilm, PianoAbbonamento	Contiene informazioni sulla classifica relativa ai Film più visti per Piano Abbonamento

Concetto	Attributi	Chiave	Descrizione
Classifica Contenitore	CodiceFilm, Contenitore, Posizione, Visualizzazioni	Contenitore, CodiceFilm	Contiene informazioni sulla classifica relativa ai Contenitori più visti per Film

Tabella 5.20: Descrizione delle materialized view Classifiche

Vediamo come cambiano gli accessi:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Classifica Stato	MV	300	R	Si hanno 10 Film per 30 Stati
2	Classifica Paese Produzione	MV	200	R	Si hanno 10 Film per 20 <i>PaesiProduzione</i>
3	Classifica Piano Abbonamento	MV	50	R	Si hanno 10 Film per 5 Piani Abbonamento
4	Classifica Contenitori	MV	2'000	R	Si hanno 10 Contenitori per 200 Film

Tabella 5.21: Tavola degli accessi in presenza di ridondanze per Classifiche

Gli accessi per mostrare le classifiche ad un **Utente** sono dunque **2'560**. Gli accessi giornalieri sono $2'560 * 5'000 = \mathbf{12'800'000}$.

Vediamo gli accessi per tenere aggiornate le classifiche:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Visualizzazione	E	2'190'000	R	Si trovano le chiavi delle Visualizzazioni dell'ultima settimana, che si stimano essere 42'000 ¹⁹
2	Azione	R	35'000	R	Si trovano le Connessioni che hanno effettuato le Visualizzazioni , si stima essere 35'000
3	Sessione	R	35'000	R	Si trovano gli Utenti che hanno fatto le Connessioni che si stimano essere 10'000 (ovvero tutti gli Utenti attivi)

¹⁹6'000 * 7 = 42'000.

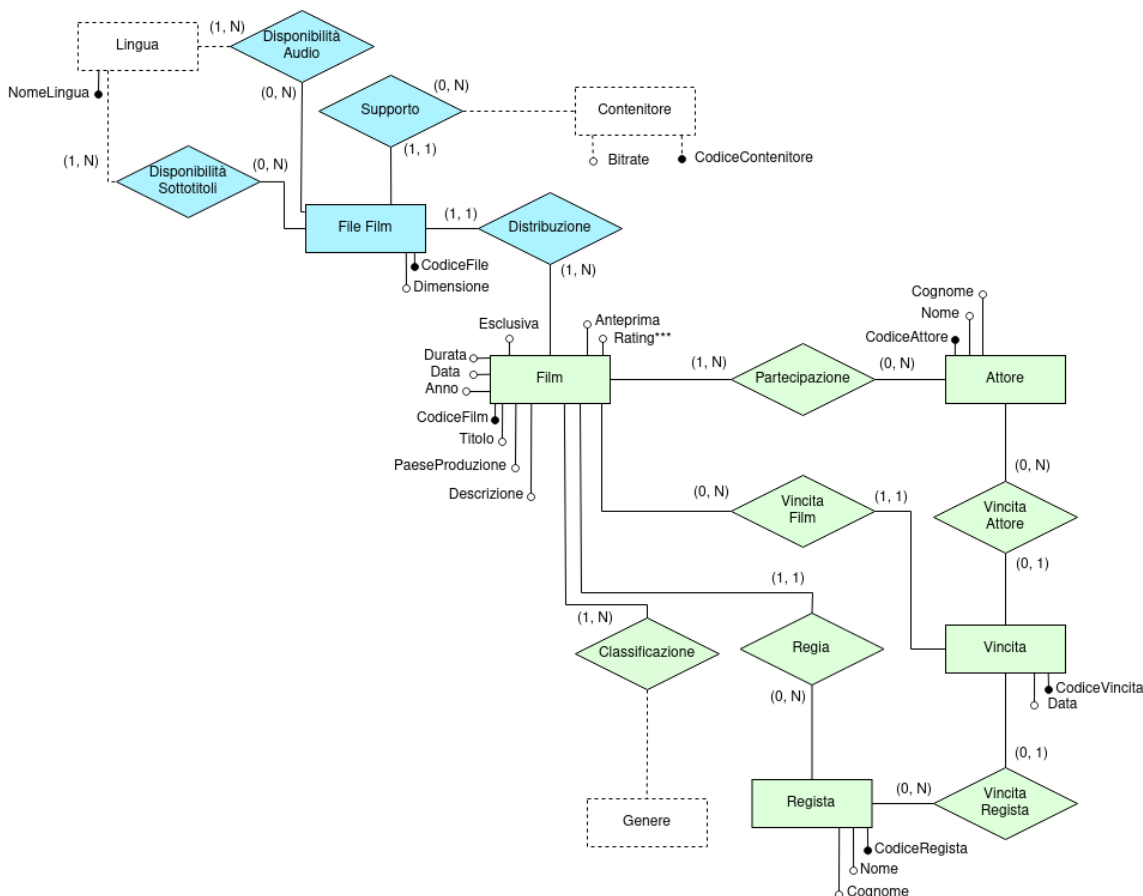
Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
4	Destinatario Effettuato	R	120'000	R	Si trovano i Pagamenti Effettuati che si stimano essere 3 per ogni Utente quindi 30'000
5	Pagamento Effettuato	E	30'000	R	Si legge l'attributo <i>Data</i> per trovare l'ultimo Pagamento Effettuato da ogni Utente
6	Saldamento Effettuato	R	10'000	R	Si trovano i Piani Abbonamento degli Utenti attivi
7	Stato	E	193	R	Si leggono gli Stati da cui provengono le Connessioni
8	Fruizione	R	3'000	R	Si trovano i File Film che sono visualizzati, si stimano essere 3'000
9	Fruizione	R	15	R	Si trovano le chiavi dei Contentori dei File Film che sono 15
10	Distribuzione	R	200	R	Si trovano i Film i cui File Film sono stati visualizzati. Si stimano essere 200
11	Film	E	200	R	Si leggono le informazioni relative al Film come il <i>Rating</i> e il <i>Paese Produzione</i>
12	Classifica Rating	MV	10	W	-
13	Classifica Stato	MV	300	W	Si hanno 10 Film per 30 Stati
14	Classifica Paese Produzione	MV	200	W	Si hanno 10 Film per 20 <i>PaesiProduzione</i>
15	Classifica Piano Abbonamento	MV	50	W	Si hanno 10 Film per 5 Piani Abbonamento
16	Classifica Contentori	MV	2'000	W	Si hanno 10 Contentori per 200 Film

Tabella 5.22: Tavola degli accessi per aggiornamento Classifiche

Gli accessi sono **2'398'761**. Gli accessi giornalieri sono dunque $2'398'761 / 7 =$
342'680.

Gli accessi totali in presenza della ridondanza **Classifiche** sono dunque **13'142'680**. Possiamo affermare che l'introduzione di **Classifiche** è vantaggiosa in quanto riduce significativamente gli accessi giornalieri dell'operazione.

5.6 Inserimento Film, Anteprime e Esclusive



Operazione che prende in input i **Film**, **Genere del Film**, **Vincite del Film**, gli **Attori** e il **Regista del Film** da inserire nella base di dati. L'operazione aggiunge i **Film** nella base di dati con i rispettivi **File Film**, gli **Attori** e il **Regista**²⁰. L'operazione inoltre modifica l'attributo *Anteprima* delle anteprima del mese precedente. L'output dell'operazione è quindi vuoto. La frequenza dell'operazione è mensile²¹.

Vediamo i volumi coinvolti:

²⁰Si suppone che il **Film** possa essere **Classificato** in uno dei **Generi** già presente nella base di dati.

²¹Si stima che si inseriscano 16 nuovi **Film** ogni mese di cui 3 *Anteprime* e 3 *Esclusive*.

Concetto	Tipo	Volume
Film	E	200
Attore	E	1'000
Regista	E	100
Vincita	E	75
Partecipazione	R	3'000
Regia	R	200
Vincita Film	E	75
Vincita Attore	R	50
Vincita Regista	R	5
Classificazione	R	200
Distribuzione	R	3'000
File Film	E	3'000
Supporto	R	3'000
Disponibilità Audio	R	9'000
Disponibilità Sottotitoli	R	9'000

Tabella 5.23: Tavola dei Volumi per Inserimento Film

5.6.1 Tavola degli Accessi

Andiamo a scrivere la tavola degli accessi per questa operazione:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Film	E	19 ²²	W	Scrivo nella base di dati i nuovi contenuti e modifico l'attributo <i>Anteprima</i> sulle <i>Anteprime</i> del mese precedente
2	Attore	E	1'000	R	Leggo se gli Attori sono già presenti nella base di dati

²²Inserisco 16 **Film** e ne modifico altri 3.

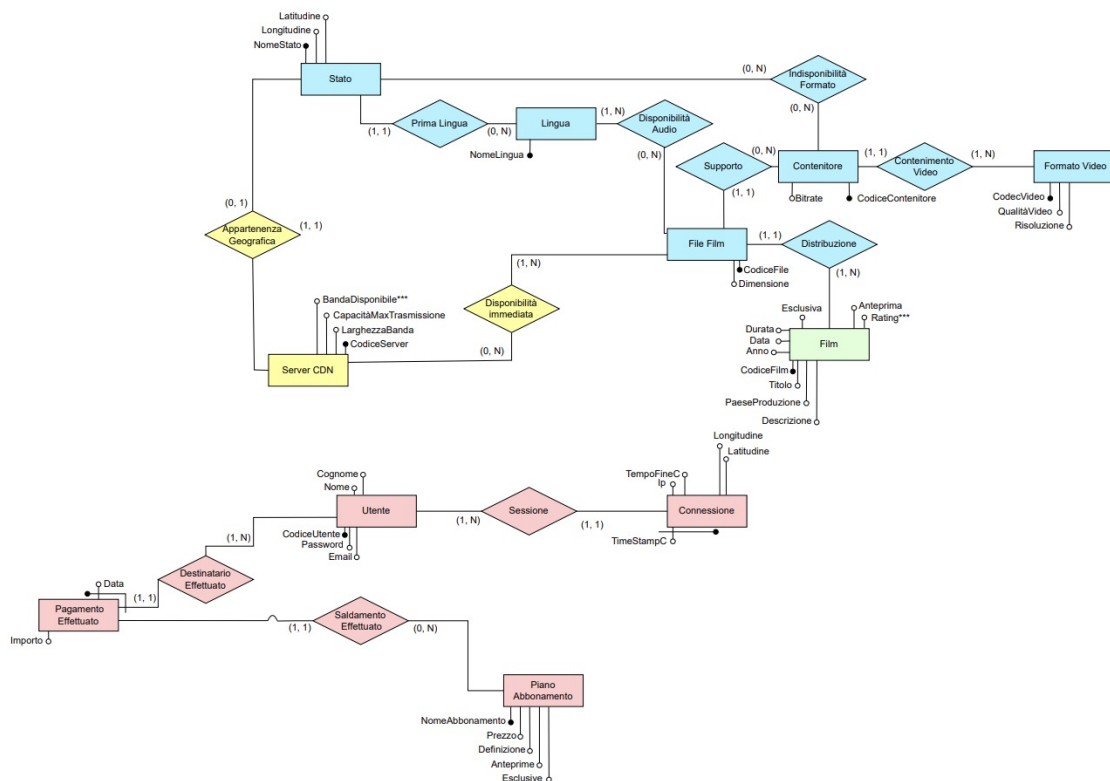
Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
3	Attore	E	151	W	Si memorizzano gli Attori non presenti. Si avrebbero 240 ($= 16 * 15$) Attori distinti ma si stima che il 30% sia già presente nella base di dati, e che i Film abbiano un 10% di Attori in comune. $16 * 15 * 0,70 * 0,9 = 151$
4	Partecipazione	R	240	W	Si scrivono le coppie Film - Attore
5	Regista	E	200	R	Leggo se i Registi sono già presenti nella base di dati
6	Regista	E	14	W	Memorizzo i Registi , si stima che 2 siano già presenti
7	Regia	R	16	W	Si scrivono le coppie Film - Regista
8	Vincita	E	15	W	Memorizzo le vincite dei Film , si stima 0,1 per Film , 0,05 per Attore e per Regista
9	Vincita Film	R	15	W	Si scrivono le coppie Film - Vincita
10	Vincita Attore	R	7	W	Si scrivono le coppie Attore - Vincita
11	Vincita Regista	R	1	W	Si scrive la coppia Regista - Vincita
12	Classificazione	R	1	W	Si scrive la coppia Film - Genere
13	File Film	E	240	W	Si scrivono i File Film dei nuovi Film , $15 * 16 = 240$
14	Distribuzione	R	240	W	Si scrivono le coppie File Film - Film
15	Supporto	R	240	W	Si scrivono le coppie File Film - Contenitore

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
16	Disponibilità Audio	R	720	W	Si scrivono le coppie File Film - Lingua , $240 \times 3 = 720$
17	Disponibilità Sottotitoli	R	720	W	Si scrivono le coppie File Film - Lingua

Tabella 5.24: Tavola dei accessi per Inserimento Film

Gli accessi mensili sono quindi **16'544** e quelli giornalieri sono $16'544 / 30 = \mathbf{552}$.

5.7 Caching



L'operazione ha come input un **Utente**, e output i **File Film** che sono stati memorizzati in ulteriori **Server CDN** e i rispettivi **Server CDN**. I **File Film** vengono spostati cercando di prevedere cosa un **Utente** guarderà. Si cerca infatti di ridurre la distanza tra l'**Utente** e il **Server CDN** che contiene il **File Film** richiesto. La frequenza dell'operazione è settimanale.

Per ogni **Utente** si leggono i suoi **Film** nella materialized view

Raccomandazioni e si spostano i **File Film** relativi al **Server CDN** più vicino all'**Utente**. Si scelgono da memorizzare nel **Server CDN** target solo i **File Film** che, per l'**Utente** in esame, rispettano i vincoli di **Lingua**, di **Risoluzione** e di **Contenitore**. Si evita di memorizzare **File Film** che l'**Utente** non può vedere a causa della **Risoluzione**, nel caso superi quella indicata dal **Piano Abbonamento**,

a causa della **Lingua**, non nota all'**Utente** o a causa del **Contenitore**, nel caso non sia disponibile nello **Stato** di "residenza" dell'**Utente**.

Per determinare il **Server CDN** più vicino all'**Utente** si considera come residenza dell'**Utente** lo **Stato** da cui provengono la maggior parte delle **Connessioni** effettuate dall'**Utente** nell'ultimo mese.

Vediamo i volumi coinvolti:

Concetto	Tipo	Volume
Utente	E	50'000
Sessione	R	1'825'000
Connessione	E	1'825'000
Raccomandazioni	MV	50'000
Destinatario Effettuato	R	120'000
Pagamento Effettuato	E	120'000
Destinatario Effettuato	R	120'000
Piano Abbonamento	E	5
Stato	E	193
Indisponibilità Formato	R	150
Prima Lingua	R	193
Lingua	E	10
Disponibilità Audio	R	9'000
File Film	E	3'000
Supporto	R	3'000
Contenitore	E	15
Contenimento Video	R	15
Formato Video	E	5
Server CDN	E	13
Disponibilità Immediata	R	9'000
Appartenenza Geografica	R	13

Tabella 5.25: Tavola dei Volumi per operazione Caching

5.7.1 Tavola degli Accessi

Scriviamo ora gli accessi dell'operazione:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Raccomandazioni	MV	50'000	R	Si leggono i Film raccomandati all' Utente nell'ultima settimana
2	Distribuzione	R	3'000	R	Si trovano tutti i File Film per i 5 Film raccomandati che sono 75

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
3	Destinatario Effettuato	R	120'000	R	Trovo la chiave dei Pagamenti Effettuati dall' Utente che stimiamo essere 3
4	Pagamento Effettuato	E	3	R	Leggo la Data per trovare la chiave dell'ultimo Pagamento Effettuato
5	Saldamento Effettuato	R	1	R	Trovo la chiave di Piano Abbonamento
6	Piano Abbonamento	E	1	R	Si legge l'attributo <i>Definizione</i> , così da scegliere il File Film utilizzabile dall' Utente
7	Sessione	R	1'825'000	R	Si trovano le Connessioni effettuate dell' Utente che sono 45
8	Connessione	E	45	R	Si trovano le coordinate da cui provenogno la maggior parte delle Connessioni dell'ultimo mese
9	Stato	E	1	R	Si fa un accesso diretto allo Stato grazie alle coordinate trovate
10	Indisponibilità Formato	R	150	R	Si trovano i Contentori non disponibili nello Stato di "residenza"
11	Prima Lingua	R	1	R	Si trova la chiave della Prima Lingua per lo Stato
12	Disponibilità Audio	R	9'000	R	Si trovano i File Film dei Film raccomandati che contengono la Lingua richiesta che si stimano essere 40, ovvero 8 per Film
13	Supporto	R	40	R	Trovo le chiavi dei Contentori dei File Film con la Lingua richiesta che si stimano essere 12

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
14	Contenimento Video	R	4	R	Trovo le chiavi dei 4 rispettivi Formati Video . Si stima che 12 Contenitori distinti abbiano 4 Formati Video distinti
15	Formato Video	R	4	R	Leggo l'attributo <i>Risoluzione</i> per verificare che rispetti il vincolo del Piano Abbonamento . Ho quindi trovato le chiavi dei Contenitori che rispetto tutti i criteri che si stimano essere 10. Tra i 75 File Film iniziali si stima che siano usufruibili dall' Utente 25 File Film ovvero 5 per Film
16	Server CDN	E	13	R	Si trovano gli Stati dei vari Server CDN , per poi trovare quello più vicino allo Stato di "residenza"
17	Appartenenza Geografica	R	13	R	Si trovano gli Stati dei vari Server CDN
18	Disponibilità Immediata	R	25	W	Si scrivono le nuove copie File Film - Server CDN

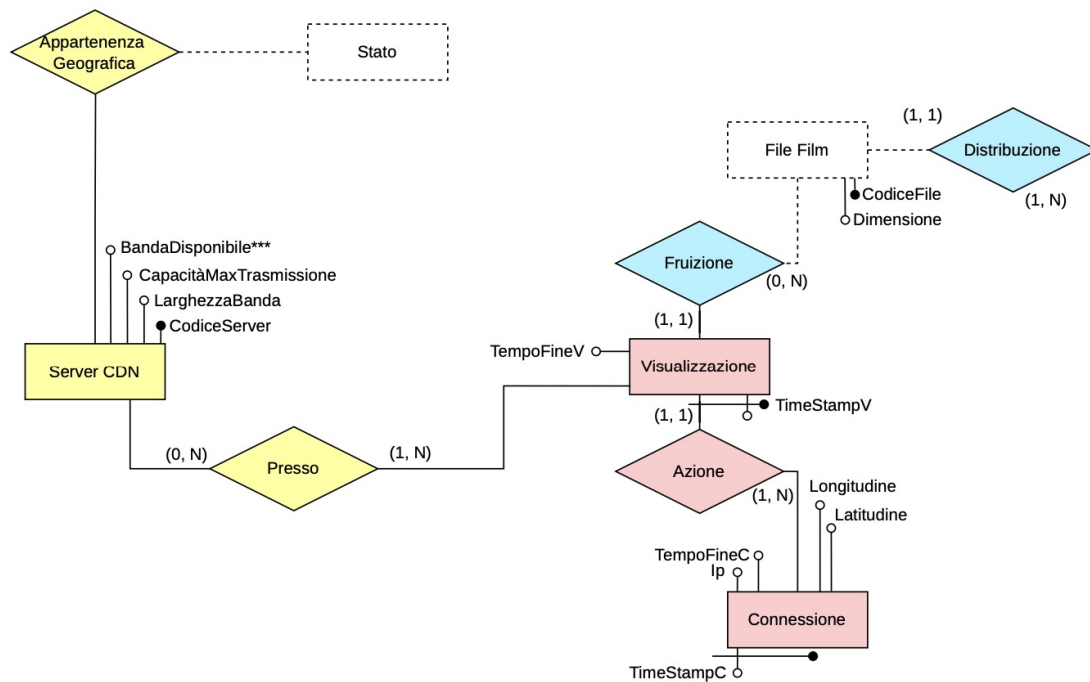
Tabella 5.26: Tavola degli accessi per Caching

Gli accessi per il Caching dei **File Film** per un **Utente** ammontano a 2'007'326.

Gli accessi totali giornalieri ammontano quindi a $(2'007'326 * 10'000)/7 =$

2'867'608'571.

5.8 Bilanciamento Del Carico



Operazione che ha un input vuoto, e che in output produce dei piani di allocazione dei **File Film** in diversi **Server CDN**. La frequenza dell'operazione è settimanale. L'operazione si divide in due fasi:

1. Fase di stima della *Larghezza di Banda* utilizzata mediamente da ogni **Server CDN**
2. Fase di scelta dei **Server CDN** destinazione in cui spostare i **File Film** e dei relativi **Server CDN** sorgente

La media del carico dei vari **Server CDN** è utilizzata per determinare i **Server CDN** sorgente, ovvero quelli che inviano i **File Film**, e i **Server CDN** destinazione, quelli che ricevono i **File Film**. Vengono etichettati come **Server CDN** sorgente i 5 **Server CDN** con carico medio maggiore, e come **Server CDN** destinazione i 5 **Server CDN** con carico medio minore.²³

Criteri di spostamento di un **File Film** da un **Server CDN** sorgente S_s a un **Server CDN** destinazione S_d :

- Se un **File Film** è molto richiesto²⁴ in un S_s , si contrallano le **Connessioni** che effettuato queste **Visualizzazioni**, se le **Connessioni** sono effettuate da **Stati** il cui **Server CDN** più vicino non è il S_s in esame, allora si sposta il **File Film** nel **Server CDN** più vicino allo **Stato** con maggiore richiesta.

²³I 7 **Server CDN** restanti verranno considerati già bilanciati.

²⁴Per **File Film** molto richiesto si intende che rientra nella top 3 dei **File Film** più richiesti nell'ultima settimana nel S_s in esame.

- Se un **File Film** collocato in un S_s è la distribuzione di un **Film** che compare in **Classifica Stato** e lo **Stato** corrispondente non ha come **Server CDN** più vicino il S_s in esame, si spostano 80% dei **File Film** del **Film** nel **Server CDN** più vicino allo **Stato**.

Vediamo i volumi coinvolti:

Concetto	Tipo	Volume
Server	E	13
Visualizzazione	E	2'190'000
Appartenenza Geografica	R	13
Fruizione	R	2'190'000
Azione	R	2'190'000
Connessione	E	1'825'000
Presso	R	2'628'000
Distribuzione	R	3'000
Classifica Stato	MV	9'000

Tabella 5.27: Tavola dei Volumi per Bilanciamento del Carico

5.8.1 Tavola degli Accessi della Fase 1

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Presso	R	2'628'000	R	Si scorre la relazione Presso per trovare le chiavi di Visualizzazione relative ai contenuti che ogni Server CDN stava erogando nell'ultima settimana, si stimano essere $6'000 * 7 = 42'000$, ottenendo l'attributo <i>TimeStampV</i> delle Visualizzazioni in esame e suddividendole in 168 intervalli t_i da 1 ora. ($24 * 7 = 168$)
2	Fruizione	R	42'000	R	Si scorre la relazione Fruizione per trovare i File Film che le Visualizzazioni stavano utilizzando per ogni t_i che si stimano essere 3'000

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
3	Supporto	R	3'000	R	Si trovano i Contenitori dei File Film erogati dal Server CDN per ogni istante t_i
4	Contenitore	R	15	R	Si legge attributo <i>Bitrate</i> dei diversi Contenitori dei File Film che si stimano essere 15

Tabella 5.28: Tavola degli accessi per Bilanciamento del Carico Fase 1

Questa fase dell'operazione calcola per ogni **Server CDN** la media della banda occupata in ogni istante dell'ultima settimana. Gli accessi per questa fase sono dunque **2'673'015**.

5.8.2 Tavola degli Accessi della Fase 2

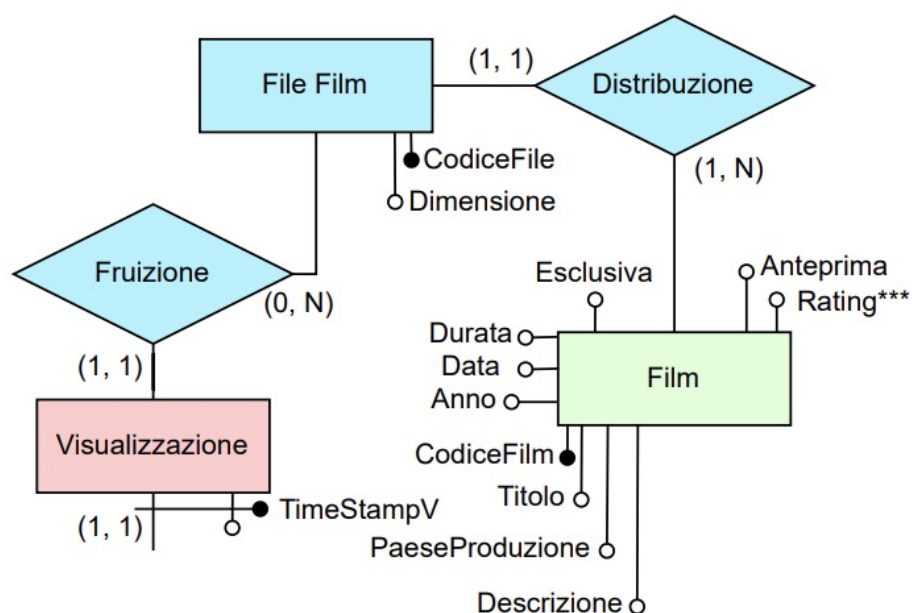
Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Distribuzione	R	3'000	R	Si scorre la relationship Distribuzione per trovare <i>CodiceFilm</i> dei File Film molto richiesti già ottenuti dagli accessi della fase 1 che si stimano essere 225, per un totale di 15 Film
2	Connessione	E	188	R	Si leggono la <i>Latitudine</i> e la <i>Longitudine</i> delle Connessioni richiedenti i File Film molto richiesti per localizzarle. Si hanno 225 Visualizzazioni che guardano File Film molto richiesti, e le conosciamo dalla fase 1. $225/1,2 = 187,5$
3	Appartenenza Geografica	R	13	R	Si trovano gli stati dei Server CDN
4	Stato	E	13	R	Si leggono gli Stati dei Server CDN per trovarne le coordinate

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
5	Classifica Stato	MV	9'000	R	Si verifica se i Film nei S_s individuati appartengono alle Classifiche Stato

Tabella 5.29: Tavola degli accessi per Bilanciamento del Carico Fase 2

Gli accessi per questa seconda fase sono **12'214**. La somma tra gli accessi delle due fasi dell'operazione ammonta dunque a: $2'673'015 + 12'214 = 2'685'229$, per un totale di $2'685'229/7 = \mathbf{383'604}$ accessi giornalieri.

5.9 Custom Analytics - Analisi Visualizzazioni



Questa operazione offre un pacchetto di informazioni che compone un'analisi delle **Visualizzazioni** effettuate dagli **Utenti** nell'ultimo trimestre. La frequenza è quindi trimestrale. L'input dell'operazione è vuoto e l'output è il seguente:

- Orario del giorno con maggiore e minore **Visualizzazioni** nel trimestre
- Giorno della settimana con maggiore e minore **Visualizzazioni** nel trimestre

- Mese con maggiore e minore **Visualizzazioni** nel trimestre
- 20 Film con maggiore e minore ritenzione²⁵ nel trimestre

Possiamo quindi individuare 2 fasi dell'operazione:

1. Calcolo orario, giorno, mese con maggiore e minore numero di **Visualizzazioni**
2. Calcolo dei 20 **Film** con maggiore e minore ritenzione del trimestre

Andiamo a vedere i volumi coinvolti:

Concetto	Tipo	Volume
Visualizzazione	E	2'190'000
Fruizione	R	2'190'000
File Film	E	3'000
Distribuzione	R	3'000
Film	E	200

Tabella 5.30: Tavola dei Volumi per Analisi delle Visualizzazioni

5.9.1 Chiarimenti sulla Ritenzione

Prima di vedere la tavola degli accessi per l'operazione andiamo a dare dei chiarimenti sulla definizione di ritenzione.

Abbiamo definito la ritenzione r come la percentuale vista di un **Film**. Chiamiamo T_v il tempo visto di un **Film**, e D la sua durata.

Il tempo visto di un **Film** è calcolato nel seguente modo:

$$T_v = (TempoFineV - TimeStampV) * 0,95$$

Si moltiplica per 0,95 per tenere conto del fatto che un **Utente** in media mette in pausa la riproduzione di un **Film** per il 5% della durata della **Visualizzazione**.

Possiamo quindi dare la definizione di ritenzione r :

$$r = \frac{T_v}{D} * 100$$

5.9.2 Tavola degli Accessi Fase 1

Vediamo gli accessi per la fase 1:

²⁵Si definisce la ritenzione come la percentuale vista di un **Film**.

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Visualizzazione	E	2'190'000	R	Si leggono gli attributi, in particolare <i>TimeStampV</i> per verificare che la Visualizzazione sia stata effettuata nell'ultimo trimestre e conoscerne l'ora, il giorno della settimana, e il mese

Tabella 5.31: Tavola degli accessi per Analisi delle Visualizzazioni Fase 1

Gli accessi per questa fase sono quindi **2'190'000**

5.9.3 Tavola degli Accessi Fase 2

Andiamo ora a vedere gli accessi per la seconda fase:

Indice	Concetto	Tipo	Accessi	TipoA	Descrizione
1	Visualizzazione	E	2'190'000	R	Si leggono gli attributi <i>TimeStampV</i> e <i>TempoFineV</i> per conoscere la durata della Visualizzazione . Si stima di trovare 540'000 Visualizzazioni che sono avvenute nell'ultimo trimestre
2	Fruizione	R	540'000	R	Si trovano i File Film visualizzati, che si stima siano 3'000
3	Distribuzione	R	3'000	R	Si trovano i Film dei File Film visualizzati che si stimano essere 200
4	Film	E	200	R	Si accede ai Film per leggere l'attributo <i>Durata</i>

Tabella 5.32: Tavola degli accessi per Analisi delle Visualizzazioni Fase 2

Gli accessi per la seconda fase sono **2'733'200**.

Gli accessi totali per l'operazione sono **4'923'200** e gli accessi giornalieri sono **54'702**.

Capitolo 6

Progettazione Logica

6.1 Traduzione dello Schema Concettuale nel Modello Logico Relazionale

Le tabelle riportate a seguito sono il risultato dell'esecuzione dell'algoritmo di traduzione.

6.1.1 Tabelle dell'Area Contenuti

- **Film**(CodiceFilm, Titolo, Anno, PaeseProduzione, Descrizione, Durata, Esclusiva, Anteprima, Rating, Regista)
- **Attore**(CodiceAttore, Nome, Cognome)
- **Regista**(CodiceRegista, Nome, Cognome)
- **Genere**(NomeGenere)
- **Recensione**(CodiceFilm, CodiceUtente, Valutazione, Descrizione, Upvote, Downvote)
- **Vincita**(CodiceVincita, Data, CodiceFilm, NomePremio, Categoria)
- **Premio**(NomePremio, Categoria)
- **VincitaAttore**(CodiceVincita, CodiceAttore)
- **VincitaRegista**(CodiceVincita, CodiceRegista)
- **Partecipazione**(CodiceFilm, CodiceAttore)
- **Classificazione**(NomeGenere, CodiceFilm)

6.1.2 Tabelle dell'Area Formati

- **FileFilm**(CodiceFile, CodiceFilm, Dimensione, CodiceContenitore)
- **Contenitore**(CodiceContenitore, CodecAudio, CodecVideo)
- **FormatoVideo**(CodecVideo, QualitàVideo, Risoluzione)
- **FormatoAudio**(CodecAudio, QualitàAudio)
- **Lingua**(NomeLingua)
- **Stato**(NomeStato, Longitudine, Latitudine, PrimaLingua)
- **DisponibilitàSottotitoli**(CodiceFile, NomeLingua)
- **DisponibilitàAudio**(CodiceFile, NomeLingua)
- **IndisponibilitàFormato**(CodiceContenitore, NomeStato)

6.1.3 Tabelle dell'Area Clienti

- **Utente**(CodiceUtente, Nome, Cognome, Email, Password, PagamentoPredefinito)
- **Dispositivo**(CodiceDispositivo, Categoria, Risoluzione, Rapporto Aspetto)
- **CartaDiCredito**(NumeroCarta, AnnoScadenza, MeseScadenza, Tipo, CVV, Circuito)
- **PagamentoEffettuato**(CodiceUtente, Data, Importo, AddebitoCarta, PianoAbbonamento)
- **PagamentoAtteso**(CodiceUtente, Data, Importo, PianoAbbonamento)
- **PianoAbbonamento**(NomeAbbonamento, Prezzo, Definizione, Anteprime, Esclusive)
- **Connessione**(TimeStampC, CodiceDispositivo, CodiceUtente, Ip, TempoFineC, Longitudine, Latitudine)
- **Visualizzazione**(TimeStampV, TimeStampC, CodiceDispositivo, TempoFineV, CodiceFile)

6.1.4 Tabelle dell'Area Streaming

- **ServerCDN**(CodiceServer, CapacitaMassimaTrasmissione, LarghezzaBanda, BandaDisponibile, NomeStato)
- **Presso**(CodiceServer, TimeStampV, TimeStampC, CodiceDispositivo)
- **DisponibilitàImmediata**(CodiceServer, CodiceFile)

6.2 Vincoli

In questa sezione andremmo a scrivere i vincoli più significativi, che comprendono i vincoli di integrità referenziale e i vincoli di tupla.

6.2.1 Vincoli di Integrità Referenziale

Andiamo a richiamare velocemente la definizione di vincolo di integrità referenziale. Un vincolo di integrità referenziale fra gli attributi X (anche più di uno) di una relazione R_1 e un'altra relazione R_2 impone ai valori su X in R_1 di comparire come valori della chiave primaria di R_2 .

Seguono i vincoli di integrità referenziale della base di dati, divisi per area.

Area Contenuti

- Vincolo di integrità referenziale tra l'attributo *CodiceFilm* di **Partecipazione** e la relazione¹ **Film**
- Vincolo di integrità referenziale tra l'attributo *CodiceAttore* di **Partecipazione** e la relazione **Attore**
- Vincolo di integrità referenziale tra l'attributo *CodiceFilm* di **Vincita** e la relazione **Film**
- Vincolo di integrità referenziale tra l'attributo *CodiceRegista* di **Film** e la relazione **Regista**
- Vincolo di integrità referenziale tra l'attributo *CodiceFilm* di **Classificazione** e la relazione **Film**
- Vincolo di integrità referenziale tra l'attributo *CodiceFilm* di **Recensione** e la relazione **Film**
- Vincolo di integrità referenziale tra l'attributo *CodiceUtente* di **Recensione** e la relazione **Utente**
- Vincolo di integrità referenziale tra l'attributo *CodiceAttore* di **VincitaAttore** e la relazione **Attore**
- Vincolo di integrità referenziale tra l'attributo *CodiceVincita* di **VincitaAttore** e la relazione **Vincita**
- Vincolo di integrità referenziale tra l'attributo *CodiceRegista* di **VincitaRegista** e la relazione **Regista**

¹In questa sezione si andranno a chiamare relazioni le tabelle prodotte dalla traduzione, da non confondere con le relationship dello schema E-R.

- Vincolo di integrità referenziale tra l'attributo *CodiceVincita* di **VincitaRegista** e la relazione **Vincita**
- Vincolo di integrità referenziale tra l'attributo *NomePremio* di **Vincita** e la relazione **Premio**

Area Formati

- Vincolo di integrità referenziale tra l'attributo *CodiceFilm* di **FileFilm** e la relazione **Film**
- Vincolo di integrità referenziale tra l'attributo *CodiceContenitore* di **FileFilm** e la relazione **Contenitore**
- Vincolo di integrità referenziale tra l'attributo *CodiceFile* di **Visualizzazione** e la relazione **FileFilm**
- Vincolo di integrità referenziale tra l'attributo *CodiceFile* di **DisponibilitàAudio** e la relazione **FileFilm**
- Vincolo di integrità referenziale tra l'attributo *CodiceFile* di **DisponibilitàSottotitoli** e la relazione **FileFilm**
- Vincolo di integrità referenziale tra l'attributo *CodiceFile* di **DisponibilitàImmediata** e la relazione **FileFilm**
- Vincolo di integrità referenziale tra l'attributo *NomeLingua* di **DisponibilitàAudio** e la relazione **Lingua**
- Vincolo di integrità referenziale tra l'attributo *NomeLingua* di **DisponibilitàSottotitoli** e la relazione **Lingua**
- Vincolo di integrità referenziale tra l'attributo *PrimaLingua* di **Stato** e la relazione **Lingua**
- Vincolo di integrità referenziale tra l'attributo *CodiceContenitore* di **IndisponibilitàFormato** e la relazione **Contenitore**
- Vincolo di integrità referenziale tra l'attributo *NomeStato* di **IndisponibilitàFormato** e la relazione **Stato**
- Vincolo di integrità referenziale tra l'attributo *CodecVideo* di **Contenitore** e la relazione **FormatoVideo**
- Vincolo di integrità referenziale tra l'attributo *CodecAudio* di **Contenitore** e la relazione **FormatoAudio**

Area Clienti

- Vincolo di integrità referenziale tra gli attributi *TimeStampV*, *TimeStampC*, *CodiceDispositivo* di **Presso** e la relazione **Visualizzazione**
- Vincolo di integrità referenziale tra gli attributi *TimeStampC*, *CodiceDispositivo* di **Visualizzazione** e la relazione **Connessione**
- Vincolo di integrità referenziale tra l'attributo *CodiceDispositivo* di **Connessione** e la relazione **Dispositivo**
- Vincolo di integrità referenziale tra l'attributo *CodiceUtente* di **Connessione** e la relazione **Utente**
- Vincolo di integrità referenziale tra l'attributo *PagamentoPredefinito* di **Utente** e la relazione **CartaDiCredito**
- Vincolo di integrità referenziale tra l'attributo *CodiceUtente* di **PagamentoEffettuato** e la relazione **Utente**
- Vincolo di integrità referenziale tra l'attributo *PianoAbbonamento* di **PagamentoEffettuato** e la relazione **PianoAbbonamento**
- Vincolo di integrità referenziale tra l'attributo *AddebitoCarta* di **PagamentoEffettuato** e la relazione **CartaDiCredito**
- Vincolo di integrità referenziale tra l'attributo *CodiceUtente* di **PagamentoAtteso** e la relazione **Utente**
- Vincolo di integrità referenziale tra l'attributo *PianoAbbonamento* di **PagamentoAtteso** e la relazione **PianoAbbonamento**

Area Streaming

- Vincolo di integrità referenziale tra l'attributo *CodiceServer* di **Presso** e la relazione **ServerCDN**
- Vincolo di integrità referenziale tra l'attributo *CodiceServer* di **DisponibilitàImmediata** e la relazione **ServerCDN**
- Vincolo di integrità referenziale tra l'attributo *NomeStato* di **ServerCDN** e la relazione **Stato**

6.2.2 Vincoli di Tupla

Andiamo a scrivere i vincoli di tupla più significativi per ogni relazione:

Film

- $Anno \leq YEAR(CURRENT_DATE())$

Recensione

- $Valutazione \geq 0 \text{ AND } Valutazione \leq 10$

Vincita

- $Data \leq CURRENT_DATE()$

FileFilm

- $Dimensione > 0$

FormatoVideo

- $QualitàVideo > 0 \text{ AND } QualitàVideo \leq 5$

FormatoAudio

- $QualitàAudio > 0 \text{ AND } QualitàAudio \leq 5$

Stato

- $Latitudine \geq -90 \text{ AND } Latitudine \leq 90$
- $Longitudine \geq -180 \text{ AND } Longitudine \leq 180$

CartaDiCredito

- $AnnoScadenza \geq 0$
- $MeseScadenza \geq 0 \text{ AND } MeseScadenza \leq 12$

PagamentoEffettuato

- $Data \leq CURRENT_DATE()$

PagamentoAtteso

- $Data > CURRENT_DATE()$

PianoAbbonamento

- $Prezzo \text{ IN } \{5, 10, 12.5, 15, 20\}$

Connessione

- $TimeStampC \leq TempoFineC$
- $Latitudine \geq -90$ AND $Latitudine \leq 90$
- $Longitudine \geq -180$ AND $Longitudine \leq 180$

Visualizzazione

- $TimeStampV \leq TempoFineV$

Capitolo 7

Analisi delle Dipendenze Funzionali e Normalizzazione

Andiamo a scrivere le dipendenze funzionali non banali di tutte le tabelle dello schema logico, e verifichiamo che siano in forma normale di Boyce-Codd. Diamo a seguire la definizione di forma normale di Boyce-Codd.

Uno schema $R(T, F)$ è in **forma normale di Boyce-Codd (BCNF)** se e solo se per ogni dipendenza funzionale non banale $X \rightarrow Y \in F^+$, X è una superchiave di R .

Ricordiamo inoltre che F^+ è la chiusura di F , ed è l'insieme di tutte le dipendenze funzionali implicate da F . Enunciamo infine il teorema che ci permette di verificare che una tabella sia o meno in BCNF.

Uno schema $R(T, F)$ è in **forma normale di Boyce-Codd (BCNF)** se e solo se per ogni dipendenza funzionale non banale $X \rightarrow Y \in F$, X è una superchiave di R .

Andiamo ad analizzare le dipendenze funzionali delle tabelle, scrivendone le dipendenze funzionali non banali, divise per area della base di dati. Se una tabella non contiene dipendenze funzionali non banali verrà omessa in quanto sicuramente in BCNF. Verranno segnate con ✓ le tabelle che rispettano la BCNF.

7.1 Verifica BCNF dell'Area Contenuti

- **Film:** $CodiceFilm \rightarrow Anno, PaeseProduzione, Titolo, Descrizione, Durata, Esclusiva, Anteprima, Rating, Regista$ ✓
- **Attore:** $CodiceAttore \rightarrow Nome, Cognome$ ✓
- **Regista:** $CodiceRegista \rightarrow Nome, Cognome$ ✓

- **Recensione:** $\text{CodiceFilm}, \text{CodiceUtente} \rightarrow \text{Upvote}, \text{Downvote}, \text{Valutazione}, \text{Descrizione} \checkmark$
- **Vincita:** $\text{CodiceVincita} \rightarrow \text{Data}, \text{CodiceFilm}, \text{NomePremio}, \text{Categoria}$ AND $\text{NomePremio}, \text{Categoria}, \text{CodiceFilm} \rightarrow \text{CodiceVincita}, \text{Data}^1 \checkmark$
- **VincitaAttore:** $\text{CodiceVincita} \rightarrow \text{CodiceAttore} \checkmark$
- **VincitaRegista:** $\text{CodiceVincita} \rightarrow \text{CodiceRegista} \checkmark$

7.2 Verifica BCNF dell'Area Formati

- **FileFilm:** $\text{CodiceFile} \rightarrow \text{Dimensione}, \text{CodiceContenitore}, \text{CodiceFilm}$ AND $\text{CodiceFilm}, \text{CodiceContenitore} \rightarrow \text{Dimensione}, \text{CodiceFile}^2 \checkmark$
- **Contenitore:** $\text{CodiceContenitore} \rightarrow \text{CodecAudio}, \text{CodecVideo} \checkmark$
- **FormatoVideo:** $\text{CodecVideo} \rightarrow \text{QualitàVideo}, \text{Risoluzione} \checkmark$
- **FormatoAudio:** $\text{CodecAudio} \rightarrow \text{QualitàAudio}$
- **Stato:** $\text{NomeStato} \rightarrow \text{Longitudine}, \text{Latitudine}, \text{PrimaLingua} \checkmark$

7.3 Verifica BCNF dell'Area Clienti

- **Utente:** $\text{CodiceUtente} \rightarrow \text{Nome}, \text{Cognome}, \text{Email}, \text{Password}, \text{PagamentoPredefinito} \checkmark$
- **CartaDiCredito:** $\text{NumeroCarta} \rightarrow \text{AnnoScadenza}, \text{MeseScadenza}, \text{Tipo}, \text{CVV}, \text{Circuito} \checkmark$
- **PagamentoEffettuato:** $\text{CodiceUtente}, \text{Data} \rightarrow \text{Importo}, \text{PianoAbbonamento}^3 \checkmark$
- **PagamentoAtteso:** $\text{CodiceUtente} \rightarrow \text{Data}, \text{Importo}, \text{PianoAbbonamento}^4 \checkmark$
- **PianoAbbonamento:** $\text{NomeAbbonamento} \rightarrow \text{Prezzo}, \text{Definizione}, \text{Anteprime}, \text{ContenutiEsclusivi} \checkmark$

¹La tabella rispetta la BCNF in quanto $\text{NomePremio}, \text{Categoria}, \text{CodiceFilm}$ rappresenta una superchiave di **Vincita**

²La tabella rispetta la BCNF in quanto $\text{CodiceFilm}, \text{CodiceContenitore}$ è superchiave di **FileFilm**

³Si noti che non esiste la dipendenza funzionale $\text{PianoAbbonamento} \rightarrow \text{Importo}$ in quanto il prezzo dei *PianiAbbonamento* potrebbe cambiare

⁴Si noti che non esiste la dipendenza funzionale $\text{PianoAbbonamento} \rightarrow \text{Importo}$ in quanto il prezzo dei *PianiAbbonamento* potrebbe cambiare

- **Connessione:** $TimeStampC, CodiceDispositivo \rightarrow CodiceUtente, Ip, TempoFineC, Longitudine, Latitudine \checkmark$
- **Visualizzazione:** $TimeStampV, TimeStampC, CodiceDispositivo \rightarrow TempoFineV \checkmark$

7.4 Verifica BCNF dell'Area Streaming

- **ServerCDN:** $CodiceServer \rightarrow CapacitaMassimaTrasmissione, LarghezzaBanda, BandaDisponibile, NomeStato \checkmark$

Capitolo 8

Implementazione su DBMS Oracle MySQL

Tutto il codice è stato implementato su MySQL.

8.1 Organizzazione dei File

Il codice è stato diviso in file come segue:

- Nel file `tabelle.sql` si trova lo script che crea la base di dati con le relative tabelle. In questo file sono anche implementati i vincoli di integrità referenziale e alcuni dei vincoli di tupla come spiegato nella sezione 8.4.
- Nel file `popolamento.sql` si trova lo script che popola la base di dati.
- Nel file `operazioni.sql` si trova lo script che implementa tutte le operazioni, implementate tramite `PROCEDURE`, `TRIGGER` e `EVENT`. Inoltre in questo file vengono anche aggiornate alcune delle ridondanze come spiegato nella sezione 8.3.
- Nel file `vincoli.sql` si trova lo script che implementa i vincoli sulle "date" come spiegato nella sezione 8.4.
- Nel file `ridondanze.sql` si trova lo script che implementa la ridondanza della *BandaDisponibile* come spiegato nella sezione 5.3.2

8.2 Popolamento

La base di dati è stata popolata con un subset variabile di dati generati casualmente.

8.3 Ridondanze

Le seguenti ridondanze: *Rating*, **Raccomandazioni** e **Classifiche** sono aggiornate direttamente dalle operazioni che le calcolano. La ridondanza *BandaDisponibile* è invece aggiornata all'interno delle operazioni quando si ha una nuova **Visualizzazione** ma viene aggiornata con **TRIGGER** quando la **Visualizzazione** termina come spiegato nella sezione 5.3.2.

8.4 Vincoli

8.4.1 Vincoli Integrità Referenziale

Tutti i vincoli di integrità referenziale sono stati implementati con il costrutto `FOREIGN KEY (...) REFERENCES ... (...)`.

8.4.2 Vincoli Di Tupla

Alcuni vincoli di tupla sono stati implementati con il costrutto `CHECK('... ' > ...)` e si trovano nel file `tabelle.sql`, mentre altri, ovvero quelli che hanno a che fare con le date, sono stati implementati con `TRIGGER BEFORE INSERT ON ...` e si trovano nel file `vincoli.sql`.

Bibliografia

- Slide del corso di Basi di Dati, Corso di Laurea in Ingegneria Informatica - A.A. 2022-23, Università di Pisa
- Basi di Dati - VI edizione. Paolo Atzeni, Stefano Ceri, Piero Fraternali, Stefano Paraboschi, Riccardo Torlone McGraw-Hill Education (Italy)
- Sito Salvatore Aranzulla usato per la stima dei *Bitrate* utilizzati da un **Film**, link: <https://www.aranzulla.it/come-funziona-netflix-2-1018542.html>
- Sito utilizzato per il dato degli utenti di Netflix, link: <https://www.demandsage.com/netflix-subscribers.html>
- Sito utilizzato per il dato dei contenuti di Netflix, link: <https://www.comparitech.com/blog/vpn-privacy/netflix-statistics-facts-figures.html>
- Cartina degli indirizzi Ip, link: https://en.m.wikipedia.org/wiki/List_of_countries_by_IPv4_address_allocation