



## Progetto Basi di Dati

Simone Gallo, Riccardo Ciucci

# Indice

<b>1</b>	<b>Introduzione</b>	<b>5</b>
<b>2</b>	<b>Progettazione Concettuale</b>	<b>6</b>
2.1	Dizionario dell'Entità . . . . .	6
2.2	Dizionario delle Relazioni . . . . .	10
2.3	Regole di Vincolo . . . . .	14
2.4	Regole di Derivazione . . . . .	14
2.5	Descrizione ER . . . . .	15
2.5.1	Entità Critico e Utente . . . . .	16
2.5.2	Vincita e Premio . . . . .	16
2.5.3	Edizione . . . . .	16
2.5.4	File . . . . .	17
2.5.5	FormatoCodifica . . . . .	17
2.5.6	Qualità di VideoFile . . . . .	17
2.5.7	Qualità di AudioFile . . . . .	17
2.5.8	IP Range . . . . .	17
2.5.9	Server . . . . .	19
2.5.10	Abbonamento . . . . .	19
2.5.11	Popolarità di Artista . . . . .	20
<b>3</b>	<b>Ristrutturazione</b>	<b>20</b>
3.1	Eliminazione delle Generalizzazioni . . . . .	20
3.1.1	Artista . . . . .	20
3.1.2	Recensore . . . . .	21
3.1.3	FormatoCodifica . . . . .	21
3.2	Eliminazione degli Attributi Multivalore . . . . .	23
3.2.1	LinguaDoppiaggio . . . . .	23
3.2.2	LinguaSottotitoli . . . . .	23
3.3	Analisi Ridondanze . . . . .	23
3.3.1	Ridondanze presenti . . . . .	23
3.3.2	Introduzione nuove Ridondanze . . . . .	23
3.4	Partizionamento e Accorpamento ER . . . . .	24
3.4.1	Qualità di VideoFile . . . . .	24
3.4.2	Qualità di AudioFile . . . . .	24
3.4.3	Posizione . . . . .	24
3.4.4	Carta di Credito . . . . .	25

<b>4</b>	<b>Tavole dei Volumi</b>	<b>26</b>
4.1	Area Contenuti . . . . .	26
4.2	Area Formato . . . . .	27
4.3	Area Clienti . . . . .	28
4.4	Area Streaming . . . . .	29
<b>5</b>	<b>Operazioni sui Dati</b>	<b>29</b>
5.1	Analisi delle Ridondanze . . . . .	29
5.1.1	CaricoAttuale . . . . .	29
5.1.2	DistanzaPrecalcolata . . . . .	32
5.1.3	MediaRecensioni NumeroRecensioni . . . . .	34
5.2	Analisi delle Operazioni . . . . .	36
5.2.1	Vincite di un Film . . . . .	36
5.2.2	Generi di un Film . . . . .	37
5.2.3	Massima Risoluzione di un Film . . . . .	38
5.2.4	Numero di Film non disponibili in un Abbonamento . . . . .	39
5.2.5	Film disponibili in Lingua specifica . . . . .	41
5.2.6	Film più visti recentemente . . . . .	42
5.2.7	Cambio Abbonamento . . . . .	43
5.2.8	Film con Recensioni migliori . . . . .	45
<b>6</b>	<b>Progettazione Logica</b>	<b>47</b>
6.1	Ridenominazione Nomi . . . . .	47
6.2	Traduzione in schema Logico . . . . .	47
6.3	Vincoli . . . . .	51
6.3.1	Vincoli di Integrità Referenziale . . . . .	51
6.3.2	Vincoli di Tupla . . . . .	52
6.3.3	Vincoli di Dominio . . . . .	53
6.3.4	Vincoli Inter-Relazionali . . . . .	53
6.4	Analisi delle Dipendenze Funzionali e Normalizzazione . . . . .	54
<b>7</b>	<b>Raccomandazione Utente</b>	<b>56</b>
7.1	Rating . . . . .	56
7.1.1	Rating del Film . . . . .	57
7.1.2	Rating per Utente . . . . .	58
7.1.3	Sistema di Raccomandazione . . . . .	58
<b>8</b>	<b>Area Analytics</b>	<b>59</b>
8.1	Classifiche . . . . .	59
8.2	Custom Analytics . . . . .	60
8.3	Bilanciamento carico . . . . .	60

<b>9</b>	<b>Area Streaming</b>	<b>61</b>
9.1	Individuazione Server . . . . .	61
9.2	Ribilanciamento del carico . . . . .	64
9.3	Caching previsionale . . . . .	65
<b>10</b>	<b>Implementazione Fisica</b>	<b>66</b>
10.1	Generazione delle Tabelle . . . . .	66
10.2	Creazione dei Vincoli . . . . .	66
10.3	API . . . . .	66
10.4	Popolamento . . . . .	66

# 1 Introduzione

Si desidera progettare e realizzare una base di dati che permetta sia la memorizzazione di dati ed informazioni relativi a FilmSphere, un servizio internazionale di streaming di contenuti video online, che l'utilizzo di operazioni e funzioni di data analytics che permettono di migliorare l'esperienza dell'utente.

Il gestore del sistema consente l'inserimento di Film e le informazioni ad essi correlate, la memorizzazione dei vari formati utilizzati insieme alle relative versioni di questi ultimi ed, infine, l'immagazzinamento di dati che riguardano gli utenti compresi abbonamento, fatture, con eventuali pagamenti effettuati, e raccomandazione di contenuti attraverso algoritmi di valutazione di contenuti.

Oltretutto il suddetto sistema sovrintende e gestisce anche la CDN, ossia la rete di distribuzione del servizio costituita da server dislocati nel globo, grazie all'implementazione di funzionalita' lato server di analisi di dati, come il caching.

Ad ogni componente dello schema ER che riguarda una delle diverse Aree della base di dati viene univocamente associato un colore:

- Area Contenuti
- Area Formati
- Area Clienti
- Area Streaming

## 2 Progettazione Concettuale

### 2.1 Dizionario dell'Entità

Entità	Attributi	Identificatore	Descrizione
<b>Film*</b>	Id, Descrizione, Titolo	ID	Film offerto dal servizio di streaming online
<b>Genere</b>	Nome	Nome	Genere dei Film offerti da FilmSpere
<b>Attore*</b>	Nome, Cognome, Popolarità	Nome, Cognome	Specializzazione di Artista, attore recitante in un Film
<b>Regista*</b>	Nome, Cognome, Popolarità	Nome, Cognome	Specializzazione di Artista, regista dirigente un Film
<b>Artista*</b>	Nome, Cognome, Popolarità	Nome, Cognome	Generalizzazione di Attore e Regista
<b>Premio*</b>	Macrotipo, Microtipo	Macrotipo, Microtipo	Premio attribuito a Film, Attori e/o Registi
<b>Vincita*</b>	Data	Data, Macrotipo(FK), Microtipo(FK)	Vincita di un Premio da parte di Film, Attori e/o Registi

Entità	Attributi	Identificatore	Descrizione
<b>Recensore*</b>	Codice, Nome, Cognome	Codice	Generalizzazione di Utente e Critico
<b>Critico*</b>	Codice, Nome, Cognome	Codice	Critico Cinematografico
<b>Casa-Produzione</b>	Nome	Nome	Casa di Produzione Cinematografica
<b>Paese*</b>	Codice, Nome, Posizione	Codice	Paese dal quale provengono dei Film
<b>File*</b>	Id, Dimensione, LingueDoppiaggio, LingueSottotitoli, BitRate, FormatoContenitore	Id	File in cui è stato salvato un Formato di un Film
<b>Edizione</b>	Id, Anno, Rapporto d'Aspetto, Lunghezza	Id	Edizione di un Film (e.g. "Godfather" fu rilasciato una seconda volta nel 1977 con alcune differenze)
<b>Formato-Codifica*</b>	Famiglia, Versione, MaxBitrate, Lossy	Famiglia, Versione	Generalizzazione di Audio e Video, Formato di Codifica e Decodifica
<b>Audio*</b>	Famiglia, Versione, MaxBitrate, Lossy	Famiglia, Versione	Specializzazione di Formato Codifica

Entità	Attributi	Identificatore	Descrizione
Video*	Famiglia, Versione, Anno, Bitrate, Lossy	Famiglia, Versione	Specializzazione di Formato Codifica
Server	Id, Lunghezza Banda, MTU, MaxConessioni, Carico Attuale, Posizione	Id	Server appartenente alla rete distribuita
IP Range	Inizio, Fine, DataInizio, DataFine	Inizio, Fine, DataInizio	Range di IP appartenenti ad una Zona Geografia
Utente	Codice, Nome, Cognome, Email, Password	Codice	Utente registrato al servizio di streaming
Connessione	IP, Inizio, Fine, Hardware	IP, Inizio, Codice(FK)	Connessione al servizio da parte di un Utente
Fattura	ID	ID	Fattura emessa da FilmSphere
Pagamento*	Data, Carta di Credito	ID(FK)	Pagamento effettuato da un Utente per una Fattura



Entità	Attributi	Identificatore	Descrizione
Visualizzazione	Timestamp	Timestamp, ID(FK), Codice(FK)	Visualizzazione di un File da parte di un Utente
Abbonamento	Tipo, Tariffa, Durata, Definizione, Offline, GB Mensili, Max Ore	Tipo	Abbonamento fornito dal servizio di streaming

Le entità indicate con "\*" sono generalizzazione, specializzazioni o altre entità che sono state modificate durante la ristrutturazione.

In particolare durante la ristrutturazione sono state eliminate le seguenti entità: **Attore, Regista, Recensore, Audio, Video**.

Sono state modificate le seguenti: **File, Pagamento, Film**.

Mentre sono state aggiunte le seguenti: **Lingua, Carta di Credito**.

Entità	Attributi	Identificatore	Descrizione
Film	ID, Titolo, Descrizione, MediaRecensioni, NumeroRecensioni	ID	File in cui è stato salvato un Film
File	ID, Dimensione, BitRate, FormatoContenitore	ID	File in cui è stato salvato un Film
Pagamento	Data	ID(FK)	Pagamento effettuato da un Utente per una Fattura
Lingua	Nome	Nome	Lingua in cui possono essere disponibili sia audio che sottotitoli
Carta di Credito	PAN, Data di Scadenza, CVV	PAN	Carta di Credito utilizzata per un Pagamento

## 2.2 Dizionario delle Relazioni

Relazione	Descrizione	Componenti	Attributi
<b>Vincita-Artista</b>	Vincita, da parte di un Artista, di un premio	Vincita (0, 1), Artista (0, N)	-
<b>Vincita-Premio</b>	Vincita di un particolare Premio	Vincita (1, 1), Premio (1, N)	-
<b>Vincita-Film</b>	Vincita, da parte di un Film, di un premio	Vincita (1, 1), Film (0, N)	-
<b>Recitazione*</b>	Recitazione da parte di un Attore in un Film	Film (1, N), Attore (1, N)	-
<b>Regia*</b>	Direzione da parte di un Regista di un Film	Regista (1, N), Film (1, 1)	-
<b>Genere-Film</b>	Appartenenza di un Film ad uno o più Generi	Genere (1, N), Film (1, N)	-
<b>Critica</b>	Valutazione, da parte di un Critico, di un Film	Critico (1, N), Film (0, N)	Testo, Data, Voto
<b>Recensione</b>	Valutazione, da parte di un Utente, di un Film	Utente (0, N), Film (0, N)	Voto
<b>Produzione</b>	Produzione, da parte di una Casa di Produzione, di un Film	CasaPro- duzione (1, N), Film (1, 1)	Anno

Relazione	Descrizione	Componenti	Attributi
Appartenenza	Residenza di una Casa di Produzione in un Paese	CasaProduzione (1, 1), Paese (0, N)	-
Restrizione	Impossibilità di vedere un Edizione in un Paese	Edizione (0, N), Paese (0, N)	-
Edizione-File	Corrispondenza di uno o piu' File ad un'Edizione di un film	File (1, 1), Edizione (1, N)	-
VideoFile*	Presenza di un Formato Video in un File	Video (1, N), File (1, 1)	Qualità
AudioFile*	Presenza di un Formato Audio in un File	Audio (1, N), File (1, 1)	Qualità
P.o.P.	Presenza di un File in un Server	File (1, N), Server (1, N)	-
Geolocal	Corrispondenza di un IP Range ad un Paese	Paese (1, N), IP Range (1, 1)	-
Erogazione	Erogazione, da parte di un Server, di un file per la Visualizzazione di un utente	Paese (1, N), IP Range (1, 1)	Inizio-Erogazione

Relazione	Descrizione	Componenti	Attributi
Connes- sioneUtente	Corrispondenza tra un Utente e la sua Connessione	Utente (0, N), Conessione (1, 1)	-
Emissione	Emissione di una Fattura per un Utente	Utente (0, N), Fattura (1, 1)	Data
Pagamen- toFattura	Corrispondenza tra una Fattura ed il suo, eventuale, Pagamento	Fattura (0, 1), Pagamento (1, 1)	-
Visualiz- zazioneCon- nessione	Corrispondenza tra la Visualizzazione e la Connessione di un utente	Connessione (0, N), Visualizzazione (1, 1)	-
Abbona- mento Utente	Corrispondenza tra un Utente ed il suo ultimo Abbonamento	Utente (1, 1), Abbonamento (0, N)	DataInizio
Visualiz- zazioneEdi- zione	Corrispondenza tra la Visualizzazione e l'Edizione visionata	Visualizzazione (1, 1), Edizione (0, N)	-
Esclusione	Esclusione di un Genere per un Abbonamento	Abbonamento (0, N), Genere (0, N)	-

Le relazioni indicate con ”\*” sono state modificate durante la ristrutturazione.

Durante la ristrutturazione sono state modificate le seguenti relazioni: **VideoFile**, **AudioFile**, **Regia**, **Recitazione**.

Mentre sono state aggiunte le seguenti: **Doppiaggio**, **Sottotitolaggio DistanzaPrecalcolata**, **Pagamento Carta**.

Relazione	Descrizione	Componenti	Attributi
<b>VideoFile</b>	Presenza di un Formato Video in un File	FormatoCodifica (0,N), File (1,1)	FPS, Risoluzione
<b>AudioFile</b>	Presenza di un Formato Audio in un File	FormatoCodifica (0, N), File (1, 1)	Frequenza, BitDepth
<b>Recitazione</b>	Recitazione da parte di un Artista in un Film	Film (1, N), Artista (0, N)	-
<b>Regia</b>	Direzione da parte di un Artista di un Film	Artista (1, N), Film (1, 1)	-
<b>Doppiaggio</b>	Doppiaggio in una Lingua disponibile in un File	Lingua (0, N), File (1, N)	-
<b>Sottotitolaggio</b>	Sottotitolaggio in una Lingua disponibile in un File	Lingua (0, N), File (1, N)	-
<b>DistanzaPre-calcolata</b>	Distanza Precalcolata tra un Paese e un Server	Paese (1, N), Server (1, N)	Valore Distanza
<b>Pagamento-Carta</b>	Corrispondenza tra un Pagamento e la Carta utilizzata	Pagamento (1, 1), Carta di Credito (1, N)	-

## 2.3 Regole di Vincolo

Le regole indicate con "\*" sono state aggiunte durante la ristrutturazione.

Area	Vincolo	Entità
<b>Contenuti*</b>	Un Artista deve aver recitato o diretto almeno un Film	<b>Film, Artista, Vincita</b>
<b>Contenuti*</b>	Una Lingua deve comparire in almeno un Doppiaggio o in almeno un Sottotitolaggio	<b>Lingua, Formato</b>

## 2.4 Regole di Derivazione

Le regole indicate con "\*" sono state aggiunte durante la ristrutturazione.

Area	Vincolo	Entità
<b>Contenuti</b>	Il rating critico di un Film è uguale alla media dei voti delle Critiche di quel film	<b>Film, Critico</b>
<b>Contenuti</b>	Il rating popolare di un Film è uguale alla media dei voti delle Recensioni di quel film ed è memorizzato come attributo ridondante nell'entità Film	<b>Film, Utente</b>
<b>Streaming*</b>	ValoreDistanza di DistanzaPrecalcolata è uguale alla distanza in km tra Posizione di Paese e Posizione di Server	<b>Server, Paese</b>
<b>Streaming</b>	CaricoAttuale di Server è una ridondanza che indica il numero di Erogazioni attive che impiegano tale Server. Quando una Erogazione viene avviata (inserita nel DB) il contatore aumenta di 1 e quando termina (viene cancellata) diminuisce di 1	<b>Server, Visualizzazione</b>
<b>Streaming</b>	CaricoAttuale di Server deve essere sempre minore di MaxConnessioni, se così non è il Server è in sovraccarico	<b>Server</b>

## 2.5 Descrizione ER

In questo paragrafo sono trattati nel dettaglio alcuni attributi, entità o relazioni il cui significato o funzionamento potrebbe non risultare chiaro ed evidente ad uno sguardo al diagramma Entità-Relazioni.

### 2.5.1 Entità Critico e Utente

Ciascuna delle due entità ha la possibilità di esprimere un giudizio su un Film tramite un Voto. Un **Critico** però non è una persona qualunque: è un delegato di un giornale o un esperto di cinema e della sua **Critica** si memorizza anche un Testo: un commento testuale riguardante il **Film** criticato. Gli utenti che guardano un Film potranno leggere queste critiche prima di guardarlo e decidere quindi se iniziarne la visione o meno.

Entrambe le entità condividono inoltre tre attributi, motivo per cui è stata effettuata la generalizzazione (esclusiva), ma poiché le medie dei voti dei critici e quelle dei voti degli utenti saranno poi distinte, assieme al fatto che solo un Critico può allegare commenti è stato deciso, in fase di Ristrutturazione, di separare le entità.

### 2.5.2 Vincita e Premio

Per **Macrotipo** e **Microtipo** si intendono, rispettivamente, la tipologia e la categoria del premio, ad esempio nel caso di "Oscar al miglior film" il **Macrotipo** sarebbe "Oscar", mentre il **Microtipo** "miglior film".

Per **Vincita** si intende un premio relativo ad uno specifico anno (e.g. Oscar al miglior film 2021)

### 2.5.3 Edizione

Uno stesso **Film** può essere ridistribuito più volte dalla casa di produzione e avere, per ogni nuova distribuzione piccoli o grandi cambiamenti, come l'aggiunta o la rimozione di scene. Un esempio in questo caso può essere *Star Wars: Una Nuova Speranza* di George Lucas che, sebbene sia stato prodotto nel 1977 ha avuto decine di riedizioni (di cui l'ultima nel 2020); tra esse sono state effettuate piccole modifiche a dettagli o anche grandi a personaggi.

Il Rapporto d'Aspetto è un altro attributo che può variare: uno stesso **Film** può averne vari, un esempio è il recente *Oppenheimer* di Christopher Nolan, che è stato rilasciato sia in rapporto d'aspetto IMAX che in 16:9.

Questo fa sì che la durata dello stesso **Film** possa cambiare tra le varie **Edizioni**: entità che ha lo scopo di raccogliere queste informazioni nella Base di Dati.



#### 2.5.4 File

Per File si intende un file in cui è stato salvato il contenuto di un'edizione di un film. Esistono file, per una stessa **Edizione**, con risoluzione, dimensione e FPS differenti per ottemperare a varie necessità, ad esempio il limite di risoluzione per alcuni abbonamenti.

Con FormatoContenitore intendiamo l'estensione del file (e.g. .AVI .mp4).

#### 2.5.5 FormatoCodifica

Per FormatoCodifica si intende lo standard utilizzato per codificare dati audio e/o video (e.g. MPEG-4 Part 10 è lo standard più utilizzato dai file MP4)

#### 2.5.6 Qualità di VideoFile

La Qualità di un **File** Video viene misurata mediante due valori: Risoluzione e FPS (Frame Per second). La prima è la risoluzione (qualità d'immagine) dei fotogrammi, ovvero il numero di pixel che ciascun fotogramma possiede; il secondo è il numero di fotogrammi che verranno visualizzati ogni secondo. Maggiori sono entrambi questi valori maggiore è la Qualità Video ma anche la dimensione del **File** e il suo BitRate.

#### 2.5.7 Qualità di AudioFile

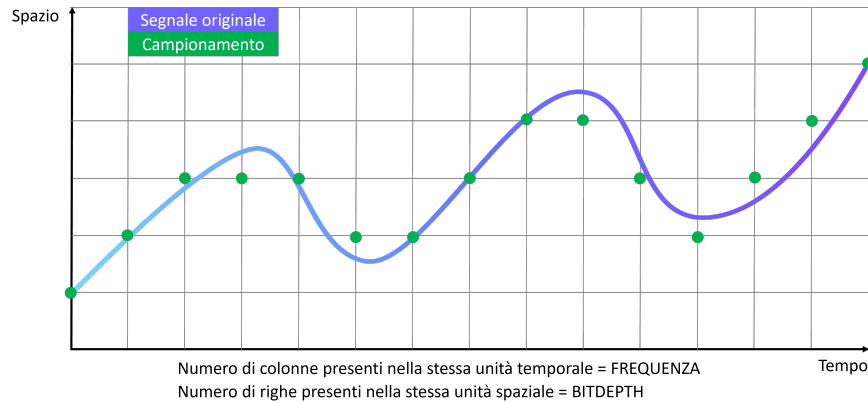
La Qualità di un File Audio viene misurata mediante due valori: Frequenza e BitDepth. La prima è la frequenza con cui il segnale audio è stato campionato (più alta è più la Qualità è maggiore); la seconda è la precisione con cui ogni punto è individuato (maggiore è il BitDepth maggiore è la Qualità del segnale).

Bisogna segnalare che, tuttavia, **File** con maggiore Qualità Audio hanno anche Dimensione maggiore e maggiore BitRate di fruizione.

#### 2.5.8 IP Range

Vi è la necessità di sapere da dove si connettono gli **Utenti** per avere una migliore fruizione dei contenuti ed anche applicare eventuali restrizioni dovute a leggi.

Se fosse l'utente stesso a comunicare il **Paese** (nazione) dove si trova ci sarebbe il problema di richieste al sistema con valori falsi: un **Utente** potrebbe



dalle impostazioni del proprio dispositivo modificare sia la posizione riportata dal GPS (che diventa quindi inutile alla base di dati) che il **Paese** dove il dispositivo pensa di trovarsi.

L'utente non può però nascondere il suo indirizzo IP (senza usare una VPN) durante la connessione e da esso si può tramite questa entità risalire al **Paese** dove si trova.

L'indirizzo IP di un dispositivo viene assegnato dal proprio ISP (Internet Service Provider, esempi sono TIM e Vodafone) al momento in cui ci si aggancia al ripetitore di quest'ultimo.

Gli indirizzi IP che un ISP può assegnare vengono comprati (e posseduti) in blocchi continui, individuabili quindi da un IP di inizio e da uno di Fine. La DataInizio di un range è la data in cui il range è stato acquistato ed è diventato quindi valido.

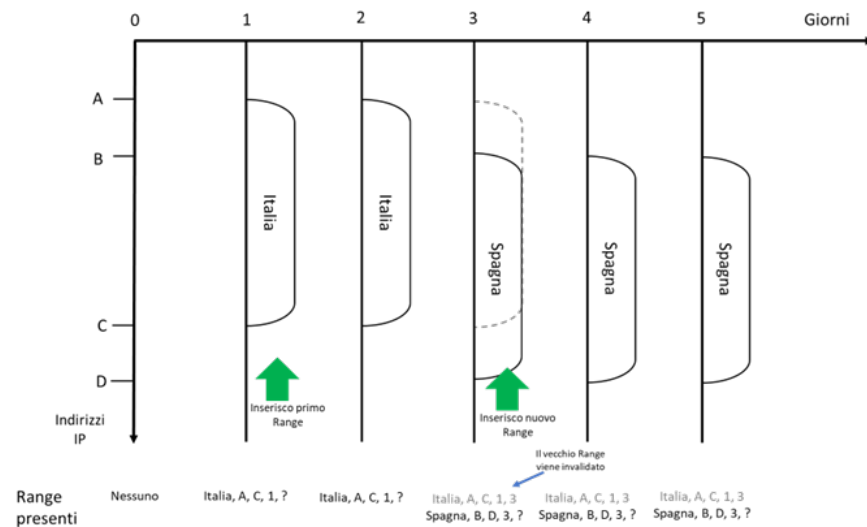
Ciascun ISP può avere sede in uno e un solo **Paese**, quindi un blocco di IP permette di risalire al **Paese** proprietario.

Un blocco (o parte) di indirizzi può però essere rivenduto ad altri ISP e quindi cambiare **Paese** di riferimento.

Dei Vincoli di Integrità faranno in modo che nella Base di Dati non vi siano range che collidono entrambi validi in uno stesso istante temporale e faranno ciò assegnando una data di fine validità (DataFine) ai range quando ne viene inserito uno nuovo che collide con uno già presente e impostando come scaduti range ineriti con data di inizio validità inferiore ad altri con

cui collidono già presenti

Due diversi range collidono se e solo se esiste un indirizzo IP che appartiene a entrambi i range.



### 2.5.9 Server

Il **CaricoAttuale** del server viene misurato in numero di connessioni attualmente aperte, mentre con **MaxConnessioni** ci riferiamo al massimo umero di connessioni che possono essere aperte in un certo istante.

Il server è considerato in sovraccarico se il numero di connessioni attualmente aperte, ossia **CaricoAttuale**, è superiore ad una certa percentuale del numero massimo di connessioni (e.g. se CaricoAttuale supera l'80% di MaxConnessioni il Server è in sovraccarico).

Maximum Transmission Unit (**MTU**), in italiano Unità massima di trasmissione, si indica la dimensione massima, misurata in byte, di un pacchetto dati che può essere inviato.

### 2.5.10 Abbonamento

La **Durata** dell'**Abbonamento** è il numero di giorni dopo il quale l'abbonamento scade, insieme all'attributo **DataInizio** della relazione **AbbonamentoUtente**

è possibile determinare se l'attuale abbonamento è in vigore o scaduto. Non viene salvato lo storico di tutti gli abbonamenti di un **Utente** ma solamente l'abbonamento più recente dell'utente, sia esso scaduto o in vigore.

Con **Definizione** si intende la definizione massima consentita da quel tipo di abbonamento. **Offline** è un booleano che indica se l'abbonamento permette all'utente di scaricare contenuti in locale.

Con **MaxOre** ci si riferisce al numero massimo di ore mensili di cui un utente può usufruire per la fruizione di contenuti. **GB Mensili** sono il massimo numero di GB utilizzabili da un utente ogni mese.

#### 2.5.11 Popolarità di Artista

La **Popolarità di Artista** è un valore numerico decimale che va da 0 a 10 e funge, per l'appunto, da metrica della fama, notorietà e celebrità di un artista.

Essa verrà, inoltre, presa in considerazione per fornire valutazioni riguardo i film in cui gli artisti prendono parte sia come attori che come registi.

## 3 Ristrutturazione

### 3.1 Eliminazione delle Generalizzazioni

#### 3.1.1 Artista

**Artista** è una generalizzazione totale sovrapposta specializzata in **Attore** e **Regista** per distinguere i diversi ruoli che una persona può assumere all'interno di un film.

Entrambe le specializzazioni di **Artista** non sono dotate di attributi propri ma possiedono unicamente gli attributi relativi ad **Artista**.

Questa generalizzazione è stata ristrutturata decidendo di conservare l'entità **Artista** la quale mantiene la relazione **VincitaArtista** ed, inoltre, guadagna le due relazioni **Recitazione** e **Regia** che erano precedentemente collegate, rispettivamente, a **Attore** e **Regista** cambiandone, tuttavia, le cardinalità.

Infatti, le cardinalità di **Regia** sono **(1, N)** rispetto a **Film**, come nell'ER non ristrutturato, e **(0, N)** verso **Artista**, diversamente da prima in quanto bisogna contemplare la presenza di un regista che non reciti in nessun film.

Mentre le cardinalità di Recitazione sono **(1, 1)** rispetto a **Film**, come nell'ER non ristrutturato, e **(0, N)** verso **Artista**, diversamente da prima in quanto bisogna contemplare la presenza di un attore che non diregga nessun film.

La nostra scelta di ristrutturazione deriva, innanzitutto, da una gestione sicuramente più efficiente del caso sporadico in cui esista un artista che ricopra sia un ruolo di **Attore** che un ruolo di **Regista** ed, oltretutto, dalla consapevolezza che non vi sia uno spreco di memoria, che si sarebbe verificato qualora avessimo ristrutturato conservando sia **Attore** che **Regista**, dovuto alla natura sovrapposta della generalizzazione.

### 3.1.2 Recensore

**Recensore** è una generalizzazione totale ed esclusiva specializzata in **Critico** e **Utente** per distinguere le diverse valutazioni, ossia critica e recensione, che vengono espresse nei confronti di un film.

La specializzazione **Critico** non è dotata di attributi propri, mentre la specializzazione **Utente** possiede due attributi propri, ossia email e password. Questa generalizzazione è stata ristrutturata decidendo di conservare le due entità **Critico** ed **Utente** eliminando l'entità **Recensore**.

La nostra scelta di ristrutturazione deriva, innanzitutto, da una gestione indubbiamente più efficiente della memoria giacché, ristrutturando solamente con l'entità **Recensore**, ci sarebbero valori NULL, sebbene sporadici, per email e password nel caso dei critici ed, oltretutto, dalla chiara separazione sia a livello concettuale e logico ma anche pratico e funzionale tra un **Critico**, che fornisce unicamente critiche, ed un utente il quale, oltre a recensire, ha un evidente ruolo nell'**Area Clienti**, ovvero tutto ciò che riguarda abbonamenti, visualizzazioni e pagamenti.

### 3.1.3 FormatoCodifica

**FormatoCodifica** è una generalizzazione totale e sovrapposta specializzata in **Audio** e **Video** per distinguere i formati utilizzati codificazione e decodificazione di dati uditivi e/o visivi.

Entrambe le specializzazioni di **FormatoCodifica** non sono dotate di attributi propri ma possiedono unicamente gli attributi relativi a **FormatoCodifica**.

Questa generalizzazione è stata tradotta decidendo di mantenere solamente

l'entità **FormatoCodifica** la quale preserva le due relazioni con **File**, ossia **AudioFile** e **VideoFile**, modificandone, però, le cardinalità.

Infatti, le cardinalità di **VideoFile** rimangono **(1,1)** rispetto a **File** ma mutano in **(0, N)** rispetto a **FormatoCodifica** poiché bisogna considerare l'esistenza di un formato unicamente audio. Invece, le cardinalità di **AudioFile** rimangono **(1,1)** rispetto a **File** ma mutano in **(0, N)** rispetto a **FormatoCodifica** poiché bisogna considerare l'esistenza di un formato unicamente video.

La nostra decisione di ristrutturazione è dovuta sia alla semplicità di introdurre una sola entità con una relazione piuttosto che due entità con due relazioni, senza nessuno spreco di memoria a cagione dell'assenza di attributi annessi alle singole specializzazioni, che alla gestione più efficiente del caso in cui un **FormatoCodifica** sia sia **Audio** che **Video**, essendo la generalizzazione sovrapposta, la quale non comporterebbe alcuno spreco di memoria.

## 3.2 Eliminazione degli Attributi Multivalore

### 3.2.1 LinguaDoppiaggio

Si considera che ogni **File** debba contenere al suo interno audio in una o più lingue. L'attributo multivalore è, pertanto, stato gestito introducendo sia un'entità **Lingua**, usata altresì per le lingue in cui sono disponibili i sottotitoli, che una relazione **Doppiaggio** che intercorre tra **Lingua**, con cardinalità **(0, N)**, in quanto una lingua potrebbe essere disponibile nei sottotitoli di qualche film ma mai disponibile come audio in nessun film, e **File**, con cardinalità **(1, N)** in quanto ogni file deve avere l'audio disponibile in almeno una lingua (si ignora la presenza del cinema muto in quanto ormai desueto).

### 3.2.2 LinguaSottotitoli

Si considera che ogni **File** possa contenere al suo interno **Sottotitoli** in varie lingue. L'attributo multivalore è stato trasformato in una relazione **Sottotitolaggio** che intercorre tra **Lingua** (creata al punto precedente), con cardinalità **(0, N)**, in quanto una lingua può essere presente come audio in un film e mai nei sottotitoli di alcun film, e **File**, con cardinalità **(0, N)**, in quanto un film non deve necessariamente avere sottotitoli.

## 3.3 Analisi Ridondanze

### 3.3.1 Ridondanze presenti

Nell'ER iniziale è presente una sola ridondanza: *CaricoAttuale* di **Server**. Essa è stata inserita in modo da facilitare (velocizzare) l'individuazione dei server sovraccarichi e, se ve ne sono, operare sulle connessioni attive su quei server per bilanciare la distribuzione degli utenti. Per garantire una fluidità del servizio questa operazione è ripetuta con alta frequenza e deve quindi completare velocemente in modo da non sovraccaricare lei stessa il DB. Essa è calcolata contando tutte le **Erogazioni** attive (presenti nel DB) sul server

### 3.3.2 Introduzione nuove Ridondanze

È stata inserita una ulteriore ridondanza sotto la forma della relazione **DistanzaPrecalcolata** tra **Server** e **Paese**, che conterrà all'interno la distanza

(espressa in chilometri) tra ogni **Server** e ogni **Paese**, in modo da velocizzare l'esecuzione della query che restituisce il server più adatto per la fruizione di un contenuto da parte dell'utente evitando il ricalcolo della distanza tra ogni **Server** ed il **Paese** dell'utente.

Sono state aggiunte all'entità Film tre attributi ridondanti: *MediaCritiche*, *MediaRecensioni*, *NumeroRecensioni*.

Essi sono necessari per evitare di calcolare la media delle critiche e delle recensioni di ogni **Film**, operazione molto comune perché avviene ogni volta che un **Utente** apre la pagina relativa ad un **Film**.

*NumeroRecensioni* permette di modificare la *MediaRecensioni* alla aggiunta/modifica/rimozione di una **Recensione** senza ricalcolare il valore dall'intero insieme dei dati.

### 3.4 Partizionamento e Accorpamento ER

#### 3.4.1 Qualità di VideoFile

L'attributo composto **Qualità** non è altro che un raggruppamento di specifiche tecniche nell'ambito video di un **File** multimediale; è quindi scomposto in due attributi *Risoluzione* e *FPS*.

#### 3.4.2 Qualità di AudioFile

L'attributo composto **Qualità** non è altro che un raggruppamento di specifiche tecniche nell'ambito audio di un **File** multimediale; è quindi scomposto in due attributi *Frequenza* e *BitDepth*.

#### 3.4.3 Posizione

L'attributo composto **Posizione** descrive una coppia Latitudine e Longitudine in modo da individuare un punto sulla superficie terrestre.

Esso, a differenza degli altri attributi composti, non necessita di essere scomposto per facilitare l'Implementazione Fisica, in quanto il DBMS in cui la Base di Dati sarà implementata (Oracle MySQL) supporta attributi di tipo *POINT*, che possono essere utilizzati per memorizzare coppie di valori numerici assieme.

Oltretutto, memorizzare una coppia *Latitudine* e *Longitudine* in questo modo facilita successivamente il calcolo di distanza tra due **Posizioni**, sfruttando la funzione nativa di MySQL *ST\_DISTANCE\_SPHERE*.



#### 3.4.4 Carta di Credito

L'attributo composto **Carta di Credito**, dell'entità **Pagamento** contiene le informazioni sulla carta impegnata nel pagamento di una **Fattura**.

Essa è stata trasformata in un'entità, poiché si prevede che un **Utente** utilizzi più volte la stessa carta; pertanto, se tutti i dati della carta fossero attribuiti di **Pagamento** non vi sarebbero una gestione efficiente della memoria a causa di valori ripetuti.

È stata quindi creata l'entità **Carta di Credito** con i seguenti attributi: *PAN*, *Scadenza*, *CVV*. La chiave dell'entità è *PAN*, in quanto ogni carta può essere individuata unicamente dal suo numero.

È stata quindi aggiunta anche una relazione **PagamentoCarta** tra le entità **Pagamento**, con cardinalità **(1, 1)**, e **Carta di Credito**, con cardinalità **(1, N)**, in quanto una carta può essere usata in più pagamenti ed in un pagamento può essere usata una sola carta.

## 4 Tavole dei Volumi

Le entità e le relazioni con segante con "\*" hanno volume assegnato per ipotesi.

### 4.1 Area Contenuti

Concetto	Tipo	Volume	Note
Film*	E	1 000	Assunzione iniziale
Produzione	R	1 000	Cardinalità (1, 1) con <b>Film</b>
Casa di Produzione*	E	20	Assunzione iniziale
Appar-tenenza	R	20	Cardinalità (1, 1) con <b>Casa di Produzione</b>
Paese	E	195+1	Nel mondo ci sono 195 stati riconosciuti, il 196-esimo rappresenta lo stato <i>Mondo</i> , utilizzato come ancora di salvezza per dare una posizione agli utenti quando non si riesce a trovare lo stato in cui si trovano
Lingua	E	141	Nel mondo ci sono 141 lingue
Doppiaggio	R	18 000	Consideriamo che ogni <b>File</b> sia doppiato, in media, in 4 <b>Lingue</b>
Sotto-titolaggio	R	13 500	Consideriamo che ogni <b>File</b> sia sottotitolato, in media, in 3 <b>Lingue</b>
Edizione Film	R	1 500	Cardinalità (1, 1) con <b>Edizione</b>
Artista*	E	4 000	Assunzione Iniziale
Regia	R	1 000	Cardinalità (1, 1) con <b>Film</b>
Recitazione	R	20 000	Consideriamo che, in media, 20 attori recitino in un <b>Film</b>
Genere*	E	27	Assunzione iniziale
GenereFilm	R	2 000	Consideriamo che, in media, un <b>Film</b> appartenga a due <b>Generi</b>

Concetto	Tipo	Volume	Note
Critico*	E	25	Assunzione iniziale
Critica	R	400	Si stima che un <b>Critico</b> faccia in media poco più di 15 critiche a <b>Film</b> diversi
Recensione	R	25 000	Si stima che in media un <b>Film</b> abbia 25 recensioni da <b>Utenti</b> diversi
Premio*	E	50	Assunzione iniziale
Vincita	E	1 500	Si stima che ogni <b>Premio</b> sia stato vinto una volta l'anno negli ultimi 30 anni
VincitaPremio	R	1 500	Cardinalità (1, 1) con <b>Vincita</b>
VincitaFilm	R	1 500	Cardinalità (1, 1) con <b>Vincita</b>
VincitaArtista	R	1 200	Si stima che la maggior parte (80%) delle vincite riguardi in maniera specifica un <b>Artista</b> (Attore o Regista)

## 4.2 Area Formato

Concetto	Tipo	Volume	Note
Edizione	E	1 500	Si stima che, in media, la metà dei <b>Film</b> abbia due edizioni
File	E	4 500	Si stima che ogni <b>Edizione</b> abbia in media 3 <b>File</b> associati, in modo da permettere la fruizione su dispositivi diversi (che possono supportare solo alcuni formati) e anche velocità di connessione deboli
Edizione File	R	4 500	Cardinalità (1, 1) con <b>File</b>
Formato Codifica*	E	100	Si stima che esistano circa 100 Formati di codifica tra video e audio
VideoFile	R	4 500	Cardinalità (1, 1) con <b>File</b>
AudioFile	R	4 500	Cardinalità (1, 1) con <b>File</b>
Restrizione	R	1 000	Si stima che ogni <b>Paese</b> bandisca, in media, 5 <b>Edizioni</b>

### 4.3 Area Clienti

Concetto	Tipo	Volume	Note
Utente*	E	1 000 000	Assunzione iniziale
Abbon-Utente	R	1 000 000	Cardinalità (1, 1) con <b>Utente</b>
Abbonamento*	E	5	Assunzione iniziale
Esclusione	R	15	Si stima che ogni <b>Abbonamento</b> escluda, in media, circa 3 <b>Generi</b>
Fattura	E	12 000 000	Fattura contiene le fatture emesse in un anno. Per ogni utente viene emessa una fattura ogni mese, quindi 12 fatture per utente in un anno.
Emissione	R	12 000 000	Cardinalità (1, 1) con <b>Fattura</b>
Pagamento	E	11 000 000	Si stima che per ogni utente le 11 ultime fatture siano state pagate mentre l'ultima no. Quindi 11 su 12, ovvero il 92% delle emissioni.
Pagamento Fattura	R	11 000 000	Cardinalità (1, 1) con <b>Pagamento</b>
Carta di credito	E	1 500 000	Si stima che la metà degli <b>Utenti</b> abbia usato, in media, due <b>Carte di Credito</b>
Pagamento Carta	R	11 000 000	Cardinalità (1, 1) con <b>Pagamento</b>
Connessione	E	30 000 000	Si stima che un <b>Utente</b> si connetta, in media, circa 1 volta al giorno. Il sistema tiene memoria di tutte le connessioni effettuate negli ultimi 30 giorni
Conn-Utente	R	30 000 000	Cardinalità (1, 1) con <b>Connessione</b>
Visualizzazione	E	15 000 000	Si stima che un <b>Utente</b> guardi, in media, un' <b>Edizione</b> ogni due giorni. Il sistema tiene traccia dei contenuti guardati negli ultimi 30 giorni
Visualizzazione Connessione	R	15 000 000	Cardinalità (1, 1) con <b>Visualizzazione</b>
Visualizzazione Edizione	R	15 000 000	Cardinalità (1, 1) con <b>Visualizzazione</b>

## 4.4 Area Streaming

Concetto	Tipo	Volume	Note
Server*	E	20	Circa il 10% dei <b>Paesi</b>
Distanza Precalcolata	R	3920	Esattamente ogni combinazione possibile tra <b>Server</b> e <b>Paese</b>
P.o.P.	R	9 000	Ogni file è presente, in media, su 2 server
Erogazione	R	40 000	Corrisponde al numero di utenti mediamente connessi in ogni istante. Stimando il numero medio di ore giornaliere di attività a 1 si deduce che il 4% degli utenti è mediamente connesso in ogni istante
Ip-Range*	E	130 000	Assunzione iniziale basata su valori reali
GeoLocal	R	130 000	Cardinalità (1, 1) con Ip-Range

## 5 Operazioni sui Dati

### 5.1 Analisi delle Ridondanze

#### 5.1.1 CaricoAttuale

*CaricoAttuale*, attributo di **Server** contiene, in ogni momento, il numero di connessioni in corso di un server CDN. Tale valore potrebbe essere calcolato contando le **Erogazioni** che puntano a quel **Server**. Per misurare il carico di un singolo server, senza la ridondanza, sarebbe necessario quindi analizzare tutte le Erogazioni e contare quelle che soddisfano il criterio.

Tale operazione avviene però quando un **Utente** è alla ricerca del **Server** migliore che gli può offrire il contenuto che cerca, rendendo necessario analizzare tutti il CaricoAttuale di tutti i Server.

Questa operazione è molto comune e deve concludersi in fretta in modo da non far aspettare l'Utente e garantire una migliore esperienza.

Il tipo di aggiornamento della ridondanza è l'**Immediate Refresh**: essa viene aggiornata ad ogni inserimento/modifica/cancellazione nella relazione **Erogazione**.

Di seguito è riportata la frequenza giornaliera delle attività che coinvolgono tale ridondanza:

Azione	FrequenzaGiornaliera	Nota
Lista Server	500 000	Si considera che ogni utente guardi un file ogni due giorni

**Input:** chiave del **File** che si intende guardare

**Output:** Lista di Server disponibili, ciascuno con il suo CaricoAttuale

I Volumi interessati sono:

Nome	Tipo	Volume
P.o.P	E	9 000
Erogazione	R	40 000

Gli accessi per Visualizzazione sarebbero in media, senza ridondanza:

Concetto	Costrutto	Accessi	Tipo	Note
P.o.P	R	9 000	R	Conoscendo la chiave di <b>File</b> si può passare direttamente a leggere i <b>P.o.P.</b> per individuare i <b>Server</b> che possiedono il <b>File</b> che si cerca. La lettura di tutte le occorrenze della relazione restituirà, in media, 2 chiavi di <b>Server</b>
Erogazione	R	40 000	R	Vengono lette tutte le occorrenze per individuare quelle che soddisfano il criterio. La chiave si <b>Server</b> è stata ottenuta da <b>P.o.P.</b>
<b>Totale:</b>		49 000		Soluzione non ottimale

Mentre, con la ridondanza sono in media:

Concetto	Costrutto	Accessi	Tipo	Note
P.o.P	R	9 000	R	Conoscendo la chiave di <b>File</b> si può passare direttamente a leggere i <b>P.o.P.</b> per individuare i <b>Server</b> che possiedono il <b>File</b> che si cerca. La lettura di tutte le occorrenze della relazione restituirà, in media, 2 chiavi di <b>Server</b>
Server	E	2	R	La chiave di <b>Server</b> è già stata individuata e accedendo all'entità si ha già il CaricoAttuale: parametro della scelta
<b>Totale:</b>		9 002		Soluzione ottimale

È importante far notare che, in questo caso, ogni volta che un'**Erogazione** viene aggiunta/rimossa sarà necessario accedere a **Server** per alterarne il CaricoAttuale. Questo costringe a sommare ai valori giornalieri ottenuti il costo ottenuto dalla periodica modifica di CaricoAttuale.

Azione	FrequenzaGiornaliera	Nota
Aggiornamento <i>CaricoAttuale</i>	750 000	Si considera che ogni <b>Visualizzazione</b> sfrutti, in media, 1,5 <b>Server</b> diversi nel corso della sua durata

I Volumi coinvolti sono:

Nome	Tipo	Volume
Server	E	20

La cui tabella degli Accessi è:

Concetto	Costrutto	Accessi	Tipo	Note
Server	E	2	W	La chiave di <b>Server</b> è già nota alla relazione <b>Erogazione</b> , permettendo di accedere al <b>Server</b> interessato e impostare il <i>CaricoAttuale</i> a +1 o -1 il suo precedente valore se l' <b>Erogazione</b> è stata aggiunta o rimossa.
<b>Totale:</b>		1		

Gli accessi giornalieri delle due tipologie sono quindi:

Metodologia	Accessi al giorno		
	Lettura	Scrittura	Totale
Senza ridondanza	$24,5 \times 10^9$	0	$24,5 \times 10^9$
Con ridondanza	$4,5 \times 10^9$	500 000	$4,5 \times 10^9$

### 5.1.2 DistanzaPrecalcolata

**DistanzaPrecalcolata** è una relazione tra **Paese** e **Server** che ha, come attributo, il valore della distanza tra un server ed un paese. Tale valore potrebbe essere calcolato a partire dalle posizioni sia di paese che di server. Qualora si stia ricercando il server ottimale per un utente ubicato in un certo paese occorre conoscere, oltre al carico del server, anche la distanza tra di esso ed il paese. Invece di ricalcolare tutte le distanze ogni volta è più efficiente andarle a leggere da **DistanzaPrecalcolata**.

**DistanzaPrecalcolata** non viene aggiornata giornalmente in quanto la posizione dei server e dei paesi è fissa.

Di seguito è riportata la frequenza giornaliera delle attività che coinvolgono tale ridondanza:

Azione	FrequenzaGiornaliera	Nota
Distanza Server	1 000 000	Si considera che ogni <b>Utente</b> guardi un file ogni due giorni, e che per farlo usi circa due <b>Server</b> , dei quali si deve calcolare la distanza dalla posizione dell' <b>Utente</b> . Per ogni richiesta si stima che ci siano due <b>Server</b> (di cui si conosce già la chiave) che ospitano il contenuto richiesto

**Input:** chiave del **Paese** dove si trova l'utente e dei Server (2 in media) che ospitano il Film richiesto

**Output:** Lista di ID di tutti i Server, ciascuno con la sua distanza dal paese dell'utente



I Volumi interessati sono:

Nome	Tipo	Volume
Distanza Precalcolata	R	3 920

Gli accessi per Distanza Server sarebbero in media, senza ridondanza:

Concetto	Costrutto	Accessi	Tipo	Note
Paese	E	1	R	Conoscendo la chiave di <b>Paese</b> vi si può accedere puntualmente
Server	E	2	R	Viene letta la posizione dei <b>Server</b> richiesti
<b>Totale:</b>		3		Soluzione non ottimale

Mentre, con la ridondanza sono in media:

Concetto	Costrutto	Accessi	Tipo	Note
Distanza- Precalco- lata	R	2	R	Conoscendo sia la chiave di <b>Paese</b> che di <b>Server</b> si può leggere la relazione in maniera puntuale.
<b>Totale:</b>		2		Soluzione ottimale

Questa soluzione è ottimale anche perché i valori di *Posizione* sia di un **Paese** che di un **Server** sono praticamente immutabili: è un evento rarissimo che una nazione nasca/scompaia o cambi notevolmente i suoi confini, come un **Server** non può cambiare posizione mentre sta erogando contenuti. Le distanze tra le *Posizioni* possono quindi essere calcolate una volta sola all'inserimento di queste entità e praticamente mai cambiate.

Gli accessi giornalieri delle due tipologie sono quindi:

Metodologia	Accessi al giorno		
	Lettura	Scrittura	Totale
Senza ridondanza	3 X 10 <sup>6</sup>	0	3 X 10 <sup>6</sup>
Con ridondanza	2 X 10 <sup>6</sup>	0	2 X 10 <sup>6</sup>

### 5.1.3 MediaRecensioni NumeroRecensioni

**MediaRecensioni** e **NumeroRecensioni** sono due attributi di **Film** che tengono conto, rispettivamente, della media delle recensioni del film e del numero di recensioni che esso ha ricevuto.

Il valore della media ha un'utilità pratica e immediata mentre il numero di recensione assolve alla funzione di alleggerimento del mantenimento della ridondanza.

Il valore della media delle recensione potrebbe essere calcolando andando a considerare tutte le recensioni del film e, poi, calcolandone la media. Qualore, però, si stia ricercando il film con valutazione media maggiore, la quale è un'operazione piuttosto comune, si andrebbe a ricalcolare molteplici volte le medie di tutti i film.

Invece di ricalcolarle ogni volta ex nihilo è più efficiente salvarle per poi rileggerle.

Il tipo di aggiornamento della ridondanza è l'**Immediate Refresh**: essa viene aggiornata ad ogni inserimento/modifica/cancellazione nella relazione **Recensione**.

Di seguito è riportata la frequenza giornaliera delle attività che coinvolgono tale ridondanza:

Azione	FrequenzaGiornaliera	Nota
Film Migliori	30 000 000	Si considera che ogniqualvolta un utente si connetta gli venga mostrato il film con recensioni migliori come raccomandazione automatica
Inserimento Recensione	70	Si stima che le 25 000 recensioni siano state accumulate nel corso di un anno

**Input:** Nulla

**Output:** Titoli e media recensioni dei 20 Film con media più alta

I Volumi interessati sono:

Nome	Tipo	Volume
Film	E	1 000
Recensione	R	25 000

Gli accessi per Visualizzazione sarebbero in media, senza ridondanza:

Concetto	Costrutto	Accessi	Tipo	Note
Recensione	R	25 000	R	Devono essere lette tutte le occorrenze, raggruppate in base alla chiave di <b>Film</b> e ordinate in base alla media di <i>Voto</i>
Film	E	20	R	Noti gli <i>ID</i> dei <b>Film</b> di cui si deve riportare il <i>Titolo</i> si accede direttamente a essi dalla chiave
<b>Totale:</b>		25 020		Soluzione non ottimale

Mentre, con la ridondanza sono in media:

Concetto	Costrutto	Accessi	Tipo	Note
Film	E	1 000	R	Si leggono tutti i Film e si trovano le migliori 20 medie
<b>Totale:</b>		1 000		Soluzione ottimale

Gli accessi relativi all'aggiornamento della ridondanza sono invece:

Concetto	Costrutto	Accessi	Tipo	Note
Film	E	1	W	Conoscendo la chiave di <b>Film</b> e memorizzando sia <i>NumeroRecensioni</i> che <i>MediaRecensioni</i> si possono calcolare i nuovi valori senza effettuare altre letture
<b>Totale:</b>		1		

Gli accessi giornalieri delle due tipologie sono quindi:

Metodologia	Accessi al giorno		
	Lettura	Scrittura	Totale
Senza ridondanza	25 X 10 <sup>9</sup>	0	25 X 10 <sup>9</sup>
Con ridondanza	1 X 10 <sup>9</sup>	70	1 X 10 <sup>9</sup>

## 5.2 Analisi delle Operazioni

### 5.2.1 Vincite di un Film

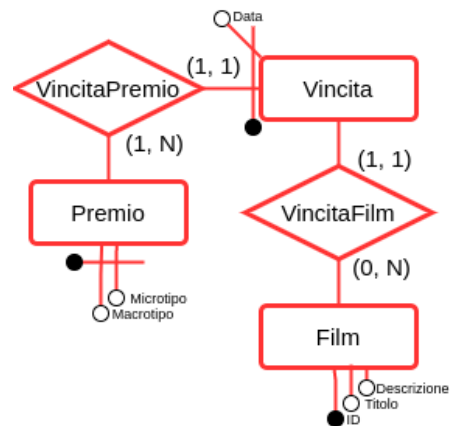
**Descrizione:** Non appena viene selezionato il film da visionare la piattaforma carica una lista contenenti sia i premi vinti dal film che la data in cui gli sono stati attribuiti.

**Input:** ID di Film

**Output:** Macrotipo, Microtipo e Data

**Frequenza Giornaliera:** 500 000 Poiché un **Utente**, in media, vede un **Film** ogni due giorni e i premi vengono mostrati nella pagina dell'app relativa al **Film**

**Porzione di Diagramma Interessato:**



**Porzione della Tavola dei Volumi:**

Concetto	Tipo	Volume
Premio	E	50
Film	E	1 000
Vincita	E	1 500
VincitaFilm	R	1 500
VincitaPremio	R	1 500

**Tavola degli Accessi:**

Concetto	Costrutto	Accessi	Tipo	Note
VincitaFilm	R	2	R	Conoscendo la chiave di <b>Film</b> si può passare direttamente a leggere <b>VincitaFilm</b> , si stima che per ogni <b>Film</b> , in media, vi siano due <b>Vincite</b> associate.
Vincita	E	2	R	Da <b>VincitaFilm</b> si può accedere direttamente a <b>Vincita</b> , dalla quale si estrae l'attributo <i>Data</i>
Vincita-Premio	R	2	R	Da <b>Vincita</b> si può accedere direttamente a <b>Vincita-Premio</b> ed estrarre da esso <i>MacroTipo</i> e <i>MicroTipo</i>
<b>Totale:</b>		6		

### 5.2.2 Generi di un Film

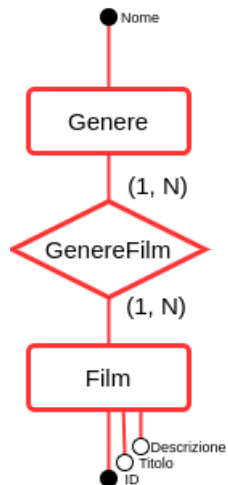
**Descrizione:** Non appena viene selezionato un film viene caricata una lista contenente i generi a cui può essere ascritto il film.

**Input:** ID

**Output:** Nome

**Frequenza Giornaliera:** 500 000 é un utente, in media, vede un film ogni due giorni

**Porzione di Diagramma Interessato:**



### Porzione della Tavola dei Volumi:

Concetto	Tipo	Volume
Genere	E	27
GenereFilm	R	2000
Film	E	1000

### Tavola degli Accessi:

Concetto	Costrutto	Accessi	Tipo	Note
Film	R		R	Conoscendo la chiave di <b>Film</b> si può passare direttamente a leggere <b>VincitaFilm</b> .
GenereFilm	E		R	Da <b>VincitaFilm</b> si può accedere direttamente a Vincita
Genere	R		R	Da <b>Vincita</b> si può accedere direttamente a Vincita
<b>Totale:</b>		20		Soluzione ottimale

### 5.2.3 Massima Risoluzione di un Film

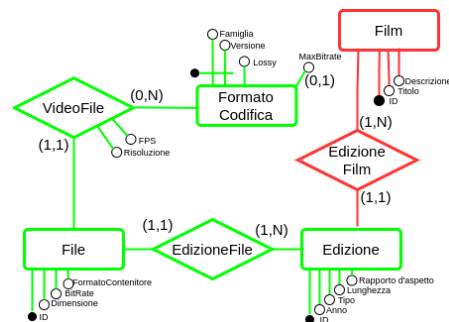
**Descrizione:** dato un **Film** restituire, tra tutte le **Edizioni**, il **File** con la *Risoluzione* maggiore.

**Input:** ID di Film

**Output:** ID di File con associata Risoluzione

**Frequenza Giornaliera:** 500 000 Perché si stima che ogni **Utente** guardi un **Film** ogni due giorni e che quando lo fa il sistema cerchi di fruirgli un **File** con la massima *Risoluzione* possibile

**Porzione di Diagramma interessato:**



**Porzione della Tabella dei Volumi:**

Concetto	Tipo	Volume
EdizioneFilm	R	1 500
EdizioneFile	R	4 500
VideoFile	R	4 500

**Tavola degli Accessi:**

Concetto	Costrutto	Accessi	Tipo	Note
Edizione-Film	R	2	R	Conoscendo la chiave di <b>Film</b> si accede alla relazione per trovare l'ID dell' <b>Edizione</b> associata, si stima che ci siano in media due riscontri
File-Edizione	E	6	R	Noti gli ID delle edizioni del <b>Film</b> target si leggono gli ID dei <b>File</b> ad esso associati, si stima che siano in media 3 per ogni <b>Edizione</b>
Video-File	R	6	R	Noti gli ID dei <b>File</b> interessati si estrae la <i>Risoluzione</i> trovandone la massima
<b>Totale:</b>		14		

**5.2.4 Numero di Film non disponibili in un Abbonamento**

**Descrizione:** Quando l'utente sceglie quale abbonamento acquistare può informarsi sul numero di film esclusi da uno specifico abbonamento

**Input:** Tipo

**Output:** Numero di Film Esclusi

**Frequenza Giornaliera:** 33 000 Poiché gli abbonamenti hanno durata mensile si stima che ogni fine mese il 20% degli utenti decida di cambiare abbonamento e, dunque, controlli il numero di film esclusi per tutti i 5 abbonamenti

**Porzione di Diagramma Interessato:**

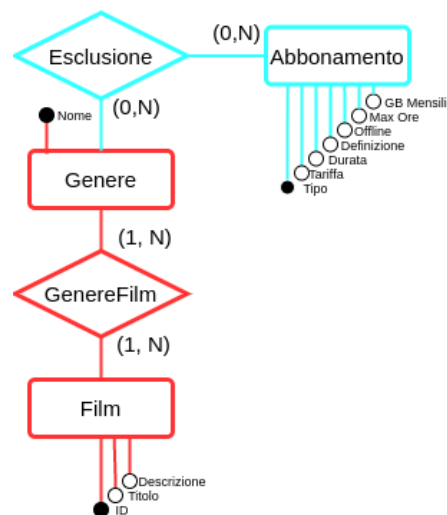


Tavola dei Volumi:

Concetto	Tipo	Volume
Abbonamento	E	5
Esclusione	R	12
Genere	E	27
GenereFilm	R	2 000
Film	E	1 000

Tavola degli Accessi:

Concetto	Costrutto	Accessi	Tipo	Note
Edizione-Film	R	2	R	Conoscendo la chiave di <b>Film</b> si accede alla relazione per trovare l'ID dell' <b>Edizione</b> , si stima che ci siano in media due riscontri
File-Edizione	E	6	R	Noti gli ID delle edizioni del <b>Film</b> target si leggono gli ID dei <b>File</b> associati, si stima che siano in media 3 per ogni <b>Edizione</b>
VideoFile	R	6	R	Noti gli ID dei <b>File</b> interessati se ne estrae la <i>Risoluzione</i> e se ne trova la massima
<b>Totale:</b>		20		Soluzione ottimale



### 5.2.5 Film disponibili in Lingua specifica

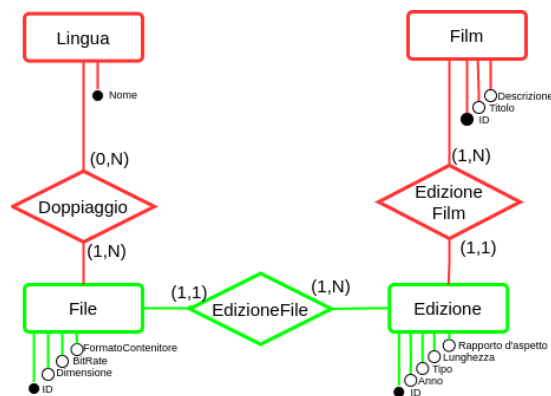
**Descrizione:** data una **Lingua** restituire, tutti i titoli di **Film** che hanno almeno un **Doppiaggio**, presente nel sistema, nella lingua richiesta.

**Input:** ID di Lingua

**Output:** Lista di ID e Titoli di Film che soddisfano il criterio

**Frequenza Giornaliera:** 100 000 Perché si stima che ogni Utente guardi un Film ogni due giorni e che un quinto delle volte ne scelga uno in base ad una Lingua specifica. La ricerca dei Film disponibili in una lingua avviene quindi, per ogni Utente, una volta ogni 10 giorni.

**Porzione di Diagramma interessato:**



**Porzione della Tabella dei Volumi:**

Concetto	Tipo	Volume
Doppiaggio	R	18 000
EdizioneFile	R	4 500
EdizioneFilm	R	1 500
Film	E	1 000

**Tavola degli Accessi:**

Concetto	Costrutto	Accessi	Tipo	Note
Doppiaggio	R	255	R	Essendo presenti 141 <b>Lingue</b> , stimate equamente distribuite, si hanno in media 255 File doppiati per ogni lingua. Da essi si estraggono gli ID dei <b>Film</b> che soddisfano il criterio

Edizione-File	R	255	R	Noti gli ID dei <b>File</b> per ciascuno si individua (legge) l'edizione associata. Si stima che ogni <b>Edizione</b> abbia in media due file associati e che quindi per ogni 2 <b>File</b> vi sia un' <b>Edizione</b>
Edizione-Film	R	123	R	Noti gli ID delle <b>Edizioni</b> interessate si risale agli ID dei Film associati, stimati come circa 2/3 delle Edizioni
Film	E	82	R	Noti gli ID dei <b>Film</b> cercati se ne estraggono i <i>Titoli</i>
<b>Totale:</b>		715		

### 5.2.6 Film più visti recentemente

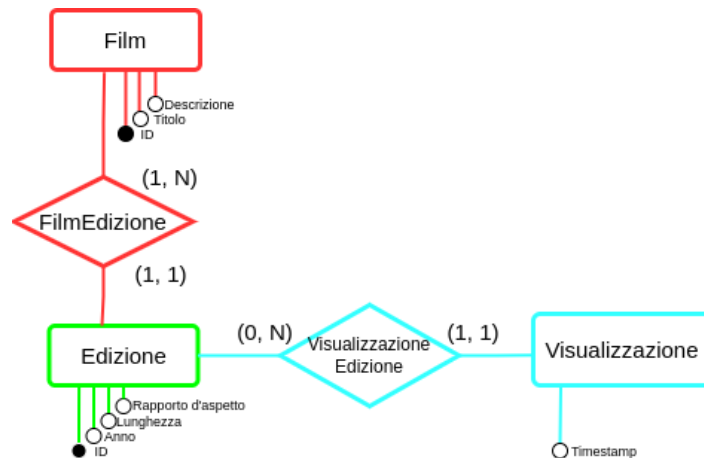
**Descrizione:** Un utente potrebbe voler conoscere gli  $n$  film più visti di recente nella piattaforma di streaming.

**Input:** Numero di Film

**Output:** Titoli di Film più visti nell'ultimo mese

**Frequenza Giornaliera:** 200 000 Si stima che il 20% delle volte che un utente si connette vada a controllare la classifica dei **Film** più visti di recente

**Porzione di Diagramma Interessato:**



**Porzione della Tabella dei Volumi:**

Concetto	Tipo	Volume
Film	E	1 000
Edizione	E	1 500
Visualizzazione	E	15 000 000
Film- Edizione	E	1 500
Visualizzazione-Edizione	R	15 000 000

**Tavola degli Accessi:**

Concetto	Costrutto	Accessi	Tipo	Note
Visualizzazione	E	0	R	Essendo tutte le <b>Visualizzazioni</b> memorizzate già riguardanti gli ultimi 31 giorni si può leggere direttamente <b>Visualizzazione-Edizione</b>
Visualizzazione-Edizione	R	15 000 000	R	Si legge ogni occorrenza e si estrae la chiave di <b>Edizione</b>
Edizione-Film	R	1 500	R	Si sommano, raggruppando per <i>ID</i> di <b>Film</b> , le visualizzazioni di ogni <b>Edizione</b>
Film	E	1 000	R	Noti gli <i>ID</i> dei <b>Film</b> se ne estraggono i <i>Titoli</i>
<b>Totale:</b>		15 000 000		

**5.2.7 Cambio Abbonamento**

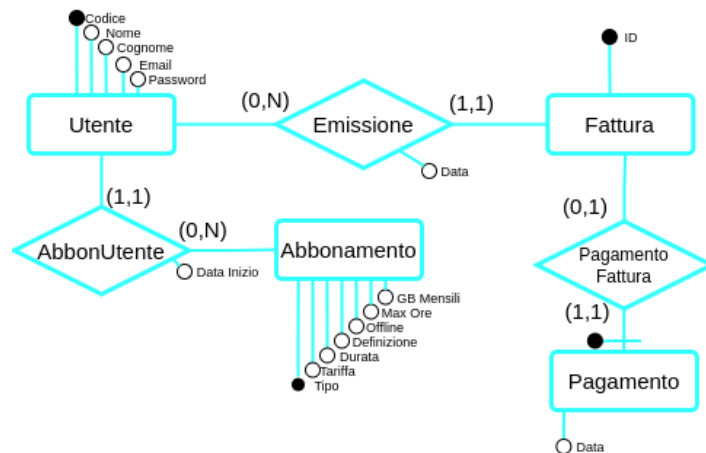
**Descrizione:** Un utente può decidere di cambiare piano di abbonamento, ovviamente prima di fare ciò occorre controllare che sia in pari coi pagamenti.

**Input:** Codice, Tipo

**Output:**

**Frequenza Giornaliera:** 7 000 Poiché gli abbonamenti hanno durata mensile si stima che ogni fine mese il 20% degli utenti decida di cambiare abbonamento

**Porzione di Diagramma Interessata:**



Porzione della Tabella dei Volumi:

Concetto	Tipo	Volume
Utente	E	1 000 000
Abbonamento	E	5
Fattura	E	12 000 000
Pagamento	E	11 000 000
Abbonamento- Utente	R	1 000 000
Emissione	R	12 000 000
Pagamento- Fattura	R	11 000 000

Tavola degli Accessi:

Concetto	Costrutto	Accessi	Tipo	Note
Abbon- Utente	R	1	R	Viene letto l' <i>ID</i> dell'attuale <b>Abbonamento</b> . Se l' <i>ID</i> ottenuto è il medesimo al quale si vuole passare (caso improbabile) si termina l'operazione
Abbona- mento	R	1	R	Viene letto il prezzo dell'attuale <b>Abbonamento</b> , se il prezzo è 0,00 allora l' <b>Utente</b> è in pari e si può saltare all'ultimo punto. Nel caso medio però questo non avviene

Emissione	R	12	R	Dall' <i>ID</i> dell' <b>Utente</b> si leggono le <b>Emissioni</b> relative a egli. Si ottengono quindi gli <i>ID</i> delle <b>Fatture</b>
Pagamento	E	12	R	Avendo <b>Pagamento</b> stessa chiave di <b>Fattura</b> , ne si controlla l'esistenza (tentativo di lettura)
Abbon- Utente	R	1	W	Si modifica la relazione sostituendo il vecchio <b>Abbonamento</b> con quello nuovo
<b>Totale:</b>		28		

### 5.2.8 Film con Recensioni migliori

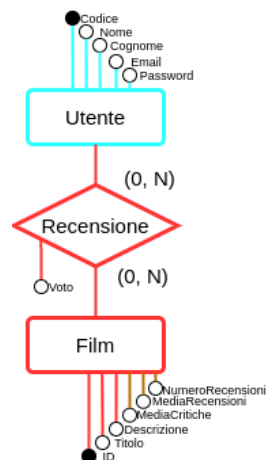
**Descrizione:** Restituire i migliori 20 **Film** con la più alta media di *Voto* nelle **Recensioni** associate.

**Input:** Nulla

**Output:** Lista di *ID* e *Titoli* di **Film**.

**Frequenza Giornaliera:** 500 000 Si considera che ogni volta che un utente si connette gli vengano mostrato i **Film** con **Recensioni** migliori come raccomandazione automatica

**Porzione di Diagramma interessato:**



**Porzione della Tabella dei Volumi:**

Concetto	Tipo	Volume
Film	E	1 000
Recensione	R	25 000

**Tavola degli Accessi:**

Concetto	Costrutto	Accessi	Tipo	Note
Film	E	1 000	R	La ridondanza <i>MediaRecensioni</i> permettere di individuare il miglior Film senza dover calcolare le recensioni medie
<b>Totale:</b>		1 000		

## 6 Progettazione Logica

### 6.1 Ridenominazione Nomi

### 6.2 Traduzione in schema Logico

Concetti ER	Traduzione	Note
Film, Produzione, Regia	Film( <u>ID</u> , Titolo, Descrizione, Anno, CasaProduzione, NomeRegista, CognomeRegista, MediaRecensioni, NumeroRecensioni)	
Artista	Artista( <u>Nome</u> , <u>Cognome</u> , Popolarità)	
Premio, VincitaPremio, Vincita, VincitaArtista, VincitaFilm	VincitaPremio(Macrotipo, Microtipo, <u>Data</u> , Film, NomeArtista, CognomeArtista)	1
Recitazione	Recitazione( <u>Film</u> , <u>NomeAttore</u> , <u>CognomeAttore</u> )	
Critico	Critico( <u>Codice</u> , Nome, Cognome)	
Critica	Critica( <u>Critico</u> , <u>Film</u> , Testo, Data, Voto)	
Recensione	Recensione( <u>Utente</u> , <u>Film</u> , Voto)	

Concetti ER	Traduzione	Note
Genere	Genere( <u>Nome</u> )	
GenereFilm	GenereFilm( <u>Genere</u> , <u>Film</u> )	
CasaProduzione, Appartenenza	CasaProduzione( <u>Nome</u> , Paese)	
Lingua	Lingua( <u>Nome</u> )	
Doppiaggio	Doppiaggio( <u>File</u> , <u>Lingua</u> )	
Sottotitolaggio	Sottotitolaggio( <u>File</u> , <u>Lingua</u> )	
Paese	Paese( <u>Codice</u> , Nome, Posizione)	
File, VideoFile, AudioFile, EdizioneFile	File( <u>ID</u> , Dimensione, BitRate, FormatoContenitore, FamigliaAudio, VersioneAudio, FamigliaVideo, VersioneVideo, Risoluzione, FPS, BitDepth, Frequenza, Edizione)	
FormatoCodifica	FormatoCodifica( <u>Famiglia</u> , <u>Versione</u> , Lossy, MaxBitrate)	
Edizione, EdizioneFilm	Edizione( <u>ID</u> , Film, Anno, Tipo, Lunghezza, Rapporto D'Aspetto)	



Concetti ER	Traduzione	Note
Restrizione	Restrizione( <u>Edizione</u> , <u>Paese</u> )	
IP Range, Geolocal	IPRange( <u>Inizio</u> , <u>Fine</u> , <u>DataInizio</u> , DataFine, Paese)	
Server	Server( <u>ID</u> , LunghezzaBanda, MTU, MaxConnessioni, CaricoAttuale, Posizione)	
P.o.P.	P.o.P.( <u>File</u> , <u>Server</u> )	
Distanza- Precalcolata	DistanzaPrecalcolata( <u>Paese</u> , <u>Server</u> , ValoreDistanza)	
Visualizzazione, Visualiz- zazioneEdizione, Visualiz- zazioneConnes- sion	Visualizzazione(Timestamp, <u>Edizione</u> , <u>Utente</u> , <u>IP</u> , <u>InizioConnessione</u> )	
Erogazione	Erogazione(Timestamp, <u>Edizione</u> , <u>Utente</u> , <u>IP</u> , <u>InizioConnessione</u> , InizioErogazione, Server)	2
Connessione, ConnessioneU- tente	Connessione( <u>Utente</u> , <u>IP</u> , <u>Inizio</u> , Fine, Hardware)	
Utente, Abbona- mentoUtente	Utente( <u>Codice</u> , Nome, Cognome, Email, Password, Abbonamento, DataInizioAbbonamento)	

Concetti ER	Traduzione	Note
Abbonamento	Abbonamento( <u>Tipo</u> , Tariffa, Durata, Definizione, Offline, MaxOre, GBMensili)	
Esclusione	Esclusione( <u>Abbonamento</u> , <u>Genere</u> )	
Fattura, Emissione, Pagamento, PagamentoCarta	Fattura( <u>ID</u> , Utente, DataEmissione, DataPagamento, CartaDiCredito)	3
Carta di Credito	CartaDiCredito( <u>PAN</u> , Scadenza, CVV)	

Note:

1. Stimiamo che la metà dei premi siano vinti da un artista, dunque è più conveniente avere due valori nulli invece di ripetere metà delle volte i 3 attributi chiave in una tabella che avrebbe tradotto la relazione VincitaArtista
2. Le Erogazioni conterranno i dati solo delle visualizzazioni attive. Se **Erogazione** fosse stato tradotto come un attributo *ServerAttuale* in **Visualizzazione**, al termine di essa sarebbe dovuto essere impostato come nullo. Visualizzazione avrebbe avuto quindi la maggior parte delle occorrenze con un campo nullo.
3. Pagamento e PagamentoCarta confluiscono dentro Fattura in quanto entrambe le entità condividono la chiave e si stima che al più 1/12 delle fatture possa non essere pagata. Si ha quindi anche un risparmio di memoria in quanto non serve memorizzare più volte la chiave di Fattura

## 6.3 Vincoli

### 6.3.1 Vincoli di Integrità Referenziale

1. *CasaProduzione* di **Film** in *Nome* di **CasaProduzione**
2. *NomeRegista, CognomeRegista* di **Film** in *Nome, Cognome* di **Artista**
3. *Film* di **VincitaPremio** in *ID* di **Film**
4. *NomeArtista, CognomeArtista* di **VincitaPremio** in *Nome, Cognome* di **Artista**
5. *Film* di **Recitazione** in *ID* di **Film**
6. *NomeAttore, CognomeAttore* di **Recitazione** in *Nome, Cognome* di **Artista**
7. *Critico* di **Critica** in *Codice* di **Critico**
8. *Film* di **Critica** in *ID* di **Film**
9. *Utente* di **Recensione** in *Codice* di **Utente**
10. *Film* di **Recensione** in *ID* di **Film**
11. *Genere* di **GenereFilm** in *Nome* di **Genere**
12. *Film* di **GenereFilm** in *ID* di **Film**
13. *Paese* di **CasaProduzione** in *Codice* di **Paese**
14. *File* di **Doppiaggio** in *ID* di **File**
15. *Lingua* di **Doppiaggio** in *Nome* di **Lingua**
16. *File* di **Sottotitolaggio** in *ID* di **File**
17. *Lingua* di **Sottotitolaggio** in *Nome* di **Lingua**
18. *FamigliaAudio, VersioneAudio* di **File** in *Famiglia, Versione* di **FormatoCodifica**
19. *FamigliaVideo, VersioneVideo* di **File** in *Famiglia, Versione* di **FormatoCodifica**
20. *Edizione* di **File** in *ID* di **Edizione**
21. *Film* di **Edizione** in *ID* di **Film**
22. *Edizione* di **Restrizione** in *ID* di **Edizione**
23. *Paese* di **Restrizione** in *Codice* di **Paese**
24. *Paese* di **IP Range** in *Codice* di **Paese**
25. *File* di **P.o.P** in *ID* di **File**
26. *Server* di **P.o.P** in *ID* di **Server**
27. *Paese* di **DistanzaPrecalcolata** in *Codice* di **Paese**
28. *Server* di **DistanzaPrecalcolata** in *Server* di **Paese**
29. *Edizione* di **Erogazione** in *ID* di **Edizione**

30. *Timestamp, Utente, IP, Edizione, InizioConnessione* di **Erogazione** in *Timestamp, Utente, IP, Edizione, InizioConnessione* di **Visualizzazione**
31. *Server* di **Erogazione** in *ID* di **Server**
32. *Edizione* di **Visualizzazione** in *ID* di **Edizione**
33. *Utente, IP, InizioConnessione* di **Visualizzazione** in *Utente, IP, Inizio* di **Connessione**
34. *Utente* di **Connessione** in *Codice* di **Utente**
35. *Abbonamento* di **Utente** in *Tipo* di **Abbonamento**
36. *Abbonamento* di **Esclusione** in *Tipo* di **Abbonamento**
37. *Genere* di **Esclusione** in *Nome* di **Genere**
38. *Utente* di **Fattura** in *Codice* di **Utente**
39. *CartaDiCredito* di **Fattura** in *PAN* di **Carta di Credito**

### 6.3.2 Vincoli di Tupla

Ogni attributo che fa parte di un identificatore primario o esterno ha il vincolo *NOT NULL*.

Fa eccezione solo *CartaDiCredito* di **Fattura**, che avrà valore *NULL* finché un pagamento non viene effettuato.

I seguenti attributi hanno la possibilità di essere *NULL*:

1. *DataFine* di **IPRange**, quando un range è in corso di validità
2. *MaxBitRate* di **FormatoCodifica**, quando la metodologia in questione non pone massimi al *BitRate* di un **File**

I seguenti attributi, invece, non potranno essere *NULL* (anche se un'altra interpretazione potrebbe dargli questa possibilità) in quanto:

1. *MaxOre* di **Abbonamento**, quando non vi è un limite sulle ore massime di streaming al giorno lo si imposta a 100
2. *GBMensili* di **Abbonamento**, quando non vi è un limite sulla mole massima di streaming al mese lo si imposta a 10 000

Questa scelta è stata effettuata nell'ottica di semplificare i controlli di tali valori durante lo straming.

Sono individuati, oltretutto, i seguenti vincoli intra-relazionali:

1. L'istante di *Fine* di una **Connessione** deve essere non minore di quello di *Inizio*
2. Se impostata, la *DataFine* di un **IPRange** deve essere non minore di *DataInizio*
3. Un **IPRange** deve avere IP di *Fine* non minore di quello di *Inizio*

### 6.3.3 Vincoli di Dominio

Sono individuati i seguenti vincoli sul dominio di alcuni attributi.

1. Una **Recensione** deve avere un *Voto* compreso tra 0.0 e 5.0
2. Una **Critica** deve avere un *Voto* compreso tra 0.0 e 5.0
3. In ogni momento, una **Critica** deve avere *Data* non maggiore della data corrente
4. In ogni momento il *Timestamp* di una **Visualizzazione** deve essere non maggiore di *CURRENT\_TIMESTAMP*
5. L'istante di *Fine* di una **Connessione** deve essere non minore di quello di *Inizio*
6. Sia l'istante di *Inizio* che quello di *Fine* di una **Connessione** devono essere, in ogni momento, non maggiori di *CURRENT\_TIMESTAMP*
7. La *DataInizio* di **IPRange** deve essere non maggiore della data corrente

### 6.3.4 Vincoli Inter-Relazionali

Sono individuati i seguenti vincoli Intra-Relazionali.

Il loro scopo è mantenere coerenza tra i vari dati memorizzati in relazioni diverse nella stessa Base di Dati.

1. Una **Critica** deve avere l'anno della *Data* non minore dell'*Anno* di produzione del **Film** criticato
2. Un **Edizione** deve avere *Anno* di pubblicazione non minore dell'*Anno* di produzione del **Film**
3. Un **File**, se ha associati **FormatoAudio** e **Videoentrambi** con un *MaxBitRate* deve avere l'attributo *BitRate* non maggiore della somma degli altri due
4. L'anno del *Timestamp* di una **Visualizzazione** deve essere non minore dell'*Anno* di pubblicazione dell'**Edizione** visualizzata
5. Il *Timestamp* di una **Visualizzazione** deve essere compreso tra gli istanti di *Inizio* e di *Fine* della **Connessione** associata

6. L'istante di *Fine* di una **Connessione** deve essere non minore di quello di *Inizio*
7. Una **CartadiCredito** utilizzata in un **Pagamento** non deve essere scaduta nella **DataPagamento**
8. Il timestamp *InizioErogazione* di **Erogazione** non può essere maggiore di *CURRENT\_TIMESTAMP* e non può essere minore del *TimeStamp* di inizio della Visualizzazione associata
9. In uno stesso istante non possono esistere due o più **IPRange** validi che collidono. Due range collidono quando esiste un indirizzo IP che appartiene ad entrambi i range

## 6.4 Analisi delle Dipendenze Funzionali e Normalizzazione

A questo punto si procede con l'analisi delle dipendenza funzionali delle relazioni ottenute dalla traduzione. Il risultato ottenuto, come di seguito, è in forma BCNF. Per ogni tabella in elenco, si analizzano le dipendenza funzionali in essa presenti e si inserisce un ✓ ove la BCNF viene soddisfatta, altrimenti si opera una decomposizione.

- **Film**(ID, Titolo, Descrizione, Anno, CasaProduzione, NomeRegista, CognomeRegista, MediaRecensioni, NumeroRecensioni) ✓  
ID → Titolo, Descrizione, Anno, CasaProduzione, NomeRegista, CognomeRegista, MediaRecensioni, NumeroRecensioni
- **Artista**(Nome, Cognome, Popolarità) ✓  
Nome, Cognome → Popolarità
- **VincitaPremio**(Macrotipo, Microtipo, Data, Film, NomeArtista, CognomeArtista) ✓  
Macrotipo, Microtipo, Data → Film, NomeArtista, CognomeArtista
- **Recitazione**(Film, NomeAttore, CognomeAttore) ✓
- **Critico**(Codice, Nome, Cognome) ✓  
Codice → Nome, Cognome
- **Critica**(Critico, Film, Testo, Data, Voto) ✓  
Critico, Film → Testo, Data, Voto

- **Recensione**(Utente, Film, Voto) ✓  
*Utente, Film → Voto*
- **Genere**(Nome) ✓
- **GenereFilm**(Genere, Film) ✓
- **CasaProduzione**(Nome, Paese) ✓  
*Nome → Paese*
- **Lingua**(Nome) ✓
- **Doppiaggio**(File, Lingua) ✓
- **Sottotitolaggio**(File, Lingua) ✓
- **Paese**(Codice, Nome, Posizione) ✓  
*Codice → Nome, Posizione*
- **File**(ID, Dimensione, BitRate, FormatoContenitore, FamigliaAudio, VersioneAudio, FamigliaVideo, VersioneVideo, Risoluzione, FPS, BitDepth, Frequenza, Edizione) ✓  
*ID → Dimensione, BitRate, FormatoContenitore, FamigliaAudio, VersioneAudio, FamigliaVideo, VersioneVideo, Risoluzione, FPS, BitDepth, Frequenza, Edizione*
- **FormatoCodifica**(Famiglia, Versione, Lossy, MaxBitrate) ✓  
*Famiglia, Versione → Lossy, MaxBitrate*
- **Edizione**(ID, Film, Anno, Tipo, Lunghezza, Rapporto D'Aspetto) ✓  
*ID → Film, Anno, Tipo, Lunghezza, Rapporto D'Aspetto*
- **Restrizione**(Edizione, Paese) ✓
- **IPRange**(Inizio, Fine, DataInizio, DataFine, Paese) ✓  
*Inizio, Fine, DataInizio → DataFine, Paese*
- **Server**(ID, Risoluzione, LunghezzaBanda, MTU, MaxConnessioni, CaricoAttuale, Posizione) ✓  
*ID → LunghezzaBanda, MTU, MaxConnessioni, CaricoAttuale, Posizione*
- **P.o.P.**(File, Server) ✓
- **DistanzaPrecalcolata**(Paese, Server, ValoreDistanza) ✓  
*Paese, Server → ValoreDistanza*

- **Erogazione**(Timestamp, Edizione, Utente, IP, InizioConnessione, InizioErogazione, Server) ✓  
Timestamp, Edizione, Utente, IP, InizioConnessione → InizioErogazione, Server
- **Visualizzazione**(Timestamp, Edizione, Utente, IP, InizioConnessione) ✓
- **Connessione**(Utente, IP, Inizio, Fine, Hardware) ✓  
Utente, IP, Inizio → Fine, Hardware
- **Utente**(Codice, Nome, Cognome, Email, Password, Abbonamento, DataInizioAbbonamento) ✓  
Codice → Nome, Cognome, Email, Password, Abbonamento, DataInizioAbbonamento
- **Abbonamento**(Tipo, Tariffa, Durata, Definizione, Offline, MaxOre, GBMensili) ✓  
Tipo → Tariffa, Durata, Definizione, Offline, MaxOre, GBMensili
- **Esclusione**(Abbonamento, Genere) ✓
- **Fattura**(ID, Utente, DataEmissione, DataPagamento, CartaDiCredito) ✓  
ID → Utente, DataEmissione, DataPagamento, CartaDiCredito
- **CartaDiCredito**(PAN, Scadenza, CVV) ✓  
PAN → Scadenza, CVV

Per quanto riguarda la tabella *Connessione* si potrebbe pensare che esista la dipendenza funzionale  $IP, Inizio \rightarrow Fine, Hardware$ , tuttavia qualora qualcuno si collegasse dallo stesso PC, aprendo due tab in due browser diversi nello stesso istante, avrei stesso IP e Inizio ma due User-Agent diversi, dunque due Hardware diversi.

## 7 Raccomandazione Utente

### 7.1 Rating

Il rating di un film è una valutazione espressa in numero di stelle, con la possibilità di averne mezze, che parte da un minimo di zero ed arriva ad un



massimo di 5.

Ogni film possiede due diverse tipologie di rating, ossia il rating del film ed il rating personalizzato, la prima è una metrica oggettiva che è, dunque, fissa per ogni film, la seconda, invece, varia sulla base del singolo utente in quanto dipendente dalla "history" di quest'ultimo.

Precedentemente si era già parlato di una forma di valutazione, espressa come media delle varie recensioni. Tuttavia, essa differisce dal rating non solo in merito al sistema di valutazione utilizzato, stelle invece che numerico, ma anche e soprattutto per le idee dietro di esse. Mentre la media valutazione è una misura molto superficiale, che corrisponde solamente al pensiero dei vari utenti, il rating si propone come una valutazione più sofisticata, complessa e sicuramente completa, infatti essa prende in considerazione una moltitudine di parametri, quali le recensioni di critici, le popolarità sia del regista che degli attori, il numero di premi vinti ed, eventualmente, le preferenze dell'utente.

### 7.1.1 Rating del Film

Il rating del film consente, come detto precedentemente, di dare una valutazione a tutto tondo e soprattutto oggettiva di un film.

Siano  $R_U$  la media delle recensioni degli utenti,  $R_C$  la media delle recensioni dei critici,  $RM_U$  la media delle recensioni degli utenti più alta tra tutti i film aventi almeno un genere in comune con il film,  $P_A$  la media delle popolarità degli attori,  $P_R$  la media delle popolarità dei registi ed, infine,  $PV$  i premi vinti dal film.

L'algoritmo è il seguente:

$$V = \frac{\lfloor 0.5 * (R_U + R_C) + 0.1 * (P_A + P_R) + 0.1 * PV + (R_U / RM_U) \rfloor}{2}$$

L'idea dietro di esso è assegnare un valore che va da 0 a 10 ai seguenti parametri: media recensioni e valutazioni, media popolarità attori e regista, numero di premi vinti e media recensioni rispetto al massimo per il genere. Assegnati i valori occorre fare una media pesata coi seguenti pesi: 0.5, 0.2, 0.2 e 0.1, prendere la parte intera del risultato e dividerla per due (in questo modo otteniamo le mezze stelle per un giudizio che arriva al massimo a 5 stelle).

### 7.1.2 Rating per Utente

Il rating per utente consente, come detto precedentemente, di dare una valutazione soggettiva di un film, basandosi sulle preferenze dell'utente.

Siano  $G_1, G_2$  i 2 generi preferiti dell'utente,  $A_1, A_2, A_3$  i 3 attori preferiti dell'utente,  $L_1, L_2$  le 2 lingue, in cui l'audio è disponibile, preferite dell'utente,  $R$  il regista preferito ed, infine,  $RA$  il rapporto d'aspetto preferito.

Questi valori valgono 1 se sono presenti nel film altrimenti 0 (e.g. Se  $G_1$  corrisponde a *Fantascienza*, in ogni film di *Fantascienza*  $G_1$  vale 1)

Per preferiti si intende i più visti nell'ultimo mese.

L'algoritmo è il seguente:

$$V = \frac{[2 * G_1 + G_2 + 1.5 * A_1 + A_2 + 0.5 * A_3 + L_1 + L_2 + R + RA]}{2}$$

L'idea dietro di esso è assegnare un bonus per ogni preferenza dell'utente che viene soddisfatta dal film in questione, ovviamente ad ogni preferenza viene attribuito uno specifico peso in base alla rilevanza. La somma di tutti questi pesi equivale a 10, ossia il massimo valore ottenibile che corrisponde ad una valutazione di 5 stelle.

Sommati tutti i bonus occorre prendere la parte intera per poi dividerla per due (in questo modo otteniamo le mezze stelle per un giudizio che arriva al massimo a 5 stelle).

### 7.1.3 Sistema di Raccomandazione

**Descrizione:** Il sistema di raccomandazione di contenuti è in grado di, dato un utente e la sua "history" di visualizzazioni, consigliare e proporre film potenzialmente graditi da quest'ultimo.

Una metrica perfetta per misurare quanto uno specifico film possa essere potenzialmente apprezzato da un particolare utente è il suddetto meccanismo di **Rating Utente** in quanto riesce a coniugare alla perfezione i gusti soggettivi con l'esigenza di raccomandazione.

**Input:** Codice dell'Utente e Numero di film richiesti

**Output:** Codici dei Film Raccomandati

**Commento Codice:** Sfruttando la funzione `RatingUtente` è possibile creare una tabella  $T(Film, RatingUtente)$ , sulla quale poi verrà operato un ordinamento decrescente in base al `RatingUtente`, usando `ORDER BY`, per poi andare a selezionare solo i primi  $N$  film graditi dall'utente, mediante `LIMIT N`.

## 8 Area Analytics

### 8.1 Classifiche

**Descrizione:** Questa funzionalità si occupa di stilare delle classifiche di **Film** e di **Edizioni** più gettonate sulla base sia del **Paese** che dell'**Abbonamento**. L'idea dietro questa funzionalità è di poter comporre una classifica degli  $N$  **Film** oppure **Edizioni** più visti in uno specifico **Paese** solo da **Utenti** con uno specifico **Abbonamento**.

**Input:**  $N$ , Codice del Paese, Tipo di Abbonamento,  $P$  (Variabile che determina cosa classificare)

**Output:** Classifica

## Immagine Codice

**Commento Codice:** La stored procedure prende in input i parametri sopra descritti ed, in base al valore di  $P$  decide come agire, ovvero se classificare i **Film** oppure le **Edizioni**.

Successivamente, crea una **CTE** *VisualizzazioniTarget* contenente tutte le Visualizzazioni che soddisfano sia i criteri riguardanti la collocazione geografica che quelli riguardanti il tipo di abbonamento.

Per controllare il primo occorre fare un *JOIN* con della tabella **Visualizzazione** con la tabella **IPRange** per controllare se il *Codice* del **Paese** corrispondente a quella visualizzazione è quello desiderato.

Invece, per controllare il secondo occorre fare un *JOIN* tra la tabella **Visualizzazione** e quella **Utente** per determinare se il tipo di abbonamento sia quello desiderato.

A questo punto, nel caso in cui si voglia stilare una classifica sui **Film** viene svolto un *JOIN* tra **Visualizzazioneitarget** ed **Edizione**, seguito da un raggruppamento su **Film** che ritorna una tabella del tipo Risultato(Film, NumeroVisualizzazioni). Si conclude ordinando in base a NumeroVisualizzazioni e restituendo gli  $N$  valori più grandi.

Qualora si volessero invece le **Edizioni** non occorre un *JOIN* ma si procede direttamente col raggruppamento, l'ordinamento e la restituzione dei migliori  $N$  valori.

## 8.2 Custom Analytics

La Custom Analytics ideata è quella di *Migliore Attore e Migliore Regista Inaspettato* della settimana.

Ogni settimana, considerando artisti aventi una popolarità minore di una certa soglia, viene stilata una classifica degli artisti e registi più gettonati.

Questa funzionalità serve a dare rilievo e lustro ad artisti debuttanti, emergenti o semplicemente meno noti al grande pubblico. Siamo convinti che questa opzione susciterebbe indubbiamente grande interesse tra gli utenti di *FilmSphere* in quanto sarebbe un modo per aprire i propri orizzonti ad artisti minori e, pertanto, ampliare la propria cultura cinematografica anche, magari, sviluppando un giudizio critico verso attori o registi debuttanti.

Innanzitutto, occorre definire la soglia massima di popolarità tra gli artisti considerati, che decidiamo di fissare a 2.5 in quanto siamo particolarmente interessati ad artisti poco rilevanti.

Per quanto concerne il punteggio da attribuire ad un attore o regista vengono presi in considerazione tutti i **Film** in cui quest'ultimo ha recitato oppure ha diretto e si somma al suo punteggio la valutazione media del **Film**, a cui vanno aggiunti eventuali 5 punti qualora il **Film** in questione risultasse vincitore di un qualsivoglia tipo di premio. Inoltre, per ogni premio eventualmente vinto da egli (artista) vengono attribuiti 100 punti.

Siano  $v_i$  la valutazione media dell' $i$ -esimo **Film** in cui ha partecipato l'**Artista**,  $p_i$  una variabile che assume come valore 1 se l' $i$ -esimo film ha vinto almeno un premio e 0 altrimenti  $n$  il numero di premi vinti dall'**Artista**, allora il suo punteggio  $P$  è uguale a:

$$P = \sum_{i=1}^n (v_i + 5 * p_i) + 50 * n$$

## 8.3 Bilanciamento carico

**Descrizione:** Per prevenire sovraccarichi sui **Server**, si cerca di prevedere le **Edizioni** che saranno più richieste nel futuro prossimo in un determinato Paese, in modo da copiare/spostare i **File** di tali **Edizioni** sui **Server** più vicini a quel **Paese**.

**Input:**  $N, M$  numeri naturali

**Output:** Lista di spostamenti da effettuare

La procedura funziona al seguente modo:

1. Per ogni **Paese** si individuano le **Visualizzazioni** avvenute in quel **Paese** dall'*IP*
2. Si raggruppa in base al **Paese** e all'**Edizione** visualizzata contando le visualizzazioni per ogni **Edizione** in ogni **Paese**
3. Per ogni **Paese** si selezionano le  $N$  **Edizioni** più visualizzate
4. Per ogni **Paese** si individuano gli  $M$  **Server** più vicini
5. Ogni coppia Edizione-Server, ignorando il **Paese** che non serve più, viene aggiunta, con attenzione a non aggiungere duplicati, alla lista da restituire
6. Per ogni **Edizione** si individuano tutti i **File** e si trasforma la lista in una File-Server; si scartano le coppie dove esiste già un **P.o.P.** tra quel **File** e quel **Server**

Allo stesso modo si può utilizzare lo stesso criterio per individuare i **File** che non è necessario tenere su un **Server**, seguendo la stessa procedura ma scegliendo, invece delle più visualizzate, le **Edizioni** meno visualizzate e, invece di prendere le coppie dove non vi è un **P.o.P.**, si prendono solo le coppie dove esiste già il **P.o.P.**

## 9 Area Streaming

### 9.1 Individuazione Server

**Descrizione:** Quando un **Utente** richiede di guardare un'**Edizione**, l'applicazione invia la richiesta al DBMS che deve restituire il miglior Server al quale connettersi.

**Input:** *Codice Utente*, *ID dell'Edizione*, *IP dell'Utente*, *DataInizio* di **Connessione**, *MaxBitRate* e *MaxRis*

**Output:** *ID* del server migliore o nulla in caso di non disponibilità

Dei parametri passati alla procedura, *Codice*, *ID*, *IP* e *DataInizio* servono ad individuare la **Connessione** dell'**Utente**, il suo **Abbonamento** e le relative permissioni, il **Paese** dove si trova, mentre *MaxBitRate* e *MaxRis* sono parametri determinati dalle caratteristiche del dispositivo e dalla potenza di segnale attuale della rete al quale si è connessi. Tali parametri sono calcolati dall'applicazione lato client e servono per scegliere il **File** dell'**Edizione** più adatto.

La procedura seguita per l'individuazione del Server migliore è la seguente:

1. Vengono lette l'entità **Utente** e il relativo *Abbonamento*, il **Film** associato all'**Edizione** e i **Generi** del **Film** e, se esistono, le **Esclusioni** tra l'**Abbonamento** e il **Genere**. Se l'interrogazione restituisce dei valori allora la procedura termina: non si è autorizzati a guardare il contenuto.
2. Dall'*IP* si risale al *Codice* di **Paese** tramite gli **IPRange** e si controlla se vi siano **Restrizioni** tra il **Paese** e l'**Edizione**. Se vi sono restrizioni la procedura termina: non si è autorizzati (per legge) a guardare il contenuto.
3. Per ogni **File** dell'**Edizione**, si calcolano i seguenti parametri:

$$\Delta Rate = \begin{cases} MaxBitRate - File.BitRate & \text{se } File.BitRate < MaxBitRate \\ -2(MaxBitRate - File.BitRate) & \text{altrimenti} \end{cases}$$
$$\Delta Ris = \begin{cases} MaxRis - File.Risoluzione & \text{se } File.Risoluzione < MaxRis \\ -2(MaxRis - File.Risoluzione) & \text{altrimenti} \end{cases}$$

Tali valori serviranno poi nella scelta del **File** migliore, più alti sono questi valori, minori saranno le probabilità che il **File** sia scelto

4. Per ogni **File** vengono individuati i **P.o.P.** e da essi i **Server** che li contengono. Dal **Paese** e i **Server**, passando per **DistanzaPre-calcolata** si ottiene la distanza, in km, tra la capitale del **Paese** dell'**Utente** e ogni **Server**. Di ogni **Server** si estrae il *CaricoPercentuale* col seguente modo:

$$CaricoPercentuale = CaricoAttuale / MaxConnessioni$$

5. Si effettuano le seguenti operazioni matematiche per riscalarare i valori trovati:

$$\Delta Rate' = map(\Delta Rate, 0, MAX\_RATE, 0, w_{Rate})$$

$$\Delta Ris' = map(\Delta Ris, 0, MAX\_RIS, 0, w_{Ris})$$

$$\Delta Pos' = map(ValoreDistanza, 0, MAX\_DIST, 0, w_{Pos})$$

$$CaricoPercentuale' = CaricoPercentuale * w_{Carico}$$

Dove *map* è una funzione che riscalda un numero appartenente ad un intervallo portandolo in un nuovo intervallo.

Essa è calcolata al seguente modo:

$$map(x, IN_{Min}, IN_{Max}, OUT_{Min}, OUT_{Max}) = OUT_{Min} + \frac{x - IN_{Min}}{IN_{Max} - IN_{Min}} \cdot (OUT_{Max} - OUT_{Min})$$

I valori *MAX\_RATE*, *MAX\_RIS* e *MAX\_DIST* rappresentano invece il massimo valore che le relative grandezze possono raggiungere.

I valori *w<sub>Rate</sub>*, *w<sub>Ris</sub>*, *w<sub>Pos</sub>* e *w<sub>Carico</sub>* sono i pesi che permetteranno di confrontare i tali parametri tra di loro, potendo cambiare facilmente l'importanza di uno rispetto all'altro.

Lo score Finale di ogni combinazione File-Server è assegnato come

$$Score = \Delta Rate' + \Delta Ris' + \Delta Pos' + CaricoPercentuale'$$

6. La combinazione con lo *Score* minore sarà restituita all'**Utente**.

Un esempio di valori dei pesi può essere:

- *w<sub>Rate</sub>* = 5 in modo da essere un parametro poco significativo, 17% del totale
- *w<sub>Ris</sub>* = 3 in modo da essere il parametro meno importante, 10% del totale
- *w<sub>Pos</sub>* = 12 in modo da essere il più importante (la distanza determina latency durante la fruizione), 40% del totale
- *w<sub>Carico</sub>* = 10 In modo da tenere il carico bilanciato tra i diversi **Server**, 33% del totale

## 9.2 Ribilanciamento del carico

Durante lo streaming di contenuti, gli **Utenti** possono cambiare **Server** dal quale stanno erogando, su richiesta del DBMS stesso, in modo da non sovraccaricare il sistema.

Ogni 10 minuti (intervalli detti tick) il DBMS calcolerà la media di *CaricoPercentuale* dei **Server** e se tale media è sopra il 70%.

Si individuano i (massimo 3) **Server** con *CaricoPercentuale* più alto e sopra la media e, per ciascuno di essi:

1. Si leggono le **Erogazioni** attive che riguardano tale **Server** e un **File** che è presente anche in almeno un altro **Server** che non sia nei 3 ineterssati e, se è presente, dove è presente.

Le **Erogazioni** considerate non devono essere iniziate negli ultimi 3 tick e si deve prevedere (in base al *TimeStamp* di **Visualizzazione** e la *Durata* di **Edizione**) che terminino entro il prossimo.

2. Per ciascuna di queste erogazioni risalgo al **Paese** dove si trova l'Utente dall'*IP* e dal **Paese** al *ValoreDistanza* tra l'**Utente** e il **Server** attuale
3. Per ciascuna di queste erogazioni in corso individuo le migliori 3 alternative (**Server** non carichi con il contenuto) sulla base della distanza dall'**Utente** (minore è, meglio è)
4. Per ciascuna alternativa calcolo uno *Score*, che permetterà di scegliere quali erogazioni modificare, al seguente modo:

$$Score = \begin{cases} (Dist_{Att} - Dist_{Altern})^2 & \text{se } Dist_{Altern} < Dist_{Att} \\ 50 \cdot (Dist_{Altern} - Dist_{Att}) & \text{altrimenti} \end{cases}$$

5. Per ogni **Erogazione** scelgo la miglio Alternativa tramite lo *Score*
6. Ordinando in base allo *Score*, scelgo le migliori coppie Erogazione-Alternativa in numero minore al 5% del *MaxConnessioni* del **Server**. Il risultato viene messo in una tabella temporanea che, tramite un'interrogazione avviata dai server (macchine) stessi, sarà svuotata e gli **Utenti** riassegnati, modificando le **Erogazioni** nella Base di Dati



### 9.3 Caching previsionale

**Descrizione:** Si ipotizza che ogni **File** sia presente in un hub centrale, macchina che ha il solo scopo di trasmettere tali contenuti ai **Server**, creando così i **P.o.P.**

In modo da facilitare le operazioni agli algoritmi di *Individuazione Server* e di *Ribilanciamento di quest'ultimi* viene creato anche un *Bilanciamento Preventivo*: algoritmo che cerca di prevedere i contenuti che avranno grande domanda nel prossimo futuro e, con questi dati, cerca di indicare i **Server** nei quali è più opportuno effettuare un **P.o.P.** di tali **File** creando così un sistema di Caching previsionale.

Allo stesso modo l'algoritmo individua i **File** che avranno minore probabilità di essere visualizzati in determinati **Server**, consigliando quindi di rimuoverli per fare spazio ad altri contenuti se lo spazio sulla macchina si dovesse esaurire.

La procedura di individuazione di contenuti da aggiungere funziona al seguente modo:

**Input:** X, N, M numeri interi non grandi

**Output:** Lista di spostamenti consigliati

1. Per ogni **Utente** si considera il **Paese** dal quale si connette di più e dal Paese gli N **Server** più vicini
2. Per ogni coppia **Utente, Paese** si considerano le M **Edizioni** con probabilità maggiore di essere guardate, ciascuna con la probabilità di essere guardate
3. si raggruppa in base al **Server** e ad ogni **File** dell'**Edizione**, sommando, per ogni **Server-File** la probabilità che sia guardato dall'**Utente** moltiplicata per un numero che scala in maniera decrescente in base al *ValoreDistanza* tra Paese e Server
4. Si restituiscono le prime X coppie Server-File con somma maggiore per le quali non esiste già un **P.o.P.**

La procedura di individuazione di contenuti da rimuovere funziona allo stesso modo di quella di individuazione di contenuti da raggiungere ma all'ultimo punto vengono restituite le X coppie con somma più bassa dove è presente un **P.o.P.**

## **10 Implementazione Fisica**

### **10.1 Generazione delle Tabelle**

### **10.2 Creazione dei Vincoli**

### **10.3 API**

### **10.4 Popolamento**