**Documentazione/relazione della progettazione del DB**

La prima cosa da fare nella modellazione di un DB è la progettazione concettuale dello stesso, durante questa fase si ricavano informazioni, dati e quanto altro per poi ricavare un modello/schema concettuale come il diagramma E-R, successivamente si passa alla progettazione logica, con la quale si ricostruisce il nostro modello concettuale (diagramma E-R) e si elabora il modello/schema logico (modello relazionale), infine si passa alla progettazione fisica, che forma lo schema fisico (il DB vero e proprio).

Prima di tutto definiamo e ricerchiamo informazioni sulle quattro entità principali sulle quali si basa il nostro DBMS:

**Genoma**:

L'entità "Genoma" rappresenta l'insieme completo di informazioni genetiche di un organismo. Può essere rappresentato come un'entità principale nel diagramma.

**Cromosoma**:

Un cromosoma è una struttura composta da DNA che contiene una parte del genoma di un organismo. All'interno del diagramma E-R, il cromosoma potrebbe essere rappresentato come un sottoinsieme del genoma, collegato ad esso tramite una relazione.

**Read**:

Un "Read" è una breve sequenza di DNA che è stata determinata sperimentalmente attraverso tecniche di sequenziamento. Questi read possono essere utilizzati per ricostruire il genoma completo. Nel diagramma, "Read" può essere considerato come un'entità collegata ai cromosomi, poiché i read vengono solitamente mappati sui cromosomi per l'analisi.

**Qualità**:

La "Qualità" in questo contesto si riferisce alla misura della precisione o affidabilità di un read. È spesso rappresentata sotto forma di punteggi di qualità associati a ciascun nucleotide del read.

Cerchiamo le relazioni tra queste quattro entità:

**Cromosoma e genoma**:

Un cromosoma può essere composto da uno o più genomi. Questa relazione è di tipo uno-a-molti, poiché un cromosoma può essere composto da più genomi, ma un genoma appartiene ad un solo cromosoma.

**Cromosoma e Read**:

Un Cromosoma può essere composto da molti Read. Anche questa relazione è di tipo uno-a-molti, poiché un Cromosoma può contenere molti Read, ma un Read è associato ad un solo Cromosoma.

**Qualità e read**:

Una qualità è posseduta da molti read. Questa relazione è di tipo uno-a-molti, poiché ogni qualità è posseduta da molti read, ma una read ha solo una qualità.

**Sviluppiamo il diagramma E-R**:

Immagine che contiene diagramma, origami, modello

Descrizione generata automaticamente

Specifico gli attributi delle quattro Entità

**Genoma**:

ID (identificativo unico del genoma)

Nome (nome del genoma, se applicabile)

Dimensione (lunghezza totale del genoma)

**Cromosoma**:

ID (identificativo unico del cromosoma)

Nome (nome del cromosoma)

Posizione (posizione all'interno del genoma)

**Read**:

ID (identificativo unico del read)

Sequenza (la sequenza di basi nucleotidiche del read)

Lunghezza (lunghezza del read)

**Qualità**:

ID (identificativo unico della qualità)

Phred\_score (il punteggio di qualità associato al read)

Passiamo alla prima e seconda fase della progettazione logica:

**Ristrutturazione del diagramma E-R**

eliminazione dallo schema E/R di tutti i costrutti che non possono essere direttamente rappresentati nel modello logico:

* Eliminazione degli attributi multivalore
* Eliminazione delle generalizzazioni
* Partizionamento/accorpamento di entità e associazioni
* Scelta degli identificatori principali

**Traduzione**

* si mappano i costrutti residui in elementi del modello relazionale

Ecco riportato sotto:

**cromosoma** (ID(chiave primaria), nome, posizione);

**qualità** (ID(chiave primaria), phred\_score);

**read** (ID (chiave primaria), sequenza, lunghezza, id\_cromosoma (chiave esterna), id\_qualità (chiave esterna));

**genoma** (ID(chiave primaria), nome, dimesione, id\_cromosoma (chiave esterna));

**relazioni** “derivato da”, “appartiene a”, “ha qualità”

La chiave esterna permette di creare un collegamento tra entità, ad esempio poiché un cromosoma deriva da tanti genomi, ma un genoma deriva da un solo cromosoma, il genoma deve possedere la chiave esterna (foreign key) per accedere al suo cromosoma.

Infine vi è la progettazione fisica:

In questa fase traduco lo schema logico in uno schema fisico (modello fisico) dei dati sotto forma di file, indici, dimensione dei campi, domini, tipi di dati e chiavi. Scelgo il DBMS, l'organizzazione fisica dei file e della base dati, nel mio caso ho scelto SQlite. Infine, implemento il database sulla macchina di test e/o di produzione seguendo le specifiche tecniche peculiari del DBMS.

La maggior parte dei database richiede un processo server, ma SQLite è serverless, ovvero l’applicazione può leggere e scrivere i dati direttamente senza architettura client-server. SQLite, come dice il nome, è leggero per quanto riguarda la configurazione, l’amministrazione e l’archiviazione. Non richiede installazione o configurazione, il che lo rende autonomo e meno dipendente dal sistema operativo

Ho dunque fatto il download di SQlite e mi sono connesso a Matlab attraverso il comando:

dbfile = "C:\Users\munar\OneDrive\Documenti\MATLAB\Examples\R2023a\bioinfo\ilmnsolexademo\test\_db.db";

conn = sqlite(dbfile);

per finire ho utilizzato i vari costrutti del linguaggio SQL per formare tabelle, campi, inserire valori e utilizzare widget.