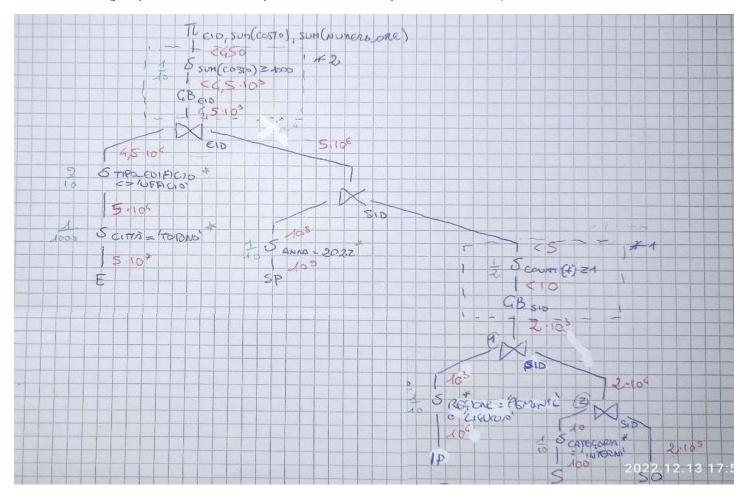
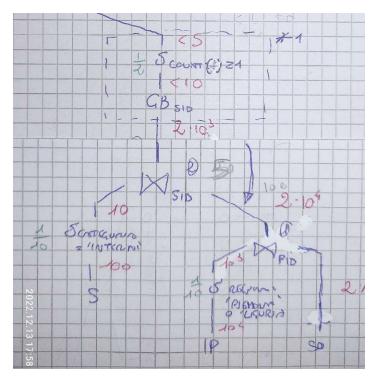
1)

Schema iniziale (join prima tra cardinalità più alta fra le tre con più bassa tra le tre)



PER POTER SPOSTARE GB(SID) occorre invertire l'ordine di join della query interna (figura2)



NOTE:

*ipotesi di distribuzione uniforme

Predicato COUNT>=1 non può avere selettività $\frac{1}{2}$ in quanto ogni tupla deve comparire almeno una volta \rightarrow seleziona tutte le tuple che compaiono, nei conti sarà considerata selettività $\frac{1}{2}$ comunque.

ANTICIPO GB:

GB *1:

Immagine1:

Impossibile anticipare senza aggiungere altri attributi a GB poiché c'è un JOIN su PID mentre la GB è su SID;

Immagine2:

Inutile anticipare a SX: raggruppamento su primary key SID;

Possibile anticipo a DX, considerazioni:

Immagine 1: Immagine 2:

costo JOIN: S-SO \rightarrow 10 join 2*10^5 NL costo JOIN: IP-SO \rightarrow 10^3 join 2*10^5 HJ SSO-IP \rightarrow 10^3 join 2*10^4 HJ

IPSO-S→10 join 50 NL/(HJ)
Costo GB: 2*10^3 GB HASH

Costo GB: 2*10^4 HASH → costoso ma posso tenere

gli hash per JOIN IPSO-S, ininfluente perché sono 50

costoTOT: calcolo circa 2*10^4 HASH costoTOT: calcolo circa 2*10^5 HASH

→ probabilmente conviene Immagine 1 NO ANTICIPO GB *1

GB *2:

Inutile anticipare a SX: raggruppamento su primary key EID

Possibile anticipo a DX, considerazioni:

Se anticipo eseguo GB su 5*(10^6) attributi, considerando una distribuzione uniforme dei dati e considerando il fatto che gli EID distinti sono 5*(10^7) → non ottengo riduzione significativa di cardinalità → nel join su EID avrò circa 5*(10^5) attributi nella OUTER TABLE: considero il join (HASH) con selettività 9/10000(circa 1/1000) "più selettivo" della GB in quanto GB(EID) su 5*(10^6) attributi che possono avere 5*(10^7) valori distinti facendo un ipotesi di distribuzione uniforme comporta una selettività di circa 1-> la selettività totale di GB+HAVING sarà circa 1/10 di 5*(10^6)=5*(10^5) e quindi una selettivita su EID di 1/100

→Se HASH JOIN è molto più dispendioso di GB HASH (almeno 10 volte più dispendioso*)

→probabilmente NON conviene ANTICIPO GB *2

^{*}bisognerebbe sapere quante tuple ci sono in ogni hash bucket: se mediamente sono più di 3 conviene anticipare GB, io considero che non ce ne siano più di 3 in questo caso

Indici:

CREATE INDEX MyIndex1 ON EDIFICO(Città); → meglio HASH

CREATE INDEX MyIndex2 ON SERVIZI-PULIZIA(Data); → meglio B+tree (eventualmente clustered)

CREATE INDEX MyIndex3 ON IMPRESA-PULIZIE(Regione); → meglio HASH

SERVIZI-OFFERTI ha già gli indici sugli attributi della primary key →NO INDICI

SERVIZIO è molto piccola → NO INDICI

Nessuno degli indici definiti da utente aiuta in GB ma aiutano nelle SELEZIONI

NO indice composto su EDIFICIO(Città,TipoEdificio) perché dovrei indicizzare 5*10^7 valori con anche TipoEdificio per poi selezionarne i 9/10→No indici coprenti

QUERY DEFINITIVA IN ALGEBRA RELAZIONALE (tiene conto di costi di join e GB)

