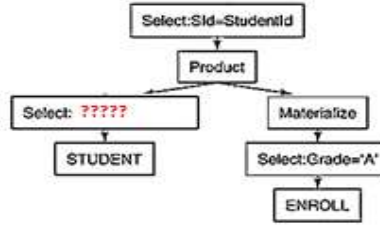


VTSG EV ÖDEVİ #4

- Her soru eşit puandır.
- Sorularda; derste kullandığımız, sunumlarda bulabileceğiniz ÖĞRENCİ KAYIT VT tablolarına ait katalog istatistik bilgilerini kullanınız.



1) Yukarıdaki sorgu ağacında Product düğümünün sağ alt ağacı somutlaştırılmıştır. Bu operasyon her zaman fayda sağlamayabilir. Bu somutlaştırma düğümünü şekilde verilen planda kullanışlı olması için ??? ile gösterilen filtre ne (ler) olabilir ?

2)

	P(b,e)	R(b,c,d)	S(b,a)
R(.)	5000	2000	1000
V(.,a)			100
V(.,b)	200	50	20
V(.,c)		100	
V(.,d)		40	
V(.,e)	50		

T1: naturaljoin (P,R) // P ve R tabloları natural join yapılıyor. Yani join yüklemi $P.b=R.b$

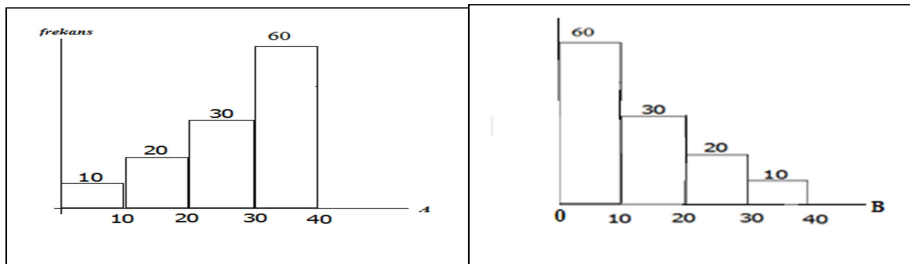
T2: naturaljoin (T1,S) // T1 ve S tabloları natural join yapılıyor. Yani join yüklemi $T1.b=S.b$

T3: select (T2, a=3 AND e=9)

T4: project (T3, {b})

- Yukarıdaki verilen işlemleri verildiği sırada sorgu tarama ağacı çiziniz. Ağaçtaki naturaljoin işlemlerine karşılık gelen düğümlere sırayla (**NJ1,NJ2, ...**) yazmanız yeterli, yüklemi yazmayın.
- Bu sorgu ağacını alan bir planlayıcı **SEZGİ 1-5 ve 6A** yı kullanarak **AĞAÇ KESTİRİMİ** yapmaktadır. Ortaya çıkaracağı ağacı çiziniz. Hangi hesaplarla bu ağacı buldunuz, **4işlem/hesap** işlemi açıklamadan yazınız.
- Önceki şıkta (b şıkkı), planlayıcının bulduğu ağacın, **tahmini** ağaç maliyetini ve ağacın en son ortaya çıkaracağı **tahmini** kayıt sayısını (**R**)'yi bulunuz.
 - Ağaç maliyeti hesabında, join maliyet kestirimi için sunumdaki yaklaşımı kullanın: giriş kayıt sayılarının toplamı
 - tahminlerde Düzenli dağılım esas alınmaktadır.
 - 4işlem/hesap gerektiren yerlerde** işlemi açıklamadan yazınız. (mesela: sonuç 40 ise ve bu değer $120/\max(3,2)$ ile bulunuyorsa; $120/\max(3,2)$ yazmanız gerekiyor.)

3)



T tablosunun **A ve B** niteliğine ait katalogda tutulan histogram bilgileri yukarıdaki graflar ile temsil edilmiştir. Yani; mesela “T tablosunda A değerinin 0-10 arasında olduğu 10 kayıt var; A değerinin 10-20 arasında olduğu 20 kayıt var” diyebiliriz. Bunun dışındaki gerekli bilgiler için, tahminler düzenli (uniform) dağılım esaslı yapılmaktadır. Aşağıdaki sorularda; **4işlem/hesap gerektiren yerlerde** işlemi açıklamadan yazınız. (mesela: sonuç 40 ise ve bu değer $120/\max(3,2)$ ile bulunuyorsa; $120/\max(3,2)$ yazmanız gerekiyor.)

- T1: **Select** $A=5$ (T) sorgu sonucu için;
 - Bu planlayıcı kaç kayıt tahmin eder; **R(T1) =?**
 - Eğer bu histogram bilgisi olmasaydı; planlayıcı **T1** sorgu sonucu için kaç kayıt tahmin ederdi; **R(T1) =?**
- T2: **T** $\bowtie_{A=B}$ **T** sorgu sonucu için;
 - Bu planlayıcı kaç kayıt tahmin eder; **R(T2) =?**
 - Eğer bu histogram bilgisi olmasaydı; planlayıcı **T2** sorgu sonucu için kaç kayıt tahmin ederdi; **R(T2) =?**