**Practica Calificada 2**

Profesor Heider Sanchez

**Alumno:**

**Instrucciones:**

* La prueba dura 100 minutos.
* Resuelva los ejercicios en un documento diferente.
* No copie los enunciados, solo lo necesario para resolver el ejercicio.
* Mantenga el orden de los enunciados en el documento de resolución.
* Queda prohibido cualquier comunicación entre alumnos.
* **Está prohibido revisar cualquier material de referencia.**
* **Sino sabe la respuesta, deje la pregunta en blanco.**

1. **(9 puntos) Optimización:**

Dada la siguiente tabla:

**TNotas**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Id** | **CodAlumno** | **CodCurso** | **CodRubrica** | **Nota** | **Fecha** |

* Asumir que la tabla tiene una cantidad 12000 tuplas y se distribuye equitativamente en cada partición que se genere.
* La tabla esta fragmentado en el **año de la fecha** con el vector de partición

[2000, 2010, 2018].

* CodRubrica es un atributo categórico y tiene las siguientes opciones {EC, PR, EP, EF}
* Cada fragmento está asignado a un servidor diferente.
* Dada la siguiente consulta de usuario:

Select CodRubrica, Count(Id) as Cont, AVG(Nota) as Prom

From TNotas

Where Extract(Month from Fecha) > 2005

Group By CodRubrica

Having Count(id) >= 5;

1. (4 pts) Diseñar el **algoritmo distribuido optimizado** mostrando gráficamente cada fase del algoritmo, y aloje las particiones intermedias en los sitios más apropiados con el fin de reducir la transmisión de datos.
2. (3 pts) Mostrar la sentencia SQL que se deriva del algoritmo distribuido.
3. (2 pts) ¿Cuántas tuplas se transmiten por la RED? Justifique su respuesta haciendo los cálculos legibles en el diseño.
4. **(7 puntos) Fragmentación Y Optimización:**

**Cliente**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DNI** | **Nombre** | **FechaRegistro** | **Ciudad** | **TotalCompras** |

**Compra**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Codigo** | **DNICliente** | **FechaRegistro** | **CodTienda** | **Monto** | **TipoPago** |

1. (3 pts) Sabiendo que la tabla Cliente esta fragmentado en el atributo ciudad (una partición por cada ciudad), se le pide realizar la fragmentación horizontal derivada en la tabla Compra. Considere dos ciudades: Lima, Trujillo.

Notación de la Fragmentación:

Equivalente en SQL:

1. (4 pts) En base a la fragmentación anterior diseñe el algoritmo distribuido optimizado para la siguiente consulta sabiendo que produce una gran cantidad de resultados imposible de procesar en el servidor que coordina la query. Asuma que contamos con 3 servidores.

**SELECT** CO.\* **FROM** Compra CO

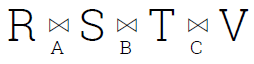
**INNER JOIN** Cliente CL **ON** CL.DNI = CO.DNICliente

**WHERE** CL.TotalCompras > 20

**ORDER BY** CO.FechaRegistro

1. **(4 puntos) Optimización y costos:**

Dada la siguiente consulta y la cantidad de tuplas por cada relación:

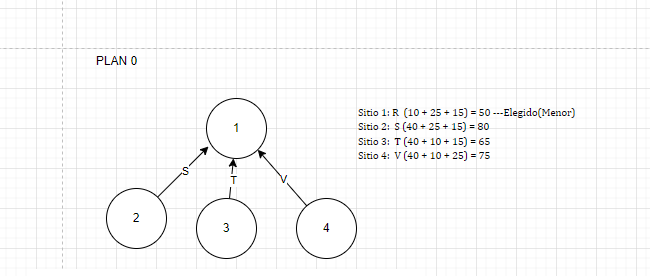


|  |  |  |
| --- | --- | --- |
| Tabla | Sitio | Tamaño |
| R | 1 | 40 |
| S | 2 | 10 |
| T | 3 | 25 |
| V | 4 | 15 |

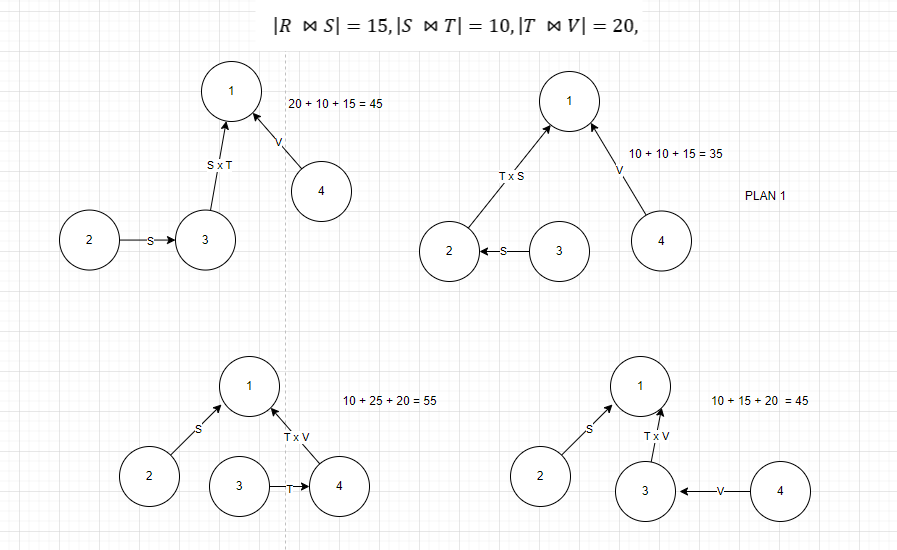
Además, considere la siguiente estimación:

Usando el método de Hill Climbing se le pide el mejor plan de transmisión de datos paso a paso. Anota las tuplas de transmisión en cada arista.

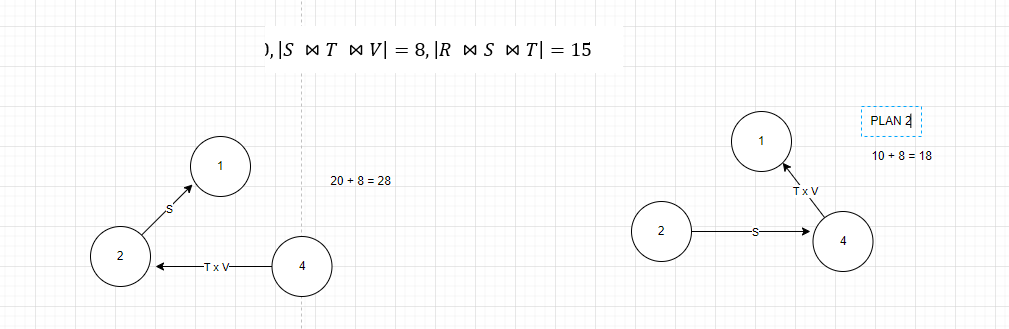
P0 (Plan Inicial):



P1 (Segundo mejor plan):



P2 (Tercer mejor plan)



Pn (Plan final):