Compsci 557 Homework 7

Kayla Goetzke

Phil Sauvey

Chapter 12 # 1, 2, & 3

1)

2) The company's database use Range partitioning based on a partition key, and that key should be the state/region code of a customer

3a) I would recommend using a homogeneous Multiple site processing, multiple-site data DBMS with the sites being located in the 4 states, FL, SC, GA, and TN. I would also recommend implementing a partially replicated database to ensure data integrity in the event of a node going down, while avoiding the costly overhead of a fully replicated DB.

Since there are several times when data must be collected from all locations, for ease of use I would suggest a DBMS that implements fragmentation level transparency, and distributed requests. It should also provide Replica Transparency, because the user does not need to know about the partial replication of data.

3b) Each table should be horizontally fragmented using range partitioning based on a partition key.

3c)

The partitioning for the invoices is easy and can be done on the INV\_REGION attribute as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Fragment | Location | Key | Node Name |
| INV1 | Florida | INV\_REGION = "FL" | FLA |
| INV2 | South Carolina | INV\_REGION = "SC" | SCA |
| INV3 | Georgia | INV\_REGION = "GA" | GEO |
| INV4 | Tennessee | INV\_REGION = "TN" | TEN |

The partitioning for the customer becomes a little trickier since according to the system description in 2c, we are not storing a user's state. There are two ways I can see to partition, the first which seems inefficient and potentially error prone is looking up a customer's corresponding invoice and getting the region from there. The second, is to use the zip code as a form key to determine the state of the customer (this is implemented below assuming the zip is stored as a number and not a string, if stored as a string, like statements and wild cards can be used to produce a similar effect )

|  |  |  |  |
| --- | --- | --- | --- |
| Fragment | Location | Key | Node Name |
| CUS1 | Florida | CUS\_ZIP BETWEEN 32000 AND 33999 OR BETWEEN 34100 AND 34999 | FLA |
| CUS2 | South Carolina | CUS\_ZIP BETWEEN 29000 AND 29999 | SCA |
| CUS3 | Georgia | CUS\_ZIP BETWEEN 30000 AND 31999 OR BETWEEN 39800 AND 39999 | GEO |
| CUS4 | Tennessee | CUS\_ZIP BETWEEN 37000 AND 38599 | TEN |

3d)

Customer Fragments:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Fragment: CUS1 | | Location: Florida | | Node: FLA | |
| CUS\_NUM | **CUS\_NAME** | **CUS ADDRESS** | **CUS\_CITY** | **CUS\_ZIP** | **CUS\_SUBSDATE** |
| 12345 | Snow While | 123 Castle Way | Jacksonville | 32205 | 16-mar-2012 |
| 13012 | Bigby Wolf | 8654 Forrest Rd | Miami | 33130 | 20-may-2013 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Fragment: CUS2 | | Location: South Carolina | | Node: SCA | |
| CUS\_NUM | **CUS\_NAME** | **CUS ADDRESS** | **CUS\_CITY** | **CUS\_ZIP** | **CUS\_SUBSDATE** |
| 10011 | Jack Horner | 13 Beanstalk St | Charleston | 29413 | 9-jun-2007 |
| 13013 | Boy Blue | 321 Music Lane Apt 312 | Columbia | 29203 | 21-may-2013 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Fragment: CUS3 | | Location: Georgia | | Node: GEO | |
| CUS\_NUM | **CUS\_NAME** | **CUS ADDRESS** | **CUS\_CITY** | **CUS\_ZIP** | **CUS\_SUBSDATE** |
| 4156 | Prince Charming | 1 Palace Blvd | Atlanta | 30310 | 6-jul-2000 |
| 254583 | Rose Red | 6538 Shadow St | Savannah | 31412 | 5-jan-2017 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Fragment: CUS4 | | Location: Tennessee | | Node: TEN | |
| CUS\_NUM | **CUS\_NAME** | **CUS ADDRESS** | **CUS\_CITY** | **CUS\_ZIP** | **CUS\_SUBSDATE** |
| 196753 | King Cole | 73 Jolly Old Way | Nashville | 37115 | 25-dec-2016 |
| 300000 | Pinocchio | 4686 Liars Lane | Memphis | 38112 | 18-mar-2018 |

Invoice Fragments:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fragment: INV1 | Location: Florida | | Node: FLA | |
| INV\_NUM | **INV\_REGION** | **CUS\_NUM** | **INV\_DATE** | **INV\_TOTAL** |
| 65465 | FL | 12345 | 16-mar-2012 | 65.99 |
| 321654 | FL | 13012 | 20-may-2013 | 12.85 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fragment: INV2 | Location: South Carolina | | Node: SCA | |
| INV\_NUM | **INV\_REGION** | **CUS\_NUM** | **INV\_DATE** | **INV\_TOTAL** |
| 63501 | SC | 10011 | 9-jun-2007 | 101.99 |
| 324840 | SC | 13013 | 21-may-2013 | 5.99 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fragment: INV3 | Location: Georgia | | Node: GEO | |
| INV\_NUM | **INV\_REGION** | **CUS\_NUM** | **INV\_DATE** | **INV\_TOTAL** |
| 8012 | GA | 4156 | 6-jul-2000 | 49.99 |
| 450362 | GA | 254583 | 5-jan-2017 | 39.99 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fragment: INV4 | Location: Tennessee | | Node: TEN | |
| INV\_NUM | **INV\_REGION** | **CUS\_NUM** | **INV\_DATE** | **INV\_TOTAL** |
| 450189 | TN | 196753 | 25-dec-2016 | 1.99 |
| 532001 | TN | 300000 | 18-mar-2018 | 42.00 |

3e) In theory, assuming all nodes are operational, a remote site should only ever need data stored locally, so no distributed database operations are needed at the remote sites.

3f) At headquarters, company management needs to be able to access the data from all 4 nodes so it must support distributed requests, and in the event of a loss of data at a node management/IT must be able to restore data from a backup or one of the other nodes where replicated data is being stored so the DBMS should also be able to handle Remote Transaction Requests, though these are not specifically required as part of the system description in the problem statement.