COMP0022: Database and information systems (PG – Term 1)

Distributed databases exercises

1. Using the 'Shop' database: This database consists of the following tables:
   Customers (custID, firstName, familyName, town, state, [country])
   Items (itemID, description, unitcost, stocklevel)
   Orders (orderID, custID, date, shippingAddr, paymentInfo)
   Lineitems (orderID, itemID, quantity, despatched)
   Supplieritems (supplierID, itemID)
   Suppliers (supplierID, supplierName, supplierAddress, phoneNo)

   We would like to distribute the shopping database.

   a. Assume that, currently, this shopping website is available in 5 different countries. There exist offices in each country for managing the item catalog and customer inquiries for items sold/customers registered in that country. Each office works with suppliers to add/update items in the catalogue. Some suppliers only sell items in one country while others sell an item in multiple countries, but often versions specific to that country (so would be a new entry in the Items table).

      Propose a scheme for fragmentation, allocation, and replication for a distributed Shop database.

      Note: These questions are very open-ended and there are many ways to design this.

      In general, you want to fragment geographically, and possible replicate for things that are used by everyone.

      Customers: split into 5 fragments on country (horizontal fragmentation).
      Items: possibly add country information and use it to horizontally fragment.
      Orders: associate with site of originating customer (derived horizontal fragmentation using custID).
      Lineitems: use orderID to do the same (derived horizontal fragmentation).
      Supplieritems: use itemID to do the same with Items' country information. (Or store full table everywhere.)
      Suppliers: store full table everywhere. (Or fragment.)

   b. Assume instead that this shopping website is only available in one country. All customer purchases go through a central server located at Site 1. Site 2 is an analysis division that looks at all customer, order, and item information and produces reports. Site 3 is a customer service office that needs information on customers and orders to address customer complaints. Site 4 handles information about suppliers and stock.

The storage-heavy but local way to do this is to store a copy of data at each site that might use it.

Site 1: Contains a full copy of Customers, Items, Orders, Lineitems.
Site 2: Contains a copy of Customers, Items, Orders, Lineitems, perhaps with personal information removed (name, shipping and payment information – vertical fragmentation). As this site is for analysis, it may not be important for this copy to be fully up to date.
Site 3: Contains a full copy of Customers, Orders, and Lineitems, and possibly Items. It is more important to keep these copies up to date.
Site 4: Contains a full copy of Items, Supplieritems, and Suppliers.

2. How would each of the proposed distributed systems handle the following operations?
   These answers are specific to the designs laid out above. Other answers are possible.

   a. Process a new order from a single customer.
      First DB: Full order is processed and stored at the site of the customer's country. All work is local and no other sites are involved.
      Second DB: The request comes in via Site 1, which needs to update all other copies as well.

   b. Update information about a supplier.
      First DB: Need to update information at all sites.
      Second DB: Update occurs at Site 4.

   c. Update item prices to new ones provided by a supplier.
      First DB: Need to decompose updates by country and send subqueries to each site.
      Second DB: Every site storing the Items table must be updated.