First Name:	Gagandeep
Last Name:	Kaur
Student ID:	1121869

ICT705 Data & Systems Integration

USC Semester 1, 2020

Final Examination

School of Business

Examination Duration: Submit your exam by 11:30am

Exam Conditions:

This is an open book exam.

Instructions to Students:

- Write your answers after each question. Please do not change the order of the questions.
- The value of each question is as shown.
- Your submission is to be made as a single Microsoft Word document – no other format is acceptable.
- Please rename this file to "*ICT705 Sem 1 2020 Task 3 Exam_**FirstName_LastName_ID**.docx*" for submission.
- Answer all questions and submit your Microsoft Word document on Blackboard in the Assessment area (Task 3) no later than **11:30am 15th June 2020**

Submission Declaration:

By submitting this assessment item, you declare that your submission is your own work and is in accordance with the University's Student Academic Integrity Policy.
Instructions to Students:

This is an individual assessment; all submissions will go through SafeAssign.

Questions (Total 40 Marks)

Answer all questions in this section **in your own words** (do not copy and paste text from sources) in the space provided below the questions. Please note space is not indicative of the length of answer required. As a guide you should attempt to answer all questions. Use the marks for each question as a guide as to the length / wordcount of your response.

- **[7 marks]** Acme Industries are looking to develop a distributed cloud-based solution for their transactions processing system, incorporating customer ordering and inventory control. Discuss CAP theorem in relation to this and explain how this might impact their choice of a Database Management System.

  Answer:

  CAP theorem applied on distributed systems that ensure system can have two of these three features: Consistency, Availability or Partition Tolerance. Only two guarantees are given to shared-data system.

  Consistency: it refers to state of distributed data system such that all system have same copy of the data, which means if data is updated in one system readers from other system will see the updated version of data. Data is same in all system all the time. Acme industry will have same inventory list every time a person buys an item, its stock is subtracted and updated in all servers across the distributed cloud.

  Availability: it means that if a user is requesting for a product description he will get the query responder every time regardless of crashes or downtimes. A valid response is given to the customer regardless of failure in system nodes. Data is available to access every time.

  Partition Tolerance: in this even if two or more servers get isolated the system operated with any delays or compromises. Partition basically mean breaking of communication between to connected nodes in the system. Partition tolerance

is despite the communication partition system continues to work.

Selection of Consistency and Partition Tolerance will allow NoSql/columns, Hadoop, Bigdata.

Selection of Consistency and Avalibility will allow RDBMS/SQL, MySql, Oracle.

Selection of Avalibilty and Partition Tolerance will allow NoSql/Document or Key-values, casandra, Voldemot.

- **[4 marks]** Compare and contrast the different Enterprise Application Integration (EAI) methods.

  EAI refers to unrestricted sharing of data and business process between any applications that are connected.

  Tere are four types of EAI:

  Data Level: in this data is shared form data stores. Data from one data store is extracted and updated in another. It is commonly done through ETL tools. Benefits of this is low cost and low risk profile because no modification is done on application code.

  Application Interface Level: this uses application interfaces to gain access to the business processes and information. It is implemented through Message Brokers; this standardize and control information flow through bus or hub. It is making interfaces between applications relatively easy with use of XML.

  Method Level: applications can access methods on an other application available. It is similar to application interface level but at a lower level of granularity. It is implemented through Web services, Distributed objects. It can beused to update customer data from different applications.

  User interface Level: it combines applications by using their user inrtface which is also known as screen scraping.

- **[6 marks]** Explain why schema matching poses significant challenges for Enterprise Information Integration (EII). In your discussion give examples of

specific challenges.

Challenges in schema matching are first same attributes are represented differently in different databases. We must identify them and change attribute name to combine. Example : address in one database can be split into different attributes such as post code, street address or state etc.

Second, fundamental conflict that means zip code in one database is string and integer in another resulting merged database can not have both.

Third merging models with different styles such as one data base cold support sub-attributes while others do not.

- **[5 marks]** Explain the purpose of wrappers in data integration and distributed query processing and give an example of each type of construction approach.

  Distributed query processing: data is distributed across multiple machines and query is processed to gain data from multiple sources

  Parallel DBMS, in this data is assumed to have homogenous nodes and network is fast. Goal of is this is to efficiently utilize all resources and balance load.

  Distributed DBMS has heterogenous nodes slow network and some sources are available to selected machines. Main purpose is to determine what queries should be used to compute at different places. This is main concern in distributed query processing.

  Example: Adaptive query processing can be used.

  Wrappers: these are components of Data integration system that communicate with the data sources. It is used to send queries from high level system to the source.

  It converts replies to format that can be manipulated by query processor. In many cases wrapper uses semi-structure data such as HTML page and transform it into set of tuples.Wrapper can built custom program that cange format from one tupe to another.

  Example of wrapper construction approach: Manual, learning, interactive,

automatic.

- **[3 marks]** An Enterprise Service Bus (ESB) provides security, monitoring and management functions. Name and describe three other core functions of an ESB.

  Protocol conversion: ESB accept messages sent in major protocols, and convert them to the format of end consumer

  Message transformation: ESB Convert messages into a format that is usable by the consumer application

  Message routing: ESB determine the appropriate end consumer based on rules and requests

- **[5 marks]** An organisation has approached you about using a data warehouse. Discuss what you would tell them a data warehouse is and why they should or should not adopt one.

  Data warehouse is repository (archive) of information gathered from various sources, stored under a unified schema, at a single site.

  Various data warehouse components are Data Extraction and Loading, Warehouse, Metadata, OLAP Tools, Data Mining tools

  Advantages of using Data warehouse:

  - Consolidation o information resources that mean various sources of information are available at single place.
  - Improved query and system performance: It simplify querying, permits study of previous and historic events. It helps in moving decision support query load away from transaction processing systems
  - It gives a foundation data mining, data visualization, advanced reporting and OLAP tools

  Features of warehouse

  - they are subject oriented data is organized how user wants.
  - They are integrated , inconsistencies regarding naming conventions

and values are removed

- Data is depend on time series and is not current

- It is non-volatile data is stored in read-only format so it can not be changed.

- **[6 marks]** Two organisations have recently merged, and between them they have several systems. Explain how an Enterprise Information Integration can help. In your discussion give examples of 6 types of tasks that can be used.

  EII tasks that are used are following:

  - Understanding the data which include:

    Discovering: getting data from various sources , focus on metadata

    Profiling: calculating quality of data received.

  - Processing data which include:

    Cleaning: removing missing values and undefined data

    Transformation: data type conversion, changing data tpe as it is moved from source to target system

    Replication: making copies of data from one system to another with hardware or software techniques such as RAID, load/unload, export, log shipping.

    Federation: providing single view for different systems

    Steaming: on demand resource allocation, managing data in real time

  - Deploying the services.

EII can help by allowing data to be stored in various formats scuch as RDBMS,NOSQL etc.

- **[4 marks]** In the table below, highlight (circle) what the outcome of the OLAP

operation is on the sample data.

<table>
<tr><th>OLAP operation</th><th>Sample Data</th></tr>
<tr><td>Roll-up</td><td>
<table>
<tr><th>C / S</th><th>S</th><th>M</th><th>L</th><th>TOT</th></tr>
<tr><td>Red</td><td>20</td><td>3</td><td>5</td><td>28</td></tr>
<tr><td>Blue</td><td>3</td><td>3</td><td>8</td><td>14</td></tr>
<tr><td>Gray</td><td>0</td><td>0</td><td>5</td><td>5</td></tr>
<tr><td>TOT</td><td>23</td><td>6</td><td>18</td><td>47</td></tr>
</table>
</td></tr>
<tr><td>Drill-down</td><td>
<table>
<tr><th>C / S</th><th>S</th><th>M</th><th>L</th><th>TOT</th></tr>
<tr><td>Red</td><td>20</td><td>3</td><td>5</td><td>28</td></tr>
<tr><td>Blue</td><td>3</td><td>3</td><td>8</td><td>14</td></tr>
<tr><td>Gray</td><td>0</td><td>0</td><td>5</td><td>5</td></tr>
<tr><td>TOT</td><td>23</td><td>6</td><td>18</td><td>47</td></tr>
</table>
</td></tr>
<tr><td>Slice</td><td>
<table>
<tr><th>C / S</th><th>S</th><th>M</th><th>L</th><th>TOT</th></tr>
<tr><td>Red</td><td>20</td><td>3</td><td>5</td><td>28</td></tr>
<tr><td>Blue</td><td>3</td><td>3</td><td>8</td><td>14</td></tr>
<tr><td>Gray</td><td>0</td><td>0</td><td>5</td><td>5</td></tr>
<tr><td>TOT</td><td>23</td><td>6</td><td>18</td><td>47</td></tr>
</table>
</td></tr>
<tr><td>Dice</td><td>
<table>
<tr><th>C / S</th><th>S</th><th>M</th><th>L</th><th>TOT</th></tr>
<tr><td>Red</td><td>20</td><td>3</td><td>5</td><td>28</td></tr>
<tr><td>Blue</td><td>3</td><td>3</td><td>8</td><td>14</td></tr>
<tr><td>Gray</td><td>0</td><td>0</td><td>5</td><td>5</td></tr>
<tr><td>TOT</td><td>23</td><td>6</td><td>18</td><td>47</td></tr>
</table>
</td></tr>
</table>

END OF EXAMINATION