



Minor-1 Answer Outlines
Department of Computer Science and Engineering
Indian Institute of Technology Jodhpur, Rajasthan, India 342030
Course: CSL4030 Data Engineering (July-December, 2023)
Duration: 1 hour
Marks: 32

Answer 1. (BCNF)

STUDENT (RegNo, Fname, Lname, DeptID, Active(Y/N))
ATTENDANCE (RegNo, CourseID, Date, Present(Y/N))
COURSE (CourseID, CourseName)
DEPT (DeptID, DeptName)
ROLLNO (RegNo, OldRollNo, NewRollNo = NULL)

Answer 2.

2PC can handle site failures and network partitions.

ACAD	ME	CSE
<p>Phase 1: ACAD initiates transaction T_i and TC_{ACAD} starts the execution. TC_{ACAD} breaks the transaction into three sub-transactions and distributes them to appropriate sites along with a "prepare T_i" message.</p> <p>TM_{ACAD} adds $\langle \text{prepare } T_i \rangle$ to its log and executes the following sub-transaction T_{i1}: lock(007); read(007); update(007); TM_{ACAD} logs $\langle \text{ready } T_i \rangle$ and sends a "ready T_i" message to TC_{ACAD}. If T_{i1} fails, TM_{ACAD} logs $\langle \text{no } T_i \rangle$ and sends an "abort T_i" message to TC_{ACAD}.</p>	<p>TM_{ME} logs $\langle \text{prepare } T_i \rangle$ and executes the following sub-transaction T_{i2}: lock(007); read(007); update(007); TM_{ME} logs $\langle \text{ready } T_i \rangle$ and sends a "ready T_i" message to TC_{ACAD}. If T_{i2} fails, TM_{ME} logs $\langle \text{no } T_i \rangle$ and sends an "abort T_i" message to TC_{ACAD}.</p>	<p>TM_{CSE} logs $\langle \text{prepare } T_i \rangle$ and executes the following sub-transaction T_{i3}: lock(007); read(007); update(007); TM_{CSE} logs $\langle \text{ready } T_i \rangle$ and sends a "ready T_i" message to TC_{ACAD}. If T_{i3} fails, TM_{CSE} logs $\langle \text{no } T_i \rangle$ and sends an "abort T_i" message to TC_{ACAD}.</p>

<p>Phase 2: If and only if TC_{ACAD} receives a “ready T_i” message from every TM before the timeout (ready state), TC_{ACAD} sends a “commit T_i” message to all TMs. Otherwise, TC_{ACAD} sends an “abort T_i” message to all TMs.</p> <p>TM_{ACAD} adds $\langle \text{commit } T_i \rangle$ or $\langle \text{abort } T_i \rangle$ to its log, and commits/rolls back its T_{i1}.</p> <p>TM_{ACAD} sends an “acknowledge T_i” message to TC_{ACAD}. unlock(007);</p> <p>If TC_{ACAD} receives the “acknowledge T_i” messages from all TMs before timeout, it logs $\langle \text{complete } T_i \rangle$.</p> <p>ACAD shows the “Successfully updated” or “Failed. Try again.” message to the Office of Academics user.</p>	<p>TM_{ME} adds $\langle \text{commit } T_i \rangle$ or $\langle \text{abort } T_i \rangle$ to its log, and commits/rolls back its T_{i2}.</p> <p>TM_{ME} sends an “acknowledge T_i” message to TC_{ACAD}. unlock(007);</p>	<p>TM_{CSE} adds $\langle \text{commit } T_i \rangle$ or $\langle \text{abort } T_i \rangle$ to its log, and commits/rolls back its T_{i3}.</p> <p>TM_{CSE} sends an “acknowledge T_i” message to TC_{ACAD}. unlock(007);</p>
--	--	--

Answer 3.

(Detailed features of the ROOT can be found in: Fons Rademakers, ‘ROOT for Big Data Analysis’, presented at the Workshop on the Future of Big Data Management, London, UK, June, 2013. <https://indico.cern.ch/event/246453/contributions/1566610/attachments/423154/587535/ROOT-BigData-Analysis-London-2013.pdf>)

The key responsibilities of the data engineers at ROOT is to make sure the data can be efficiently stored and queried.

The ROOT data does qualify as a big data because of its volume and velocity (hundreds of petabytes of data being continuously generated by the LHC sensors) as well as variety (numerous sensors collecting numerical data, researchers generating image plots and videos for visualization, subsequently discussing about the results on their internal forums which are accumulating text data).

We might need to combine all three data models -- relational, document-based, graph-based – in order to build such a complex system. Numerical transactional data from sensors that is constantly being accumulated can be saved as relational tables. The results, reports, images, and videos can be linked to each other in a graph database. The self-containing documents such as user profiles on the internal forums can be stored in a document database.