[Music]
[Applause]
[Music]
hi everyone myself navat RI working as assistant professor in the department of cyber security and data science in mlr instit of technology today I'm going to deal with the topic schema refinement and the overview of presentation includes about schema refinement that is the definition of schema refinement and why what is the purpose of schema refinement problems caused by data redundency
decompositions and problems related to decompositions let us discuss about schema refinement first so as we have already discussed about schema it's nothing but the design of a database what is refinement we used to refine the database
so here let us discuss one relation which consists of S ID and S name we can call this as a schema that is the design of the data here I didn't mention any kind of data in it that means whatever the records in this table I didn't mention it's just a design that means which represents the attributes or Columns of the particular data that is what we call as schema here how we can able to refine that one without any redundancies so in the context of dbms what is a redundancy nothing but the duplication of data for example let us say the ideas one name as XY
Z IDE as two the name as XY Z so here we have the same names this is what we called as
reny duplication of data so sometimes for example the IDS General sense the ID may not be uh duplicated why because for each and every person will have unique ID in the same manner each and every student will have unique role number so if at all if at all here the the same ID has been given to or else by Chance the same ID by mistake the same ID has been given to two different persons then it is called as rden
that is nothing but duplication of data

how we could able to refine this kind of data right so that is what about the schema
[Music]
refinement so let us discuss this one in detail so it is a technique of all organizing the data in a database that is have to have to maintain the data in a database without any duplications that is what the schema refinement is meant for it refers to refine the schema by using some techniques and one of the best techniques is decomposition how we can refine that data how we can refine the data by using some methodologies or techniques one of those techniques is nothing but
decomposition one of those techniques is nothing but decomposition what is decomposition uh splitting a table into number of small subtes that is what we call as decomposition we'll discuss this topic in detail in further slides the next point is it is nothing but the systematic approach of decomposing tables to eliminate data redundancy so what is the need of decomposition what is the need of decomposing the given table into subtables because to remove redundancy here we are using this technique what is the purpose of removing redundency to maintain database in an in in an consistent manner that is what the thing and undesirable characteristics like insertion updation and deletion anomalies so here we have uh we have to know about anomalies anomaly is nothing but a problem caus it to the database while inserting the data or updating the data or deleting the data why because we used to interact with the database with these three operations only right while inserting while updating and while deleting so by doing these operations some problems may occur those are nothing but the anomalies here we have three kinds of anomalies so what are those insertion anomaly deltion anomaly and updation anomaly let us discuss about anomalies in detail so what is the problems what

are the problems caused by here data redundancy insertion anomaly deltion anomaly and updation anomaly as I said what is an anomaly in simple words it is nothing but a problem so problem while doing insertion operation problem while doing deletion operation and problem while doing updation operation those are called as anomalies right so here what is an anomaly it refers to inconsistencies or errors that occur when manipulating or querying the data in a database that means whatever the operation that we are going to interact with the data datase while doing that operation some problems may occur that is what we called as anomaly so if at all we are inserting a data into a database if at problem occurs then it is called insertion anomaly right so next these anomalies can lead to Incorrect and inconsistent results and can negatively impact the overall functionality of a database and it impacts the accuracy of a database also it's just like nothing but an error so let us discuss each and every anomaly in detail with a simple example so let us discuss insertion anomaly so what is insertion anomaly here if at all I want to insert a new data or a new row or a new record into an existing table I can't insert it
directly and if at all I want to insert that I want to insert that particular row with that along with that row I want to insert some unnecessary data also that is what my intention here so what is here it may not be possible to store some information unless some other information is stored as well let us discuss with this an example here I have considered one table which consists of five columns S ID yes name student ID student name let us discuss this is a student table student ID student name and course ID C ID is nothing but course ID course name and fee so here in this Sid I have some redundancy of data here I took some of the redundancies that is nothing but duplication of even though

we know that in general sense IDs can't be repeated but to explain to explore this concept brancy concept here particularly I going to take the duplication of data in SIDS also right so here SIDS that means S1 has been repeated next S2 next S3 has been repeated here we have redundancies in Sid and the duplication of data in s name the duplication of data in course duplication of data in course name and duplication of data in fee so how we can make this data as consistent which result this is in inconsistent data right because it contains duplication of data we have to remove these duplications right so what is the solution for this let us discuss first the problem then we'll go for solution so here this Sid which contains the S1 S2 and S3 right if at all I want to insert I want to insert some other course new course C4 some other new course C4
and for
example database this is the subject I'm going to uh insert in the in this row 12K will be the fee right so till now no student has been enrolled in this course so I have to uh keep blank for this one Sid and S name but but in case in some cases we used to take Sid as primary key we used to represent Sid as primary key
we know that if at all if at all a column or an attribute has been represented as primary key then it must not be
null and it must not be null and it must contain unique values and it must contain unique values so these are the two constraints that we follow if at all I used to declare Sid as primary key then it must not be null and it must contain values for suppose let us discuss for all I used to consider Sid as primary key and here I have introduced I want to introduce a new course but no student has been enrolled in that course then what I want to do I can't I can't insert the data here why

because primary key if at all I used to represent Sid as primary key then it will restrict what is the Restriction you are inserting null values so it cannot inserted so what is the solution for this here for temporary purpose I have to represent a dummy data for example some this is an ID and ABC this is what we call as dummy data to insert some useful information we need to insert some other dummy data by forcefully so this is what we called as insertion anomaly so what is insertion anomaly here if a new course C4 has been introduced but no student enrolled in this particular course then we have to enter some dummy data forcefully as if Sid is consider as primary key it could not be null that is why we have to insert some other unnecessary data some dummy data which is not very useful not very much useful so the only option is to enter dummy data which may cause problems in future so because of insertion of some data it is forced to insert some other data this is called as insertion insertion anomaly insertion anomaly that means we are facing a problem while doing insertion operation into the database next let us consider deletion anomaly so let us discuss this deletion Anomaly by using the following same table if at all I want to delete yes3 if at all I want to delete yes3 so delete S3 this is my command to the database so what the rows which contains S3 has been deleted this row this row will be deleted right so if at all I used to delete these two rows here some useful information I'm going to delete that means the Java course which I have already there right that also the course itself has been deleted the is not correct our job is to delete the student S3 along with that here the Java course the useful information has also been deleted so to delete one information forcefully we have to delete some other

information that is what we call as deletion anomaly the problem in deletion so it may not be possible to delete some information without losing some other information as well I want to delete S3 student information and along with that I'm going to delete Java course also that is not correct so deletion of S3 student caus the deletion of course Java so because of deletion of some data forc to delete some other useful data so this is the problem that is nothing but division anomaly let us discuss updation anomaly updation anomaly with the same example so what what is updation anomaly here I want to I want to uh update I want to update the fee for example 5K from 7,000 I want to update the key 5K from 5,000 to 7,000 then for each and every row I want to check and I want to update if at all I didn't update I have updated for this row and I forgot to update for this second row then what will happen the data remains inconsistent the data in the database is not correct and accurate why because I have updated in mon row and I forgot to update in the second row this is not correct so if at all I want to update in one row I want to update for all the rows also which results in more IO cost which results in more IO cost that is also a problem that is what we called as updation anomaly so if one copy of repeated data is updated and inconsistency is created unless all copies are similarly updated until and unless all copies are updated then only the data will be consistent otherwise it remains inconsistent if there is updation in fee from 7 5K to 7K the same example which I have discussed now then we have to update fee column in all the rows else the data will be inconsistent so to update the data in all the rows it will be time consuming process and IO cost also will be increased to update all the row which reflects to more IO cost so these are the three different anomalies these are the problems caused

due to maintenance of data redundancies in the database so what is the solution what is the solution of these problems The One and Only Solution is nothing but decomposition so what is decomposition decomposition means the splitting the parent table into small subtes that is what we call as decomposition so here let us discuss the definition of decomposition first C it is nothing but the solution to the problem caused by data
redundancy it means what is the decomposition meant for it means breaking up the large schema large data into smaller multiple schemas it removes all the anomalies and also helps to maintain data Integrity that is what the purpose of decomposition
so we have already discussed about this example the same example how we can able to remove all the anomalies which we have discussed in the earlier slides by using
decomposition right so let us see this one I have decomposed this table into these two subtables and I have removed all the anomalies also how so see here I have splitted the one Common Table into two tables which consists of three columns each that is nothing but this is table a this is table
B so in table a for example here I have considered primary key as primary key as S ID and C ID primary key as S ID and C ID and for this table the primary key as C ID what is a primary key primary key is a key that contains unique values and which does not contain null values that is what primary key so here in this table if if I used to consider only Sid as primary key is that possible no why because if at all only Sid I used to consider only
Sid here Sid may contain the redundancy data that is nothing but S1 S2 S2 is unique and next S3 we have duplication of data so Sid single by considering single Sid we can't take it as primary key why because it has
duplications and in the same manner I

can't consider single C single column c as primary key why because it has also duplication data so by combining two columns both Sid and C the combination of these two can be taken can be preferred as can be preferred as primary key how see here S1
C1 combinely consider as one S2 C1 S1 C2 S3 C2 S3 C3 no repetitions allowed right so we can take these two columns as primary key and next here in this we we can take C ID as primary key single why because it contains no duplication course IDs so we can take C ID so if at all I want to insert I want to insert some C4 course I can insert C4 course here in this course table or if at all I want to insert C5 course C5 for example CF and data structures DS course 12K can insert simply previous table if I want to insert a course see this is the table which has redundancies before decomposition if at all I want to insert a course so for example C4 I want to insert so here I can't I will get some null values why because no student has been enrolled but if at all I decomposed those tables then simply I can insert that course table and courses in this course table without considering any student in
information in that manner I can able to remove insertion anomaly right next if at all I want to delete I want to delete Java I want to delete Java course so here I can simply delete this Java C3 course and no other information has been deleted in previous table if I want to delete S3 student the course will be deleted if I want to delete so see here see here if I want to delete
S3 if I want to delete
S3 so S3 can be deleted directly and here B C2 C3 courses will be delet CS only C IDs will be deleted so if at all I want to delete S3 from here the Java course will be deleted why because only one Java course has been available if at all I used to delete here from here the Java course will be available in this table so right so if at all I used to

delete from this table the information must reside in other table that is what the purpose of decomposition right next if at all I want to update I want to update 5K to 7K so here I can easily update in the course table in this manner by decomposition by decomposing one table into multiple tables not only two two or more multiple tables we can able to we can able to uh modify or we can able to remove the redundancies that is what the beauty of decomposition here and the problems related to decomposition some of the minor problems related to decomposition that's mainly of if at all we used to split more number of tables obviously we must require more memory right so it produces number of tables with a relatively small number of columns these columns then have to be joined using their primary key or Fallen key relationship so if we want to split into different tables to combine or to join those two tables by using primary key or foreign key Concepts which may results in some complexity so it has more uh major two major disadvantages what are those disadvantages performance and complex queries what is performance so here to join or to me all the data to mge the data slow processing and place additional stress on your Hardware in terms of in terms of memory and processing processing time next complex queries developers have to code complex queries in order to merge data from different tables that is nothing but to split of to splitting the tables from a single table is very easy but while merging them it will be very complex why because we have to see the foreign foreign key Concepts that means which table we have to refer and from which table we have to link the data from which table we have to retrieve the data so this might results in some complexity that is what the problems related to decompositions so till now what we have

discussed in this session so
here about schema
refinement that means we used to refine
the schema whatever the schema is there
how we can able to refine that schema
that means to remove to remove
redundancies to remove redundancies so
what is a redundancy nothing but
duplication of data uh what is a
redundancy and what are the problems
caus to due to redundancy we have three
problems we have discussed three
problems what are those three problems
insertion anomaly deletion anomaly and
updation anomaly updation anomaly how we
can able to overcome these three
problems by using one technique that is
what we called as
decomposition and we have discussed some
of the drawbacks of decomposition also
performance like that thank you