Sample SQLSERVER Interview Questions

by

Shivprasad Koirala

Including SQLCLR, XML Integration, Database optimization, Data warehousing, Data mining and reporting services

The Table of contents is different from what is available in traditional books.So rather than reading through the whole book just look at what questions you feel uncomfortable and revise that.

Contents

“Cheers to the true fighting spirit of IT professionals”

Introduction

Dedication

This book is dedicated to my kid Sanjana, whose dad’s play time has been stolen and given to this book. I am thankful to my wife for constantly encouraging me and also to BPB Publication to give new comer a platform to perform. Finally at the top of all thanks to two old eyes my mom and dad for always blessing me. I am blessed to have Raju as my brother who always keeps my momentum moving on.

I am grateful to Bhavnesh Asar who initially conceptualized the idea I believe concept thinking is more important than execution. Tons of thanks to my reviewers whose feedback provided an essential tool to improve my writing capabilities.

Just wanted to point out Miss Kadambari . S. Kadam took all the pain to review for the left outs with out which this book would have never seen the quality light.

About the author

Author works in a big multinational company and has over 8 years of experience in software industry. He is working presently as project lead and in past has led projects in Banking, travel and financial sectors.

But on the top of all , I am a simple developer like you all guys there doing an 8 hour job. Writing is something I do extra and I love doing it. No one is perfect and same holds true for me .So anything you want to comment, suggest, point typo / grammar mistakes or technical mistakes regarding the book you can mail me at shiv_koirala@yahoo.com. Believe me guys your harsh words would be received with love and treated to the top most priority. Without all you guys I am not an author.

Writing an interview question book is really a great deal of responsibility. I have tried to cover maximum questions for the topic because I always think probably leaving one silly question will cost someone’s job there. But huge natural variations in an interview are something difficult to cover in this small book. So if you have come across such questions during interview which is not addressed in this book do mail at shiv_koirala@yahoo.com .Who knows probably that question can save some other guys job.

Features of the book

√ This book goes in best combination with my previous book ".NET Interview questions". One takes care of your front end aspect and this one the back end which will make you really stand out during .NET interviews.

√ Around 400 plus SQL Server Interview questions sampled from real SQL Server Interviews conducted across IT companies.

√ Other than core level interview question, DBA topics like database optimization and locking are also addressed.

√ Replication section where most of the developer stumble, full chapter is dedicated to replication so that during interview you really look a champ.

√ SQLCLR that is .NET integration which is one of the favorites of every interviewer is addressed with great care .This makes developer more comfortable during interview.

√ XML is one of the must to be answered questions during interview. All new XML features are covered with great elegance.

√ Areas like data warehousing and data mining are handled in complete depth.

√ Reporting and Analysis services which can really surprise developers during interviews are also dealt with great care.

√ A complete chapter on ADO.NET makes it more stronger from a programmer aspect. In addition new ADO.NET features are also highlighted which can be pain points for the new features released with SQL Server.

√ Must for developers who are looking to crack SQL Server interview for DBA position or programmer position.

√ Must for freshers who want to avoid some unnecessary pitfall during interview.

√ Every answer is precise and to the point rather than hitting around the bush. Some questions are answered to greater detail with practical implementation in mind.

√ Every question is classified in DB and NON-DB level. DB level question are mostly for guys who are looking for high profile DBA level jobs. All questions other than DB level are NON-DB level which is must for every programmer to know.

√ Tips and tricks for interview, resume making and salary negotiation section takes this book to a greater height.

Introduction

When my previous book ".NET Interview Questions" reached the readers, the only voice heared was more “SQL Server”. Ok guys we have heard it louder and clearer, so here’s my complete book on SQL Server: - “SQL Server Interview Questions”. But there’s a second stronger reason for writing this book which stands taller than the readers demand and that is SQL Server itself. Almost 90 % projects in software industry need databases or persistent data in some or other form. When it comes to .NET persisting data SQL Server is the most preferred database to do it. There are projects which use ORACLE, DB2 and other database product, but SQL Server still has the major market chunk when language is .NET and especially operating system is windows. I treat this great relationship between .NET, SQL Server and Windows OS as a family relationship.

In my previous book we had only one chapter which was dedicated to SQL Server which is complete injustice to this beautiful product.

So why an interview question book on SQL Server? If you look at any .NET interview conducted in your premises both parties (Employer and Candidate) pay no attention to SQL Server even though when it is such an important part of development project. They will go talking about stars (OOP, AOP, Design patterns, MVC patterns, Microsoft Application blocks, Project Management etc.) but on database side there would be rare questions. I am not saying these things are not important but if you see in development or maintenance majority time you will be either in your IDE or in SQL Server.

Secondly many candidates go really as heroes when answering questions of OOP , AOP , Design patterns , architecture , remoting etc etc but when it comes to simple basic question on SQL Server like SQL , indexes (forget DBA level questions) they are completely out of track.

Third very important thing IT is changing people expect more out of less. That means they expect a programmer should be architect, coder, tester and yes and yes a DBA also. For mission critical data there will always be a separate position for a DBA. But now many interviewers expect programmers to also do a job of DBA, Data warehousing etc. This is the major place where developers lack during facing these kinds of interview.

So this book will make you walk through those surprising questions which can sprang from SQL Server aspect. I have tried to not go too deep as that will defeat the complete purpose of an Interview Question book. I think that an interview book should make you run through those surprising question and make you prepare in a small duration (probably with a night or so). I hope this book really points those pitfalls which can come during SQL Server Interview's.

I hope this book takes you to a better height and gives you extra confidence boost during interviews.Best of Luck and Happy Job-Hunting............

How to read this book

If you can read English, you can read this book....kidding. In this book there are some legends which will make your reading more effective. Every question has simple tags which mark the rating of the questions.

These rating are given by Author and can vary according to companies and individuals.

Compared to my previous book ".NET Interview Questions" which had three levels (Basic, Intermediate and Advanced) this book has only two levels (DBA and NON-DBA) because of the subject. While reading you can come across section marked as "Note" , which highlight special points of that section. You will also come across tags like "TWIST", which is nothing , but another way of asking the same question, for instance "What is replication?" and "How do I move data between two SQL Server database?" , point to the same answer.

All questions with DBA level are marked with (DB) tag. Questions which do not have tags are NON-DBA levels. Every developer should have a know how of all NON-DBA levels question. But for DBA guys every question is important. For instance if you are going for a developer position and you flunk in simple ADO.NET question you know the result. Vice versa if you are going for a DBA position and you can not answer basic query optimization questions probably you will never reach the HR round.

So the best way to read this book is read the question and judge yourself do you think you will be asked these types of questions? For instance many times you know you will be only asked about data warehousing and rather than hitting the bush around you would like to target that section more. And Many times you know your weakest area and you would only like to brush up those sections. You can say this book is not a book which has to be read from start to end you can start from a chapter or question and when you think you are ok close it.

Software Company hierarchy

It’s very important during interview to be clear about what position you are targeting. Depending on what positions you are targeting the interviewer shoots you questions. Example if you are looking for a DBA position you will be asked around 20% ADO.NET questions and 80% questions on query optimization, profiler, replication, data warehousing, data mining and others.

> *Note:- In small scale software house and mid scale software companies there are chances where they expect a developer to a job of programming , DBA job , data mining and everything. But in big companies you can easily see the difference where DBA job are specifically done by specialist of SQL Server rather than developers. But now a days some big companies believe in a developer doing multitask jobs to remove dependencies on a resource.*

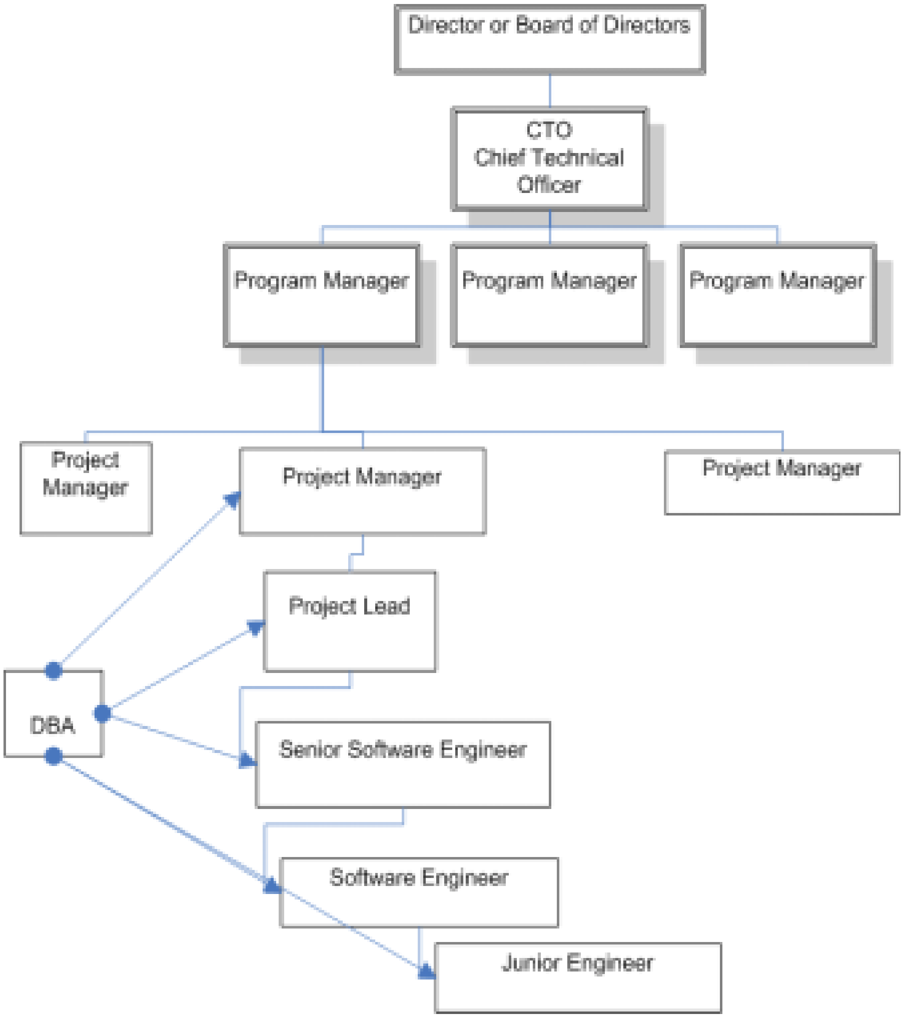


Figure :- 0.1 IT Company hierarchy

Above is a figure of a general hierarchy across most IT companies (Well not always but I hope most of the time). Because of inconsistent HR way of working you will see difference between companies.

Note: - There are many small and medium software companies which do not follow this hierarchy and they have there own ADHOC way of defining positions in the company.

So why there is a need of hierarchy in an interview?

"Interview is a contract between the employer and candidate to achieve specific goals."

So employer is looking for a suitable candidate and candidate for a better career. Normally in interviews the employer is very clear about what type of candidate he is looking for. But 90% times the candidate is not clear about the positions he is looking for.

How many times has it happened with you that you have given a whole interview and when you mentioned the position you are looking for...pat comes the answer we do not have any requirements for this position. So be clarified about the position right when you start the interview.

Following are the number of years of experience according to position.

√ Junior engineers are especially fresher and work under software engineers.

√ Software engineers have around 1 to 2 years of experience. Interviewer expects software engineers to have know how of how to code ADO.NET with SQL Server.

√ Senior Software Engineers have around 2 to 4 years of experience. Interviewer expect them to be technically very strong.

√ Project leads should handle majority technical aspect of project and should have around 4 to 8 years of experience. They are also actively involved in to defining architect of the project. Interviewer expects them to be technically strong plus should have managerial skills.

√ Project Managers are expected to be around 40% technically strong and should have experience above 10 years plus. But they are more interviewed from aspect of project management, client interaction, people management, proposal preparation etc.

√ Pure DBA's do not come in hierarchy as such in pure development projects. They do report to the project managers or project leads but they are mainly across the hierarchy helping every one in a project. In small companies software developers can also act as DBA's depending on companies policy. Pure DBA's have normally around 6 and above years of experience in that particular database product.

- √ When it comes to maintenance projects where you have special DBA positions lot of things are ADHOC. That means one or two guys work fulfilling maintenance tickets.

So now judge where you stand where you want to go..........

Resume Preparation Guidelines

First impression the last impression

Before even the interviewer meets you he will first meet your resume. Interviewer looking at your resume is almost a 20% interview happening with out you knowing it. I was always a bad guy when it comes to resume preparation. But when I looked at my friends resume they where gorgeous. Now that I am writing series of book on interviews I thought this will be a good point to put in. You can happily skip it if you are confident about your resume. There is no hard and fast rule that you have to follow the same pattern but just see if these all check list are attended.

- √ Use plain text when you are sending resumes through email. For instance you sent your resume using Microsoft word and what if the interviewer is using Linux he will never be able to read your resume. You can not be sure both wise , you sent your resume in Word 2000 and the guy has Word 97…uuhhh.
- √ Attach a covering letter it really impresses and makes you look traditionally formal. Yes even if you are sending your CV through email send a covering letter.

Check list of content you should have in your resume :-

- √ Start with an objective or summary, for instance, “Working as a Senior Database administrator for more than 4 years. Implemented quality web based application. Followed the industry’s best practices and adhered and implemented processes, which enhanced the quality of technical delivery. Pledged to deliver the best technical solutions to the industry.”
- √ Specify your Core strengths at the start of the resume by which the interviewer can make a quick decision are you eligible for the position. For example :-
  - Looked after data mining and data warehousing department independently. Played a major role in query optimization.
  - Worked extensively in database design and ER diagram implementation.
  - Well versed with CMMI process and followed it extensively in projects.

- Looking forward to work on project manager or senior manager position.

This is also a good position to specify your objective or position which makes it clear to the interviewer that should he call you for an interview. For instance if you are looking for senior position specify it explicitly looking for this job profile. Any kind of certification like MCP, MCSD etc you can make it visible in this section.

√ Once you have specified briefly your goals and what you have done its time to specify what type of technology you have worked with. For instance RDBMS, TOOLS, Languages, Web servers, process (Six sigma, CMMI).

√ After that you can make a run through of your experience company wise that is what company you have worked with, year / month joining and year / month left. This will give an overview to the interviewer what type of companies you have associated your self.

Now its time to mention all your projects you have worked till now. Best is to start in descending order that is from your current project and go backwards. For every project try to put these things :-

√ Project Name / Client name (It's sometimes unethical to mention clients name; I leave it to the readers).

√ Number of team members.

√ Time span of the project.

√ Tools, language, RDBMS and technology used to complete the project.

√ Brief summary of the project.

Senior people who have huge experience will tend to increase there CV with putting in summary for all project. Best for them is to just put description of the first three projects in descending manner and rest they can say verbally during interview. I have seen CV above 15 pages… I doubt who can read it.

√ Finally comes your education and personal details.

√ Trying for onsite, do not forget to mention your passport number.

√ Some guys tend to make there CV large and huge. I think an optimal size should be not more than 4 to 5 pages.

√ Do not mention your salary in CV. You can talk about it during interview with HR or the interviewer.

√ When you are writing your summary for project make it effective by using verbs like managed a team of 5 members, architected the project from start to finish etc. It brings huge weight.

√ This is essential very essential take 4 to 5 Xerox copies of your resume you will need it now and then.

√ Just in case take at least 2 passport photos with you. You can escape it but many times you will need it.

√ Carry you’re all current office documents specially your salary slips and joining letter.

Salary Negotiation

Ok that’s what we all do it for money… not every one right. This is probably the weakest area for techno savvy guys. They are not good negotiators. I have seen so many guys at the first instance they will smile say “NEGOTIABLE SIR”. So here are some points:-

√ Do a study of what‘s the salary trend? For instance have some kind of baseline. For example what’s the salary trend on number of year of experience? Discuss this with your friends out.

√ Do not mention your expected salary on the resume?

√ Let the employer first make the salary offer. Try to delay the salary discussion till the end.

√ If they say what you expect ? , come with a figure with a little higher end and say negotiable. Remember never say negotiable on something which you have aimed, HR guys will always bring it down. So negotiate on AIMED SALARY + some thing extra.

√ The normal trend is that they look at your current salary and add a little it so that they can pull you in. Do your home work my salary is this much and I expect this much so whatever it is now I will not come below this.

√ Do not be harsh during salary negotiations.

√ It’s good to aim high. For instance I want 1 billion dollars / month but at the same time be realistic.

- √ Some companies have those hidden cost attached in salary clarify that rather to be surprised at the first salary package.
- √ Many of the companies add extra performance compensation in your basic which can be surprising at times. So have a detail break down. Best is to discuss on hand salary rather than NET.
- √ Talk with the employer in what frequency does the hike happen.
- √ Take everything in writing , go back to your house and have a look once with a cool head is the offer worth it of what your current employer is giving.
- √ Do not forget once you have job in hand you can come back to your current employer for negotiation so keep that thing in mind.
- √ Remember the worst part is cribbing after joining the company that your colleague is getting this much. So be careful while interview negotiations or be sportive to be a good negotiator in the next interview.
- √ One very important thing the best negotiation ground is not the new company where you are going but the old company which you are leaving. So once you have offer on hand get back to your old employee and show them the offer and then make your next move. It's my experience that negotiating with the old employer is easy than with the new one….Frankly if approached properly rarely any one will say no. Just do not be aggressive or egoistic that you have an offer on hand.

Top of all some time some things are worth above money :- JOB SATISFACTION. So whatever you negotiate if you think you can get JOB SATISFACTION aspect on higher grounds go for it. I think its worth more than money.

Points to remember

- √ One of the first questions asked during interview is "Can you say something about yourself "?
- √ Can you describe about your self and what you have achieved till now?
- √ Why you want to leave the current company?
- √ Where do you see yourself after three years?
- √ What are your positive and negative points?
- √ How much do you rate yourself in .NET and SQL Server in one out of ten?

- √ Are you looking for onsite opportunities? (Be careful do not show your desperation of abroad journeys)
- √ Why have you changed so many jobs? (Prepare a decent answer do not blame companies and individuals for your frequent change).
- √ Never talk for more than 1 minute straight during interview.
- √ Have you worked with previous version of SQL Server?
- √ Would you be interested in a full time Database administrator job?
- √ Do not mention client name's in resume. If asked say that it's confidential which brings ahead qualities like honesty
- √ When you make your resume keep your recent projects at the top.
- √ Find out what the employer is looking for by asking him questions at the start of interview and best is before going to interview. Example if a company has projects on server products employer will be looking for BizTalk, CS CMS experts.
- √ Can you give brief about your family background?
- √ As you are fresher do you think you can really do this job?
- √ Have you heard about our company ? Say five points about our company? Just read at least once what company you are going for?
- √ Can you describe your best project you have worked with?
- √ Do you work on Saturday and Sunday?
- √ Which is the biggest team size you have worked with?
- √ Can you describe your current project you have worked with?
- √ How much time will you need to join our organization? What's notice period for your current company?
- √ What certifications have you cleared?
- √ Do you have pass port size photos, last year mark sheet, previous companies employment letter, last months salary slip, pass port and other necessary documents.
- √ What's the most important thing that motivates you?
- √ Why you want to leave the previous organization?

- √ Which type of job gives you greatest satisfaction?
- √ What is the type of environment you are looking for?
- √ Do you have experience in project management?
- √ Do you like to work as a team or as individual?
- √ Describe your best project manager you have worked with?
- √ Why should I hire you?
- √ Have you been ever fired or forced to resign?
- √ Can you explain some important points that you have learnt from your past project experiences?
- √ Have you gone through some unsuccessful projects, if yes can you explain why did the project fail?
- √ Will you be comfortable with location shift? If you have personal problems say no right at the first stage.... or else within two months you have to read my book again.
- √ Do you work during late nights? Best answer if there is project deadline yes. Do not show that it's your culture to work during nights.
- √ Any special achievements in your life till now...tell your best project which you have done best in your career.
- √ Any plans of opening your own software company...Beware do not start pouring your bill gate's dream to him.....can create a wrong impression.

1. Database Concepts

What is database or database management systems (DBMS)?

Twist: - What's the difference between file and database? Can files qualify as a database?

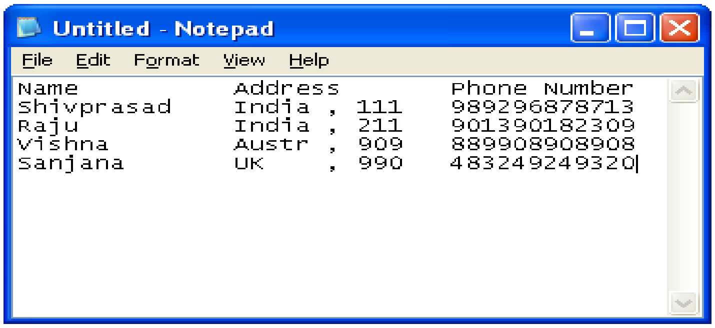
Note: - Probably these questions are too basic for experienced SQL SERVER guys. But from freshers point of view it can be a difference between getting a job and to be jobless.

Database provides a systematic and organized way of storing, managing and retrieving from collection of logically related information.

Secondly the information has to be persistent, that means even after the application is closed the information should be persisted.

Finally it should provide an independent way of accessing data and should not be dependent on the application to access the information.

Ok let me spend a few sentence more on explaining the third aspect. Below is a simple figure of a text file which has personal detail information. The first column of the information is Name, Second address and finally the phone number. This is a simple text file which was designed by a programmer for a specific application.



```
Name              Address              Phone Number
Shivprasad        India , 111          989296878713
Raju              India , 211          901390182309
Vishna            Austr , 909          889908908908
Sanjana           UK    , 990          483249249320
```

Figure 1.1:- Non-Uniform Text File

It works fine in the boundary of the application. Now some years down the line a third party application has to be integrated with this file , so in order the third party application integrates properly it has the following options :-

- √ Use interface of the original application.
- √ Understand the complete detail of how the text file is organized, example the first column is Name, then address and finally phone number. After analyzing write a code which can read the file, parse it etc ….Hmm lot of work right.

That's what the main difference between a simple file and database; database has independent way (SQL) of accessing information while simple files do not (That answers my twisted question defined above). File meets the storing, managing and retrieving part of a database but not the independent way of accessing data.

Note: - Many experienced programmers think that the main difference is that file can not provide multi-user capabilities which a DBMS provides. But if you look at some old COBOL and C programs where file where the only means of storing data, you can see functionalities like locking, multi-user etc provided very efficiently. So it's a matter of debate if some interviewers think this as a main difference between files and database accept it… going in to debate is probably loosing a job.

(Just a note for fresher's multi-user capabilities means that at one moment of time more than one user should be able to add, update, view and delete data. All DBMS provides this as in built functionalities but if you are storing information in files it's up to the application to write a logic to achieve these functionalities)

What's difference between DBMS and RDBMS ?

Ok as said before DBMS provides a systematic and organized way of storing, managing and retrieving from collection of logically related information. RDBMS also provides what DBMS provides but above that it provides relationship integrity. So in short we can say

RDBMS = DBMS + REFERENTIAL INTEGRITY

Example in above figure 1.1 every person should have an address this is a referential integrity between "Name" and "Address". If we break this referential integrity in DBMS and File's it will not complain, but RDBMS will not allow you to save this data if you have defined the relation integrity between person and addresses. These relations are defined by using "Foreign Keys" in any RDBMS.

Many DBMS companies claimed there DBMS product was a RDBMS compliant, but according to industry rules and regulations if the DBMS fulfills the twelve CODD rules it's truly a RDBMS. Almost all DBMS (SQL SERVER, ORACLE etc) fulfills all the twelve CODD rules and are considered as truly RDBMS.

> *Note: - One of the biggest debate, Is Microsoft Access a RDBMS? We will be answering this question in later section.*

(DB)What are CODD rules?

> *Twist: - Does SQL SERVER support all the twelve CODD rules?*
>
> *Note: - This question can only be asked on two conditions when the interviewer is expecting you to be at a DBA job or you are complete fresher, yes and not to mention the last one he treats CODD rules as a religion. We will try to answer this question from perspective of SQL SERVER.*

In 1969 Dr. E. F. Codd laid down some 12 rules which a DBMS should adhere in order to get the logo of a true RDBMS.

Rule 1: Information Rule.

"All information in a relational data base is represented explicitly at the logical level and in exactly one way - by values in tables."

In SQL SERVER all data exists in tables and are accessed only by querying the tables.

Rule 2: Guaranteed access Rule.

"Each and every datum (atomic value) in a relational data base is guaranteed to be logically accessible by resorting to a combination of table name, primary key value and column name."

In flat files we have to parse and know exact location of field values. But if a DBMS is truly RDBMS you can access the value by specifying the table name, field name, for instance Customers.Fields ['Customer Name']

SQL SERVER also satisfies this rule in ADO.NET we can access field information using table name and field names.

Rule 3: Systematic treatment of null values.

"Null values (distinct from the empty character string or a string of blank characters and distinct from zero or any other number) are supported in fully relational DBMS for representing missing information and inapplicable information in a systematic way, independent of data type."

In SQL SERVER if there is no data existing NULL values are assigned to it. Note NULL values in SQL SERVER do not represent spaces, blanks or a zero value; it's a distinct representation of missing information and thus satisfying rule 3 of CODD.

Rule 4: Dynamic on-line catalog based on the relational model.

"The data base description is represented at the logical level in the same way as ordinary data, so that authorized users can apply the same relational language to its interrogation as they apply to the regular data."

The Data Dictionary is held within the RDBMS, thus there is no-need for off-line volumes to tell you the structure of the database.

Rule 5: Comprehensive data sub-language Rule.

"A relational system may support several languages and various modes of terminal use (for example, the fill-in-the-blanks mode). However, there must be at least one language whose statements are expressible, per some well-defined syntax, as character strings and that is comprehensive in supporting all the following items

- √ Data Definition
- √ View Definition
- √ Data Manipulation (Interactive and by program).
- √ Integrity Constraints
- √ Authorization.
- √ Transaction boundaries (Begin , commit and rollback)

SQL SERVER uses SQL to query and manipulate data which has well-defined syntax and is being accepted as an international standard for RDBMS.

> *Note: - According to this rule CODD has only mentioned that some language should be present to support it, but not necessary that it should be SQL. Before 80's different*

database vendors where providing there own flavor of syntaxes until in 80 ANSI-SQL came into standardize this variation between vendors. As ANSI-SQL is quiet limited, every vendor including Microsoft introduced there additional SQL syntaxes in addition to the support of ANSI-SQL. You can see SQL syntaxes varying from vendor to vendor.

Rule 6: .View updating Rule

"All views that are theoretically updatable are also updatable by the system."

In SQL SERVER not only views can be updated by user, but also by SQL SERVER itself.

Rule 7: High-level insert, update and delete.

"The capability of handling a base relation or a derived relation as a single operand applies not only to the retrieval of data but also to the insertion, update and deletion of data."

SQL SERVER allows you to update views which in turn affect the base tables.

Rule 8: Physical data independence.

"Application programs and terminal activities remain logically unimpaired whenever any changes are made in either storage representations or access methods."

Any application program (C#, VB.NET, VB6 VC++ etc) Does not need to be aware of where the SQL SERVER is physically stored or what type of protocol it's using, database connection string encapsulates everything.

Rule 9: Logical data independence.

"Application programs and terminal activities remain logically unimpaired when information-preserving changes of any kind that theoretically permit un-impairment are made to the base tables."

Application programs written in C# or VB.NET does not need to know about any structure changes in SQL SERVER database example: - adding of new field etc.

Rule 10: Integrity independence.

"Integrity constraints specific to a particular relational data base must be definable in the relational data sub-language and storable in the catalog, not in the application programs."

In SQL SERVER you can specify data types (integer, nvarchar, Boolean etc) which puts in data type checks in SQL SERVER rather than through application programs.

Rule 11: Distribution independence.

"A relational DBMS has distribution independence."

SQL SERVER can spread across more than one physical computer and across several networks; but from application programs it has not big difference but just specifying the SQL SERVER name and the computer on which it is located.

Rule 12: Non-subversion Rule.

"If a relational system has a low-level (single-record-at-a-time) language, that low level cannot be used to subvert or bypass the integrity Rules and constraints expressed in the higher level relational language (multiple-records-at-a-time)."

In SQL SERVER whatever integrity rules are applied on every record are also applicable when you process a group of records using application program in any other language (example: - C#, VB.NET, J# etc...).

Reader's can see from the above explanation SQL SERVER satisfies all the CODD rules, some database guru's consider SQL SERVER as not truly RDBMS, but that's a matter of debate.

Is access database a RDBMS?

Access fulfills all rules of CODD, so from this point of view yes it's truly RDBMS. But many people can contradict it as a large community of Microsoft professional thinks that access is not.

What's the main difference between ACCESS and SQL SERVER?

As said before access fulfills all the CODD rules and behaves as a true RDBMS. But there's a huge difference from architecture perspective, due to which many developers prefer to use SQL SERVER as major database rather than access. Following is the list of architecture differences between them:-

√ Access uses file server design and SQL SERVER uses the Client / Server model. This forms the major difference between SQL SERVER and ACCESS.

Note: - Just to clarify what is client server and file server I will make quick description of widely accepted architectures. There are three types of architecture:-

- *Main frame architecture (This is not related to the above explanation but just mentioned it as it can be useful during interview and also for comparing with other architectures)*
- *File sharing architecture (Followed by ACCESS).*
- *Client Server architecture (Followed by SQL SERVER).*

In Main Frame architecture all the processing happens on central host server. User interacts through dump terminals which only sends keystrokes and information to host. All the main processing happens on the central host server. So advantage in such type of architecture is that you need least configuration clients. But the disadvantage is that you need robust central host server like Main Frames.

In File sharing architecture which is followed by access database all the data is sent to the client terminal and then processed. For instance you want to see customers who stay in INDIA, in File Sharing architecture all customer records will be send to the client PC regardless whether the customer belong to INDIA or not. On the client PC customer records from India is sorted/filtered out and displayed, in short all processing logic happens on the client PC. So in this architecture the client PC should have heavy configuration and also it increases network traffic as lot of data is sent to the client PC. But advantage of this architecture is that your server can be of low configurations.

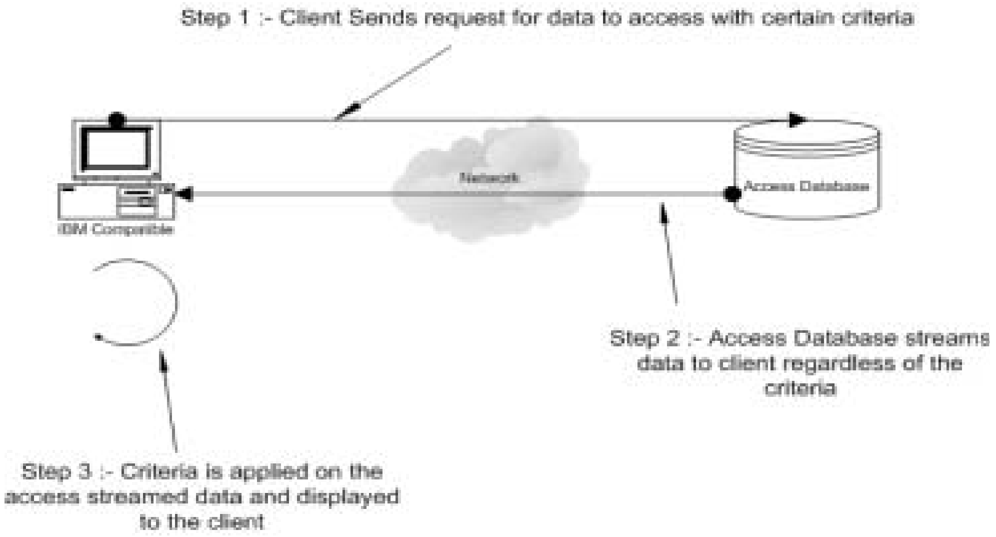

Step 1 :- Client Sends request for data to access with certain criteria
Network
Access Database
IBM Compatible
Step 2 :- Access Database streams data to client regardless of the criteria
Step 3 :- Criteria is applied on the access streamed data and displayed to the client

Figure 1.2:- File Server Architecture of Access

In client server architecture the above limitation of the file server architecture is removed. In Client server architecture you have two entities client and the database server. File server is now replaced by database server. Database server takes up the load of processing any database related activity and the client any validation aspect of database. As the work is distributed between the entities it increases scalability and reliability. Second the network traffic also comes down as compared to file server. For example if you are requesting customers from INDIA, database server will sort/ filter and send only INDIAN customer details to the client, thus bringing down the network traffic tremendously. SQL SERVER follows the client-server architecture.

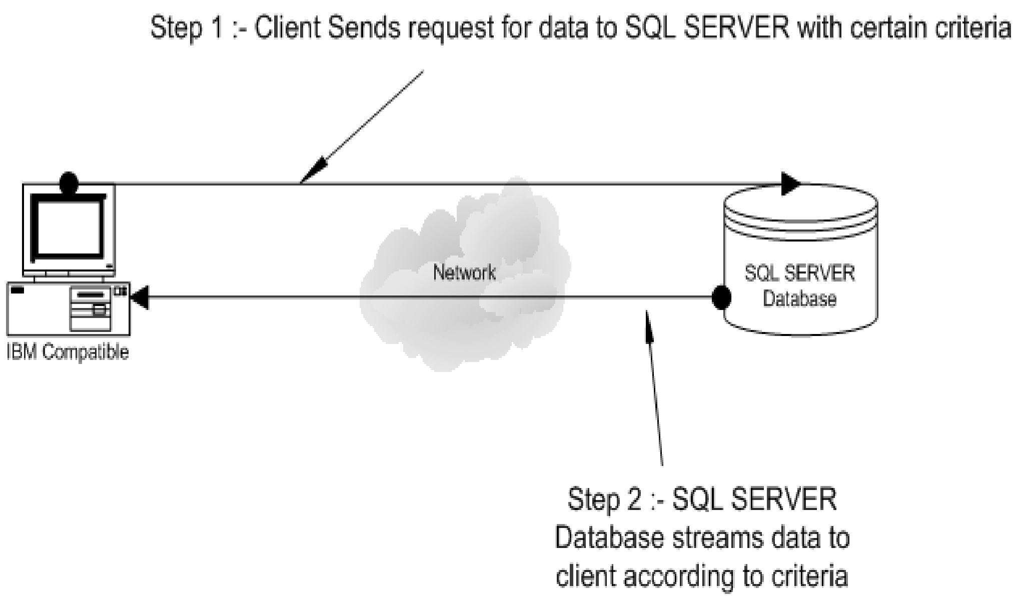

Step 1 :- Client Sends request for data to SQL SERVER with certain criteria
Network
SQL SERVER Database
IBM Compatible
Step 2 :- SQL SERVER Database streams data to client according to criteria

Figure 1.3:- Client Server Architecture of SQL SERVER

√ Second issue comes in terms of reliability. In Access the client directly interacts with the access file, in case there is some problem in middle of transaction there are chances that access file can get corrupt. But in SQL SERVER the engine sits in between the client and the database, so in case of any problems in middle of the transaction it can revert back to its original state.

Note: - SQL SERVER maintains a transaction log by which you can revert back to your original state in case of any crash.

√ When your application has to cater to huge load demand, highly transactional environment and high concurrency then its better to for SQL SERVER or MSDE.

√ But when it comes to cost and support the access stands better than SQL SERVER.In case of SQL SERVER you have to pay for per client license, but access runtime is free.

Summarizing: - SQL SERVER gains points in terms of network traffic, reliability and scalability vice-versa access gains points in terms of cost factor.

What’s the difference between MSDE and SQL SERVER 2000?

MSDE is a royalty free, redistributable and cut short version of the giant SQL SERVER database. It’s primarily provided as a low cost option for developers who need database server which can easily be shipped and installed. It can serve as good alternative for Microsoft Access database as it over comes quiet a lot of problems which access has.

Below is a complete list which can give you a good idea of differences:-

- √ Size of database: - MS ACCESS and MSDE have a limitation of 2GB while SQL SERVER has 1,048,516 TB1.
- √ Performance degrades in MSDE 2000 when maximum number of concurrent operations goes above 8 or equal to 8. It does not mean that you can not have more than eight concurrent operations but the performance degrades. Eight connection performance degradation is implemented by using SQL SERVER 2000 work load governor (we will be looking in to more detail of how it works). As compared to SQL SERVER 2000 you can have 32,767 concurrent connections.
- √ MSDE does not provide OLAP and Data ware housing capabilities.
- √ MSDE does not have support facility for SQL mail.
- √ MSDE 2000 does not have GUI administrative tool such as enterprise manager, Query analyzer or Profiler. But there are round about ways by which you can manage MSDE 2000 :-
  - √ Old command line utility OSQL.EXE.
  - √ VS.NET IDE Server Explorer: - Inside VS.NET IDE you have a functionality which can give you a nice GUI administrative tool to manage IDE.
  - √ SQL SERVER WEB Data administrator installs a web based GUI which you can use to manage your database. For any details refer http://www.microsoft.com/downloads/details.aspx?familyid=c039a798-c57a-419e-acbc-2a332cb7f959&displaylang=en
  - √ SQL-DMO objects can be used to build your custom UI.

  - √ There are lots of third party tools which provide administrative capability GUI, which is out of scope of the book as it's only meant for interview questions.

- √ MSDE does not support Full text search.

Summarizing: - There are two major differences first is the size limitation (2 GB) of database and second are the concurrent connections (eight concurrent connections) which are limited by using the work load governor. During interview this answer will suffice if he is really testing your knowledge.

What is SQL SERVER Express 2005 Edition?

Twist: - What's difference between SQL SERVER Express 2005 and MSDE 2000?

Note: - Normally comparison is when the product is migrating from one version to other version. When SQL SERVER 7.0 was migrating to SQL 2000, asking differences was one of the favorite questions.

SQL SERVER Express edition is a scaled down version of SQL SERVER 2005 and next evolution of MSDE.

Below listed are some major differences between them:-

- √ MSDE maximum database size is 2GB while SQL SERVER Express has around 4GB.
- √ In terms of programming language support MSDE has only TSQL, but SQL SERVER Express has TSQL and .NET. In SQL SERVER Express 2005 you can write your stored procedures using .NET.
- √ SQL SERVER Express does not have connection limitation which MSDE had and was controlled through the work load governor.
- √ There was no XCOPY support for MSDE, SQL SERVER Express has it.
- √ DTS is not present in SQL SERVER express while MSDE has it.
- √ SQL SERVER Express has reporting services while MSDE does not.
- √ SQL SERVER Express has native XML support and MSDE does not.

Note: - Native XML support mean now in SQL SERVER 2005 :-

- √ You can create a field with data type "XML".

- √ You can provide SCHEMA to the SQL SERVER fields with “XML” data type.
- √ You can use new XML manipulation techniques like “XQUERY” also called as “XML QUERY”.

There is complete chapter on SQL SERVER XML Support so till then this will suffice.

> *Summarizing: - Major difference is database size (2 GB and 4 GB), support of .NET support in stored procedures and native support for XML. This is much can convince the interviewer that you are clear about the differences.*

(DB) What is SQL Server 2000 Workload Governor?

Workload governor limits performance of SQL SERVER Desktop engine (MSDE) if the SQL engine receives more load than what is meant for MSDE. MSDE was always meant for trial purpose and non-critical projects. Microsoft always wanted companies to buy there full blow version of SQL SERVER so in order that they can put limitation on MSDE performance and number of connections they introduced Workload governor.

Workload governor sits between the client and the database engine and counts number of connections per database instance. If Workload governor finds that the number of connections exceeds eight connections, it starts stalling the connections and slowing down the database engine.

> *Note: - It does not limit the number of connections but makes the connection request go slow. By default 32,767 connections are allowed both for SQL SERVER and MSDE. But it just makes the database engine go slow above eight connections.*

What’s the difference between SQL SERVER 2000 and 2005?

> *Twist: - What’s the difference between Yukon and SQL SERVER 2000?*
>
> *Note:-This question will be one of the favorites during SQL SERVER Interviews. I have marked the points which should be said by developers as PG and DBA for Database Administrator.*

Following are the some major differences between the two versions:-

- √ (PG) The most significant change is the .NET integration with SQL SERVER 2005. Stored procedures, User-defined functions, triggers, aggregates, and user-defined types can now be written using your own favorite .NET language (VB.NET, C#, J# etc .). This support was not there in SQL SERVER 2000 where the only language was T-SQL. In SQL 2005 you have support for two languages T-SQL and .NET.

√ (PG) SQL SERVER 2005 has reporting services for reports which is a newly added feature and does not exist for SQL SERVER 2000.It was a seperate installation for SQL Server 2000

√ (PG) SQL SERVER 2005 has introduced two new data types varbinary (max) and XML. If you remember in SQL SERVER 2000 we had image and text data types. Problem with image and text data types is that they assign same amount of storage irrespective of what the actual data size is. This problem is solved using varbinary (max) which acts depending on amount of data. One more new data type is included "XML" which enables you to store XML documents and also does schema verification. In SQL SERVER 2000 developers used varchar or text data type and all validation had to be done programmatically.

√ (PG) SQL SERVER 2005 can now process direct incoming HTTP request with out IIS web server. Also stored procedure invocation is enabled using the SOAP protocol.

√ (PG) Asynchronous mechanism is introduced using server events. In Server event model the server posts an event to the SQL Broker service, later the client can come and retrieve the status by querying the broker.

√ For huge databases SQLSERVER has provided a cool feature called as “Data partitioning”. In data partitioning you break a single database object such as a table or an index into multiple pieces. But for the client application accessing the single data base object “partitioning” is transparent.

√ In SQL SERVER 2000 if you rebuilt clustered indexes even the non-clustered indexes where rebuilt. But in SQL SERVER 2005 building the clustered indexes does not built the non-clustered indexes.

√ Bulk data uploading in SQL SERVER 2000 was done using BCP (Bulk copy program’s) format files. But now in SQL SERVER 2005 bulk data uploading uses XML file format.

- √ In SQL SERVER 2000 there where maximum 16 instances, but in 2005 you can have up to 50 instances.
- √ SQL SERVER 2005 has support of "Multiple Active Result Sets" also called as "MARS". In previous versions of SQL SERVER 2000 in one connection you can only have one result set. But now in one SQL connection you can query and have multiple results set.
- √ In previous versions of SQL SERVER 2000, system catalog was stored in master database. In SQL SERVER 2005 it's stored in resource database which is stored as sys object , you can not access the sys object directly as in older version wewhere accessing master database.
- √ This is one of hardware benefits which SQL SERVER 2005 has over SQL SERVER 2000 – support of hyper threading. WINDOWS 2003 supports hyper threading; SQL SERVER 2005 can take the advantage of the feature unlike SQL SERVER 2000 which did not support hyper threading.

Note: - Hyper threading is a technology developed by INTEL which creates two logical processors on a single physical hardware processor.

- √ SMO will be used for SQL Server Management.
- √ AMO (Analysis Management Objects) to manage Analysis Services servers, data sources, cubes, dimensions, measures, and data mining models. You can map AMO in old SQL SERVER with DSO (Decision Support Objects).
- √ Replication is now managed by RMO (Replication Management Objects).

Note: - SMO, AMO and RMO are all using .NET Framework.

- √ SQL SERVER 2005 uses current user execution context to check rights rather than ownership link chain, which was done in SQL SERVER 2000.

Note: - There is a question on this later see for execution context questions.

- √ In previous versions of SQL SERVER the schema and the user name was same, but in current the schema is separated from the user. Now the user owns schema.

Note: - There are questions on this, refer – "Schema" later.

Note:-Ok below are some GUI changes.

- √ Query analyzer is now replaced by query editor.
- √ Business Intelligence development studio will be used to create Business intelligence solutions.
- √ OSQL and ISQL command line utility is replaced by SQLCMD utility.
- √ SQL SERVER Enterprise manager is now replaced by SQL SERVER Management studio.
- √ SERVER Manager which was running in system tray is now replaced by SQL Computer manager.
- √ Database mirror concept supported in SQL SERVER 2005 which was not present in SQL SERVER 2000.
- √ In SQL SERVER 2005 Indexes can be rebuild online when the database is in actual production. If you look back in SQL SERVER 2000 you can not do insert, update and delete operations when you are building indexes.
- √ (PG) Other than Serializable, Repeatable Read, Read Committed and Read Uncommitted isolation level there is one more new isolation level "Snapshot Isolation level".

Note: - We will see "Snapshot Isolation level" in detail in coming questions.

Summarizing: - The major significant difference between SQL SERVER 2000 and SQL SERVER 2005 is in terms of support of .NET Integration, Snap shot isolation level, Native XML support, handling HTTP request, Web service support and Data partitioning. You do not have to really say all the above points during interview a sweet summary and you will rock.

What are E-R diagrams?

E-R diagram also termed as Entity-Relationship diagram shows relationship between various tables in the database. Example: - Table "Customer" and "CustomerAddresses" have a one to many relationships (i.e. one customer can have multiple addresses) this can be shown using the ER diagram. ER diagrams are drawn during the initial stages of project to forecast how the database structure will shape up. Below is a screen shot of a sample ER diagram of "Asset Management" which ships free with access.

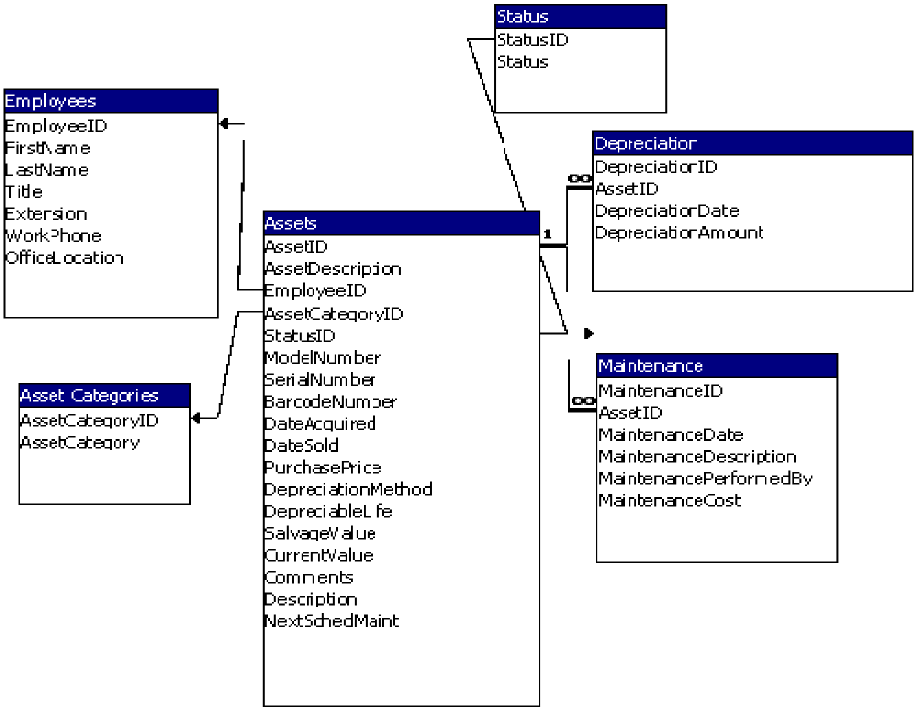


Figure 1.4: - Asset management ER diagram.

How many types of relationship exist in database designing?

There are three major relationship models:-

√ One-to-one

As the name suggests you have one record in one table and corresponding to that you have one record in other table. We will take the same sample ER diagram defined for asset management. In the below diagram "Assets" can only have one "Status" at moment of time (Outdated / Need Maintenance and New Asset). At any moment of time "Asset" can only have one status of the above, so there is one-to-one relationship between them.

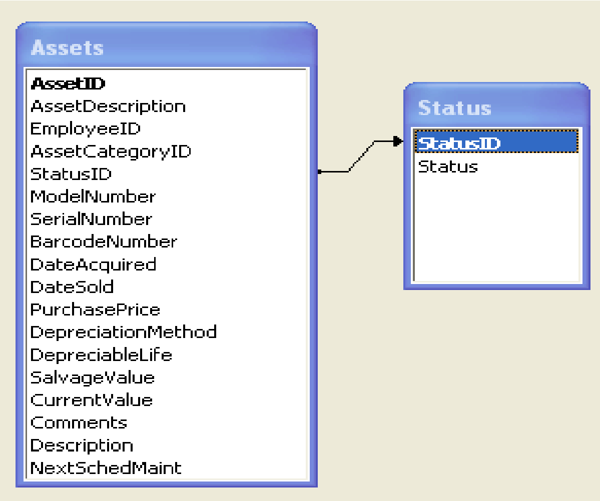


Figure 1.5 : - One-to-One relationship ER diagram

√ One-to-many

In this many records in one table corresponds to the one record in other table. Example: - Every one customer can have multiple sales. So there exist one-to-many relationships between customer and sales table.

One “Asset” can have multiple “Maintenance”. So “Asset” entity has one-to-many relationship between them as the ER model shows below.

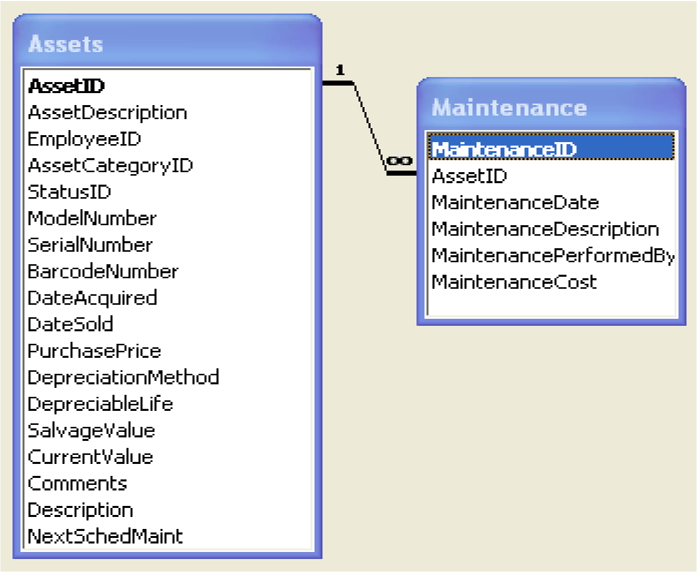


Figure 1.6 : - One-to-Many Relationship ER diagram

√ Many-to-many

In this one record in one table corresponds to many rows in other table and also vice-versa. For instance :- In a company one employee can have many skills like java , c# etc and also one skill can belong to many employees.

Given below is a sample of many-to-many relationship. One employee can have knowledge of multiple “Technology”. So in order to implement this we have one more table “EmployeeTechnology” which is linked to the primary key of “Employee” and “Technology” table.

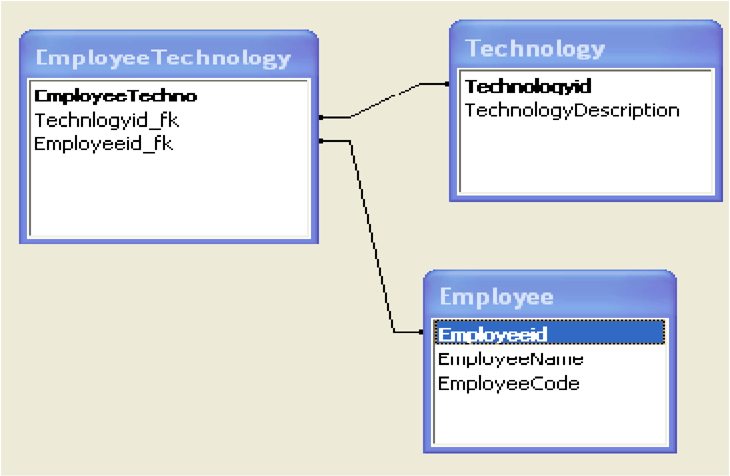


Figure 1.7 : - Many-to-Many Relationship ER diagram

What is normalization? What are different type of normalization?

Note :- A regular .NET programmer working on projects often stumbles in this question ,which is but obvious.Bad part is sometimes interviewer can take this as a very basic question to be answered and it can be a turning point for the interview.So let's cram it.

It is set of rules that has been established to aid in the design of tables that are meant to be connected through relationships. This set of rules is known as Normalization.

Benefits of Normalizing your database include:

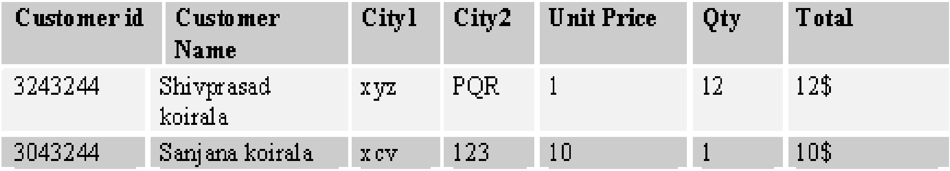
- √ Avoiding repetitive entries
- √ Reducing required storage space
- √ Preventing the need to restructure existing tables to accommodate new data.
- √ Increased speed and flexibility of queries, sorts, and summaries.

Note :- During interview people expect to answer maximum of three normal forms and thats what is expected practically.Actually you can normalize database to fifth normal *form.But believe this book answering three normal forms will put you in decent shape during interview.*

Following are the three normal forms :-

First Normal Form

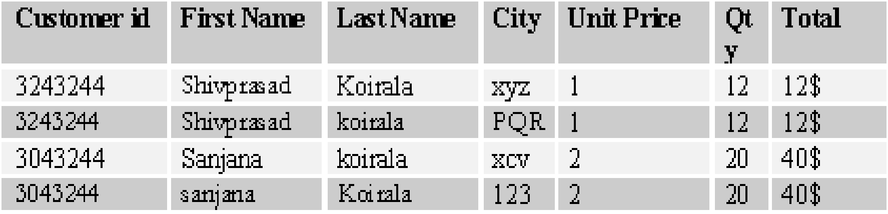
For a table to be in first normal form, data must be broken up into the smallest units possible.In addition to breaking data up into the smallest meaningful values, tables in first normal form should not contain repetitions groups of fields.



Customer id	Customer Name	City1	City2	Unit Price	Qty	Total
3243244	Shivprasad koirala	xyz	PQR	1	12	12$
3043244	Sanjana koirala	xcv	123	10	1	10$

Figure 1.8 :- Repeating groups example

For in the above example city1 and city2 are repeating.In order this table to be in First normal form you have to modify the table structure as follows.Also not that the Customer Name is now broken down to first name and last name (First normal form data should be broken down to smallest unit).



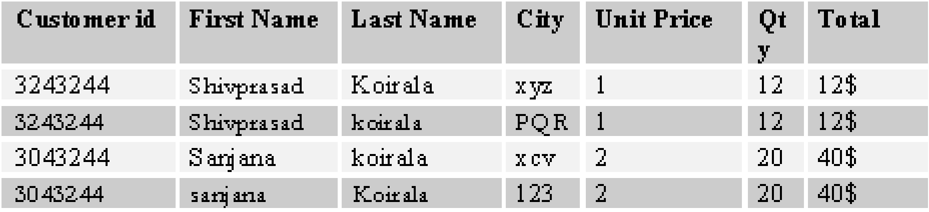
Customer id	First Name	Last Name	City	Unit Price	Qty	Total
3243244	Shivprasad	Koirala	xyz	1	12	12$
3243244	Shivprasad	koirala	PQR	1	12	12$
3043244	Sanjana	koirala	xcv	2	20	40$
3043244	sanjana	Koirala	123	2	20	40$

Figure 1.9 :- Customer table normalized to first normal form

Second Normal form

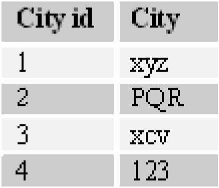
The second normal form states that each field in a multiple field primary keytable must be directly related to the entire primary key. Or in other words,each non-key field should be a fact about all the fields in the primary key.

In the above table of customer , city is not linked to any primary field.



Customer id	First Name	Last Name	City	Unit Price	Qt y	Total
3243244	Shivprasad	Koirala	xyz	1	12	12$
3243244	Shivprasad	koirala	PQR	1	12	12$
3043244	Sanjana	koirala	xcv	2	20	40$
3043244	sanjana	Koirala	123	2	20	40$

Figure 1.10 :- Normalized customer table.



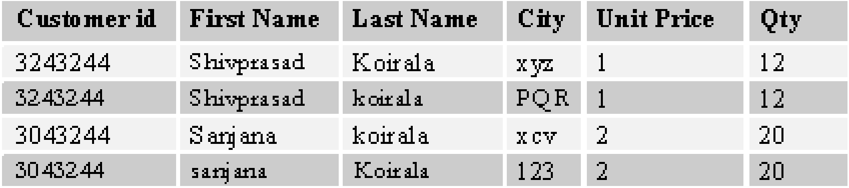
City id	City
1	xyz
2	PQR
3	xcv
4	123

Figure 1.11 :- City is now shifted to a different master table.

That takes our database to a second normal form.

Third normal form

A non-key field should not depend on other Non-key field.The field "Total" is dependent on "Unit price" and "qty".



Customer id	First Name	Last Name	City	Unit Price	Qty
3243244	Shivprasad	Koirala	xyz	1	12
3243244	Shivprasad	koirala	PQR	1	12
3043244	Sanjana	koirala	xcv	2	20
3043244	sanjana	Koirala	123	2	20

Figure 1.12 :- Fill third normal form

So now the "Total" field is removed and is multiplication of Unit price * Qty.

What is denormalization ?

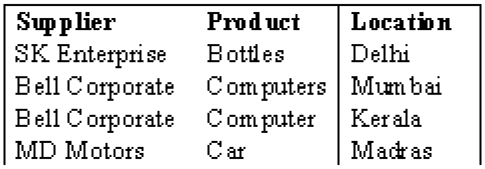
Denormalization is the process of putting one fact in numerous places (its vice-versa of normalization).Only one valid reason exists for denormalizing a relational design - to enhance performance.The sacrifice to performance is that you increase redundancy in database.

(DB) Can you explain Fourth Normal Form?

> *Note: - Whenever interviewer is trying to go above third normal form it can have two reasons ego or to fail you. Three normal forms are really enough, practically anything more than that is an overdose.*

In fourth normal form it should not contain two or more independent multi-valued facts about an entity and it should satisfy "Third Normal form".

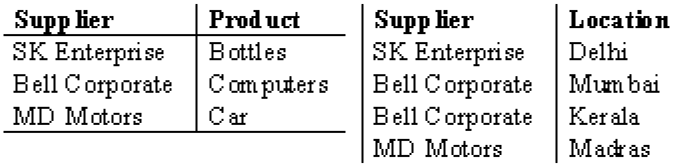
So let's try to see what multi-valued facts are. If there are two or more many-to-many relationship in one entity and they tend to come to one place is termed as "multi-valued facts".



Supplier	Product	Location
SK Enterprise	Bottles	Delhi
Bell Corporate	Computers	Mumbai
Bell Corporate	Computer	Kerala
MD Motors	Car	Madras

Figure 1.13 : - Multi-valued facts

In the above table you can see there are two many-to-many relationship between "Supplier" / "Product" and "Supplier" / "Location" (or in short multi-valued facts). In order the above example satisfies fourth normal form, both the many-to-many relationship should go in different tables.



Supplier	Product
SK Enterprise	Bottles
Bell Corporate	Computers
MD Motors	Car

Supplier	Location
SK Enterprise	Delhi
Bell Corporate	Mumbai
Bell Corporate	Kerala
MD Motors	Madras

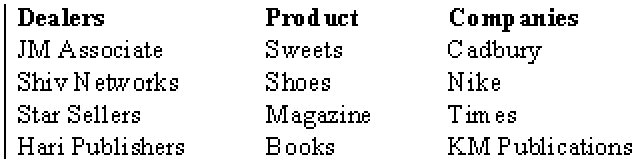
Figure 1.14 : - Normalized to Fourth Normal form.

(DB) Can you explain Fifth Normal Form?

Note: - UUUHHH if you get his question after joining the company do ask him, did he himself really use it?

Fifth normal form deals with reconstructing information from smaller pieces of information. These smaller pieces of information can be maintained with less redundancy.

Example: - “Dealers” sells “Product” which can be manufactured by various “Companies”. “Dealers” in order to sell the “Product” should be registered with the “Company”. So these three entities have very much mutual relationship within them.



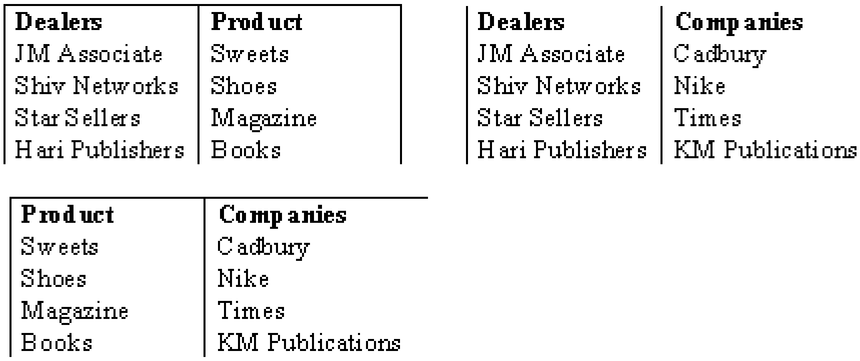
Dealers	Product	Companies
JM Associate	Sweets	Cadbury
Shiv Networks	Shoes	Nike
Star Sellers	Magazine	Times
Hari Publishers	Books	KM Publications

Figure 1.15 : - Not in Fifth Normal Form.

The above table shows some sample data. If you observe closely a single record is created using lot of small information’s. For instance: - “JM Associate” can sell sweets in the following two conditions:-

- √ “JM Associate” should be an authorized dealer of “Cadbury”.
- √ “Sweets” should be manufactured by “Cadbury” company.

These two smaller information forms one record of the above given table. So in order that the above information to be “Fifth Normal Form” all the smaller information should be in three different places. Below is the complete fifth normal form of the database.



Dealers	Product
JM Associate	Sweets
Shiv Networks	Shoes
Star Sellers	Magazine
Hari Publishers	Books

Dealers	Companies
JM Associate	Cadbury
Shiv Networks	Nike
Star Sellers	Times
Hari Publishers	KM Publications

Product	Companies
Sweets	Cadbury
Shoes	Nike
Magazine	Times
Books	KM Publications

Figure 1.16 : - Complete Fifth Normal Form

(DB) What’s the difference between Fourth and Fifth normal form?

> *Note: - There is huge similarity between Fourth and Fifth normal form i.e. they address the problem of “Multi-Valued facts”.*

“Fifth normal form” multi-valued facts are interlinked and “Fourth normal form” values are independent. For instance in the above two questions “Supplier/Product” and “Supplier/Location” are not linked. While in fifth form the “Dealer/Product/Companies” are completely linked.

(DB) Have you heard about sixth normal form?

> *Note: - Arrrrggghhh yes there exists a sixth normal form also. But note guys you can skip this statement or just in case if you want to impress the interviewer.*

If you want relational system in conjunction with time you use sixth normal form. At this moment SQL Server does not supports it directly.

What is Extent and Page?

Twist: - What's the relationship between Extent and Page?

Extent is a basic unit of storage to provide space for tables. Every extent has number of data pages. As new records are inserted new data pages are allocated. There are eight data pages in an extent. So as soon as the eight pages are consumed it allocates new extent with data pages.

While extent is basic unit storage from database point of view, page is a unit of allocation within extent.

(DB)What are the different sections in Page?

Page has three important sections:-

- √ Page header
- √ Actual data i.e. Data row
- √ Row pointers or Row offset

Page header has information like timestamp, next page number, previous page number etc.

Data rows are where your actual row data is stored. For every data row there is a row offset which point to that data row.

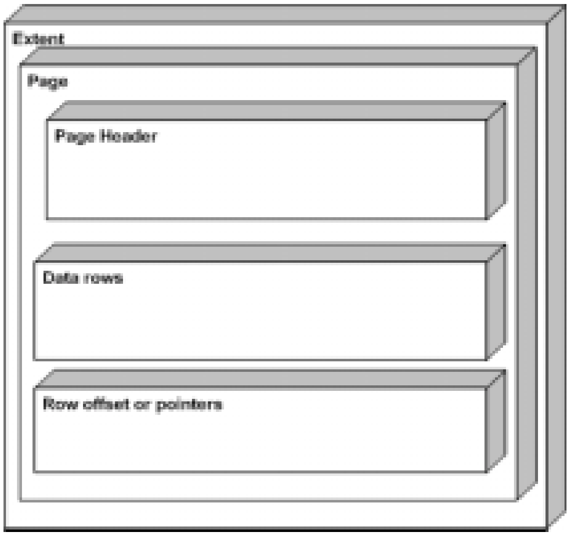


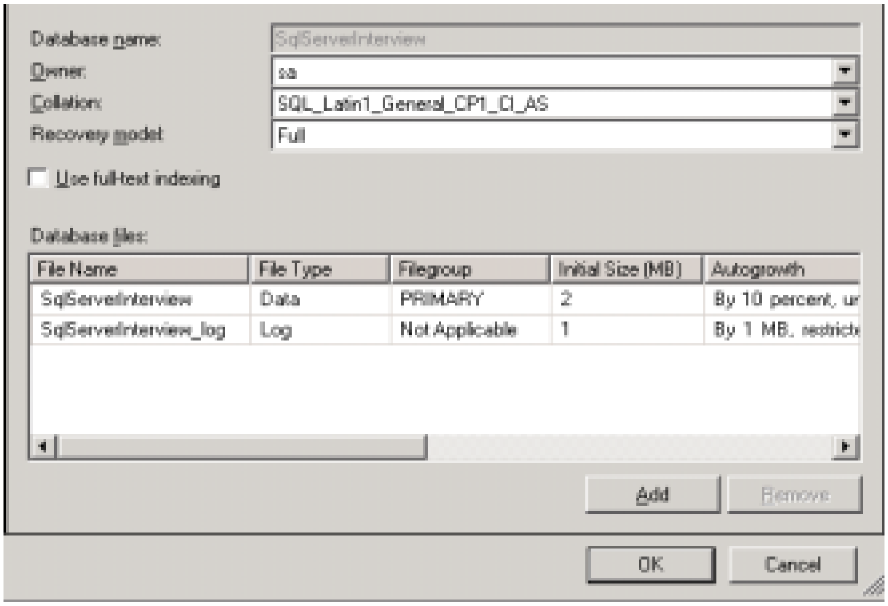
Figure 1.17 : - General view of a Extent

What are page splits?

Pages are contained in extent. Every extent will have around eight data pages. But all the eight data pages are not created at once; it's created depending on data demand. So when a page becomes full it creates a new page, this process is called as "Page Split".

In which files does actually SQL Server store data?

Any SQL Server database is associated with two kinds of files: - .MDF and .LDF..MDF files are actual physical database file where your data is stored finally. .LDF (LOG) files are actually data which is recorded from the last time data was committed in database.



Database name: SqlServerInterview

Owner: sa

Collation: SQL_Latin1_General_CP1_CI_AS

Recovery model: Full

☐ Use full-text indexing

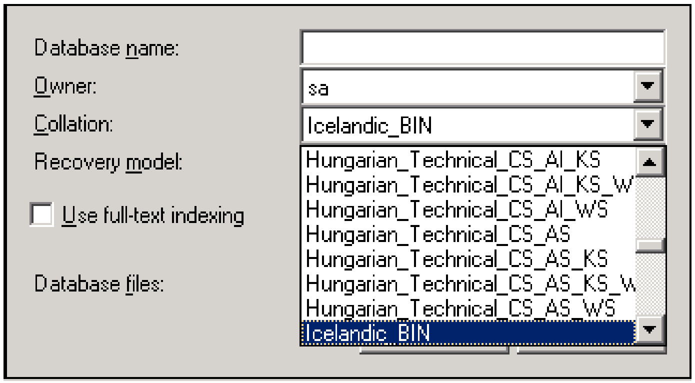
Database files:

File Name	File Type	Filegroup	Initial Size (MB)	Autogrowth
SqlServerInterview	Data	PRIMARY	2	By 10 percent, ur
SqlServerInterview_log	Log	Not Applicable	1	By 1 MB, restricte

Figure 1.18 : - MDF and LDF files.

What is Collation in SQL Server?

Collation refers to a set of rules that determine how data is sorted and compared. Character data is sorted using rules that define the correct character sequence, with options for specifying case-sensitivity, accent marks, kana character types and character width.



Database name:

Owner: sa

Collation: Icelandic_BIN

Recovery model:

□ Use full-text indexing

Database files:

Hungarian_Technical_CS_AI_KS
Hungarian_Technical_CS_AI_KS_W
Hungarian_Technical_CS_AI_WS
Hungarian_Technical_CS_AS
Hungarian_Technical_CS_AS_KS
Hungarian_Technical_CS_AS_KS_W
Hungarian_Technical_CS_AS_WS
Icelandic_BIN

Figure 1.19 : - Collation according to language

Note:- Different language will have different sort orders.

Case sensitivity

If A and a, B and b, etc. are treated in the same way then it is case-insensitive. A computer treats A and a differently because it uses ASCII code to differentiate the input. The ASCII value of A is 65, while a is 97. The ASCII value of B is 66 and b is 98.

Accent sensitivity

If “a” and “A”, o and “O” are treated in the same way, then it is accent-insensitive. A computer treats “a” and “A” differently because it uses ASCII code for differentiating the input. The ASCII value of “a” is 97 and “A” 225. The ASCII value of “o” is 111 and “O” is 243.

Kana Sensitivity

When Japanese kana characters Hiragana and Katakana are treated differently, it is called Kana sensitive.

Width sensitivity

When a single-byte character (half-width) and the same character when represented as a double-byte character (full-width) are treated differently then it is width sensitive.

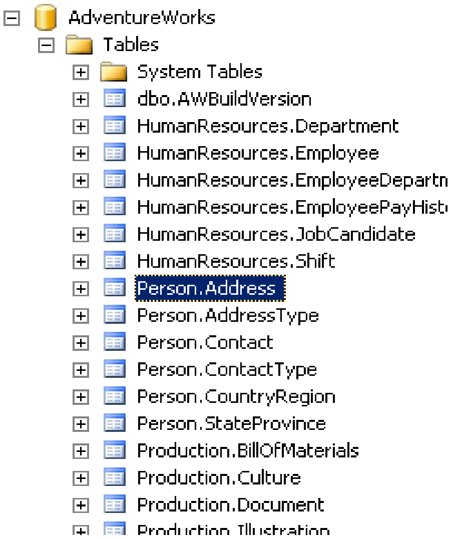
(DB)Can we have a different collation for database and table?

Yes you can specify different collation sequence for both the entity differently.

2. SQL

Note: - This is one of the crazy things which I did not want to put in my book. But when I did sampling of some real interviews conducted across companies I was stunned to find some interviewer judging developers on syntaxes. I know many people will conclude this is childish but it's the interviewer's decision. If you think that this chapter is not useful you can happily skip it. But I think on fresher's level they should not

Note: - I will be heavily using the "AdventureWorks" database which is a sample database shipped (in previous version we had the famous'NorthWind' database sample) with SQL Server 2005. Below is a view expanded from "SQL Server Management Studio".



AdventureWorks
Tables
System Tables
dbo.AWBuildVersion
HumanResources.Department
HumanResources.Employee
HumanResources.EmployeeDepartn
HumanResources.EmployeePayHist
HumanResources.JobCandidate
HumanResources.Shift
Person.Address
Person.AddressType
Person.Contact
Person.ContactType
Person.CountryRegion
Person.StateProvince
Production.BillOfMaterials
Production.Culture
Production.Document
Production.Illustration

Figure 2.1 : - AdventureWorks

Revisiting basic syntax of SQL?

```
CREATE TABLE ColorTable
(code VARCHAR(2),
 ColorValue VARCHAR(16)
)
```

```
INSERT INTO ColorTable (code, colorvalue) VALUES ('b1', 'Brown')

DELETE FROM ColorTable WHERE code = ‘b1'

UPDATE ColorTable SET colorvalue =’Black’ where code=’bl’

DROP TABLE table-name {CASCADE|RESTRICT}

GRANT SELECT ON ColorTable TO SHIVKOIRALA WITH GRANT
OPTION

REVOKE SELECT, INSERT, UPDATE (ColorCode) ON ColorTable FROM
Shivkoirala

COMMIT [WORK]

ROLLBACK [WORK]

Select * from Person.Address

Select AddressLine1, City from Person.Address

Select AddressLine1, City from Person.Address where city ='Sammamish'
```

What are “GRANT” and “REVOKE’ statements?

GRANT statement grants rights to the objects (table). While revoke does the vice-versa of it, it removes rights from the object.

What is Cascade and Restrict in DROP table SQL?

Twist: - What is “ON DELETE CASCADE” and “ON DELETE RESTRICT”?

RESTRICT specifies that table should not be dropped if any dependencies (i.e. triggers, stored procedure, primary key, foreign key etc) exist. So if there are dependencies then error is generated and the object is not dropped.

CASCADE specifies that even if there dependencies go ahead with the drop. That means drop the dependencies first and then the main object also. So if the table has stored procedures and keys (primary and secondary keys) they are dropped first and then the table is finally dropped.

How to import table using “INSERT” statement?

I have made a new temporary color table which is flourished using the below SQL. Structures of both the table should be same in order that this SQL executes properly.

```
INSERT INTO TempColorTable

SELECT code,Color Value

FROM ColorTable
```

What is a DDL, DML and DCL concept in RDBMS world?

DDL (Data definition language) defines your database structure. CREATE and ALTER are DDL statements as they affect the way your database structure is organized.

DML (Data Manipulation Language) lets you do basic functionalities like INSERT, UPDATE, DELETE and MODIFY data in database.

DCL (Data Control Language) controls you DML and DDL statements so that your data is protected and has consistency. COMITT and ROLLBACK are DCL control statements. DCL guarantees ACID fundamentals of a transaction.

Note: - Refer to “Transaction and Locks” chapter.

What are different types of joins in SQL?

INNER JOIN

Inner join shows matches only when they exist in both tables. Example in the below SQL there are two tables Customers and Orders and the inner join in made on Customers.Customerid and Orders.Customerid. So this SQL will only give you result with customers who have orders. If the customer does not have order it will not display that record.

```
SELECT Customers.*, Orders.* FROM Customers INNER JOIN Orders ON
Customers.CustomerID =Orders.CustomerID
```

LEFT OUTER JOIN

Left join will display all records in left table of the SQL statement. In SQL below customers with or without orders will be displayed. Order data for customers without orders appears as NULL values. For example, you want to determine the amount ordered by each customer and you need to see who has not ordered anything as well. You can also see the LEFT OUTER JOIN as a mirror image of the RIGHT OUTER JOIN (Is covered in the next section) if you switch the side of each table.

SELECT Customers., Orders.* FROM Customers LEFT OUTER JOIN Orders ON Customers.CustomerID =Orders.CustomerID*

RIGHT OUTER JOIN

Right join will display all records in right table of the SQL statement. In SQL below all orders with or without matching customer records will be displayed. Customer data for orders without customers appears as NULL values. For example, you want to determine if there are any orders in the data with undefined CustomerID values (say, after a conversion or something like it). You can also see the RIGHT OUTER JOIN as a mirror image of the LEFT OUTER JOIN if you switch the side of each table.

SELECT Customers., Orders.* FROM Customers RIGHT OUTER JOIN Orders ON Customers.CustomerID =Orders.CustomerID*

What is “CROSS JOIN”?

Twist: - What is Cartesian product?

“CROSS JOIN” or “CARTESIAN PRODUCT” combines all rows from both tables. Number of rows will be product of the number of rows in each table. In real life scenario I can not imagine where we will want to use a Cartesian product. But there are scenarios where we would like permutation and combination probably Cartesian would be the easiest way to achieve it.

You want to select the first record in a given set of rows?

*Select top 1 * from sales.salesperson*

How do you sort in SQL?

Using the “ORDER BY” clause, you either sort the data in ascending manner or descending manner.

```
select  * from sales.salesperson order by salespersonid asc

select  * from sales.salesperson order by salespersonid desc
```

How do you select unique rows using SQL?

Using the “DISTINCT” clause. For example if you fire the below give SQL in “AdventureWorks” , first SQL will give you distinct values for cities , while the other will give you distinct rows.

```
select distinct city  from person.address

select distinct *  from person.address
```

Can you name some aggregate function is SQL Server?

Some of them which every interviewer will expect:-

- √ AVG: Computes the average of a specific set of values, which can be an expression list or a set of data records in a table.
- √ SUM: Returns the sum of a specific set of values, which can be an expression list or a set of data records in a table.
- √ COUNT: Computes the number of data records in a table.
- √ MAX: Returns the maximum value from a specific set of values, which can be an expression list or a set of data records in a table.
- √ MIN: Returns the minimum value from a specific set of values, which can be an expression list or a set of data records in a table.

What is the default “SORT” order for a SQL?

ASCENDING

What is a self-join?

If you want to join two instances of the same table you can use self-join.

What's the difference between DELETE and TRUNCATE ?

Following are difference between them:

- √ DELETE TABLE syntax logs the deletes thus making the delete operations low. TRUNCATE table does not log any information but it logs information about deallocation of data page of the table. So TRUNCATE table is faster as compared to delete table.
- √ DELETE table can be rolled back while TRUNCATE can not be.
- √ DELETE table can have criteria while TRUNCATE can not.
- √ TRUNCATE table can not have triggers.

Select addresses which are between '1/1/2004' and '1/4/2004'?

Select * from Person.Address where modifieddate between '1/1/2004' and '1/4/2004'

What are Wildcard operators in SQL Server?

Twist: - What is like clause in SQL?

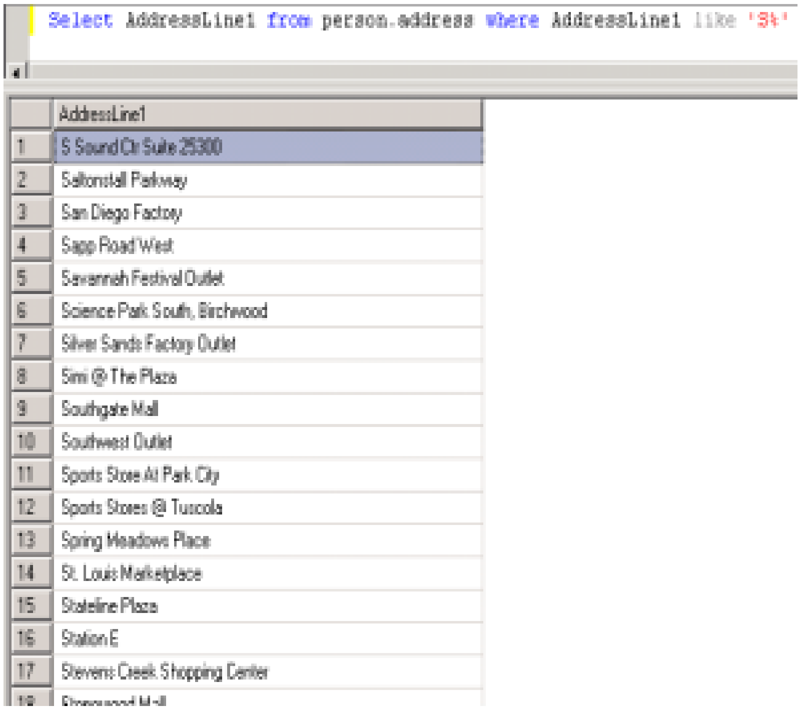
Note: - For showing how the wildcards work I will be using the "person.address" table in adventureworks.

There are basically two types of operator:-

"%" operator (During Interview you can spell it as "Percentage Operator").

"%" operator searches for one or many occurrences. So when you fire a query using "%" SQL Server searches for one or many occurrences. In the below SQL I have applied "%" operator to "S" character.

```
Select AddressLine1 from person.address where AddressLine1 like 'S%'
```



```
Select AddressLine1 from person.address where AddressLine1 like 'S%'
```

	AddressLine1
1	S Sound Ctr Suite 25300
2	Saltonstall Parkway
3	San Diego Factory
4	Sapp Road West
5	Savannah Festival Outlet
6	Science Park South, Birchwood
7	Silver Sands Factory Outlet
8	Simi @ The Plaza
9	Southgate Mall
10	Southwest Outlet
11	Sports Store At Park City
12	Sports Stores @ Tuscola
13	Spring Meadows Place
14	St. Louis Marketplace
15	Stateline Plaza
16	Station E
17	Stevens Creek Shopping Center
18	Stonewood Mall

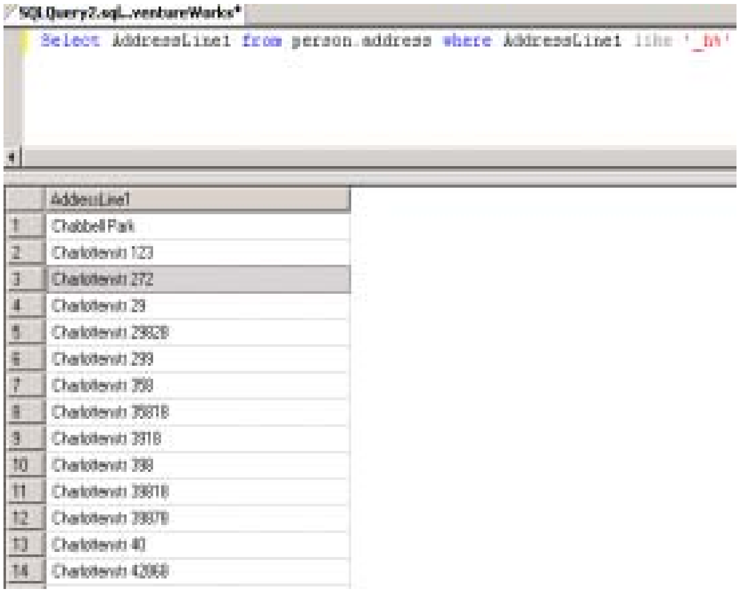
Figure 2.2 : - "%" operator in action.

"_" operator (During Interview you spell it as "Underscore Operator").

"_" operator is the character defined at that point. In the below sample I have fired a query

Select AddressLine1 from person.address where AddressLine1 like '_h%'

So all data where second letter is "h" is returned.



```
Select AddressLine1 from person.address where AddressLine1 like '_h%'
```

```
    AddressLine1
1   Chabbell Park
2   Charlottenstr 123
3   Charlottenstr 272
4   Charlottenstr 29
5   Charlottenstr 29828
6   Charlottenstr 299
7   Charlottenstr 358
8   Charlottenstr 35818
9   Charlottenstr 3818
10  Charlottenstr 398
11  Charlottenstr 39818
12  Charlottenstr 39878
13  Charlottenstr 40
14  Charlottenstr 42068
```

Figure 2.3 : - "_" operator in action

What's the difference between "UNION" and "UNION ALL" ?

UNION SQL syntax is used to select information from two tables. But it selects only distinct records from both the table. , while UNION ALL selects all records from both the tables.

To explain it practically below are two images one fires "UNION" and one "UNION ALL" on the "person.address" table of the "AdventureWorks" database.

*Select * from person.address*

Union

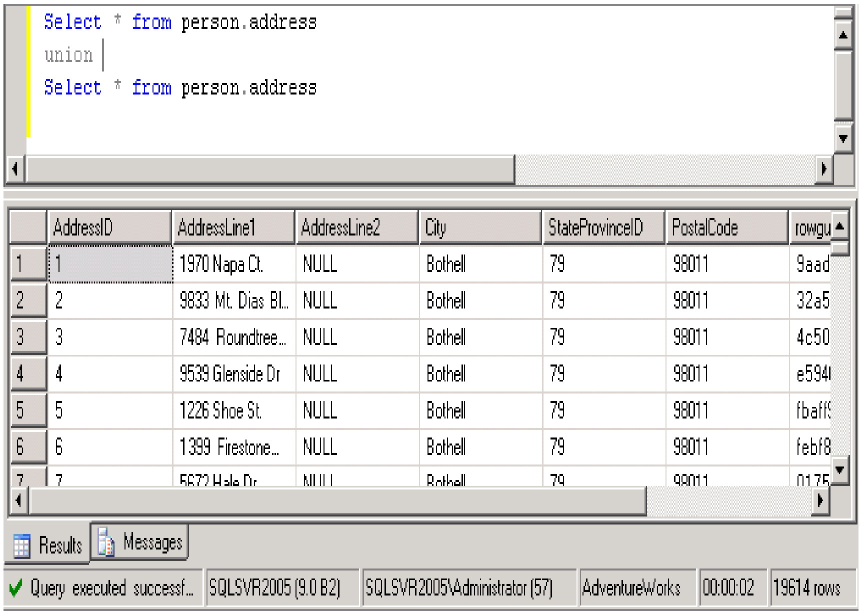
*Select * from person.address*

This returns 19614 rows (that's mean it removes duplicates)

*Select * from person.address*

```
union all
Select * from person.address
```

This returns 39228 rows (“unionall” does not check for duplicates so returns double the record show up)

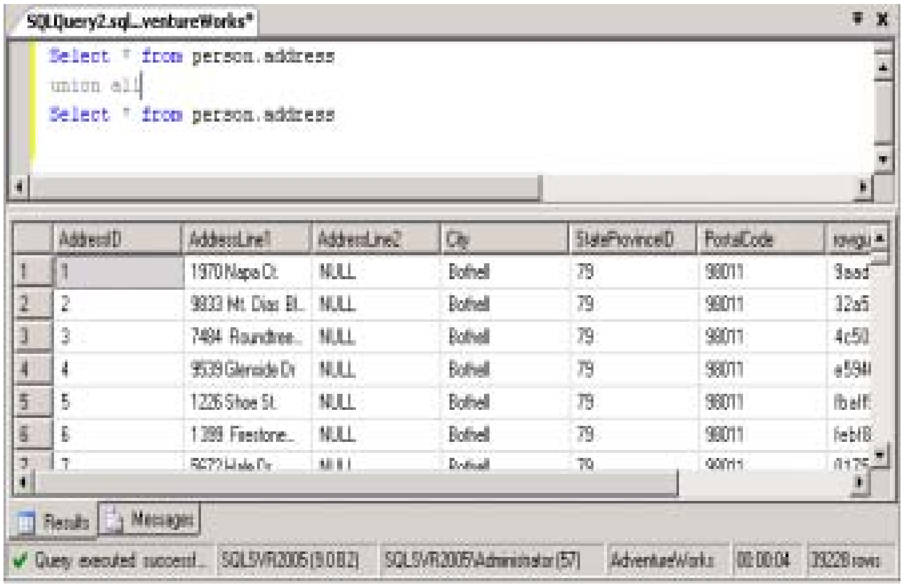


```
Select * from person.address
union
Select * from person.address
```

	AddressID	AddressLine1	AddressLine2	City	StateProvinceID	PostalCode	rowgu
1	1	1970 Napa Ct.	NULL	Bothell	79	98011	9aad
2	2	9833 Mt. Dias Bl...	NULL	Bothell	79	98011	32a5
3	3	7484 Roundtree...	NULL	Bothell	79	98011	4c50
4	4	9539 Glenside Dr	NULL	Bothell	79	98011	e594
5	5	1226 Shoe St.	NULL	Bothell	79	98011	fbaf
6	6	1399 Firestone...	NULL	Bothell	79	98011	febf8
7	7	5672 Hale Dr	NULL	Bothell	79	98011	0175

✓ Query executed successf... | SQLSVR2005 (9.0 B2) | SQLSVR2005\Administrator (57) | AdventureWorks | 00:00:02 | 19614 rows

Figure 2.4 : - Union keyword in action (19614 rows)



SQLQuery2.sql...ventureWorks*

```
Select * from person.address
union all
Select * from person.address
```

	AddressID	AddressLine1	AddressLine2	City	StateProvinceID	PostalCode	rowgu
1	1	1970 Napa Ct.	NULL	Bothell	79	98011	9aad
2	2	9833 Mt. Dias Bl...	NULL	Bothell	79	98011	32a5
3	3	7484 Roundtree...	NULL	Bothell	79	98011	4c50
4	4	9539 Glenside Dr	NULL	Bothell	79	98011	e594
5	5	1226 Shoe St.	NULL	Bothell	79	98011	fbaff
6	6	1399 Firestone...	NULL	Bothell	79	98011	febf8
7	7	5672 Hale Dr.	NULL	Bothell	79	98011	0175

Query executed successf... | SQLSVR2005 (9.0 SP2) | SQLSVR2005\Administrator (57) | AdventureWorks | 00:00:04 | 39228 rows

Figure 2.5 : - Union All in action (39228 rows)

Note: - Selected records should have same data type or else the syntax will not work.

Note: - In the coming questions you will see some 5 to 6 questions on cursors. Though not a much discussed topic but still from my survey 5% of interviews have asked questions on cursors. So let's leave no stone for the interviewer to reject us.

What are cursors and what are the situations you will use them?

SQL statements are good for set at a time operation. So it is good at handling set of data. But there are scenarios where you want to update row depending on certain criteria. You will loop through all rows and update data accordingly. There's where cursors come in to picture.

What are the steps to create a cursor?

Below are the basic steps to execute a cursor.

- √ Declare
- √ Open
- √ Fetch
- √ Operation
- √ Close and Deallocate

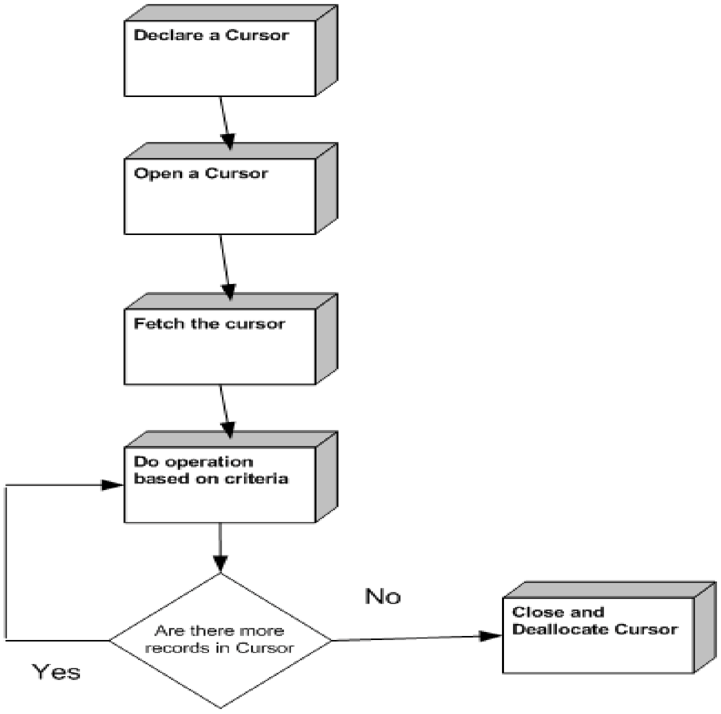

Declare a Cursor
Open a Cursor
Fetch the cursor
Do operation based on criteria
Are there more records in Cursor
No
Yes
Close and Deallocate Cursor

Figure 2.6 : - Steps to process a cursor

This is a small sample which uses the "person.address" class. This T-SQL program will only display records which have "@Provinceid" equal to "7".

DECLARE @provinceid int

-- *Declare Cursor*

```
DECLARE provincecursor CURSOR FOR
SELECT stateprovinceid
FROM Person.Address
-- Open cursor
OPEN provincecursor
-- Fetch data from cursor in to variable
FETCH NEXT FROM provincecursor
INTO @provinceid
WHILE @@FETCH_STATUS = 0
BEGIN
-- Do operation according to row value
if @Provinceid=7
begin
PRINT @Provinceid
end
-- Fetch the next cursor
FETCH NEXT FROM provincecursor
INTO @provinceid
END
-- Finally do not forget to close and deallocate the cursor
CLOSE provincecursor
DEALLOCATE provincecursor
```

What are the different Cursor Types?

Cursor types are assigned when we declare a cursor.

```
DECLARE cursor_name CURSOR
```

```
[LOCAL | GLOBAL]
[FORWARD_ONLY | SCROLL]
[STATIC | KEYSET | DYNAMIC | FAST_FORWARD]
[READ_ONLY | SCROLL_LOCKS | OPTIMISTIC]
[TYPE_WARNING]
FOR select_statement
[FOR UPDATE [OF column_list]]
```

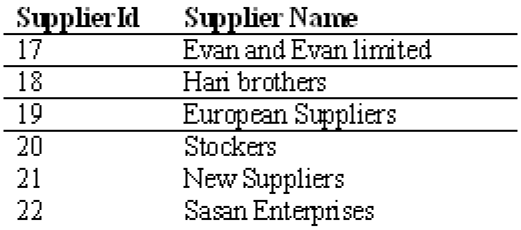
STATIC

STATIC cursor is a fixed snapshot of a set of rows. This fixed snapshot is stored in a temporary database. As the cursor is using private snapshot any changes to the set of rows external will not be visible in the cursor while browsing through it. You can define a static cursor using “STATIC” keyword.

```
DECLARE cusorname CURSOR STATIC
FOR SELECT * from tablename
WHERE column1 = 2
```

KEYSET

In KEYSET the key values of the rows are saved in tempdb. For instance let’s say the cursor has fetched the following below data. So only the “supplierid” will be stored in the database. Any new inserts happening is not reflected in the cursor. But any updates in the key-set values are reflected in the cursor. Because the cursor is identified by key values you can also absolutely fetch them using “FETCH ABSOLUTE 12 FROM mycursor”



SupplierId	Supplier Name
17	Evan and Evan limited
18	Hari brothers
19	European Suppliers
20	Stockers
21	New Suppliers
22	Sasan Enterprises

Figure 2.7 : - Key Set Data

DYNAMIC

In DYNAMIC cursor you can see any kind of changes happening i.e. either inserting new records or changes in the existing and even deletes. That's why DYNAMIC cursors are slow and have least performance.

FORWARD_ONLY

As the name suggest they only move forward and only a one time fetch is done. In every fetch the cursor is evaluated. That means any changes to the data are known, until you have specified "STATIC" or "KEYSET".

FAST_FORWARD

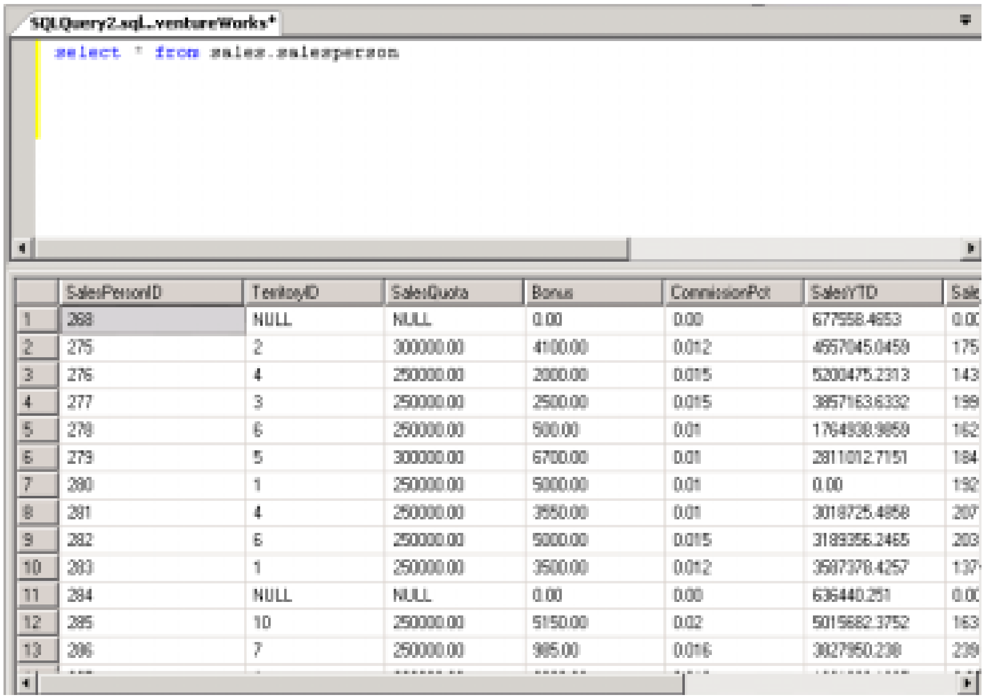
These types of cursor are forward only and read-only and in every fetch they are not re-evaluated again. This makes them a good choice to increase performance.

What are "Global" and "Local" cursors?

Cursors are global for a connection. By default cursors are global. That means you can declare a cursor in one stored procedure and access it outside also. Local cursors are accessible only inside the object (which can be a stored procedure, trigger or a function). You can declare a cursor as "Local" or "Global" in the "DECLARE" cursor syntax. Refer the "DECLARE" statement of the cursor in the previous sections.

What is "Group by" clause?

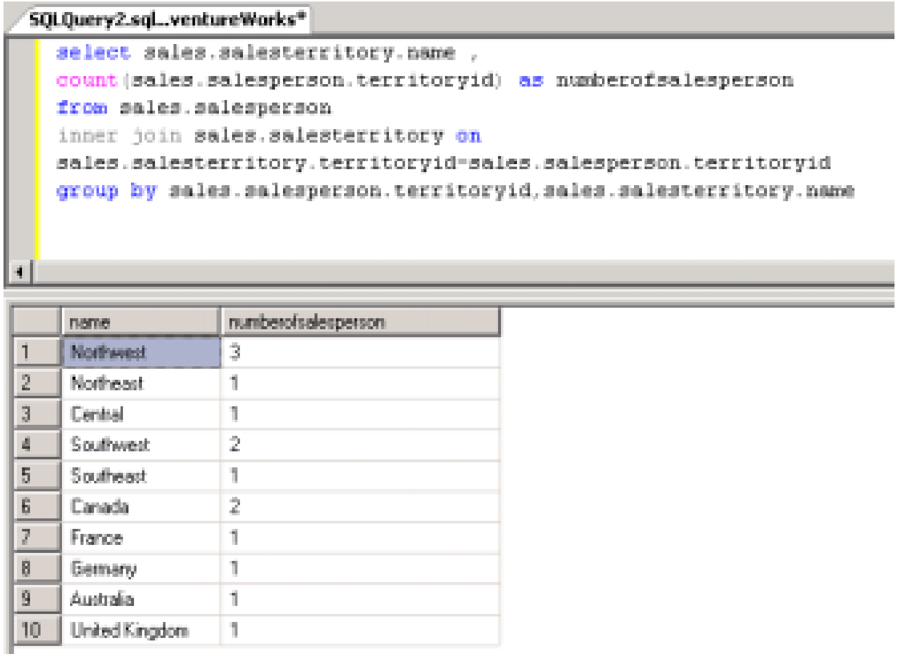
"Group by" clause group similar data so that aggregate values can be derived. In "AdventureWorks" there are two tables "Salesperson" and "Salesterritory". In below figure "Actual data" is the complete view of "Salesperson". But now we want a report that per territory wise how many sales people are there. So in the second figure I made a group by on territory id and used the "count" aggregate function to see some meaningful data. "Northwest" has the highest number of sales personnel.



```
select * from sales.salesperson
```

	SalesPersonID	TerritoryID	SalesQuota	Bonus	CommissionPct	SalesYTD	Sale
1	268	NULL	NULL	0.00	0.00	677558.4653	0.00
2	275	2	300000.00	4100.00	0.012	4557045.0459	175
3	276	4	250000.00	2000.00	0.015	5200475.2313	143
4	277	3	250000.00	2500.00	0.015	3857163.6332	199
5	278	6	250000.00	500.00	0.01	1764938.9859	162
6	279	5	300000.00	6700.00	0.01	2811012.7151	184
7	280	1	250000.00	5000.00	0.01	0.00	192
8	281	4	250000.00	3550.00	0.01	3018725.4858	207
9	282	6	250000.00	5000.00	0.015	3189356.2465	203
10	283	1	250000.00	3500.00	0.012	3587378.4257	137
11	284	NULL	NULL	0.00	0.00	636440.251	0.00
12	285	10	250000.00	5150.00	0.02	5015682.3752	163
13	286	7	250000.00	985.00	0.016	3827950.238	239

Figure 2.8 : - Actual Data



```
select sales.salesterritory.name ,
count(sales.salesperson.territoryid) as numberofsalesperson
from sales.salesperson
inner join sales.salesterritory on
sales.salesterritory.territoryid=sales.salesperson.territoryid
group by sales.salesperson.territoryid,sales.salesterritory.name
```

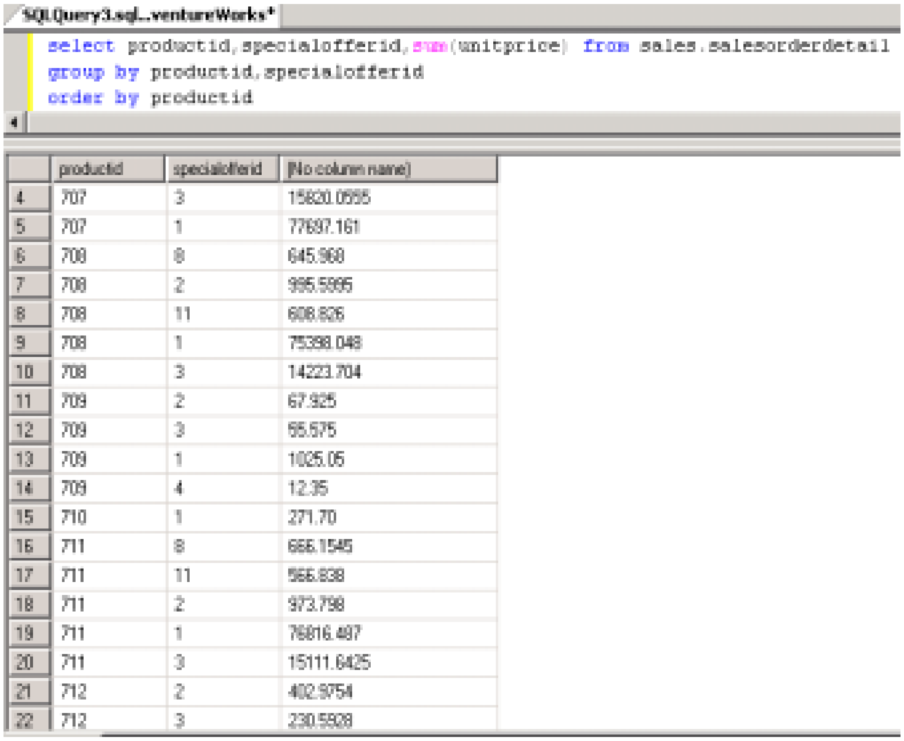
	name	numberofsalesperson
1	Northwest	3
2	Northeast	1
3	Central	1
4	Southwest	2
5	Southeast	1
6	Canada	2
7	France	1
8	Germany	1
9	Australia	1
10	United Kingdom	1

Figure 2.9 : - Group by applied

What is ROLLUP?

ROLLUP enhances the total capabilities of "GROUP BY" clause.

Below is a GROUP BY SQL which is applied on "SalesorderDetail" on "Productid" and "Specialofferid". You can see 707,708,709 etc products grouped according to "Specialofferid" and the third column represents total according to each pair of "Productid" and "Specialofferid". Now you want to see sub-totals for each group of "Productid" and "Specialofferid".



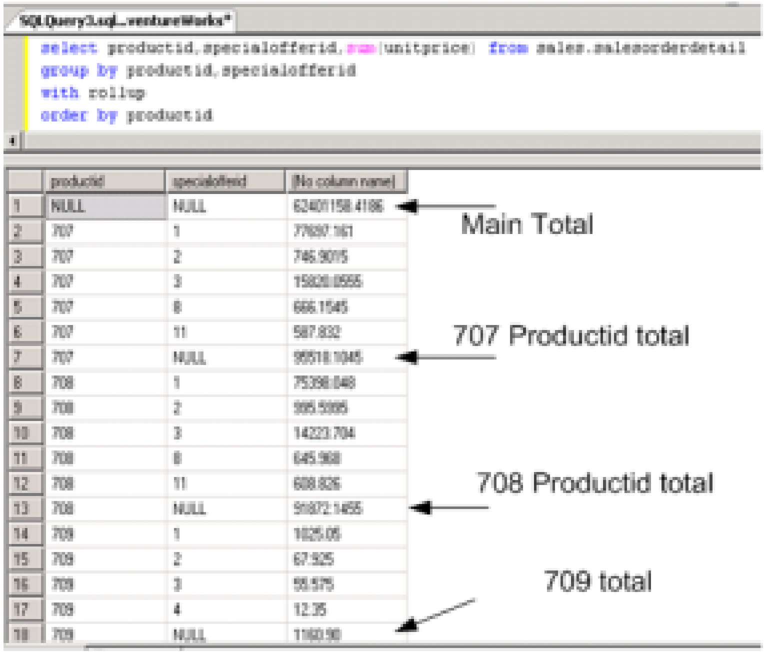
SQLQuery3.sql...ventureWorks*

```
select productid,specialofferid,sum(unitprice) from sales.salesorderdetail
group by productid,specialofferid
order by productid
```

	productid	specialofferid	(No column name)
4	707	3	15820.0555
5	707	1	77697.161
6	708	8	645.968
7	708	2	995.5995
8	708	11	608.826
9	708	1	75398.048
10	708	3	14223.704
11	709	2	67.925
12	709	3	95.575
13	709	1	1025.05
14	709	4	12.35
15	710	1	271.70
16	711	8	666.1545
17	711	11	566.838
18	711	2	973.798
19	711	1	76816.487
20	711	3	15111.6425
21	712	2	402.9754
22	712	3	230.5928

Figure 2.10: - Salesorder displayed with out ROLLUP

So after using ROLLUP you can see the sub-total. The first row is the grand total or the main total, followed by sub-totals according to each combination of "Productid" and "Specialofferid". ROLLUP retrieves a result set that contains aggregates for a hierarchy of values in selected columns.



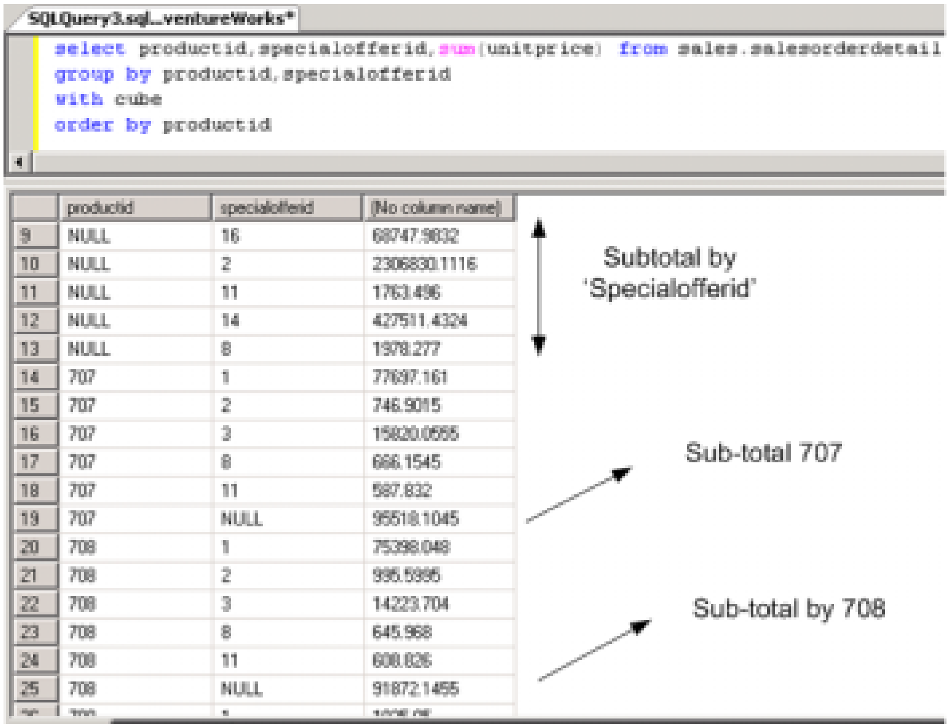
```
select productid,specialofferid,sum(unitprice) from sales.salesorderdetail
group by productid,specialofferid
with rollup
order by productid
```

	productid	specialofferid	(No column name)
1	NULL	NULL	62401158.4186
2	707	1	77697.161
3	707	2	746.9015
4	707	3	15820.0555
5	707	8	666.1545
6	707	11	587.832
7	707	NULL	95518.1045
8	708	1	75398.048
9	708	2	995.5995
10	708	3	14223.704
11	708	8	645.968
12	708	11	608.826
13	708	NULL	91872.1455
14	709	1	1025.05
15	709	2	67.925
16	709	3	55.575
17	709	4	12.35
18	709	NULL	1160.90

Figure 2.11: - Subtotal according to product using ROLLUP

What is CUBE?

CUBE retrieves a result set that contains aggregates for all combinations of values in the selected columns. ROLLUP retrieves a result set that contains aggregates for a hierarchy of values in selected columns.



```
select productid,specialofferid,sum(unitprice) from sales.salesorderdetail
group by productid,specialofferid
with cube
order by productid
```

	productid	specialofferid	(No column name)
9	NULL	16	68747.9832
10	NULL	2	2306830.1116
11	NULL	11	1763.496
12	NULL	14	427511.4324
13	NULL	8	1978.277
14	707	1	77697.161
15	707	2	746.9015
16	707	3	15820.0555
17	707	8	666.1545
18	707	11	587.832
19	707	NULL	95518.1045
20	708	1	75398.048
21	708	2	995.5995
22	708	3	14223.704
23	708	8	645.968
24	708	11	608.826
25	708	NULL	91872.1455

Figure 2.12: - CUBE in action

What is the difference between "HAVING" and "WHERE" clause?

"HAVING" clause is used to specify filtering criteria for "GROUP BY", while "WHERE" clause applies on normal SQL.

In the above example we if we want to filter on territory which has sales personnel count above 2.

select sales.salesterritory.name ,

count(sales.salesperson.territoryid) as numberofsalesperson

from sales.salesperson

inner join sales.salesterritory on

sales.salesterritory.territoryid=sales.salesperson.territoryid

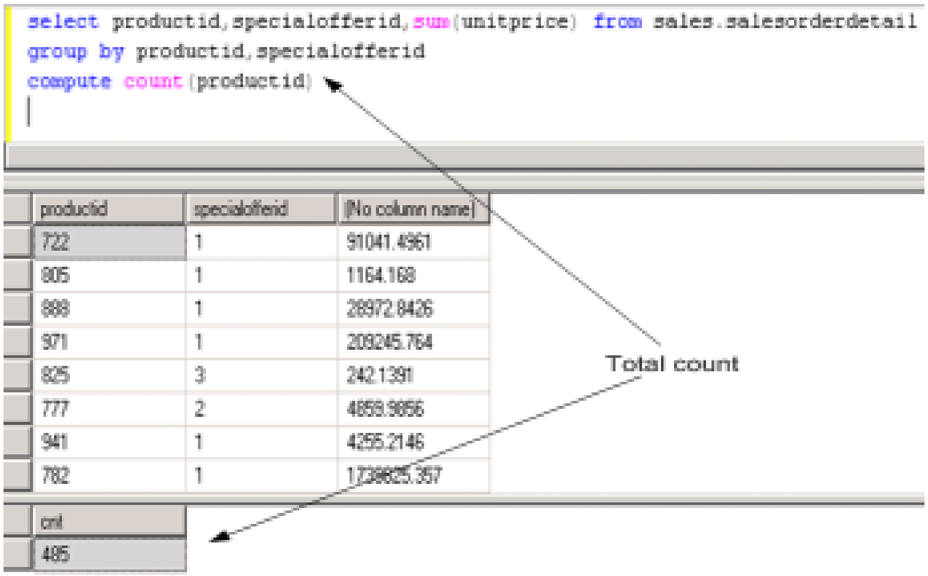
group by sales.salesperson.territoryid,sales.salesterritory.name

having count(sales.salesperson.territoryid) >= 2

Note:- You can see the having clause applied. In this case you can not specify it with "WHERE" clause it will throw an error. In short "HAVING" clause applies filter on a group while "WHERE" clause on a simple SQL.

What is "COMPUTE" clause in SQL?

"COMPUTE "clause is used in SQL to produce subtotals for each group.



```
select productid,specialofferid,sum(unitprice) from sales.salesorderdetail
group by productid,specialofferid
compute count(productid)
```

productid	specialofferid	(No column name)
722	1	91041.4961
805	1	1164.168
888	1	28972.8426
971	1	209245.764
825	3	242.1391
777	2	4859.9856
941	1	4255.2146
782	1	1738825.357

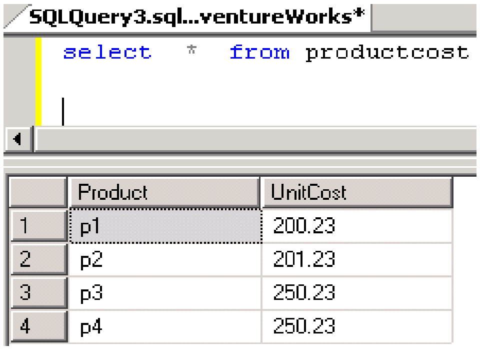
cnt
485

Figure 2.13 : - "Compute" in action

What is "WITH TIES" clause in SQL?

"WITH TIES" clause specifies that additional rows be returned from the base result set with the same value in the ORDER BY columns appearing as the last of the TOP n

(PERCENT) rows. So what does that sentence mean? See the below figure there are four products p1,p2,p3 and p4. “UnitCost” of p3 and p4 are same.

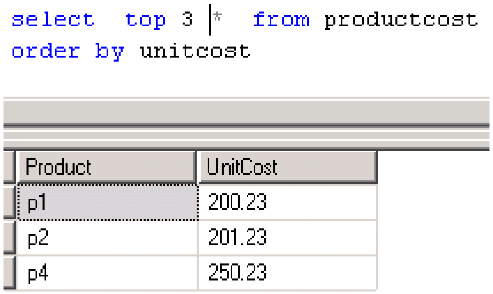


```
select  *  from productcost
```

	Product	UnitCost
1	p1	200.23
2	p2	201.23
3	p3	250.23
4	p4	250.23

Figure 2.14 : - Actual Data

So when we do a TOP 3 on the “ProductCost” table we will see three rows as show below. But even p3 has the same value as p4. SQL just took the TOP 1. So if you want to display tie up data like this you can use “WITH TIES”.

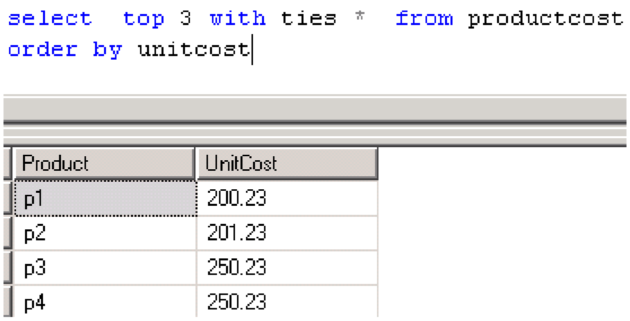


```
select  top 3 *  from productcost
order by unitcost
```

Product	UnitCost
p1	200.23
p2	201.23
p4	250.23

Figure 2.15 : - TOP 3 from the “productcost” table

You can see after firing SQL with “WITH TIES” we are able to see all the products properly.



```
select  top 3 with ties *  from productcost
order by unitcost
```

Product	UnitCost
p1	200.23
p2	201.23
p3	250.23
p4	250.23

Figure 2.16: - WITH TIES in action

Note: - You should have an "ORDER CLAUSE" and "TOP" keyword specified or else "WITH TIES" is not of much use.

What does "SET ROWCOUNT" syntax achieves?

Twist: - What's the difference between "SET ROWCOUNT" and "TOP" clause in SQL?

"SET ROWCOUNT" limits the number of rows returned. Its looks very similar to "TOP" clause, but there is a major difference the way SQL is executed. The major difference between "SET ROWCOUNT" and "TOP" SQL clause is following:-

"SET ROWCOUNT is applied before the order by clause is applied. So if "ORDER BY" clause is specified it will be terminated after the specified number of rows are selected. ORDER BY clause is not executed"

What is a Sub-Query?

A query nested inside a SELECT statement is known as a subquery and is an alternative to complex join statements. A subquery combines data from multiple tables and returns results that are inserted into the WHERE condition of the main query. A subquery is always enclosed within parentheses and returns a column. A subquery can also be referred to as an inner query and the main query as an outer query. JOIN gives better performance than a subquery when you have to check for the existence of records.

For example, to retrieve all EmployeeID and CustomerID records from the ORDERS table that have the EmployeeID greater than the average of the EmployeeID field, you can create a nested query, as shown:

```
SELECT DISTINCT EmployeeID, CustomerID

FROM ORDERS

WHERE EmployeeID > (SELECT AVG(EmployeeID)

FROM ORDERS)
```

What is “Correlated Subqueries”?

A simple subquery retrieves rows that are then passed to the outer query to produce the desired result set. Using Correlated Subqueries, the outer query retrieves rows that are then passed to the subquery. The subquery runs for each row that is processed by the outer query. Below is an example of a simple co-related subquery. You execute it in “AdventureWorks” to see the results.

```
Select salespersonid ,

(Select name from sales.salesterritory where sales.salesterritory.territoryid=
sales.Salesperson.Territoryid)

from sales.Salesperson
```

Note: - Below are some homework questions, you can discuss with your friends for better insight.

What is “ALL” and “ANY” operator?

What is a “CASE” statement in SQL?

What does COLLATE Keyword in SQL signify?

What is CTE (Common Table Expression)?

CTE is a temporary table created from a simple SQL query. You can say it’s a view. Below is a simple CTE created “PurchaseOrderHeaderCTE” from “PurchaseOrderHeader”.

```
WITH PURCHASEORDERHEADERCTE(Orderdate,Status) as

(

Select orderdate,Status from purchasing.PURCHASEORDERHEADER

)
```

```
Select * from PURCHASEORDERHEADERCTE
```

The WITH statement defines the CTE and later using the CTE name I have displayed the CTE data.

Why should you use CTE rather than simple views?

With CTE you can use a recursive query with CTE itself. That’s not possible with views.

What is TRY/CATCH block in T-SQL?

No I am not referring to .NET TRY/CATCH block this is the new way of handling error in SQL Server. For instance in the below T-SQL code any error during delete statement is caught and the necessary error information is displayed.

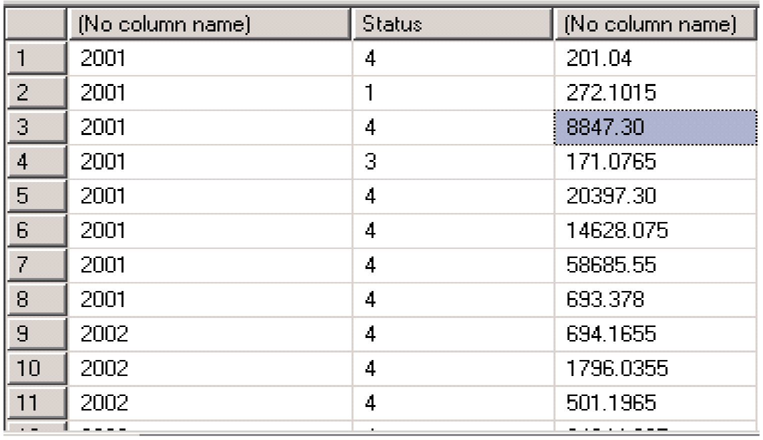
```
BEGIN TRY
  DELETE table1 WHERE id=122
END TRY
BEGIN CATCH
  SELECT
  ERROR_NUMBER() AS ErrNum,
    ERROR_SEVERITY() AS ErrSev,
    ERROR_STATE() as ErrSt,
    ERROR_MESSAGE() as ErrMsg;
END CATCH
```

What is PIVOT feature in SQL Server?

PIVOT feature converts row data to column for better analytical view. Below is a simple PIVOT fired using CTE. Ok the first section is the CTE which is the input and later PIVOT is applied over it.

```
WITH PURCHASEORDERHEADERCTE(Orderdate,Status,Subtotal) as
(
        Select year(orderdate),Status,isnull(Subtotal,0) from
purchasing.PURCHASEORDERHEADER

)

Select Status as OrderStatus,isnull([2001],0) as 'Yr 2001' ,isnull([2002],0) as 'Yr
2002' from PURCHASEORDERHEADERCTE

pivot (sum(Subtotal) for  Orderdate in ([2001],[2002])) as pivoted
```

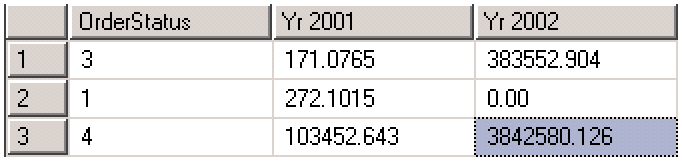
You can see from the above SQL the top WITH statement is the CTE supplied to the PIVOT. After that PIVOT is applied on subtotal and orderdate. You have to specify in what you want the pivot (here it is 2001 and 2002). So below is the output of CTE table.



	(No column name)	Status	(No column name)
1	2001	4	201.04
2	2001	1	272.1015
3	2001	4	8847.30
4	2001	3	171.0765
5	2001	4	20397.30
6	2001	4	14628.075
7	2001	4	58685.55
8	2001	4	693.378
9	2002	4	694.1655
10	2002	4	1796.0355
11	2002	4	501.1965

Figure 2.17 : - CTE output

After the PIVOT is applied you can see the rows are now grouped column wise with the subtotal assigned to each. You can summarize that PIVOT summarizes your data in cross tab format.



	OrderStatus	Yr 2001	Yr 2002
1	3	171.0765	383552.904
2	1	272.1015	0.00
3	4	103452.643	3842580.126

Figure 2.18 : - Pivoted table

What is UNPIVOT?

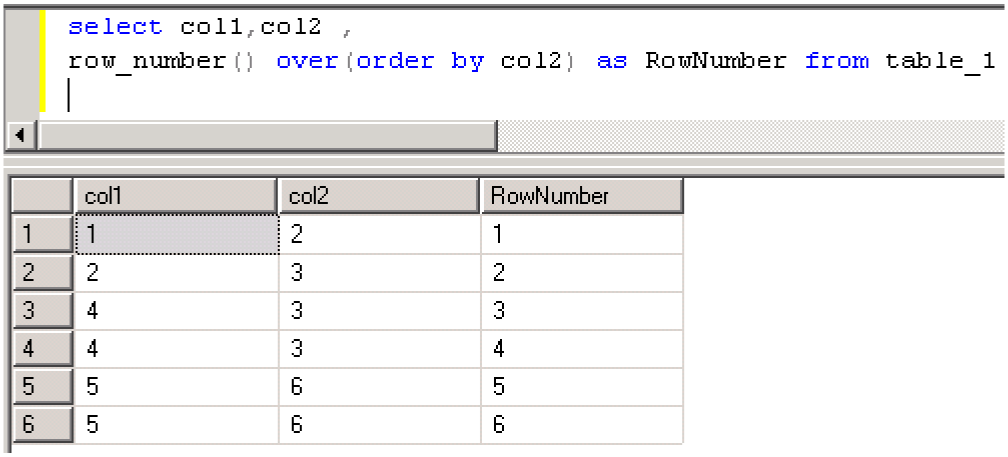
It’s exactly the vice versa of PIVOT. That means you have a PIVOTED data and you want to UNPIVOT it.

What are RANKING functions?

They add columns that are calculated based on a ranking algorithm. These functions include ROW_NUMBER(), RANK(), DENSE_RANK(), and NTILE().

What is ROW_NUMBER()?

The ROW_NUMBER() function adds a column that displays a number corresponding the row's position in the query result . If the column that you specify in the OVER clause is not unique, it still produces an incrementing column based on the column specified in the OVER clause. You can see in the figure below I have applied ROW_NUMBER function over column col2 and you can notice the incrementing numbers generated.



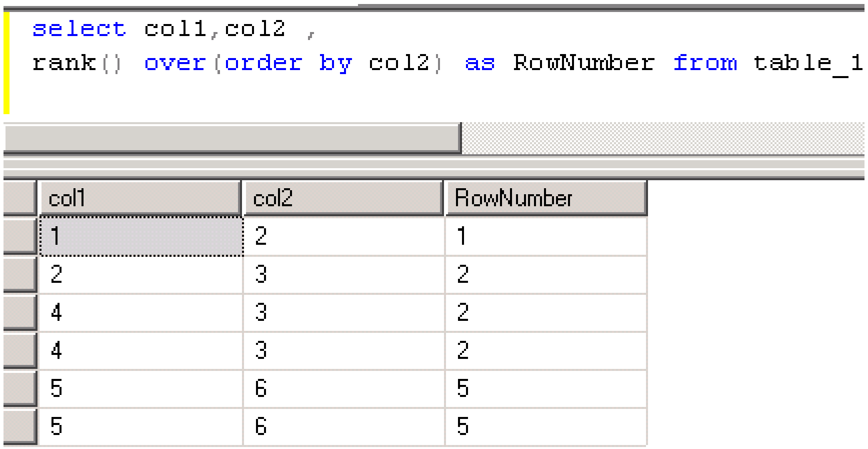
```
select col1,col2 ,
row_number() over(order by col2) as RowNumber from table_1
```

	col1	col2	RowNumber
1	1	2	1
2	2	3	2
3	4	3	3
4	4	3	4
5	5	6	5
6	5	6	6

Figure 2.19 :- ROW_NUMBER in action

What is RANK() ?

The RANK() function works much like the ROW_NUMBER() function in that it numbers records in order. When the column specified by the ORDER BY clause contains unique values, then ROW_NUMBER() and RANK() produce identical results. They differ in the way they work when duplicate values are contained in the ORDER BY expression. ROW_NUMBER will increment the numbers by one on every record, regardless of duplicates. RANK() produces a single number for each value in the result set. You can see for duplicate value it does not increment the row number.



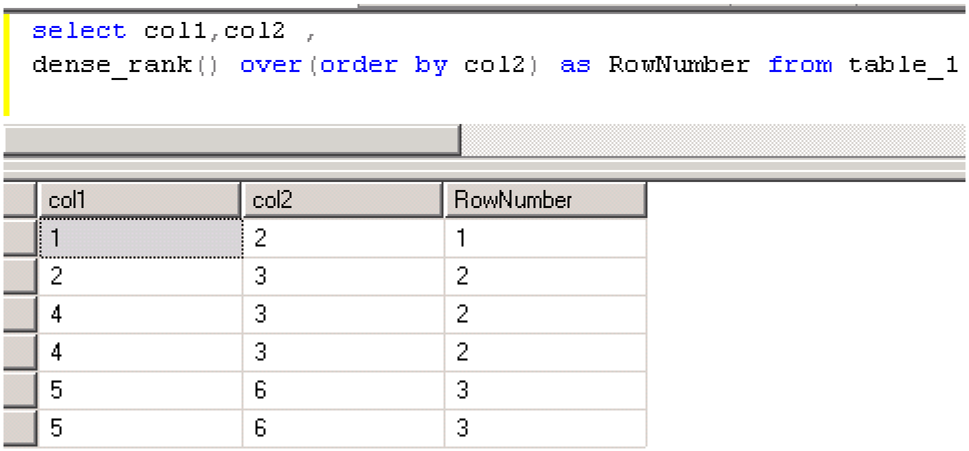
```
select col1,col2 ,
rank() over(order by col2) as RowNumber from table_1
```

col1	col2	RowNumber
1	2	1
2	3	2
4	3	2
4	3	2
5	6	5
5	6	5

Figure 2.20 : - RANK

What is DENSE_RANK()?

DENSE_RANK() works the same way as RANK() does but eliminates the gaps in the numbering. When I say GAPS you can see in previous results it has eliminated 4 and 5 from the count because of the gap in between COL2. But for dense_rank it overlooks the gap.



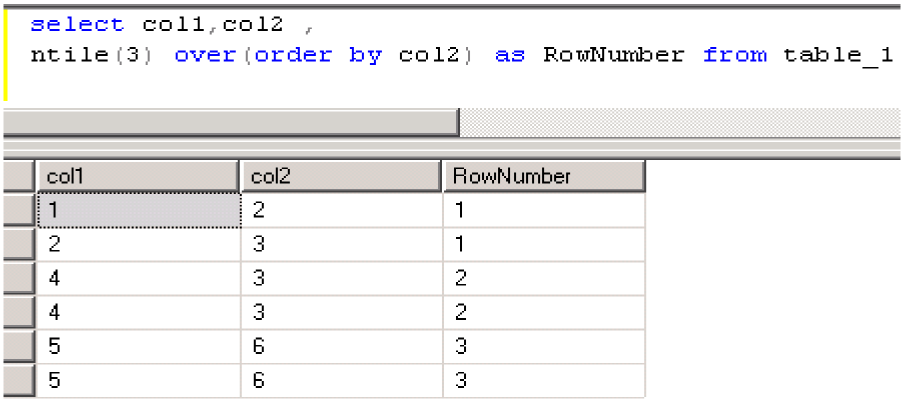
```
select col1,col2 ,
dense_rank() over(order by col2) as RowNumber from table_1
```

	col1	col2	RowNumber
	1	2	1
	2	3	2
	4	3	2
	4	3	2
	5	6	3
	5	6	3

Figure 2.21 :- DENSE_RANK() in action

What is NTILE()?

NTILE() breaks the result set into a specified number of groups and assigns the same number to each record in a group. Ok NTILE just groups depending on the number given or you can say divides the data. For instance I have said to NTILE it to 3. It has 6 total rows so it grouped in number of 2.



```
select col1,col2 ,
ntile(3) over(order by col2) as RowNumber from table_1
```

	col1	col2	RowNumber
	1	2	1
	2	3	1
	4	3	2
	4	3	2
	5	6	3
	5	6	3

Figure 2.22 : - NTILE in Action

(DB)What is SQl injection ?

It is a Form of attack on a database-driven Web site in which the attacker executes unauthorized SQL commands by taking advantage of insecure code on a system connected to the Internet, bypassing the firewall. SQL injection attacks are used to steal information from a database from which the data would normally not be available and/or to gain access to an organization's host computers through the computer that is hosting the database.

SQL injection attacks typically are easy to avoid by ensuring that a system has strong input validation.

As name suggest we inject SQL which can be relatively dangerous for the database. Example this is a simple SQL

```
SELECT email, passwd, login_id, full_name
  FROM members
 WHERE email = 'x'
```

Now somebody does not put "x" as the input but puts "x ; DROP TABLE members;". So the actual SQL which will execute is :-

```
SELECT email, passwd, login_id, full_name
  FROM members
 WHERE email = 'x'  ; DROP TABLE members;
```

Think what will happen to your database.

What's the difference between Stored Procedure (SP) and User Defined Function (UDF)?

Following are some major differences between a stored procedure and user defined functions:-

- √ UDF can be executed using the "SELECT" clause while SP's can not be.
- √ UDF can not be used in XML FOR clause but SP's can be used.
- √ UDF does not return output parameters while SP's return output parameters.

- √ If there is an error in UDF its stops executing. But in SP's it just ignores the error and moves to the next statement.
- √ UDF can not make permanent changes to server environments while SP's can change some of the server environment.

3. .NET Integration

What are steps to load a .NET code in SQL SERVER 2005?

Following are the steps to load a managed code in SQL SERVER 2005:-

- √ Write the managed code and compile it to a DLL / Assembly.
- √ After the DLL is compiled using the "CREATE ASSEMBLY" command you can load the assembly in to SQL SERVER. Below is the create command which is loading "mycode.dll" in to SQL SERVER using the "CREATE ASSEMBLY" command.

```
CREATE ASSEMBLY mycode FROM 'c:/mycode.dll'
```

How can we drop an assembly from SQL SERVER?

```
DROP ASSEMBLY mycode
```

Are changes made to assembly updated automatically in database?

No, it will not synchronize the code automatically. For that you have to drop the assembly (using the DROP ASSEMBLY) and create (using the CREATE ASSEMBLY) it again.

Why do we need to drop assembly for updating changes?

When we load the assembly in to SQL SERVER, it persist it in sys.assemblies. So any changes after that to the external DLL / ASSEMBLY will not reflect in SQL SERVER. So you have to DROP and then CREATE assembly again in SQL SERVER.

How to see assemblies loaded in SQL Server?

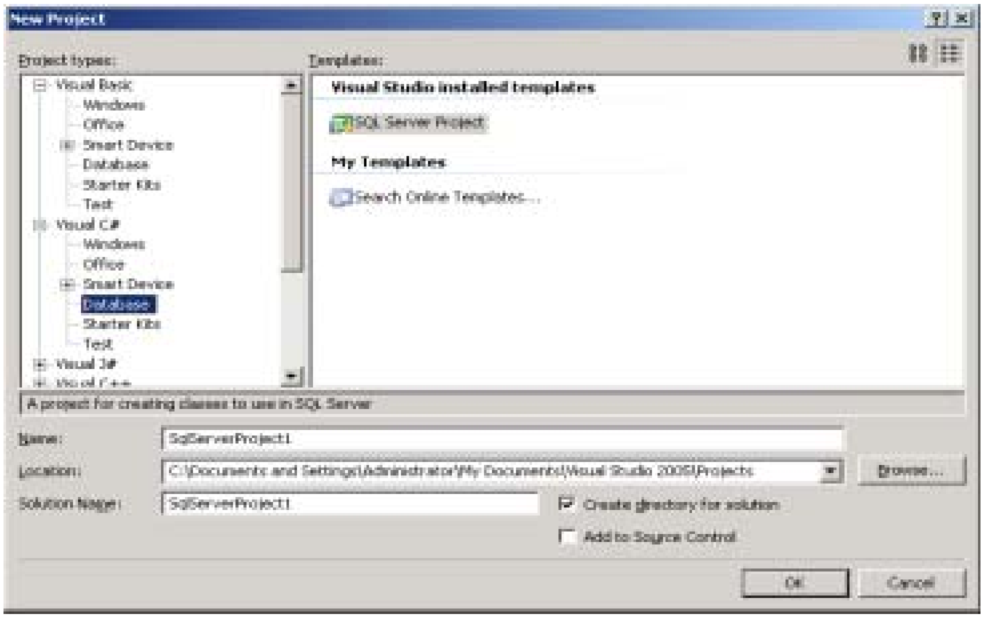
```
Select * from sys.assemblies.
```

I want to see which files are linked with which assemblies?

Assembly_files system tables have the track about which files are associated with what assemblies.

*SELECT * FROM sys.assembly_files*

Note :- You can create SQL SERVER projects using VS 2005 which provides ready made templates to make development life easy.



A project for creating classes to use in SQL Server

Name: SqlServerProject1

Location: C:\Documents and Settings\Administrator\My Documents\Visual Studio 2005\Projects

Solution Name: SqlServerProject1

☑ Create directory for solution

☐ Add to Source Control

Figure 3.1 : - Creating SQL SERVER Project using VS2005

Does .NET CLR and SQL SERVER run in different process?

.NET CLR engine (hence all the .NET applications) and SQL SERVER run in the same process or address space. This "Same address space architecture" is implemented so that there no speed issues. If the architecture was implemented the other way (i.e. SQL SERVER and .NET CLR engine running in different memory process areas) there would have been reasonable speed issue.

Does .NET controls SQL SERVER or is it vice-versa?

SQL Server controls the way .NET application will run. Normally .NET framework controls the way application should run. But in order that we have high stability and good security SQL Server will control the way .NET framework works in SQL Server environment. So lot of things will be controlled through SQL Server example threads, memory allocations, security etc .

SQL Server can control .NET framework by "Host Control" mechanism provided by .NET Framework 2.0. Using the "Host Control" framework external application's can control the way memory management is done, thread allocation's are done and lot more. SQL Server uses this "Host Control" mechanism exposed by .NET 2.0 and controls the framework.

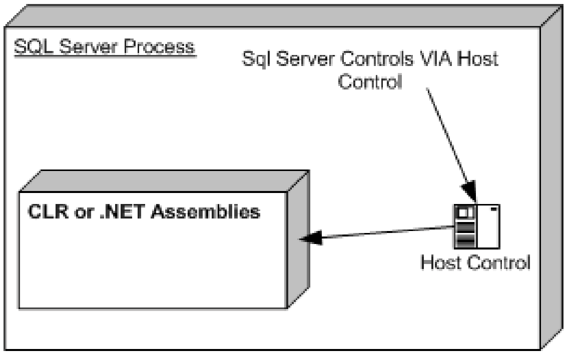


Figure 3.2 :- CLR Controlled by Host Control

Is SQLCLR configured by default?

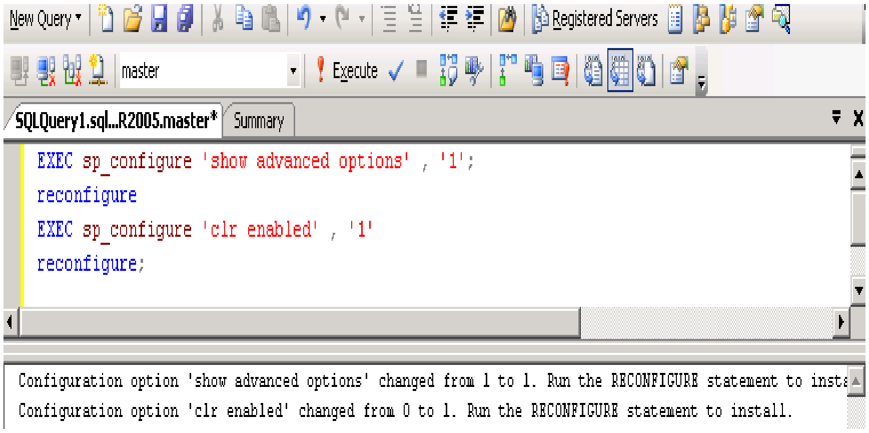
SQLCLR is not configured by default. If developers want to use the CLR integration feature of SQL SERVER it has to be enabled by DBA.

How to configure CLR for SQL SERVER?

It's a advanced option you will need to run the following code through query analyzer.

```
EXEC sp_configure 'show advanced options', '1';
go
reconfigure
```

```
go
EXEC sp_configure 'clr enabled' , '1'
go
reconfigure;
go
```



```
EXEC sp_configure 'show advanced options' , '1';
reconfigure
EXEC sp_configure 'clr enabled' , '1'
reconfigure;
```

```
Configuration option 'show advanced options' changed from 1 to 1. Run the RECONFIGURE statement to insta
Configuration option 'clr enabled' changed from 0 to 1. Run the RECONFIGURE statement to install.
```

Figure 3.3 :- sp_configure in action

Note :- You can see after running the SQL "clr enabled" property is changed from 0 to 1 , which indicates that the CLR was successfully configured for SQL SERVER.

Is .NET feature loaded by default in SQL Server?

No it will not be loaded, CLR is lazy loaded that means it's only loaded when needed. It goes one step ahead where the database administrator has to turn the feature on using "sp_configure".

Note: - Loading .NET runtime consumes some memory resources around 20 to 30 MB (it may vary depending on lot of situations). So if you really need .NET Integration then only go for this option.

How does SQL Server control .NET run-time?

.NET CLR exposes interfaces by which an external host can control the way .NET run time runs.

In previous versions of .NET it was done via COM interface "ICorRuntimeHost".

In pervious version you can only do the following with the COM interface.

- √ Specify that whether its server or workstation DLL.
- √ Specify version of the CLR (e.g. version 1.1 or 2.0)
- √ Specify garbage collection behavior.
- √ Specify whether or not jitted code may be shared across AppDomains.

In .NET 2.0 it's done by "ICLRRuntimeHost".

But in .NET 2.0 you can do much above what was provided by the previous COM interface.

- √ Exceptional conditions
- √ Code loading
- √ Class loading
- √ Security particulars
- √ Resource allocation

SQL Server uses the "ICLRRuntimeHost" to control .NET run-time as the flexibility provided by this interface is far beyond what is given by the previous .NET version, and that's what exactly SQL Server needs, a full control of the .NET run time.

What's a "SAND BOX" in SQL Server 2005?

> *Twist: - How many types of permission levels are there and explain in short there characteristics?*

Ok here's a general definition of sand box:-

> *"Sandbox is a safe place for running semi-trusted programs or scripts, often originating from a third party."*

Now for SQL Server it's .NET the external third party which is running and SQL Server has to ensure that .NET runtime crashes does not affect his working. So in order that SQL Server runs properly there are three sandboxes that user code can run :-

Safe Access sandbox

This will be the favorite setting of DBA's if they are every compelled to run CLR - Safe access. Safe means you have only access to in-proc data access functionalities. So you can create stored procedures, triggers, functions, data types, triggers etc. But you can not access memory, disk, create files etc. In short you can not hang the SQL Server.

External access sandbox

In external access you can use some real cool features of .NET like accessing file systems outside the box, you can leverage you classes etc. But here you are not allowed to play around with threading , memory allocation etc.

Unsafe access sand box

In Unsafe access you have access to memory management, threading etc. So here developers can write unreliable and unsafe code which destabilizes SQL Server. In the first two access levels of sand box its difficult to write unreliable and unsafe code.

What is an application domain?

Previously "PROCESS" where used as security boundaries. One process has its own virtual memory and does not over lap the other process virtual memory, due to this one process can not crash the other process. So any problem or error in one process does not affect the other process.In .NET they went one step ahead introducing application domains. In application domains multiple application can run in same process with out influencing each other.If one of the application domains throws error it does not affect the other application domains. To invoke method in a object running in different application domain .NET remoting is used.

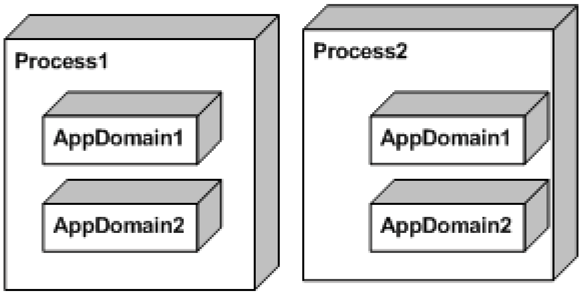


Figure 3.4 :- Application Domain architecture

How are .NET Appdomain allocated in SQL SERVER 2005?

In one line it's "One Appdomain per Owner Identity per Database".

That means if owner "A" owns "Assembly1" and "Assembly2" which belong to one database. They will be created in one Appdomain. But if they belong to different database two Appdomains will be created.

Again if there are different owners for every the same assembly then every owner will have its own Appdomain.

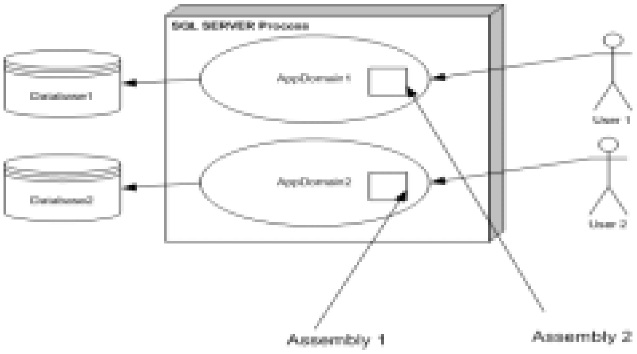


Figure 3 .5: - One Appdomain / Owner / Database

Note: - This can be pretty confusing during interviews so just make one note "One Appdomain per Owner Identity per Database".

What is Syntax for creating a new assembly in SQL Server 2005?

```
CREATE ASSEMBLY customer FROM 'c:\customers\customer.dll'
```

Do Assemblies loaded in database need actual .NET DLL?

No, once the assembly is loaded you do not need the source. SQL Server will load the DLL from the catalog.

You have a assembly which is dependent on other assemblies, will SQL Server load the dependent assemblies?

Ok. Let me make the question clearer. If you gave “Assembly1.dll” who is using “Assembly2.dll” and you try cataloging “Assembly1.dll” in SQL Server will it catalog “Assembly2.dll” also? Yes it will catalog it. SQL Server will look in to the manifest for the dependencies associated with the DLL and load them accordingly.

> *Note: - All Dependent assemblies have to be in the same directory, do not expect SQL Server to go to some other directory or GAC to see the dependencies.*

Does SQL Server handle unmanaged resources?

SQL Server does not handle the unmanaged resource of a framework. It has to be guaranteed by the framework DLL that it will clean up the unmanaged resource. SQL Server will not allow you to load .NET framework DLL which do not have clean up code for unmanaged resources.

What is Multi-tasking?

It’s a feature of modern operating systems with which we can run multiple programs at same time example Word,Excel etc.

What is Multi-threading?

Multi-threading forms subset of Multi-tasking instead of having to switch between programs this feature switches between different parts of the same program. Example you are writing in word and at the same time word is doing a spell check in background.

What is a Thread ?

A thread is the basic unit to which the operating system allocates processor time.

Can we have multiple threads in one App domain?

One or more threads run in an AppDomain. An AppDomain is a runtime representation of a logical process within a physical process. Each AppDomain is started with a single thread, but can create additional threads from any of its threads.

Note :- All threading classes are defined in System.Threading namespace.

What is Non-preemptive threading?

In Non-preemptive threading every thread gives control to other threads to execute. So for example we have "Thread1" and "Thread2", let's say for that instance "Thread1" is running. After some time it will give the control to "Thread2" for execution.

What is pre-emptive threading?

In pre-emptive threading operating system schedules which thread should run, rather threads making there own decisions.

Can you explain threading model in SQL Server?

SQL Server uses the "Non-preemptive" threading model while .NET uses "pre-emptive" threading model.

How does .NET and SQL Server thread work?

Note: - From hence onwards I will refer .NET assemblies running on SQL SERVER as SQLCLR that's more of industry acronym.

As said in the previous section threading model of .NET and SQL Server is completely different. So SQL Server has to handle threads in a different way for SQLCLR. So a little different threading architecture is implemented termed as "Tasking of Threads". In tasking thread architecture there is switching between SQLCLR threads and SQL Server threads, so that .NET threads do not consume full resource and go out of control. SQL Server introduced blocking points which allows this transition to happen between SQLCLR and SQL Server threads.

How is exception in SQLCLR code handled?

If you remember in the previous section's we had mentioned that there is One Appdomain / User / Database. So if there is an error in any of the Appdomain SQL Server will shut down the Appdomain , release all locks if the SQLCLR was holding and rollback the transaction in case if there are any.

So the Appdomain shut down policy ensures that all other Appdomain including SQL Server process is not affected.

Are all .NET libraries allowed in SQL Server?

No, it does not allow all .NET assemblies to execute example :- System.Windows.Forms, System.Drawing, System.Web etc are not allowed to run.SQL Server maintains a list of .NET namespaces which can be executed, any namespaces other than that will be restricted by SQL Server policy. This policy checks are made on two instances:-

- √ When you are cataloging the assembly.
- √ When you are executing the assembly.

Readers must be wondering why a two time check. There are many things in .NET which are building on runtime and can not be really made out from IL code. So SQL Server makes check at two point while the assembly is cataloged and while it's running which ensures 100 % that no runaway code is going to execute.

Note : - Read Hostprotectionattribute in next questions.

What is "Hostprotectionattribute" in SQL Server 2005?

As said previously .NET 2.0 provides capability to host itself and that's how SQL Server interacts with the framework. But there should be some mechanism by which the host who is hosting .NET assemblies should be alerted if there is any out of serious code running like threading, synchronization etc. This is what exactly the use of "Hostprotectionattribute". It acts like a signal to the outer host saying what type of code it has. When .NET Framework 2.0 was in development Microsoft tagged this attribute on many assemblies , so that SQL Server can be alerted to load those namespaces or not. Example if you look at System.Windows you will see this attribute.

So during runtime SQL Server uses reflection mechanism to check if the assembly has valid protection or not.

> *Note :- HostProtection is checked only when you are executing the assembly in SQL Server 2005.*

How many types of permission level are there for an assembly?

There are three types of permission levels for an assembly:-

Safe permission level

Safe assemblies can only use pre-defined framework classes, can call any COM based components or COM wrapper components, can not access network resources and, can not use PInvoke or platform invoke

External Access

It's like safe but you can access network resources like files from network, file system, DNS system, event viewer's etc.

Unsafe

In unsafe code you can run anything you want. You can use PInvoke; call some external resources like COM etc. Every DBA will like to avoid this and every developer should avoid writing unsafe code unless very much essential. When we create an assembly we can give the permission set at that time.

> *Note: - We had talked about sand boxes in the previous question. Just small note sandboxes are expressed by using the permission level concepts.*

In order that an assembly gets loaded in SQL Server what type of checks are done?

SQL Server uses the reflection API to determine if the assembly is safe to load in SQL Server.

Following are the checks done while the assembly is loaded in SQL Server:-

- √ It does the META data and IL verification, to see that syntaxes are appropriate of the IL.
- √ If the assembly is marked as safe and external then following checks are done
  - ■ Check for static variables, it will only allow read-only static variables.
  - ■ Some attributes are not allowed for SQL Server and those attributes are also checked.
  - ■ Assembly has to be type safe that means no unmanaged code or pointers are allowed.
  - ■ No finalizer's are allowed.

Note: - SQL Server checks the assembly using the reflection API, so the code should be IL compliant.

You can do this small exercise to check if SQL Server validates your code or not. Compile the simple below code which has static variable defined in it. Now because the static variable is not read-only it should throw an error.

```
using System; namespace StaticDll { public class Class1 { static int i; } }
```

After you have compiled the DLL, use the Create Assembly syntax to load the DLL in SQL Server. While cataloging the DLL you will get the following error:-

```
Msg 6211, Level 16, State 1, Line 1 CREATE ASSEMBLY failed because type 'StaticDll.Class1' in safe assembly 'StaticDll' has a static field 'i'. Attributes of static fields in safe assemblies must be marked readonly in Visual C#, ReadOnly in Visual Basic, or initonly in Visual C++ and intermediate language.
```

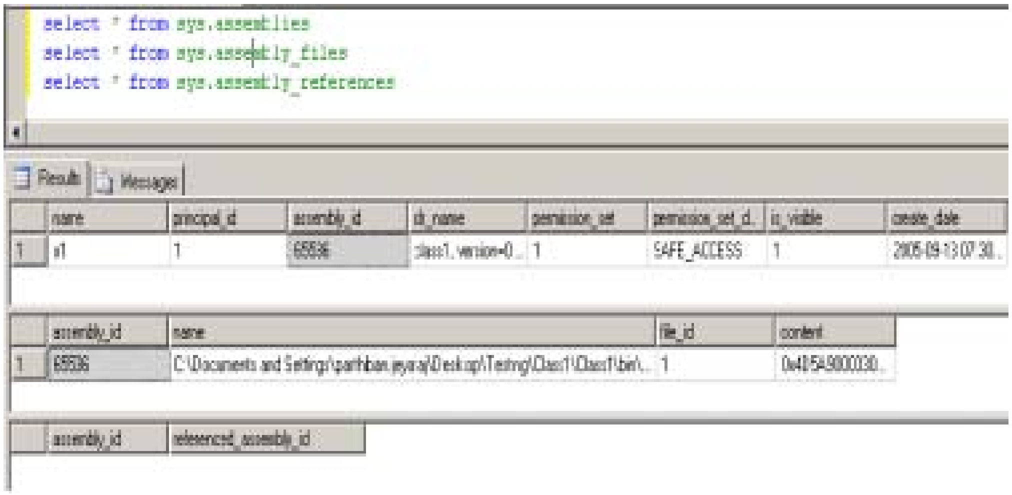
Can you name system tables for .NET assemblies?

There are three mail files which are important :-

- √ sys.assemblies :- They store information about the assembly.
- √ sys.assembly_files :- For every assembly you will can one or more files and this is where actual file raw data , file name and path is stored.
- √ sys.assembly_references :– All references to assemblies are stored in this table.

We can do a small practical hand on to see how the assembly table looks like. Let's try to create a simple class class1. Code is as shown below.

```
using System;
using System.Collections.Generic;
using System.Text;
namespace Class1
{
        public class Class1
        {

        }
}
```



```
select * from sys.assemblies
select * from sys.assembly_files
select * from sys.assembly_references
```

	name	principal_id	assembly_id	clr_name	permission_set	permission_set_d...	is_visible	create_date
1	x1	1	65536	class1, version=0...	1	SAFE_ACCESS	1	2005-09-13 07:30...

	assembly_id	name	file_id	content
1	65536	C:\Documents and Settings\parthiban.jeyaraj\Desktop\Testing\Class1\Class1\bin\...	1	0x4D5A9000030...

	assembly_id	referenced_assembly_id

Figure 3.6: - Assembly related system files.

Then we create the assembly by name "X1" using the create assembly syntax. In above image is the query output of all three main tables in this sequence sys.assemblies, sys.assembly_files and sys.assembly_references.

> *Note :- In the second select statement we have a content field in which the actual binary data stored. So even if we do not have the actual assembly it will load from this content field.*

Are two version of same assembly allowed in SQL Server?

You can give different assembly name in the create statement pointing to the different file name version.

How are changes made in assembly replicated?

ALTER ASSEMBLY CustomerAsm ADD FILE FROM 'c:\mydir\CustomerAsm.pdb'

Note:- You can drop and recreate it but will not be a good practice to do that way. Also note I have set the file reference to a "PDB" file which will enable my debugging just in case if I want it.

Is it a good practice to drop a assembly for changes?

Dropping an assembly will lead to loose following information:-

- √ You will loose all permissions defined to the assembly.
- √ All stored procedure, triggers and UDF (User defined functions) or any SQL Server object defined from them.

Note: - If you are doing bug fixes and modifications its good practice to Alter rather than Drop? Create assembly.

In one of the projects following steps where done, will it work?

Twist :- Are public signature changes allowed in "Alter assembly" syntax ?

Following are the steps:-

- √ Created the following class and method inside it.

```
public class clscustomer
{
Public void add()
{
}
}
```

Compiled the project success fully.

- √ Using the create assembly cataloged it in SQL Server.
- √ Later made the following changes to the class

```
public class clscustomer
{
Public void add(string code)
{
}
}
```

Note: - The add method signature is now changed.

- √ After that using "Alter" we tried to implement the change.

Using alter syntax you can not change public method signatures, in that case you will have to drop the assembly and re-create it again.

What does Alter assembly with unchecked data signify?

```
"ALTER ASSEMBLY CustomerAssembly FROM 'c:\cust.dll'   WITH
UNCHECKED DATA"
```

This communicates with SQL Server saying that you have not made any changes to serialization stuff, no data types are changed etc. Only you have changed is some piece of code for fixing bugs etc.

How do I drop an assembly?

```
DROP ASSEMBLY cust
```

Note: - If the assembly is referencing any other objects like triggers, stored procedures, UDT, other assemblies then the dependents should be dropped first, or else the drop assembly will fail.

Can we create SQLCLR using .NET framework 1.0?

No at this moment only .NET 2.0 version and above is supported with SQL Server.

While creating .NET UDF what checks should be done?

Following are some checks are essential for UDF (User Defined Function):-

√ The class in which the function is enclosed should be public.

√ .NET Function must be static.

Note: - When we want to catalog the assembly and function. Then first we catalog the class and the function. In short we use "Create Assembly" and then "Create Function".

How do you define a function from the .NET assembly?

Below is a sample "create function "statement and following are the legends defined in it:-

SortCustomer :- Name of the stored procedure function and can be different from the .NET function.In this cas the .NET function is by name sort which is defined in the external name.

CustomerAssembly :- The name of the assembly.

CustomerNameSpace :- The Namespace in which the class is lying.

Sort :- The .NET function name.

Other things are self explanatory.

Create Function SortCustomer(@Strcustcode int) returns int As EXTERNAL NAME CustomerAssembly.[CustomerNameSpace.CustomerClass].Sort

Note: - One important thing about the function parametersis allinput parameter will go by the order mapping. So what does that mean if my .NET function name func1 has the following definitions:-

func1 (int i, double x,string x)

Then my stored procedure should be defined accordingly and in the same order. That means in the stored procedure you should define it in the same order.

Can you compare between T-SQL and SQLCLR?

Note: - This will be one the favorite questions during interview. Interviewer will want to know if you know when is the right decision to take T-SQL or .NET for writing SQL Server objects. When I say SQL Server objects I am referring to stored procedure, functions or triggers.

√ Pure Data access code should always be written using T-SQL as that's what they where meant for. T-SQL does not have to load any runtime which make the access much faster. Also to note T-SQL will have access directly to internal buffers for SQL Server and they where written in probably assembly and "C" which makes them much faster for data access.

√ Pure Non-data access code like computation, string parsing logic etc should be written in .NET. If you want to access webservices or want to exploit OOP's programming for better reusability and read external files its good to go for .NET.

We can categorize our architect decision on there types of logic:-

√ Pure Data access functionality – Go for T-SQL.

√ Pure NON-Data access functionality – Go for .NET.

√ Mixture of data access and NON-Data access – Needs architecture decision.

If you can see the first two decisions are straight forward. But the third one is where you will have do a code review and see what will go the best. Probably also run it practically, benchmark and see what will be the best choice.

With respect to .NET is SQL SERVER case sensitive?

Following are some points to remember regarding case sensitiveness:-

√ Assembly names are not case sensitive.

√ Class and Function names are case sensitive.

So what does that mean? Well if you define a .NET DLL and catalog it in SQL Server. All the methods and class name are case sensitive and assembly is not case sensitive. For instance I have cataloged the following DLL which has the following details:-

- √ Assembly Name is “CustomerAssembly”.
- √ Class Name in the “CustomerAssembly” is “ClsCustomer”.
- √ Function “GetCustomerCount()” in class “ClsCustomer”.

When we catalog the above assembly in SQL Server. We can not address the “ClsCustomer” with “CLSCUSTOMER” or function “GetCustomerCount()” with “getcustomercount()” in SQL Server T-SQL language. But assembly “CustomerAssembly” can be addressed by “customerassembly” or “CUSTOMERASSEMBLY”, in short the assemblies are not case sensitive.

Does case sensitive rule apply for VB.NET?

The above case sensitive rules apply irrespective of whether the .NET language is case sensitive or not. So even if VB.NET is not case sensitive the rule will apply.

Can nested classes be accessed in T-SQL?

No, you can not access nested class in T-SQL its one of the limitation of SQLCLR.

Can we have SQLCLR procedure input as array?

SQL Server has no data types like arrays so it will not be able to map array datatype of .NET. This will throw an error.

Can object datatype be used in SQLCLR?

You can pass object datatype to SQLCLR but then that object should be defined as UDF in SQL Server.

How’s precision handled for decimal datatypes in .NET?

> *Note: - Precision is actually the number of digits after point which determines how accurate you want to see the result. Example in “9.29” we have precision of two decimal places (.29).*

In .NET we declare decimal datatypes with out precision. But in SQL Server you can define the precision part also.

decimal i; --> .NET Definition

decimal(9,2) --> SQL Server Definition

This creates a conflict when we want the .NET function to be used in T-SQL as SQLCLR and we want the precision facility.

Here's the answer you define the precision in SQL Server when you use Create syntax. So even if .NET does not support the precision facility we can define the precision in SQL Server.

```
.NET definition
func1(decimal x1)
{
}
SQL Server definition
create function func1(@x1 decimal(9,2))
returns decimal
as external name CustomerAssembly.[CustomerNameSpace.ClsCustomer].func1
```

If you see in the above code sample func1 is defined as simple decimal but later when we are creating the function definition in SQL Server we are defining the precision.

How do we define INPUT and OUTPUT parameters in SQLCLR?

.NET has following type of variable directions byvalue, byref and out (i.e. for C# only).Following is how the mapping goes:-

- √ Byval definition for function maps as input parameters for SQL Server.
- √ Byref definition maps to input and output parameters for SQL Server.

But for “out” types of parameters there are no mappings defined. Its logical “out” types of parameter types does not have any equivalents in SQL Server.

Note: - When we define byref in .NET that means if variable value is changed it will be reflected outside the subroutine, so it maps to SQL Server input/output (OUT) parameters.

Is it good to use .NET datatypes in SQLCLR?

No, it’s always recommended to use SQL Server datatypes for SQLCLR code as it implements better integration. Example for “int” datatype in .NET we can not assign NULL value it will crash, but using SQL datatype SqlInt32 NULLS will be handled. All SQL datatypes are available in “system.data.SQLtypes”, so you have to refer this namespace in order to get advantage of SQL datatypes.

Note: - NULL is a valid data in SQL Server which represents no data, but .NET datatype does not accept it.

How to move values from SQL to .NET datatypes?

You have to use the value property of SQL datatypes.

```
SqlInt32 x = 3;

int y = x.Value;
```

Note :- Direct assigning the values will crash your program.

What is System.Data.SqlServer?

When you have functions, stored procedures etc written in .NET you will use this provider rather than the traditional System.Data.SQLClient. If you are accessing objects created using T-SQL language then you will need a connection to connect them. Because you need to specify which server you will connect, what is the password and other credentials? But if you are accessing objects made using .NET itself you are already residing in SQL Server so you will not need a connection but rather a context.

What is SQLContext?

As said previously when use ADO.NET to execute a T-SQL created stored procedure we are out of the SQL Server boundary. So we need to provide SQLConnection object to connect to the SQLServer. But when we need to execute objects which are created using .NET language we only need the context in which the objects are running.

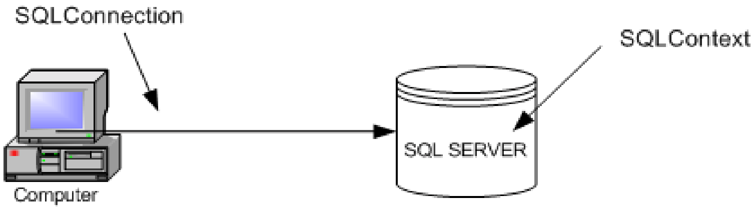


Figure 3.7 : - SQLConnection and SQLContext

So you can see in the above figure SQLConnection is used because you are completely outside SQL Server database. While SQLContext is used when you are inside SQL Server database. That means that there is already a connection existing so that you can access the SQLContext. And any connections created to access SQLContext are a waste as there is already a connection opened to SQL Server.

These all things are handled by SQLContext.

Which are the four static methods of SQLContext?

Below are the four static methods in SQLContext:-

GetConnection() :- This will return the current connection

GetCommand() :- Get reference to the current batch

GetTransaction() :- If you have used transactions this will get the current transaction

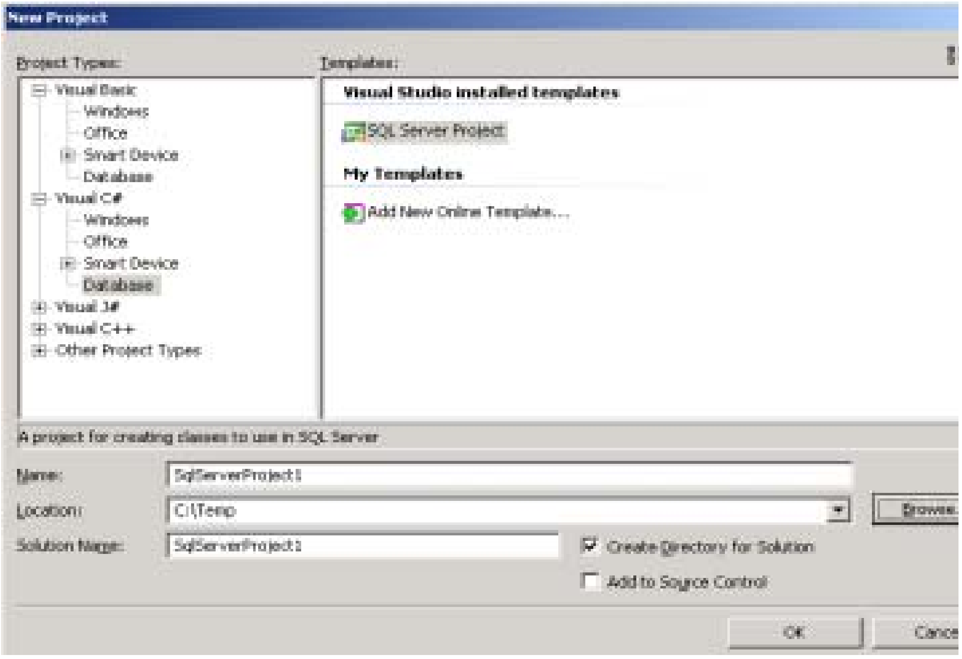
GetPipe() :- This helps us to send results to client. The output is in Tabular Data stream format. Using this method you can fill in datareader or data set, which can later be used by client to display data.

> *Note: - In top question I had shown how we can manually register the DLL's in SQL Server but in real projects no body would do that rather we will be using the VS.NET studio to accomplish the same. So we will run through a sample of how to deploy DLL's using VS.NET and paralelly we will also run through how to use SQLContext.*

Can you explain essential steps to deploy SQLCLR?

This example will make a simple walk through of how to create a stored procedure using visual studio.net editor. During interview you can make the steps short and explain to the interviewer. So we will create a stored procedure which will retrieve all products from adventureworks database. All products are stored in “Production.product” table.

Let start step1 go to visual studio --> new project --> expand the Visual C# (+)--> select database, you will see SQL Server project. Select SQL Server project template and give a name to it, then click ok.



A project for creating classes to use in SQL Server

Name: SqlServerProject1

Location: C:\Temp

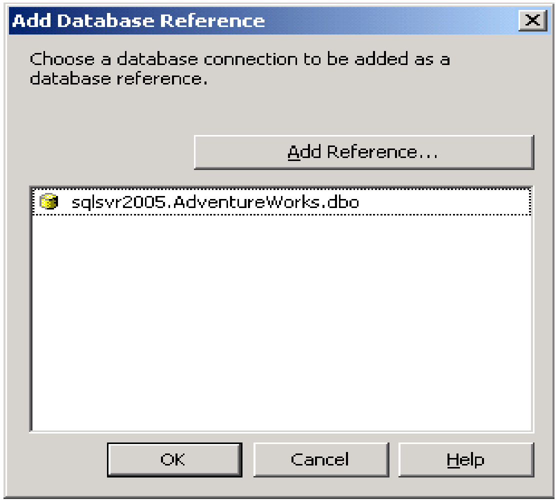
Solution Name: SqlServerProject1

☑ Create Directory for Solution

□ Add to Source Control

Figure 3.8 : - Template dialog box

As these DLL's need to be deployed on the server you will need to specify the server details also. So for the same you will be prompted to specify database on which you will deploy the .NET stored procedure. Select the database and click ok. In case you do not see the database you can click on "Add reference" to add the database to the list.



Choose a database connection to be added as a database reference.

sqlsvr2005.AdventureWorks.dbo

Figure 3.9 : - Select database

Once you specify the database you are inside the visual studio.net editor. At the right hand side you can see the solution explorer with some basic files created by visual studio in order to deploy the DLL on the SQL Server. Right click on SQL Server project and click on ADD --> New items are displayed as shown in figure below.

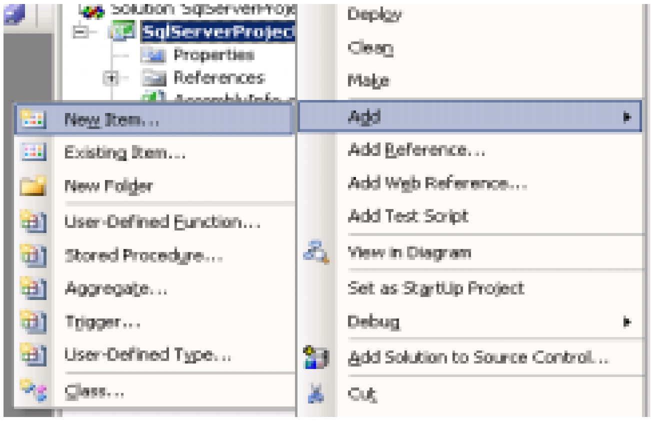
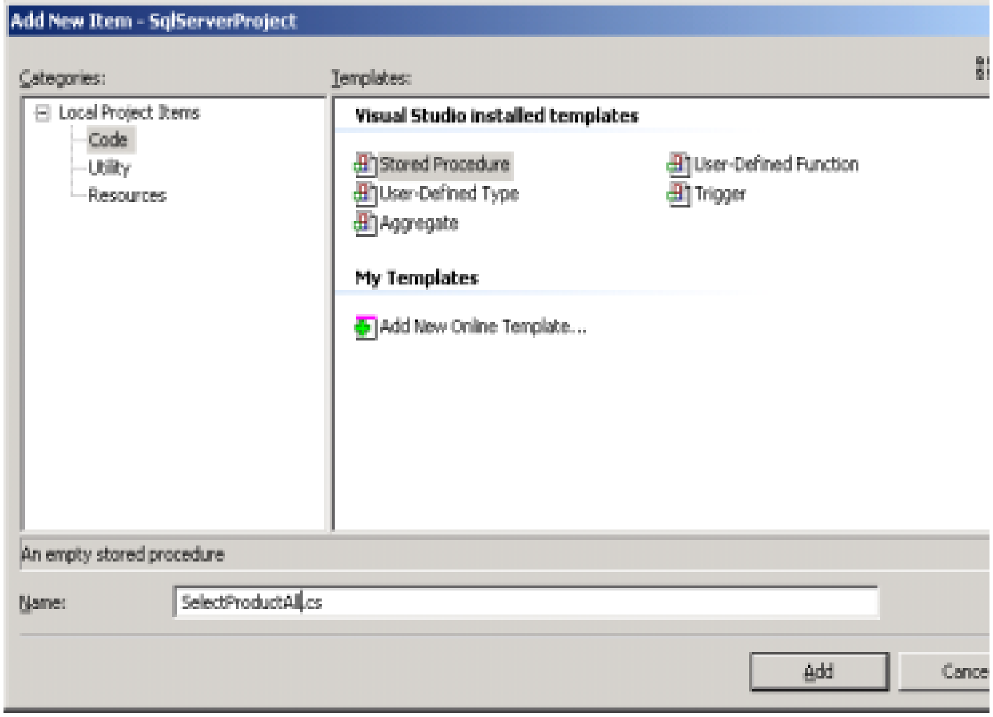


Figure 3.10: - Click add item

You can see in the below figure you can create different objects using VS.NET. For this point of time we need to only create a stored procedure which will fetch data from "Product.Product".



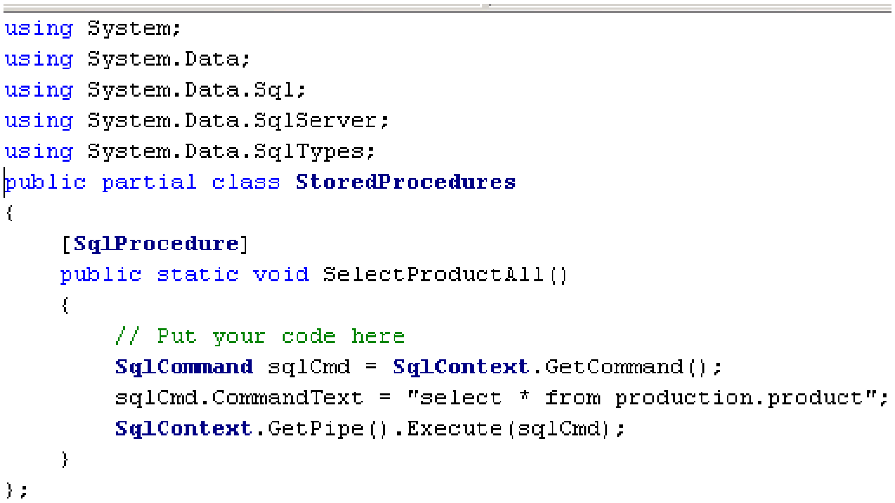
An empty stored procedure

Name: SelectProductAll.cs

Figure 3.11 : - Select stored procedure template

This section is where the real action will happen. As said previously you do not need to open a connection but use the context. So below are the three steps:-

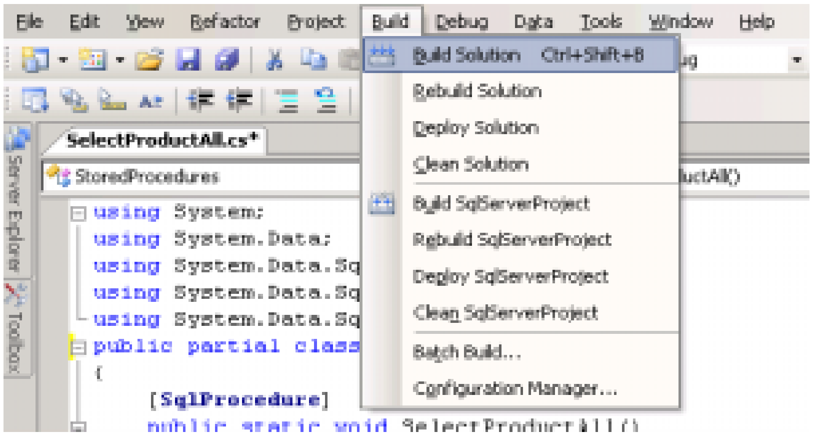
- Get the reference of the context.
- Get the command from the context.
- Set the command text, at this moment we need to select everything from "Production.Product" table.
- Finally get the Pipe and execute the command.



```
using System;
using System.Data;
using System.Data.Sql;
using System.Data.SqlServer;
using System.Data.SqlTypes;
public partial class StoredProcedures
{
    [SqlProcedure]
    public static void SelectProductAll()
    {
        // Put your code here
        SqlCommand sqlCmd = SqlContext.GetCommand();
        sqlCmd.CommandText = "select * from production.product";
        SqlContext.GetPipe().Execute(sqlCmd);
    }
};
```

Figure 3.12 : - Simple code to retrieve product table

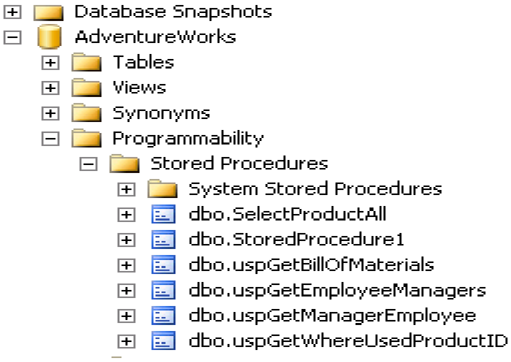
After that you need to compile it to a DLL form and then deploy the code in SQL Server. You can compile using “Build Solution” menu to compile and “Deploy Solution” to deploy it on SQL Server.



```
using System;
using System.Data;
using System.Data.Sq
using System.Data.Sq
using System.Data.Sq
public partial class
{
    [SqlProcedure]
    public static void SelectProductAll()
```

Figure 3.13 : - Finally build and deploy solution

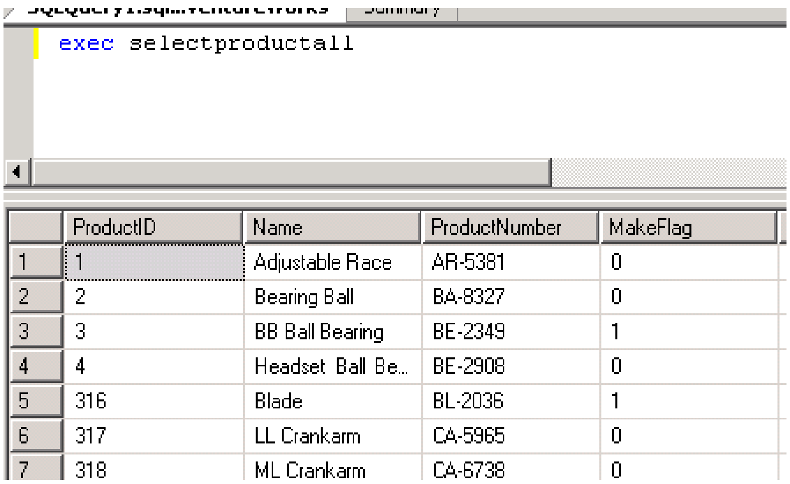
After deploying the solution you can see the stored procedure “SelectProductAll” in the stored procedure section as shown below.



```
⊞ Database Snapshots
⊟ AdventureWorks
    ⊞ Tables
    ⊞ Views
    ⊞ Synonyms
    ⊟ Programmability
        ⊟ Stored Procedures
            ⊞ System Stored Procedures
            ⊞ dbo.SelectProductAll
            ⊞ dbo.StoredProcedure1
            ⊞ dbo.uspGetBillOfMaterials
            ⊞ dbo.uspGetEmployeeManagers
            ⊞ dbo.uspGetManagerEmployee
            ⊞ dbo.uspGetWhereUsedProductID
```

Figure 3.14 : - SelectProductall listed in database

Just to test I have executed the stored procedure and everything working fine.



```
exec selectproductall
```

	ProductID	Name	ProductNumber	MakeFlag
1	1	Adjustable Race	AR-5381	0
2	2	Bearing Ball	BA-8327	0
3	3	BB Ball Bearing	BE-2349	1
4	4	Headset Ball Be...	BE-2908	0
5	316	Blade	BL-2036	1
6	317	LL Crankarm	CA-5965	0
7	318	ML Crankarm	CA-6738	0

Figure 3.15 : - Execute stored procedure selectproductall

How do create function in SQL Server using .NET?

In order to create the function you have to select in Visual studio installed templates, User defined function template. Below is the sample code. Then follow again the same procedure of compiling and deploying the solution.

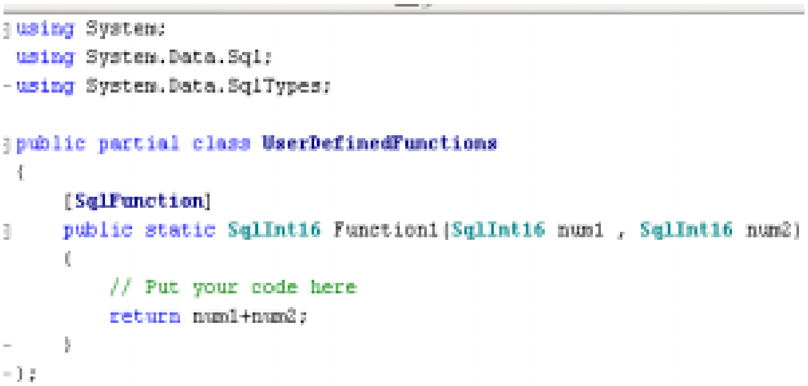
How do we create trigger using .NET?

For trigger you have to select trigger template. But you can see some difference in code here. You have to specify on which object and which event will this method fire. The first attribute specifies the name , target (on which the trigger will fire) and event (insert , update or delete).

```
[SqlTrigger (Name="Trigger1", Target="Table1", Event="FOR INSERT")]
      public static void Trigger1()
      {
        // Put your code here
        SqlTriggerContext objtrigcontext = SqlContext.GetTriggerContext();
        SqlPipe objsqlpipe = SqlContext.GetPipe();
        SqlCommand objcommand = SqlContext.GetCommand();
        if (objtrigcontext.TriggerAction == TriggerAction.Insert)
        {
                objcommand.CommandText = "insert into table1 values('Inserted')";
                objsqlpipe.Execute(objcommand);
        }
      }
```

How to create User Define Functions using .NET?

The below code is self explanatory. Compiling and deploying remains same for all object created using .NET.



```
using System;
using System.Data.Sql;
using System.Data.SqlTypes;

public partial class UserDefinedFunctions
{
    [SqlFunction]
    public static SqlInt16 Function1(SqlInt16 num1 , SqlInt16 num2)
    {
        // Put your code here
        return num1+num2;
    }
};
```

Figure 3.16 :- Function source code

Note :- Some home work for readers down.

How to create aggregates using .NET?

What is Asynchronous support in ADO.NET?

One of features which was missing in ADO.NET was asynchronous processing. That means once your SQL command is executed your UI has to wait until it's finished. ADO.NET provides support where you do not have to wait until the SQL is executed in database. You can see from the code below you have to issue "BeginExecuteReader" and then proceed ahead with some other process. After you finish the process you can come back and see your results. Long running queries can really be benefited from Asynchronous support.

```
conn.Open();
   IAsyncResult myResult = mycommand.BeginExecuteReader();
   while (!myResult.IsCompleted)
   {
      // execute some other process
   }
   // Finally process the data reader output
```

```
SqlDataReader rdr = mycommand.EndExecuteReader(myResult);
```

Note: - Here's a small project which you can do with Asynchronous processing. Fire a heavy duty SQL and in UI show how much time the SQL Server took to execute that query.

What is MARS support in ADO.NET?

In previous versions of ADO.NET you should one connection on every result set. But this new feature allows you to execute multiple commands on the same connection. You can also switch back and forth between the command objects in a connection. There is nothing special to do for MARS (Multiple Active Result Sets) just you can allocate multiple command objects on a single connection.

What is SQLbulkcopy object in ADO.NET?

With SQlbulkcopy you can insert bulk records in to table. The command is pretty simple you can see we have provided the datareader object to the SQlbulkcopy object and he will take care of the rest.

```
SqlBulkCopy objbulkData = new SqlBulkCopy(conn);

objbulkData.DestinationTableName = "table1";

objbulkData.WriteToServer(datareader1);
```

How to select range of rows using ADO.NET?

Twist: - What is paging in ADO.NET?

By paging you can select a range of rows from a result set. You have to specify starting row in the result set and then how many rows you want after that.

```
command.ExecutePageReader(CommandBehavior.Default, 1, 10);
```

You can see in the above example I have selected 10 rows and starts from one. This functionality will be used mainly when you want to do paging on UI side. For instance you want to show 10 records at a time to the user this can really ease of lot of pain.

What are different types of triggers in SQl SERVER 2000 ?

There are two types of triggers :-

INSTEAD OF triggers

INSTEAD OF triggers fire in place of the triggering action. For example, if an INSTEAD OF UPDATE trigger exists on the Sales table and an UPDATE statement is executed against the Salestable, the UPDATE statement will not change a row in the sales table. Instead, the UPDATE statement causes the INSTEAD OF UPDATE trigger to be executed, which may or may not modify data in the Sales table.

AFTER triggers

AFTER triggers execute following the SQL action, such as an insert, update, or delete.This is the traditional trigger which existed in SQL SERVER.

INSTEAD OF triggers gets executed automatically before the Primary Key and the Foreign Key constraints are checked, whereas the traditional AFTER triggers gets executed after these constraints are checked.

Unlike AFTER triggers, INSTEAD OF triggers can be created on views.

If we have multiple AFTER Triggers on table how can we define the sequence of the triggers ?

If a table has multiple AFTER triggers, then you can specify which trigger should be executed first and which trigger should be executed last using the stored procedure sp_settriggerorder. All the other triggers are in an undefined order which you cannot control.

How can you raise custom errors from stored procedure ?

The RAISERROR statement is used to produce an ad hoc error message or to retrieve a custom message that is stored in the sysmessages table. You can use this statement with the error handling code presented in the previous section to implement custom error messages in your applications. The syntax of the statement is shown here.

```
RAISERROR ({msg_id |msg_str }{,severity ,state }
  [ ,argument [ ,,...n ] ] ))
```

```
          [ WITH option [ ,,...n ] ]
```

A description of the components of the statement follows.

msg_id :-The ID for an error message, which is stored in the error column in sysmessages.

msg_str :-A custom message that is not contained in sysmessages.

severity :- The severity level associated with the error. The valid values are 0–25. Severity levels 0–18 can be used by any user, but 19–25 are only available to members of the fixed-server role sysadmin. When levels 19–25 are used, the WITH LOG option is required.

state A value that indicates the invocation state of the error. The valid values are 0–127. This value is not used by SQL Server.

Argument, . . .

One or more variables that are used to customize the message. For example, you could pass the current process ID (@@SPID) so it could be displayed in the message.

WITH option, . . .

The three values that can be used with this optional argument are described here.

LOG - Forces the error to logged in the SQL Server error log and the NT application log.

NOWAIT - Sends the message immediately to the client.

SETERROR - Sets @@ERROR to the unique ID for the message or 50,000.

The number of options available for the statement make it seem complicated, but it is actually easy to use. The following shows how to create an ad hoc message with a severity of 10 and a state of 1.

```
RAISERROR ('An error occured updating the NonFatal table',10,1)
```

```
--Results--

An error occured updating the NonFatal table
```

The statement does not have to be used in conjunction with any other code, but for our purposes it will be used with the error handling code presented earlier. The following alters the ps_NonFatal_INSERT procedure to use RAISERROR.

```
USE tempdb

go
```

```
ALTER PROCEDURE ps_NonFatal_INSERT
@Column2 int =NULL
AS
DECLARE @ErrorMsgID int
INSERT NonFatal VALUES (@Column2)
SET @ErrorMsgID =@@ERROR
IF @ErrorMsgID <>0
 BEGIN
  RAISERROR ('An error occured updating the NonFatal table',10,1)
 END
```

When an error-producing call is made to the procedure, the custom message is passed to the client. The following shows the output generated by Query Analyzer.

4. ADO.NET

Which are namespaces for ADO.NET?

Following are the namespaces provided by .NET for data management:-

System.data

This contains the basic objects used for accessing and storing relational data, such as DataSet, DataTable and DataRelation. Each of these is independent of the type of data source and the way we connect to it.

System.Data.OleDB

This contains the objects that we use to connect to a data source via an OLE-DB provider, such as OleDbConnection, OleDbCommand, etc. These objects inherit from the common base classes and so have the same properties, methods, and events as the SqlClient equivalents.

System.Data.SqlClient:

This contains the objects that we use to connect to a data source via the Tabular Data Stream (TDS) interface of Microsoft SQL Server (only). This can generally provide better performance as it removes some of the intermediate layers required by an OLE-DB connection.

System.XML

This Contains the basic objects required to create, read, store, write, and manipulate XML documents according to W3C recommendations.

Can you give a overview of ADO.NET architecture?

The most important section in ADO.NET architecture is "Data Provider".Data Provider provides access to datasource (SQL SERVER, ACCESS , ORACLE).In short it provides object to achieve functionalities like opening and closing connection , retrieve data and update data.In the below figure you can see the four main sections of a data provider :-

- √ Connection.
- √ Command object (This is the responsible object to use stored procedures)

- √ Data Adapter (This object acts as a bridge between datastore and dataset).
- √ Datareader (This object reads data from data store in forward only mode).

Dataset object represents disconnected and cached data.If you see the diagram it is not in direct connection with the data store (SQL SERVER , ORACLE etc) rather it talks with Data adapter , who is responsible for filling the dataset.Dataset can have one or more Datatable and relations.

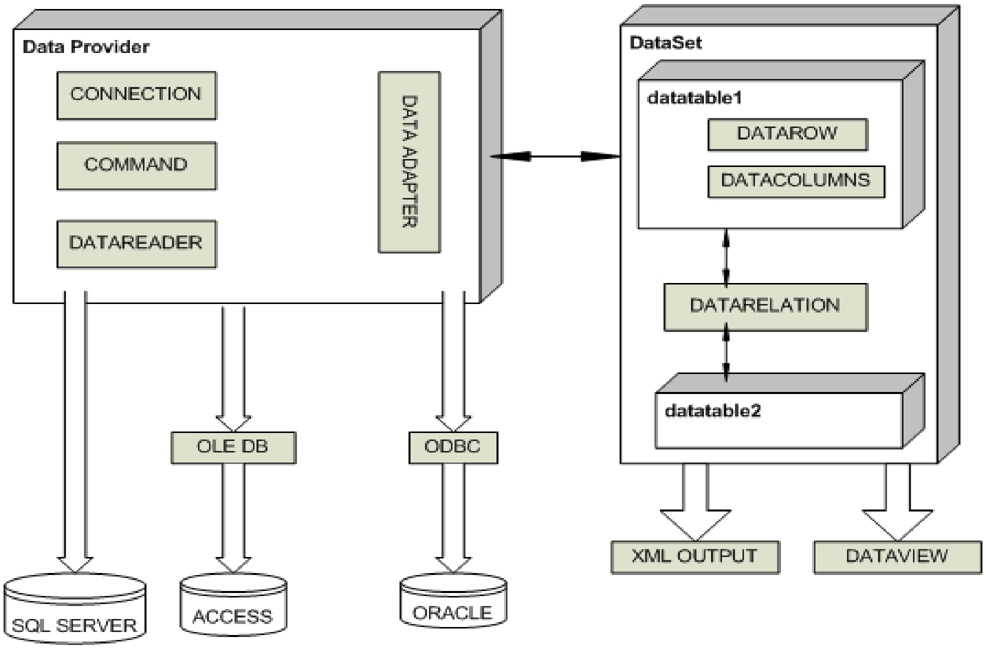


Figure :- 4.1 ADO.NET Architecture

"DataView" object is used to sort and filter data in Datatable.

> *Note:- This is one of the favorite questions in .NET.Just paste the picture in your mind and during interview try to refer that image.*

What are the two fundamental objects in ADO.NET?

Datareader and Dataset are the two fundamental objects in ADO.NET.

What is difference between dataset and datareader?

Following are some major differences between dataset and datareader :-

- √ DataReader provides forward-only and read-only access to data , while the DataSet object can hold more than one table (in other words more than one rowset) from the same data source as well as the relationships between them.
- √ Dataset is a disconnected architecture while datareader is connected architecture.
- √ Dataset can persist contents while datareader can not persist contents, they are forward only.

What are major difference between classic ADO and ADO.NET?

Following are some major differences between both

- √ As in classic ADO we had client and server side cursors they are no more present in ADO.NET.Note it's a disconnected model so they are no more applicable.
- √ Locking is not supported due to disconnected model.
- √ All data is persisted in XML as compared to classic ADO where data was persisted in Binary format also.

What is the use of connection object?

They are used to connect a data to a Command object.

- √ An OleDbConnection object is used with an OLE-DB provider
- √ A SqlConnection object uses Tabular Data Services (TDS) with MS SQL Server

What are the methods provided by the command object?

They are used to connect connection object to Datareader or dataset.Following are the methods provided by command object :-

- √ ExecuteNonQuery :- Executes the command defined in the CommandText property against the connection defined in the Connection property for a query that does not return any rows (an UPDATE, DELETE or INSERT). Returning an Integer indicating the number of rows affected by the query.
- √ ExecuteReader :- Executes the command defined in the CommandText property against the connection defined in the Connection property. Returns a "reader" object that is connected to the resulting rowset within the database, allowing the rows to be retrieved.
- √ ExecuteScalar :- Executes the command defined in the CommandText property against the connection defined in the Connection property. Returns only a single value (effectively the first column of the first row of the resulting rowset). Any other returned columns and rows are discarded. Fast and efficient when only a "singleton" value is required

What is the use of “Dataadapter”?

These are objects that connect one or more Command objects to a Dataset object..They provide logic that gets the data from the data store and populates the tables in the DataSet, or pushes the changes in the DataSet back into the data store.

- √ An OleDbDataAdapter object is used with an OLE-DB provider
- √ A SqlDataAdapter object uses Tabular Data Services with MS SQL Server.

What are basic methods of “Dataadapter”?

There are three most commonly used methods of Dataadapter :-

Fill :- Executes the SelectCommand to fill the DataSet object with data from the data source. Can also be used to update (refresh) an existing table in a DataSet with changes made to the data in the original datasource if there is a primary key in the table in the DataSet.

FillSchema :- Uses the SelectCommand to extract just the schema for a table from the data source, and creates an empty table in the DataSet object with all the corresponding constraints.

Update:- Calls the respective InsertCommand, UpdateCommand, or DeleteCommand for each inserted, updated,or deleted row in the DataSet so as to update the original data source with the changes made to the content of the DataSet. This is a little like the

UpdateBatch method provided by the ADO Recordset object, but in the DataSet it can be used to update more than one table.

What is Dataset object?

The DataSet provides the basis for disconnected storage and manipulation of relational data. We fill it from a data store,work with it while disconnected from that data store, then reconnect and flush changes back to the data store if required.

What are the various objects in Dataset?

Dataset has a collection of DataTable object within the Tables collection. Each DataTable object contains a collection of DataRow objects and a collection of DataColumn objects. There are also collections for the primary keys,constraints, and default values used in this table which is called as constraint collection, and the parent and child relationships between the tables.Finally, there is a DefaultView object for each table. This is used to create a DataView object based on the table, so that the data can be searched, filtered or otherwise manipulated while displaying the data.

> *Note :- Look back again to the main diagram for ADO.NET architecture for visualizing this answer in pictorial form*

How can we connect to Microsoft Access, FoxPro, Oracle etc?

Microsoft provides System.Data.OleDb namespace to communicate with databases like Access, oracle etc.In short any OLE DB-Compliant database can be connected using System.Data.OldDb namespace.

```
 Private Sub loadData()
        Dim strPath As String
        strPath = AppDomain.CurrentDomain.BaseDirectory
        Dim objOLEDBCon As New
OleDbConnection(“Provider=Microsoft.Jet.OLEDB.4.0;Data Source =” &
strPath & “Nwind.mdb”)
        Dim objOLEDBCommand As OleDbCommand
        Dim objOLEDBReader As OleDbDataReader
        Try
```

```
            objOLEDBCommand = New OleDbCommand(“Select FirstName
            from Employees”)
            objOLEDBCon.Open()
            objOLEDBCommand.Connection = objOLEDBCon
            objOLEDBReader = objOLEDBCommand.ExecuteReader()
            Do While objOLEDBReader.Read()
                lstNorthwinds.Items.Add(objOLEDBReader.GetString(0))
            Loop
        Catch ex As Exception
            Throw ex
        Finally
            objOLEDBCon.Close()
        End Try

    End Sub
```

What’s the namespace to connect to SQL Server?

Below is a sample code which shows a simple connection with SQL Server.

```
Private Sub LoadData()
        ‘ note :- with and end with makes your code more readable
        Dim strConnectionString As String
        Dim objConnection As New SqlConnection
        Dim objCommand As New SqlCommand
        Dim objReader As SqlDataReader
        Try
            ‘ this gets the connectionstring from the app.config
file.
            ‘ note if this gives error see where the MDB file is
stored in your pc and point to that
            strConnectionString =
AppSettings.Item(“ConnectionString”)
            ‘ take the connectiostring and initialize the connection
object
            With objConnection
                .ConnectionString = strConnectionString
                .Open()
            End With
            objCommand = New SqlCommand(“Select FirstName from
Employees”)
            With objCommand
                .Connection = objConnection
                objReader = .ExecuteReader()
            End With
            ‘ looping through the reader to fill the list box
            Do While objReader.Read()
                lstData.Items.Add(objReader.Item(“FirstName”))
            Loop
        Catch ex As Exception
            Throw ex
        Finally
            objConnection.Close()
        End Try
```

Now from interview point of view definitely you are not going to say the whole source code which is given in book. Interviewer expects only the broader answer of what are the steps needed to connect to SQL SERVER. For fundamental sake author has explained the whole source code. In short you have to explain the “LoadData” method in broader way. Following are the steps to connect to SQL SERVER :-

- √ First is import the namespace “System.Data.SqlClient”.
- √ Create a connection object as shown in “LoadData” method.

```
       With objConnection
             .ConnectionString = strConnectionString
             .Open()
       End Withs
```

- √ Create the command object with the SQL.Also assign the created connection object to command object. and execute the reader.

```
objCommand = New SqlCommand(“Select FirstName from Employees”)
            With objCommand
                .Connection = objConnection
                objReader = .ExecuteReader()
            End With
```

- √ Finally loop through the reader and fill the list box.If old VB programmers are expecting the movenext command it’s replaced by Read() which returns true if there is any data to be read.If the .Read() return’s false that means that it’s end of datareader and there is no more data to be read.

```
Do While objReader.Read()
lstData.Items.Add(objReader.Item(“FirstName”))
Loop
```

- √ Finally do not forget to close the connection object.

How do we use stored procedure in ADO.NET?

ADO.NET provides the SqlCommand object which provides the functionality of executing stored procedures.

```
If txtEmployeeName.Text.Length = 0 Then
                objCommand = New SqlCommand(“SelectEmployee”)
            Else
                objCommand = New SqlCommand(“SelectByEmployee”)
                objCommand.Parameters.Add(“@FirstName”,
Data.SqlDbType.NVarChar, 200)
                objCommand.Parameters.Item(“@FirstName”).Value =
txtEmployeeName.Text.Trim()
            End If
```

In the above sample not lot has been changed only that the SQL is moved to the stored procedures. There are two stored procedures one is “SelectEmployee” which selects all the employees and the other is “SelectByEmployee” which returns employee name starting with a specific character. As you can see to provide parameters to the stored procedures we are using the parameter object of the command object. In such question interviewer expects two simple answers one is that we use command object to execute stored procedures and the parameter object to provide parameter to the stored procedure. Above sample is provided only for getting the actual feel of it. Be short, be nice and get a job.

How can we force the connection object to close?

Command method Executereader takes a parameter called as CommandBehavior where in we can specify saying close connection automatically after the Datareader is close.

pobjDataReader = pobjCommand.ExecuteReader(CommandBehavior.CloseConnection)

I want to force the datareader to return only schema?

pobjDataReader = pobjCommand.ExecuteReader(CommandBehavior.SchemaOnly)

Can we optimize command object when there is only one row?

Again CommandBehaviour enumeration provides two values SingleResult and SingleRow. If you are expecting a single value then pass “CommandBehaviour.SingleResult” and the query is optimized accordingly, if you are expecting single row then pass “CommandBehaviour.SingleRow” and query is optimized according to single row.

Which is the best place to store connectionstring?

Config files are the best place to store connection strings. If it's a web-based application "Web.config" file will be used and if it's a windows application "App.config" files will be used.

What are steps involved to fill a dataset?

Twist :- How can we use dataadapter to fill a dataset?

Below is a simple code which loads a dataset and then finally loads the listbox.

```
 Private Sub LoadData()
        Dim strConnectionString As String
        strConnectionString = AppSettings.Item("ConnectionString")
        Dim objConn As New SqlConnection(strConnectionString)
        objConn.Open()
        Dim objCommand As New SqlCommand("Select FirstName from
Employees")
        objCommand.Connection = objConn
        Dim objDataAdapter As New SqlDataAdapter()
        objDataAdapter.SelectCommand = objCommand
        Dim objDataSet As New DataSet

    End Sub
```

In such type of question's interviewer is looking from practical angle, that have you worked with dataset and datadapters. Let me try to explain the above code first and then we move to what steps to say suring interview.

Dim objConn As New SqlConnection(strConnectionString)

objConn.Open()

First step is to open the connection.Again note the connection string is loaded from config file.

Dim objCommand As New SqlCommand("Select FirstName from Employees")

objCommand.Connection = objConn

Second step is to create a command object with appropriate SQL and set the connection object to this command.

Dim objDataAdapter As New SqlDataAdapter()

objDataAdapter.SelectCommand = objCommand

Third step is to create the Adapter object and pass the command object to the adapter object.

```
objDataAdapter.Fill(objDataSet)
```

Fourth step is to load the dataset using the “Fill” method of the dataadapter.

```
lstData.DataSource = objDataSet.Tables(0).DefaultView
lstData.DisplayMember = “FirstName”
lstData.ValueMember = “FirstName”
```

Fifth step is to bind to the loaded dataset with the GUI.At this moment sample has listbox as the UI. Binding of the UI is done by using DefaultView of the dataset.Just to revise every dataset has tables and every table has views. In this sample we have only loaded one table i.e. Employees table so we are referring that with a index of zero.

Just say all the five steps during interview and you will see the smile in the interviewer’s face.....Hmm and appointment letter in your hand.

What are the methods provided by the dataset for XML?

> *Note:- XML is one of the most important leap between classic ADO and ADO.NET. So this question is normally asked more generally how can we convert any data to XML format. Best answer is convert in to dataset and use the below methods.*

√ ReadXML

Read’s a XML document in to Dataset.

√ GetXML

This is function’s which return’s a string containing XML document.

√ WriteXML

This writes a XML data to disk.

How can we save all data from dataset?

Dataset has “AcceptChanges” method which commits all the changes since last time “Acceptchanges” has been executed.

How can we check for changes made to dataset?

Twist :- How can we cancel all changes done in dataset?, How do we get values which are changed in a dataset?

For tracking down changes Dataset has two methods which comes as rescue "GetChanges" and "HasChanges".

GetChanges

Return's dataset which are changed since it was loaded or since Acceptchanges was executed.

HasChanges

This property indicates has any changes been made since the dataset was loaded or "acceptchanges" method was executed.

If we want to revert or abandon all changes since the dataset was loaded use "RejectChanges".

Note:- One of the most misunderstood things about these properties is that it tracks the changes of actual database. That's a fundamental mistake; actually the changes are related to only changes with dataset and has nothing to with changes happening in actual database. As dataset are disconnected and do not know anything about the changes happening in actual database.

How can we add/remove row's in "DataTable" object of "DataSet"?

"Datatable" provides "NewRow" method to add new row to "DataTable"."DataTable" has "DataRowCollection" object which has all rows in a "DataTable" object. Following are the methods provided by "DataRowCollection" object :-

Add

Add's a new row in DataTable

Remove

Remove's a "DataRow" object from "DataTable"

RemoveAt

Remove's a "DataRow" object from "DataTable" depending on index position of the "DataTable".

What's basic use of "DataView"?

"DataView" represent's a complete table or can be small section of rows depending on some criteria. It's best used for sorting and finding data with in "datatable".

Dataview has the following methods :-

Find

Take's an array of values and returns the index of the row.

FindRow

This also takes array of values but returns a collection of "DataRow".

If we want to manipulate data of "DataTable" object create "DataView" (Using the "DefaultView" we can create "DataView" object) of the "DataTable" object. and use the following functionalities :-

AddNew

Add's a new row to the "DataView" object.

Delete

Delete the specified row from "DataView" object.

What's difference between "DataSet" and "DataReader"?

> *Twist :- Why is DataSet slower than DataReader?*
>
> *Fourth point is the answer to the twist.*
>
> *Note:- This is my best question and I expect everyone to answer it. It's asked almost 99% in all companies....Basic very Basic cram it.*

Following are the major difference between "DataSet" and "DataReader" :-

- √ “DataSet” is a disconnected architecture, while “DataReader” has live connection while reading data. So if we want to cache data and pass to a different tier “DataSet” forms the best choice and it has decent XML support.
- √ When application needs to access data from more than one table “DataSet” forms the best choice.
- √ If we need to move back while reading record’s, “datareader” does not support this functionality.
- √ But one of the biggest drawbacks of DataSet is speed. As “DataSet” carry considerable overhead because of relations, multiple tables etc speed is slower than “DataReader”.Always try to use “DataReader” wherever possible, as it’s meant specially for speed performance.

How can we load multiple tables in a DataSet?

```
objCommand.CommandText = "Table1"

objDataAdapter.Fill(objDataSet, "Table1")

objCommand.CommandText = "Table2"

objDataAdapter.Fill(objDataSet, "Table2")
```

Above is a sample code which shows how to load multiple “DataTable” object’s in one “DataSet” object.Sample code shows two tables “Table1” and “Table2” in object ObjDataSet.

```
lstdata.DataSource = objDataSet.Tables("Table1").DefaultView
```

In order to refer “Table1” DataTable, use Tables collection of DataSet and the Defaultview object will give you the necessary output.

How can we add relation’s between table in a DataSet?

```
Dim objRelation As DataRelation

objRelation=New

DataRelation("CustomerAddresses",objDataSet.Tables("Customer").Columns("Custid")

,objDataSet.Tables("Addresses").Columns("Custid_fk"))

objDataSet.Relations.Add(objRelation)
```

Relation's can be added between "DataTable" objects using the "DataRelation" object. Above sample code is trying to build a relationship between "Customer" and "Addresses" "Datatable" using "CustomerAddresses" "DataRelation" object.

What's the use of CommandBuilder?

"CommandBuilder" builds "Parameter" objects automatically. Below is a simple code which uses commandbuilder to load its parameter objects.

```
Dim pobjCommandBuilder As New OleDbCommandBuilder(pobjDataAdapter)

pobjCommandBuilder.DeriveParameters(pobjCommand)
```

Be careful while using "DeriveParameters" method as it needs a extra trip to the Datastore which can be very inefficient.

What's difference between "Optimistic" and "Pessimistic" locking?

In pessimistic locking when user wants to update data it locks the record and till then no one can update data. Other user's can only view the data when there is pessimistic locking.

In optimistic locking multiple user's can open the same record for updating, thus increase maximum concurrency. Record is only locked when updating the record. This is the most preferred way of locking practically. Now a days browser based application are very common and having pessimistic locking is not a practical solution.

How many way's are there to implement locking in ADO.NET?

Following are the ways to implement locking using ADO.NET :-

√ When we call "Update" method of DataAdapter it handles locking internally.If the DataSet values are mot matching with current data in Database it raises Concurrency exception error. We can easily trap this error using Try-Catch block and raise appropriate error message to the user.

√ Define Datetime stamp field in the table. When actually you are firing the UPDATE SQL statements compare the current timestamp with one existing in the database. Below is a sample SQL which checks for timestamp before updating and any mismatch in timestamp it will not update the records. This is the best practice used by industries for locking.

Update table1 set field1=@test where LastTimeStamp=@CurrentTimeStamp

√ Check for original values stored in SQL SERVER and actual changed values. In stored procedure check before updating that the old data is same as the current. Example in the below shown SQL before updating field1 we check that is the old field1 value same. If not then some one else has updated and necessary action has to be taken.

Update table1 set field1=@test where field1 = @oldfield1value

Locking can be handled at ADO.NET side or at SQL SERVER side i.e. in stored procedures.for more details of how to implementing locking in SQL SERVER read "What are different locks in SQL SERVER?" in SQL SERVER chapter.

Note:- This is one of the favorite question's of interviewer, so cram it....When I say cram it i do not mean it.... I mean understand it. This book has tried to cover ADO.NET as much as possible, but indeterminist nature of ADO.NET interview questions makes it difficult to make full justice. But hope so that the above questions will make you quiet confident during interviews.

How can we perform transactions in .NET?

The most common sequence of steps that would be performed while developing a transactional application is as follows:

√ Open a database connection using the Open method of the connection object.

√ Begin a transaction using the Begin Transaction method of the connection object. This method provides us with a transaction object that we will use later to commit or rollback the transaction. Note that changes caused by any queries executed before calling the Begin Transaction method will be committed to the database immediately after they execute. Set the Transaction property of the command object to the above mentioned transaction object.

√ Execute the SQL commands using the command object. We may use one or more command objects for this purpose, as long as the Transaction property of all the objects is set to a valid transaction object.

- √ Commit or roll back the transaction using the Commit or Rollback method of the transaction object.
- √ Close the database connection.

What’s difference between Dataset. clone and Dataset. copy?

Clone: - It only copies structure, does not copy data.

Copy: - Copies both structure and data.

Whats the difference between Dataset and ADO Recordset?

There two main basic differences between recordset and dataset :-

- √ With dataset you can retrieve data from two databases like oracle and SQL Server and merge them in one dataset, with recordset this is not possible
- √ All representation of Dataset is using XML while recordset uses COM.
- √ Recordset can not be transmitted on HTTP while Dataset can be.

5. Notification Services

What are notification services?

Notification services help you to deliver messaging application which can deliver customized messages to huge group of subscribers.

In short it's a software application which sits between the information and the recipient.

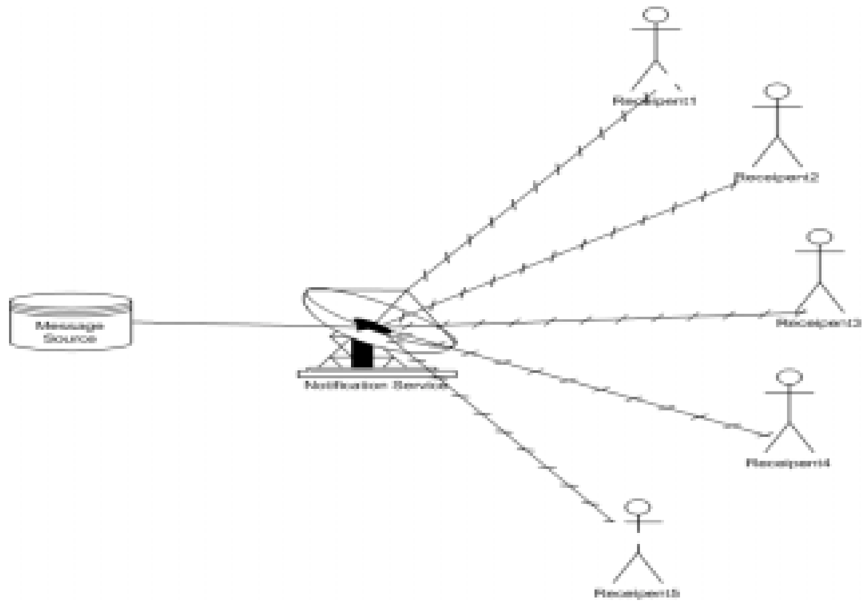


Figure 5.1: - Overall Notification Service architecture

(DB)What are basic components of Notification services?

Following are three basic components of SQL Server notification services:-

- √ Events: - Events are triggers on which user wants relevant information. Example: - Stock exchange can have "stocks price down" events or a weather department has "heavy rainfall" events.

√ Subscriptions: - Subscriptions are nothing but user showing interest in certain events and registering them to the events for information. For instance a user may subscribe to "heavy rainfall" event. In short subscription links the user and the event.

√ Notifications: - Notification is the actual action which takes place that is message is sent to the actual user who has shown interest in the event. Notification can be in various formats and to variety of devices.

√ Notification engine: - This is the main coordinator who will monitor for any events; if any event occurs it matches with the subscribers and sends the notifications.

In short notification engine is the central engine which manages "Events", "Subscriptions" and "Notifications".

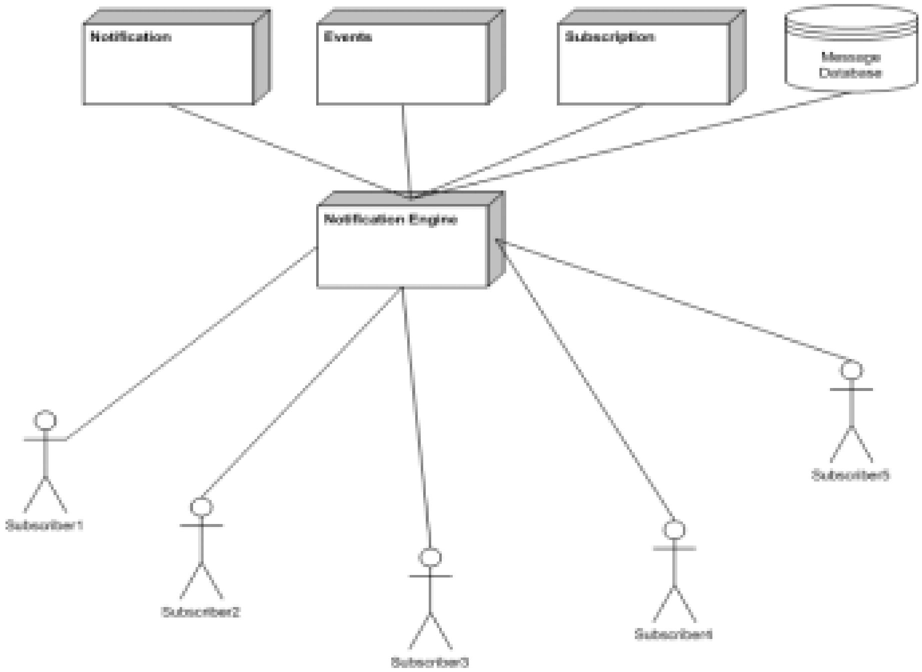


Figure 5.2 : - Events, Subscription, Notification and Notification engine in action.

(DB)Can you explain architecture of Notification Services?

Note: - This will go more as a DBA question.

Following are the detail sections in SQL notification services:-

Notification Service application:- It's a simple user application which will be used to add subscription to the subscription database.

Event providers :- All events reside in Event providers. There are two event providers which are provided by default "File System watcher" and "SQL Server event provider". "File System watcher" detects changes in operating system files. "SQL Server event provider" watches for SQL Server or analysis service database for change. You can also plug in custom event providers. When event providers find any change in database they send the event in "Event" table.

Generator :- Generator checks the event database, when it finds any event it tries to match with the subscription and sends it to the notification database. So generator in short is the decision maker between the subscribers and the events.

Distributor: - Distributor continuously pools the "Notification" database for any "Notification's" to be processed. If the distributor finds any entry it retrieves it and formats it so that it can be delivered to the end recipient. Formatting is normally done using "XML" and "XSLT" for rendering purpose. After the formatting is done it is then pushed to the "distribution providers". They are nothing but medium of delivery. There are three built-in providers:-

- √ SMTP provider
- √ File provider
- √ HTTP provider

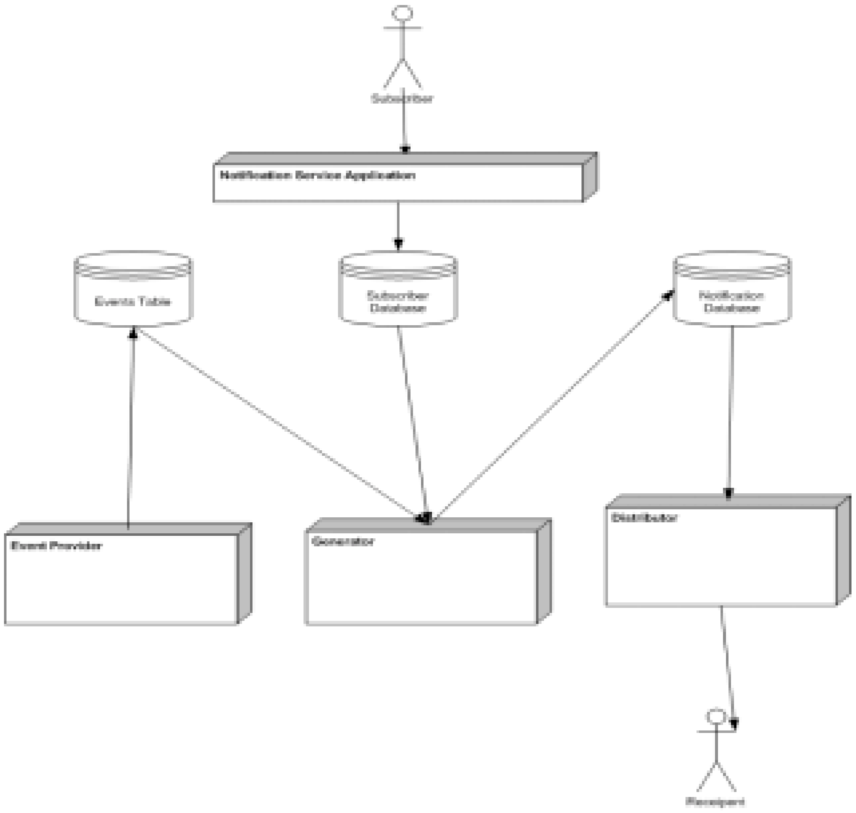


Figure 5.3 : - Detail architecture of SQL notification services.

(DB)Which are the two XML files needed for notification services?

What are ADF and ACF XML files for?

ADF (Application definition file) and ACF (Application configuration file) are core XML files which are needed to configure “Notification Services application”.

ACF file defines the instance name and the application’s directory path of the application. This is the application which runs the notification service.

ADF file describes the event, subscription, rules, and notification structure that will be employed by the Notification Services application.

After these files have been defined you have to load the ADF file using the command line utility or the UI provided by SQL Server 2005. Click on the server browser as show below and expand the "Notification Services", the right click to bring up the "Notification Services" dialog box.

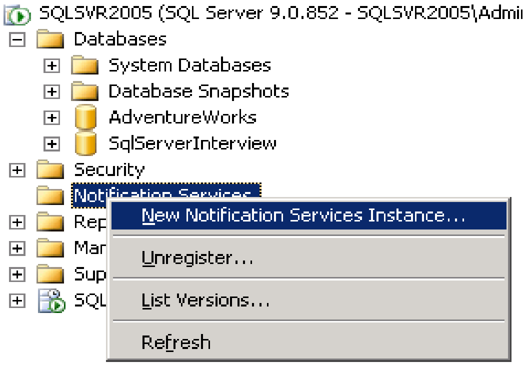
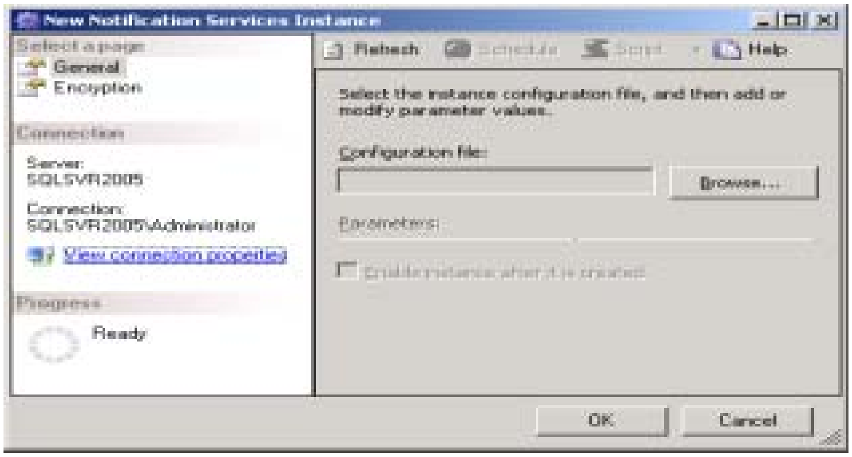


Figure 5.4 : - Notification Services Explorer



Select the instance configuration file, and then add or modify parameter values.

Server:
SQLSVR2005

Connection:
SQLSVR2005\Administrator

View connection properties

□ Enable instance after it is created

Ready

Figure 5.5 : - Notification Services dialog box

You will also have to code the logic to add subscription here's a small sample.

```
using Microsoft.SqlServer.NotificationServices;
using System.Text;
public class NSSubscriptions
{
    private string AddSubscription(string instanceName, string
      applicationName, string subscriptionClassName, string subscriberId)
   {
       NSInstance myNSInstance = new NSInstance(instanceName);
       NSApplication myNSApplication = new NSApplication
        (myNSInstance, applicationName);
       Subscription myNSSubscription = new Subscription
         (myNSApplication, subscriptionClassName);
       myNSSubscription.Enabled = true;
       myNSSubscription.SubscriberId = subscriberId;
       myNSSubscription["Emailid"] = "shiv_koirala@yahoo.com";
       string subscriptionId = myNSSubscription.Add();
       return subscriptionId;
   }
}
```

Note: - As this been an interview book it's beyond the scope of the book to go in to detail of how to create notification. It's better to create a small sample using MSDN and get some fundamentals clear of how practically "Notification services" are done. Try to understand the full format of both the XML files.

(DB)What is Nscontrols command?

For creating notification services you can either use the dialog box of notification service or use the command line utility "Nscontrols" command. So just in short "Nscontrol" is a command-line tool that's used to create and administer Notification Services applications.

Note: - You can refer MSDN for "Nscontrol" commands.

What are the situations you will use "Notification" Services?

Note: - This question is to judge if you can practically map the use of notification service with real world.

If you have read the answers I am sure you can map to lot of practical application. But still in case you want some decent answer you can say about weather forecast, stock ticker etc. Let's say you want a build an application where the user wants to be alerted on the mobile if the weather reaches XYZ degrees or PQR cm rainfall. Then you can create a notification service, provide the appropriate provider and map to it to the devices.

6. Service Broker

What do we need Queues?

There are instances when we expect that the other application with which we are interacting are not available. For example when you chat on messaging system like yahoo, MSN, ICQ etc, you do not expect that the other users will be guaranteed online. So there is where we need queues. So during chatting if the user is not online all the messages are sent to a queue. Later when the user comes online he can read all messages from the queue.

What is “Asynchronous” communication?

Once a client has send messages to the other end application, he can continue with some other task without waiting for any notifications from the end client. For instance take an example of any online email systems. Once you have sent a mail to the end user, you do not have to wait for notification from the ends user. User just sends the message to queue which is later picked up by the mailing system and sent to the desired end-user.

> *Note: - MSMQ does the messaging and queuing, but now the queuing functionality is leveraged to SQL Server 2005, due to its practical needs.*

What is SQL Server Service broker?

SQL Server Service broker provides asynchronous queuing functionality to SQL Server. So now the end client will not have to wait. He can just say add these 1000 records and then come back after one hour or so to see has the work been done or not.

What are the essential components of SQL Server Service broker?

Following are the essential components of SQL Server:-

√ End-Points

The endpoints can be two applications running on different servers or instances, or they can be two applications running on the same server.

√ Message

A message is an entity that is exchanged between Server Brokers. A message must have a name and data type. Optionally, a message can have a validation on that type of data. A message is part of a conversation and it has a unique identifier as well as a unique sequence number to enforce message ordering.

- √ Dialog

Dialog ensure messages to be read in the same order as they where put in to queue between endpoints. In short it ensures proper ordered sequence of events at both ends for a message.

- √ Conversation Group

Conversation Group is a logical grouping of Dialog. To complete a task you can need one or more dialog. For instance an online payment gateway can have two Dialog's first is the "Address Check" and second is the "Credit Card Number" validation, these both dialog form your complete "Payment process". So you can group both the dialogs in one Conversation Group.

- √ Message Transport

Message transport defines how the messages will be send across networks. Message transport is based on TCP/IP and FTP. There are two basic protocols "Binary Adjacent Broker Protocol" which is like TCP/IP and "Dialog Protocol" which like FTP.

What is the main purpose of having Conversation Group?

There two main purpose of having conversation group:-

- √ You can lock a conversation group during reading, so that no other process can read those queue entries.
- √ The most difficult thing in an asynchronous message system is to maintain states. There is huge delay between arrivals of two messages. So conversation groups maintains state using state table. Its uses instance ID to identify messages in a group.

How to implement Service Broker?

Below are the steps for practical implementation:-

- √ Create a Messagetype which describes how the message is formed. If the message type is XML you can also associate a schema with it.

- Further you have to assign these Messagetype to Contract. Messagetype is grouped in Contracts. Contract is an entity which describes messages for a particular Dialog. So a contract can have multiple messagetype's.
- Contracts are further grouped in service. Service has all the dialogs needed to complete one process.
- Service can further be attached to multiple queues. Service is the basic object from SQL Server Service broker point of view.
- So when any client wants to communicate with a queue he opens a dialog with the service.

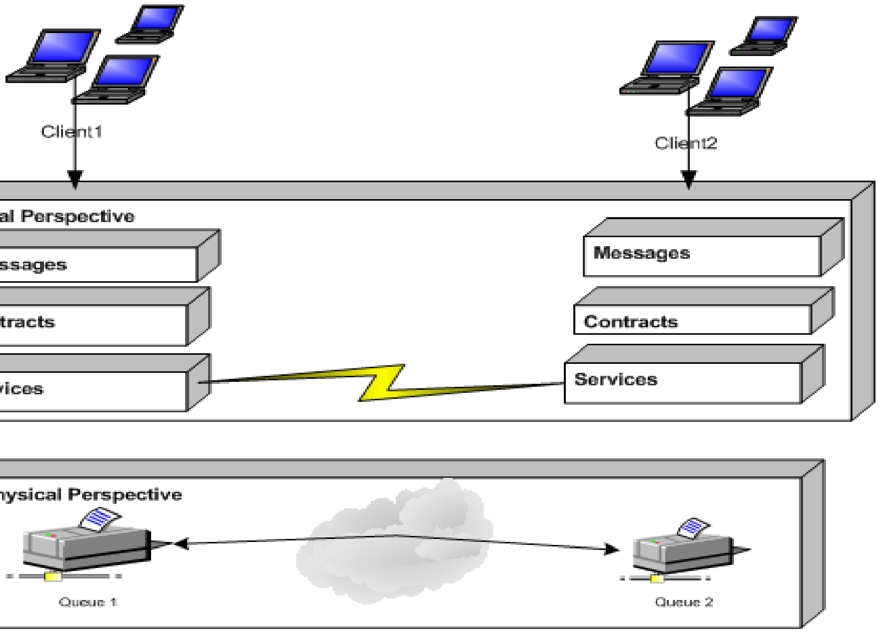


Figure 6.1 : - SQL Server Service Broker in Action.

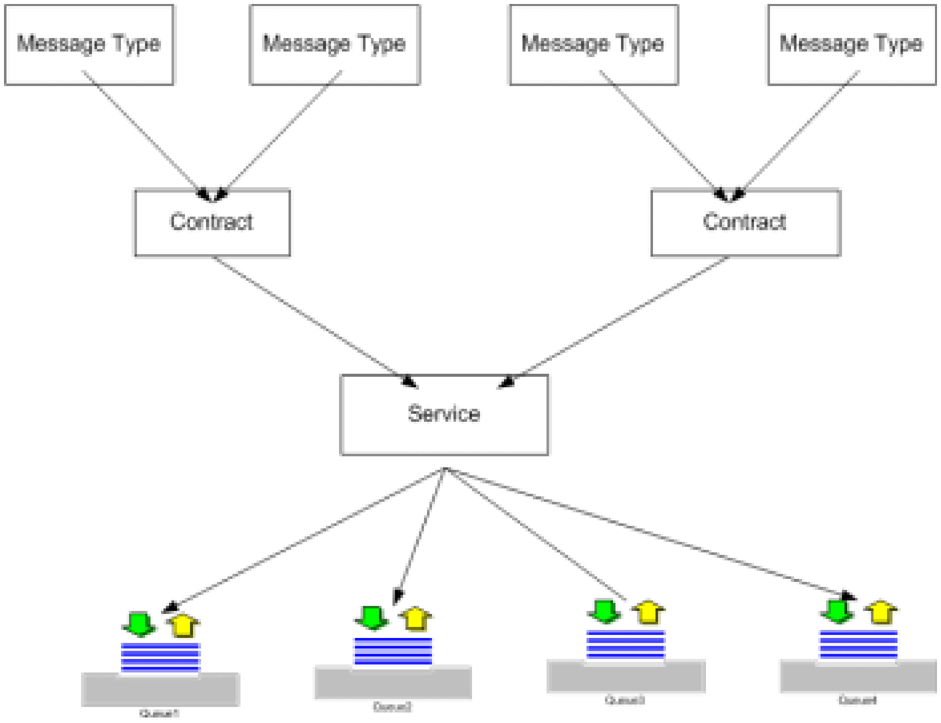


Figure 6.2 : - Message, contract and service

Above figure shows how SQL Server Service broker works. Client who want to use the queues do not have to understand the complexity of queues. They only communicate with the logical view of SQL Server Service broker objects (Messages, Contracts and Services). In turn these objects interact with the queues below and shield the client from any physical complexities of queues.

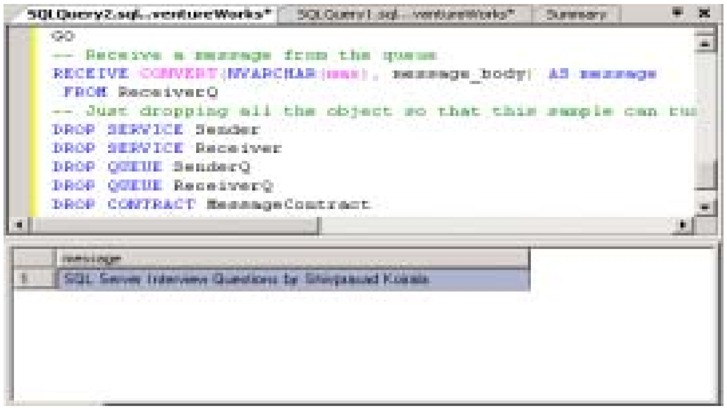
Below is a simple practical implementation of how this works. Try running the below statements from a T-SQL and see the output.

```
-- Create a Message type and do not do any data type validation for this

CREATE MESSAGE TYPE MessageType

VALIDATION = NONE
```

```
GO
-- Create Message contract what type of users can send these messages at this moment we are
defining current as an initiator
CREATE CONTRACT MessageContract
(MessageType SENT BY INITIATOR)
GO
-- Declare the two end points that's sender and receive queues
CREATE QUEUE SenderQ
CREATE QUEUE ReceiverQ
GO
-- Create service and bind them to the queues
CREATE SERVICE Sender
  ON QUEUE SenderQ
CREATE SERVICE Receiver
  ON QUEUE ReceiverQ (MessageContract)
GO
-- Send message to the queue
DECLARE @conversationHandle UNIQUEIDENTIFIER
DECLARE @message NVARCHAR(100)
BEGIN
  BEGIN TRANSACTION;
  BEGIN DIALOG @conversationHandle
       FROM SERVICE Sender
       TO SERVICE 'Receiver'
       ON CONTRACT MessageContract
```

```
-- Sending message
  SET @message = N'SQL Server Interview Questions by Shivprasad Koirala';
  SEND  ON CONVERSATION @conversationHandle
      MESSAGE TYPE MessageType (@message)
 COMMIT TRANSACTION
END
GO
-- Receive a message from the queue
RECEIVE CONVERT(NVARCHAR(max), message_body) AS message
 FROM ReceiverQ
-- Just dropping all the object so that this sample can run successfully
DROP SERVICE Sender
DROP SERVICE Receiver
DROP QUEUE SenderQ
DROP QUEUE ReceiverQ
DROP CONTRACT MessageContract
DROP MESSAGE TYPE MessageType
GO
```

After executing the above T-SQL command you can see the output below.



```sql
GO
-- Receive a message from the queue
RECEIVE CONVERT(NVARCHAR(max), message_body) AS message
 FROM ReceiverQ
-- Just dropping all the object so that this sample can ru
DROP SERVICE Sender
DROP SERVICE Receiver
DROP QUEUE SenderQ
DROP QUEUE ReceiverQ
DROP CONTRACT MessageContract
```

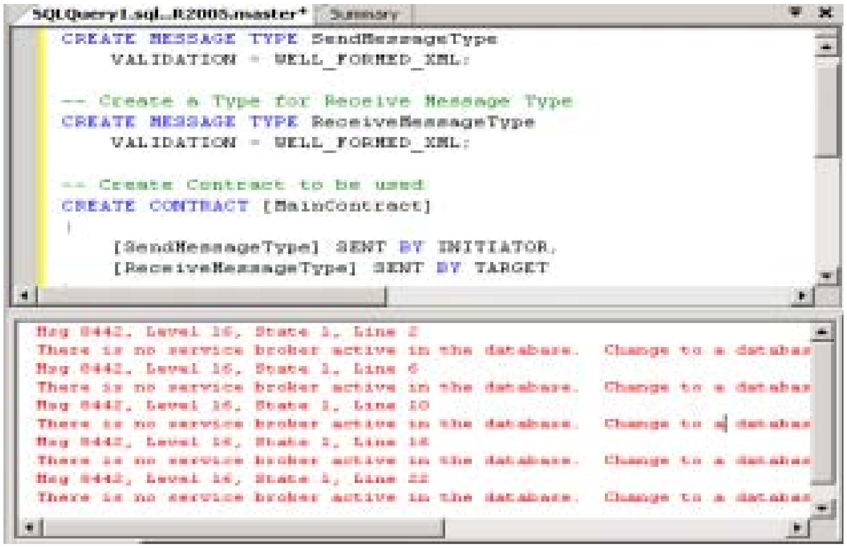
	message
1	SQL Server Interview Questions by Shivprasad Koirala

Figure 6.3 : - Output of the above sample

Note:- In case your SQL Server service broker is not active you will get the following error as shown below. In order to remove that error you have to enable the service broker by using

Alter Database [DatabaseName] set Enable_broker

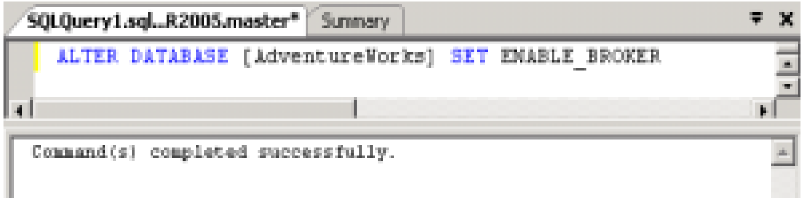
At this moment I have created all these samples in the sample database "AdventureWorks".



```
CREATE MESSAGE TYPE SendMessageType
    VALIDATION = WELL_FORMED_XML;

-- Create a Type for Receive Message Type
CREATE MESSAGE TYPE ReceiveMessageType
    VALIDATION = WELL_FORMED_XML;

-- Create Contract to be used
CREATE CONTRACT [MainContract]
(
    [SendMessageType] SENT BY INITIATOR,
    [ReceiveMessageType] SENT BY TARGET
```

```
Msg 8447, Level 16, State 1, Line 2
There is no service broker active in the database.  Change to a databas
Msg 8447, Level 16, State 1, Line 6
There is no service broker active in the database.  Change to a databas
Msg 8447, Level 16, State 1, Line 10
There is no service broker active in the database.  Change to a databas
Msg 8447, Level 16, State 1, Line 18
There is no service broker active in the database.  Change to a databas
Msg 8447, Level 16, State 1, Line 22
There is no service broker active in the database.  Change to a databas
```

Figure 6.4 : - Error Service broker not active



```
ALTER DATABASE [AdventureWorks] SET ENABLE_BROKER
```

```
Command(s) completed successfully.
```

Figure 6.5 : - Enabling Service broker

How do we encrypt data between Dialogs?

If you create a dialog using "WITH ENCRYPTION” clause a session key is created that’s used to encrypt the messages sent between dialog.

7. XML Integration

Note: - In this chapter we will first just skim through basic XML interview questions so that you do not get stuck up with simple questions.

What is XML?

XML (Extensible markup language) is all about describing data. Below is a XML which describes invoice data.

```
<?xml version="1.0" encoding="ISO-8859-1"?>

<invoice>

<productname>Shoes</productname>

<qty>12</qty>

<totalcost>100</totalcost>

<discount>10</discount>

</invoice>
```

An XML tag is not something predefined but it is something you have to define according to your needs. For instance in the above example of invoice all tags are defined according to business needs. The XML document is self explanatory, any one can easily understand looking at the XML data what exactly it means.

What is the version information in XML?

“version” tag shows which version of XML is used.

What is ROOT element in XML?

In our XML sample given previously <invoice></invoice> tag is the root element. Root element is the top most element for a XML.

If XML does not have closing tag will it work?

No, every tag in XML which is opened should have a closing tag. For instance in the top if I remove </discount> tag that XML will not be understood by lot of application.

Is XML case sensitive?

Yes, they are case sensitive.

What’s the difference between XML and HTML?

XML describes data while HTML describes how the data should be displayed. So HTML is about displaying information while XML is about describing information.

Is XML meant to replace HTML?

No they both go together one is for describing data while other is for displaying data.

Can you explain why your project needed XML?

> *Note: - This is an interview question where the interviewer wants to know why you have chosen XML.*

Remember XML was meant to exchange data between two entities as you can define your user friendly tags with ease. In real world scenarios XML is meant to exchange data. For instance you have two applications who want to exchange information. But because they work in two complete opposite technologies it’s difficult to do it technically. For instance one application is made in JAVA and the other in .NET. But both languages understand XML so one of the applications will spit XML file which will be consumed and parsed by other application.s

You can give a scenario of two applications which are working separately and how you chose XML as the data transport medium.

What is DTD (Document Type definition)?

It defines how your XML should structure. For instance in the above XML we want to make it compulsory to provide “qty” and “totalcost”, also that these two elements can only contain numeric. So you can define the DTD document and use that DTD document with in that XML.

What is well formed XML?

If a XML document is confirming to XML rules (all tags started are closed, there is a root element etc) then it’s a well formed XML.

What is a valid XML?

If XML is confirming to DTD rules then it's a valid XML.

What is CDATA section in XML?

All data is normally parsed in XML but if you want to exclude some elements you will need to put those elements in CDATA.

What is CSS?

With CSS you can format a XML document.

What is XSL?

XSL (the eXtensible Stylesheet Language) is used to transform XML document to some other document. So its transformation document which can convert XML to some other document. For instance you can apply XSL to XML and convert it to HTML document or probably CSV files.

What is Element and attributes in XML?

In the below example invoice is the element and the invnumber the attribute.

<invoice invnumber=1002></invoice>

Can we define a column as XML?

Yes, this is a new feature provided by SQL Server. You can define a column data type as XML for a table.

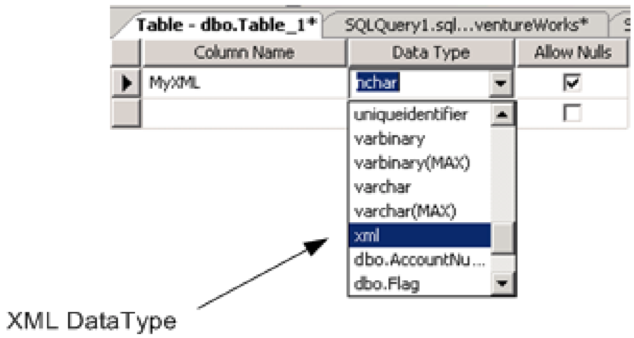


Figure 7.1 : - Specify XML data type

How do we specify the XML data type as typed or untyped?

If there is a XSD schema specified to the data type then it’s typed or else it’s untyped. If you specify XSD then with every insert SQL Server will try to validate and see that is the data adhering to XSD specification of the data type.

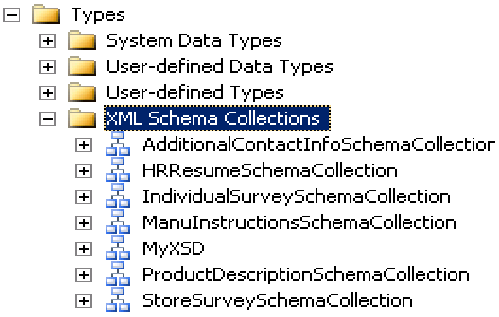
How can we create the XSD schema?

Below is the DDL statement for creating XML schema.

```
CREATE XML SCHEMA COLLECTION MyXSD AS
N'<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
 elementFormDefault="qualified" targetNamespace="http://MyXSD">
  <xs:element name="MyXSD">
   <xs:complexType>
    <xs:sequence>
      <xs:element name="Orderid" type="xs:string" />
      <xs:element name="CustomerName" type="xs:string" />
    </xs:sequence>
```

```
        </xs:complexType>
      </xs:element>
    </xs:schema>'
```

After you have created the schema you see the MYXSD schema in the schema collections folder.



- Types
  - System Data Types
  - User-defined Data Types
  - User-defined Types
  - XML Schema Collections
    - AdditionalContactInfoSchemaCollectior
    - HRResumeSchemaCollection
    - IndividualSurveySchemaCollection
    - ManuInstructionsSchemaCollection
    - MyXSD
    - ProductDescriptionSchemaCollection
    - StoreSurveySchemaCollection

Figure 7.2 : - You can view the XSD in explorer of Management Studio

When you create the XML data type you can assign the MyXsd to the column.

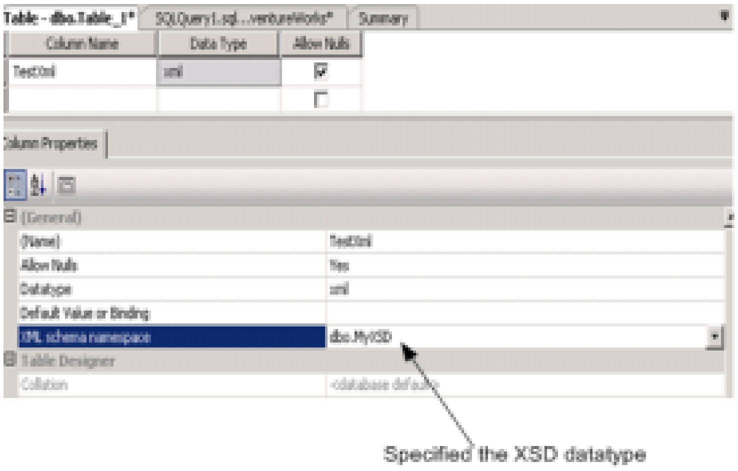


Table - dbo.Table_1* | SQLQuery1.sql...ventureWorks* | Summary

Column Name	Data Type	Allow Nulls
TestXml	xml	☑
		□

Column Properties

(General)

(Name)	TestXml
Allow Nulls	Yes
Datatype	xml
Default Value or Binding	
XML schema namespace	dbo.MyXSD

Table Designer

Collation	<database default>

Specified the XSD datatype

Figure 7.3 : - MyXSD assigned to a column

How do I insert in to a table which has XSD schema attached to it?

I know many developers will just say what the problem with simple insert statement. Well guys its not easy with attaching the XSD its now a well formed datatype.The above table I have named as xmltable. So we had specified in the schema two nodes one is ordered and the other customername. So here's the insert.

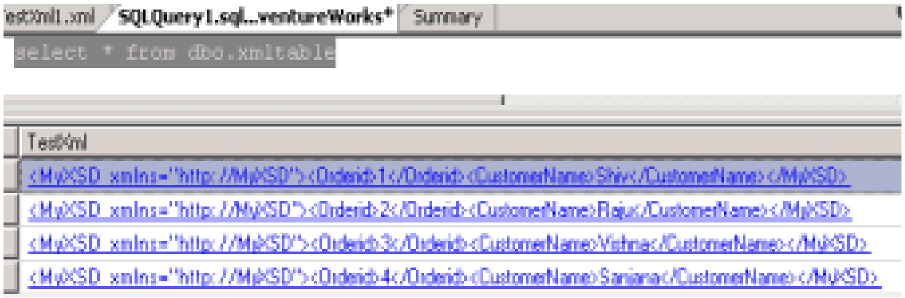
> *Insert into xmltable values ('<MyXSD xmlns="http://MyXSD"><Orderid>1</Orderid><CustomerName>Shiv</CustomerName></MyXSD>')*

What is maximum size for XML datatype?

2 GB and is stored like varbinary.

What is Xquery?

In a typical XML table below is the type of data which is seen. Now I want to retrieve orderid "4". I know many will jump up with saying use the "LIKE" keyword. Ok you say that interviewer is very sure that you do not know the real power of XML provided by SQL Server.



```
select * from dbo.xmltable
```

TestXml
<MyXSD xmlns="http://MyXSD"><Orderid>1</Orderid><CustomerName>Shiv</CustomerName></MyXSD>
<MyXSD xmlns="http://MyXSD"><Orderid>2</Orderid><CustomerName>Raju</CustomerName></MyXSD>
<MyXSD xmlns="http://MyXSD"><Orderid>3</Orderid><CustomerName>Vishnu</CustomerName></MyXSD>
<MyXSD xmlns="http://MyXSD"><Orderid>4</Orderid><CustomerName>Sanjana</CustomerName></MyXSD>

Figure 7.4 : - XML data

Well first thing XQUERY is not that something Microsoft invented, it's a language defined by W3C to query and manipulate data in a XML. For instance in the above scenario we can use XQUERY and drill down to specific element in XML.

So to drill down here's the XQUERY

```
SELECT * FROM xmltable
WHERE TestXml.exist('declare namespace
xd=http://MyXSD/xd:MyXSD[xd:Orderid eq "4"]') = 1
```

Note: - It's out of the scope of this book to discuss XQUERY. I hope and only hope guys many interviewers will not bang in this section. In case you have doubt visit www.w3c.org or SQL Server books online they have a lot of material in to this.

What are XML indexes?

XML data types have huge size 2 GB. But first thing is that you should have a primary key on the XML data type column. Then you can use the below SQL statement to create index on the XML column:-

```
CREATE PRIMARY XML INDEX xmlindex ON xmltable(TestXML)
```

What are secondary XML indexes?

Secondary indexes are built on document attributes.

What is FOR XML in SQL Server?

FOR XML clause returns data in XML rather than simple rows and columns. For instance if you fire the below query on any table you will get XML output:-

```
SELECT * FROM MyTable FOR XML AUTO
```

Can I use FOR XML to generate SCHEMA of a table and how?

The below SQL syntax will return the SCHEMA of the table.

```
SELECT * FROM MyTable FOR XML AUTO, XMLSCHEMA
```

What is the OPENXML statement in SQL Server?

We had seen that FOR XML returns a XML format of a table data. And FOR XML does the vice versa of it. If you pass XML document to it will convert it to rows and columns.

I have huge XML file which we want to load in database?

Twist: - Can I do a BULK load of XML in database?

Below is the SQL statement which will insert from “MyXml.xml” in to “MyTable”.

```
INSERT into MyTable(MyXMlColumn) SELECT * FROM OPENROWSET
   (Bulk 'c:\MyXml.xml', SINGLE_CLOB) as abc
```

How to call stored procedure using HTTP SOAP?

Twist: - Can I create web services for SQL Server objects?

Note: - Ok every one reading this answer out of dedication I have switched off my mobile and I am writing this answer.

You can call a stored procedure using HTTP SOAP. This can be done by creating END POINTS using the “CREATE ENDPOINT” DDL statement. I have created a TotalSalesHttpEndPoint which can be called later through “webservices”.

```
CREATE ENDPOINT TotalSalesHttpEndPOint
STATE = STARTED
AS HTTP(
   PATH = '/sql',
   AUTHENTICATION = (INTEGRATED ),
   PORTS = ( CLEAR ),
   SITE = 'server'
   )
FOR SOAP (
    WEBMETHOD 'http://tempUri.org/'.'GetTotalSalesOfProduct'
            (name='AdventureWorks.dbo.GetTotalSalesOfProduct',
           schema=STANDARD ),
BATCHES = ENABLED,
WSDL = DEFAULT,
DATABASE = 'AdventureWorks',
 NAMESPACE = 'http://AdventureWorks/TotalSales'
)
```

What is XMLA?

XMLA stand for XML for Analysis Services. Analysis service is covered in depth in data mining and data ware housing chapters. Using XMLA we can expose the Analysis service data to the external world in XML. So that any data source can consume it as XML is universally known.

8. Data Warehousing/Data Mining

Note: - "Data mining" and "Data Warehousing" are concepts which are very wide and it's beyond the scope of this book to discuss it in depth. So if you are specially looking for a "Data mining / warehousing" job its better to go through some reference books. But below questions can shield you to some good limit.

What is "Data Warehousing"?

"Data Warehousing "is a process in which the data is stored and accessed from central location and is meant to support some strategic decisions. "Data Warehousing" is not a requirement for "Data mining". But just makes your Data mining process more efficient.

Data warehouse is a collection of integrated, subject-oriented databases designed to support the decision-support functions (DSF), where each unit of data is relevant to some moment in time.

What are Data Marts?

Data Marts are smaller section of Data Warehouses. They help data warehouses collect data. For example your company has lot of branches which are spanned across the globe. Head-office of the company decides to collect data from all these branches for anticipating market. So to achieve this IT department can setup data mart in all branch offices and a central data warehouse where all data will finally reside.

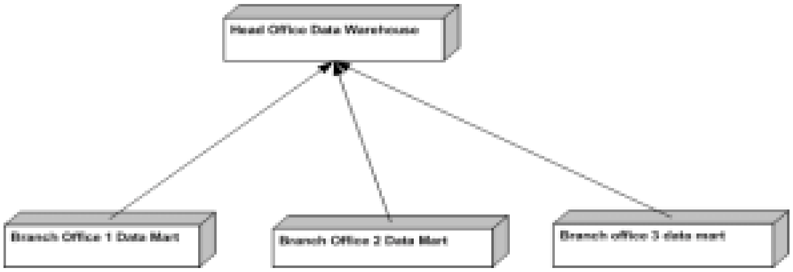

Head Office Data Warehouse
Branch Office 1 Data Mart
Branch Office 2 Data Mart
Branch office 3 data mart

Figure 8.1: - Data Mart in action

What are Fact tables and Dimension Tables?

Twist: - What is Dimensional Modeling?

Twist: - What is Star Schema Design?

When we design transactional database we always think in terms of normalizing design to its least form. But when it comes to designing for Data warehouse we think more in terms of "denormalizing" the database. Data warehousing databases are designed using "Dimensional Modeling". Dimensional Modeling uses the existing relational database structure and builds on that.

There are two basic tables in dimensional modeling:-

- √ Fact Tables.
- √ Dimension Tables.

Fact tables are central tables in data warehousing. Fact tables have the actual aggregate values which will be needed in a business process. While dimension tables revolve around fact tables. They describe the attributes of the fact tables. Let's try to understand these two conceptually.

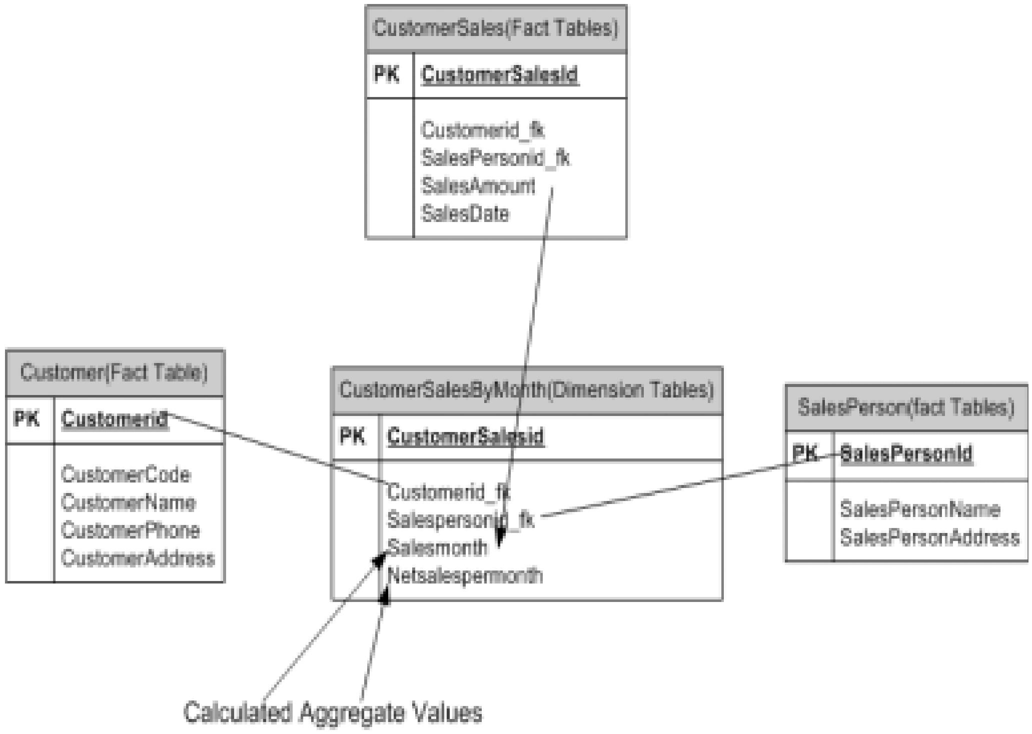


Figure 8.2 : - Dimensional Modeling

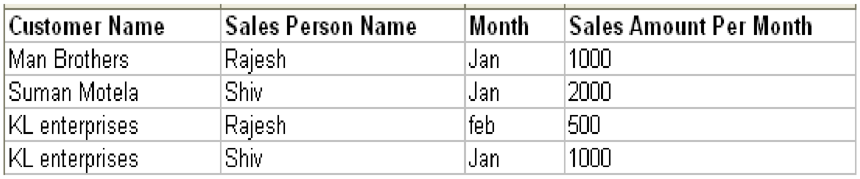
In the above example we have three tables which are transactional tables:-

√ Customer: - It has the customer information details.

√ Salesperson: - Sales person who are actually selling products to customer.

√ CustomerSales: - This table has data of which sales person sold to which customer and what was the sales amount.

Below is the expected report Sales / Customer / Month. You will be wondering if we make a simple join query from all three tables we can easily get this output. But imagine if you have huge records in these three tables it can really slow down your reporting process. So we introduced a third dimension table "CustomerSalesByMonth" which will have foreign key of all tables and the aggregate amount by month. So this table becomes the dimension table and all other tables become fact tables. All major data warehousing design use Fact and Dimension model.



Customer Name	Sales Person Name	Month	Sales Amount Per Month
Man Brothers	Rajesh	Jan	1000
Suman Motela	Shiv	Jan	2000
KL enterprises	Rajesh	feb	500
KL enterprises	Shiv	Jan	1000

Figure 8.3: - Expected Report.

The above design is also called as Star Schema design.

Note: - For a pure data warehousing job this question is important. So try to understand why we modeled out design in this way rather than using the traditional approach - normalization.

(DB)What is Snow Flake Schema design in database?

Twist: - What's the difference between Star and Snow flake schema?

Star schema is good when you do not have big tables in data warehousing. But when tables start becoming really huge it is better to denormalize. When you denormalize star schema it is nothing but snow flake design. For instance below "customeraddress" table is been normalized and is a child table of "Customer" table. Same holds true for "Salesperson" table.

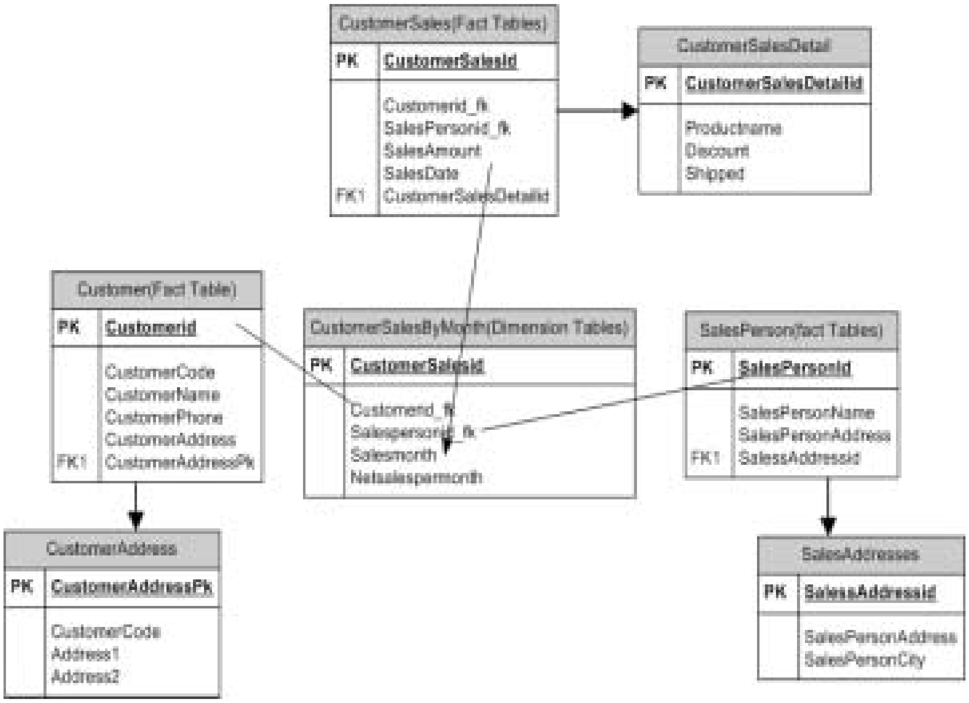


Figure 8.4 : - Snow Flake Schema

(DB)What is ETL process in Data warehousing?

Twist: - What are the different stages in "Data warehousing"?

ETL (Extraction, Transformation and Loading) are different stages in Data warehousing. Like when we do software development we follow different stages like requirement gathering, designing, coding and testing. In the similar fashion we have for data warehousing.

Extraction:-

In this process we extract data from the source. In actual scenarios data source can be in many forms EXCEL, ACCESS, Delimited text, CSV (Comma Separated Files) etc. So extraction process handle's the complexity of understanding the data source and loading it in a structure of data warehouse.

Transformation:-

This process can also be called as cleaning up process. It's not necessary that after the extraction process data is clean and valid. For instance all the financial figures have NULL values but you want it to be ZERO for better analysis. So you can have some kind of stored procedure which runs through all extracted records and sets the value to zero.

Loading:-

After transformation you are ready to load the information in to your final data warehouse database.

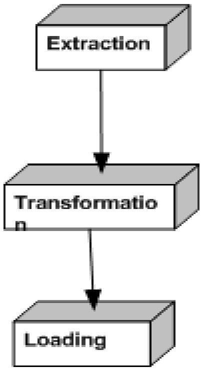


Figure 8.5 : - ETL stages

(DB)How can we do ETL process in SQL Server?

> *I can hear that scream: - Words and words, show us where does this ETL practically fit in SQL Server.*

SQL Server has following ways with which we can import or export data in SQL Server:-

- √ BCP (Bulk Copy Program).
- √ Bulk Insert
- √ DTS (Data Transformation Services).DTS is now called as Integration Services.

What is "Data mining"?

“Data mining” is a concept by which we can analyze the current data from different perspectives and summarize the information in more useful manner. It’s mostly used either to derive some valuable information from the existing data or to predict sales to increase customer market.

There are two basic aims of “Data mining”:-

- √ Prediction: - From the given data we can focus on how the customer or market will perform. For instance we are having a sale of 40000 $ per month in India, if the same product is to be sold with a discount how much sales can the company expect.
- √ Summarization: - To derive important information to analyze the current business scenario. For example a weekly sales report will give a picture to the top management how we are performing on a weekly basis?

Compare “Data mining” and “Data Warehousing”?

“Data Warehousing” is technical process where we are making our data centralized while “Data mining” is more of business activity which will analyze how good your business is doing or predict how it will do in the future coming times using the current data.

As said before “Data Warehousing” is not a need for “Data mining”. It’s good if you are doing “Data mining” on a “Data Warehouse” rather than on an actual production database. “Data Warehousing” is essential when we want to consolidate data from different sources, so it’s like a cleaner and matured data which sits in between the various data sources and brings then in to one format.

“Data Warehouses” are normally physical entities which are meant to improve accuracy of “Data mining” process. For example you have 10 companies sending data in different format, so you create one physical database for consolidating all the data from different company sources, while “Data mining” can be a physical model or logical model. You can create a database in “Data mining” which gives you reports of net sales for this year for all companies. This need not be a physical database as such but a simple query.

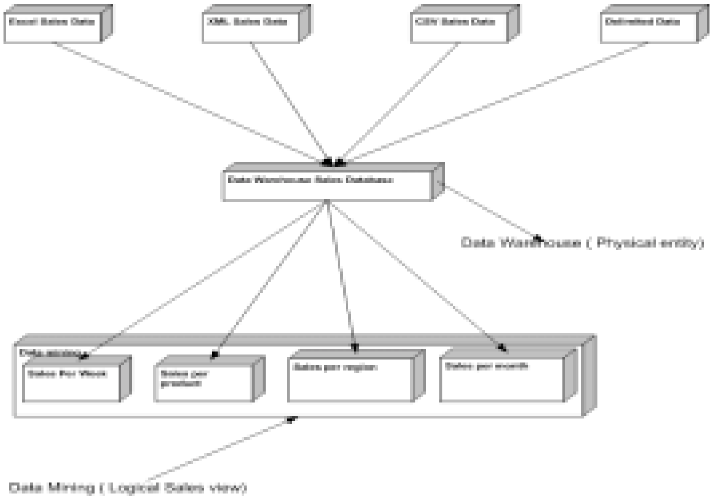


Figure 8.6 : - Data Warehouse and Data mining

The above figure gives a picture how these concepts are quiet different. “Data Warehouse” collects cleans and filters data through different sources like “Excel”, “XML” etc. But “Data Mining” sits on the top of “Data Warehouse” database and generates intelligent reports. Now either it can export to a different database or just generate report using some reporting tool like “Reporting Services”.

What is BCP?

> *Note: - It’s not necessary that this question will be asked for data mining. But if a interviewer wants to know your DBA capabilities he will love to ask this question. If he is a guy who has worked from the old days of SQL Server he will expect this to be answered.*

There are times when you want to move huge records in and out of SQL Server, there’s where this old and cryptic friend will come to use. It’s a command line utility. Below is the detail syntax:-

```
bcp {[[<database name>.][<owner>].]{<table name>|<view name>}|"<query>"}
    {in | out | queryout | format} <data file>
    [-m <maximum no. of errors>] [-f <format file>] [-e <error file>]
    [-F <first row>] [-L <last row>] [-b <batch size>]
```

```
[-n] [-c] [-w] [-N] [-V (60 | 65 | 70)] [-6]
[-q] [-C <code page>] [-t <field term>] [-r <row term>]
[-i <input file>] [-o <output file>] [-a <packet size>]
[-S <server name>[\<instance name>]] [-U <login id>] [-P <password>]
[-T] [-v] [-R] [-k] [-E] [-h "<hint> [,...n]"]
```

UUUHH Lot of attributes there. But during interview you do not have to remember so much. Just remember that BCP is a utility with which you can do import and export of data.

How can we import and export using BCP utility?

In the first question you can see there is huge list of different command. We will try to cover only the basic commands which are used.

-T: - signifies that we are using windows authentification

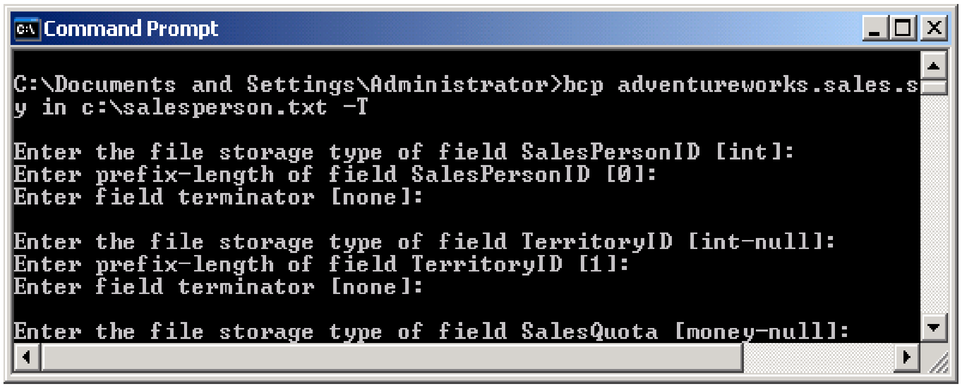
-t: - By default every record is tab separated. But if you want to specify comma separated you can use this command.

-r :- This specifies how every row is separated. For instance specifying –r/n specifies that every record will be separated by ENTER.

```
bcp adventureworks.sales.salesperson out c:\salesperson.txt -T
```

```
bcp adventureworks.sales.salespersondummy in c:\salesperson.txt –T
```

When you execute the BCP syntax you will be prompted to enter the following values (data type, length of the field and the separator) as shown in figure below. You can either fill it or just press enter to escape it. BCP will take in the default values.

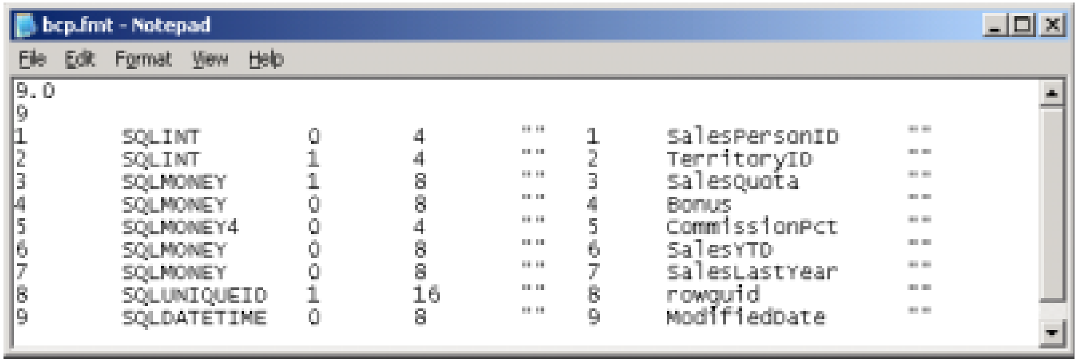


```
C:\Documents and Settings\Administrator>bcp adventureworks.sales.s
y in c:\salesperson.txt -T

Enter the file storage type of field SalesPersonID [int]:
Enter prefix-length of field SalesPersonID [0]:
Enter field terminator [none]:

Enter the file storage type of field TerritoryID [int-null]:
Enter prefix-length of field TerritoryID [1]:
Enter field terminator [none]:

Enter the file storage type of field SalesQuota [money-null]:
```

Figure 8.7 : - After executing BCP command prompts for some properties

During BCP we need to change the field position or eliminate some fields how can we achieve this?

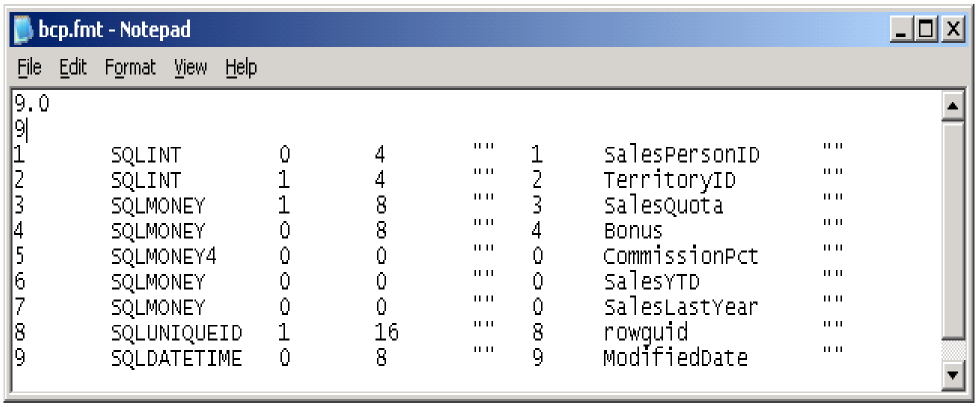
For some reason during BCP you want some fields to be eliminated or you want the positions to be in a different manner. For instance you have field1, field2 and field3. You want that field2 should not be imported during BCP. Or you want the sequence to be changed as field2, field1 and then finally field3. This is achieved by using the format file. When we ran the BCP command in the first question it has generated a file with ".fmt" extension. Below is the FMT file generated in the same directory from where I ran my BCP command.



```
9.0
9
1       SQLINT          0       4       ""      1       SalesPersonID       ""
2       SQLINT          1       4       ""      2       TerritoryID         ""
3       SQLMONEY        1       8       ""      3       SalesQuota          ""
4       SQLMONEY        0       8       ""      4       Bonus               ""
5       SQLMONEY4       0       4       ""      5       CommissionPct       ""
6       SQLMONEY        0       8       ""      6       SalesYTD            ""
7       SQLMONEY        0       8       ""      7       SalesLastYear       ""
8       SQLUNIQUEID     1       16      ""      8       rowguid             ""
9       SQLDATETIME     0       8       ""      9       ModifiedDate        ""
```

Figure 8.8 : - Format file generated due to BCP.

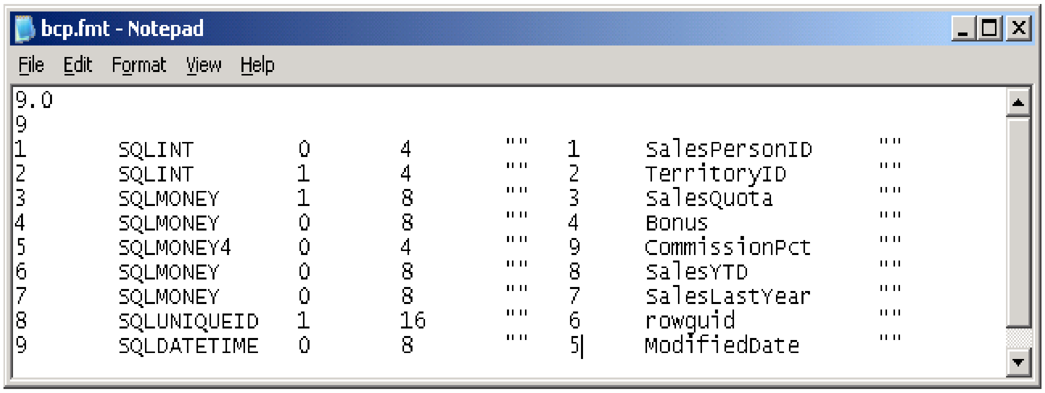
FMT file is basically the format file for BCP to govern how it should map with tables. Lets say, in from our salesperson table we want to eliminate commissionpct, salesytd and saleslastyear. So you have to modify the FMT file as shown below. We have made the values zero for the fields which has to be eliminated.



```
9.0
9
1       SQLINT        0     4     ""   1     SalesPersonID     ""
2       SQLINT        1     4     ""   2     TerritoryID       ""
3       SQLMONEY      1     8     ""   3     SalesQuota        ""
4       SQLMONEY      0     8     ""   4     Bonus             ""
5       SQLMONEY4     0     0     ""   0     CommissionPct     ""
6       SQLMONEY      0     0     ""   0     SalesYTD          ""
7       SQLMONEY      0     0     ""   0     SalesLastYear     ""
8       SQLUNIQUEID   1     16    ""   8     rowguid           ""
9       SQLDATETIME   0     8     ""   9     ModifiedDate      ""
```

Figure 8.9 : - FMT file with fields eliminated

If we want to change the sequence you have to just change the original sequence number. For instance we have changed the sequence from 9 to 5 --> 5 to 9 , see the figure below.



```
9.0
9
1       SQLINT        0     4     ""   1     SalesPersonID     ""
2       SQLINT        1     4     ""   2     TerritoryID       ""
3       SQLMONEY      1     8     ""   3     SalesQuota        ""
4       SQLMONEY      0     8     ""   4     Bonus             ""
5       SQLMONEY4     0     4     ""   9     CommissionPct     ""
6       SQLMONEY      0     8     ""   8     SalesYTD          ""
7       SQLMONEY      0     8     ""   7     SalesLastYear     ""
8       SQLUNIQUEID   1     16    ""   6     rowguid           ""
9       SQLDATETIME   0     8     ""   5     ModifiedDate      ""
```

Figure 8.10 : - FMT file with field sequence changed

Once you have changed the FMT file you can specify the .FMT file in the BCP command arguments as shown below.

```
bcp adventureworks.sales.salesperson in c:\salesperson.txt -
c:\bcp.fmt -T
```

Note: - we have given the .FMT file in the BCP command.

What is Bulk Insert?

Bulk insert is very similar to BCP command but we can not do export with the command. The major difference between BCP and Bulk Insert:-

- √ Bulk Insert runs in the same process of SQL Server, so it can avail to all performance benefits of SQL Server.
- √ You can define Bulk insert as part of transaction. That means you can use the Bulk Insert command in BEGIN TRANS and COMMIT TRANS statements.

Below is a detailed syntax of BULK INSERT. You can run this from “SQL Server Management Studio”, TSQL or ISQL.

```
BULK INSERT [[‘database_name‘.][‘owner‘].]
        {‘table_name‘ | ‘view_name‘ FROM ‘data_file’ }
  [WITH (
       [BATCHSIZE [ = batch_size ]]
       [[,] CHECK_CONSTRAINTS ]
       [[,] CODEPAGE [ = ‘ACP’ | ‘OEM’ | ‘RAW’ | ‘code_page’ ]]
       [[,] DATAFILETYPE [ = {‘char’|’native’|
                    ’widechar’|’widenative’ }]]
       [[,] FIELDTERMINATOR [ = ‘field_terminator’ ]]
       [[,] FIRSTROW [ = first_row ]]
       [[,] FIRETRIGGERS [ = fire_triggers ]]
       [[,] FORMATFILE [ = ‘format_file_path’ ]]
```

```
        [[,] KEEPIDENTITY ]
        [[,] KEEPNULLS ]
        [[,] KILOBYTES_PER_BATCH [ = kilobytes_per_batch ]]
        [[,] LASTROW [ = last_row ]]
        [[,] MAXERRORS [ = max_errors ]]
        [[,] ORDER ( { column [ ASC | DESC ] }[ ,...n ])]
        [[,] ROWS_PER_BATCH [ = rows_per_batch ]]
        [[,] ROWTERMINATOR [ = 'row_terminator' ]]
        [[,] TABLOCK ]
       )]
```

Below is a simplified version of bulk insert which we have used to import a comma separated file in to “SalesPersonDummy”. The first row is the column name so we specified start importing from the second row. The other two attributes define how the fields and rows are separated.

```
bulk insert adventureworks.sales.salespersondummy from 'c:\salesperson.txt' with
(
FIRSTROW=2,
FIELDTERMINATOR = ',',
ROWTERMINATOR = '\n'
)
```

What is DTS?

Note :- It’s now a part of integration service in SQL Server 2005.

DTS provides similar functionality as we had with BCP and Bulk Import. There are two major problems with BCP and Bulk Import:-

√ BCP and Bulk import do not have user friendly User Interface. Well some DBA does still enjoy using those DOS prompt commands which makes them feel doing something worthy.

√ Using BCP and Bulk imports we can import only from files, what if we wanted to import from other database like FoxPro, access, and oracle. That is where DTS is the king.

√ One of the important things that BCP and Bulk insert misses is transformation, which is one of the important parts of ETL process. BCP and Bulk insert allows you to extract and load data, but does not provide any means by which you can do transformation. So for example you are getting sex as “1” and “2”, you would like to transform this data to “M” and “F” respectively when loading in to data warehouse.

√ It also allows you do direct programming and write scripts by which you can have huge control over loading and transformation process.

√ It allows lot of parallel operation to happen. For instance while you are reading data you also want the transformation to happen in parallel , then DTS is the right choice.

You can see DTS Import / Export wizard in the SQL Server 2005 menu.

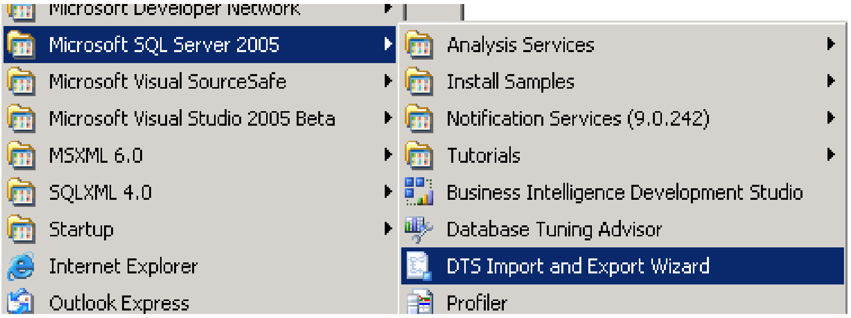

Microsoft Developer Network
Microsoft SQL Server 2005
Microsoft Visual SourceSafe
Microsoft Visual Studio 2005 Beta
MSXML 6.0
SQLXML 4.0
Startup
Internet Explorer
Outlook Express
Analysis Services
Install Samples
Notification Services (9.0.242)
Tutorials
Business Intelligence Development Studio
Database Tuning Advisor
DTS Import and Export Wizard
Profiler

Figure 8.11 : - DTS Import Export

Note: - DTS is the most used technology when you are during Data warehousing using SQL Server. In order to implement the ETL fundamental properly Microsoft has rewritten the whole DTS from scratch using .NET and named it as “Integration Services”. There is a complete chapter which is dedicated to “Integration Services” which will cover DTS indirectly in huge details. Any interviewer who is looking for data warehousing professional in SQL Server 2005 will expect that candidates should know DTS properly.

(DB)Can you brief about the Data warehouse project you worked on?

> *Note: - This question is the trickiest and shoot to have insight, from where the interviewer would like to spawn question threads. If you have worked with a data warehouse project you can be very sure of this. If not then you really have to prepare a project to talk about…. I know it's unethical to even talk in books but?*

I leave this to readers as everyone would like to think of a project of his own. But just try to include the ETL process which every interviewer thinks should be followed for a data warehouse project.

What is an OLTP (Online Transaction Processing) System?

Following are the characteristics of an OLTP system:-

√ They describe the actual current data of the system

√ Transactions are short. For example user fills in data and closes the transaction.

√ Insert/Update/Delete operation is completely online.

√ System design expected to be in the maximum Normalized form.

√ Huge volume of transactions. Example lots of online users are entering data in to an online tax application.

√ Backup of transaction is necessary and needs to be recovered in case of problems

> *Note: - OLTP systems are good at putting data in to database system but serve no good when it comes to analyzing data.*

What is an OLAP (On-line Analytical processing) system?

Following are characteristic of an OLAP system:-

√ It has historical as well as current data.

√ Transactions are long. They are normally batch transaction which is executed during night hours.

√ As OLAP systems are mainly used for reporting or batch processing, so "Denormalization" designs are encouraged.

- √ Transactions are mainly batch transactions which are running so there are no huge volumes of transaction.
- √ Do not need to have recovery process as such until the project specifies specifically.

What is Conceptual, Logical and Physical model?

Depending on clients requirement first you define the conceptual model followed by logical and physical model.

Conceptual model involves with only identifying entities and relationship between. Fields / Attributes are not planned at this stage. It's just an identifying stage but not in detail.

Logical model involves in actually identifying the attributes, primary keys, many-to-many relationships etc of the entity. In short it's the complete detail planning of what actually has to be implemented.

Physical model is where you develop your actual structure tables, fields, primary keys, foreign leys etc. You can say it's the actual implementation of the project.

> *Note: - To Design conceptual and logical model mostly VISIO is used and some company combine this both model in one time. So you will not be able to distinguish between both models.*

(DB)What is Data purging?

You can also call this as data cleaning. After you have designed your data warehouse and started importing data, there is always a possibility you can get in lot of junk data. For example you have some rows which have NULL and spaces, so you can run a routine which can delete these kinds of records. So this cleaning process is called as "Data Purging".

What is Analysis Services?

Analysis Services (previously known as OLAP Services) was designed to draw reports from data contained in a "Data Warehouses"." Data Warehouses" do not have typical relational data structure (fully normalized way), but rather have snowflake or star schema (refer star schema in the previous sections).

The data in a data warehouse is processed using online analytical processing (OLAP) technology. Unlike relational technology, which derives results by reading and joining data when the query is issued, OLAP is optimized to navigate the summary data to quickly return results. As we are not going through any joins (because data is in denormalized form) SQL queries are executed faster and in more optimized way.

(DB)What are CUBES?

As said in previous question analysis services do not work on relation tables, but rather use "CUBES". Cubes have two important attributes dimensions and measures. Dimensions are data like Customer type, country name and product type. While measures are quantitative data like dollars, meters and weight. Aggregates derived from original data are stored in cubes.

(DB)What are the primary ways to store data in OLAP?

There are primary three ways in which we store information in OLAP:-

MOLAP

Multidimensional OLAP (MOLAP) stores dimension and fact data in a persistent data store using compressed indexes. Aggregates are stored to facilitate fast data access. MOLAP query engines are usually proprietary and optimized for the storage format used by the MOLAP data store. MOLAP offers faster query processing than ROLAP and usually requires less storage. However, it doesn't scale as well and requires a separate database for storage.

ROLAP

Relational OLAP (ROLAP) stores aggregates in relational database tables. ROLAP use of the relational databases allows it to take advantage of existing database resources, plus it allows ROLAP applications to scale well. However, ROLAP's use of tables to store aggregates usually requires more disk storage than MOLAP, and it is generally not as fast.

HOLAP

As its name suggests, hybrid OLAP (HOLAP) is a cross between MOLAP and ROLAP. Like ROLAP, HOLAP leaves the primary data stored in the source database. Like MOLAP, HOLAP stores aggregates in a persistent data store that's separate from the primary relational database. This mix allows HOLAP to offer the advantages of both MOLAP and ROLAP. However, unlike MOLAP and ROLAP, which follow well-defined standards, HOLAP has no uniform implementation.

(DB)What is META DATA information in Data warehousing projects?

META DATA is data about data. Well that's not an enough definition for interviews we need something more than that to tell the interviewer. It's the complete documentation of a data warehouse project. From perspective of SQL Server all Meta data is stored in Microsoft repository. It's all about way the structure is of data ware house, OLAP, DTS packages.

Just to summarize some elements of data warehouse Meta data are as follows:-

√ Source specifications — such as repositories, source schemas etc.

√ Source descriptive information — such as ownership descriptions, updates frequencies, legal limitations, access methods etc.

√ Process information — such as job schedules, extraction code.

√ Data acquisition information — such as data transmission scheduling, results and file usage.

√ Dimension table management — such as definitions of dimensions, surrogate key.

√ Transformation and aggregation — such as data enhancement and mapping, DBMS load scripts, aggregate definitions &c.

√ DMBS system table contents,

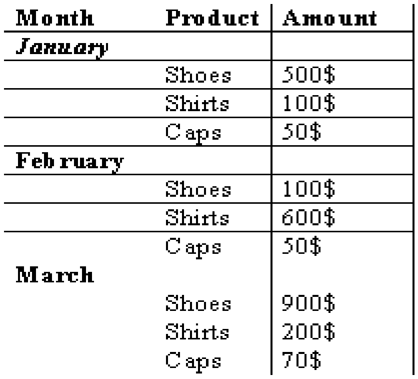
√ descriptions for columns

√ network security data

All Meta data is stored in system tables MSDB. META data can be accessed using repository API, DSO (Decision Support Objects).

(DB)What is multi-dimensional analysis?

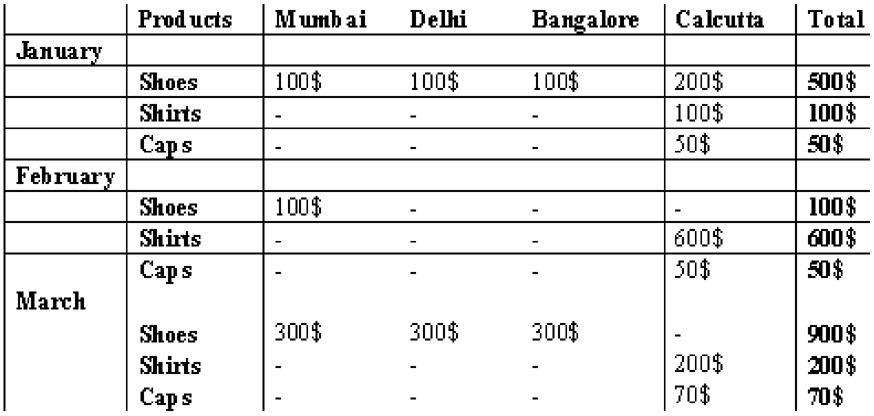
Multi-dimensional is looking data from different dimensions. For example we can look at a simple sale of a product month wise.



Month	Product	Amount
January		
	Shoes	500$
	Shirts	100$
	Caps	50$
February		
	Shoes	100$
	Shirts	600$
	Caps	50$
March		
	Shoes	900$
	Shirts	200$
	Caps	70$

Figure 8.12 : - Single Dimension view.

But let’s add one more dimension “Location” wise.



	Products	Mumbai	Delhi	Bangalore	Calcutta	Total
January						
	Shoes	100$	100$	100$	200$	**500$**
	Shirts	-	-	-	100$	**100$**
	Caps	-	-	-	50$	**50$**
February						
	Shoes	100$	-	-	-	**100$**
	Shirts	-	-	-	600$	**600$**
	Caps	-	-	-	50$	**50$**
March						
	Shoes	300$	300$	300$	-	**900$**
	Shirts	-	-	-	200$	**200$**
	Caps	-	-	-	70$	**70$**

Figure 8.13 : - Multi-Dimension View

The above table gives a three dimension view; you can have more dimensions according to your depth of analysis. Like from the above multi-dimension view I am able to predict that “Calcutta” is the only place where “Shirts” and “Caps” are selling, other metros do not show any sales for this product.

(DB)What is MDX?

MDX stands for multi-dimensional expressions. When it comes to viewing data from multiple dimensions SQL lacks many functionalities, there's where MDX queries are useful. MDX queries are fired against OLAP data bases. SQL is good for transactional databases (OLTP databases), but when it comes to analysis queries MDX stands the top.

> *Note: - If you are planning for data warehousing position using SQL Server 2005, MDX will be the favorite of the interviewers. MDX itself is such a huge and beautiful beast that we cannot cover in this small book. I will suggest at least try to grab some basic syntaxes of MDX like select before going to interview.*

(DB)How did you plan your Data ware house project?

> *Note: - This question will come up if the interviewer wants to test that had you really worked on any data warehouse project. Second if he is looking for a project manager or team lead position.*

Below are the different stages in Data warehousing project:-

√ System Requirement Gathering

This is what every traditional project follows and data warehousing is no different. What exactly is this complete project about? What is the client expecting? Do they have existing data base which they want to data warehouse or do we have to collect from lot of places. If we have to extract from lot of different sources, what are they and how many are they?. For instance you can have customer who will say this is the database now data warehouse it. Or customer can say consolidate data from EXCEL, ORACLE, SQL Server, CSV files etc etc. So if more the disparate systems more are the complications. Requirement gathering clears all these things and gives a good road map for the project ahead.

> *Note: - Many data warehouse projects take requirement gathering for granted. But I am sure when customer will come up during execution with, I want that (Sales by month) and also that (consolidate data from those 20 excels) and that (prepare those extra two reports) and that (migrate that database)…. and the project goes there (programmer work over time) and then there (project goes over budget) and then (Client looses interest)…. Somewhere (software company goes under loss).*

√ Selecting Tool.

Once you are ok with requirement its time to select which tools can do good work for you. This book only focuses on SQL Server 2005, but in reality there are many tools for data warehousing. Probably SQL Server 2005 will sometimes not fit your project requirement and you would like to opt for something else.

√ Data Modeling and design

This where the actual designing takes place. You do conceptual and logical designing of your database, star schema design.

√ ETL Process

This forms the major part for any data warehouse project. Refer previous section to see what an ETL process is. ETL is the execution phase for a data warehouse project. This is the place where you will define your mappings, create DTS packages, define work flow, write scripts etc. Major issue when we do ETL process is about performance which should be considered while executing this process.

> *Note: - Refer "Integration Services" for how to do the ETL process using SQL Server 2005.*

√ OLAP Cube Design

This is the place where you define your CUBES, DIMENSIONS on the data warehouse database which was loaded by the ETL process. CUBES and DIMENSIONS are done by using the requirement specification. For example you see that customer wants a report "Sales Per month" so he can define the CUBES and DIMENSIONS which later will be absorbed by the front end for viewing it to the end user.

√ Front End Development

Once all your CUBES and DIMENSIONS are defined you need to present it to the user. You can build your front ends for the end user using C#, ASP.NET, VB.NET any language which has the ability to consume the CUBES and DIMENSIONS. Front end stands on top of CUBES and DIMENSION and delivers the report to the end users. With out any front end the data warehouse will be of no use form user's perspective.

√ Performance Tuning

Many projects tend to overlook this process. But just imagine a poor user sitting to view "Yearly Sales" for 10 minutes….frustrating no. There are three sections where you can really look why your data warehouse is performing slow:-

- While data is loading in database “ETL” process.

  This is probably the major area where you can optimize your database. The best is to look in to DTS packages and see if you can make it better to optimize speed.

- OLAP CUBES and DIMENSIONS.

  CUBES and DIMENSIONS are something which will be executed against the data warehouse. You can look in to the queries and see if some optimization can be done.

- Front end code.

  Front end are mostly coded by programmers and this can be a major bottle neck for optimization. So you can probably look for loops and you also see if the front end is running too far away from the CUBES.

√ User Acceptance Test (UAT)

UAT means saying to the customer “Is this product ok with you?”. It’s a testing phase which can be done either by the customer (and mostly done by the customer) or by your own internal testing department to ensure that its matches with the customer requirement which was gathered during the requirement phase.

√ Rolling out to Production

Once the customer has approved your UAT, its time to roll out the data ware house in production so that customer can get the benefit of it.

√ Production Maintenance

I know the most boring aspect from programmer’s point of view, but the most profitable for an IT company point of view. In data warehousing this will mainly involve doing back ups, optimizing the system and removing any bugs. This can also include any enhancements if the customer wants it.

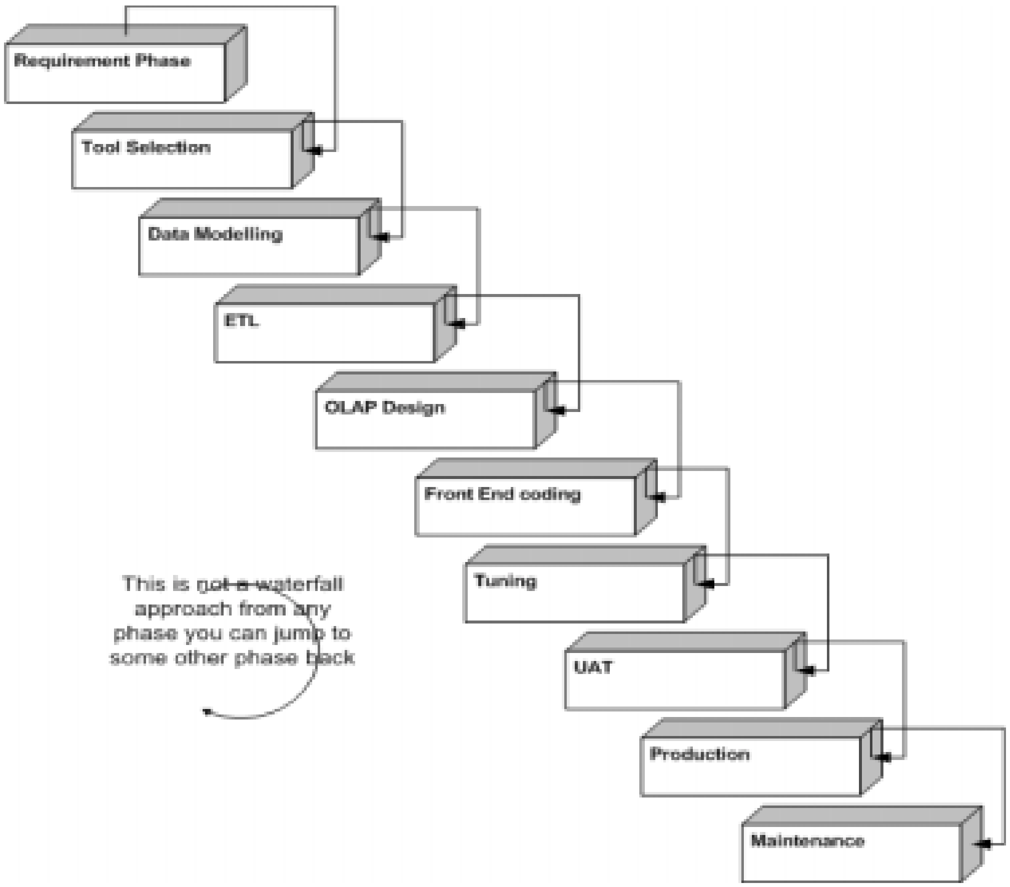


Figure 8.14 : - Data ware house project life cycle

What are different deliverables according to phases?

> *Note: - Deliverables means what documents you will submit during each phase. For instance Source code is deliverable for execution phase, Use Case Documents or UML documents are a deliverable for requirement phase. In short what will you give to client during each phase?.*

Following are the deliverables according to phases:-

- √ Requirement phase: - System Requirement documents, Project management plan, Resource allocation plan, Quality management document, Test plans and Number of reports the customer is looking at. I know many people from IT will start raising there eye balls hey do not mix the project management with requirement gathering. But that's a debatable issue I leave it to you guys if you want to further split it.
- √ Tool Selection: - POC (proof of concept) documents comparing each tool according to project requirement.

  Note: - POC means can we do?. For instance you have a requirement that, 2000 users at a time should be able to use your data warehouse. So you will probably write some sample code or read through documents to ensure that it does it.

- √ Data modeling: - Logical and Physical data model diagram. This can be ER diagrams or probably some format which the client understands.
- √ ETL: - DTS packages, Scripts and Metadata.
- √ OLAP Design:-Documents which show design of CUBES / DIMENSIONS and OLAP CUBE report.
- √ Front end coding: - Actual source code, Source code documentation and deployment documentation.
- √ Tuning: - This will be a performance tuning document. What performance level we are looking at and how will we achieve it or what steps will be taken to do so. It can also include what areas / reports are we targeting performance improvements.
- √ UAT: - This is normally the test plan and test case document. It can be a document which has steps how to create the test cases and expected results.
- √ Production: - In this phase normally the entire data warehouse project is the deliverable. But you can also have handover documents of the project, hardware, network settings, in short how is the environment setup.
- √ Maintenance: - This is an on going process and mainly has documents like error fixed, issues solved, within what time the issues should be solved and within what time it was solved.

(DB)Can you explain how analysis service works?

Note: - Ok guys this question is small but the answer is going to be massive. You are going to just summarize them but I am going to explain analysis services in detail, step by step *with a small project. For this complete explanation I am taking the old sample database of Microsoft "NorthWind".*

First and foremost ensure that your service is started so go to control panel, services and start the "Analysis Server "service.

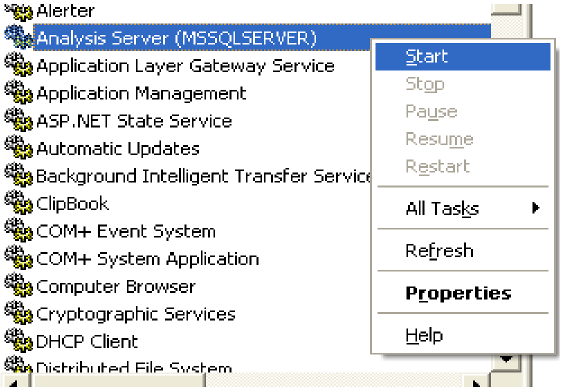


Figure 8.15 : - Start Analysis Server

As said before we are going to use "NorthWind" database for showing analysis server demo.

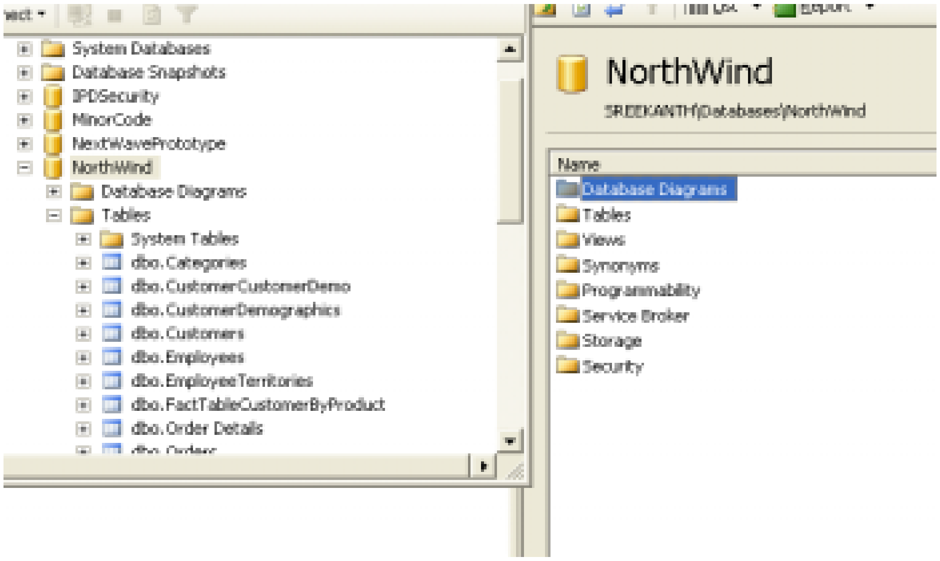


Figure 8.16 : - NorthWind Snapshot.

We are not going use all tables from “NorthWind”. Below are the only tables we will be operating using. Leaving the “FactTableCustomerByProduct” all other tables are self explanatory. Ok I know I have still not told you what we want to derive from this whole exercise. We will try to derive a report how much products are bought by which customer and how much products are sold according to which country. So I have created the fact table with three fields Customerid , Productid and the TotalProducts sold. All the data in Fact table I have loaded from “Orders” and “Order Details”. Means I have taken all customerid and productid with there respective totals and made entries in Fact table.

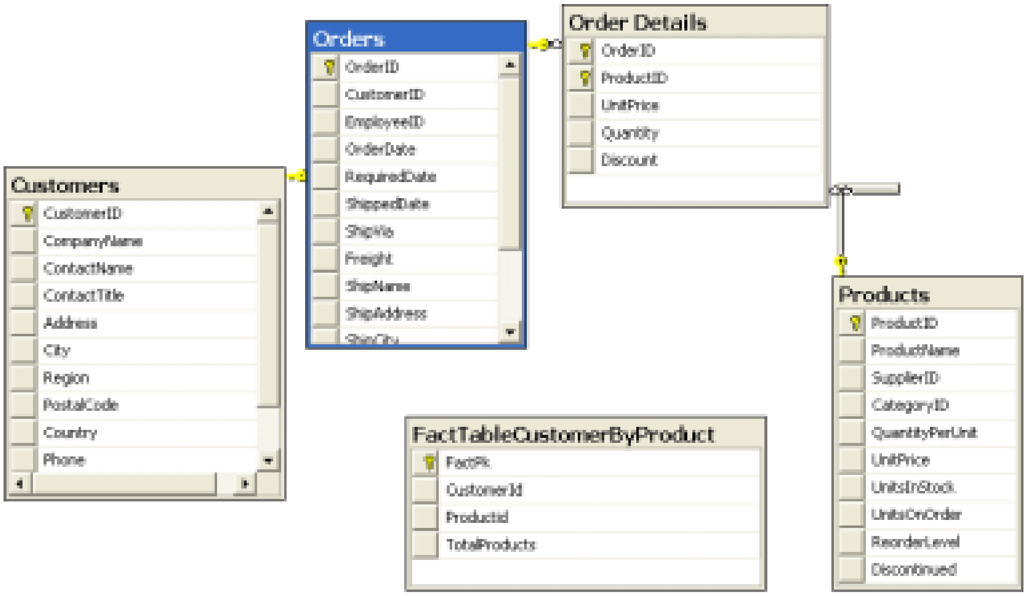


Figure 8.17: - Fact Table

Ok I have created my fact table and also populated using our ETL process. Now its time to use this fact table to do analysis.

So let's start our BI studio as shown in figure below.

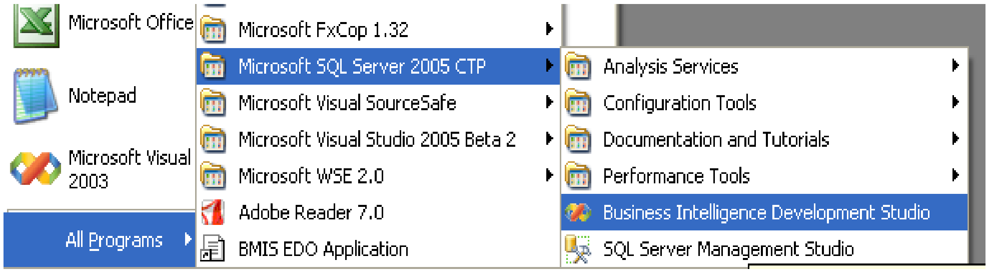
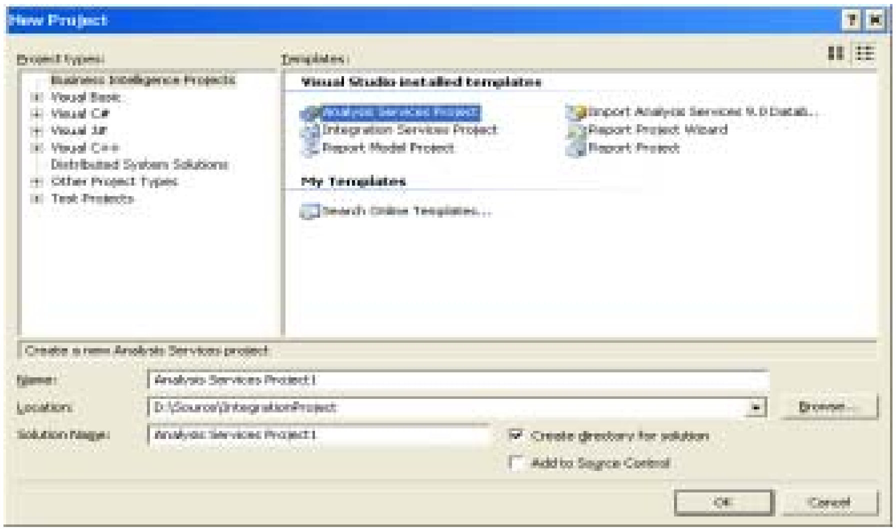

Microsoft Office
Notepad
Microsoft Visual 2003
All Programs
Microsoft FxCop 1.32
Microsoft SQL Server 2005 CTP
Microsoft Visual SourceSafe
Microsoft Visual Studio 2005 Beta 2
Microsoft WSE 2.0
Adobe Reader 7.0
BMIS EDO Application
Analysis Services
Configuration Tools
Documentation and Tutorials
Performance Tools
Business Intelligence Development Studio
SQL Server Management Studio

Figure 8.18 : - Start the Business Development Studio

Select "Analysis" project from the project types.



New Project
Project types:
Business Intelligence Projects
Visual Basic
Visual C#
Visual J#
Visual C++
Distributed System Solutions
Other Project Types
Test Projects
Templates:
Visual Studio installed templates
Analysis Services Project
Integration Services Project
Report Model Project
Import Analysis Services 9.0 Datab...
Report Project Wizard
Report Project
My Templates
Search Online Templates...
Browse...
OK
Cancel

Create a new Analysis Services project

Name: Analysis Services Project1

Location: D:\Source\IntegrationProject

Solution Name: Analysis Services Project1

☑ Create directory for solution

☐ Add to Source Control

Figure 8.19 : - Select Analysis Services Project

I have name the project as “AnalysisProject”. You can see the view of the solution explorer.

Data Sources :- This is where we will define our database and connection.

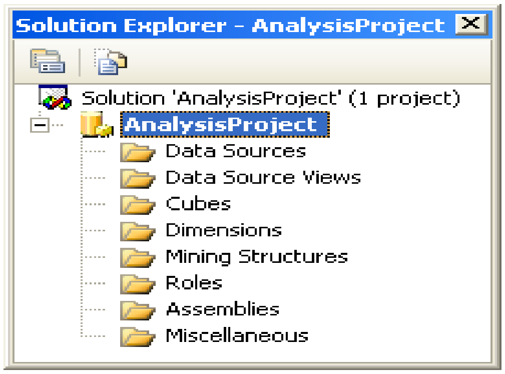


Figure 8.20 : - Solution Explorer

To add a new “data Source” right click and select “new Data Source”.

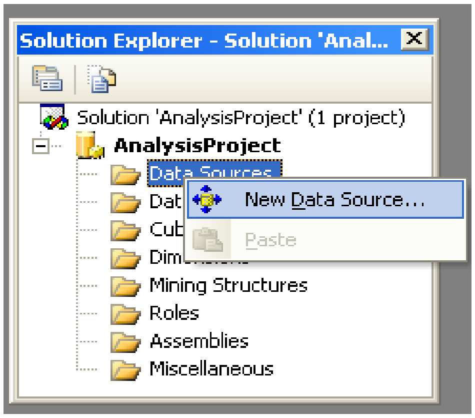
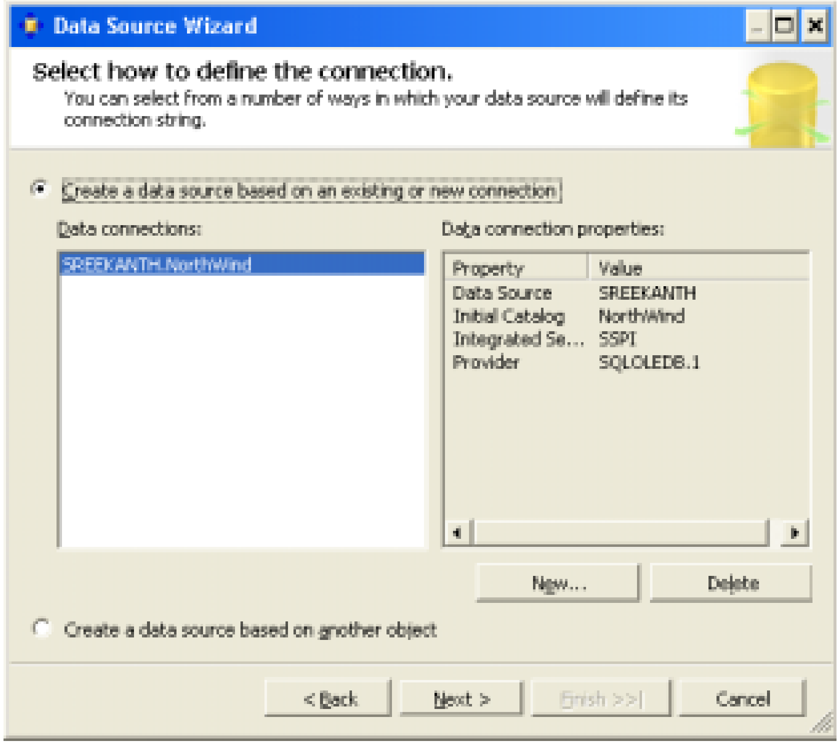


Figure 8.21 : - Create new data Source

After that Click next and you have to define the connection for the data source which you can do by clicking on the new button. Click next to complete the data source process.



Select how to define the connection.

You can select from a number of ways in which your data source will define its connection string.

● Create a data source based on an existing or new connection

Data connections:

SREEKANTH.NorthWind

Data connection properties:

Property	Value
Data Source	SREEKANTH
Initial Catalog	NorthWind
Integrated Se...	SSPI
Provider	SQLOLEDB.1

○ Create a data source based on another object

Figure 8.22 : - Define Data source connection details

After that its time to define view.

Data Source View: - It’s an abstraction view of data source. Data source is the complete database. It’s rare that we will need the complete database at any moment of time. So in “data source view” we can define which tables we want to operate on. Analysis server never operates on data source directly but it only speaks with the “Data Source” view.

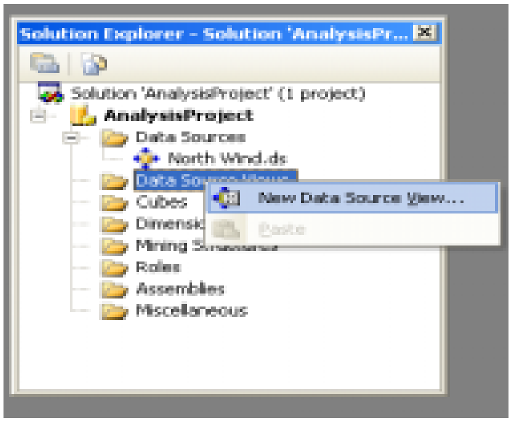
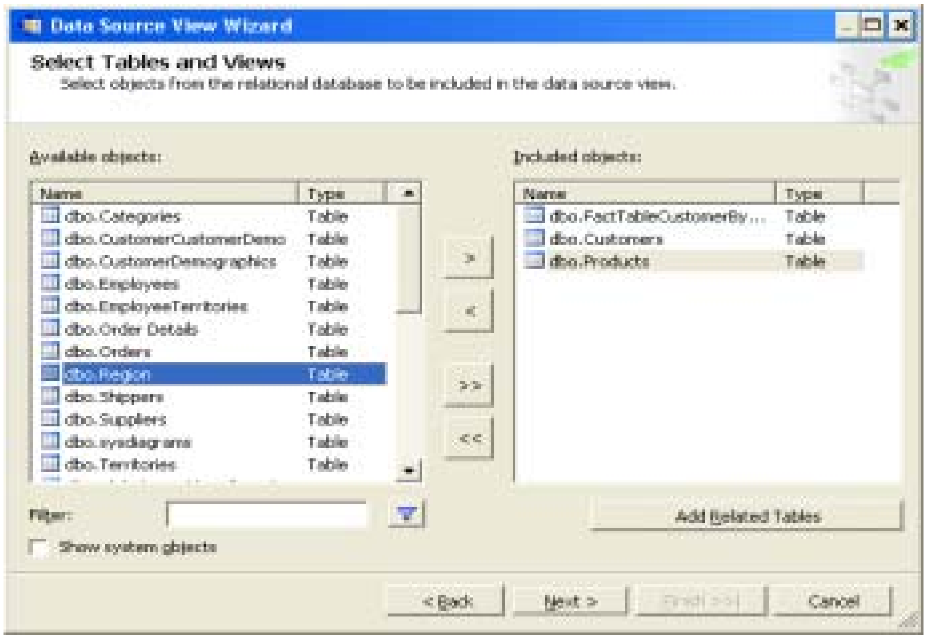


Figure 8.23 : - Create new Data source view

So here we will select only two tables "Customers", "Products" and the fact table.



Select Tables and Views

Select objects from the relational database to be included in the data source view.

Available objects:

Name	Type
dbo.Categories	Table
dbo.CustomerCustomerDemo	Table
dbo.CustomerDemographics	Table
dbo.Employees	Table
dbo.EmployeeTerritories	Table
dbo.Order Details	Table
dbo.Orders	Table
dbo.Region	Table
dbo.Shippers	Table
dbo.Suppliers	Table
dbo.sysdiagrams	Table
dbo.Territories	Table

Included objects:

Name	Type
dbo.FactTableCustomerBy...	Table
dbo.Customers	Table
dbo.Products	Table

Figure 8.24 : - Specify tables for the view

We had said previously fact table is a central table for dimension table. You can see products and customers table form the dimension table and fact table is the central point. Now drag and drop from the “Customerid” of fact table to the “Customerid” field of the customer table. Repeat the same for the “productid” table with the products table.

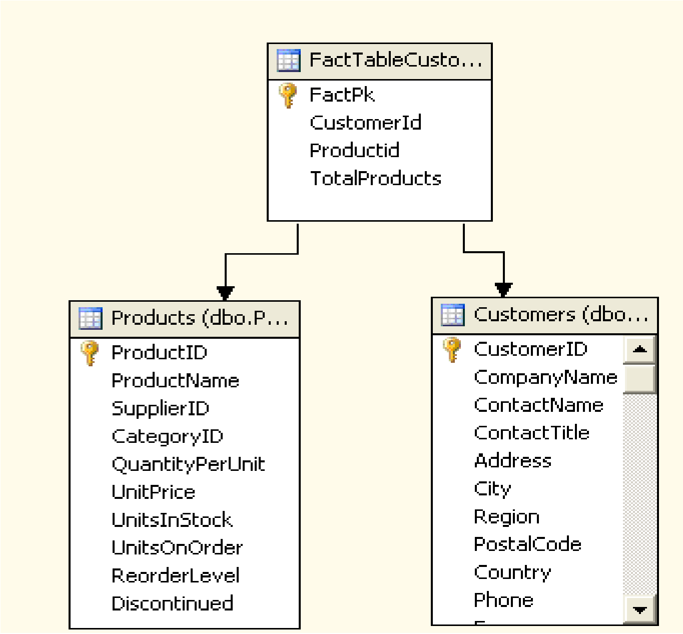
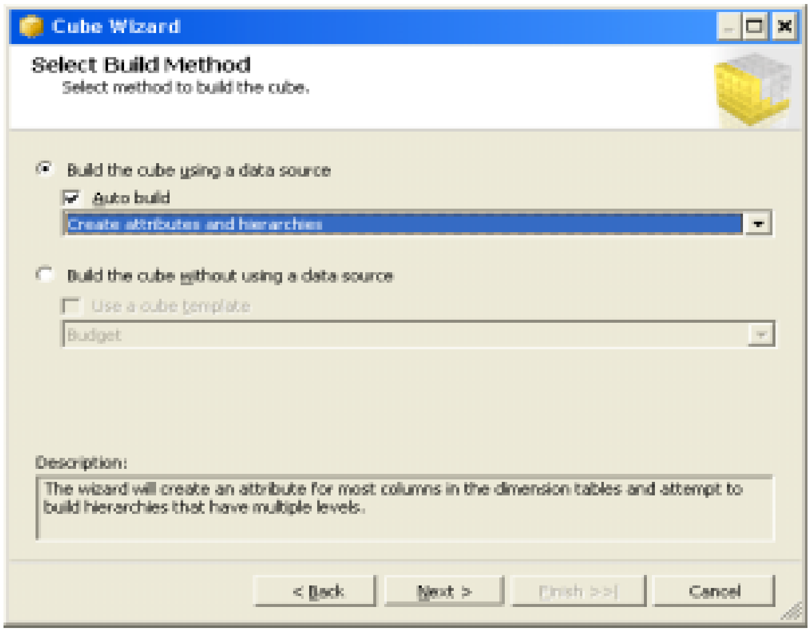


Figure 8.25 : - Final Data Source view

Check “Autobuild” as we are going to let the analysis service decide which tables he want to decide as “fact” and “Dimension” tables.



Cube Wizard

Select Build Method

Select method to build the cube.

● Build the cube using a data source

☑ Auto build

Create attributes and hierarchies

○ Build the cube without using a data source

□ Use a cube template

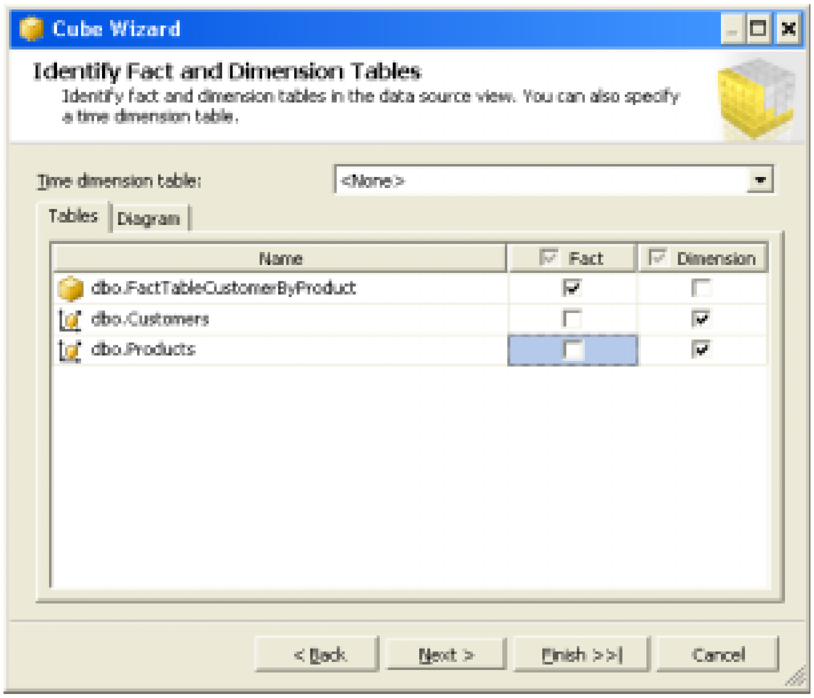
Budget

Description:

The wizard will create an attribute for most columns in the dimension tables and attempt to build hierarchies that have multiple levels.

Figure 8.26 : - Check Auto build

After that comes the most important step which are the fact tables and which are dimension tables. SQL Analysis services decides by itself, but we will change the values as shown in figure below.



Identify Fact and Dimension Tables

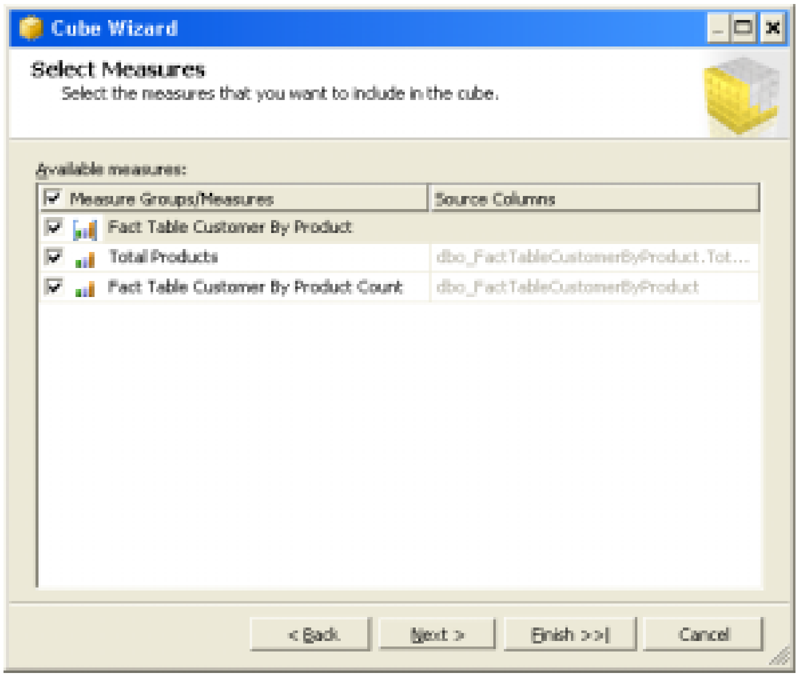
Identify fact and dimension tables in the data source view. You can also specify a time dimension table.

Time dimension table: <None>

Name	☑ Fact	☑ Dimension
dbo.FactTableCustomerByProduct	☑	□
dbo.Customers	□	☑
dbo.Products	□	☑

Figure 8.27 : - Specify Fact and Dimension Tables

This screen defines measures.



Cube Wizard

Select Measures

Select the measures that you want to include in the cube.

Available measures:

☑ Measure Groups/Measures	Source Columns
☑ Fact Table Customer By Product	
☑ Total Products	dbo_FactTableCustomerByProduct.Tot...
☑ Fact Table Customer By Product Count	dbo_FactTableCustomerByProduct

Figure 8.28 : - Specify measures

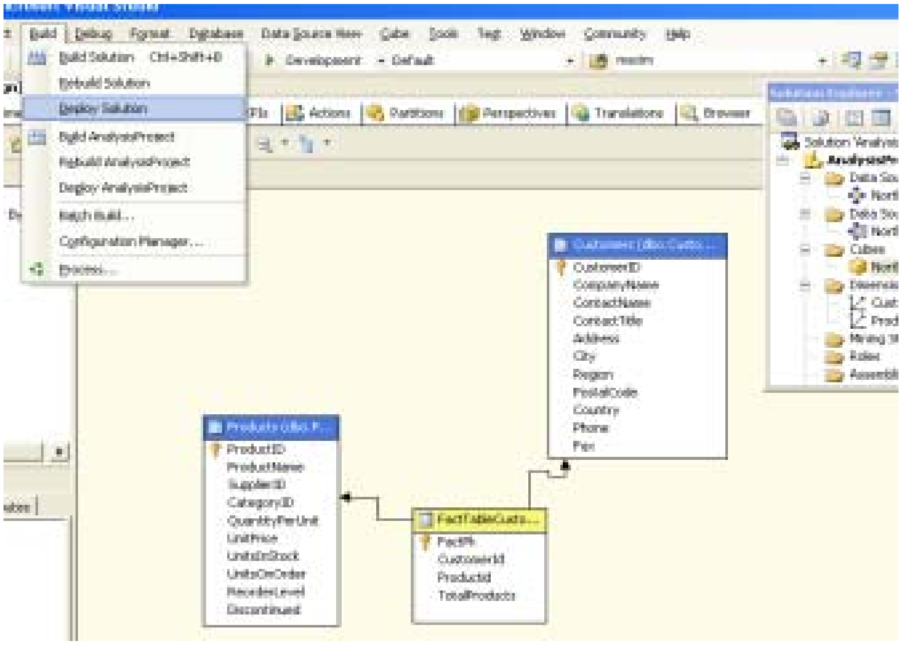
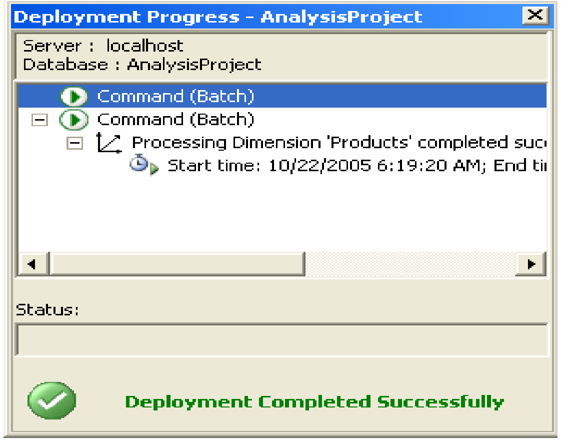


Figure 8.29 : - Deploy Solution



Deployment Progress - AnalysisProject

Server : localhost
Database : AnalysisProject

```
Command (Batch)
Command (Batch)
    Processing Dimension 'Products' completed suc
        Start time: 10/22/2005 6:19:20 AM; End ti
```

Status:

Deployment Completed Successfully

Figure 8.30 : - Deployment Successful

√ Cube Builder Works with the cube measures

√ Dimensions Works with the cube dimensions

√ Calculations Works with calculations for the cube

√ KPIs Works with Key Performance Indicators for the cube

√ Actions Works with cube actions

√ Partitions Works with cube partitions

√ Perspectives Works with views of the cube

√ Translations Defines optional transitions for the cube

√ Browser Enables you to browse the deployed cube

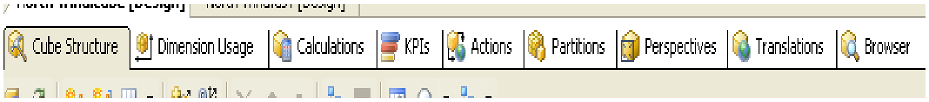
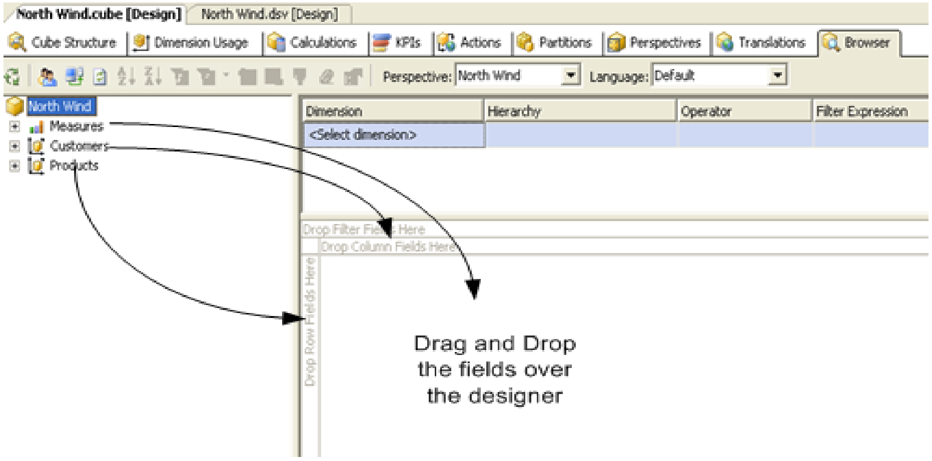


Figure 8.31: - View of top TAB

Once you are done with the complete process drag drop the fields as shown by the arrows below.

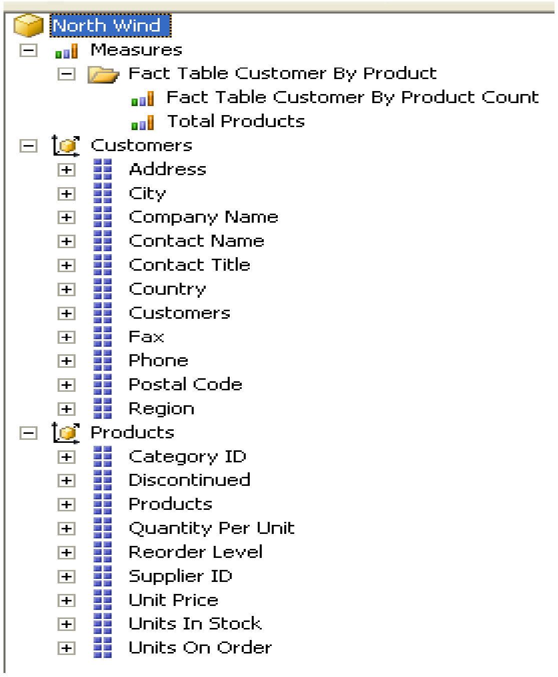


Perspective: North Wind
Language: Default

Dimension	Hierarchy	Operator	Filter Expression
<Select dimension>			

Drag and Drop
the fields over
the designer

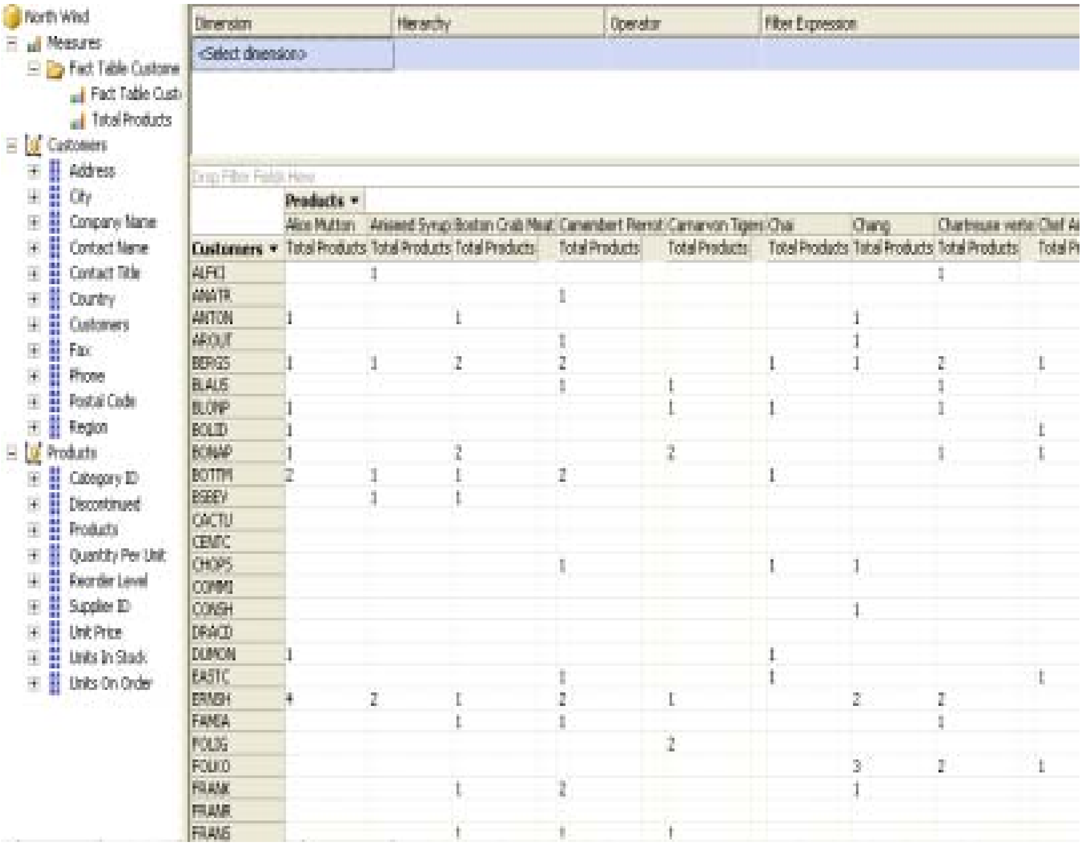
Figure 8.32: - Drag and Drop the fields over the designer



- North Wind
  - Measures
    - Fact Table Customer By Product
      - Fact Table Customer By Product Count
      - Total Products
  - Customers
    - Address
    - City
    - Company Name
    - Contact Name
    - Contact Title
    - Country
    - Customers
    - Fax
    - Phone
    - Postal Code
    - Region
  - Products
    - Category ID
    - Discontinued
    - Products
    - Quantity Per Unit
    - Reorder Level
    - Supplier ID
    - Unit Price
    - Units In Stock
    - Units On Order

Figure 8.33: - Final look of the CUBE

Once you have dragged dropped the fields you can see the wonderful information unzipped between which customer has bought how many products.



North Wind
- Measures
  - Fact Table Custome
    - Fact Table Cust
    - Total Products
- Customers
  - Address
  - City
  - Company Name
  - Contact Name
  - Contact Title
  - Country
  - Customers
  - Fax
  - Phone
  - Postal Code
  - Region
- Products
  - Category ID
  - Discontinued
  - Products
  - Quantity Per Unit
  - Reorder Level
  - Supplier ID
  - Unit Price
  - Units In Stock
  - Units On Order

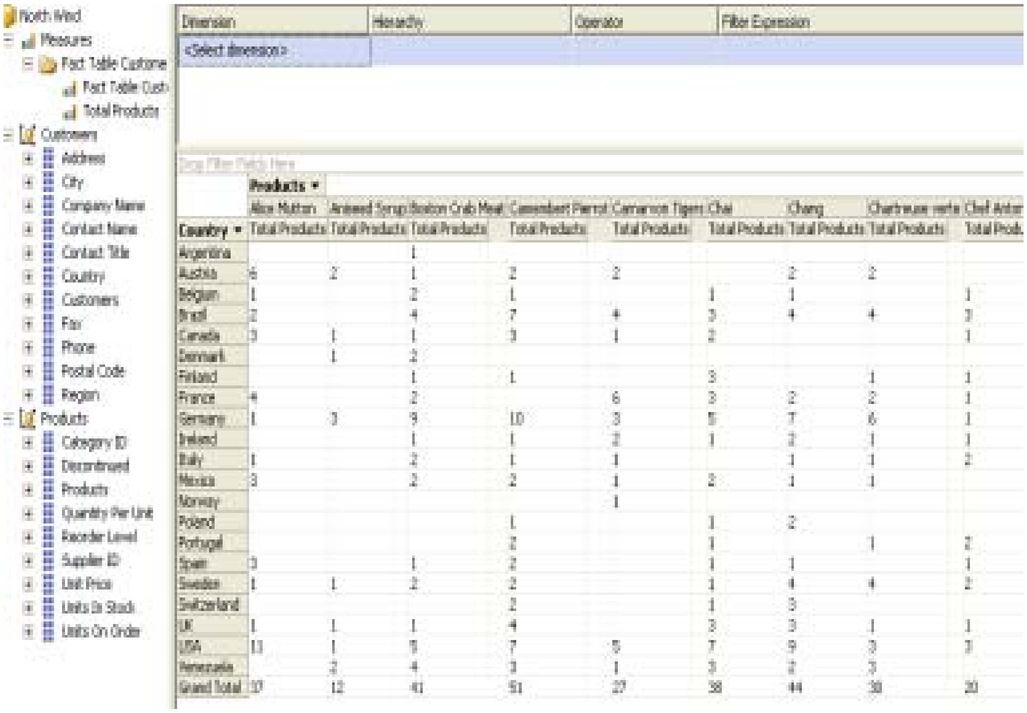
Dimension	Hierarchy	Operator	Filter Expression
<Select dimension>			

Drop Filter Fields Here

	Products ▾								
	Alice Mutton	Aniseed Syrup	Boston Crab Meat	Camembert Pierrot	Carnarvon Tigers	Chai	Chang	Chartreuse verte	Chef A
Customers ▾	Total Products	Total Products	Total Products	Total Products	Total Products	Total Products	Total Products	Total Products	Total Pr
ALFKI		1						1	
ANATR				1					
ANTON	1		1				1		
AROUT				1			1		
BERGS	1	1	2	2		1	1	2	1
BLAUS				1	1			1	
BLONP	1				1	1		1	
BOLID	1								1
BONAP	1		2		2			1	1
BOTTM	2	1	1	2		1			
BSBEV		1	1						
CACTU									
CENTC									
CHOPS				1		1	1		
COMMI									
CONSH							1		
DRACD									
DUMON	1					1			
EASTC				1		1			1
ERNSH	4	2	1	2	1		2	2	
FAMIA			1	1				1	
FOLIG					2				
FOLKO							3	2	1
FRANK			1	2			1		
FRANR									
FRANS			1	1	1				

Figure 8.34: - Product and Customer Report

This is the second report which says in which country I have sold how many products.



	Products ▾								
	Alice Mutton	Aniseed Syrup	Boston Crab Meat	Camembert Pierrot	Carnarvon Tigers	Chai	Chang	Chartreuse verte	Chef Antor
Country ▾	Total Products	Total Products	Total Products	Total Products	Total Products	Total Products	Total Products	Total Products	Total Prod
Argentina			1						
Austria	6	2	1	2	2		2	2	
Belgium	1		2	1		1	1		1
Brazil	2		4	7	4	3	4	4	3
Canada	3	1	1	3	1	2			1
Denmark		1	2						
Finland			1	1		3		1	1
France	4		2		6	3	2	2	1
Germany	1	3	9	10	3	5	7	6	1
Ireland			1	1	2	1	2	1	1
Italy	1		2	1	1		1	1	2
Mexico	3		2	2	1	2	1	1	
Norway					1				
Poland				1		1	2		
Portugal				2		1		1	2
Spain	3		1	2		1	1		1
Sweden	1	1	2	2		1	4	4	2
Switzerland				2		1	3		
UK	1	1	1	4		3	3	1	1
USA	11	1	5	7	5	7	9	3	3
Venezuela		2	4	3	1	3	2	3	
Grand Total	37	12	41	51	27	38	44	38	20

Figure 8.35: - Product Sales by country

Note: - I do not want my book to increase pages just because of images but sometimes the nature of the explanation demands it. Now you can just summarize to the interviewer from the above steps how you work with analysis services.

What are the different problems that “Data mining” can solve?

There are basically four problems that “Data mining” can solve:-

Analyzing Relationships

This term is also often called as “Link Analysis”. For instance one of the companies who sold adult products did an age survey of his customers. He found his entire products where bought by customers between age of 25 – 29. He further became suspicious that all of his customers must have kids around 2 to 5 years as that's the normal age of marriage. He analyzed further and found that maximum of his customers where married with kids. Now the company can also try selling kid products to the same customer as they will be interested in buying it, which can tremendously boost up his sales. Now here the link analysis was done between the "age" and "kids" decide a marketing strategy.

Choosing right Alternatives

If a business wants to make a decision between choices data mining can come to rescue. For example one the companies saw a major resignation wave in his company. So the HR decided to have a look at employee's joining date. They found that major of the resignations have come from employee's who have stayed in the company for more than 2 years and there where some resignation's from fresher. So the HR made decision to motivate the freshers rather than 2 years completed employee's to retain people. As HR thought it's easy to motivate freshers rather than old employees.

Prediction

Prediction is more about forecasting how the business will move ahead. For instance company has sold 1000 Shoe product items, if the company puts a discount on the product sales can go up to 2000.

Improving the current process.

Past data can be analyzed to view how we can improve the business process. For instance for past two years company has been distributing product "X" using plastic bags and product "Y" using paper bags. Company has observed closely that product "Y" sold the same amount as product "X" but has huge profits. Company further analyzed that major cost of product "X" was due to packaging the product in plastic bags. Now the company can improve the process by using the paper bags and bringing down the cost and thus increasing profits.

What are different stages of "Data mining"?

Problem Definition.

This is the first step in "Data mining" define your metrics by which the model will be evaluated. For instance if it's a small travel company he would like to measure his model on number of tickets sold , but if it's a huge travel companies with lot of agents he would like to see it with number of tickets / Agent sold. If it's a different industry together like bank they would like to see actual amount of transactions done per day.

There can be several models which a company wants to look into. For instance in our previous travel company model, they would like to have the following metrics:-

- √ Ticket sold per day
- √ Number of Ticket sold per agent
- √ Number of ticket sold per airlines
- √ Number of refunds per month

So you should have the following check list:-

- √ What attribute you want to measure and predict?
- √ What type of relationship you want to explore? In our travel company example you would like to explore relationship between Number of tickets sold and Holiday patterns of a country.

Preprocessing and Transforming Data

This can also be called as loading and cleaning of data or to remove unnecessary information to simplify data. For example you will be getting data for title as "Mr.", "M.r.", "Miss", "Ms" etc ... Hmm can go worst if these data are maintained in numeric format "1", "2", "6" etc...This data needs to be cleaned for better results.

You also need to consolidate data from various sources like EXCEL, Delimited Text files; any other databases (ORACLE etc).

Microsoft SQL Server 2005 Integration Services (SSIS) contains tools which can be used for cleaning and consolidating from various services.

Note: - Data warehousing ETL process is a subset of this section.

Exploring Models

Data mining / Explore models means calculating the min and max values, look in to any serious deviations that are happening, and how is the data distrubuted. Once you see the data you can look in to if the data is flawed or not. For instance normal hours in a day is

24 and you see some data has more than 24 hours which is not logical. You can then look in to correcting the same.

Data Source View Designer in BI Development Studio contains tools which can let you analyze data.

Building Models

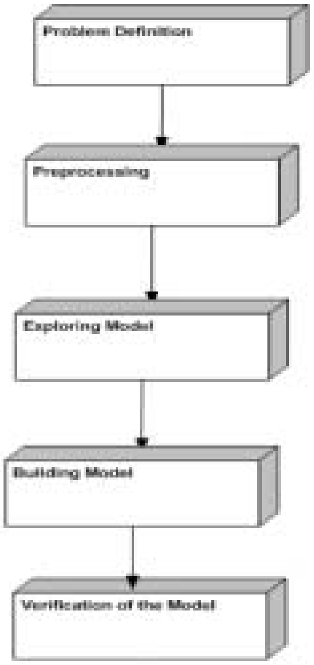
Data derived from Exploring models will help us to define and create a mining model. A model typically contains input columns, an identifying column, and a predictable column. You can then define these columns in a new model by using the Data Mining Extensions (DMX) language or the Data Mining Wizard in BI Development Studio.

After you define the structure of the mining model, you process it, populating the empty structure with the patterns that describe the model. This is known as training the model. Patterns are found by passing the original data through a mathematical algorithm. SQL Server 2005 contains a different algorithm for each type of model that you can build. You can use parameters to adjust each algorithm.

A mining model is defined by a data mining structure object, a data mining model object, and a data mining algorithm.

Verification of the models.

By using viewers in Data Mining Designer in BI Development Studio you can test / verify how well these models are performing. If you find you need any refining in the model you have to again iterate to the first step.



Note :- We can move from any of down process to upper process it's a iterative model rather than waterfall model. Did not make the arrow direction just to avoid confusion. So you can move from building model to problem definition.

Figure 8.36 : - Data mining life Cycle.

(DB)What is Discrete and Continuous data in Data mining world?

Discrete: - A data item that has a finite set of values. For example Male or Female.

Continuous: - This does not have finite set of value, but rather continuous value. For instance sales amount per month.

(DB)What is MODEL is Data mining world?

MODEL is extracting and understanding different patterns from a data. Once the patterns and trends of how data behaves are known we can derive a model from the same. Once these models are decided we can see how these models can be helpful for prediction / forecasting, analyzing trends, improving current process etc.

DB)How are models actually derived?

Twist: - What is Data Mining Algorithms?

Data mining Models are created using Data mining algorithm's. So to derive a model you apply Data mining algorithm on a set of data. Data mining algorithm then looks for specific trends and patterns and derives the model.

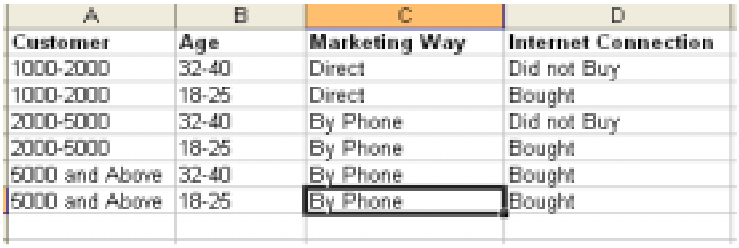
Note : - Now we will go through some algorithms which are used in "Data Mining" world. If you are looking out for pure "Data Mining" jobs, these basic question will be surely asked. Data mining algorithm is not Microsoft proprietary but is old math's which is been used by Microsoft SQL Server. The below section will look like we are moving away from SQL Server but trust me…if you are looking out for data mining jobs these questions can be turning point.

(DB)What is a Decision Tree Algorithm?

Note: - As we have seen in the first question that to derive a model we need algorithms. The further section will cover basic algorithms which will be asked during interviews.

"Decision Tree" is the most common method used in "data mining". In a decision tree structure leaves determine classification and the branches represent the reason of classifications.

For instance below is a sample data collected for an ISP provider who is in supplying "Home Internet Connection".



A	B	C	D
Customer	Age	Marketing Way	Internet Connection
1000-2000	32-40	Direct	Did not Buy
1000-2000	18-25	Direct	Bought
2000-5000	32-40	By Phone	Did not Buy
2000-5000	18-25	By Phone	Bought
5000 and Above	32-40	By Phone	Bought
5000 and Above	18-25	By Phone	Bought

Figure 8.37 : - Sample Data for Decision Tree

Based on the above data we have made the following decision tree. So you can see decision tree takes data and then start applying attribute comparison on every node recursively.

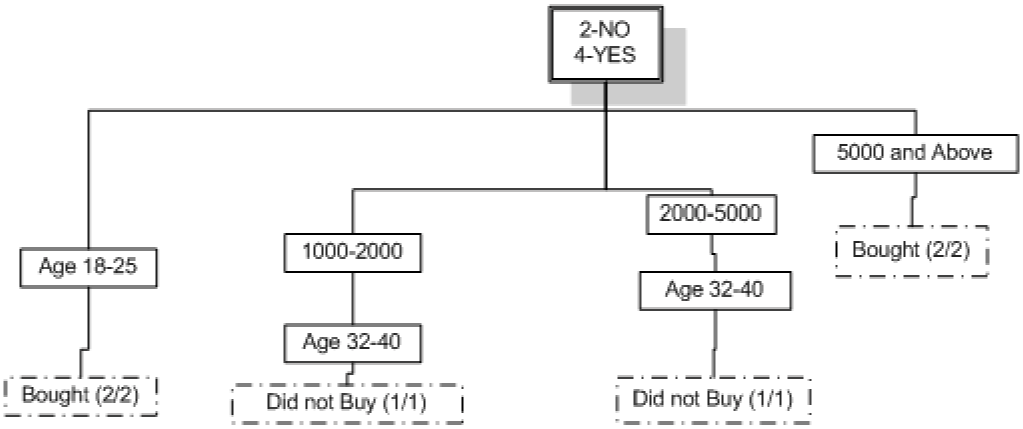


Figure 8.38 : - First Iteration Decision Tree

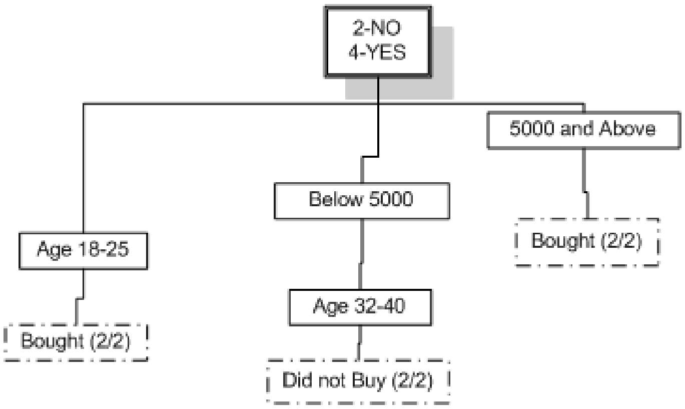


Figure 8.39 : - Conclusion from the Decision Tree

From the “Decision Tree” diagram we have concluded following predictions “-

- √ Age 18-25 always buys internet connection, irrelevant of income.
- √ Income drawers above 5000 always buy internet connection, irrelevant of age.

Using this data we have made predictions that if we market using the above criteria's we can make more "Internet Connection" sales.

So we have achieved two things from "Decision tree":-

Prediction

- √ If we market to age groups between 32-40 and income below 5000 we will not have decent sales.
- √ If we target customer with Age group 18-25 we will have good sales.
- √ All income drawers above 5000 will always have sales.

Classification

- √ Customer classification by Age.
- √ Customer classification depending on income amount.

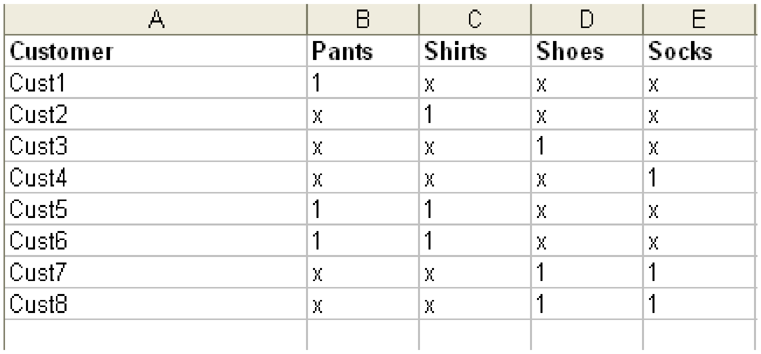
(DB)Can decision tree be implemented using SQL?

With SQL you can only look through one angle point of view. But with decision tree as you traverse recursively through all data you can have multi-dimensional view. For example give above using SQL you could have made the conclusion that age 18-25 has 100 % sales result. But "If we market to age groups between 32-40 and income below 5000 we will not have decent sales." Probably a SQL can not do (we have to be too heuristic).

(DB)What is Naïve Bayes Algorithm?

"Bayes' theorem can be used to calculate the probability that a certain event will occur or that a certain proposition is true, given that we already know a related piece of information."

Ok that's a difficult things to understand lets make it simple. Let's take for instance the sample data down.



A	B	C	D	E
Customer	**Pants**	**Shirts**	**Shoes**	**Socks**
Cust1	1	x	x	x
Cust2	x	1	x	x
Cust3	x	x	1	x
Cust4	x	x	x	1
Cust5	1	1	x	x
Cust6	1	1	x	x
Cust7	x	x	1	1
Cust8	x	x	1	1

Figure 8.40 : - Bayesian Sample Data

If you look at the sample we can say that 80 % of time customer who buy pants also buys shirts.

P (Shirt | Pants) = 0.8

Customer who buys shirts are more than who buys pants , we can say 1 of every 10 customer will only buy shirts and 1 of every 100 customer will buy only pants.

P (Shirts) = 0.1

P (Pants) = 0.01

Now suppose we a customer comes to buys pants how much is the probability he will buy a shirt and vice-versa. According to theorem:-

Probability of buying shirt if bought pants = 0.8-0.01 / 0.1=7.9

Probability of buying pants if bought shirts = 0.8-0.1 / 0.01=70

So you can see if the customer is buying shirts there is a huge probability that he will buy pants also. So you can see naïve bayes algorithm is use for predicting depending on existing data.

(DB)Explain clustering algorithm?

"Cluster is a collection of objects which have similarity between then and are dissimilar from objects different clusters."

Following are the ways a clustering technique works:-

- √ Exclusive: A member belongs to only one cluster.
- √ Overlapping: A member can belong to more than one cluster.
- √ Probabilistic: A member can belong to every cluster with a certain amount of probability.
- √ Hierarchical: Members are divided into hierarchies, which are sub-divided into clusters at a lower level.

(DB)Explain in detail Neural Networks?

Humans always wanted to beat god and neural networks is one of the step towards that. Neural network was introduced to mimic the sharpness of how brain works. Whenever human see something, any object for instance an animal. Many inputs are sent to his brains for example it has four legs, big horns, long tail etc etc. With these inputs your brain concludes that it's an animal. From childhood your brain has been trained to understand these inputs and your brain concludes output depending on that. This all happens because of those 1000 neurons which are working inside your brain inter-connected to decide the output.

That's what human tried to devise neural network. So now you must be thinking how it works.

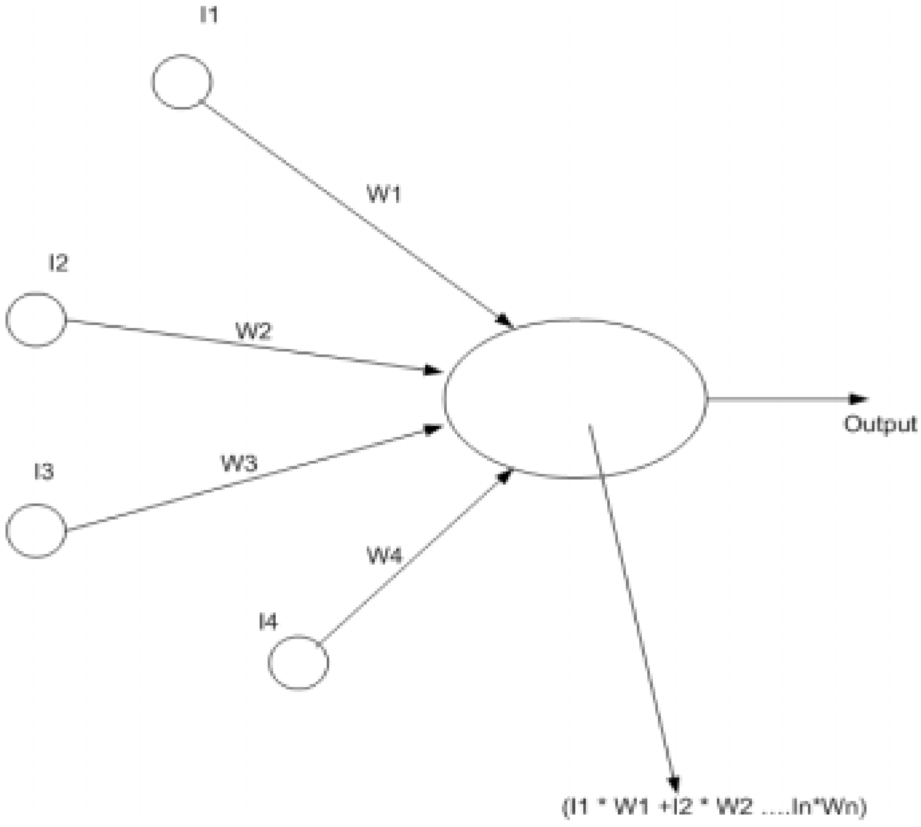


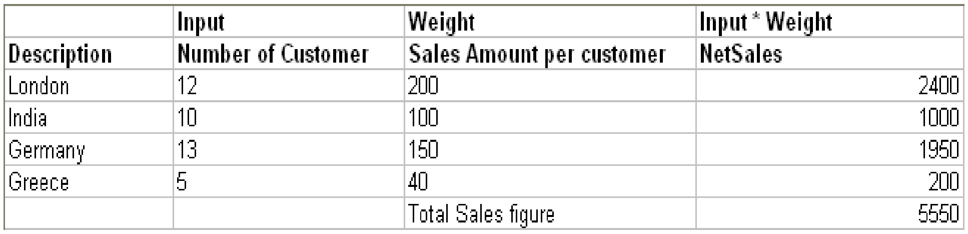
Figure 8.41 : - Artificial Neuron Model

Above is the figure which shows a neuron model. We have inputs (I1, I2 … IN) and for every input there are weights (W1, W2 …. WN) attached to it. The ellipse is the “NEURON”. Weights can have negative or positive values. Activation value is the summation and multiplication of all weights and inputs coming inside the nucleus.

Activation Value = I1 * W1 + I2 * W2+ I3 * W3+ I4 * W4…… IN * WN

There is threshold value specified in the neuron which evaluates to Boolean or some value, if the activation value exceeds the threshold value.

So probably feeding a customer sales records we can come out with an output is the sales department under profit or loss.



	Input	Weight	Input * Weight
Description	Number of Customer	Sales Amount per customer	NetSales
London	12	200	2400
India	10	100	1000
Germany	13	150	1950
Greece	5	40	200
		Total Sales figure	5550

Figure 8.42 : - Neural Network Data

For instance take the case of the top customer sales data. Below is the neural network defined for the above data.

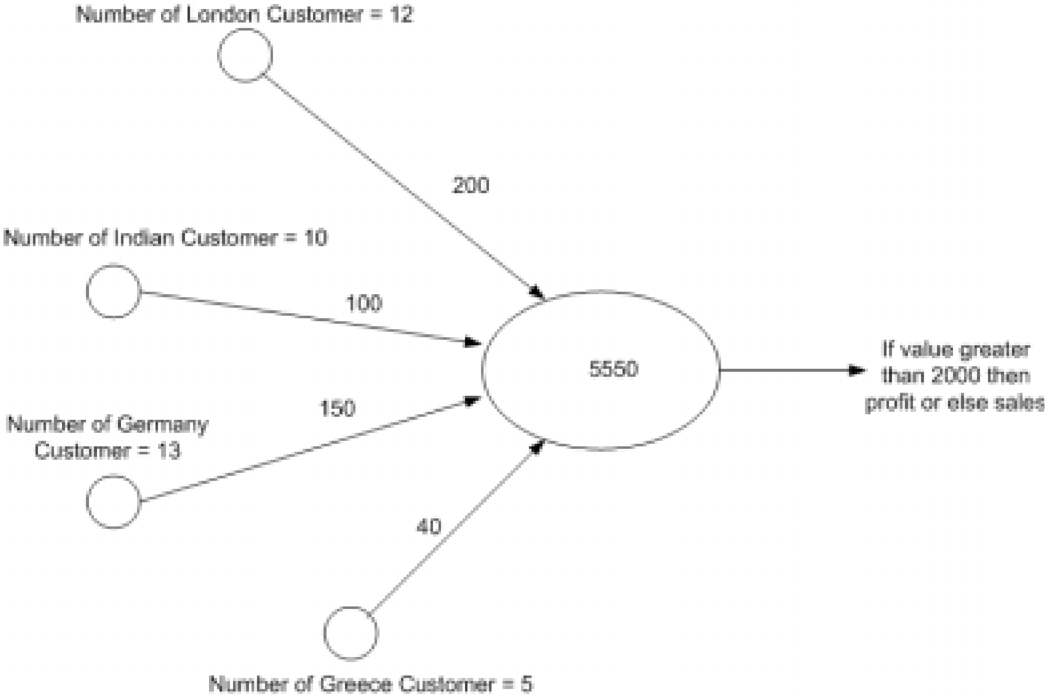

Number of London Customer = 12
200
Number of Indian Customer = 10
100
Number of Germany Customer = 13
150
5550
If value greater than 2000 then profit or else sales
40
Number of Greece Customer = 5

Figure 8.43: - Neural Network for Customer Sales Data

You can see neuron has calculated the total as 5550 and as it's greater than threshold 2000 we can say the company is under profit.

The above example was explained for simplification point of view. But in actual situation there can many neurons as shown in figure below. It's a complete hidden layer from the data miner perspective. He only looks in to inputs and outputs for that scenario.

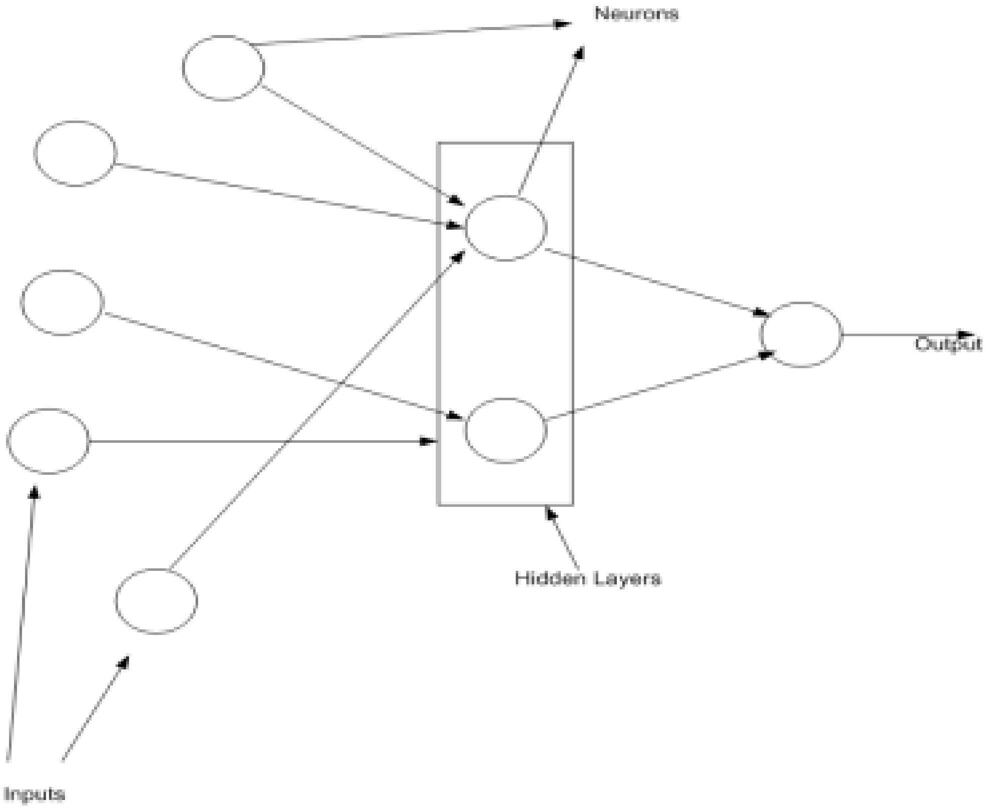


Figure 8.44: - Practical Neural Network

(DB)What is Back propagation in Neural Networks?

Back propagation helps you minimize error and optimize your network. For instance in our top example we get neuron summation as 80000000000, which is a weird figure (as you are expecting values between 0 to 6000 maximum). So you can always go back and look at whether you have some wrong input or weights. So the error is again Fed back to the neural network and the weights are adjusted accordingly. This is also called training the model.

(DB)What is Time Series algorithm in data mining?

The Microsoft Time Series algorithm allows you to analyze and forecast any time-based data, such as sales or inventory. So the data should be continuous and you should have some past data on which it can predict values.

(DB)Explain Association algorithm in Data mining?

Association algorithm tries to find relation ship between specific categories of data. In Association first it scans for unique values and then the frequency of values in each transaction is determined. For instance if lets say we have city master and transactional customer sales table. Association algorithm first find unique instance of all cities and then see how many city occurrences have occurred in the customer sales transactional table.

(DB)What is Sequence clustering algorithm?

Sequence clustering algorithm analyzes data that contains discrete-valued series. It looks for how the past data is transitioning and then makes future predictions. It's a hybrid of clustering and sequencing algorithm

> *Note: - UUUh I understand algorithm are dreaded level question and will never be asked for programmer level job, but guys looking for Data mining jobs these questions are basic. It's difficult to cover all algorithms existing in data mining world, as its complete area by itself. As been an interview question book I have covered algorithm which are absolutely essential from SQL Server point of view. Now we know the algorithms we can classify where they can be used. There are two important classifications in data mining world Prediction / Forecasting and grouping. So we will classify all algorithms which are shipped in SQL server in these two sections only.*

(DB)What are algorithms provided by Microsoft in SQL Server?

Predicting an attribute, for instance how much will be the product sales next year.

√ Microsoft Decision Trees Algorithm

√ Microsoft Naive Bayes Algorithm

√ Microsoft Clustering Algorithm

√ Microsoft Neural Network Algorithm

Predicting a continuous attribute, for example, to forecast next year's sales.

√ Microsoft Decision Trees Algorithm

√ Microsoft Time Series Algorithm

Predicting a sequence, for example, to perform a click stream analysis of a company's Web site.

√ Microsoft Sequence Clustering Algorithm

Finding groups of common items in transactions, for example, to use market basket analysis to suggest additional products to a customer for purchase.

√ Microsoft Association Algorithm

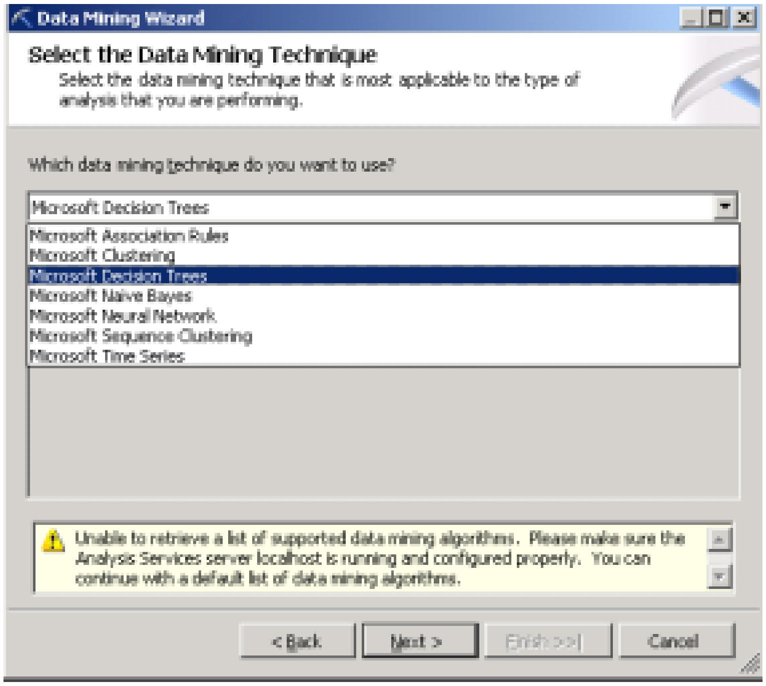
√ Microsoft Decision Trees Algorithm

Finding groups of similar items, for example, to segment demographic data into groups to better understand the relationships between attributes.

√ Microsoft Clustering Algorithm

√ Microsoft Sequence Clustering Algorithm

Why we went through all these concepts is when you create data mining model you have to specify one the algorithms. Below is the snapshot of all SQL Server existing algorithms.



Select the Data Mining Technique

Select the data mining technique that is most applicable to the type of analysis that you are performing.

Which data mining technique do you want to use?

Microsoft Decision Trees

- Microsoft Association Rules
- Microsoft Clustering
- Microsoft Decision Trees
- Microsoft Naive Bayes
- Microsoft Neural Network
- Microsoft Sequence Clustering
- Microsoft Time Series

Unable to retrieve a list of supported data mining algorithms. Please make sure the Analysis Services server localhost is running and configured properly. You can continue with a default list of data mining algorithms.

Figure 8.45: - Snapshot of the algorithms in SQL Server

Note: - During interviewing it's mostly the theory that counts and the way you present. For datamining I am not showing any thing practical as such probably will try to cover this thing in my second edition. But it's a advice please do try to run make a small project and see how these techniques are actually used.

(DB)How does data mining and data warehousing work together?

Twist: - What is the difference between data warehousing and data mining?

This question will be normally asked to get an insight how well you know the whole process of data mining and data warehousing. Many new developers tend to confuse data mining with warehousing (especially freshers). Below is the big picture which shows the relation between "data warehousing" and "data mining".

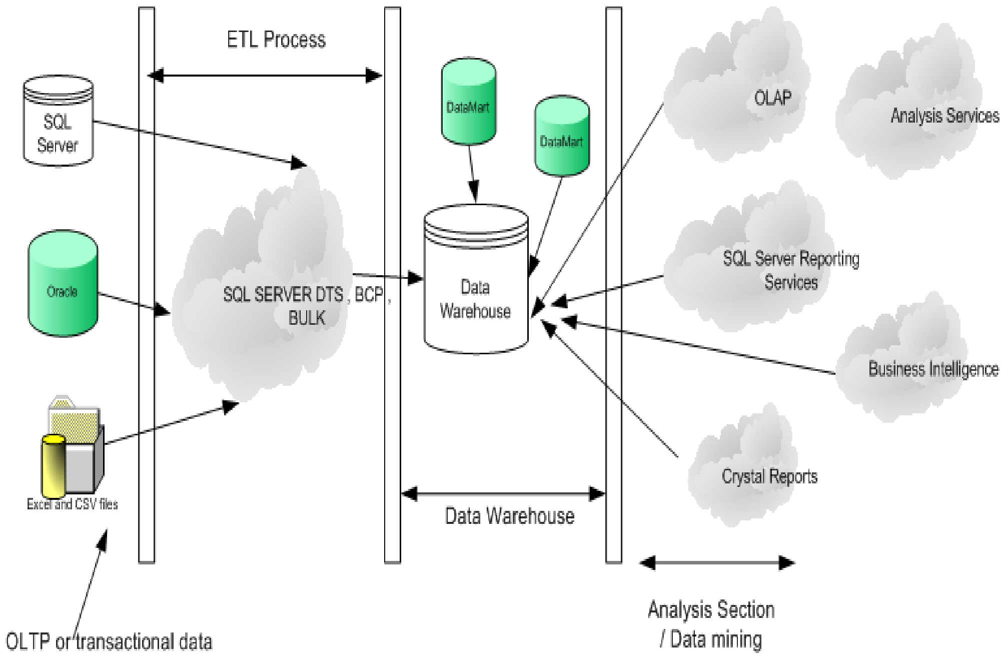


Figure 8.46 : - Data mining and Data Warehousing

Let's start from the most left hand side of the image. First section comes is the transaction database. This is the database in which you collect data. Next process is the ETL process. This section extracts data from the transactional database and sends to your data warehouse which is designed using STAR or SNOW FLAKE model. Finally when your data warehouse data is loaded in data warehouse, you can use SQL Server tools like OLAP, Analysis Services, BI, Crystal reports or reporting services to finally deliver the data to the end user.

> *Note: - Interviewer will always try goof you up saying why should not we run OLAP, Analysis Services, BI, Crystal reports or reporting services directly on the transactional data. That is because transactional database are in complete normalized form which can make the data mining process complete slow. By doing data warehousing we denormalize the data which makes the data mining process more efficient.*

What is XMLA?

XML for Analysis (XMLA) is fundamentally based on web services and SOAP. Microsoft SQL Server 2005 Analysis Services uses XMLA to handle all client application communications to Analysis Services.

XML for Analysis (XMLA) is a Simple Object Access Protocol (SOAP)-based XML protocol, designed specifically for universal data access to any standard multidimensional data source residing on the Web. XMLA also eliminates the need to deploy a client component that exposes Component Object Model (COM) or Microsoft .NET Framework.

What is Discover and Execute in XMLA?

The XML for Analysis open standard describes two generally accessible methods: Discover and Execute. These methods use the loosely-coupled client and server architecture supported by XML to handle incoming and outgoing information on an instance of SSAS.

The Discover method obtains information and metadata from a Web service. This information can include a list of available data sources, as well as information about any of the data source providers. Properties define and shape the data that is obtained from a data source. The Discover method is a common method for defining the many types of information a client application may require from data sources on Analysis Services instances. The properties and the generic interface provide extensibility without requiring you to rewrite existing functions in a client application.

The Execute method allows applications to run provider-specific commands against XML for Analysis data sources.

9. Integration Services/DTS

Note: - We had seen some question on DTS in the previous chapter "Data Warehousing". But in order to just make complete justice with this topic I have included them in integration services.

What is Integration Services import / export wizard?

Note :- What is DTS import / Export Wizard ?

Note: - Try to do this practically as it can be useful if the interviewer wants to visualize the whole stuff.

DTS import / export wizard lets us import and export from external data sources. There are seven steps which you can just go through of how to use the wizard.

You can find DTS import and export wizard as shown below.

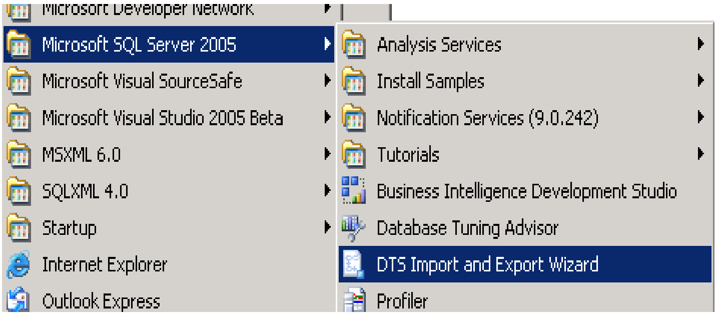
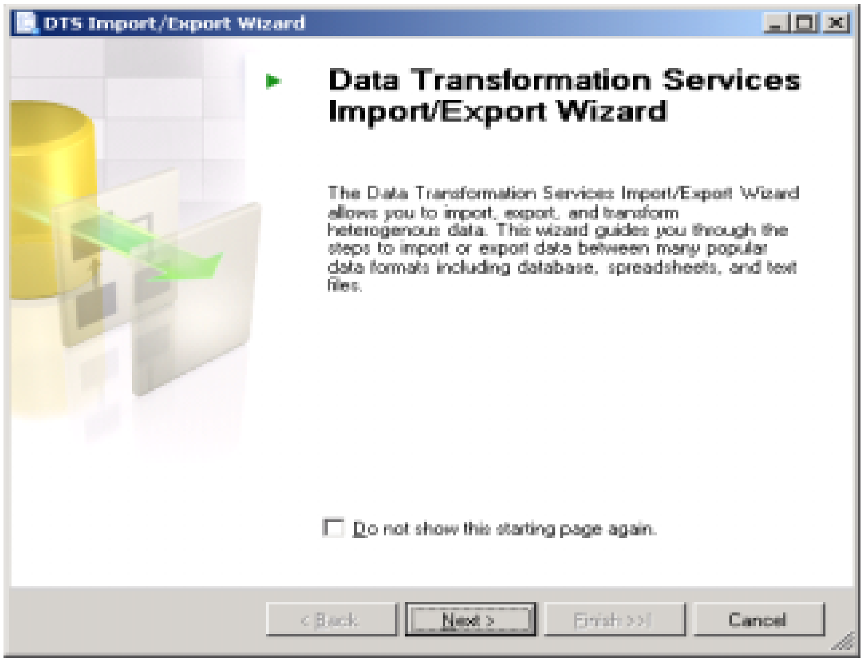

Microsoft Developer Network
Microsoft SQL Server 2005
Microsoft Visual SourceSafe
Microsoft Visual Studio 2005 Beta
MSXML 6.0
SQLXML 4.0
Startup
Internet Explorer
Outlook Express
Analysis Services
Install Samples
Notification Services (9.0.242)
Tutorials
Business Intelligence Development Studio
Database Tuning Advisor
DTS Import and Export Wizard
Profiler

Figure 9.1 : - Location of DTS Import and Export Wizard

You will be popped with a screen as below click "Next"



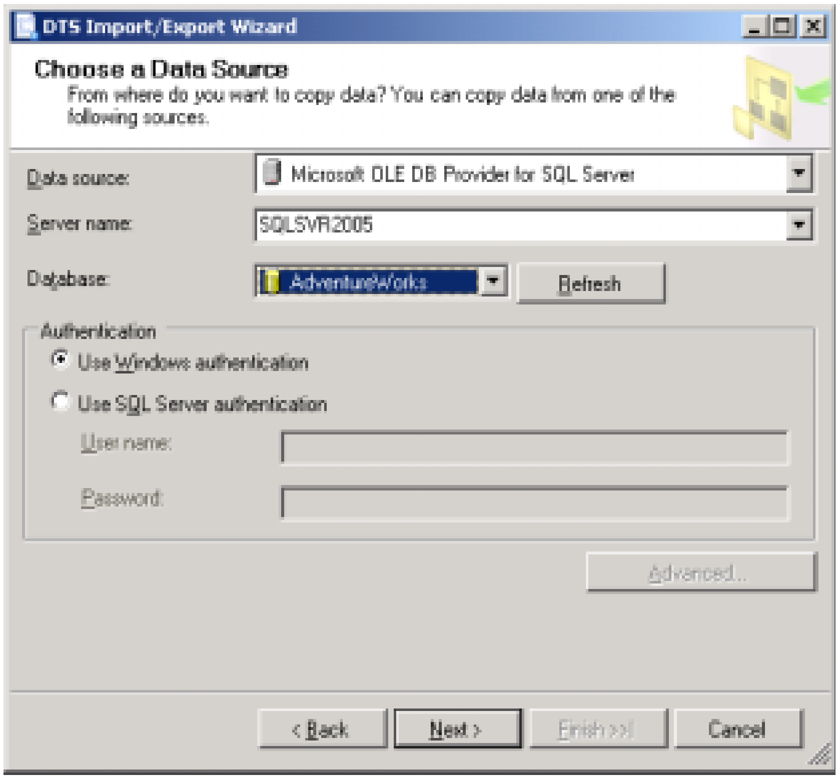
Data Transformation Services Import/Export Wizard

The Data Transformation Services Import/Export Wizard allows you to import, export, and transform heterogenous data. This wizard guides you through the steps to import or export data between many popular data formats including database, spreadsheets, and text files.

□ Do not show this starting page again.

Figure 9.2 : - Import / Export Wizard

Next step is to specify from which source you want to copy data. You have to specify the Data source name and server name. For understanding purpose we are going to move data between "AdventureWork" databases. I have created a dummy table called as "SalesPersonDummy" which has the same structure as that of "SalesPerson" table. But the only difference is that "SalesPersonDummy" does not have data.



Choose a Data Source

From where do you want to copy data? You can copy data from one of the following sources.

Data source: Microsoft OLE DB Provider for SQL Server

Server name: SQLSVR2005

Database: AdventureWorks

Authentication

● Use Windows authentication

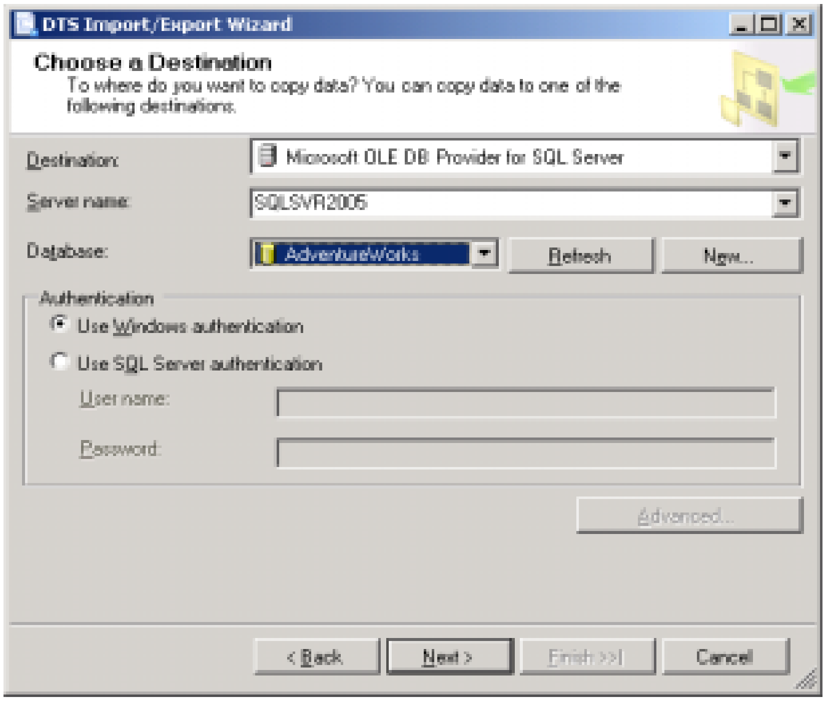
○ Use SQL Server authentication

User name:

Password:

Figure 9.3 : - Specify the Data Source.

Next step is to specify the destination where the source will be moved. At this moment we are moving data inside “AdventureWorks” itself so specify the same database as the source.



Choose a Destination

To where do you want to copy data? You can copy data to one of the following destinations.

Destination: Microsoft OLE DB Provider for SQL Server

Server name: SQLSVR2005

Database: AdventureWorks

Authentication

● Use Windows authentication

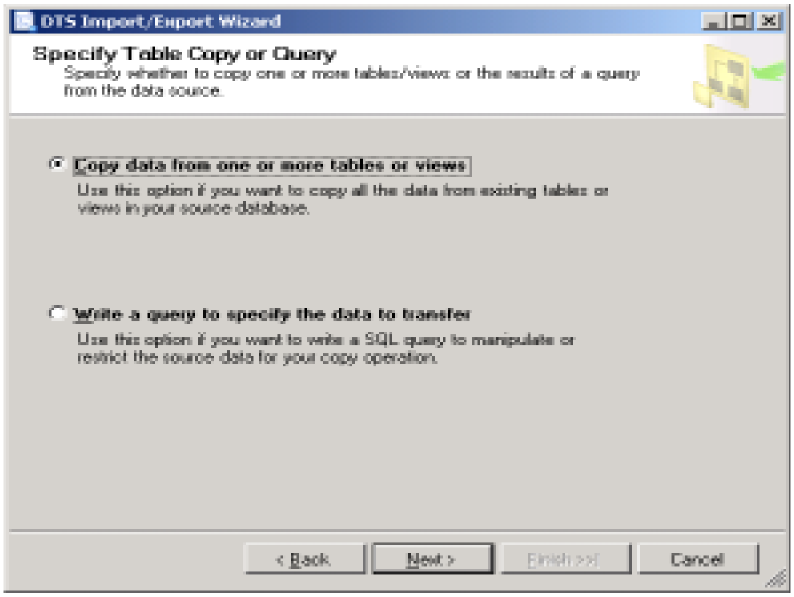
○ Use SQL Server authentication

User name:

Password:

Figure 9.4 : - Specify Destination for DTS

Next step is to specify option from where you want to copy data. For the time being we going to copy from table, so selected the first option.



DTS Import/Export Wizard

Specify Table Copy or Query

Specify whether to copy one or more tables/views or the results of a query from the data source.

● Copy data from one or more tables or views

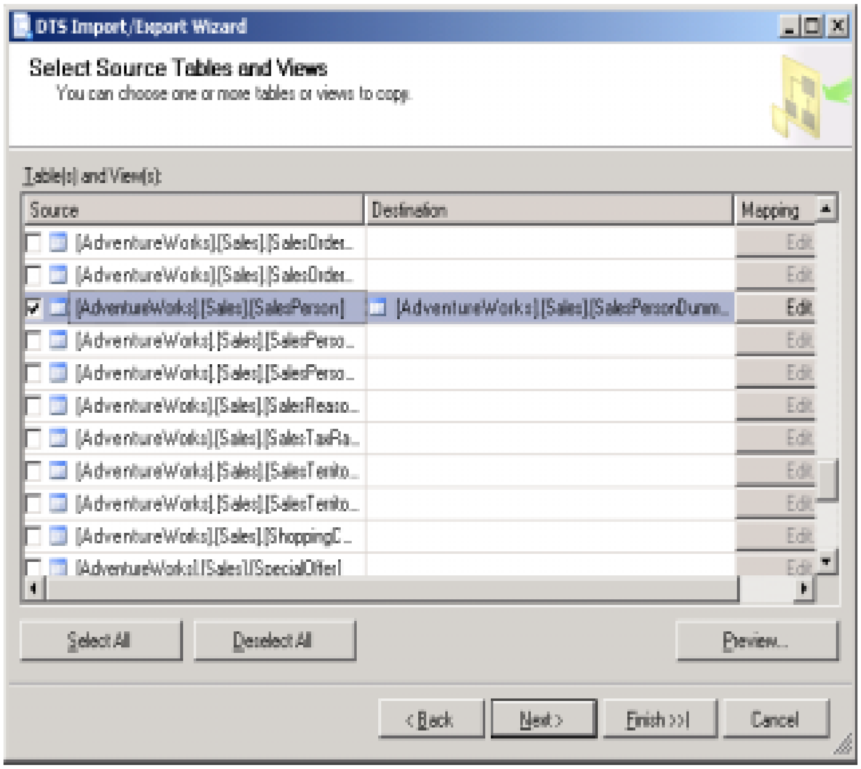
Use this option if you want to copy all the data from existing tables or views in your source database.

○ Write a query to specify the data to transfer

Use this option if you want to write a SQL query to manipulate or restrict the source data for your copy operation.

Figure 9.5 : - Specify option

Finally choose which object you want to map where. You can map multiple objects if you want.



Select Source Tables and Views

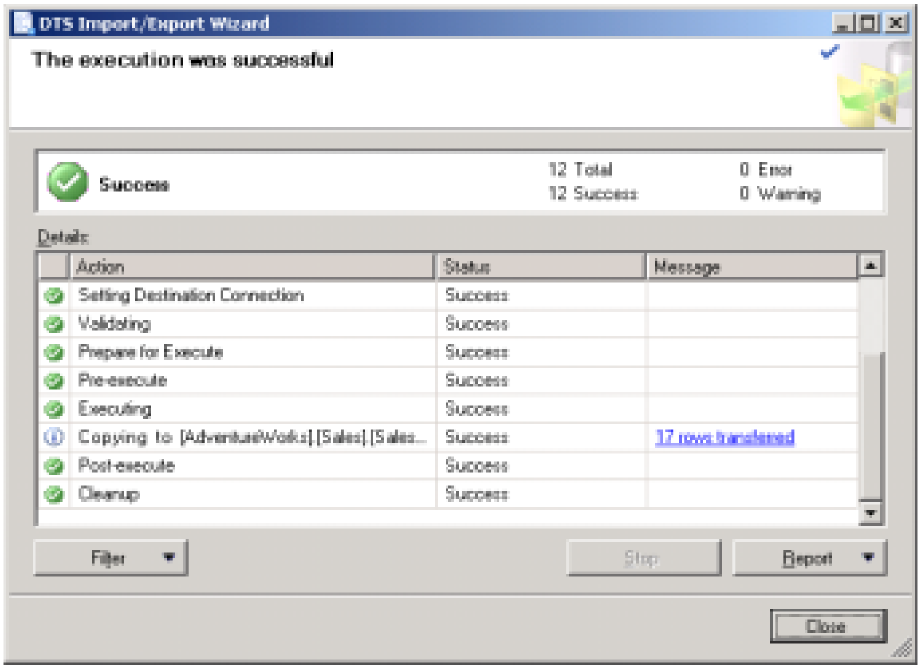
You can choose one or more tables or views to copy.

Table(s) and View(s):

Source	Destination	Mapping
□ [AdventureWorks].[Sales].[SalesOrder...		Edit
□ [AdventureWorks].[Sales].[SalesOrder...		Edit
☑ [AdventureWorks].[Sales].[SalesPerson]	[AdventureWorks].[Sales].[SalesPersonDumm...	Edit
□ [AdventureWorks].[Sales].[SalesPerso...		Edit
□ [AdventureWorks].[Sales].[SalesPerso...		Edit
□ [AdventureWorks].[Sales].[SalesReaso...		Edit
□ [AdventureWorks].[Sales].[SalesTaxRa...		Edit
□ [AdventureWorks].[Sales].[SalesTerrito...		Edit
□ [AdventureWorks].[Sales].[SalesTerrito...		Edit
□ [AdventureWorks].[Sales].[ShoppingC...		Edit
□ [AdventureWorks].[Sales].[SpecialOffer]		Edit

Figure 9.6 : - "Salesperson" is mapped to "SalesPersonDummy"

When everything goes successful you can see the below screen, which shows the series of steps DTS has gone through.



The execution was successful

✓ Success

12 Total
12 Success

0 Error
0 Warning

Details:

	Action	Status	Message
✓	Setting Destination Connection	Success	
✓	Validating	Success	
✓	Prepare for Execute	Success	
✓	Pre-execute	Success	
✓	Executing	Success	
ⓘ	Copying to [AdventureWorks].[Sales].[Sales...	Success	17 rows transferred
✓	Post-execute	Success	
✓	Cleanup	Success	

Figure 9.7 : - Successful execution after series of checks

What are prime components in Integration Services?

There are two important components in Integration services:-

√ DTP (Data transformation pipeline)

DTP is a bridge which connects the source (CSV, any other Database etc) and the destination (SQL Server Database). While DTP moves data between source and destination, transformation takes place between input and output columns. Probably some column will go as one to one mapping and some with some manipulations.

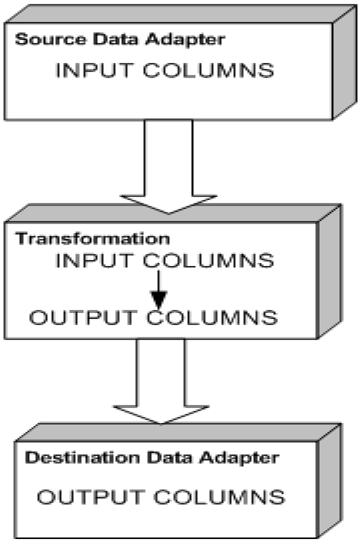


Figure 9.8 : - Data Transformation Pipeline

√ DTR (Data transformation runtime)

While DTP acts as bridge DTR controls you integration service. They are more about how will be the workflow and different components during transformation. Below are different components associated with DTR:-

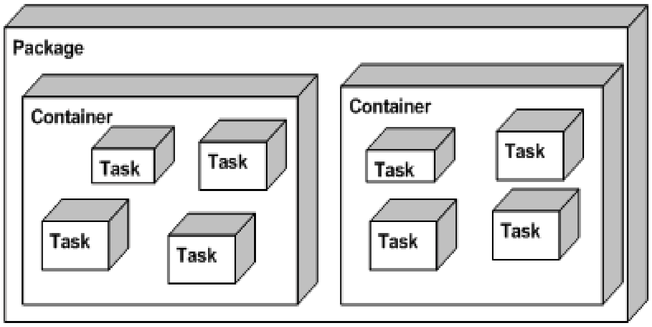


Figure 9.9 : - Data Transformation Runtime

√ Task: - It's the smallest unit which you want to execute.

- √ Container: - Container logically groups task. For instance you have a task to load CSV file in to database. So you will have two or three task probably :-
  - √ Parse the CSV file.
  - √ Check for field data type
  - √ Map the source field to the destination.

So you can define all the above work as task and group them logically in to a container called as Container.

- √ Package: - Package are executed to actually do the data transfer.

DTP and DTR model expose API which can be used in .NET language for better control.

> *Note : -I can hear the shout practical.. practical. I think I have confused you guys over there. So let's warm up on some practical DTS stuff. 1000 words is equal to one compiled program – Shivprasad Koirala ? I really want to invent some proverbs if you do not mind it.*

How can we develop a DTS project in Integration Services?

> *Twist: - Can you say how have you implemented DTS in your project and for what?*

> *Note: - We had visited DTS import / export wizard in previous section of this chapter. But for a real data transformation or a data warehousing (ETL process) it's not enough. You will need to customize the project, there's where we can use this beautiful thing called as "BI development project". If possible just try to go step by step in creating this sample project.*

You can get the development studio as shown below.

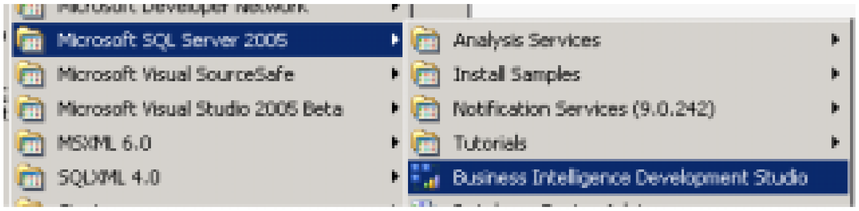

Microsoft Developer Network
Microsoft SQL Server 2005
Microsoft Visual SourceSafe
Microsoft Visual Studio 2005 Beta
MSXML 6.0
SQLXML 4.0
Analysis Services
Install Samples
Notification Services (9.0.242)
Tutorials
Business Intelligence Development Studio

Figure 9.10 : - Location of BI development studio

Click File—New – Project and select “Data Transformation Project”.

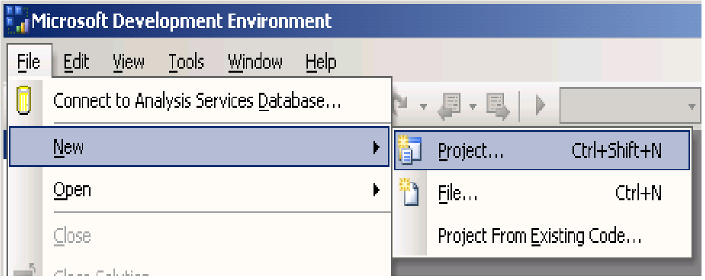

Microsoft Development Environment
File
Edit
View
Tools
Window
Help
Connect to Analysis Services Database...
New
Open
Close
Project...
Ctrl+Shift+N
File...
Ctrl+N
Project From Existing Code...

Figure 9.11 : - New Project DTS

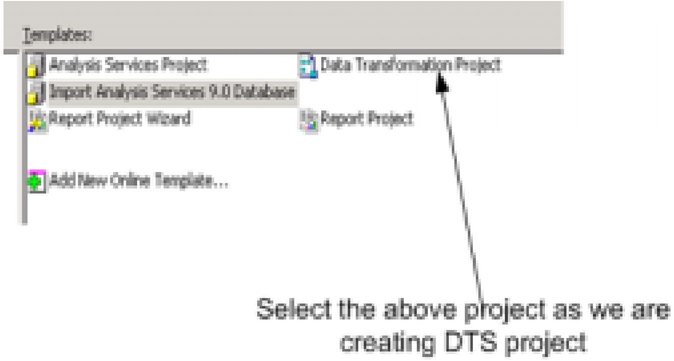

Templates:
Analysis Services Project
Data Transformation Project
Import Analysis Services 9.0 Database
Report Project Wizard
Report Project
Add New Online Template...
Select the above project as we are creating DTS project

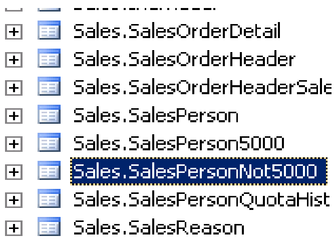
Figure 9.12 : - Dialog DTS project

Give name to the project as “Salesperson” project. Before moving ahead let me give a brief about what we are trying to do. We are going to use “Sales.SalesPerson” table from the “adventureworks” database. “Sales.Salesperson” table has field called as “Bonus”. We have the following task to be accomplished:-

> *Note: - These both tables have to be created manually by you. I will suggest to use the create statements and just make both tables. You can see in the image below there are two tables “SalesPerson5000” and “SalesPersonNot5000”.*

√ Whenever “Bonus” field is equal to 5000 it should go in“Sales.Salesperson5000”.

√ Whenever “Bonus” field is not equal to 5000 it should go in “Sales.SalespersonNot5000”.

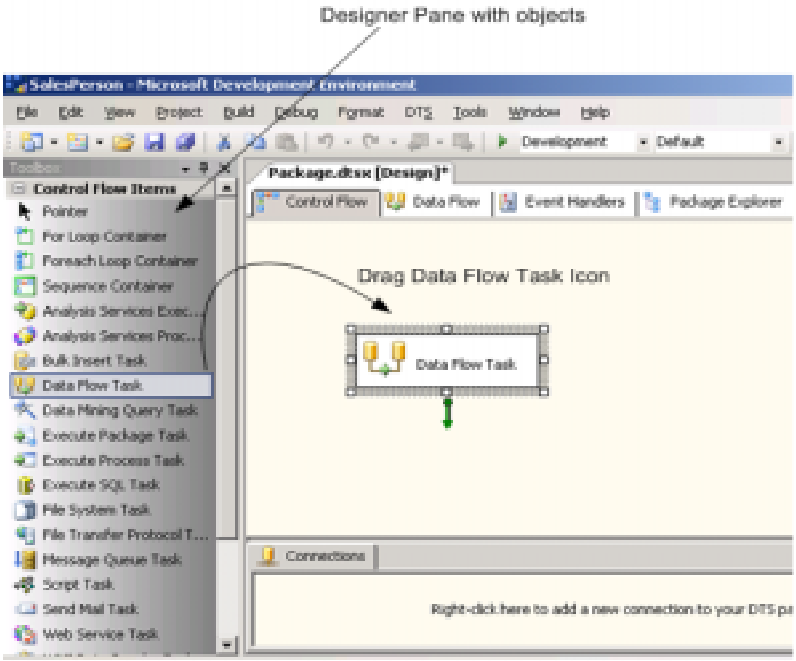


Sales.SalesOrderDetail
Sales.SalesOrderHeader
Sales.SalesOrderHeaderSale
Sales.SalesPerson
Sales.SalesPerson5000
Sales.SalesPersonNot5000
Sales.SalesPersonQuotaHist
Sales.SalesReason

Figure 9.13 : - Snapshot of my database with both tables

One you selected the “Data transformation project” , you will be popped with a designer explorer as show below. I understand you must be saying its cryptic…it is. But let’s try to simplify it. At the right hand you can see the designer pane which has lot of objects on it. At right hand side you can see four tabs (Control flow, Data Flow, Event handlers and Package Explorer).

Control flow: - It defines how the whole process will flow. For example if you loading a CSV file. Probably you will have task like parsing, cleaning and then loading. You can see lot of control flow items which can make your data mining task easy. But first we have to define a task in which we will define all our data flows. So you can see the curve arrow which defines what you have to drag and drop on the control flow designer. You can see the arrow tip which defines the output point from the task.


Designer Pane with objects

Drag Data Flow Task Icon

Right-click here to add a new connection to your DTS pa

Figure 9.14 : - Data Flow Task

In this project I have only define one task, but in real time project something below like this can be seen (Extraction, Transformation and Loading: - ETL). One task points as a input to other task and the final task inputs data in SQL Server.

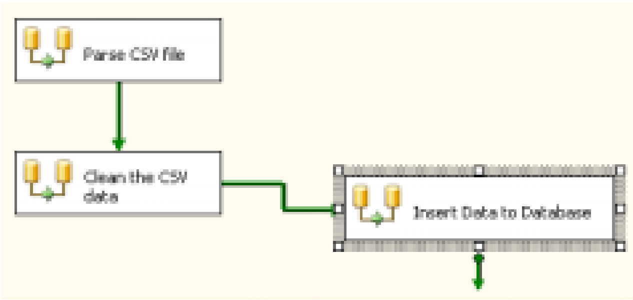


Figure 9.15 : - Multiple Task CSV

Data Flow: - Data flow say how the objects will flow inside a task. So Data flow is subset of a task defining the actual operations.

Event Handlers: - The best of part of DTS is that we can handle events. For instance if there is an error what action do you want it to do. Probably log your errors in error log table, flat file or be more interactive send a mail.

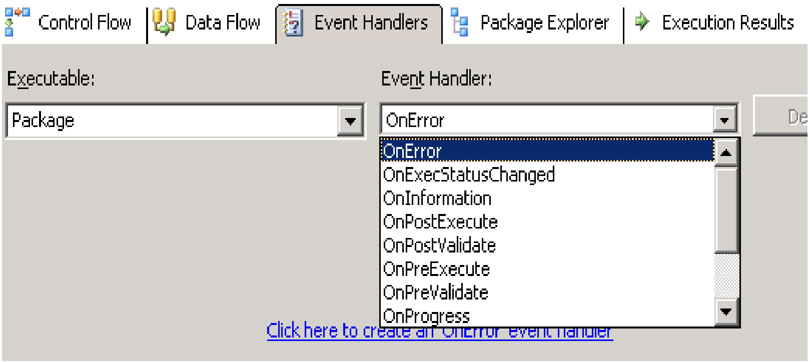


Figure 9.16 : - Event Handlers

Package Explorer: - It shows all objects in a DTS in hierarchical way.

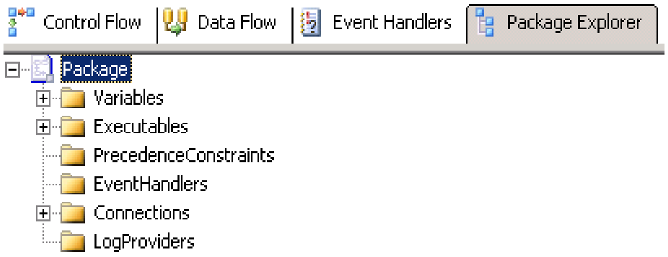
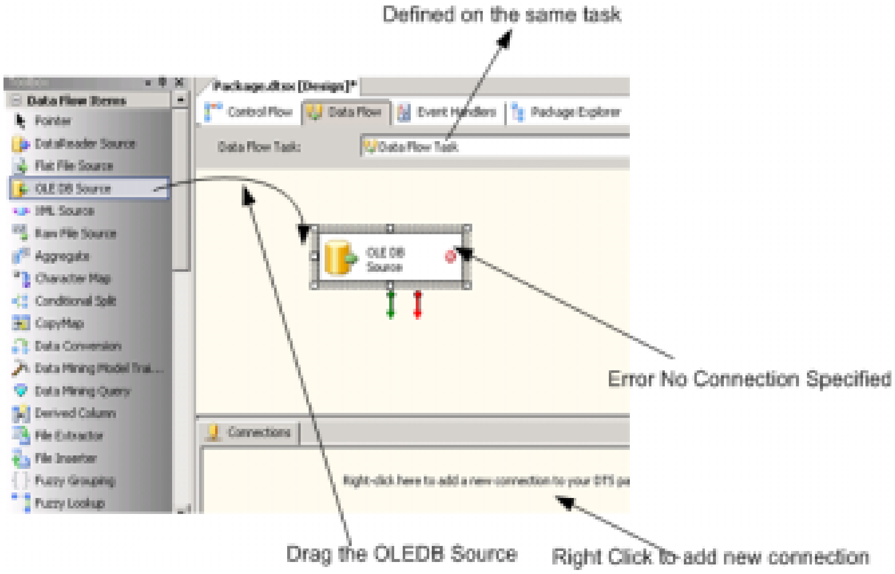


Figure 9.17 :- Package Explorer

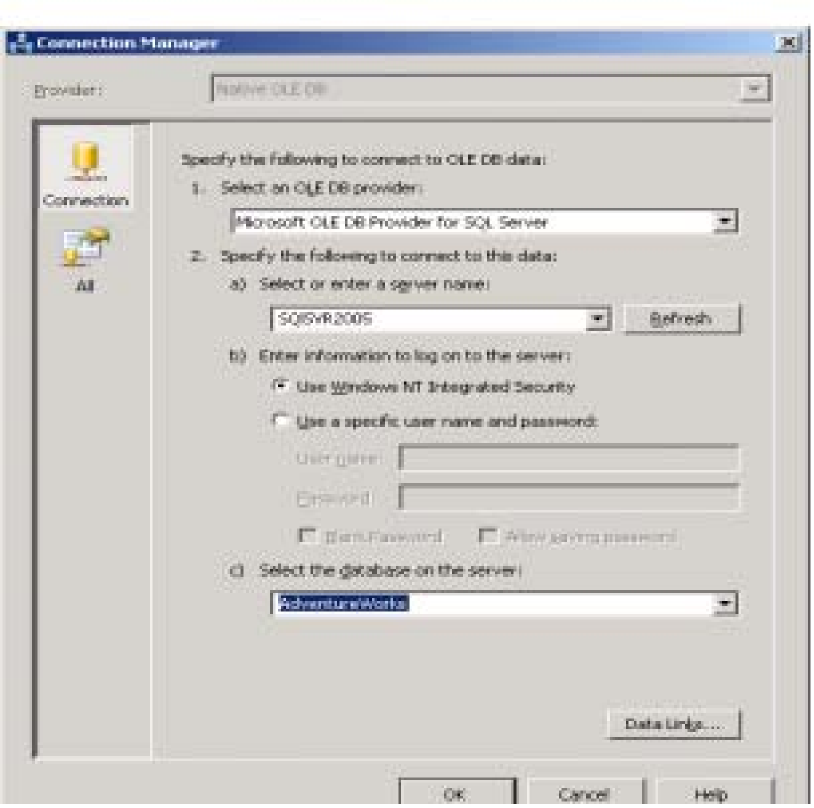
Now that you have defined your task its time to define the actual operation that will happen with in the task. We have to move data from "Sales.SalesPerson" to "Sales.SalesPerson5000" (if their "Bonus" fields are equal to 5000) and "Sales.SalesPersonNot5000" (if their "Bonus" fields are not equal to 5000). In short we have "Sales.SalesPerson" as the source and other two tables as Destination. So click on the “Data Flow” tab and drag the OLEDB Source data flow item on the designer, we will define source in this item. You can see that there is some error which is shown by a cross on the icon. This signifies that you need to specify the source table that is “Sales.Salesperson”.



Right-click here to add a new connection to your DTS pa

Figure 9.18 : - Adding OLEDB Source

In order to specify source tables we need to specify connections for the OLEDB source. So right click on the below tab “Connections” and select “New OLEDB Connection”. You will be popped up with a screen as show below. Fill in all details and specify the database as “AdventureWorks” and click “OK”.



Provider: Native OLE DB

Specify the following to connect to OLE DB data:

1. Select an OLE DB provider:

Microsoft OLE DB Provider for SQL Server

2. Specify the following to connect to this data:

a) Select or enter a server name:

SQISVR2005

b) Enter information to log on to the server:

● Use Windows NT Integrated Security

○ Use a specific user name and password:

User name:

Password:

□ Blank Password □ Allow saving password

c) Select the database on the server:

AdventureWorks

Figure 9.19 : - Connection Manager

If the connection credentials are proper you can see the connection in the “Connections” tab as shown in below figure.

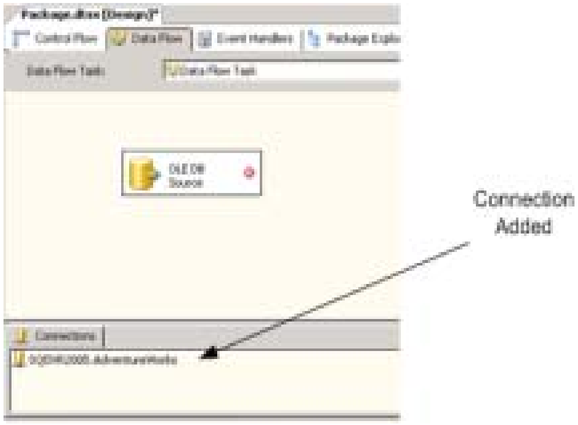


Figure 9.20 : - Connection Added Successfully

Now that we have defined the connection we have to associate that connection with the OLE DB source. So right click and select the "Edit" menu.

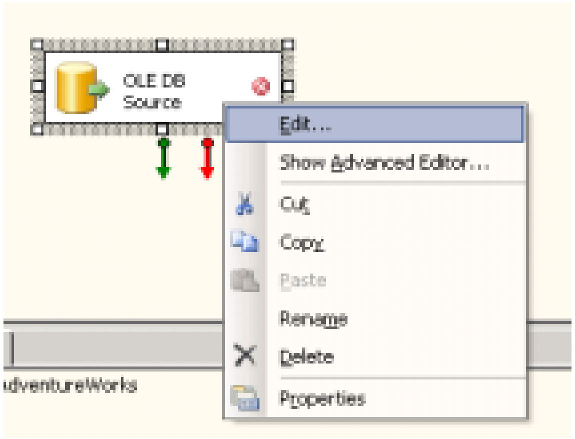
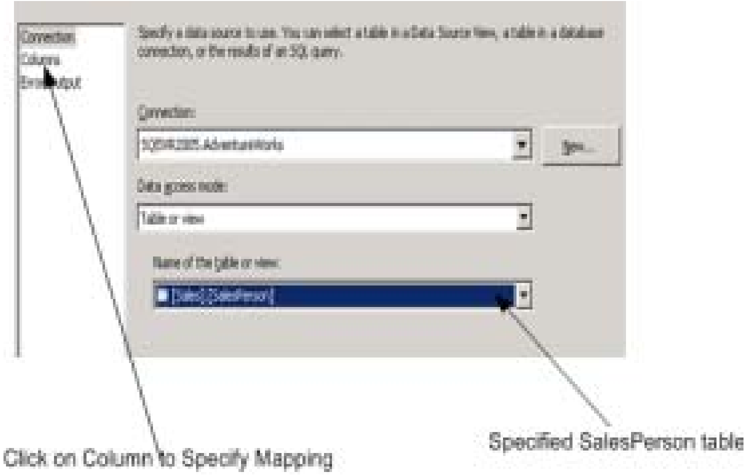


Figure 9.21 : - Edit OleDB

Once you click edit you will see a dialog box as shown below. In data access mode select "Table or View" and select the "Sales.Salesperson" table. To specify the mapping click on "Columns" tab and then press ok.


Connection
Columns
Error Output
New...
Click on Column to Specify Mapping
Specified SalesPerson table

Specify a data source to use. You can select a table in a Data Source View, a table in a database connection, or the results of an SQL query.

Connection:
SQDVR2005.AdventureWorks

Data access mode:
Table or view

Name of the table or view:
[Sales].[SalesPerson]

Figure 9.22 : - Specify Connection Values

If the credentials are ok you can see the red Cross is gone and the OLE DB source is not ready to connect further. As said before we need to move data to appropriate tables on condition that "Bonus" field value. So from the data flow item drag and drop the "Conditional Split" data flow item.

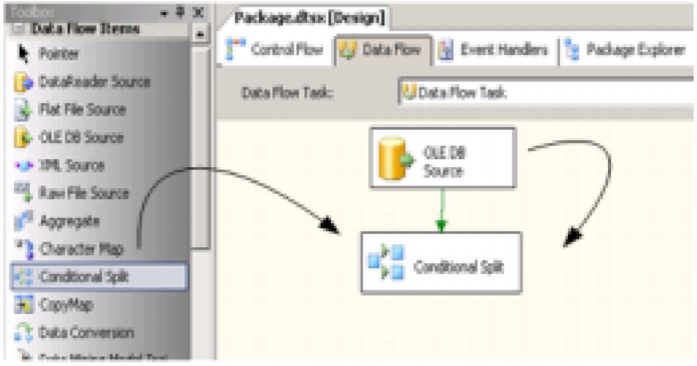
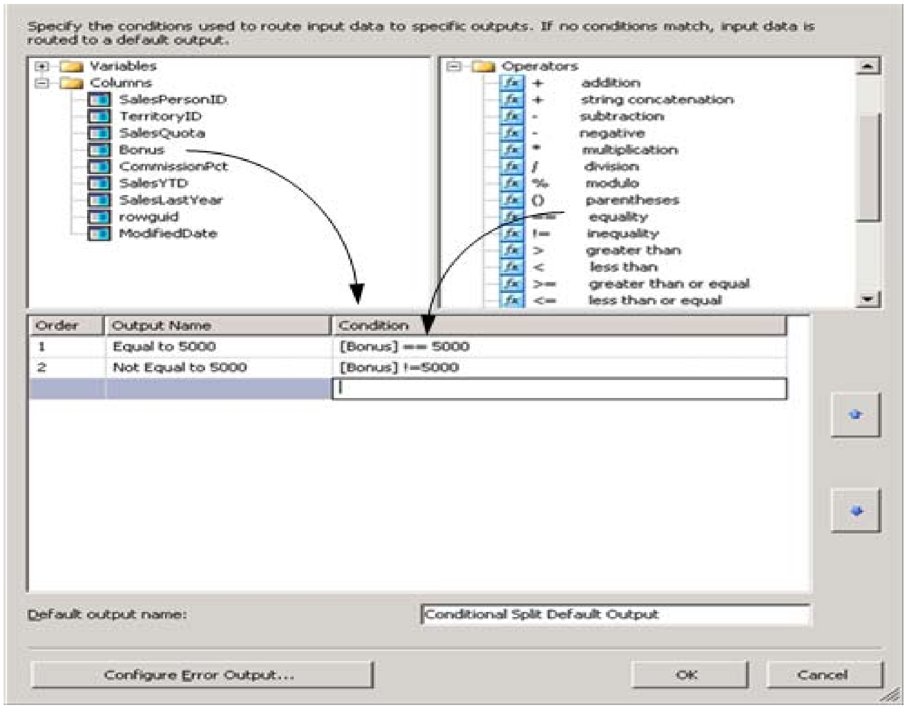


Figure 9.23 : - Conditional Split

Right click on the "Conditional Split" data flow item so that you can specify the criteria. It also gives you a list of fields in the table which you can drag drop. You can also drag drop the operators and specify the criteria. I have made two outputs from the conditional split one which is equal to 5000 and second not equal to 5000.


Specify the conditions used to route input data to specific outputs. If no conditions match, input data is routed to a default output.

Order	Output Name	Condition
1	Equal to 5000	[Bonus] == 5000
2	Not Equal to 5000	[Bonus] !=5000

Default output name: Conditional Split Default Output

Figure 9.24 : - Specifying Conditional Split Criteria

Conditional split now has two outputs one which will go in “Sales.SalesPerson5000” and other in “Sales.SalesPersonNot5000”. So you have to define two destination and the associate respective tables to it. So drag two OLE DB destination data flow items and connect it the two outputs of conditional split.

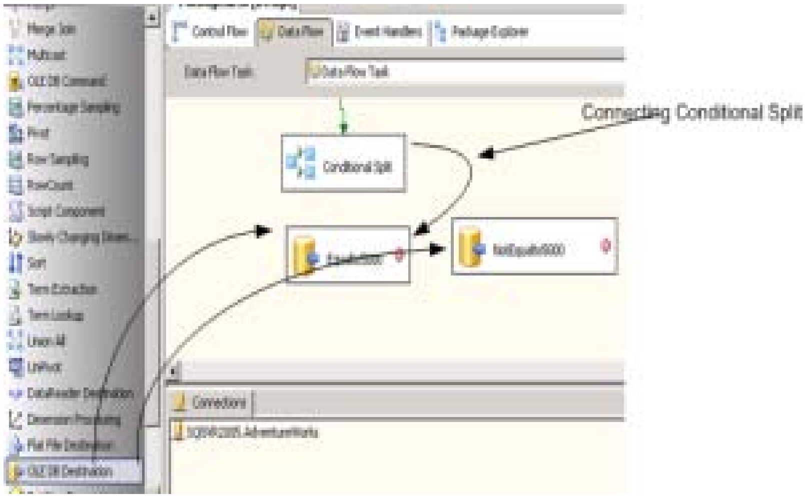
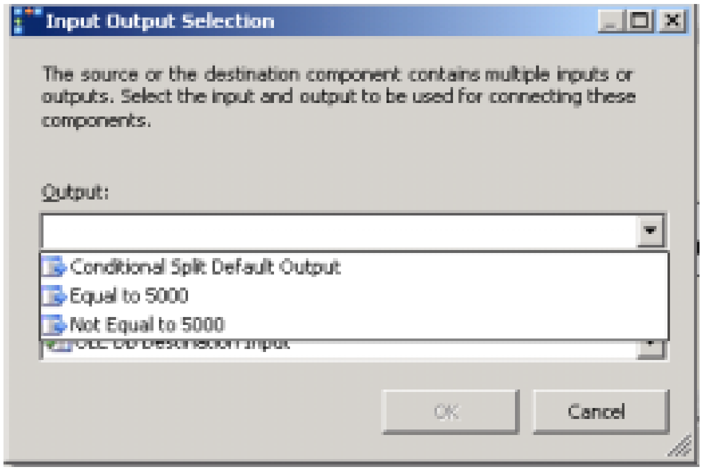


Figure 9.25 : - Specify Destination

When you drag from the conditional split items over OLEDB destination items it will pop up a dialog to specify which output this destination has to be connected. Select the one from drop down and press ok. Repeat this step again for the other destination object.



Input Output Selection

The source or the destination component contains multiple inputs or outputs. Select the input and output to be used for connecting these components.

Output:

Conditional Split Default Output
Equal to 5000
Not Equal to 5000

OLE DB Destination Input

Figure 9.26 : - Specify Input and output Selection

That's the final Data flow structure expected.

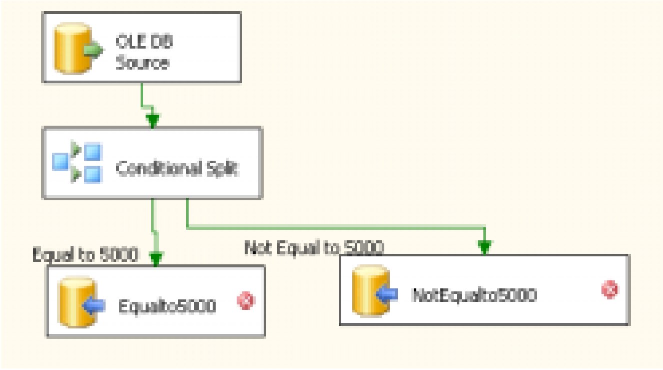


Figure 9.27 : - Final DTS

Its time to build and run the solution which you can do from the drop down. To run the DTS you press the green icon as pointed by arrow in the below figure. After you run query both the tables have the appropriate values or not.

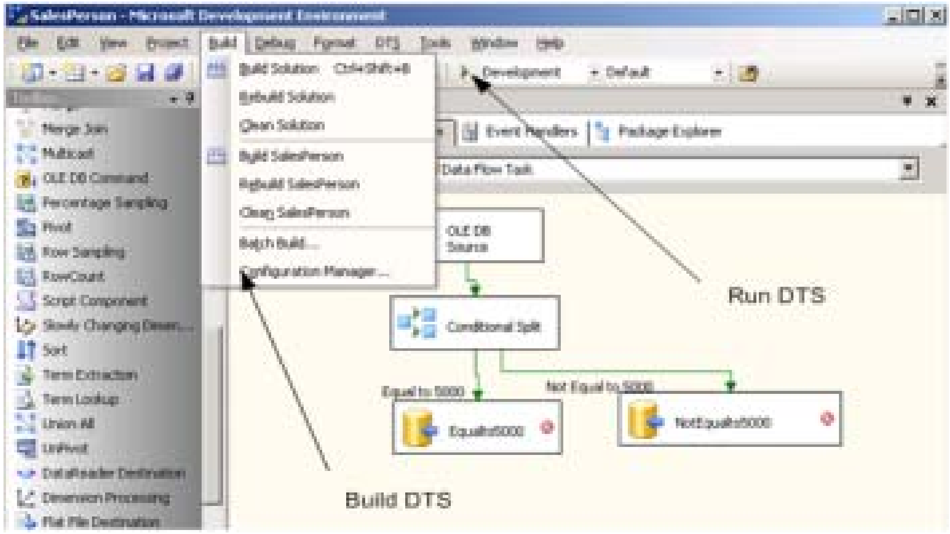


Figure 9.28 : - Build and Run

Note: - You can see various data flow items on the right hand side; it's out of the scope to cover all items (You must be wondering how much time this author will say out of scope , but its fact guys something you have to explore). In this sample project we needed the conditional split so we used it. Depending on projects you will need to explore the toolbox. It's rare that any interviewer will ask about individual items but rather ask fundamentals or general overview of how you did DTS.

10. Replication

Whats the best way to update data between SQL Servers?

By using Replication we can solve this problem. Many of the developers end up saying DTS, BCP or distributed transaction management. But this is one of the most reliable ways to maintain consistency between databases.

What are the scenarios you will need multiple databases with schema?

Following are the situations you can end up in to multi-databases architecture:-

24x7 Hours uptime systems for online systems

This can be one of the major requirements for duplicating SQL Server's across network. For instance you have a system which is supposed to be 24 hours online. This system is hosted in a central database which is far away in terms of Geographic's. As said first that this system should be 24 hours online, in case of any break over from the central server we hosted one more server which is inside the premises. So the application detects that it can not connect to the online server so it connects to the premises server and continues working. Later in the evening using replication all the data from the local SQL Server is sent to the central server.

License problems

SQL Server per user usage has a financial impact. So many of the companies decide to use MSDE which is free, so that they do not have to pay for the client licenses. Later every evening or in some specific interval this all data is uploaded to the central server using replication.

Note: - MSDE supports replication.

Geographical Constraints

It is if the central server is far away and speed is one of the deciding criteria.

Reporting Server

In big multi-national sub-companies are geographically far away and the management wants to host a central reporting server for the sales, which they want to use for decision making and marketing strategy. So here the transactional SQL Server's entire database is scattered across the sub-companies and then weekly or monthly we can push all data to the central reporting server.

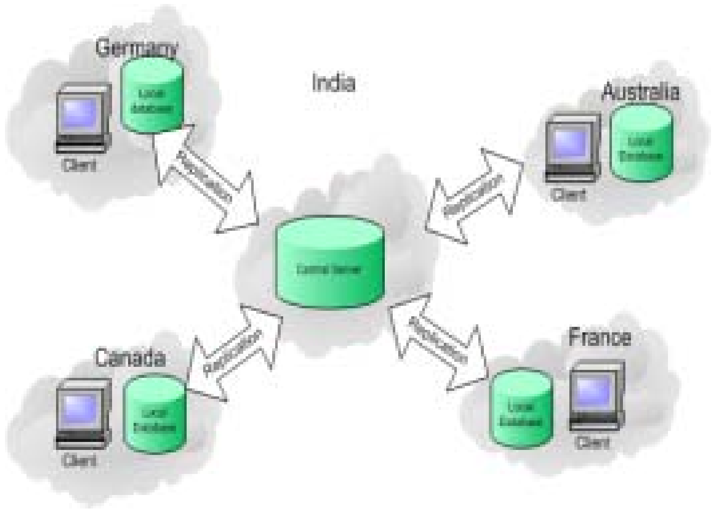


Figure 10.1 : - Replication in Action

You can see from the above figure how data is consolidated in to a central server which is hosted in India using replication.

(DB)How will you plan your replication?

Following are some important questions to be asked before going for replication.

Data planning

It's not necessary that you will need to replicate the complete database. For example you have Sales database which has Customer, Sales, Event logs and History tables. You have requirement to host a centralized reporting server which will be used by top management to know “Sales by Customer”. To achieve this you do not need the whole database on reporting server, from the above you will only need “Sales” and “Customer” tables.

Frequency planning

As defined in the top example let’s say management wants only “Sales by Customer weekly”, so you do not need to update every day , rather you can plan weekly. But if the top management is looking for “Sales by Customer per day” then probably your frequency of updates would be every night.

Schema should not have volatile “baseline”

> *Note: - I like this word “baseline” it really adds weight while speaking as a project manager. It’s mainly used to control change management in projects. You can say “Baseline” is a process by which you can define a logical commit to a document. For example you are coding a project and you have planned different versions for the project. So after every version you do a baseline and create a setup and deploy to the client side. Any changes after this will be a new version.*

One of the primary requirements of a replication is that the schemas which should be replicated across should be consistent. If you are keeping on changing schema of the server then replication will have huge difficulty in synchronizing. So if you are going to have huge and continuous changes in the database schema rethink over replication option. Or else a proper project management will help you solve this.

What are publisher, distributor and subscriber in “Replication”?

Publisher is the one who owns the database and is the main source for data. Publisher identifies what data should be distributed across.

Distributor is a bridge between publisher and subscriber. Distributor gathers all the published data and holds it until it sends it across to all subscriber. So as it’s a bridge who sits in between publisher and subscriber, it supports multiple publisher and subscriber concept.

Subscriber is the end source or the final destination to which data has to be transmitted.

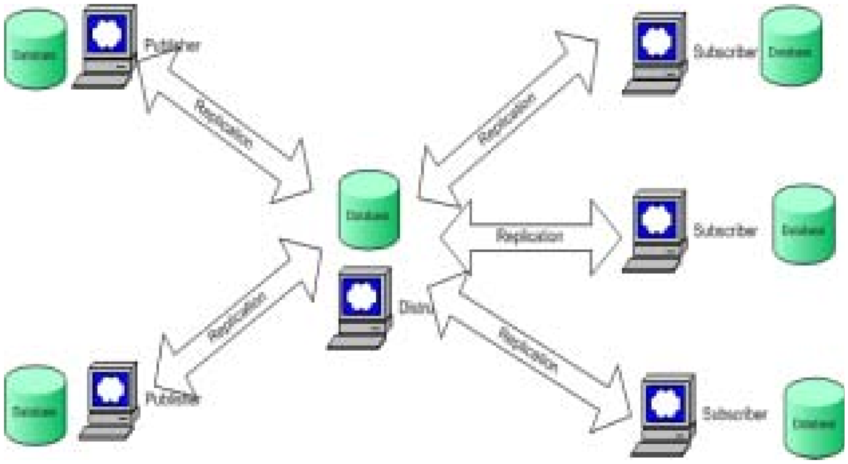


Figure 10.2 : - Publisher, Distributor and Subscriber in action

What is “Push” and “Pull” subscription?

Subscription can be configured in two ways:-

√ Push subscription

In push subscription the publisher has full rights when to update data to subscriber. Subscriber completely plays a passive role in this scenario. This model is needed when we want full control of data in hand of publisher.

√ Pull subscription

In pull subscription the subscriber requests for new or changed data. Subscriber decides when to update himself. This model is needed when we want the control to be on hands of subscriber rather than publisher.

(DB)Can a publication support push and pull at one time?

A publication mechanism can have both. But a subscriber can have only one model for one publication. In short a subscriber can either be in push mode or pull mode for a publication, but not both.

What are different models / types of replication?

√ Snapshot replication

√ Merge replication

√ Transactional replication

Note: - Below I will go through each of them in a very detail way.

What is Snapshot replication?

A complete picture of data to be replicated is taken at the source. Depending on the schedule defined when the replication should happen, destination data is completely replaced by this. So over a period of time changes are accumulated at the publisher end and then depending on the schedule it's sent to the destination.

Note: - In snapshot you will also be sending data which has not changed.

What are the advantages and disadvantages of using Snapshot replication?

Advantages:-

√ Simple to setup. If the database is small or you only want to replicate master data (State code, Pin code etc) it's the best approach, as these values do not change heavily.

√ If you want to keep a tight control over when to schedule the data this is the best approach. For example you will like to replicate when the network traffic is low (probably during Saturday and Sunday).

Disadvantages:

Note: - This is probably the least used approach. So definitely the interviewer is expecting the disadvantages points to be clearer, rather than advantages.

√ As data start growing the time taken to complete the replication will go on increasing.

What type of data will qualify for "Snapshot replication"?

- √ Read-only data are the best candidates for snapshot replication.
- √ Master tables like zip code, pin code etc are some valid data for snapshot replication.

What's the actual location where the distributor runs?

You can configure where the distributor will run from SQL Server. But normally if it's a pull subscription it runs at the subscriber end and for push subscription it runs on the publisher side.

Can you explain in detail how exactly "Snapshot Replication" works?

Following are the basic steps for "Snapshot Replication" to work:-

> *Note: - There are two important components "Snapshot Agent" and "Distribution Agent" which we will define first. Snapshot agent creates image of the complete published data and copies it to the distributor. Distribution Agent sends the copied image and replaces the data on the subscriber side.*

- √ Snapshot agent places a shared lock on the data to be published.
- √ Whole snapshot is then copied to the distributor end. There are three files which are created one for database schema, BCP files and the index data.
- √ Finally the snapshot agent releases lock over the published data.
- √ Distribution agent then replaces the subscriber data using files created by snapshot agent.

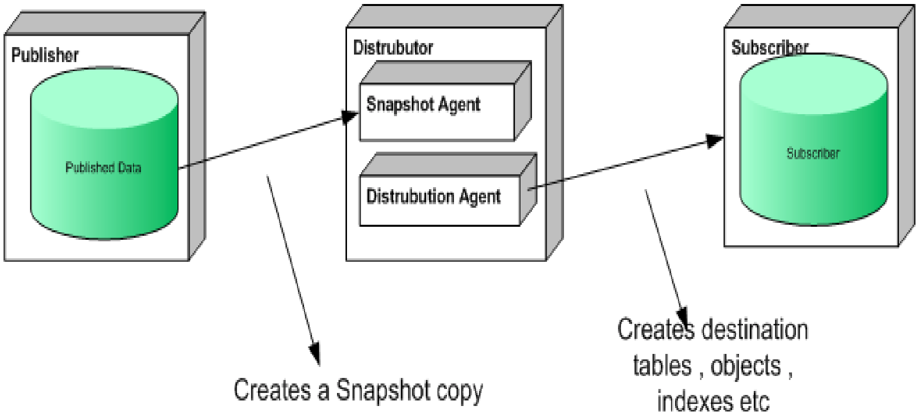


Figure 10.3 : - Snapshot replication in Action

What is merge replication?

If you are looking forward to manage changes on multiple servers which need to be consolidated merge replication is the best design.

How does merge replication works?

Merge Agent component is one of the important components which makes merge replication possible. It consolidates all data from subscriber and applies them to publisher. Merge agent first copies all data from publishers to the subscribers and then replicates them vice-versa so that all stakeholders have consistent data.

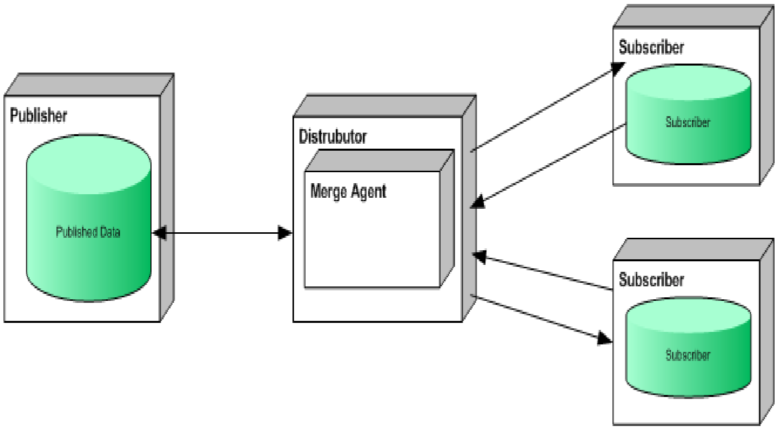


Figure 10.4 : - Merge Replication

Merge agent stands in between subscriber and publisher. Any conflicts are resolved through merge agent in turn which uses conflict resolution. Depending how you have configured the conflict resolution the conflicts are resolved by merge agent.

What are advantages and disadvantages of Merge replication?

Advantages:

- √ This is the only way you can manage consolidating multiple server data.

Disadvantage:

- √ It takes lot of time to replicate and synchronize both ends.
- √ There is low consistency as lot of parties has to be synchronized.
- √ There can be conflicts while merge replication if the same rows are affected in more than once subscriber and publisher. Definitely there is conflict resolution in place but that adds complication.

What is conflict resolution in Merge replication?

There can be practical situations where same row is affected by one or many publishers and subscribers. During such critical times Merge agent will look what conflict resolution is defined and make changed accordingly.

SQL Server uniquely identifies a column using globally unique identifier for each row in a published table. If the table already has a uniqueidentifier column, SQL Server will automatically use that column. Else it will add a rowguid column to the table and create an index based on the column.

Triggers will be created on the published tables at both the Publisher and the Subscribers. These are used to track data changes based on row or column changes.

What is a transactional replication?

Transactional replication as compared to snapshot replication does not replicate full data, but only replicates when anything changes or something new is added to the database. So whenever on publisher side we have INSERT, UPDATE and DELETE operations, these changes are tracked and only these changes are sent to the subscriber end. Transactional Replication is one of the most preferred replication methodologies as they send least amount of data across network.

Can you explain in detail how transactional replication works?

- √ Any change made to the publisher's database is logged in to a log file.
- √ Later log reader agent reads the changes and sends it to the distribution agent.
- √ Distribution agent sends the data across to the subscribers.

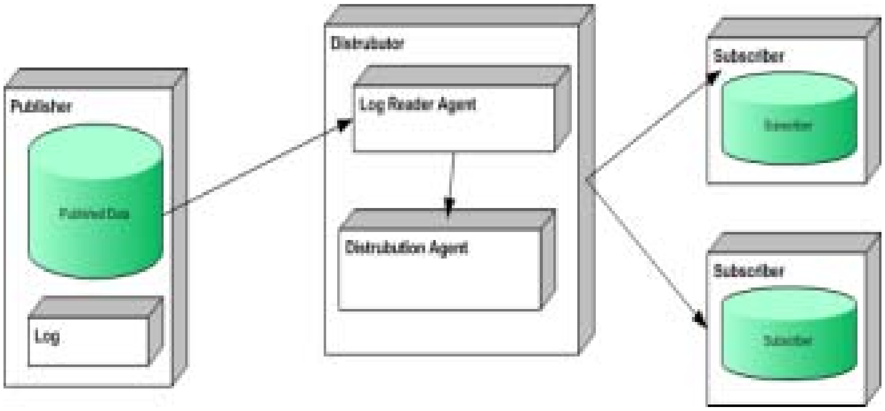


Figure 10.5 : - Transactional Replication

What are data type concerns during replications?

√ If it's a transactional replication you have to include a "Timestamp" column.

√ If it's merge replication you will need a "uniqueidentifier" column. If you do not have one replication creates one.

> *Note: - As this is an interview question book we will try to limit only to theoretical basis. The best way is to practically do one sample of replication with a sample project. But just for your knowledge I will show some important screen's of replication wizard.*

In the "SQL Server Management studio" you can see the publication folder. When you right click on it you can see the "New Publication" menu.

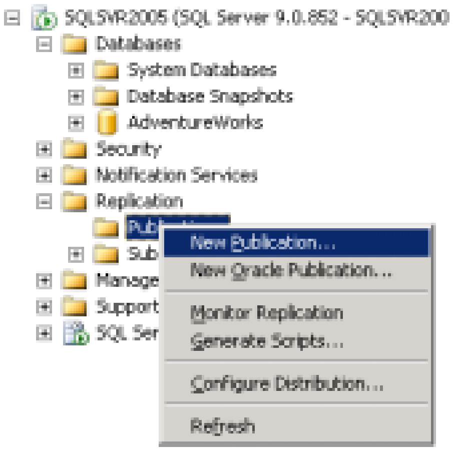
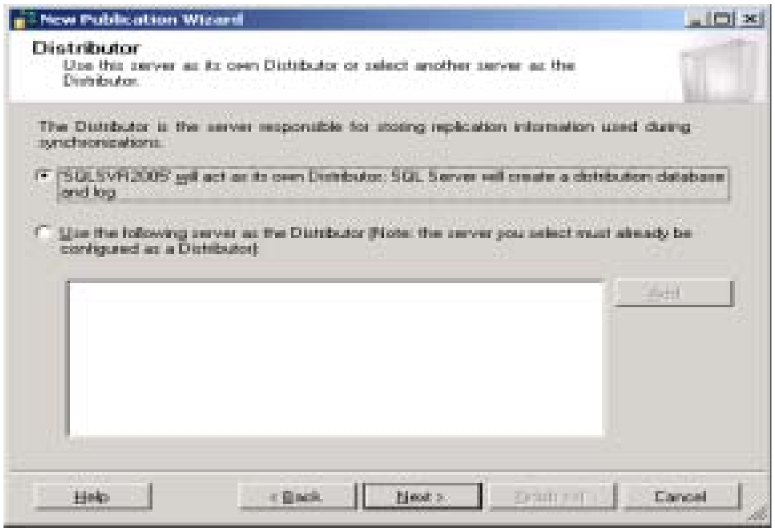


Figure 10. 6 : - Create new publication

This wizard will be used to specify the "Distributor".



Distributor

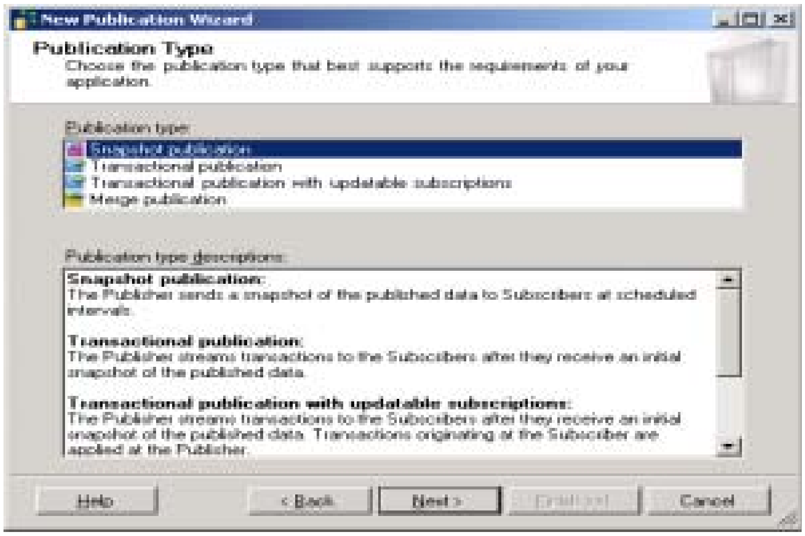
Use this server as its own Distributor or select another server as the Distributor.

The Distributor is the server responsible for storing replication information used during synchronizations.

● 'SQLSVR2005' will act as its own Distributor; SQL Server will create a distribution database and log

○ Use the following server as the Distributor (Note: the server you select must already be configured as a Distributor)

Figure 10.7 : - Specify the Server as distributor



Publication Type

Choose the publication type that best supports the requirements of your application.

Publication type:

- Snapshot publication
- Transactional publication
- Transactional publication with updatable subscriptions
- Merge publication

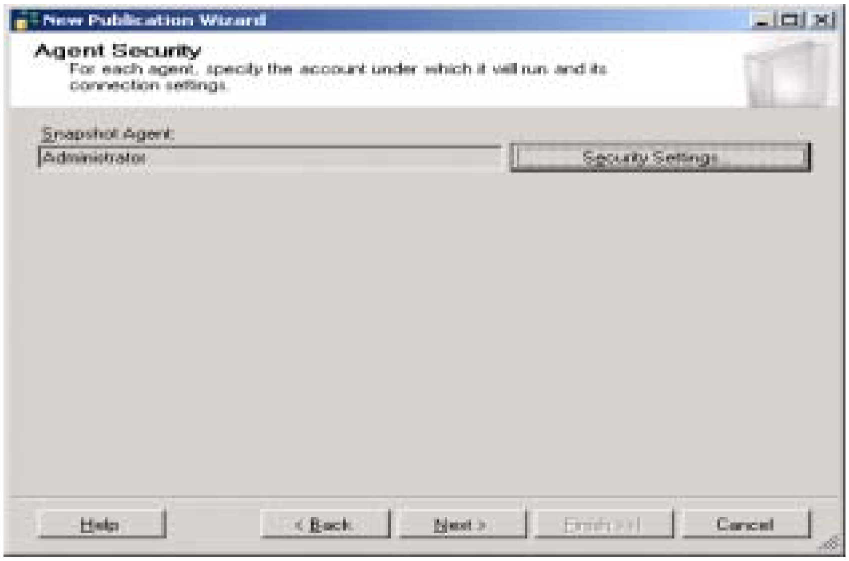
Publication type descriptions:

Snapshot publication:
The Publisher sends a snapshot of the published data to Subscribers at scheduled intervals.

Transactional publication:
The Publisher streams transactions to the Subscribers after they receive an initial snapshot of the published data.

Transactional publication with updatable subscriptions:
The Publisher streams transactions to the Subscribers after they receive an initial snapshot of the published data. Transactions originating at the Subscriber are applied at the Publisher.

Figure 10.8 : - Specify Type of replication

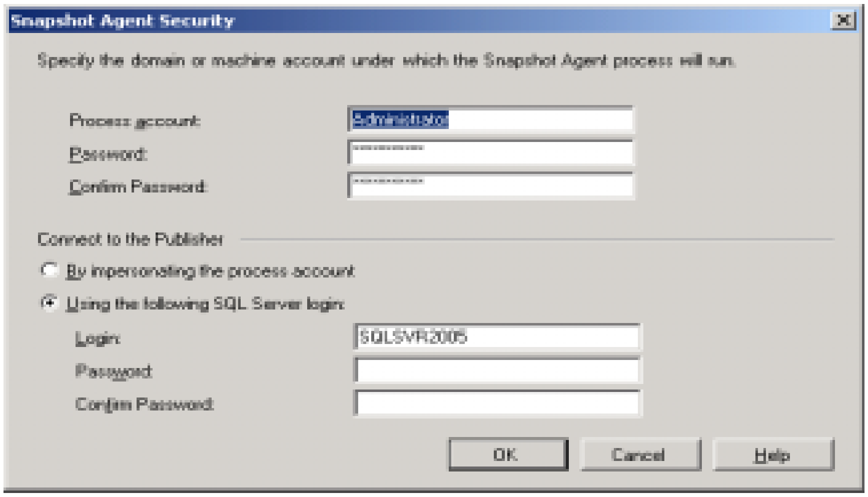


Agent Security

For each agent, specify the account under which it will run and its connection settings.

Snapshot Agent:
Administrator

Figure 10.9 : - Specify under which agent it will run under



Snapshot Agent Security

Specify the domain or machine account under which the Snapshot Agent process will run.

Process account: Administrator

Password: **********

Confirm Password: **********

Connect to the Publisher

○ By impersonating the process account

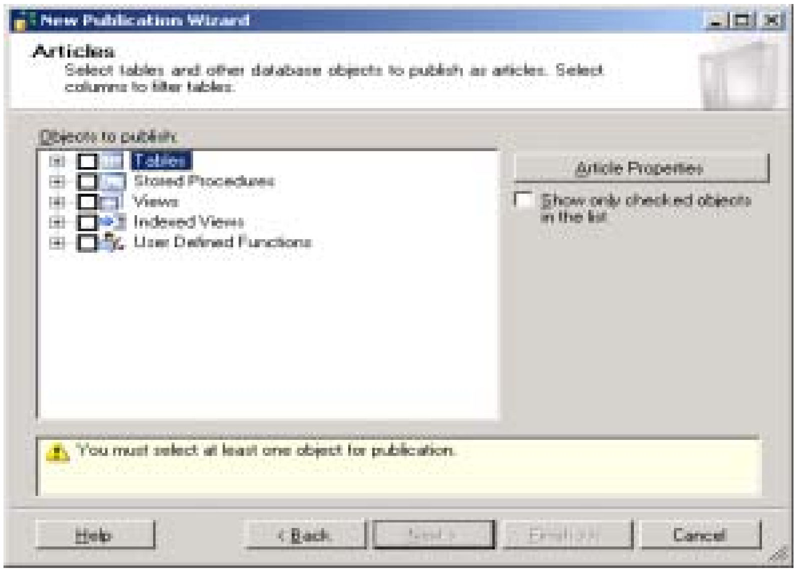
● Using the following SQL Server login:

Login: SQLSVR2005

Password:

Confirm Password:

Figure 10.10 : - Security Details



New Publication Wizard

Articles

Select tables and other database objects to publish as articles. Select columns to filter tables.

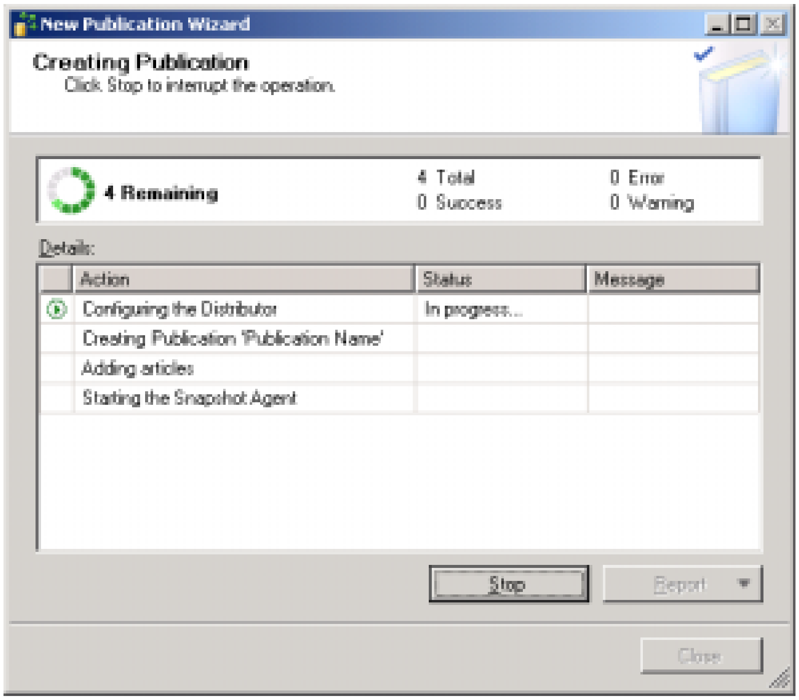
Objects to publish:

- □ Tables
- □ Stored Procedures
- □ Views
- □ Indexed Views
- □ User Defined Functions

□ Show only checked objects in the list

You must select at least one object for publication.

Figure 10.11 : - Specify which objects you want to replicate



Creating Publication

Click Stop to interrupt the operation.

4 Remaining

4 Total
0 Success

0 Error
0 Warning

Details:

Action	Status	Message
Configuring the Distributor	In progress...	
Creating Publication 'Publication Name'		
Adding articles		
Starting the Snapshot Agent		

Figure 10.12 : - Replication in Action

11. Reporting Services

Note: - I know every one screaming this is a part of Data mining and warehousing. I echo the same voice with you my readers, but not necessarily. When you want to derive reports on OLTP systems this is the best way to get your work done. Secondly reporting services is used so much heavily in projects now a day that it will be completely unfair to discuss this topic in a short way as subsection of some chapter.

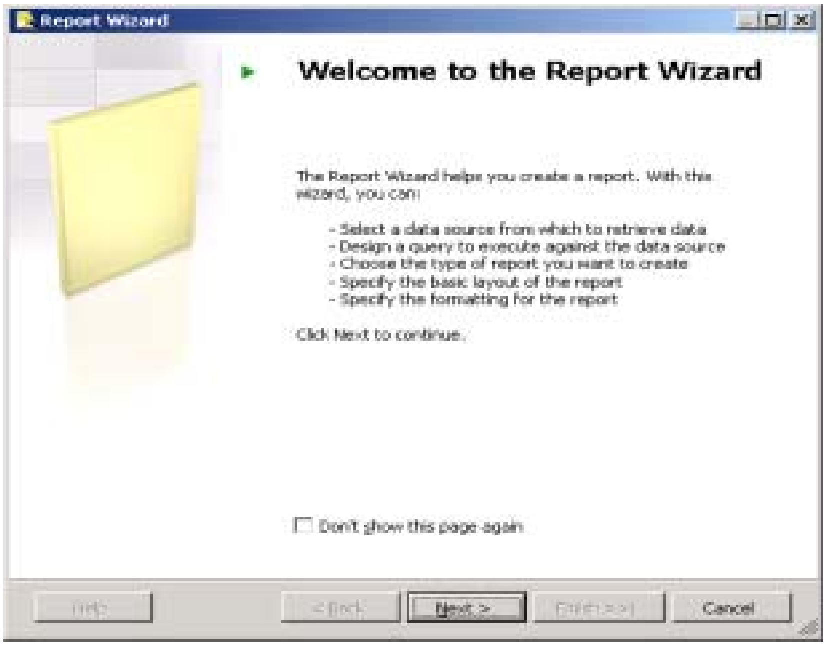
Can you explain how can we make a simple report in reporting services?

We will be using “AdventureWorks” database for this sample. We would like to derive a report how much quantity sales were done per product. For this sample we will have to refer three tables Salesorderdetails, Salesorderheader and product table. Below is the SQL which also shows what the relationship between those tables is:-

```
select production.product.Name  as ProductName, count(*) as TotalSales from
sales.salesorderdetail

inner join Sales.Salesorderheader

on Sales.Salesorderheader.salesorderid= Sales.Salesorderdetail.Salesorderid

inner join production.product

on production.product.productid=sales.salesorderdetail.productid

group by production.product.Name
```

So we will be using the above SQL and trying to derive the report using reporting services.

First click on business intelligence studio menu in SQL Server 2005 and say File --> New --> Project. Select the “Report” project wizard. Let’s give this project name “TotalSalesByProduct”. You will be popped with a startup wizard as shown below.



Report Wizard

Welcome to the Report Wizard

The Report Wizard helps you create a report. With this wizard, you can:

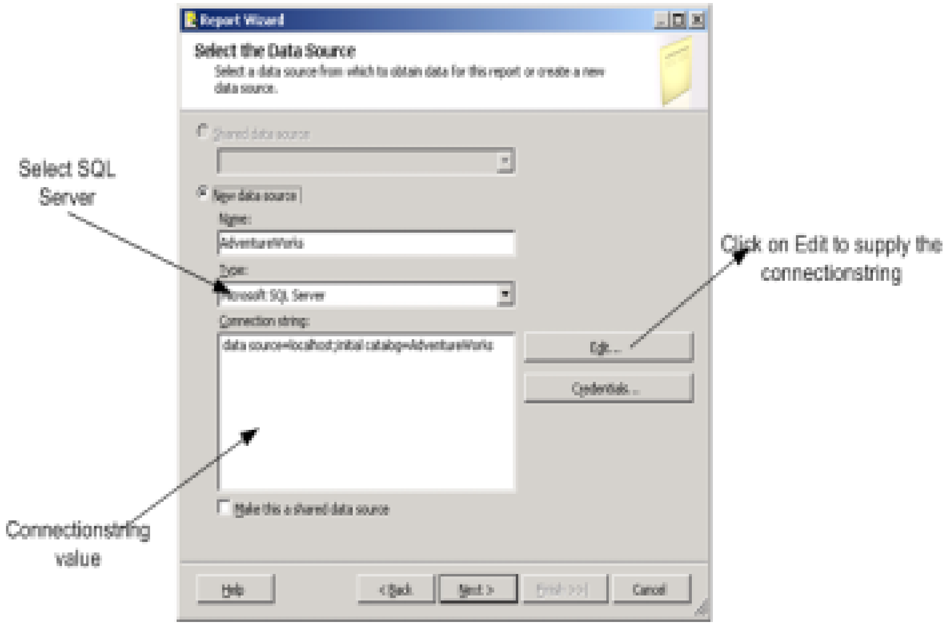
- Select a data source from which to retrieve data
- Design a query to execute against the data source
- Choose the type of report you want to create
- Specify the basic layout of the report
- Specify the formatting for the report

Click Next to continue.

□ Don't show this page again

Figure 11. 1: - Welcome reporting services wizard

Click next and you will be prompted to input data source details like type of server, connection string and name of data source. If you have the connection string just paste it on the text area or else click edit to specify connection string values through GUI.



Report Wizard

Select the Data Source

Select a data source from which to obtain data for this report or create a new data source.

○ Shared data source

● New data source

Name: AdventureWorks

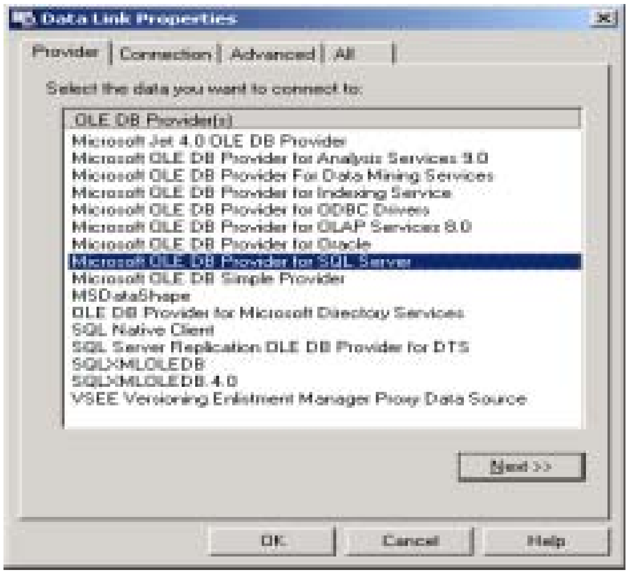
Type: Microsoft SQL Server

Connection string: data source=localhost;initial catalog=AdventureWorks

□ Make this a shared data source

Figure 11.2: - Specify Data Source Details

As we are going to use SQL Server for this sample specify OLEDB provider for SQL Server and click next.



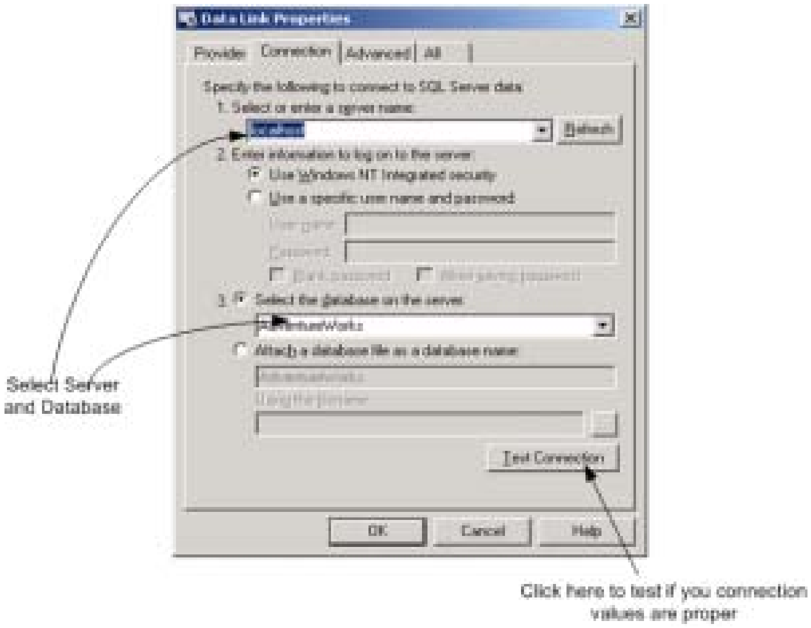
Select the data you want to connect to:

OLE DB Provider(s)

Microsoft Jet 4.0 OLE DB Provider
Microsoft OLE DB Provider for Analysis Services 9.0
Microsoft OLE DB Provider For Data Mining Services
Microsoft OLE DB Provider for Indexing Service
Microsoft OLE DB Provider for ODBC Drivers
Microsoft OLE DB Provider for OLAP Services 8.0
Microsoft OLE DB Provider for Oracle
Microsoft OLE DB Provider for SQL Server
Microsoft OLE DB Simple Provider
MSDataShape
OLE DB Provider for Microsoft Directory Services
SQL Native Client
SQL Server Replication OLE DB Provider for DTS
SQLXMLOLEDB
SQLXMLOLEDB.4.0
VSEE Versioning Enlistment Manager Proxy Data Source

Figure 11.3: - Specify provider

After selecting the provider specify the connection details which will build your connection string. You will need to specify the following details Server Name, Database name and security details.



Specify the following to connect to SQL Server data:

1. Select or enter a server name:
   localhost

2. Enter information to log on to the server:
   ● Use Windows NT Integrated security
   ○ Use a specific user name and password:
   User name:
   Password:
   □ Blank password □ Allow saving password

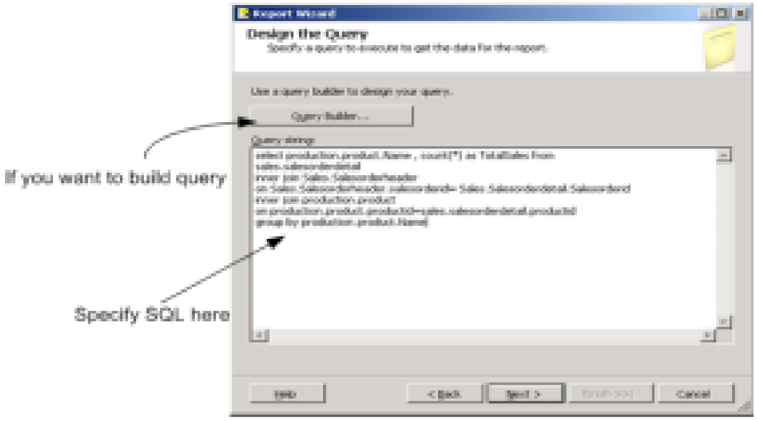
3. ● Select the database on the server:
   AdventureWorks
   ○ Attach a database file as a database name:
   AdventureWorks
   Using the filename:

Select Server and Database

Click here to test if you connection values are proper

Figure 11.4 : - Specify Connection Details

This is the most important step of reporting services, specifying SQL. You remember the top SQL we had specified the same we are pasting it here. If you are not sure about the query you can use the query builder to build your query.


Report Wizard
Design the Query
Specify a query to execute to get the data for the report.
Use a query builder to design your query.
Query Builder...
Query string:
If you want to build query
Specify SQL here
Help
< Back
Next >
Finish >>|
Cancel

```
select production.product.Name , count(*) as TotalSales from
sales.salesorderdetail
inner join Sales.Salesorderheader
on Sales.Salesorderheader.salesorderid= Sales.Salesorderdetail.Salesorderid
inner join production.product
on production.product.productid=sales.salesorderdetail.productid
group by production.product.Name
```

Figure 11.5 : - SQL Query

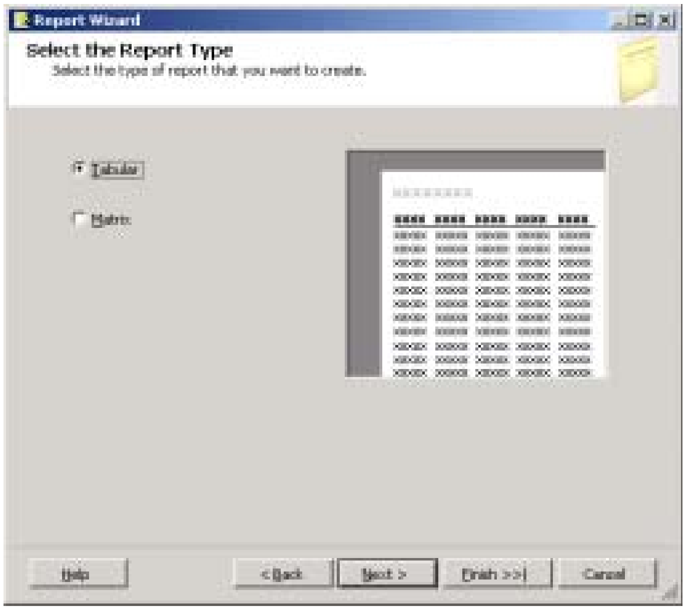
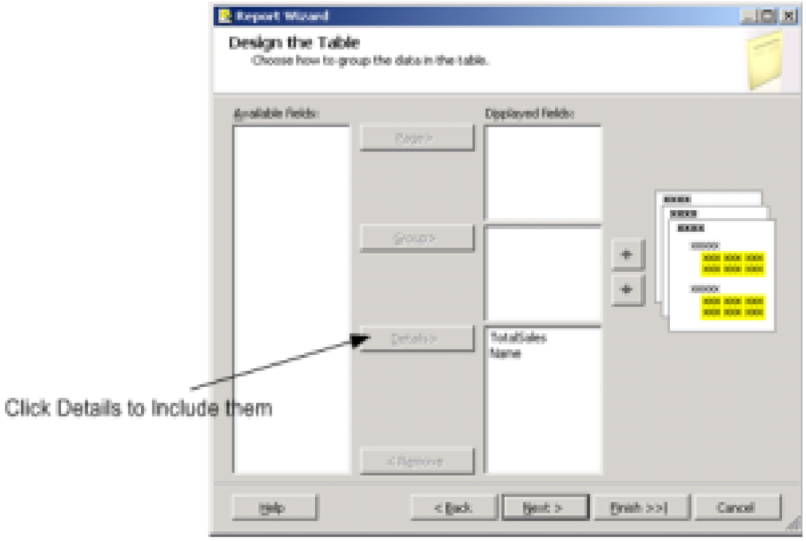


Figure 11.6 : - Type of report

Now it's the time to include the fields in reports. At this moment we have only two fields name of product and total sales.

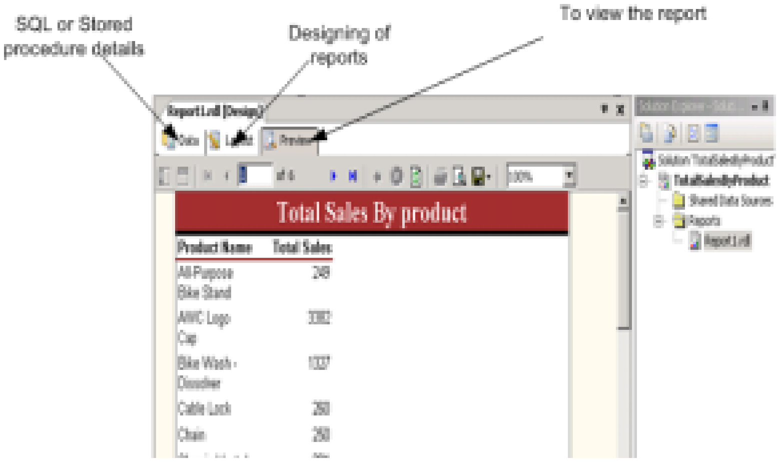


Displayed fields:
TotalSales
Name

Click Details to Include them

Figure 11.7 : - Specify field positions

Finally you can preview your report. In the final section there are three tabs data, layout and preview. In data tab you see your SQL or the data source. In layout tab you can design your report most look and feel aspect is done in this section. Finally below is the preview where you can see your results.



Total Sales By product

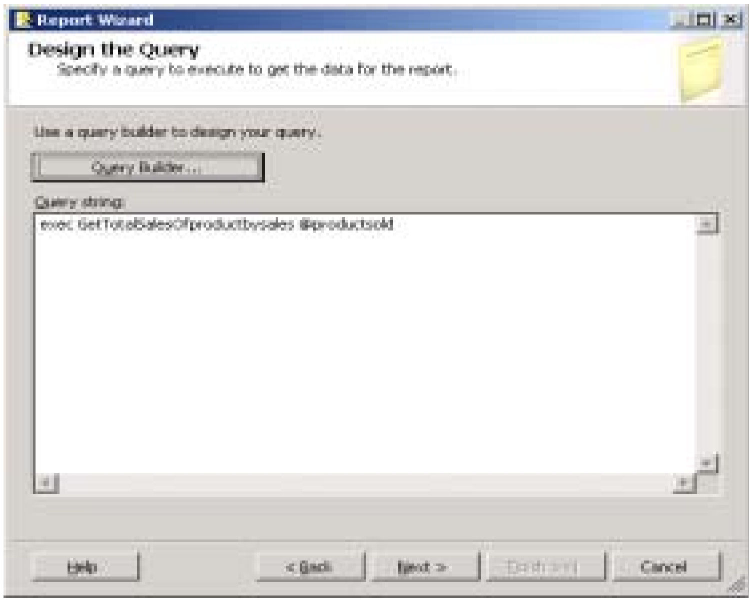
Product Name	Total Sales
All-Purpose Bike Stand	248
AWC Logo Cap	3382
Bike Wash - Dissolver	1327
Cable Lock	260
Chain	250

Figure 11.8 : - Final view of the report

How do I specify stored procedures in Reporting Services?

There are two steps to specify stored procedures in reports of reporting services:-

Specify it in the query string. For instance I have a stored procedure “GettotalSalesofproductsbySales” which has “@ProductSold” as the input parameter.



Report Wizard

Design the Query

Specify a query to execute to get the data for the report.

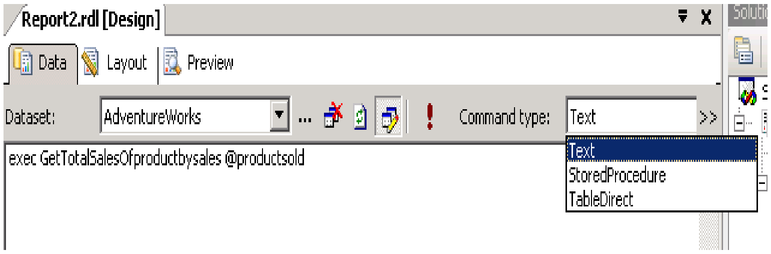
Use a query builder to design your query.

Query string:

```
exec GetTotalSalesOfproductbysales @productsold
```

Figure 11.9 : - stored procedure in the query builder

You have to also specify the command type from the data tab.



Report2.rdl [Design]

Dataset: AdventureWorks

Command type: Text

```
exec GetTotalSalesOfproductbysales @productsold
```

Figure 11.10 : - Specify the command type from the Data tab.

What is the architecture for “Reporting Services “?

"Reporting Services" is not a stand alone system but rather a group of server sub-system which work together for creation, management, and deployment of reports across the enterprise.

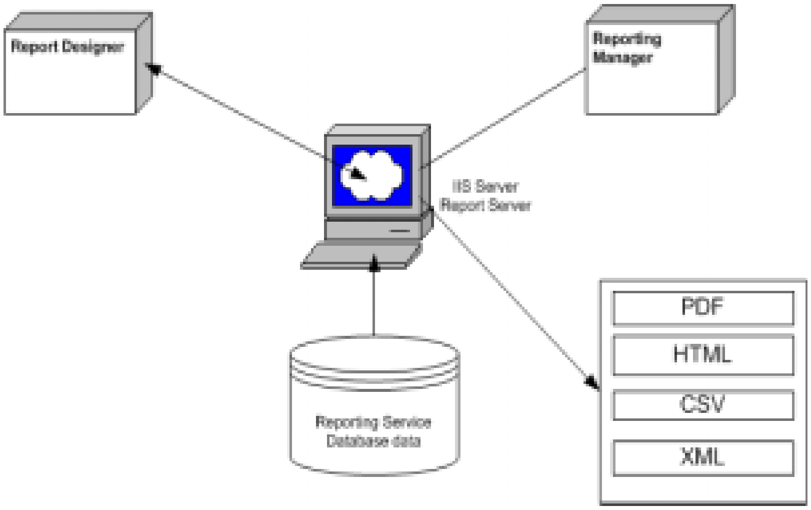


Figure 11.11 : - Reporting Services Architecture

Report designer

This is an interactive GUI which will help you to design and test your reports.

Reporting Service Database

After the report is designed they are stored in XML format. These formats are in RDL (Report Design Layout) formats. These entire RDL format are stored in Report Service Database.

Report Server

Report Server is nothing but an ASP.NET application running on IIS Server. Report Server renders and stores these RDL formats.

Report Manager

It’s again an ASP.NET web based application which can be used by administrators to control security and managing reports. From administrative perspective who have the authority to create the report, run the report etc...

You can also see the various formats which can be generated XML, HTML etc using the report server.

12. Database Optimization

What are indexes?

Index makes your search faster. So defining indexes to your database will make your search faster.

What are B-Trees?

Most of the indexing fundamentals use “B-Tree” or “Balanced-Tree” principle. It’s not a principle that is something is created by SQL Server but is a mathematical derived fundamental.

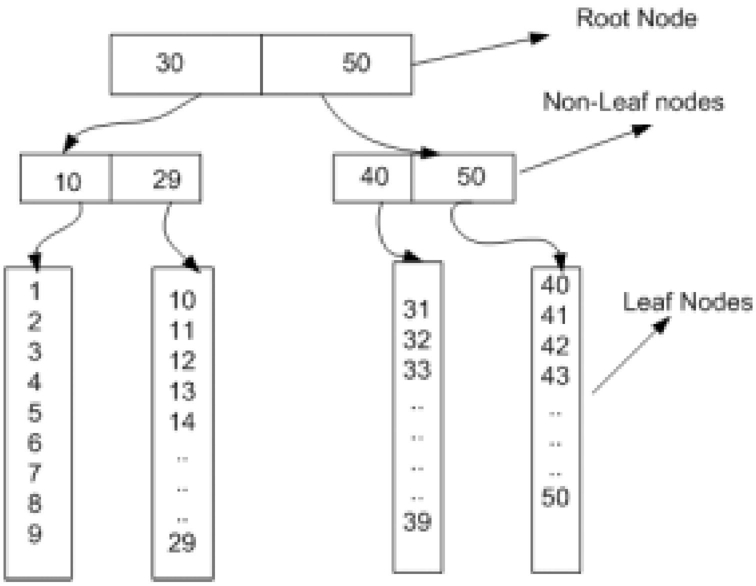


Figure 12.1: - B-Tree principle.

Above is a sample diagram which explains how B-Tree fundamental works. The above diagram is showing how index will work for number from 1-50. Let's say you want to search 39. SQL Server will first start from the first node i.e. root node.

- √ It will see that the number is greater than 30, so it moves to the 50 node.
- √ Further in Non-Leaf nodes it compares is it more than 40 or less than 40. As it's less than 40 it loops through the leaf nodes which belong to 40 nodes.

You can see that this is all attained in only two steps…faster aaah. That is how exactly indexes work in SQL Server.

I have a table which has lot of inserts, is it a good database design to create indexes on that table?

Twist: - Insert's are slower on tables which have indexes, justify it?

Twist: - Why do page splitting happen?

"B-Tree" stands for balanced tree. In order that "B-tree" fundamental work properly both of the sides should be balanced. All indexing fundamentals in SQL Server use "B-tree" fundamental. Now whenever there is new data inserted or deleted the tree tries to become unbalance. In order that we can understand the fundamental properly let's try to refer the figure down.

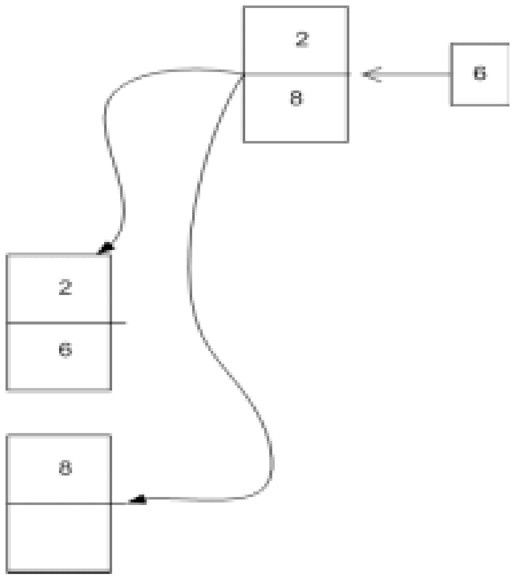


Figure 12.2 : - Page split for Indexed tables

If you see the first level index there is “2” and “8”, now let say we want to insert “6”. In order to balance the “B-TREE” structure rows it will try to split in two pages, as shown. Even though the second page split has some empty area it will go ahead because the primary thing for him is balancing the “B-TREE” for fast retrieval.

Now if you see during the split it is doing some heavy duty here:-

- √ Creates a new page to balance the tree.
- √ Shuffle and move the data to pages.

So if your table is having heavy inserts that means it’s transactional, then you can visualize the amount of splits it will be doing. This will not only increase insert time but will also upset the end-user who is sitting on the screen.

So when you forecast that a table has lot of inserts it’s not a good idea to create indexes.

What are “Table Scan’s” and “Index Scan’s”?

These are ways by which SQL Server searches a record or data in table. In “Table Scan” SQL Server loops through all the records to get to the destination. For instance if you have 1, 2, 5, 23, 63 and 95. If you want to search for 23 it will go through 1, 2 and 5 to reach it. Worst if it wants to search 95 it will loop through all the records.

While for “Index Scan’s” it uses the “B-TREE” fundamental to get to a record. For “B-TREE” refer previous questions.

> *Note: - Which way to search is chosen by SQL Server engine. Example if it finds that the table records are very less it will go for table scan. If it finds the table is huge it will go for index scan.*

What are the two types of indexes and explain them in detail?

> *Twist: - What’s the difference between clustered and non-clustered indexes?*

There are basically two types of indexes:-

- √ Clustered Indexes.
- √ Non-Clustered Indexes.

Ok every thing is same for both the indexes i.e. it uses “B-TREE” for searching data. But the main difference is the way it stores physical data. If you remember the previous figure (give figure number here) there where leaf level and non-leaf level. Leaf level holds the key which is used to identify the record. And non-leaf level actually point to the leaf level.

In clustered index the non-leaf level actually points to the actual data.

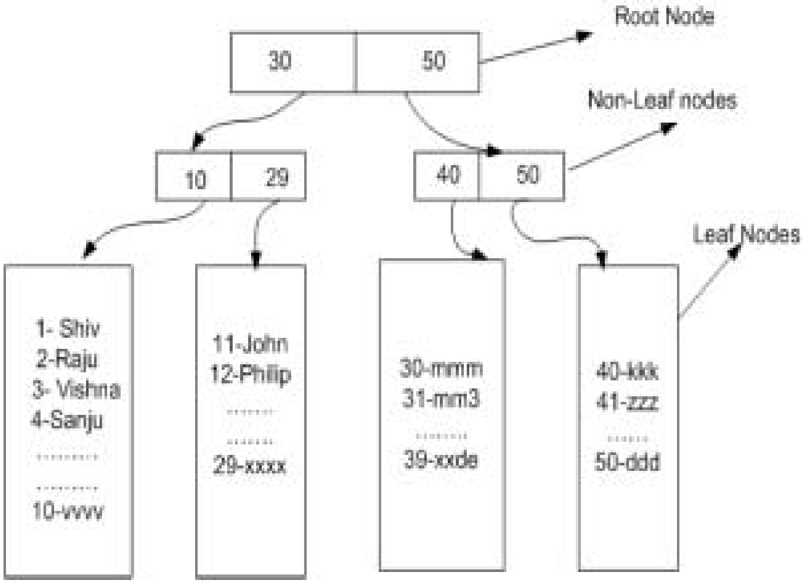


Figure 12.3 : - Clustered Index Architecture

In Non-Clustered index the leaf nodes point to pointers (they are rowid's) which then point to actual data.

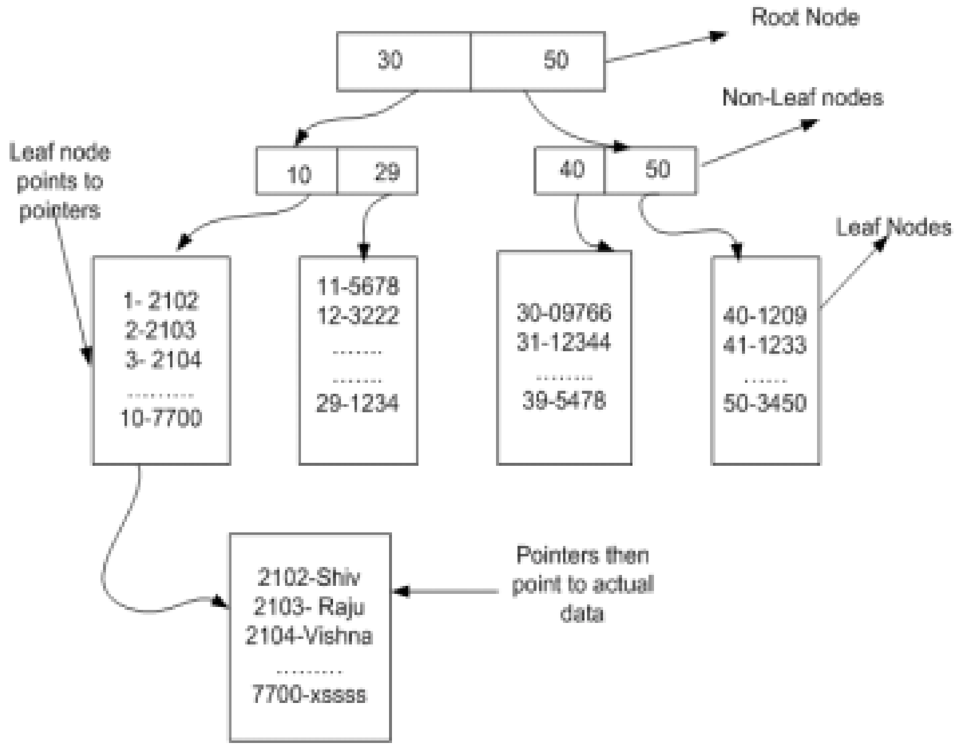


Figure 12.4 : - Non-Clustered Index has pointers.

So here's what the main difference is in clustered and non-clustered , in clustered when we reach the leaf nodes we are on the actual data. In non-clustered indexes we get a pointer, which then points to the actual data.

So after the above fundamentals following are the basic differences between them:-

- √ Also note in clustered index actual data as to be sorted in same way as the clustered indexes are. While in non-clustered indexes as we have pointers which is logical arrangement we do need this compulsion.
- √ So we can have only one clustered index on a table as we can have only one physical order while we can have more than one non-clustered indexes.

If we make non-clustered index on a table which has clustered indexes, how does the architecture change?

The only change is that the leaf node point to clustered index key. Using this clustered index key can then be used to finally locate the actual data. So the difference is that leaf node has pointers while in the next half it has clustered keys. So if we create non-clustered index on a table which has clustered index it tries to use the clustered index.

(DB)What is “FillFactor” concept in indexes?

When SQL Server creates new indexes, the pages are by default full. “FillFactor” is a percentage value (from 1 – 100) which says how much full your pages will be. By default “FillFactor” value is zero.

(DB) What is the best value for “FillFactor”?

“FillFactor” depends on how transactional your database is. Example if your database is highly transactional (i.e. heavy insert’s are happening on the table), then keep the fill factor less around 70. If it’s only a read-only database probably used only for reports you specify 100%.

Remember there is a page split when the page is full. So fill factor will play an important role.

What are “Index statistics”?

Statistics are something the query optimizer will use to decide what type of index (table scan or index scan) to be used to search data. Statistics change according to inserts and updates on the table, nature of data on the table etc...In short “Index statistics” are not same in all situations. So DBA has to run statistics again and again after certain interval to ensure that the statistics are up-to-date with the current data.

> *Note: - If you want to create index you can use either the “Create Index” statement or you can use the GUI.*

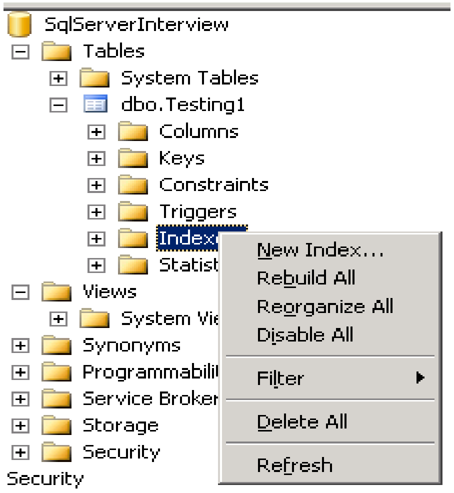


Figure 12.5 : - Index creation in Action.

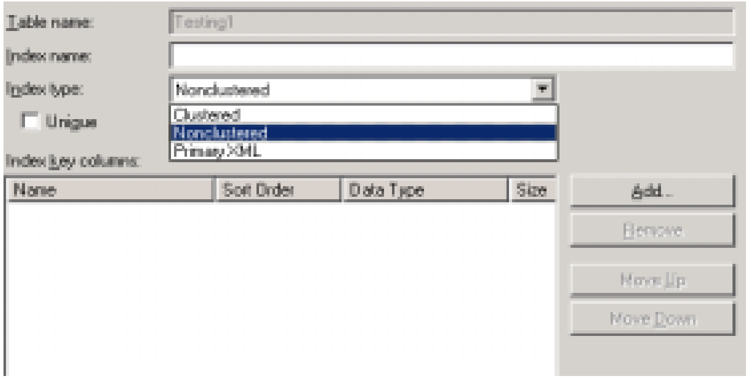


Table name: Testing1

Index name:

Index type: Nonclustered

Clustered
Nonclustered
Primary XML

□ Unique

Index key columns:

Name	Sort Order	Data Type	Size

Figure 12.6 : - Index Details

(DB)How can we see statistics of an index?

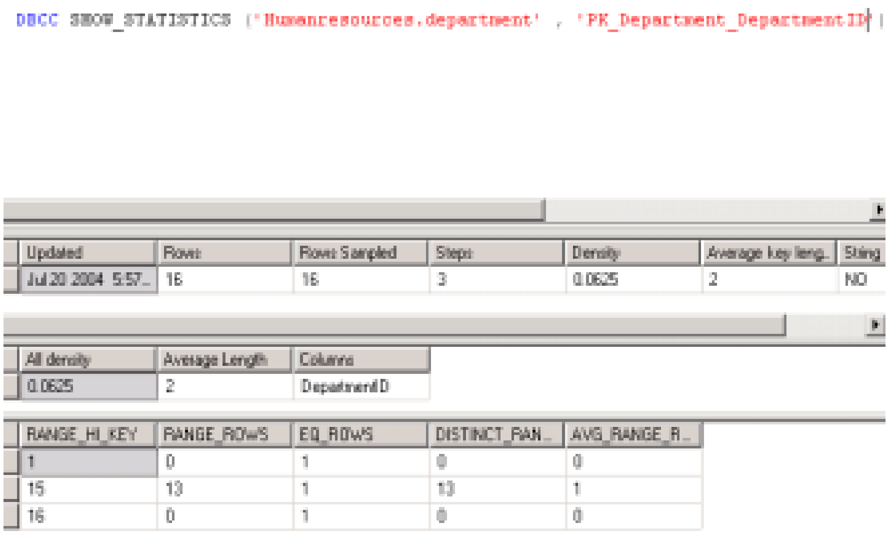
Twist: - How can we measure health of index?

In order to see statistics of any index following the T-SQL command you will need to run.

Note: - Before reading this you should have all the answers of the pervious section clear. Especially about extent, pages and indexes.

```
DECLARE
@ID int,
@IndexID int,
@IndexName varchar(128)
-- input your table and index name
SELECT @IndexName = 'AK_Department_Name'
SET @ID = OBJECT_ID('HumanResources.Department')
SELECT @IndexID = IndID
FROM sysindexes
WHERE id = @ID AND name = @IndexName
--run the DBCC command
DBCC SHOWCONTIG (@id, @IndexID)
```

Just a short note here “DBCC” i.e. “Database consistency checker” is used for checking heath of lot of entities in SQL Server. Now here we will be using it to see index health. After the command is run you will see the following output. You can also run “DBCC SHOWSTATISTICS” to see when was the last time the indexes rebuild.



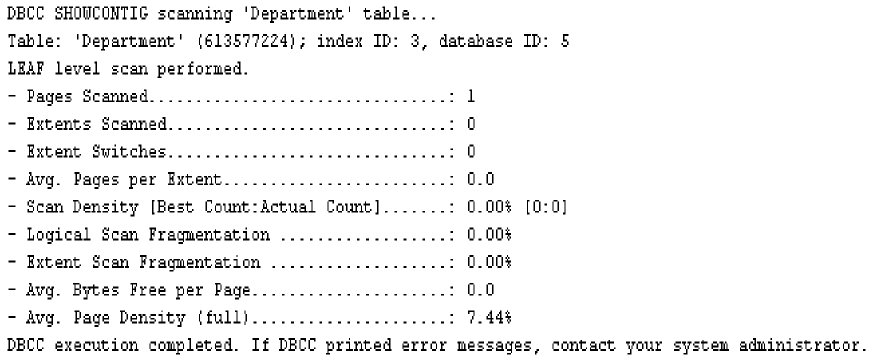
```
DBCC SHOW_STATISTICS ('Humanresources.department' , 'PK_Department_DepartmentID')
```

Updated	Rows	Rows Sampled	Steps	Density	Average key leng...	String
Jul 20 2004 5:57...	16	16	3	0.0625	2	NO

All density	Average Length	Columns
0.0625	2	DepartmentID

RANGE_HI_KEY	RANGE_ROWS	EQ_ROWS	DISTINCT_RAN...	AVG_RANGE_R...
1	0	1	0	0
15	13	1	13	1
16	0	1	0	0

Figure 12.7 DBCC SHOWSTATISTICS



```
DBCC SHOWCONTIG scanning 'Department' table...
Table: 'Department' (613577224); index ID: 3, database ID: 5
LEAF level scan performed.
- Pages Scanned................................: 1
- Extents Scanned..............................: 0
- Extent Switches..............................: 0
- Avg. Pages per Extent........................: 0.0
- Scan Density [Best Count:Actual Count].......: 0.00% [0:0]
- Logical Scan Fragmentation ..................: 0.00%
- Extent Scan Fragmentation ...................: 0.00%
- Avg. Bytes Free per Page.....................: 0.0
- Avg. Page Density (full).....................: 7.44%
DBCC execution completed. If DBCC printed error messages, contact your system administrator.
```

Figure 12.8: - DBCC SHOWCONTIG.

Pages Scanned

The number of pages in the table (for a clustered index) or index.

Extents Scanned

The number of extents in the table or index. If you remember we had said in first instance that extent has pages. More extents for the same number of pages the higher will be the fragmentation.

Extent Switches

The number of times SQL Server moves from one extent to another. More the switches it has to make for the same amount of pages, the more fragmented it is.

Avg. Pages per Extent

The average number of pages per extent. There are eight pages / extent so if you have an extent full with the eight you are in a better position.

Scan Density [Best Count: Actual Count]

This is the percentage ratio of Best count / Actual count. Best count is number of extent changes when everything is perfect. It's like a baseline. Actual count is the actual number of extent changes on that type of scenario.

Logical Scan Fragmentation

Percentage of out-of-order pages returned from scanning the leaf pages of an index. An out of order page is one for which the next page indicated is different page than the page pointed to by the next page pointer in the leaf page. .

Extent Scan Fragmentation

This one is telling us whether an extent is not physically located next to the extent that it is logically located next to. This just means that the leaf pages of your index are not physically in order (though they still can be logically), and just what percentage of the extents this problem pertains to.

Avg. Bytes free per page

This figure tells how many bytes are free per page. If it's a table with heavy inserts or highly transactional then more free space per page is desirable, so that it will have less page splits.

If it's just a reporting system then having this closer to zero is good as SQL Server can then read data with less number of pages.

Avg. Page density (full)

Average page density (as a percentage). It's is nothing but:-

1 - (Avg. Bytes free per page / 8096)

8096 = one page is equal to 8096 bytes

> *Note: - Read every of the above sections carefully, incase you are looking for DBA job you will need the above fundamentals to be very clear. Normally interviewer will try to shoot questions like "If you see the fill factor is this much, what will you conclude? , If you see the scan density this much what will you conclude?*

(DB) How do you reorganize your index, once you find the problem?

You can reorganize your index using "DBCC DBREINDEX". You can either request a particular index to be re-organized or just re-index the all indexes of the table.

This will re-index your all indexes belonging to "HumanResources.Department".

```
DBCC DBREINDEX ([HumanResources.Department])
```

This will re-index only "AK_Department_Name".

```
DBCC DBREINDEX ([HumanResources.Department],[AK_Department_Name])
```

This will re-index with a "fill factor".

```
DBCC DBREINDEX ([HumanResources.Department],[AK_Department_Name],70)
```

You can then again run DBCC SHOWCONTIG to see the results.

What is Fragmentation?

Speed issues occur because of two major things

√ Fragmentation.

√ Splits.

Splits have been covered in the first questions. But one other big issue is fragmentation. When database grows it will lead to splits, but what happens when you delete something from the database…HeHeHe life has lot of turns right. Ok let's say you have two extents and each have two pages with some data. Below is a graphical representation. Well actually that's now how things are inside but for sake of clarity lot of things have been removed.

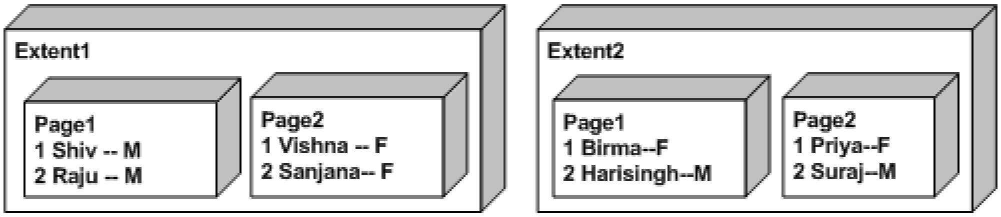


Figure 12.9 : - Data Distribution in Initial Stages

Now over a period of time some Extent and Pages data undergo some delete. Here's the modified database scenario. Now one observation you can see is that some page's are not removed even when they do not have data. Second If SQL server wants to fetch all "Females" it has to span across to two extent and multiple pages within them. This is called as "Fragmentation" i.e. to fetch data you span across lot of pages and extents. This is also termed as "Scattered Data".

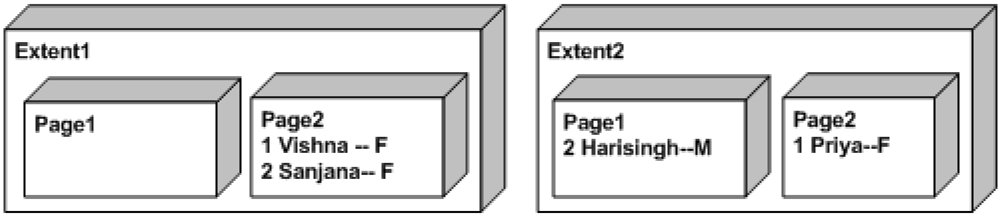


Figure 12.10 : - Data Distribution after Deletes

What if the fragmentation is removed, you only have to search in two extent and two pages. Definitely this will be faster as we are spanning across less entities.

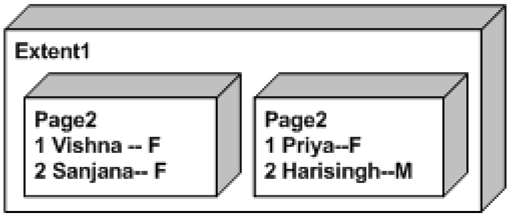


Figure 12.11 : - Fragmentation removed

(DB)How can we measure Fragmentation?

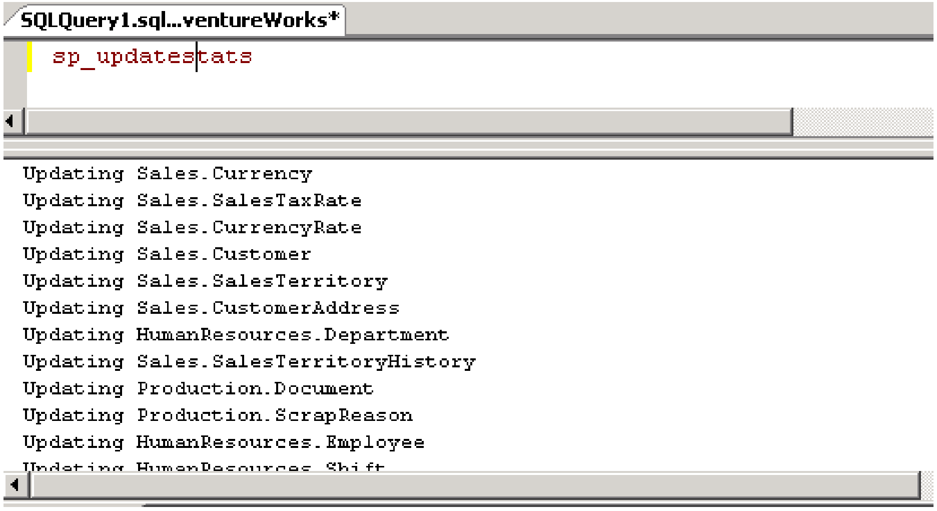
Using "DBCC SHOWCONTIG ".

(DB)How can we remove the Fragmented spaces?

√ Update Statistics :- The most used way by DBA's

√ Sp_updatestats. :- It's same as update statistics , but update statistics applies only for specified object and indexes , while "sp_updatestats" loops through all tables and applies statistics updates to each and every table. Below is a sample which is run on "AdventureWorks" database.

> *Note: - "AdventureWorks" is a sample database which is shipped with SQL Server 2005.*



```
sp_updatestats
```

```
Updating Sales.Currency
Updating Sales.SalesTaxRate
Updating Sales.CurrencyRate
Updating Sales.Customer
Updating Sales.SalesTerritory
Updating Sales.CustomerAddress
Updating HumanResources.Department
Updating Sales.SalesTerritoryHistory
Updating Production.Document
Updating Production.ScrapReason
Updating HumanResources.Employee
Updating HumanResources.Shift
```

Figure 12.12: - sp_updatestats in action

√ DBCC INDEXFRAG: - This is not the effective way of doing fragmentation it only does fragmenting on the leaf nodes.

What are the criteria you will look in to while selecting an index?

Note: - Some answers what I have got for this question.

√ I will create index wherever possible.

√ I will create clustered index on every table.

That's why DBA's are always needed.

√ How often the field is used for selection criteria. For example in a "Customer" table you have "CustomerCode" and "PinCode". Most of the searches are going to be performed on "CustomerCode" so it's a good candidate for indexing rather than using "PinCode". In short you can look in to the "WHERE" clauses of SQL to figure out if it's a right choice for indexing.

- √ If the column has higher level of unique values and is used in selection criteria again is a valid member for creating indexes.
- √ If "Foreign" key of table is used extensively in joins (Inner, Outer, and Cross) again a good member for creating indexes.
- √ If you find the table to be highly transactional (huge insert, update and deletes) probably not a good entity for creating indexes. Remember the split problems with Indexes.
- √ You can use the "Index tuning wizard" for index suggestions.

(DB)What is "Index Tuning Wizard"?

Twist: - What is "Work Load File"?

In the previous question the last point was using the "Index Tuning wizard". You can get the "Index Tuning Wizard" from "Microsoft SQL Server Management Studio" – "Tools" – "SQL Profiler".

> *Note: - This book refers to SQL Server 2005, so probably if you have SQL Server 2000 installed you will get the SQL Profiler in Start – Programs – Microsoft SQL Server -- Profiler. But in this whole book we will refer only SQL Server 2005.We will going step by step for this answer explaining how exactly "Index Tuning Wizard" can be used.*

Ok before we define any indexes let's try to understand what is "Work Load File". "Work Load File" is the complete activity that has happened on the server for a specified period of time. All the activity is entered in to a ".trc" file which is called as "Trace File". Later "Index Tuning Wizard" runs on the "Trace File" and on every query fired it tries to find which columns are valid candidates for indexes depending on the Indexes.

Following are the step's to use "Index Tuning Wizard":-

- √ Create the Trace File using "SQL Profiler".
- √ Then use "Database Tuning Advisor" and the "Trace File" for what columns to be indexed.

Create Trace File

Once you have opened the "SQL Profiler" click on "New Trace".

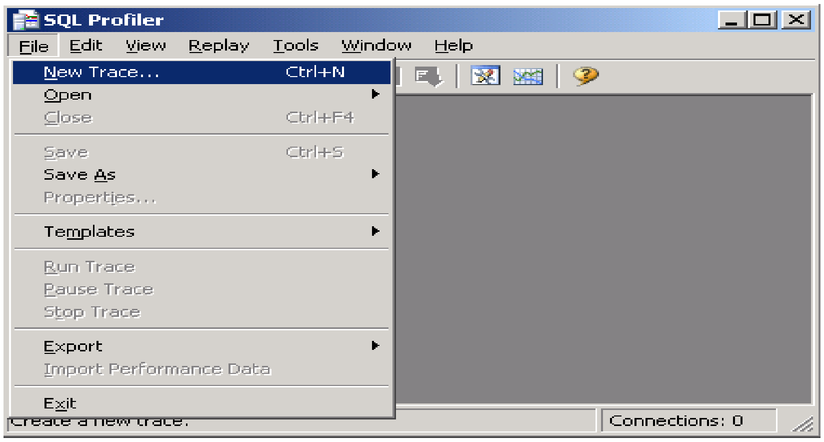
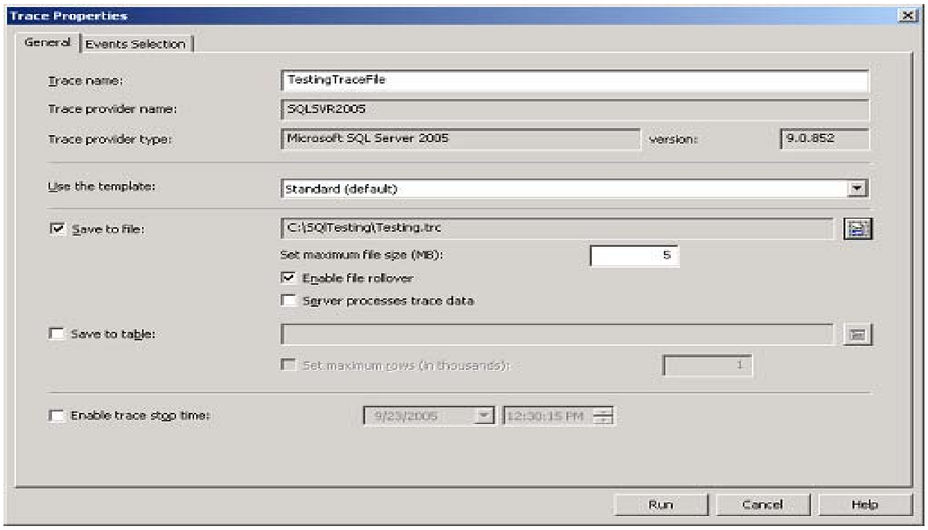


Figure 12.13 : - Create New Trace File.

It will alert for giving you all trace file details for instance the “Trace Name”, “File where to save”. After providing the details click on “Run” button provided below. I have provided the file name of the trace file as “Testing.trc” file.



Trace name: TestingTraceFile

Trace provider name: SQLSVR2005

Trace provider type: Microsoft SQL Server 2005 version: 9.0.852

Use the template: Standard (default)

☑ Save to file: C:\SQlTesting\Testing.trc

Set maximum file size (MB): 5

☑ Enable file rollover

□ Server processes trace data

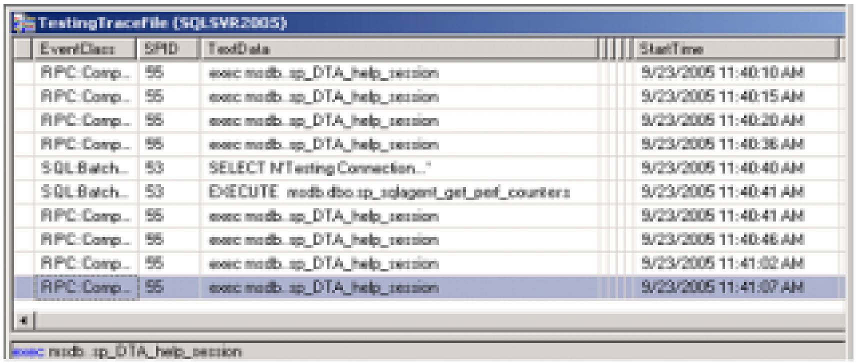
□ Save to table:

□ Set maximum rows (in thousands): 1

□ Enable trace stop time: 9/23/2005 12:30:15 PM

Figure 12.14 : - Trace File Details

HUH and the action starts. You will notice that profiler has started tracing queries which are hitting "SQL Server" and logging all those activities in to the "Testing.trc" file. You also see the actual SQL and the time when the SQL was fired.



EventClass	SPID	TextData	StartTime
RPC:Comp...	55	exec msdb..sp_DTA_help_session	9/23/2005 11:40:10 AM
RPC:Comp...	55	exec msdb..sp_DTA_help_session	9/23/2005 11:40:15 AM
RPC:Comp...	55	exec msdb..sp_DTA_help_session	9/23/2005 11:40:20 AM
RPC:Comp...	55	exec msdb..sp_DTA_help_session	9/23/2005 11:40:36 AM
SQL:Batch...	53	SELECT N'Testing Connection...'	9/23/2005 11:40:40 AM
SQL:Batch...	53	EXECUTE msdb.dbo.sp_sqlagent_get_perf_counters	9/23/2005 11:40:41 AM
RPC:Comp...	55	exec msdb..sp_DTA_help_session	9/23/2005 11:40:41 AM
RPC:Comp...	55	exec msdb..sp_DTA_help_session	9/23/2005 11:40:46 AM
RPC:Comp...	55	exec msdb..sp_DTA_help_session	9/23/2005 11:41:02 AM
RPC:Comp...	55	exec msdb..sp_DTA_help_session	9/23/2005 11:41:07 AM

exec msdb..sp_DTA_help_session

Figure 12.15 : - Tracing in Actions

Let the trace run for some but of time. In actually practical environment I run the trace for almost two hours in peak to capture the actual load on server. You can stop the trace by clicking on the red icon given above.

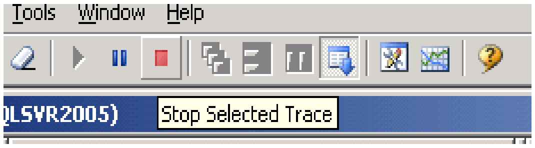


Figure 12.16 : - Stop Trace File.

You can go the folder and see your ".trc" file created. If you try to open it in notepad you will see binary data. It can only be opened using the profiler. So now that we have the load file we have to just say to the advisor hey advisor here's my problem (trace file) can you suggest me some good indexes to improve my database performance.

Using Database Tuning Advisor

In order to go to “Database Tuning Advisor” you can go from “Tools” – “Database Tuning Advisor”.

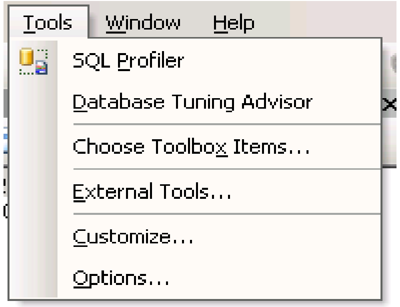


Figure 12.17 : - Menu for SQL Profiler and Database advisor

In order to supply the work load file you have to start a new session in “Database tuning advisor”.

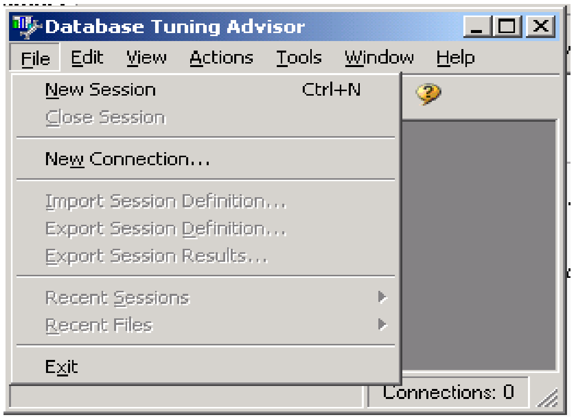


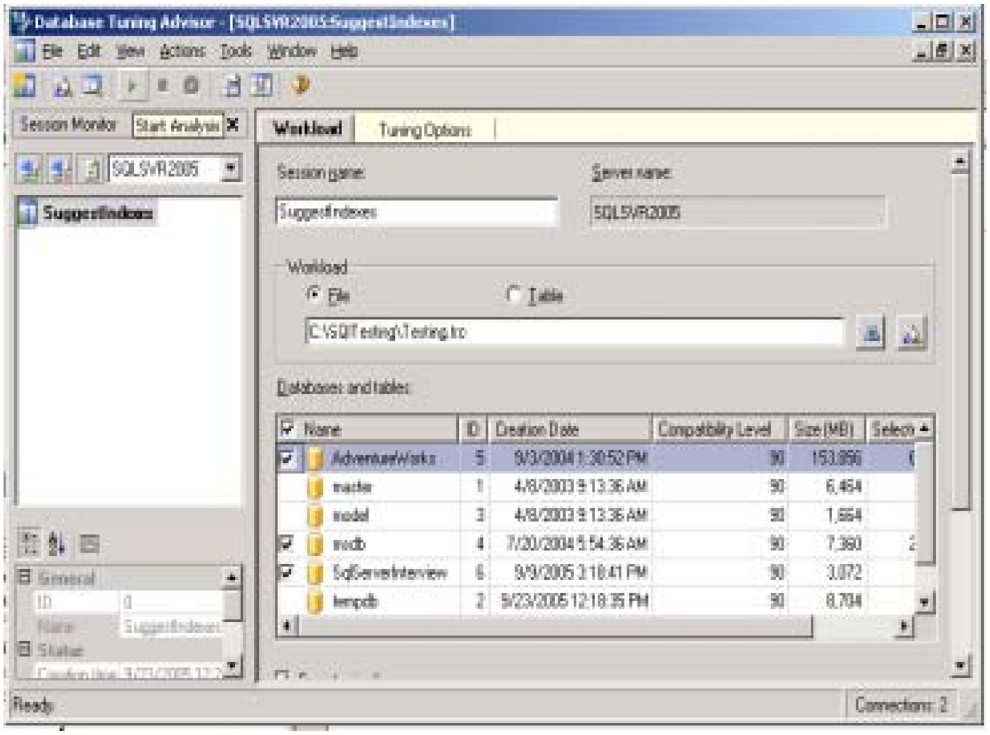
Figure 12.18 : - Creating New Session in Advisor

After you have said “New Session” you have to supply all details for the session. There are two primary requirements you need to provide to the Session:-

√ Session Name

√ “Work Load File” or “Table” (Note you can create either a trace file or you can put it in SQL Server table while running the profiler).

I have provided my “Testing.trc” file which was created when I ran the SQL profiler. You can also filter for which database you need index suggestions. At this moment I have checked all the databases. After all the details are filled in you have to click on “Green” icon with the arrow. You can see the tool tip as “Start analysis” in the image below.



Session name: SuggestIndexes

Server name: SQLSVR2005

Workload

● File ○ Table

C:\SQLTesting\Testing.trc

Databases and tables:

☑ Name	ID	Creation Date	Compatibility Level	Size (MB)	Select
☑ AdventureWorks	5	9/3/2004 1:30:52 PM	90	153,856	
□ master	1	4/8/2003 9:13:36 AM	90	6,464	
□ model	3	4/8/2003 9:13:36 AM	90	1,664	
☑ msdb	4	7/20/2004 5:54:36 AM	90	7,360	
☑ SqlServerInterview	6	9/9/2005 3:18:41 PM	90	3,072	
□ tempdb	2	9/23/2005 12:18:35 PM	90	8,704	

Figure 12.19 : - Session Details for Advisor

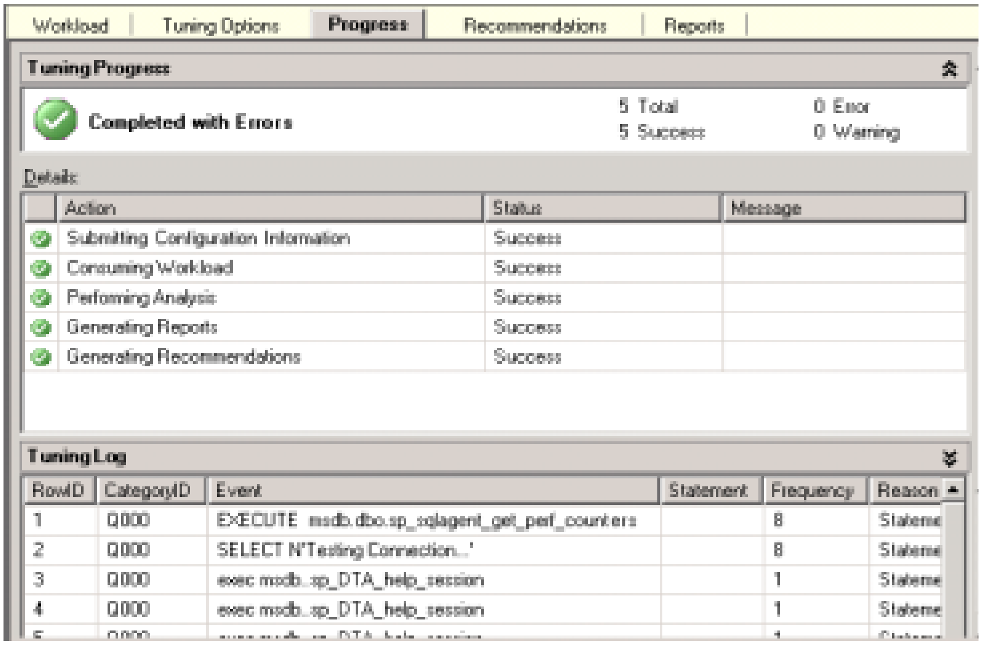
While analyzing the trace file it performs basic four major steps:-

√ Submits the configuration information.

√ Consumes the Work load data (that can be in format of a file or a database table).

- √ Start performing analysis on all the SQL executed in the trace file.
- √ Generates reports based on analysis.
- √ Finally give the index recommendations.

You can see all the above steps have run successfully which is indicated by "0 Error and 0 Warning".



Completed with Errors

5 Total
5 Success
0 Error
0 Warning

Details:

	Action	Status	Message
✓	Submitting Configuration Information	Success	
✓	Consuming Workload	Success	
✓	Performing Analysis	Success	
✓	Generating Reports	Success	
✓	Generating Recommendations	Success	

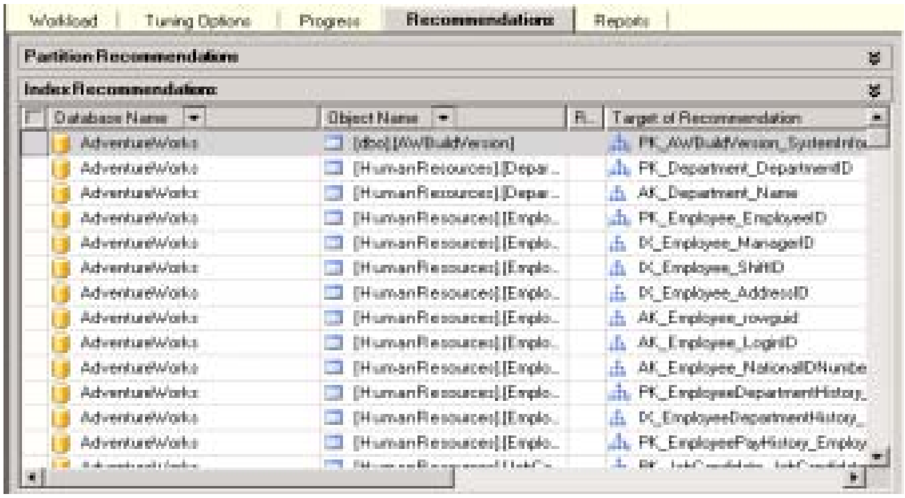
Tuning Log

RowID	CategoryID	Event	Statement	Frequency	Reason
1	Q000	EXECUTE msdb.dbo.sp_sqlagent_get_perf_counters		8	Stateme
2	Q000	SELECT N'Testing Connection...'		8	Stateme
3	Q000	exec msdb..sp_DTA_help_session		1	Stateme
4	Q000	exec msdb..sp_DTA_help_session		1	Stateme
5	Q000	exec msdb..sp_DTA_help_session		1	Stateme

Figure 12.20 : - Session completed with out Errors

Now its time to see what index recommendations SQL Server has provided us. Also note it has included two new tabs after the analysis was done "Recommendations" and "Reports".

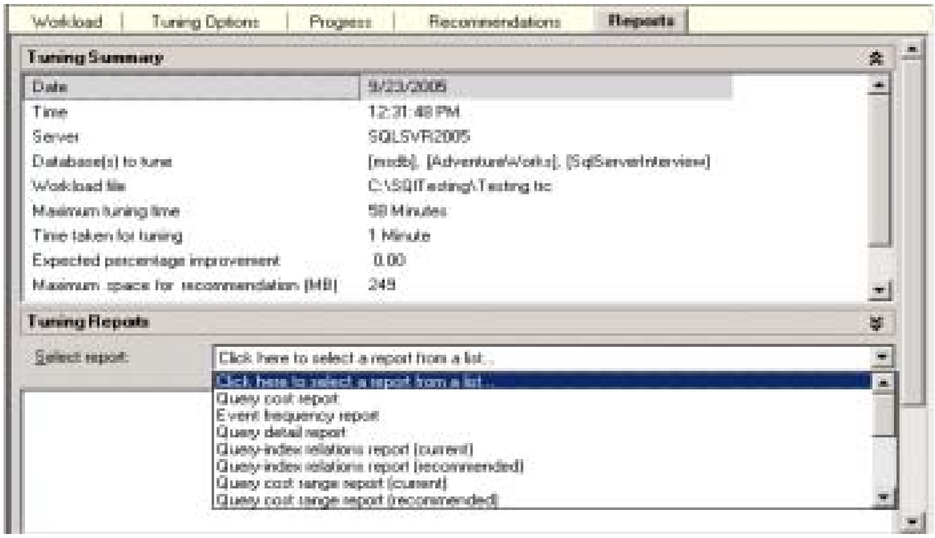
You can see on "AdventureWorks" SQL Server has given me huge recommendations. Example on "HumanResources.Department" he has told me to create index on "PK_Department_DepartmentId".



	Database Name	Object Name	R...	Target of Recommendation
	AdventureWorks	[dbo].[AWBuildVersion]		PK_AWBuildVersion_SystemInfo...
	AdventureWorks	[HumanResources].[Depar...		PK_Department_DepartmentID
	AdventureWorks	[HumanResources].[Depar...		AK_Department_Name
	AdventureWorks	[HumanResources].[Emplo...		PK_Employee_EmployeeID
	AdventureWorks	[HumanResources].[Emplo...		IX_Employee_ManagerID
	AdventureWorks	[HumanResources].[Emplo...		IX_Employee_ShiftID
	AdventureWorks	[HumanResources].[Emplo...		IX_Employee_AddressID
	AdventureWorks	[HumanResources].[Emplo...		AK_Employee_rowguid
	AdventureWorks	[HumanResources].[Emplo...		AK_Employee_LoginID
	AdventureWorks	[HumanResources].[Emplo...		AK_Employee_NationalIDNumbe
	AdventureWorks	[HumanResources].[Emplo...		PK_EmployeeDepartmentHistory_
	AdventureWorks	[HumanResources].[Emplo...		IX_EmployeeDepartmentHistory_
	AdventureWorks	[HumanResources].[Emplo...		PK_EmployeePayHistory_Employ

Figure 12.21 : - Recommendations by SQL Server

In case you want to see detail reports you can click on the “Reports” tab and there are wide range of reports which you can use to analyze how you database is performing on that “Work Load” file.



Date	9/23/2005
Time	12:31:48 PM
Server	SQLSVR2005
Database(s) to tune	[msdb], [AdventureWorks], [SqlServerInterview]
Workload file	C:\SQlTesting\Testing.trc
Maximum tuning time	58 Minutes
Time taken for tuning	1 Minute
Expected percentage improvement	0.00
Maximum space for recommendation (MB)	249

Select report: Click here to select a report from a list...

Click here to select a report from a list...
Query cost report
Event frequency report
Query detail report
Query-index relations report (current)
Query-index relations report (recommended)
Query cost range report (current)
Query cost range report (recommended)

Figure 12.22 : - Reports by Advisor

Note: - The whole point of putting this all step by step was that you have complete understanding of how to do "automatic index decision" using SQL Server. During interview one of the question's that is very sure "How do you increase speed performance of SQL Server? " and talking about the "Index Tuning Wizard" can fetch you some decent points.

(DB)What is an Execution plan?

Execution plan gives how the query optimizer will execute a give SQL query. Basically it shows the complete plan of how SQL will be executed. The whole query plan is prepared depending on lot of data for instance:-

√ What type of indexes do the tables in the SQL have?

√ Amount of data.

√ Type of joins in SQL (Inner join , Left join , Cross join , Right join etc)

Click on the ICON in SQL Server management studio as shown in figure below.

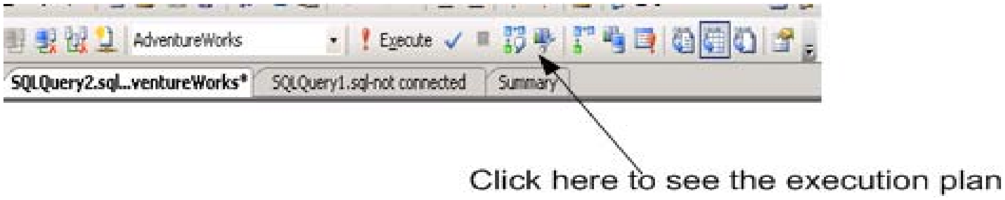

AdventureWorks
Execute
SQLQuery2.sql...ventureWorks*
SQLQuery1.sql-not connected
Summary
Click here to see the execution plan

Figure 12.23 : - Click here to see execution plan

In bottom window pane you will see the complete break up of how your SQL Query will execute. Following is the way to read it:-

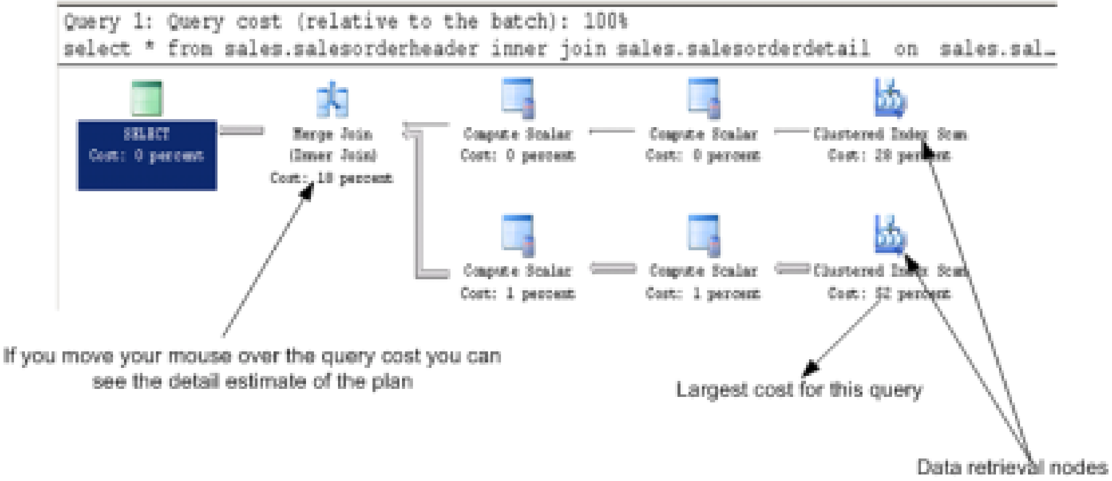
√ Data flows from left to right.

√ Any execution plan sums to total 100 %. For instance in the below figure it is 18 + 28 + 1 + 1 + 52. So the highest is taken by Index scan 52 percent. Probably we can look in to that logic and optimize this query.

√ Right most nodes are actually data retrieval nodes. I have shown them with arrows the two nodes.

√ In below figure you can see some arrows are thick and some are thin. More the thickness more the data is transferred.

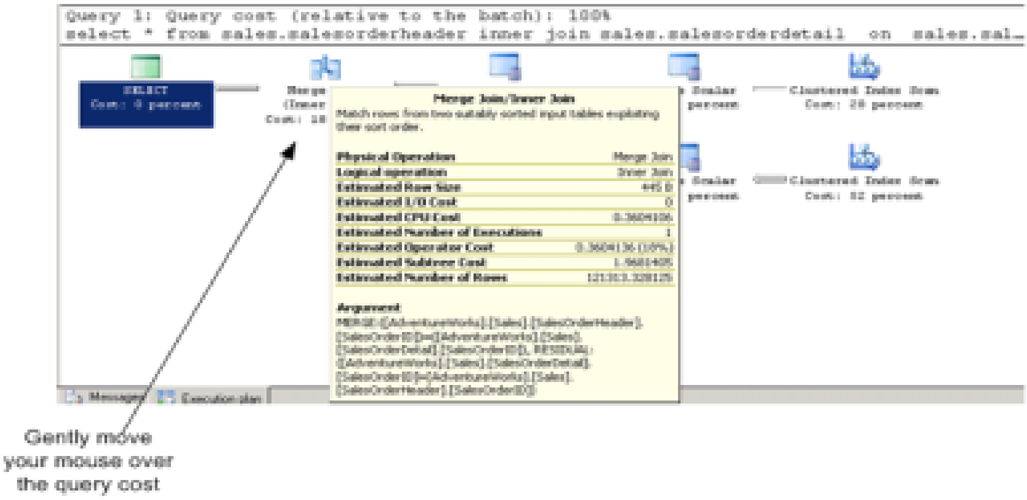
√ There are three types of join logic nested join, hash join and merge join.



```
Query 1: Query cost (relative to the batch): 100%
select * from sales.salesorderheader inner join sales.salesorderdetail  on  sales.sal…
```

Figure 12.24 : - Largest Cost Query

If you move your mouse gently over any execution strategy you will see a detail breakup of how that node is distributed.



```
Query 1: Query cost (relative to the batch): 100%
select * from sales.salesorderheader inner join sales.salesorderdetail  on  sales.sal…
```

Merge Join/Inner Join

Match rows from two suitably sorted input tables exploiting their sort order.

Physical Operation	Merge Join
Logical operation	Inner Join
Estimated Row Size	445 B
Estimated I/O Cost	0
Estimated CPU Cost	0.3609106
Estimated Number of Executions	1
Estimated Operator Cost	0.3609136 (18%)
Estimated Subtree Cost	1.9681405
Estimated Number of Rows	121313.328125

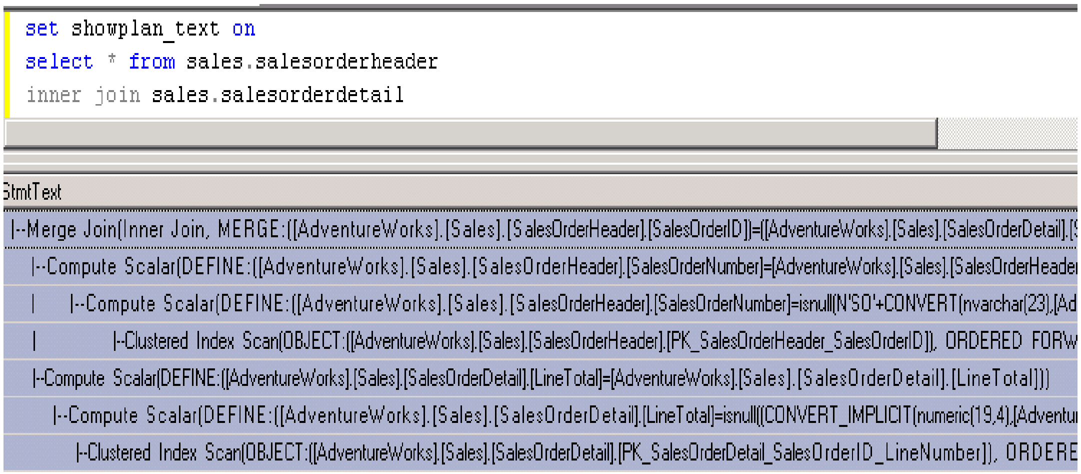
Argument

MERGE:([AdventureWorks].[Sales].[SalesOrderHeader].[SalesOrderID])=([AdventureWorks].[Sales].[SalesOrderDetail].[SalesOrderID]), RESIDUAL:([AdventureWorks].[Sales].[SalesOrderDetail].[SalesOrderID]=[AdventureWorks].[Sales].[SalesOrderHeader].[SalesOrderID])

Figure 12.25 : - Complete Break up estimate

How do you see the SQL plan in textual format?

Execute the following “set showplan_text on” and after that execute your SQL, you will see a textual plan of the query. In the first question what I discussed was a graphical view of the query plan. Below is a view of how a textual query plan looks like. In older versions of SQL Server where there was no way of seeing the query plan graphically “SHOWPLAN” was the most used. Today if any one is using it that I think he is doing a show business or a new come learner.



```
set showplan_text on
select * from sales.salesorderheader
inner join sales.salesorderdetail
```

```
StmtText
|--Merge Join(Inner Join, MERGE:([AdventureWorks].[Sales].[SalesOrderHeader].[SalesOrderID])=([AdventureWorks].[Sales].[SalesOrderDetail].[S
   |--Compute Scalar(DEFINE:([AdventureWorks].[Sales].[SalesOrderHeader].[SalesOrderNumber]=[AdventureWorks].[Sales].[SalesOrderHeade
   |    |--Compute Scalar(DEFINE:([AdventureWorks].[Sales].[SalesOrderHeader].[SalesOrderNumber]=isnull(N'SO'+CONVERT(nvarchar(23),[Ad
   |         |--Clustered Index Scan(OBJECT:([AdventureWorks].[Sales].[SalesOrderHeader].[PK_SalesOrderHeader_SalesOrderID]), ORDERED FORW
   |--Compute Scalar(DEFINE:([AdventureWorks].[Sales].[SalesOrderDetail].[LineTotal]=[AdventureWorks].[Sales].[SalesOrderDetail].[LineTotal]))
        |--Compute Scalar(DEFINE:([AdventureWorks].[Sales].[SalesOrderDetail].[LineTotal]=isnull((CONVERT_IMPLICIT(numeric(19,4),[Adventu
             |--Clustered Index Scan(OBJECT:([AdventureWorks].[Sales].[SalesOrderDetail].[PK_SalesOrderDetail_SalesOrderID_LineNumber]), ORDERE
```

Figure 12.26 : - Textual Query Plan View

(DB)What is nested join, hash join and merge join in SQL Query plan?

A join is whenever two inputs are compared to determine and output. There are three basic types of strategies for this and they are: nested loops join, merge join and hash join. When a join happens the optimizer determines which of these three algorithms is best to use for the given problem, however any of the three could be used for any join. All of the costs related to the join are analyzed the most cost efficient algorithm is picked for use. These are in-memory loops used by SQL Server.

Nested Join

If you have less data this is the best logic. It has two loops one is the outer and the other is the inner loop. For every outer loop, its loops through all records in the inner loop. You can see the two loop inputs given to the logic. The top index scan is the outer loop and bottom index seek is the inner loop for every outer record.

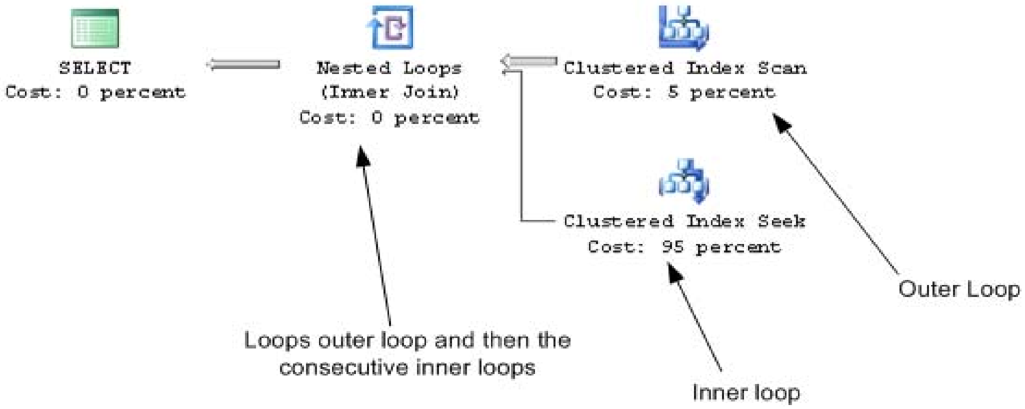


Figure 12.27 : - Nested joins

It's like executing the below logic:-

```
For each outer records
    For each inner records
    Next
Next
```

So you visualize that if there fewer inner records this is a good solution.

Hash Join

Hash join has two input "Probe" and "Build" input. First the "Build" input is processed and then the "Probe" input. Which ever input is smaller is the "Build" input. SQL Server first builds a hash table using the build table input. After that he loops through the probe input and finds the matches using the hash table created previously using the build table and does the processing and gives the output.

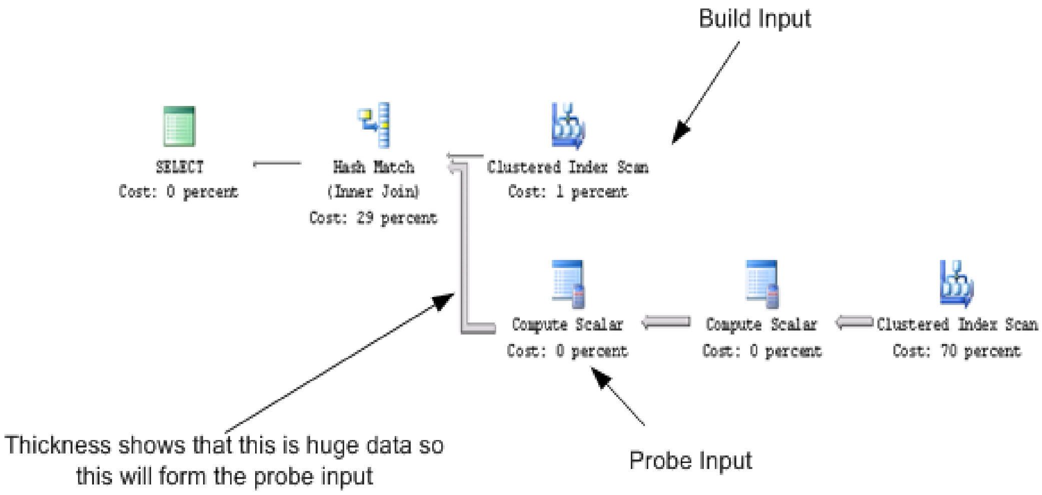


Figure 12.28 : - Hash Join

Merge Join

In merge joins both the inputs are sorted on the merge columns. Merge columns are determined depending on the inner join defined in SQL. Since each input join is sorted merge join takes input and compares for equality. If there is equality then matching row is produced. This is processed till the end of rows.

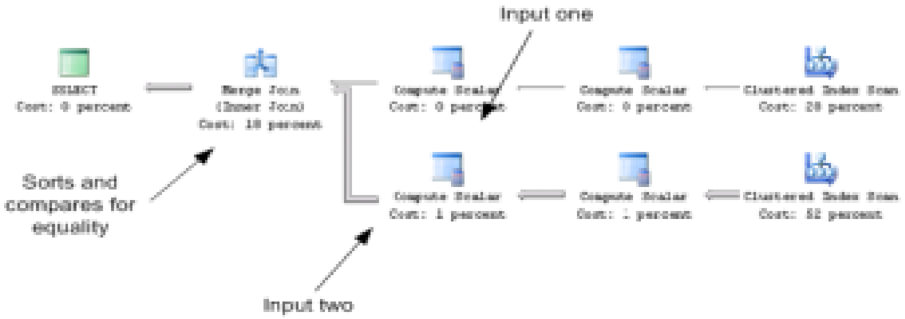


Figure 12.29: - Merge Join

What joins are good in what situations?

Nested joins best suited if the table is small and it's a must the inner table should have an index.

Merge joins best of large tables and both tables participating in the joins should have indexes.

Hash joins best for small outer tables and large inner tables. Not necessary that tables should have indexes, but would be better if outer table has indexes.

> *Note: - Previously we have discussed about table scan and index scan do revise it which is also important from the aspect of reading query plan.*

(DB)What is RAID and how does it work ?

Redundant Array of Independent Disks (RAID) is a term used to describe the technique of improving data availability through the use of arrays of disks and various data-striping methodologies. Disk arrays are groups of disk drives that work together to achieve higher data-transfer and I/O rates than those provided by single large drives. An array is a set of multiple disk drives plus a specialized controller (an array controller) that keeps track of how data is distributed across the drives. Data for a particular file is written in segments to the different drives in the array rather than being written to a single drive.

For speed and reliability, it's better to have more disks. When these disks are arranged in certain patterns and use a specific controller, they are called a Redundant Array of Inexpensive Disks (RAID) set. There are several numbers associated with RAID, but the most common are 1, 5 and 10.

RAID 1 works by duplicating the same writes on two hard drives. Let's assume you have two 20 Gigabyte drives. In RAID 1, data is written at the same time to both drives. RAID1 is optimized for fast writes.

RAID 5 works by writing parts of data across all drives in the set (it requires at least three drives). If a drive failed, the entire set would be worthless. To combat this problem, one of the drives stores a "parity" bit. Think of a math problem, such as 3 + 7 = 10. You can think of the drives as storing one of the numbers, and the 10 is the parity part. By removing any one of the numbers, you can get it back by referring to the other two, like this: 3 + X = 10. Of course, losing more than one could be evil. RAID 5 is optimized for reads.

RAID 10 is a bit of a combination of both types. It doesn't store a parity bit, so it's fast, but it duplicates the data on two drives to be safe. You need at least four drives for RAID 10. This type of RAID is probably the best compromise for a database server.

Note :- It's difficult to cover complete aspect of RAID in this book.It's better to take some decent SQL SERVER book for in detail knowledge , but yes from interview aspect you can probably escape with this answer.

13. Transaction and Locks

What is a "Database Transactions "?

It's a unit of interaction within a database which should be independent of other transactions.

What is ACID?

"ACID" is a set of rule which are laid down to ensure that "Database transaction" is reliable. Database transaction should principally follow ACID rule to be safe. "ACID" is an acronym which stands for:-

√ Atomicity

A transaction allows for the grouping of one or more changes to tables and rows in the database to form an atomic or indivisible operation. That is, either all of the changes occur or none of them do. If for any reason the transaction cannot be completed, everything this transaction changed can be restored to the state it was in prior to the start of the transaction via a rollback operation.

√ Consistency

Transactions always operate on a consistent view of the data and when they end always leave the data in a consistent state. Data may be said to be consistent as long as it conforms to a set of invariants, such as no two rows in the customer table have the same customer id and all orders have an associated customer row. While a transaction executes these invariants may be violated, but no other transaction will be allowed to see these inconsistencies, and all such inconsistencies will have been eliminated by the time the transaction ends.

√ Isolation

To a given transaction, it should appear as though it is running all by itself on the database. The effects of concurrently running transactions are invisible to this transaction, and the effects of this transaction are invisible to others until the transaction is committed.

√ Durability

Once a transaction is committed, its effects are guaranteed to persist even in the event of subsequent system failures. Until the transaction commits, not only are any changes made by that transaction not durable, but are guaranteed not to persist in the face of a system failure, as crash recovery will rollback their effects.

The simplicity of ACID transactions is especially important in a distributed database environment where the transactions are being made simultaneously.

What is “Begin Trans”, “Commit Tran”, “Rollback Tran” and “Save Tran”?

Begin Tran: - It’s a point which says that from this point onwards we are starting the transaction.

Commit Tran: - This is a point where we say we have completed the transaction. From this point the data is completely saved in to database.

Rollback Tran: - This point is from where we go back to the start point that i.e. “Begin Tran” stage.

Save Tran: - It’s like a bookmark for rollback to come to some specified state. When we say “rollback Tran” we go back directly to “Begin Tran”, but what if we want to go back to some specific point after “Begin Tran”. So “Save Tran” is like book marks which can be used to come back to that state rather than going directly to the start point.

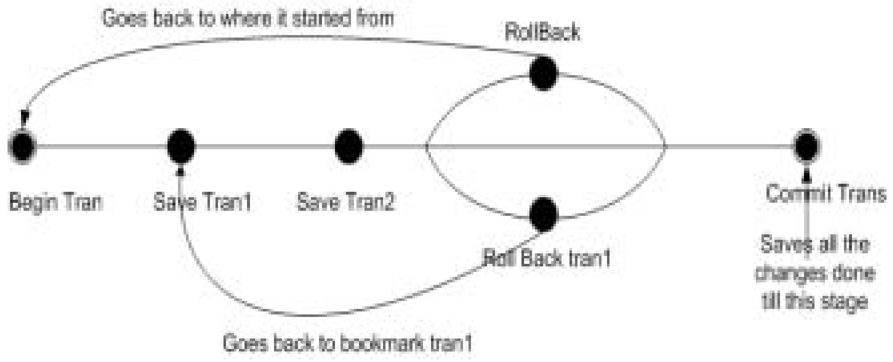


Figure 13.1 : - Different Types of Transaction Points

There are two paths defined in the transaction one which rollbacks to the main state and other which rollbacks to a “tran1”. You can also see “tran1” and “tran2” are planted in multiple places as book mark to roll-back to that state.

Brushing up the syntaxes

To start a transaction

BEGIN TRAN Tran1

Creates a book point

SAVE TRAN PointOne

This will roll back to point one

ROLLBACK TRAN PointOne

This commits complete data right when Begin Tran point

COMMIT TRAN Tran1

(DB)What are “Checkpoint’s” in SQL Server?

In normal operation everything that is done by SQL Server is not committed directly to database. All operation is logged in to “Transaction Log” first. “CheckPoint” is a point which signals SQL Server to save all data to main database. If there are no “CheckPoints” then the log file will get full.

You can use the “CHECKPOINT” command to commit all data in to SQL SERVER. “CheckPoint” command is also fired when you shut the SQL Server, that’s why it takes long time to shut down.

(DB)What are “Implicit Transactions”?

In order to initiate a transaction we use “Begin Tran Tran1” and later when we want to save complete data we use “Commit Tran <TransactionName>”. In SQL Server you can define to start transaction by default i.e. with out firing “Begin Tran Tr1”. You can set this by using:-

SET IMPLICIT_TRANSACTIONS ON

So after the above command is fired any SQL statements that are executed will be by default in transaction. You have to only fire “Commit Tran <Transaction Name>” to close the transaction.

(DB)Is it good to use “Implicit Transactions”?

No. If in case developer forgets to shoot the “Commit Tran” it can open lot of transaction’s which can bring down SQL Server Performance.

What is Concurrency?

In multi-user environment if two users are trying to perform operations (Add, Modify and Delete) at the same time is termed as “Concurrency”. In such scenarios there can be lot of conflicts about the data consistency and to follow ACID principles.

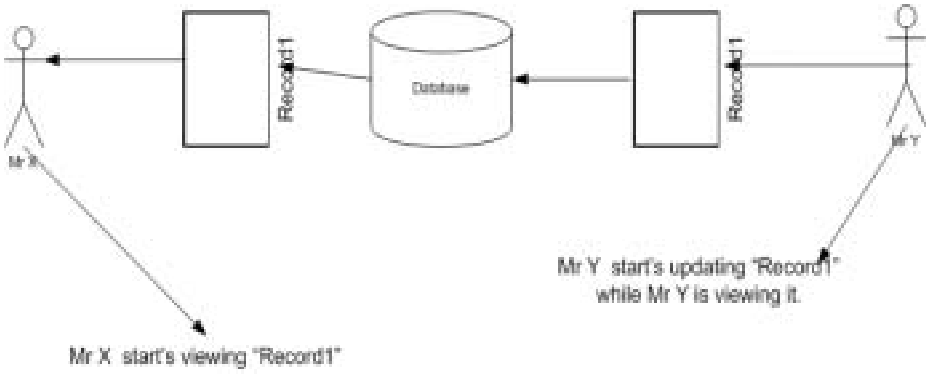

Record1
Database
Record1
Mr X
Mr Y
Mr Y start's updating "Record1" while Mr Y is viewing it.
Mr X start's viewing "Record1"

Figure 13.2 : - Concurrency Problem

For instance the above figure depicts the concurrency problem. “Mr X” started viewing “Record1” after some time “MR Y” picks up “Record1” and starts updating it. So “Mr X” is viewing data which is not consistent with the actual database.

How can we solve concurrency problems?

Concurrency problems can be solved by implementing proper “Locking strategy”. In short by “Locking”. Locks prevents action on a resource to be performed when some other resource is already performing some action on it.

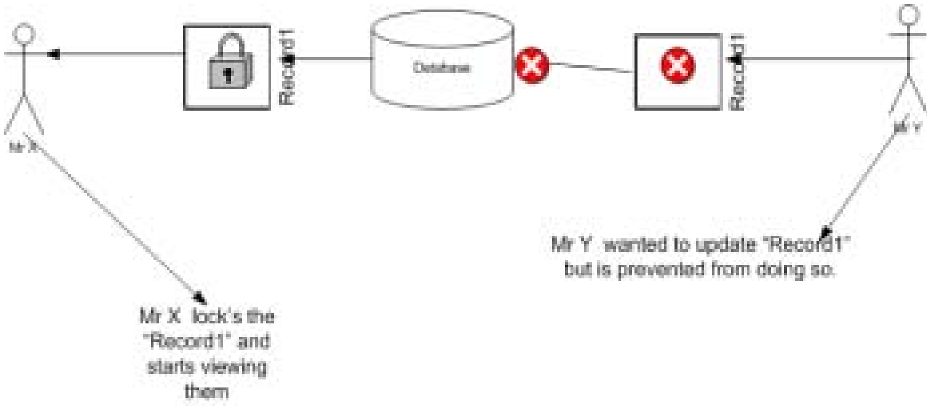

Record1
Database
Record1
Mr X
Mr Y
Mr Y wanted to update "Record1" but is prevented from doing so.
Mr X lock's the "Record1" and starts viewing them

Figure 13.3: - Locking implemented

In our first question we saw the problem above is how locking will work. "Mr. X" retrieves "Record1" and locks it. When "Mr Y" comes in to update "Record1" he can not do it as it's been locked by "Mr X".

> *Note: - What I have showed is small glimpse, in actual situations there are different types of locks we will going through each in the coming questions.*

What kind of problems occurs if we do not implement proper locking strategy?

There are four major problems that occur:-

√ Dirty Reads

√ Unrepeatable reads

√ Phantom reads

√ Lost updates

What are "Dirty reads"?

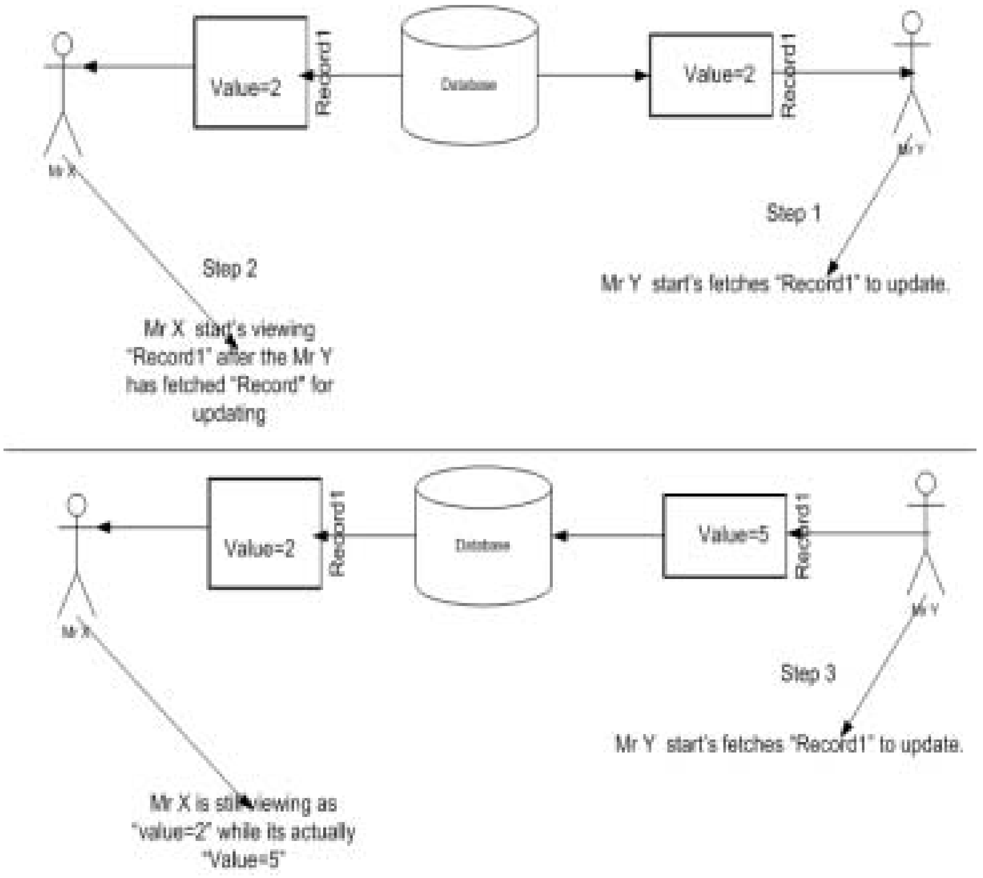

Value=2
Record1
Database
Value=2
Record1
Mr X
Mr Y
Step 1
Mr Y start's fetches "Record1" to update.
Step 2
Mr X start's viewing "Record1" after the Mr Y has fetched "Record" for updating
Value=2
Record1
Database
Value=5
Record1
Mr X
Mr Y
Step 3
Mr Y start's fetches "Record1" to update.
Mr X is still viewing as "value=2" while its actually "Value=5"

Figure 13.4 : - Dirty Reads

“Dirty Read” occurs when one transaction is reading a record which is part of a half finished work of other transaction. Above figure defines the “Dirty Read” problem in a pictorial format. I have defined all activities in Step’s which shows in what sequence they are happening (i.e. Step1, Step 2 etc).

√ Step1: -“Mr. Y” Fetches “Record” which has “Value=2” for updating it.

√ Step2:- In mean time “Mr. X” also retrieves “Record1” for viewing. He also sees it as “Value=2”.

- √ Step3:- While “Mr. X” is viewing the record, concurrently “Mr. Y” updates it as “Value=5”. Boom… the problem “Mr. X” is still seeing it as “Value=3”, while the actual value is “5”.

What are “Unrepeatable reads”?

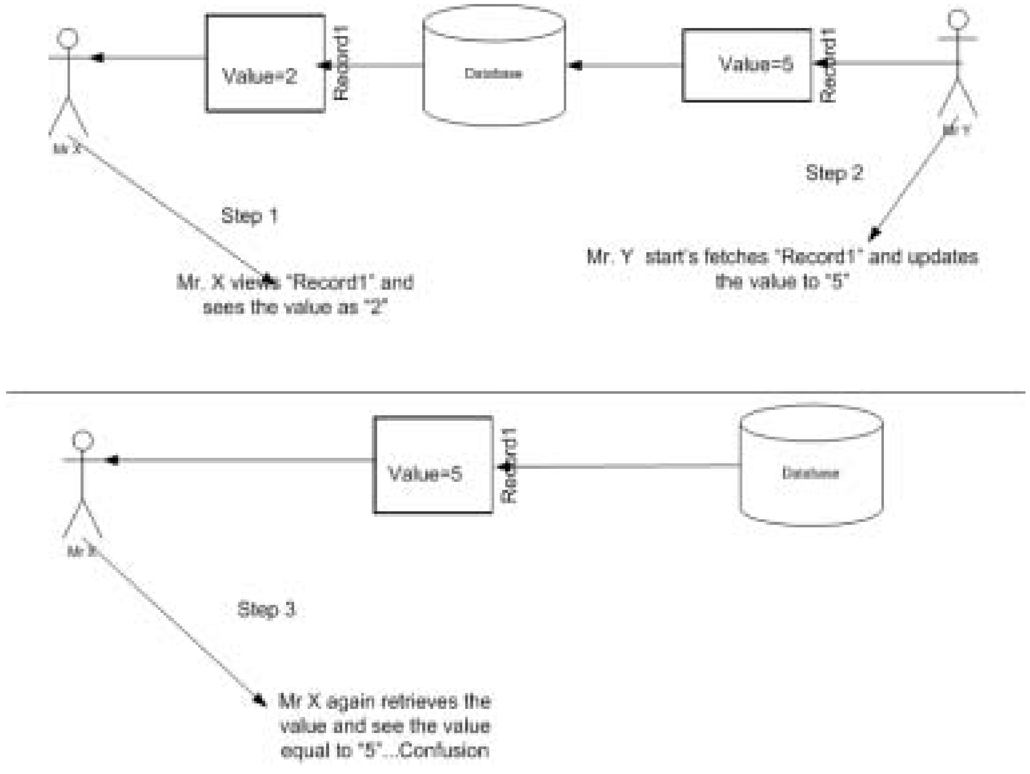

Value=2
Record1
Database
Value=5
Record1
Mr X
Mr Y
Step 1
Mr. X views "Record1" and sees the value as "2"
Step 2
Mr. Y start's fetches "Record1" and updates the value to "5"
Value=5
Record1
Database
Mr X
Step 3
Mr X again retrieves the value and see the value equal to "5"...Confusion

Figure 13.5 : - Unrepeatable Read

In every data read if you get different values then it’s an “Unrepeatable Read” problem. Lets try to iterate through the steps of the above given figure:-

- √ Step1:- “Mr. X” get “Record” and sees “Value=2”.
- √ Step2:- “Mr. Y” meantime comes and updates “Record1” to “Value=5”.

- √ Step3:- “Mr. X” again gets “Record1” ohh... values are changed “2” … Confusion.

What are “Phantom rows”?

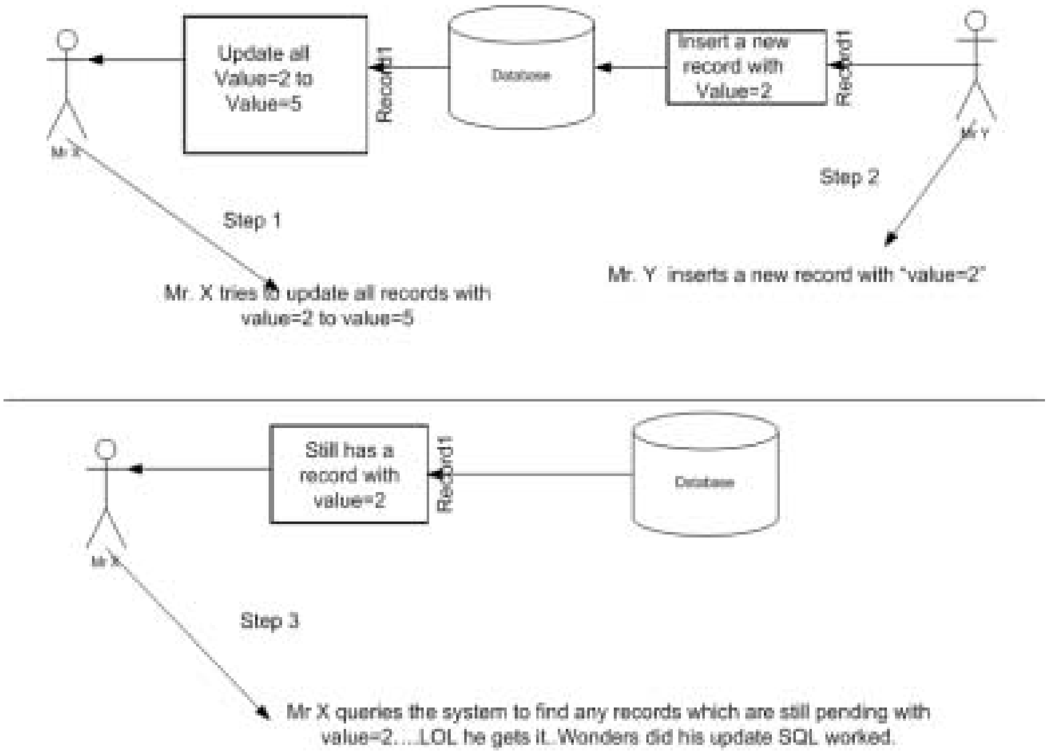

Update all Value=2 to Value=5
Record1
Database
Insert a new record with Value=2
Record1
Mr X
Mr Y
Step 1
Step 2
Mr. X tries to update all records with value=2 to value=5
Mr. Y inserts a new record with "value=2"
Still has a record with value=2
Record1
Database
Mr X
Step 3
Mr X queries the system to find any records which are still pending with value=2....LOL he gets it..Wonders did his update SQL worked.

Figure 13.6 : - Phantom Rows

If “UPDATE” and “DELETE” SQL statements seems to not affect the data then it can be “Phantom Rows” problem.

- √ Step1:- “Mr. X” updates all records with “Value=2” in “record1” to “Value=5”.
- √ Step2:- In mean time “Mr. Y” inserts a new record with “Value=2”.

- √ Step3:- “Mr. X” wants to ensure that all records are updated, so issues a select command for “Value=2”….surprisingly find records which “Value=2”…

So “Mr. X” thinks that his “UPDATE” SQL commands are not working properly.

What are “Lost Updates”?

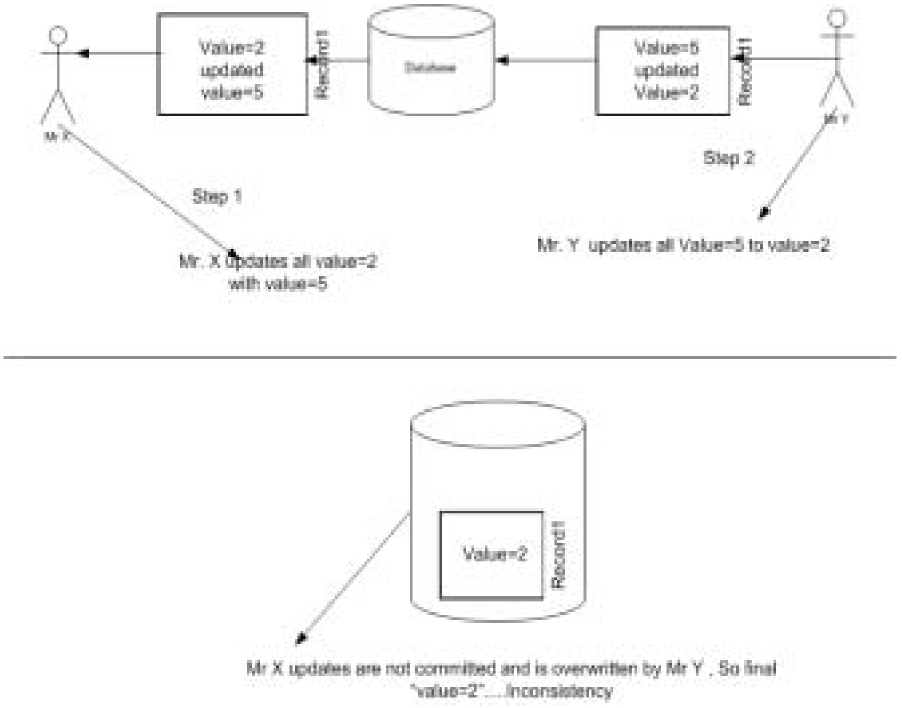

Mr X
Value=2
updated
value=5
Record1
Database
Value=5
updated
Value=2
Record1
Mr Y
Step 1
Mr. X updates all value=2
with value=5
Step 2
Mr. Y updates all Value=5 to value=2
Value=2
Record1
Mr X updates are not committed and is overwritten by Mr Y . So final
"value=2"....Inconsistency

Figure 13.7 : - Lost Updates

“Lost Updates” are scenario where one updates which is successfully written to database is over-written with other updates of other transaction. So let’s try to understand all the steps for the above figure:-

- √ Step1:- “Mr. X” tries to update all records with “Value=2” to “Value=5”.
- √ Step2:- “Mr. Y” comes along the same time and updates all records with “Value=5” to “Value=2”.

- √ Step3 :- Finally the “Value=2” is saved in database which is inconsistent according to “Mr. X” as he thinks all the values are equal to “2”.

What are different levels of granularity of locking resources?

Extent:-Extent is made of one or more pages. So all pages are locked and data inside those pages are also locked.

Page: - Page lock puts lock on all data, table and indexes in the page.

Database:-If you are making database structure changes then whole database will be locked.

Table:-We can also lock object at a table level. That means indexes related to it also are locked.

Key: - If you want to lock a series a key of indexes you can place lock on those group of records.

Row or Row Identifier (RID):-This is the lowest level of locking. You can lock data on a row level.

What are different types of Locks in SQL Server?

Below are the different kinds of locks in SQL Server:-

- √ Shared Locks (S): - These types of locks are used while reading data from SQL Server. When we apply a Shared lock on a record, then other users can only read the data, but modifying the data is not allowed. Other users can add in new records to the table but can not modify the row which has shared lock applied to it.
- √ Exclusive Locks (X):- These types of lock are not compatible with any other type of locks. As the name suggests any resource which is having exclusive locks will not allow any locks to take over it. Nor it can take over any other type of locks. For instance if a resource is having “Shared” lock on a resource you can not make an “Exclusive lock” over the resource. They are specially used for “Insert”, “Update” and “Delete” operations.
- √ Update Locks (U):- “Update” locks are in a mid-level between “Shared” and “Exclusive” locks. When SQL Server wants to modify data and later promote

the “Update” locks to “Exclusive” locks then “Update” locks are used. “Update” locks are compatible with “Shared” locks.

Ok just to give a brief of how the above three locks will move in actual environment. Below is the figure which shows sequence of “SQL” steps executed and the locks they are trying to acquire on it.

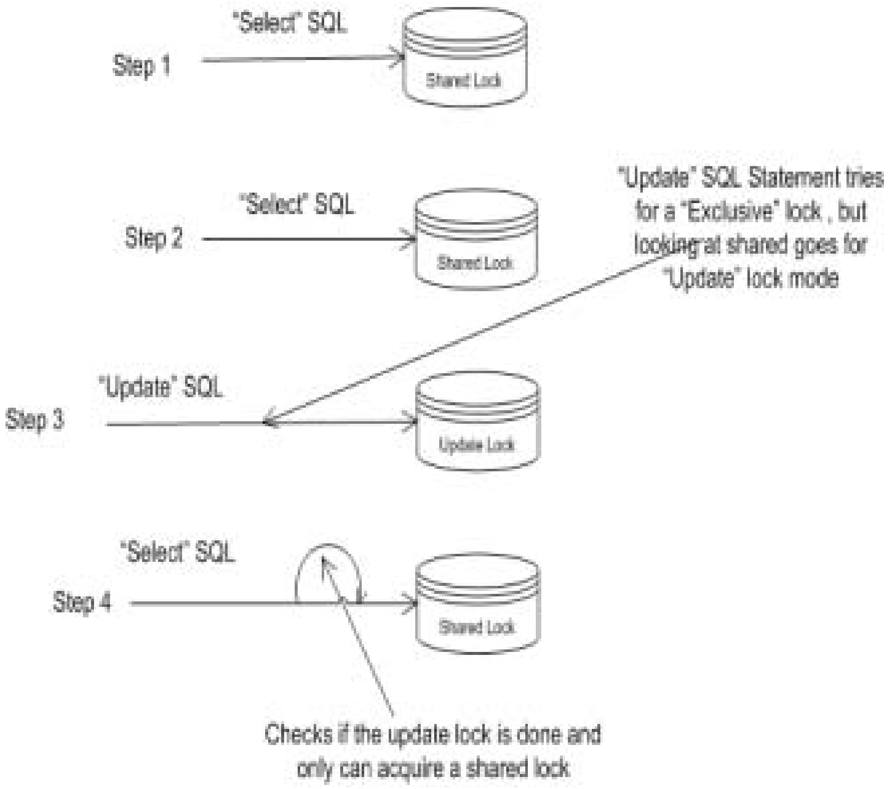

"Select" SQL
Step 1
Shared Lock
"Select" SQL
Step 2
Shared Lock
"Update" SQL Statement tries for a "Exclusive" lock , but looking at shared goes for "Update" lock mode
"Update" SQL
Step 3
Update Lock
"Select" SQL
Step 4
Shared Lock
Checks if the update lock is done and only can acquire a shared lock

Figure 13.8 : - Different Lock sequence in actual scenarios

Step1:- First transaction issues a “SELECT” statement on the resource, thus acquiring a “Shared Lock” on the data.

Step2:- Second transaction also executes a “SELECT” statement on the resource which is permitted as “Shared” lock is honored by “Shared” lock.

Step3:- Third transaction tries to execute an “Update” SQL statement. As it’s a “Update” statement it tries to acquire an “Exclusive”. But because we already have a “Shared” lock on it, it acquires a “Update” lock.

Step4:- The final transaction tries to fire “Select” SQL on the data and try to acquire a “Shared” lock. But it can not do until the “Update” lock mode is done.

So first “Step4” will not be completed until “Step3” is not executed. When “Step1” and “Step2” is done “Step3” make the lock in to “Exclusive” mode and updates the data. Finally “Step4” is completed.

- √ Intent Locks: - When SQL Server wants to acquire a “Shared” lock or an “Exclusive” lock below the hierarchy you can use “Intent” locks. For instance one of the transactions has acquired as table lock and you want to have row level lock you can use “Intent” locks. Below are different flavors of “Intent” locks but with one main intention to acquire locks on lower level:-
  - Intent locks include:
  - Intent shared (IS)
  - Intent exclusive (IX)
  - Shared with intent exclusive (SIX)
  - Intent update (IU)
  - Update intent exclusive (UIX)
  - Shared intent update (SIU)
- √ Schema Locks: - Whenever you are doing any operation which is related to “Schema” operation this lock is acquired. There are basically two types of flavors in this :-
  - Schema modification lock (Sch-M):- Any object structure change using ALTER, DROP, CREATE etc will have this lock.
  - Schema stability lock (Sch-S) – This lock is to prevent “Sch-M” locks. These locks are used when compiling queries. This lock does not block any transactional locks, but when the Schema stability (Sch-S) lock is used, the DDL operations cannot be performed on the table.

- √ Bulk Update locks:-Bulk Update (BU) locks are used during bulk copying of data into a table. For example when we are executing batch process in midnight over a database.
- √ Key-Range locks: - Key-Range locks are used by SQL Server to prevent phantom insertions or deletions into a set of records accessed by a transaction.

Below are different flavors of "Key-range" locks

- ■ RangeI_S
- ■ RangeI_U
- ■ RangeI_X
- ■ RangeX_S
- ■ RangeX_U

What are different Isolation levels in SQL Server?

Twist: - What is an Isolation level in SQL Server?

Locking protects your data from any data corruption or confusion due to multi-user transactions. Isolation level determines how sensitive are your transaction in respect to other transactions. How long the transaction should hold locks to protect from changes done by other transactions. For example if you have long exclusive transaction, then other transactions who want to take over the transaction have to wait for quiet long time. So, isolation level defines the contract between two transactions how they will operate and honor each other in SQL Server. In short how much is on transaction isolated from other transaction.

What are different types of Isolation levels in SQL Server?

Following are different Isolation levels in SQL Server:-

- √ READ COMMITTED
- √ READ UNCOMMITTED
- √ REPEATABLE READ
- √ SERIALIZABLE

Note: - By default SQL Server has "READ COMMITTED" Isolation level.

Read Committed

Any "Shared" lock created using "Read Committed" will be removed as soon as the SQL statement is executed. So if you are executing several "SELECT" statements using "Read Committed" and "Shared Lock", locks are freed as soon as the SQL is executed.

But when it comes to SQL statements like "UPDATE / DELETE AND INSERT" locks are held during the transaction.

With "Read Committed" you can prevent "Dirty Reads" but "Unrepeatable" and "Phantom" still occurs.

Read Uncommitted

This Isolation level says "do not apply any locks". This increases performance but can introduces "Dirty Reads". So why is this Isolation level in existence?. Well sometimes when you want that other transaction do not get affected and you want to draw some blurred report , this is a good isolation level to opt for.

Repeatable Read

This type of read prevents "Dirty Reads" and "Unrepeatable reads".

Serializable

It's the king of everything. All concurrency issues are solved by using "Serializable" except for "Lost update". That means all transactions have to wait if any transaction has a "Serializable" isolation level.

Note: - Syntax for setting isolation level:-

SET TRANSACTION ISOLATION LEVEL <READ COMMITTED|READ UNCOMMITTED|REPEATABLE READ|SERIALIZABLE>

If you are using COM+ what "Isolation" level is set by default?

In order to maintain integrity COM+ and MTS set the isolation level to "SERIALIZABLE".

What are “Lock” hints?

This is for more control on how to use locking. You can specify how locking should be applied in your SQL queries. This can be given by providing optimizer hints. “Optimizer” hints tells SQL Server that escalate me to this specific lock level. Example the below query says to put table lock while executing the SELECT SQL.

*SELECT * FROM MasterCustomers WITH (TABLOCKX)*

What is a “Deadlock” ?

Deadlocking occurs when two user processes have locks on separate objects and each process is trying to acquire a lock on the object that the other process has. When this happens, SQL Server ends the deadlock by automatically choosing one and aborting the process, allowing the other process to continue. The aborted transaction is rolled back and an error message is sent to the user of the aborted process. Generally, the transaction that requires the least amount of overhead to rollback is the transaction that is aborted.

What are the steps you can take to avoid “Deadlocks” ?

Below are some guidelines for avoiding “Deadlocks” :-

- √ Make database normalized as possible. As more small pieces the system is better granularity you have to lock which can avoid lot of clashing.
- √ Do not lock during user is making input to the screen, keep lock time as minimum as possible by good design.
- √ As far as possible avoid cursors.
- √ Keep transactions as short as possible. One way to help accomplish this is to reduce the number of round trips between your application and SQL Server by using stored procedures or keeping transactions with a single batch. Another way of reducing the time a transaction takes to complete is to make sure you are not performing the same reads over and over again. If you do need to read the same data more than once, cache it by storing it in a variable or an array, and then re-reading it from there.
- √ Reduce lock time. Try to develop your application so that it grabs locks at the latest possible time, and then releases them at the very earliest time.

√ If appropriate, reduce lock escalation by using the ROWLOCK or PAGLOCK.

√ Consider using the NOLOCK hint to prevent locking if the data being locked is not modified often.

√ If appropriate, use as low of isolation level as possible for the user connection running the transaction.

√ Consider using bound connections.

(DB)How can I know what locks are running on which resource?

In order to see the current locks on an “object” or a “process” expand the management tree and right click on “Activity” tab. So in case you want to see “dead locks” or you want to terminate the “dead lock” you can use this facility to get a bird-eye view.

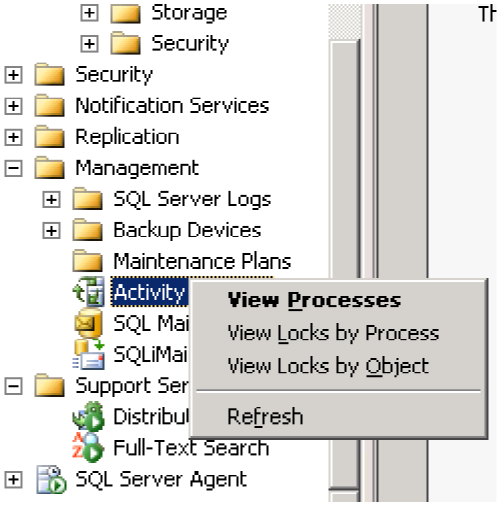


Figure 13.9 : - View current locks in SQL Server