Project Name: Backup, Restore, Snapshot, Jobs, Import/Export, shrinking, Compression, Partitioning and Indexes

Deliverables:

* Class Presentation
* Documentation

Due Date: 2/26/2021

# Overview: Project for all the groups

Your group has been given a project to create a database for a sales company. As a database administrator, you need to put in place a disaster recovery solution for this database. See below information.

Database name: SaleSpan

RPO: 15 Minutes

RTO: 45 Minutes

Retention Policy: 30days

# Part one: Understanding of key concepts

1. Explain the process in creating a database including the tables

a. Name of the table (should reflect its content).

b. Will a schema be implemented on the table

c. Number of columns that will be present on the table

d. What type of data will be inserted into the table

e. What server/database will this table be in

f. Who will have access to this table

g. What type of recovery model is the database on in which the table will be stored

h. What type of constraints will be implemented on the table

1. With the SLA above, explain the difference between RPO and RTO

An RPO (Recovery point objective) is an agreement that states in the event of a database for file corruption or other disaster, the maximum amount of data loss within a time span as dictated by this agreement (15 minutes). An RTO (recovery time objective) is an agreement that states that in the event of a database corruption or other disaster, the maximum amount of downtime the application can tolerate is a set amount of time (45 minutes).

1. What are Constraints? List the various types of constraint in a SQL database

a. Primary Key constraint

b. Foreign Key constraint

c. Unique Key constraint

d. Not Null

e. Check

f. Default

1. What is database normalization: Explain the various normal forms?

Normalization is the process by which data is organized in a database. It organizes data into tables to prevent redundancy and maximize space. We do this using constraints most of the time.

1. What do you understand by referential integrity and how is it enforced in a SQL database?

Referential integrity requires that a foreign key must have a matching primary key or it must be null. This constraint is specified between two tables (parent and child); it maintains the correspondence between rows in these tables. It means the reference from a row in one table to another table must be valid.

1. From your understanding what problems do RDBMS solve and how?

RDBMS help organize data using the interrelationship they have. They help organize these data in tables who are in turn saved with related files. Related files are stored together in related filegroups and all these filegroups are stored in a database.

1. Can you list and explain the properties of a transaction?

A transaction is a very small unit of a program and it may contain several low-level tasks. A transaction in a database system must maintain Atomicity (each statement in a transaction is treated as a single unit), Consistency (ensures that transactions only make changes to tables in predefined, predictable ways), Isolation (when multiple users are reading and writing from the same table all at once, isolation of their transactions ensures that the concurrent transactions don’t interfere with or affect one another), and Durability (ensures that changes to your data made by successfully executed transactions will be saved, even in the event of system failure) − commonly known as ACID properties − in order to ensure accuracy, completeness, and data integrity.

1. Can you give us a detailed overview of some of the projects you have worked on in your current environment in relation to one of the key concepts in this section?

Worked on creating tables using best practices of naming them depending on the contents it holds. We also linked tables using foreign key constraints. To ensure just single values are added onto the table, we used primary key and unique key constraints to ensure each entry is unique.

LAB:

Consider the Students table, with the primary key underlined, and the following data:

Students:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Alpha | Name | Email | Courses | GradePoints |
| 100111 | John Doe | doe@usna.edu | NN204, SI204, IT221 | 2,3,3 |
| 092244 | Matt Smith | smith@usna.edu | SM223, EE301 | 4,4 |
| 113221 | Melinda Black | black@usna.edu | SI204 | 3 |
| 090112 | Tom Johnson | Johnson@usna.edu | NN204, SI204, IT221 | 4,2,3 |

1. Is the Students table in 1NF? Why?

No. because the data within each table isn’t stored in individual columns and definitely not in its most reduced form.

1. If the Students table is not in 1NF, redesign the tables such that all the information currently in the Students table is found in the resulting tables, and the resulting tables are in 1NF. For each of the resulting tables, give the table name, column names, primary keys, and foreign keys.

# Part Two Database backup and restore

1. Explain the concept of backups and the various types of SQL backup

Backup is a way in which data is stored and accessed during times of disasters. There are 8 types of different database backups: Full, Differential, Transactional (or T-log), Tail log, Partial backup, files and filegroups, split backup and copy only backup.

2. Based-off the SLA, what would be your backup strategy for the SaleSpan. Create a maintenance plan for the backup strategy (Your backups should be well organized to reflect a good practice)

First for recovery model, I’ll schoose a Full recovery model. We setup recovery models that were adequate to our environment and followed our SLA. Setup jobs to take full backups everyday, differentials every hour and T-log every 15 minutes. And also setup a job that at the end of every full backup, we delete the previous backup to save up some space.

1. Assuming a disaster occurs, (Determine the disaster time and explain how you would bring back the database to a point in time)

If it’s within the day when the database is usually very busy, we have 45 minutes to bring back our database and tables back to working. Have a restore of the latest differential backup done as soon as possible.

If disaster happens at night (or times during minimal use), we can go ahead and rollback the latest full backup taken.

1. Explain the various types of disasters you have come across as a database administrator and how you recovered from them.

Natural disasters; disasters caused by natural events such as hurricanes, tornados, earthquakes etc. for these types of events, to avoid data loss we setup up our backup server to be at least 800 miles from the primary server according to Microsoft best practices.

Man-made disasters; disasters caused by human interference of the database. To avoid any data loss, we do the same plan as the Natural disaster.

1. How does the database recovery model affect the type of backup?

Depending on the type of recovery model, speed of backup and amount of data backed up depends on it. Simple backup recovery model only supports full and differential backups, doesn’t take T-log. Full and bulk-log backup everything so takes significantly amount of time in comparison to the simple recovery model.

1. What are the different compatibility levels supported by each server instance from 2000 to 2019 and can you restore a backup from an instance of 2008 to 2019 explain?

You can restore a backup from a 2008 instance onto a 2019 instance. This is because compatibility levels increase with higher versions of SQL, meaning newer versions can support older backup files while older versions can’t support newer backup files/databases.

1. What are your recommendations to design a backup and recovery solution?

This depends on a lot factors and how we are willing to affect performance of the database. best practice, I’d recommend a full backup everyday when database isn’t in use around midnight, a differential every hour and T-log every 10 minutes.

LAB:

* Restore AdventureWorks2016 on your default instance with the name Rxie-prj-001
* Create a daily full differential and transaction log backup job based on the SLA provided above?
* Demonstrate how you would check if your backups are running.

a. go to event viewer and see at designated times the backups ran

b. go to saved folder and see if the backups are there

c. have your service agent send you alerts through email when the job is done/failed.

* What are some the possible reasons a backup job can fail and how would you troubleshoot and resolve that?

a. service agent is off

b. database deleted

c. database is not online

d. server down

solutions:

1. Analysis of the problem (reading error, log error etc) can be viewed in Event viewer.

2. Troubleshooting (finding solutions)

3. Implement

4. Document.

# Part Three: Performance and Large databases

1. What are some of the challenges faced with large databases?

Performance. They tend to be really slow and give tedious work to find data in the database.

Lack of space: huge amounts of data can prove to be space consuming

1. How do you manage large databases in your environment?

a. Setup Indexes to improve performance

b. Data compression and shrinking kills performance but increases available space.

1. Explain in detail the following 

* Compression: there are 2 types of compression; data compression and backup compression. Note: data compression is only available in the Enterprise version of SQL server from 2012. Compression is the process in which we reduce the number of bits needed to represent data. It helps economize disk space, speed up file transfer and decrease cost for storage hardware. Data can be compressed at the row or page level. Row compression identifies the data type of each column and squeezes out any available space it finds e.g. converts char to varchar.
* Table partitioning: Table partitioning is the process by which very large tables are divided into multiple smaller parts. These smaller parts (tables) are then saved into separate files which are then saved into different filegroups. This is done to enhance performance of the database during searches and aid in maintaining large databases. Steps to partitioning include creating filegroups, add files into file group, create partition function, create partition scheme and apply the partition to the table. There are 2 different types of partitioning; vertical partitioning which partitions columns into multiple tables containing the same rows or horizontal partitioning which partitions rows into multiple tables with the same column.
* What is a database snapshot and when is it recommended?

Snapshots is a rollback plan that takes a picture of how the database looks like at that exact moment. It is a static read-only copy of an existing database. the snapshot taken is the exact replica of that database at that point in time. If the source database changes (original database), the snapshot doesn’t. This is a great rollback plan if and only if you are applying on a relatively slow database. It is also recommended as a short restore option i.e restore from snapshots when data loss is within your SLA and your snapshot is within that time range. It can also be used to repair damaged data from the source database. it can also be used to correct administrative errors done on the source database. although transactions committed on the source database show on that database, it is not reflected on the snapshot but the snapshot does grow in size.

* What is the difference between a database snapshot, and a database backup?

Snapshot are instantaneous, backups take a while

Cannot delete source database when snapshot is present but can when backup is taken

Snapshots can only be created using T-SQL but backups can be done using GUI as well.

Snapshots cannot be shrunk nor compressed, backups can

Snapshots grow when source database grows in size only while backups don’t grow when taken

4. Walk us through the process of deploying changes in your production environment

Understand what the script will do the database.

a. Take a full backup of your product environment

b. install this backup in your test environment

c. deploy changes in the test environment and make sure the right thing is done

d. if correct thing is done, you can safely run script on product environment. If not, report issue to source code producer.

1. How do indexes affect performance, give examples?

They help find information a lot faster especially in very large databases. Instead of going through the entire table, indexes are able to accurately pinpoint the location of the desired data in a small amount of time.

1. How would you configure your database for optimal performance?

By creating different indexes depending on the data in it.

1. Define a stripe backup and in what situation/scenario would you take a stripe backup.

A striped backup is a backup of your database that is spread across multiple files

1. What is capacity planning and how would you implement it in your environment.

Server capacity planning is the process of determining the specific resources your SQL Server will need over time, as well as how to tailor your approach when rolling out a new SQL Server instance. Example, Microsoft best practices is to allocate just 80% of your RAM to SQL server so that during server crashes, we are still able to use the system.

LAB:

* On your default instance take a striped backup of Rxie-prj-001
* Restore the striped backup on your test instance as Rxie-prj-Dev

# Part Four : Indexes

1. From your understanding what is a good analogy for database index.
2. What is fill factor? When do you choose a high or low Fill factor value?
3. Why is an index described as a double-edge-sword?
4. What is the difference between index Rebuild and Index Reorganize operations?
5. From an execution plan what does a table scan, index scan and index seek tell you.
6. What would influence your decision to use
   1. Clustered index
   2. Non-clustered index
   3. Filtered index
   4. Covering index
7. What is the difference between OLTP and OLAP workloads and how do they affect index creation decisions?
8. What are some index creation best practices?

LAB:

* Give us a script to check all the index fragmentation in your environment.
* Give us a script to check all the unused indexes in your environment.
* How do specify the fill factor in an index