

Programming Interview Notes - Database

Isolation Levels:

Read Uncommitted: No shared locks while reading.

Read Committed(Default): Won't read uncommitted data – shared locks for duration of read.

Repeatable Read: Holds shared locks until transaction completes.

Serializable: Key range locking.

Snapshot - Optimistic

Read Committed Snapshot Isolation: Same as Read Committed, but no shared locks for reads - Statement level consistency - less tempdb space than snapshot isolation.

Snapshot Isolation: Transactional consistency

ACID: Atomic, Consistent, Isolated, Durable

Process Query: Parse, Algebrize (Binding/Name Resolution/Semantic Check), Optimize, Compile, Execute

Normal Forms:

1NF – Has a primary key (each entity is different) – each column is atomic – no repeating groups – Each entity has same number of values.

2NF – Every non-key column is fully functionally dependent on the primary key - Each attribute describes the entire key (Only relevant with composite key).

3NF - Every non-key column is functionally dependent only on the primary key - Eliminate columns not dependant on key.

Boyce-Codd Normal Form (BCNF): Every Determinant is a candidate key. - full functional dependency

Latches – Lightweight locks. Used when managing access to internal data structures – data page, index page.

Locking levels – Row, page, extent, table, database - The large majority of latches are acquired when a data page is moved from the storage engine to the data cache

Spinlocks: A light-weight lock mechanism that doesn't lock data but it waits for a short period of time for a lock to be free if a lock already exists on the data a transaction is trying to lock. It's a mutual exclusion mechanism to reduce context switching between threads in SQL Server.

Execution Plan:

Spool: Temp copy of rows - TempDB - Typically between two other operators

Table Spool - Reads rows from child table and writes them to Temp DB

Loop Join: Rebind: Execution needing to calculate new value - Rewind: Execution using same value as preceding execution

Hash Match: Hash table for each row - Distinct, Aggregates may need to remove duplicates

Bitmap Operator: Filters out rows with key values that cannot produce joins

Bitwise: & (AND):~ (NOT):| (OR):^ (OR)

Joins: Nested Loop : Merge(ordered input) : Hash(Hash smaller table values)

2014 SQL Server Maxes: Bytes/index: 900 - Row Size: 8060 bytes - Columns/Index: 16 - Columns/table: 1024 - Instances: 50 - Nonclustered indexes/table: 999

RAID: RAID 0 – Striping : RAID 1 – Mirroring : RAID 5 – Striping with a parity bit : RAID 10 – Striped set of mirrors : RAID 0+1 - Mirrored pair of stripes

Disaster Recovery / High Availability:

Log Shipping: DR - Can't use target DB during restore - Multiple secondaries - Manual failover.

AlwaysOn: Enterprise - HA(local-sync) or DR(remote datacenter-async) - Groups(Containers) - Replicas(Host instances) - Databases(users DBs) - Group Listener(for user connections)

Failover Clustering: Standard: 2 node - Enterprise: 16 node - HA - Per instance

Mirroring: Async(Enterprise) and Sync - HA - Per DB - Can't read data on mirror

Replication: Must use PKs - Redundancy and geographic distribution - Types: Transactional(can use for non-SS - allows table-level), Snapshot, Peer-To-Peer(Enterprise), Merge.

Data Warehousing:

Schemas: Star : Snowflake : Galaxy

Conforming Dimension: Used in multiple schemas

Slowly Changing Dimensions: Type 1-Current : Type 2-Historical : Type 3- Horizontal Change Tracking

Junk Dimension: Multiple dimensions in 1 table

Inmon (3rd Normal Form) vs Kimball (Dimensional)

SSIS:

Package=>Project=>Solution

Control Flow: Tasks and containers - Order workflow - Precedence Constraints(Success, Failure, Complete)

Data Flow Task: Sources, Transformations, Destinations - OLE DB Connection, Flat File Source - Paths(Failed, OnMatch, NonMatch)

Connection Manager: OLE DB, Flat File, etc.

Data source - Data source view(subset of data from source)

Transformations: Aggregate, Conditional Split, Data Conversion, Derived Column, Lookup, Merge(Sorted datasets), Merge Join, Multicast, Sort, Union All

Containers: Foreach Loop, For Loop, Sequence, Task Host

Tasks: Expression(Set variable values), Custom, Maintenance, Execute SQL, Execute package

Script Component: Run custom script code, access .Net Assemblies, etc.

Event Handler: OnError, OnPostExecute, OnPostValidate

Logging: Text File, Database, XML File, Windows Event Log

Deployment: Package(Configurations) and Project(SSIS Catalog, .ISPAC file, Parameters)

Flat File Source - OLE DB Destination(Connection Manager - Mapping)

Save Package: Package location, you can select 'SQL Server' to store the package in a msdb database, 'File System' to create a .dtsx file, or 'SSIS Package Store' to save a file that will be accessible from SQL Server.

Data flow task: Source, transformations, Destinations

Variable: Pass values between multiple tasks

Lookup transformation: reference data - Cache: Full, Partial, No

SSRS:

Project: 'Business Intelligence Projects' and 'Report Server Project'

Create data source - Create data set(model of query results) can change column names, supply default procedure parameters, etc, - Create new report - Report Type(Tabular or Matrix(specify rows and columns)

Publish Reports: Project Properties: 'TargetReportUrl', 'TargetServerVersion' - Build.

Security: Browser, Content Manager, My Reports, Publisher, Report Builder

SSAS:

Create Cube: 1) project – 'Business Intelligence Projects' – 'Analysis Services Project' : 2) Create data source : 3) Create a data source view – Use data source – Select tables for the view - Select measure group tables – Fact tables - Select Measures - Select new Dimensions.

Dimensions: By default, only the primary key of each dimension table is brought in - Drag additional columns to 'Attributes'.

Deploy Cube: Right click on project – Properties - 'Deployment' - 'Target' specify server and DB name.

Deploy: Right click project – 'Process'.

Browse Results in AS: 'Browser' tab - 'Detail Fields','Column Fields', 'Row Fields'.

Multidimensional(MDX) : Tabular(DAX) - to use Power View

Calculations with MDX - Business logic : Roles - To configure user access

DMX: Data Mining Extensions

DAX: Data Analysis Expression - Tabular

Tabular: Storage Modes: In-Memory, Direct Query(DB Storage)

MDX: Multi Dimensional Expression

User-Defined Hierarchy: Must be manually added - Dimensions

Role Playing Dimension: A dimension used multiple times for different purposes

UDM: Unified Dimensional Model - Cube metadata

Fact data stored in partitions.

MOLAP: Multi Dimensional OLAP

HOLAP: Hybrid OLAP - Only aggregation data and UDM are stored on the server.

ROLAP: Relational OLAP - Only UDM stored on server - Data and aggregates in relational DB.

Perspective: Subset of cube

Power BI:

Power Pivot: Create data models

Power View: Data exploration tool - Excel viewing - Select from Power Pivot model.

Power Map: Visualisations of geographic data

Power Query: Data discovery and query tool - Transformations - Can import more types of sources than P Pivot.

Power Q&A: Natural language query environment Use Power Pivot model.

Power BI: Cloud-based dashboards and reporting - Pro to share with teams