Data Base Administration

Data and Database Administration

- Data Administration →
  - Total Organization's Data
  - Concerned with data as an organizational asset
  - Data Consistency
  - Data Utilization
- Database Administration (DBA)
  - One (or more) particular databases
  - Concerned with particular platform and operational aspects
  - Database security

Data Administrator vs Database Administrator

Data Administrator (DA)	**Database Administrator (DBA)**
Performs strategic planning	Controls and supervises
Sets long-term goals	Executes plans to reach goals
Sets policies and standards	Enforces policies and procedures Enforces programming standards
Job is broad in scope	Job is narrow in scope
Focuses on the long term	Focuses on the short term (daily operations)
Has a managerial orientation	Has a technical orientation
Is DBMS-independent	Is DBMS-specific

Data Administration

- Global Domains for data consistency
  - Standard names
  - Standard semantic domain (meaning)
  - Standard methods of representation
  - Standard formats (display)
  - Standard constraints (data integrity)
- Data Policies
  - Organizational “ownership”
  - Usage & Security
  - Back-up and Disaster recovery
- Data Conflict (consistency) Resolution →

Data Consistency

- Same term for same data
  - Customer vs Account
- Same representation for same data
  - Phone as character vs numeric representation
  - Three digit code for department in one part of an organization vs five digit code in another
- Same internal representation for data
  - ASCII vs EBCIDIC
- Same format for same data
  - MM/DD/YY vs MonthName DD, YYYY

DBA Roles

- Application
- Operational

Application DBA

- Database Design (*including* views)
- SQL related coding standards
- RDBMS Data Dictionary (active)
- Stored Procedures
- Triggers
- Constraints
- Database/application/query optimization
- Application Level performance statistics
- Media allocation/optimization
- Data distribution optimization

Operational DBA

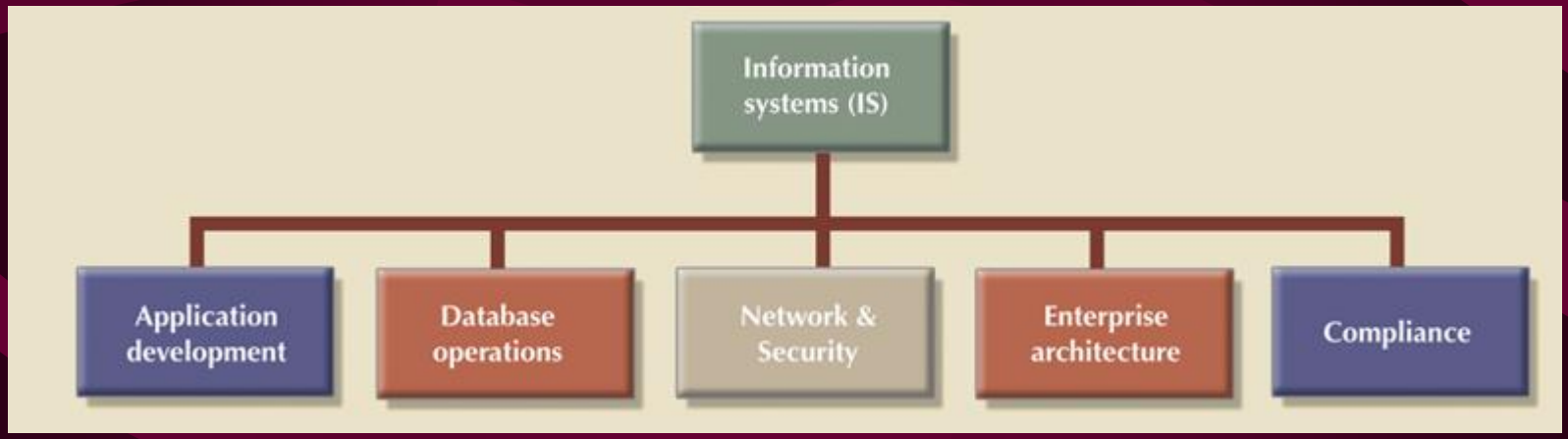
- Database level performance and reliability statistics, and reporting thereon
- Tracking service level agreements
- Backup
- Recovery
- Security
- Disaster and Contingency Plan development and testing
- Replication/distribution operations
- RDBMS configuration control & documentation

Preferred DBA Skills

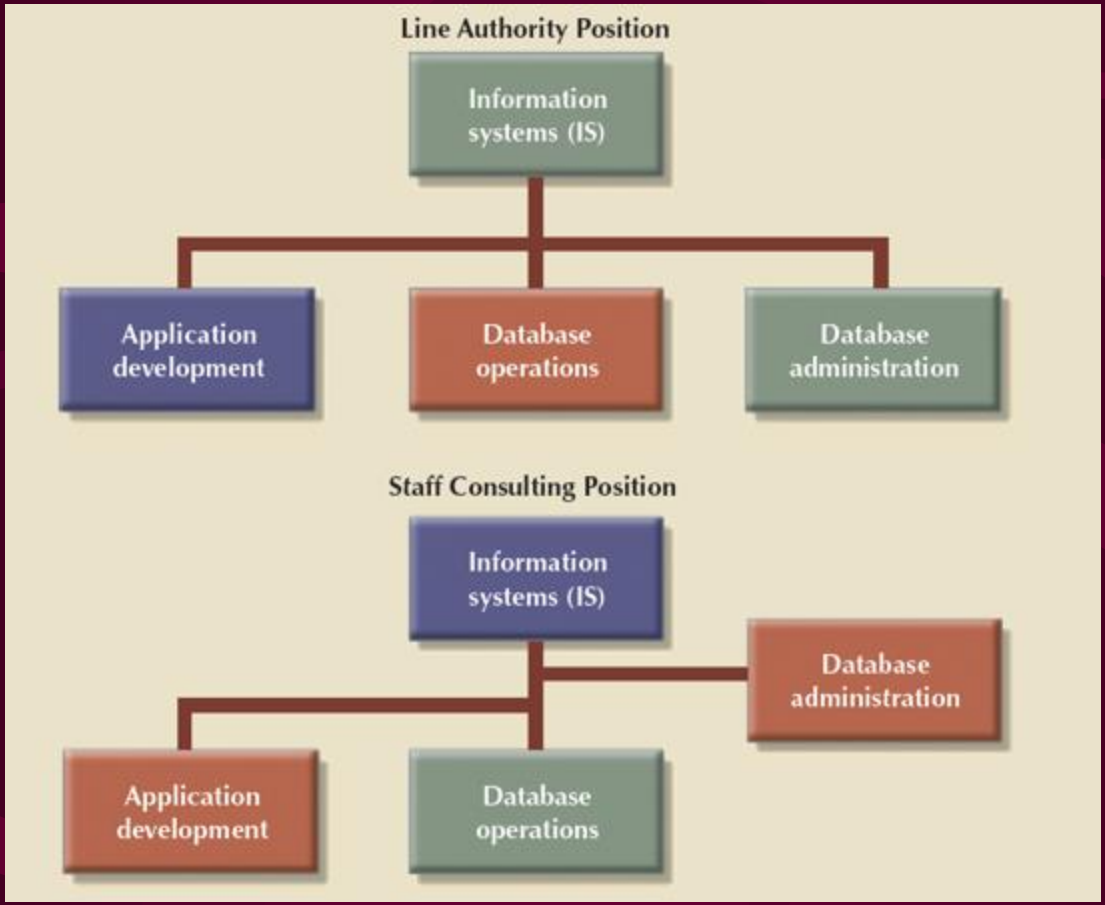
Managerial	**Technical**
Broad business understanding	Broad data-processing background and up-to-date knowledge of database technologies
Coordination skills	Understanding of Systems Development Life Cycle
Analytical skills	Structured methodologies • Data flow diagrams • Structure charts • Programming languages
Conflict resolution skills	Knowledge of Database Life Cycle
Communication skills (oral and written)	Database modeling and design skills • Conceptual • Logical • Physical
Negotiation skills	Operational skills: database implementation, data dictionary management, security, and so on
Experience: 10 years in a large DP department	Experience: 10 years in a large DP department

DBA in the IS Department

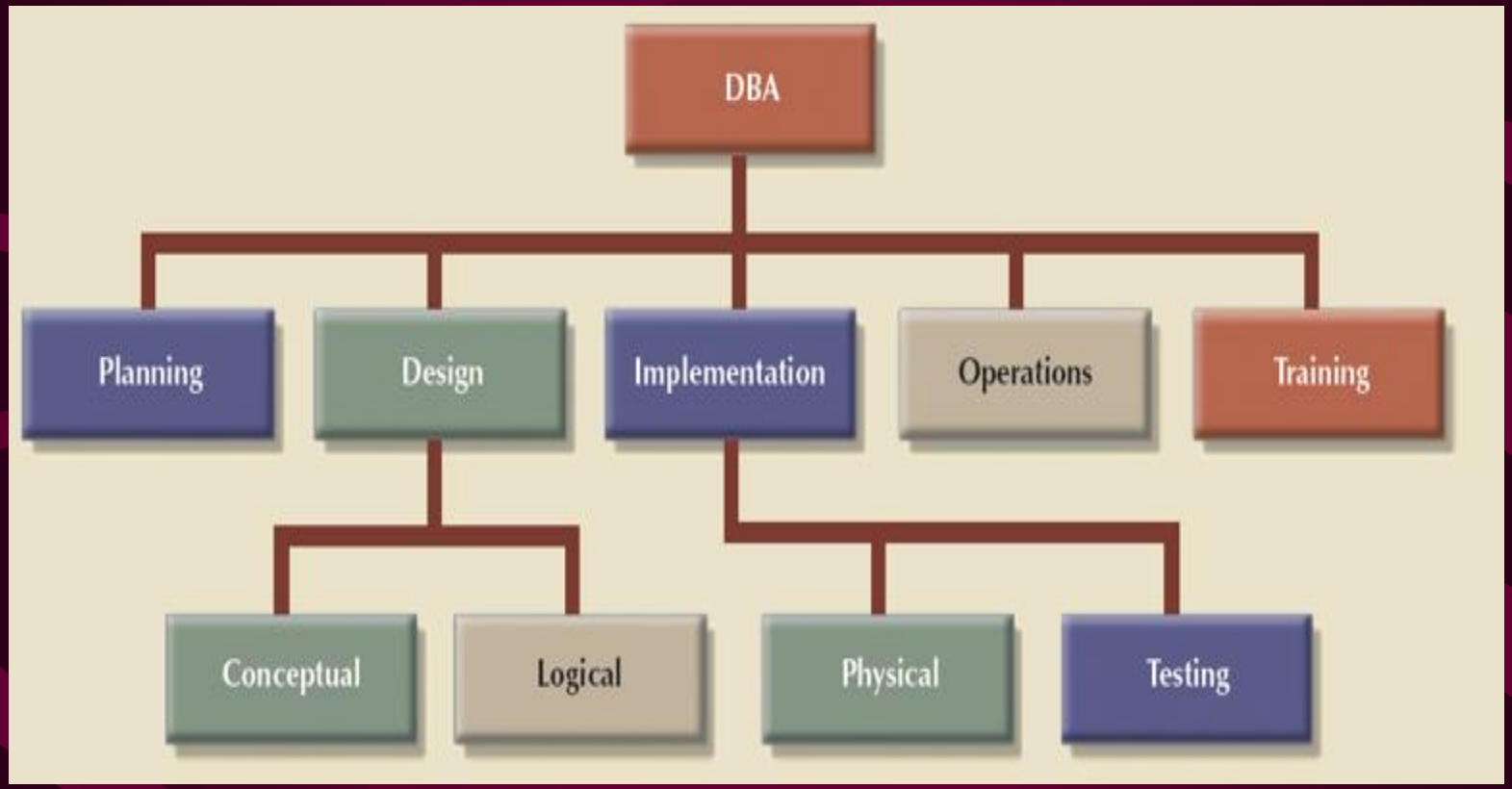
- The DBA functions are often in the applications development group, but sometimes there is a separate group for DBA's



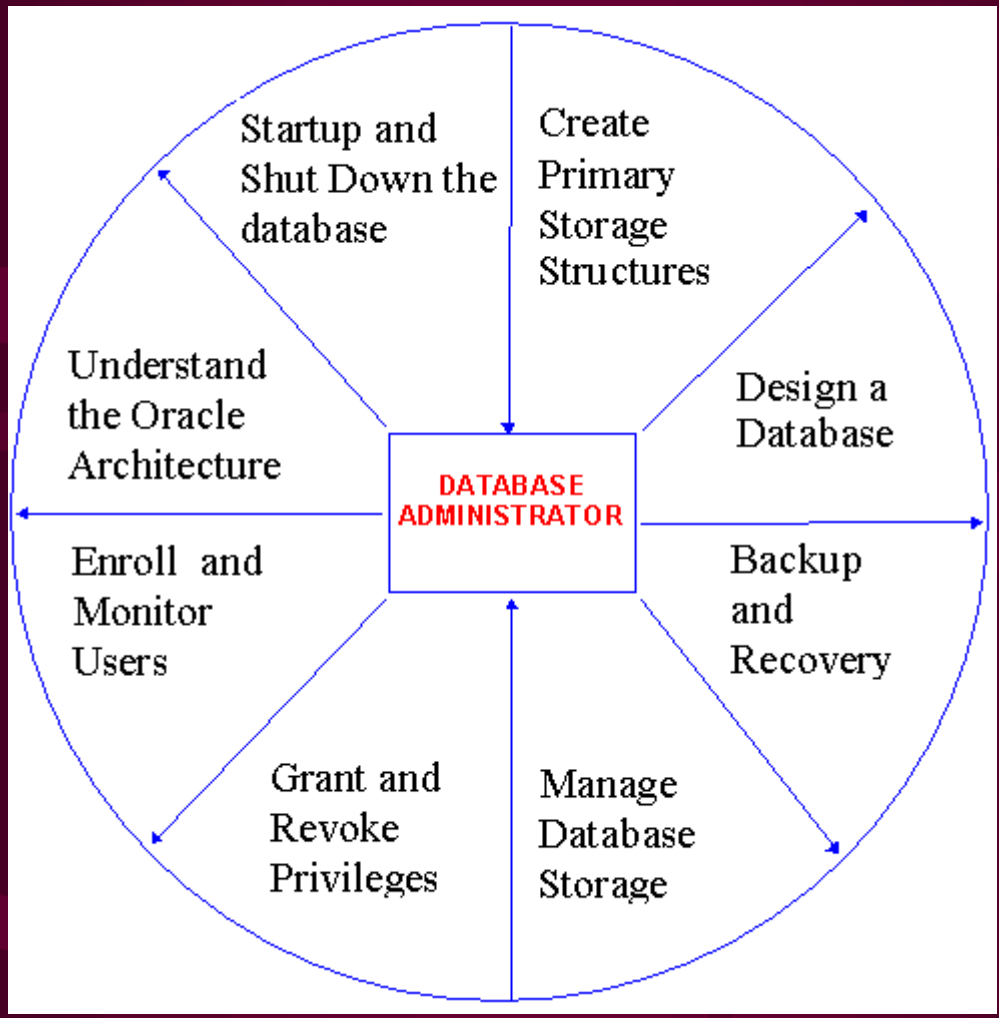
Line vs Staff Arrangement



DBA Group Specialization

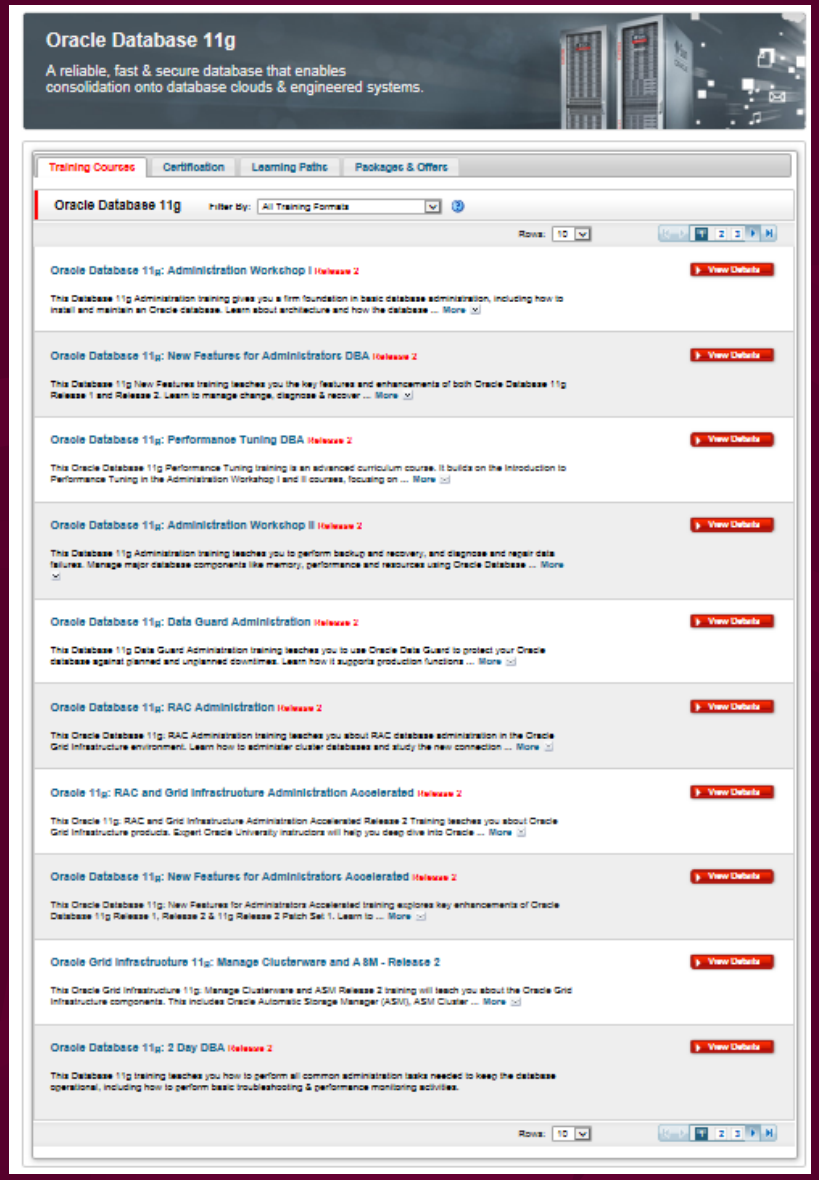


DBA Roles & Tasks

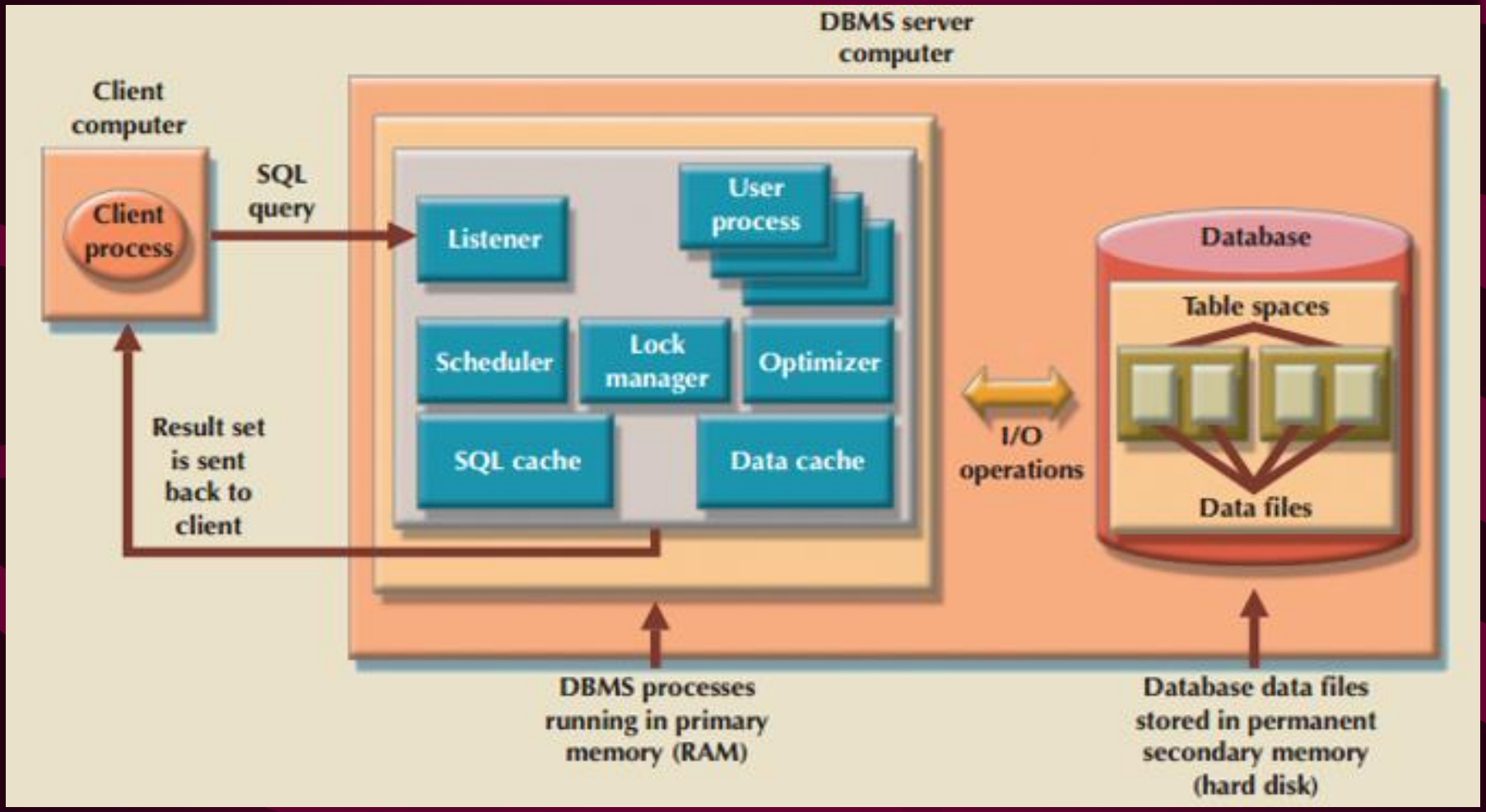


Oracle DBA Training Classes

(1 of 8 pages)



Typical RDBMS Environment



Typical Database Constructs

- Database
- Storage Group
- Tablespace
- Tables
- Domains
- Constraints
- Indexspace
- Indexes
- Views
- Triggers
- Stored Procedures

Database

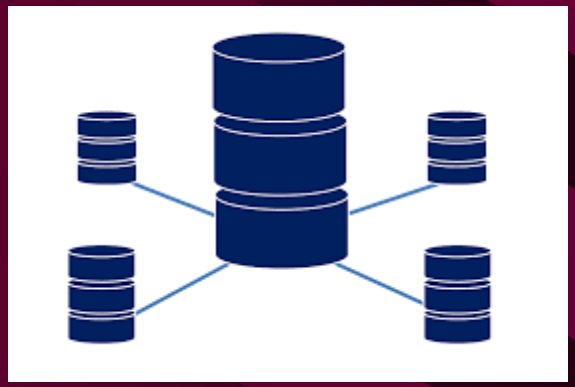
- A database may be composed of storage groups (or logical volumes)
- A database can be started (connect) or stopped (disconnect)
- Authorization to administer a database can be assigned, and authorization to use a database can be assigned (defaults assigned at this level)

Database Subdivision – Why ?

- Access speed
  - High contention tables spread across multiple physical drives
  - ID (generated key) contention
  - Clustering
- Recoverability and Backup
- Separating data from indexes

Storage Group

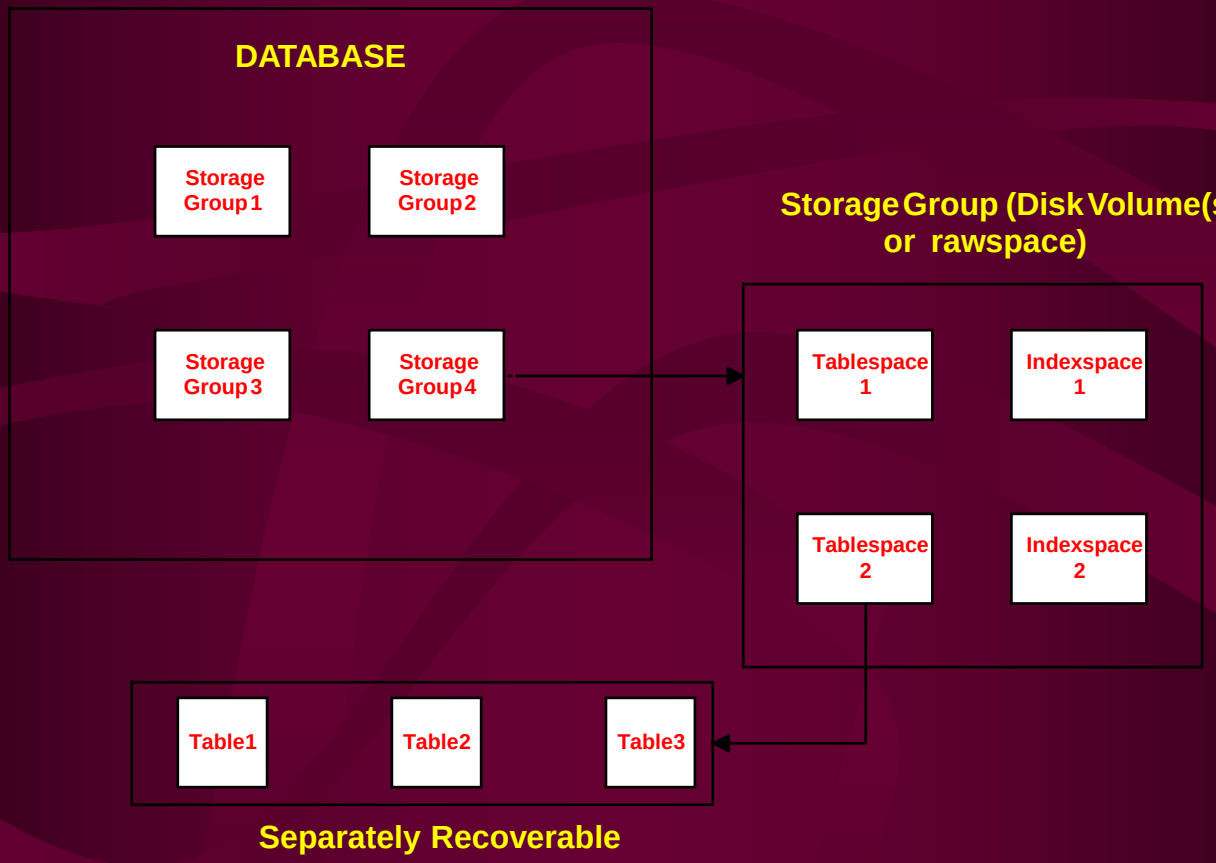
- A storage group is a set of tablespaces and indexspaces
- Storage groups may span disk units (physical volumes)



Tablespace

- A tablespace is either a collection of one or more files (data sets) or “raw” space, normally on one disk unit (or RAID unit)
- Both user data and metadata are stored in table spaces
- A tablespace may contain one or many tables (simple tablespace), or part of a table (partitioned tablespace)
- Normally tablespaces are separately recoverable units

Database Subdivision



DBMS Performance Tuning

- Recommendations for physical storage of databases
  - Utilize I/O accelerators
  - Use RAID (Redundant Array of Independent Disks) to provide a balance between performance improvement and fault tolerance
  - Minimize disk contention
  - Put high-usage tables in their own table spaces
  - Assign separate data files in separate storage volumes for indexes, system, and high-usage tables
  - Take advantage of the various table storage organizations in the database
  - Partition tables based on usage
  - Apply denormalized tables where appropriate
  - Store computed and aggregate attributes in tables

MySQL

- For MySql, multiple “databases” can be supported per server
- For ISAM organization:
  - Each database resides in a directory (folder)
  - Each table is a file
    - The limit on table size is typically the same limit on file size as set by the operating system
- User authority is granted by server origin, database (s), table(s), permissions

MySQL InnoDB

- Fully integrated with MySQL Server, the **InnoDB** storage engine maintains its own buffer pool for caching data and indexes in main memory
- **InnoDB** stores its tables and indexes in a tablespace, which may consist of several files (or raw disk partitions)
  - This is different from, for example, **MyISAM** tables where each table is stored using separate files. **InnoDB** tables can be of any size even on operating systems where file size is currently limited to 2GB.
- You can elect to store each **InnoDB** table and its indexes in its own file - this feature is called “multiple tablespaces” because in effect each table has its own tablespace
- Using multiple tablespaces can be beneficial to users who want to move specific tables to separate physical disks or who wish to restore backups of single tables quickly without interrupting the use of the remaining **InnoDB** tables
- You can enable multiple tablespaces by adding the appropriate line to the **[mysqld]** section of my.cnf

- What things can happen to compromise or destroy one’s database ?

Don’t look ahead !

Christian Brothers University

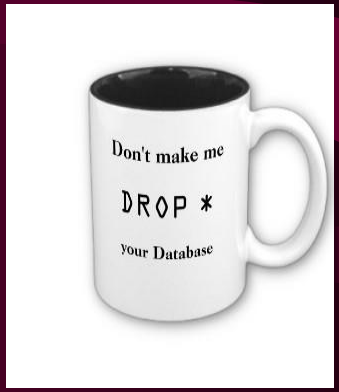
Failure Scenarios

- Database gets corrupted due to I/O issues
- Need to roll back the database to a previous timeframe because of user error
- Meet audit compliances
- Power failure (resulting in corrupted data)
- Systems software failure
- Malicious deletion or modification of data
- Cyber security issues – viruses, etc.
- Natural disasters (that is, fire, flood, earthquake, and so on)
- Theft or sabotage

Backup and Recovery Measures

- Periodic data and application backups
- Proper backup identification
  - Internal labels
  - External labels
- Convenient and safe backup storage
- Physical protection of both hardware and software
- Personal access control to the software of a database installation
- Insurance coverage for the data in the database
- Data recovery and contingency plans must be tested, evaluated, and practiced frequently

Backup Types



- Full - Entire database
- Tablespace - Part of a database
- Incremental - Changed tables/rows
- Log – Change logs

Backup Strategy & SLA

- Let's say that an organization determines that 2 hrs is acceptable for data loss
- Then SLA can be calculated as:
  - SLA = 2 hours + time required to restore the database + time required to release the db to production
- So a backup strategy has to be designed in such a way that at any scenario, they will lose maximum of 2 hours of data
- To arrive at a <2 hrs data loss, here is a sample plan:
  - Full back up every day at 12:00 am
  - Differential backup every 8 hours
  - Log backup every 2 hours
- So if there is something wrong at 2:51 in the afternoon on Wednesday then they can restore Wednesday morning 12:00 am full backup, at 8:00 am differential backup, at 10:00 am log backup, at 12:00 am log backup, and at 2:00 pm log backup
- So above plan guarantees that the database can be brought to a point as it was at 2:00 PM

Data Deduplication

- In computing, **data deduplication** is a specialized data compression technique for eliminating duplicate copies of repeating data
- Related and somewhat synonymous terms are **intelligent (data) compression** and **single-instance (data) storage**
- The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent
- In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis
- As the analysis continues, other chunks are compared to the stored copy and whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk
- Given that the same byte pattern may occur dozens, hundreds, or even thousands of times (the match frequency is dependent on the chunk size), the amount of data that must be stored or transferred can be greatly reduced

Security

- Compliance
  - Activities that meet data privacy and security reporting guidelines
- Confidentiality
  - Protecting data against unauthorized access
- Integrity
  - Keeping data consistent and free of errors or anomalies
- Availability
  - Accessibility of data whenever required by authorized users and for authorized purposes

Security Policies

- Collection of standards, policies, and procedures created to guarantee security
  - Ensures auditing and compliance
  - Sprecific policies in regard to: users, workstations, servers, operating systems, applicatios, networks, data
- Security audit process
  - Identifies security vulnerabilities
  - Finds measures to protect the system

Security Vulnerabilities

- Weakness in a system component that could allow unauthorized access or cause service disruptions
  - Categories: technical, managerial, cultural, and procedural
  - Security threat: imminent security violation
  - Security breach: occurs when a security threat is exploited and could lead to a database whose integrity is preserved or corrupted
    - Preserved or corrupted

Database Security

- DBMS features and related measures that comply with the security requirements
  - Authorization management: procedures to protect database security and integrity
    - User access management
    - View definition
    - DBMS access control
    - DBMS usage monitoring
      - Audit log: automatically records description of database operations performed by all users

Database Administration Tools

- There are many sophisticated data administration tools
  - Monitoring
  - Load testing
  - Performance tuning
  - SQL code optimization
  - Bottleneck identification and remediation
  - Modeling and design
  - Data extraction, transformation, and loading

Microsoft DBA Tools

Current state of SQL Tools portfolio

Tool	Code Repo	Build & Test	Deploy	Monitor	Develop
Graphical User Interface					
SSMS			●	●	●
SSDT	●	●	●		●
mssql for VS Code	●				●
SQL Ops Studio preview	●	● coming	●	●	●
Command Line Interface					
mssql-tools (sqlcmd & bcp)				●	
mssql-scripter preview					●
mssql-cli preview				●	●
sqlpackage coming soon		●	●		

Platforms

The Data Dictionary

- Main types
  - Integrated: included with the DBMS
  - Standalone: third-party systems
  - Active: automatically updated by the DBMS with every database access
  - Passive: requires running a batch process
- Main function: store description of all objects that interact with the database
  - Provides database designers and end users with a much-improved ability to communicate
  - Helps the DBA resolve data conflicts

The Data Dictionary (con't)

- Key element of information resource management
  - Can be described as the information resource dictionary
- Metadata is the basis for monitoring database use and for assigning access rights to users
  - Information stored in the data dictionary is usually based on a relational table format, thus enabling the DBA to query the database with SQL commands
- DBA uses data dictionary to support data analysis and design

LOGS

- Application
  - Audit Trails
  - Error Logs
  - Need to keep copies both in database (tables) and outside of database (sequential files or tape)
- RDBMS
  - Before Image
  - After Image
  - Should maintain multiple copies (one copy on tape)

Before Image

- The before image log keeps a copy of every database row before it was changed by a transaction, including the transaction id, type of change, and timestamp
- This log is used for rollback (undo) of transactions
- The log may also contain indexing (or physical) pointers to thread transactions
- This log may also be used for “read consistency” services

After Image

- The after image log keeps a copy of every database row after it was changed by a transaction, including the transaction id, type of change, and timestamp
- This log used for rollforward (redo)
- Checkpoints are noted in the log
- The log may also contain indexing (or physical) pointers to thread transactions
- This log may also be used for replication services

Checkpoint

- For maximum efficiency, database changes are not written to disk as they occur
- Database is paged into memory, and pages are written back to disk only as necessary
- A checkpoint synchronizes the disk copy(s) and memory (paged) copy of a database; all database activity is paused for a checkpoint
- The disk copy is typically mirrored (or RAID used) to minimize the effect of disk hardware problems

Media (or other platform) Failure (rollforward)

- To rollforward (or redo transactions), “after images” since the last checkpoint are applied to the database

Application Failure (Rollback)

- To rollback (or undo one or more transactions), each changed row is replaced by its before image version

SQL Transaction and Data Control

Transaction Control Language	
COMMIT	Permanently saves data changes
ROLLBACK	Restores data to its original values
Data Control Language	
GRANT	Gives a user permission to take a system action or access a data object
REVOKE	Removes a previously granted permission from a user

Database Security

- SUBJECTS - A person, group, or function (App) that can access the database
- OBJECTS - Those table (or view) columns that can be accessed
- PRIVILEGES - Those action that can be performed by the subjects on the objects

- Which subjects have what kind of access to which objects ?

Subjects
Identification/Authentication

- User Identification - USER
- User Authentication - PASSWORD
- May be same USER and PASSWORD for platform (operating system) logon as database login (“single signon”), or two separate sets of codes

Privileges

- The typical privileges that may be granted to subjects include:
  - ALL - all privileges
  - USAGE - use a certain domain
  - DELETE - delete rows
  - INSERT - insert rows
  - INSERT (x) - insert into column(s) x
  - SELECT - select rows (or create views)
  - UPDATE - update
  - UPDATE (x) - update column(s) x

Objects

- The objects that may be operated upon include:
  - Tables
  - Views
- Note that the privileges also includes reference to column objects for some privileges (ie INSERT & UPDATE)
- Some RDBMS may offer password control on objects as well as subjects

SQL GRANT

- GRANT INSERT, UPDATE ON SP TO JOE
- GRANT SELECT ON P TO PUBLIC
- GRANT ALL ON P TO SAM, DAVE
- GRANT UPDATE (Qty) ON SP TO ED
- GRANT DELETE ON S TO BONNIE, CLYDE

CASCADE

- GRANT SELECT ON SP TO ALICE WITH GRANT OPTION
- ALICE can now pass on the select on SP to anyone else

REVOKE

- REVOKE SELECT ON SP FROM ALICE
- Revokes Alice's select privilege on SP, plus anyone else that Alice has passed on that privilege to

Views & Security

- One can create views to only allow subjects to see certain parts of certain tables
- Then grant authority for those views
- CREATE VIEW MYPRODUCTS AS
  - SELECT * FROM P WHERE P.PID IN
    - (SELECT SP.PID FROM SP
    - WHERE SP.SID = USER)
- GRANT SELECT ON MYPRODUCTS TO PUBLIC

Database Certifications

- Microsoft Certified Database Administrator (MCDBA)
- MySql Certified Professional
- IBM Information Management Certification
- ICCP – Certified Data Management Professional
- MongoDB NoSQL Certifications
- Oracle Certifications →

ORACLE®
Certifications

Database Administration

- Oracle Database 11g Administrator Certified Associate
- Oracle Database 11g Administrator Certified Professional
- Oracle Database 11g Administrator Certified Master
- Oracle Certified Expert, Oracle Real Application Clusters 11g and Grid Infrastructure Administrator
- Oracle Database: SQL Certified Expert

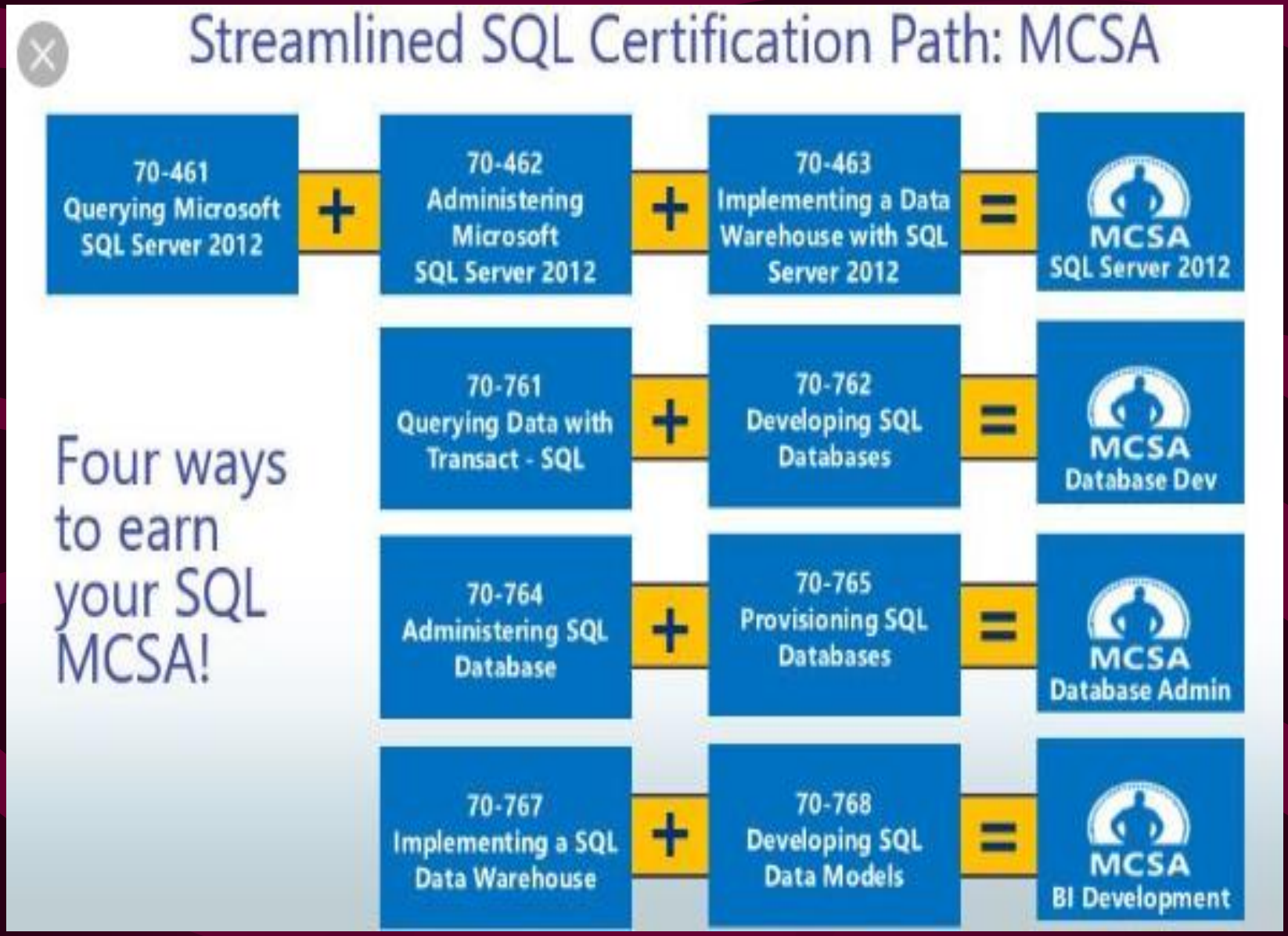
Application Development

- Oracle Application Express (APEX) Developer Certified Expert

Middleware

- Oracle PL/SQL Developer Certified Associate
- Oracle Advanced PL/SQL Developer Certified Professional

Streamlined SQL Certification Path: MCSA



References

- Fundamental of Database Administration: DBA by Aatif Jamshed Namrata Dr. Krishna kant Agrawal Editor:Mukesh KumarMohanty, Ms.Namrata, et al.
- MCSA SQL Database Administration Exam Ref 2-pack: Exam Refs 70-764 and 70-765 by Victor Isakov, Paul Marquardt, et al.
- Expert Oracle Database Administration (Expert's Voice in Oracle) by Sam Alapat

Homework

- Textbook Chapter 16
- Review Questions 1, 8, 9, 11, 15, 17, 21, 22, 26
- Project Evaluation Form