

The Management Discipline of Database Administration:


DBA Best Practices.

Craig S. Mullins

Mullins Consulting, Inc.

<http://www.mullinsconsultinginc.com>



Sponsored by 

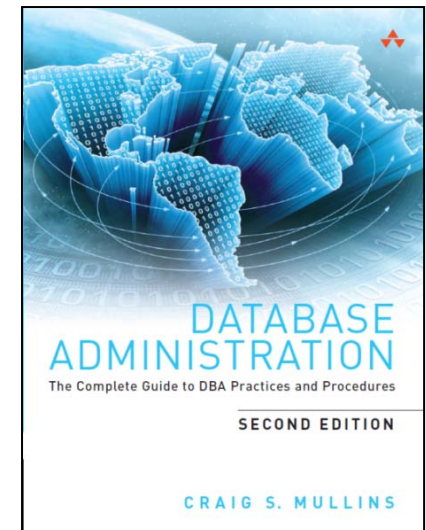
Objectives

- This presentation will clarify the roles and responsibilities of the DBA and offer guidance for an effective DBA function. It will offer the DB2 DBA a database administration roadmap, of sorts. Topics to be addressed include:
 - A framework for database administration
 - Control objectives for DBAs to ensure an effective database environment
 - Best practices for DBA tasks such as database design, performance management, backup & recovery, etc.
 - Advice on expanding from being viewed only as technicians
 - Guidance on how to interact with other IT professionals



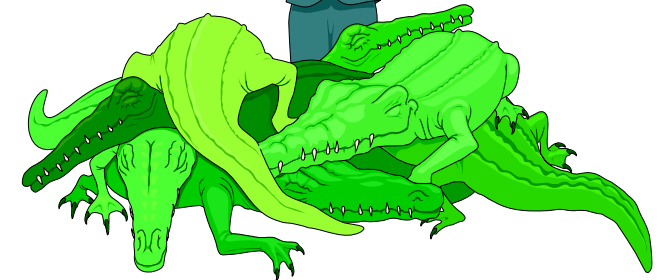
Agenda

- What is a DBA?
 - A framework for DB2 database administration
- DBA Tasks and Roles
 - Control objectives for DBAs to ensure an effective database environment
 - Best practices for DBA tasks such as database design, performance management, backup & recovery, etc.
- Advice on expanding from being viewed only as technicians
- Guidance on how to interact with other IT professionals
- Summary

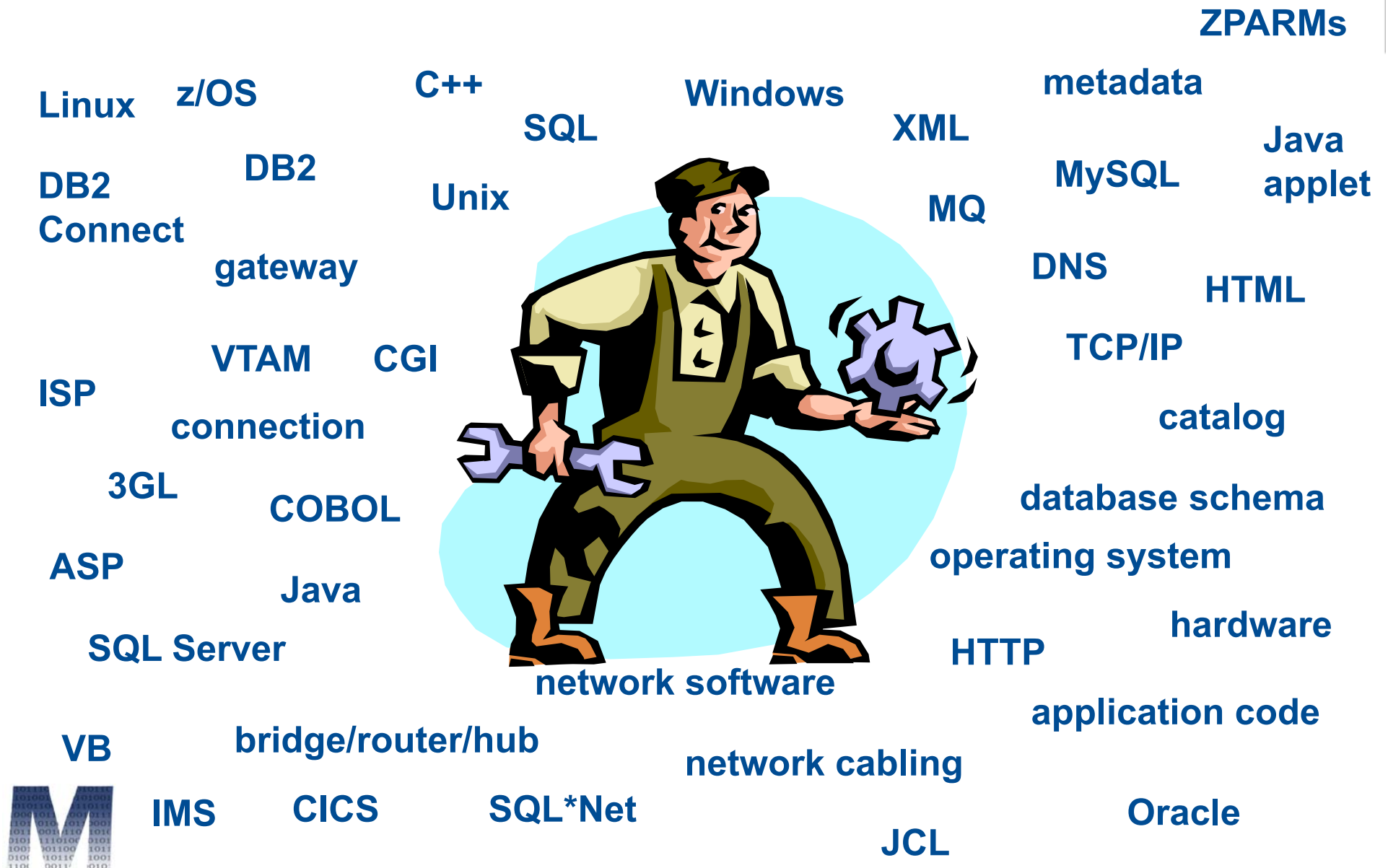


What is a DBA?

- A day in the life of a DBA...



The DBA is a “Jack of all Trades”



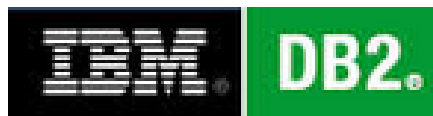
DBA as a Management Discipline

- The term *discipline* implies planning and then implementing according to that plan.
 - ◆ Creating the Database Environment
 - ◆ Database Design
 - ◆ Application Design
 - ◆ Design Reviews
 - ◆ Database Change Management
 - ◆ Data Availability
 - ◆ Performance Management
 - System Performance
 - Database Performance
 - Application Performance
 - ◆ Data Integrity
 - ◆ Database Security
 - ◆ Backup and Recovery
 - ◆ Disaster Planning
 - ◆ Storage Management
 - ◆ Distributed Database Management
 - ◆ Data Warehouse Administration
 - ◆ Database Utility Management
 - ◆ Database Connectivity
 - ◆ Procedural DBA
 - ◆ Soft Skills



Creating the Database Environment

- Choosing a DBMS
 - vendor, platform, and architecture of DBMS



Mainframe: z/OS

Distributed:

Windows

Unix

AIX

Sun Solaris

HP-UX

Linux

Others (VSE, VMS, MPE, iSeries, etc.)

Desktop OS

Windows

Linux

Mac

Enterprise

Departmental

Personal

Mobile

Free version?

Parallel Edition



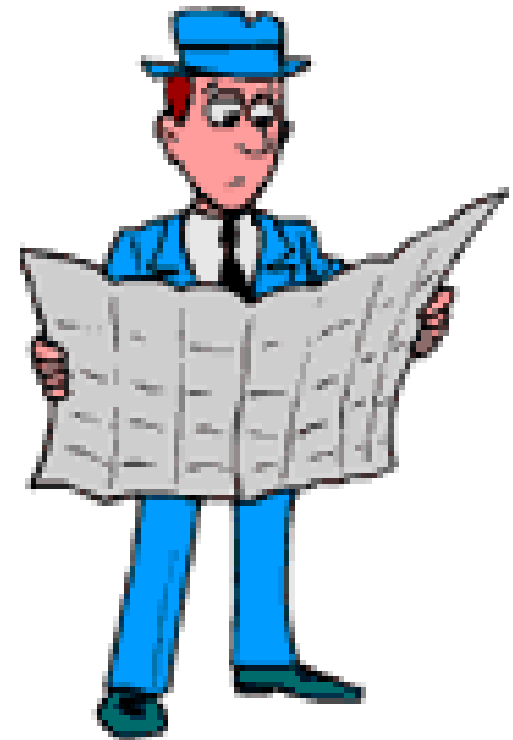
Choosing the DBMS

- Operating System Support
- Benchmarks (*TPC, homegrown*)
- Scalability
- Availability of Tools
- Availability of Technicians (*DBAs, Programmers, SAs, etc.*)
- Cost of Ownership
- Release Schedule (*Versions, Releases*)
- Reference Customers



Installing the DBMS

- Hardware Requirements
 - CPU (version/speed), firmware, memory, etc.
- Storage Requirements
 - System, Applications
- Software Requirements
 - Allied Agents (TP, MQ, middleware)
 - Languages and Compilers
- Configuration
 - ...of the DBMS
 - ...of connecting software
- Verification



Upgrading the DBMS

- Analysis of New Features
 - Check all Requirements
 - Hardware and Software (see Installation Checklist)
- Planning the Upgrade
 - Impact to system, applications
 - Scheduling
- Fallback Strategy
- Migration Verification

Every 18 to 36 months...



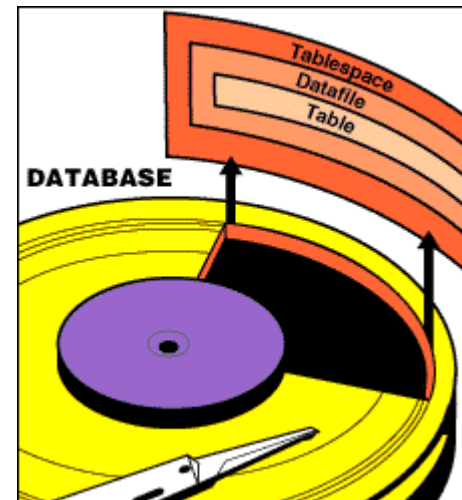
Database Standards & Procedures

- Naming Conventions (in conjunction with DA)
- Roles & Responsibilities (in conjunction with DA and SA)
- Programming Guidelines (in conjunction with App Dev)
- Database Guidelines
 - Security (in conjunction with Security Admin or SA)
 - Migration & Turnover Procedures
 - Design Review Guidelines
(in conjunction with DA, SA, App Dec)
 - Operational Support (in conjunction with Operations)
- DBMS Education
 - DBMS Overview ; Data Modeling and Database Design; Database Administration ; Introduction to SQL; Advanced SQL; Database Programming

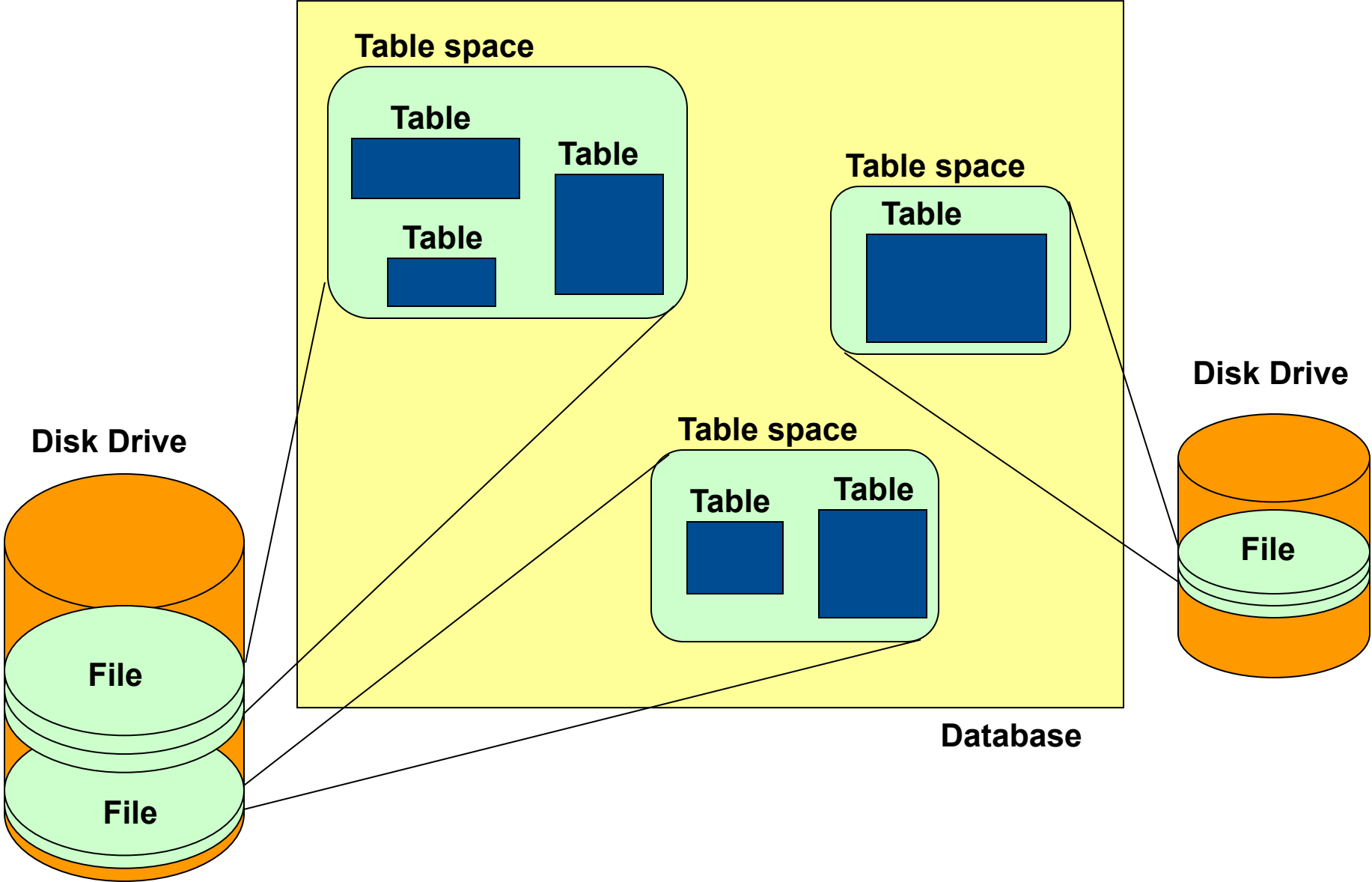


Database Design

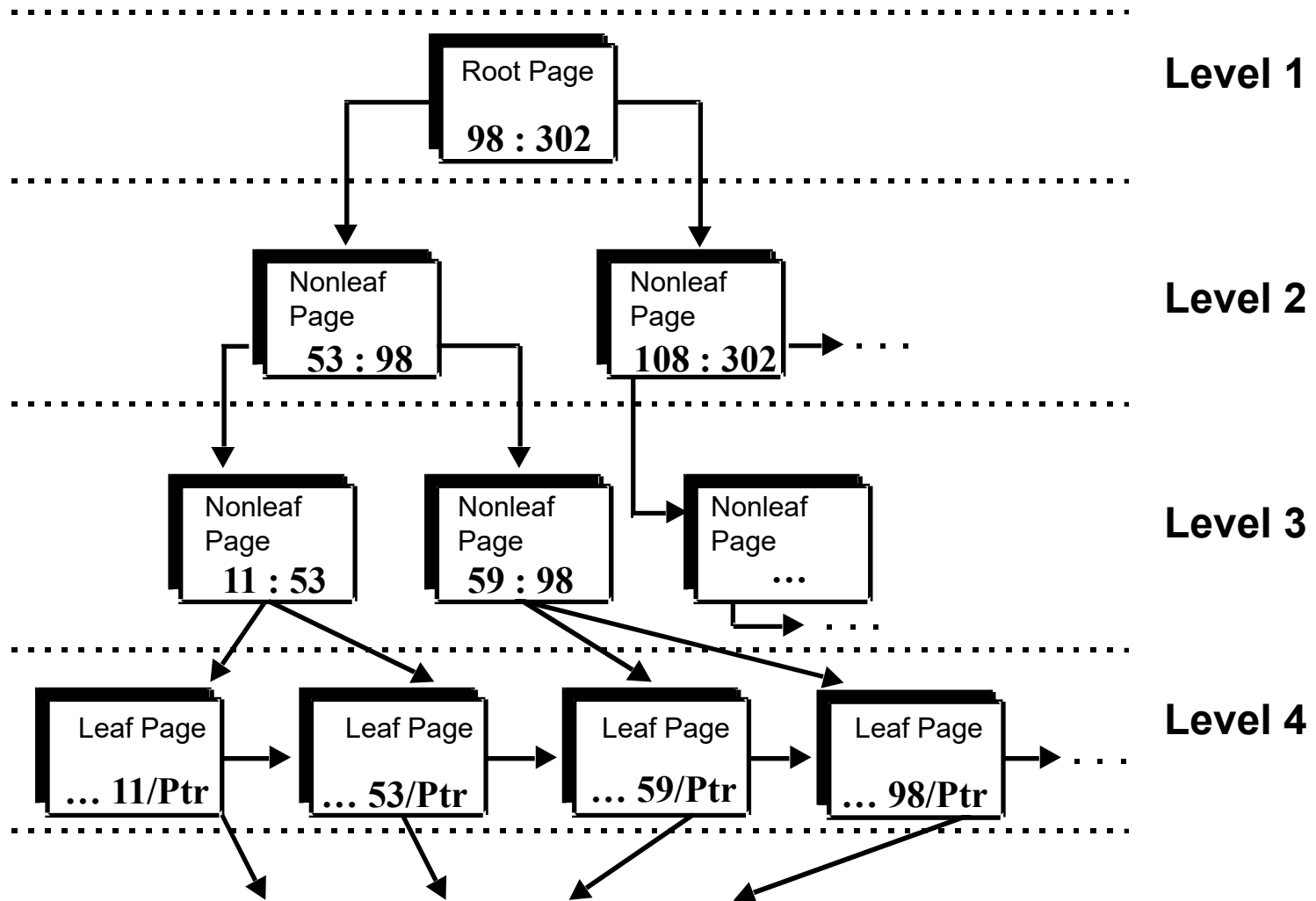
- Translation of Logical Model to Physical Database
 - Entities to Tables, Attributes to Columns, etc.
 - ...But differences CAN and WILL occur
- Create DDL
- Create Storage Structures for Database
 - Files for data and indexes
 - Raw Files versus OS Files
 - Partitioning
 - Clustering
 - Placement
 - Interleaving Data



Basic Physical DBMS Constructs



Indexing for Performance

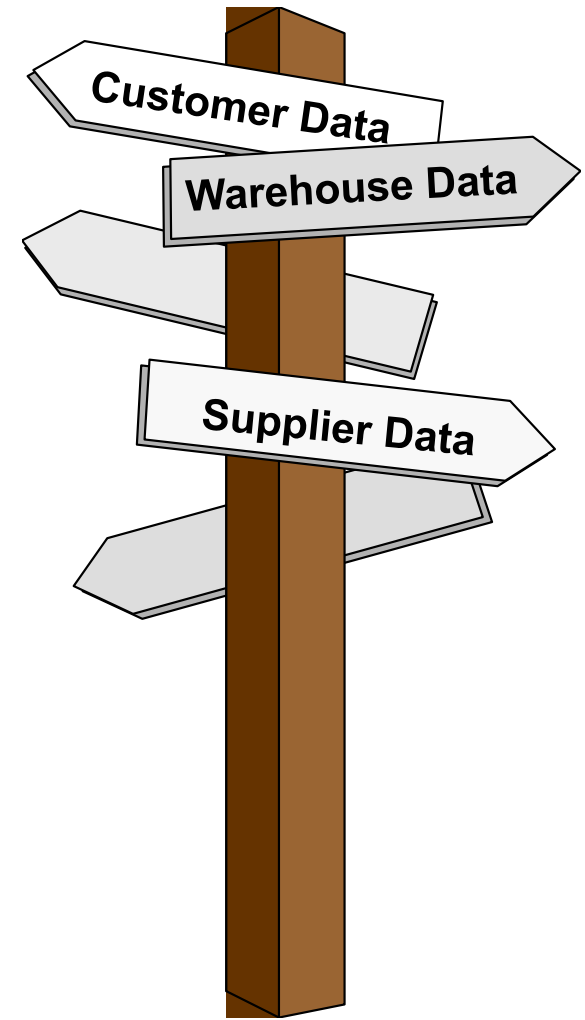


...to the data in the table.



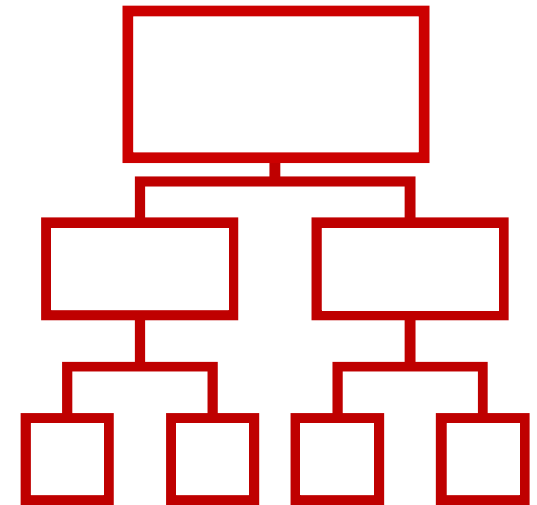
Types of Indexes

- B-Tree Indexes
- Bit-map Indexes
- Reverse Key Indexes
- Partitioning Indexes
- Clustering Indexes
- Unique/Non-unique Indexes
- Hashing



When and How to Index

- A proper indexing strategy can be the #1 factor to ensure optimal performance
 - First take care of unique & PK constraints
 - Then for foreign keys (*usually*)
 - Heavily used queries - predicates
 - Overloading of columns for IXO
 - Index to avoid sorting
 - ORDER BY, GROUP BY, DISTINCT
 - Consider I/U/D implications
 - Choose first column (high cardinality)
 - Indexing variable columns?



Denormalization?

- Pre-Joined Tables
- Report Tables
- Mirror Tables
- Split Tables (Splitting Long Text Columns)
- Combined Tables
- Redundant Data
- Repeating Groups
- Derivable Data
- Hierarchies
- Special Physical Implementation Needs

Only for performance reasons

Continuously re-evaluate



Application Design

- Database Application Development and SQL
 - SQL
 - *Set-at-a-Time Processing and Relational Closure*
 - *Embedding SQL in a Program*
 - *SQL Middleware and APIs*
 - *Code Generators*
 - *Object Orientation and SQL*
 - *Types of SQL*
 - ad hoc versus planned
 - embedded versus stand-alone
 - static versus dynamic
 - *SQL Coding for Performance*
 - Hints and Tips



Defining Transactions

- **A transaction is an atomic unit of work with respect to recovery and consistency.**
 - Defining Transactions
 - **A**tomicity
 - **C**onsistency
 - **I**solation
 - **D**urability
 - Unit of Work
 - Ensure proper definition and coding



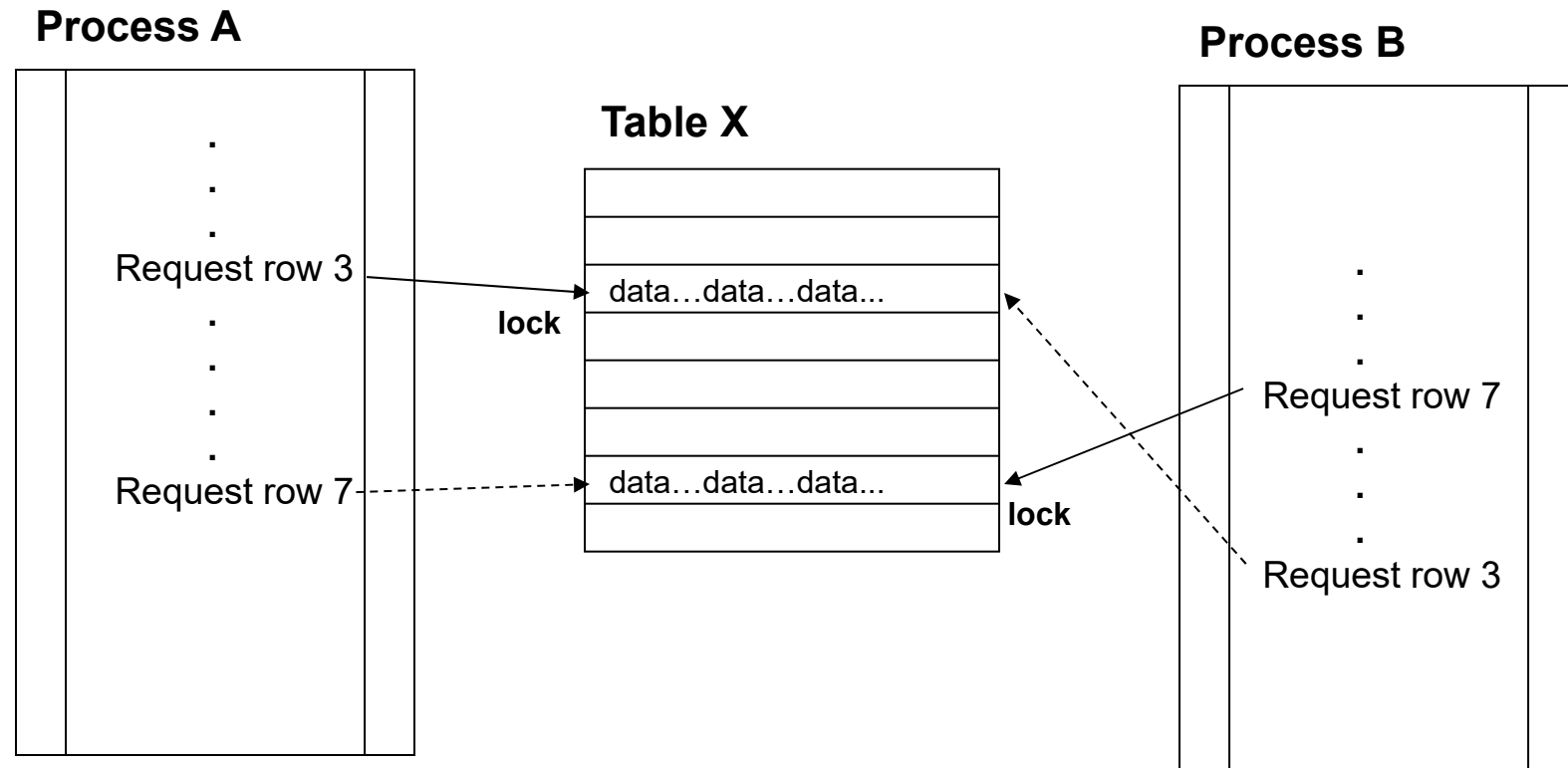
Transactions and Locking

- Locking
 - Types of Locks
 - Share, eXclusive
 - Intent Locks
 - Row, Page, Table, Table Space
 - Timeouts
 - Deadlocks
 - Example on next page



Locks are used to ensure the integrity of data.

Deadlocks



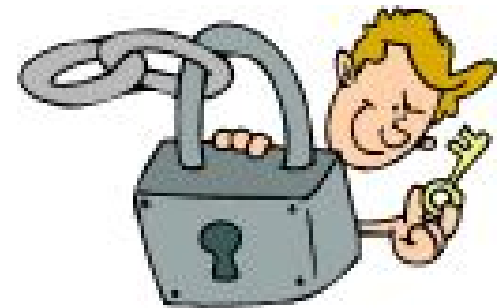
Process A is waiting on Process B

Process B is waiting on Process A



Lock Duration

- Isolation Level
 - Read uncommitted
 - Read committed
 - Repeatable read
 - Serializable
- Lock Escalation
- Techniques to Minimize Lock Problems
 - standardize the sequence of updates within all programs
 - save all data modification requests until the end of the UOW



Design Reviews

- *Design reviews* are conducted to analyze and review all aspects of the database and application code for efficiency, effectiveness, and accuracy.
 - Types of Design Reviews
 - *The Conceptual Design Review*
 - *The Logical Design Review*
 - *The Physical Design Review*
 - *The Organizational Design Review*
 - *SQL and Application Code Review*
 - *The Pre-Implementation Design Review*
 - *The Post-Implementation Design Review*



Database Change Management

- The DBA is the custodian of database changes.
- Types of Changes
 - DBMS Software
 - Hardware Configuration
 - Logical and Physical Design
 - Applications
 - Physical Database Structures
- Compliance



The Limitations of ALTER

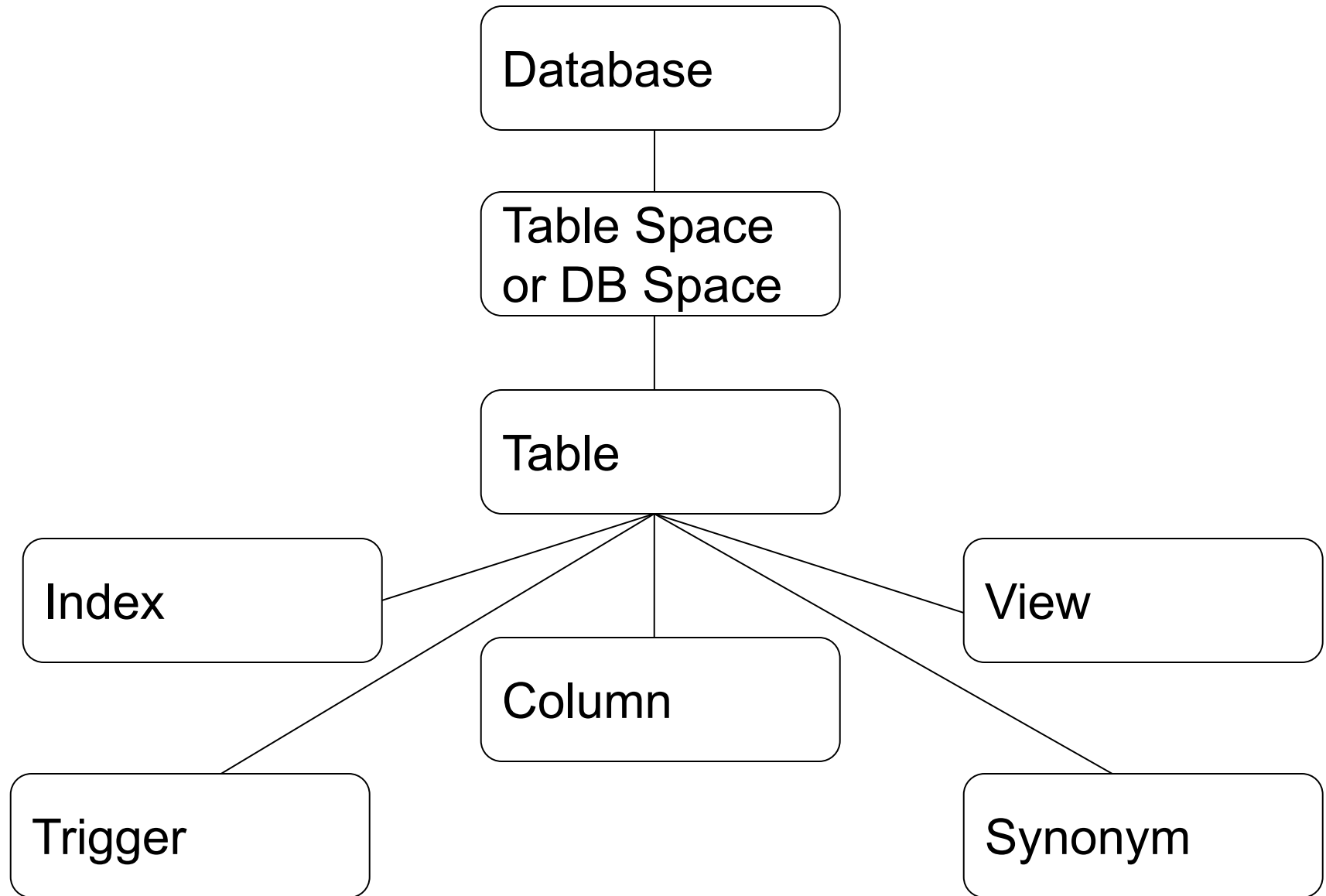
Many Types of Changes are not Likely to be Supported

- Changing the name of some objects.
- Moving an object to another database.
- Changing the number of table space partitions or data files or removing a partition.
- Moving a table from one table space to another.
- Rearranging the order of columns in a table.
- Removing columns from a table.
- Changing the definition of a key (P or F)
- Changing a view (adding or removing cols).
- Modifying the contents of a trigger.
- Changing a hash key.
- Changing table clustering.
- And so on...

DROP and re-create



The Hierarchy of Database Objects



Adding a Column to Middle of a Table

- Retrieve current table def by querying sys catalog.
- Retrieve def for any views on that table.
- Retrieve def for all indexes on that table.
- Retrieve def for all triggers on that table.
- Capture all referential constraints for the table & related tables.
- Retrieve all security.
- Obtain a list of all programs that access the table.
- Unload the data in the table.
- Drop the table, which in turn drops any related objects.
- Re-create table adding the new column.
- Reload the table using the unloaded data from step 8.
- Re-create any referential constraints.
- Re-create any triggers, views and indexes for the table.
- Re-create the security authorizations from step 6.
- Examine each application program to determine if changes are required for it to continue functioning appropriately.



Change Management Guidelines

- Requesting Database Changes
 - Standardized Change Requests
 - Automated or Paper Forms
 - DBA Checklists
 - Service Level Standards
- Communication
 - To DBA from Requesters
 - ...and from DBA back to Requesters
- Audit
 - Who...what...why...when...how...



Data Availability

- *Availability* is the condition where a given resource can be accessed by its consumers.
 - Availability can be broken down into four distinct components:
 - Manageability – the ability to create and maintain an effective environment that delivers service to users.
 - Recoverability – the ability to re-establish service in the event of an error or component failure.
 - Reliability – the ability to deliver service to specified levels for a stated period of time.
 - Serviceability – the ability to effectively determine the existence of problems, diagnosis their cause(s), and repair the problem.

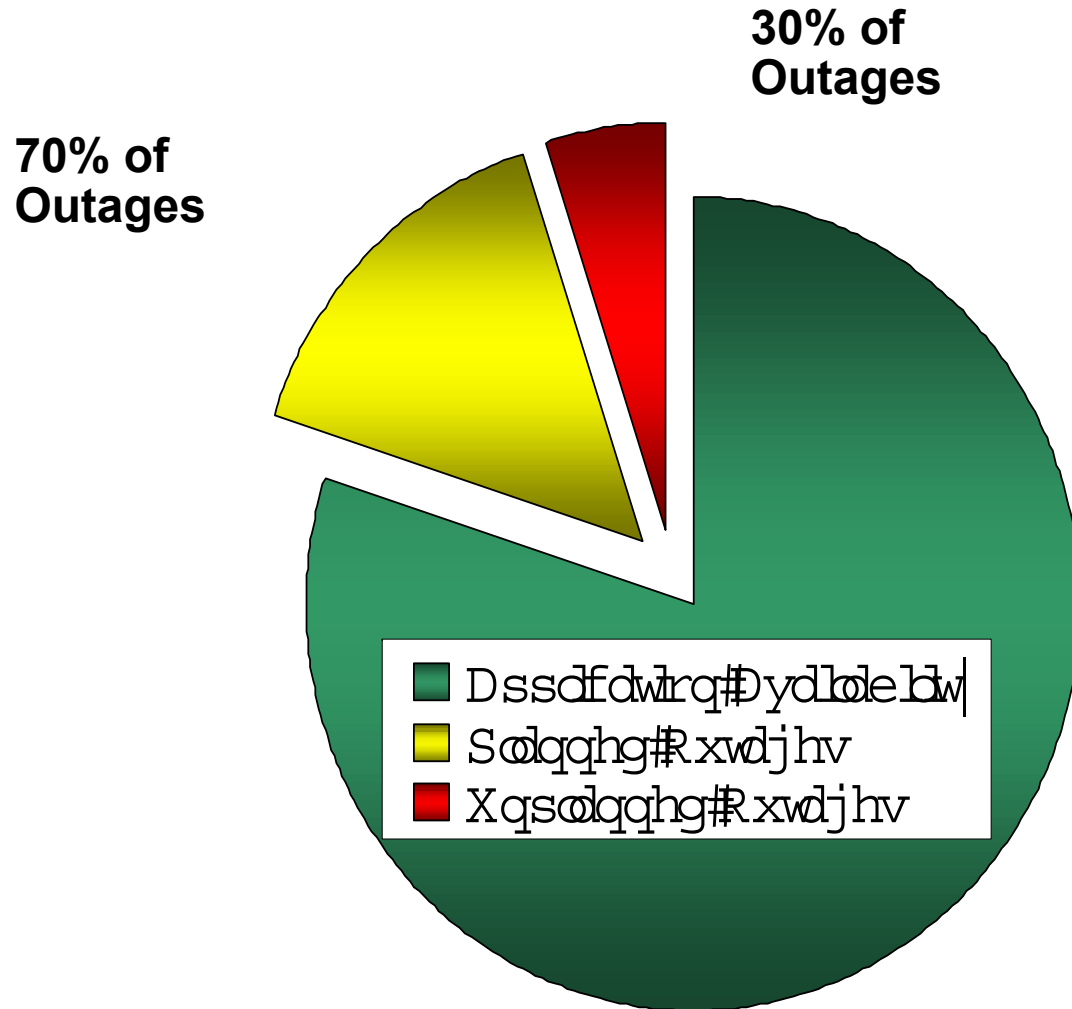


Causes of Availability Problems

- Loss of the Data Center
- Network Problems
- Loss of the Server Hardware
- Disk-Related Outages
- Operating System Failure
- Failure of the DBMS Software
- Application Problems
- Security and Authorization Problems
- Corruption of Data
- Loss of Database Objects
- Loss of Data
- Data Replication and Propagation Failures
- Severe Performance Problems
- Recovery Issues
- DBA Mistakes
- Planned Versus Unplanned Outages



Planned Versus Unplanned Outages



How Much Availability is Enough?

- Five 9's?
 - 99.999% equals 5 minutes per year!
- The Cost of Downtime
 - Varies by Industry
 - Varies by Applications
 - Requires Analysis
 - Cost of Assuring Availability vs. Cost of Downtime
- Service Level Agreements



Techniques to Improve Availability

- Perform Routine Maintenance While Systems Remain Operational
 - Online Utilities
- Automate DBA Functions
- Exploit High Availability Features of the DBMS
 - Parallelism
 - Clustering and data sharing
- Exploit Hardware Technologies
 - Storage technologies (e.g. RAID)

99.999 %

99.99 %

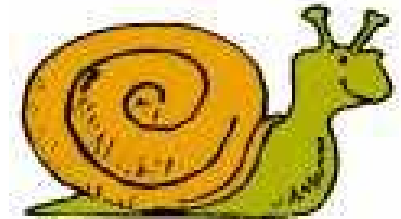
99.9 %

99.9999 %



Performance Management

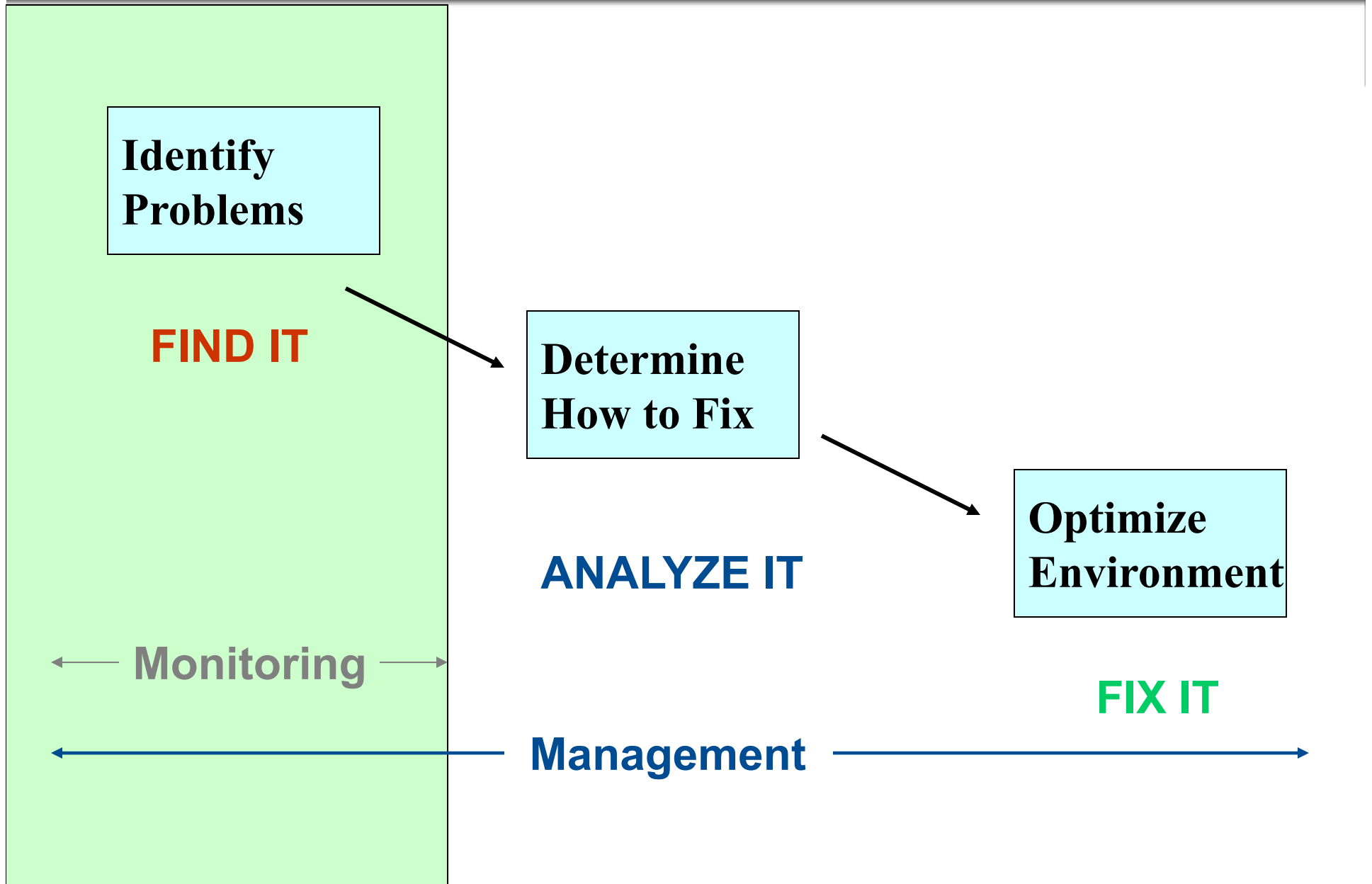
- Defining Performance
 - (Database) performance is the optimization of resource use to increase throughput and minimize contention, enabling the largest possible workload to be processed.
- Monitoring Versus Management
 - Reactive Versus Proactive
 - Identification Versus Correction
 - Historical Trending



YBWJ methodology is common



Monitoring Versus Management



General Performance Guidelines

- 80% of the results of tuning come from 20% of the tuning effort *-and-*
 - 20% of your DB2 applications cause 80% of your problems
- Tune one thing at a time
 - How else do you know whether the action helped or not?
- All tuning optimizes:
 - **CPU, I/O or concurrency**



Service Level Agreements

- Create and manage to service levels
- Without SLAs how do you know when you have tuned “enough”?
 - After all, there is always more you can do.
 - The question is, should you?
- Examples of SLAs:
 - Sub-second transaction response time
 - 99.95% up time, during the hours of 9:00 AM to 10:00 PM on weekdays
 - Batch window from Midnight to 6AM



DB Performance Management Tools

- Performance Monitors
- Performance Estimation
- Capacity Planning
- SQL Analysis
 - explain
 - query re-write
- System Analysis & Tuning
 - buffer pools
 - cache
 - memory
- Reorganization
- Compression
- Sorting
- Autonomics



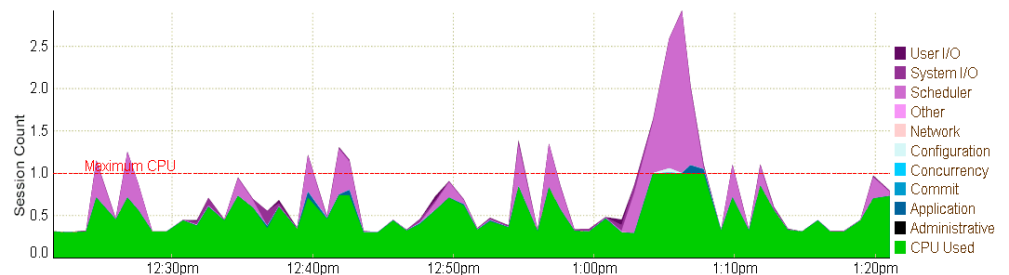
Types of Performance Management

- System
 - DBMS System
 - Interaction with O/S and other software
- Database
 - Database Design
 - Database Structure
- Application
 - Program Design
 - SQL



System Performance

- Interaction w/the OS and Allied Agents
- Hardware configuration
 - *Disk Storage and I/O*
- DBMS Installation and Configuration Issues
 - *Memory Usage*
 - *Data Cache Details*
 - *“Open” Database Objects*
 - *Database Logs*
 - *Locking and Contention*
 - *The System Catalog*
 - *Other Configuration Options*
- System Monitoring



Database Performance Issues

- Partitioning
- Indexing
- Denormalization
- Clustering
- Interleaving Data
- Free Space
- Compression
- Page Size (Block Size)
- File Placement and Allocation
 - Database Log Placement
 - Distributed Data Placement
 - Disk Allocation
- Database Reorganization
 - Determining When to Reorganize
 - Online Reorganization
 - Automation

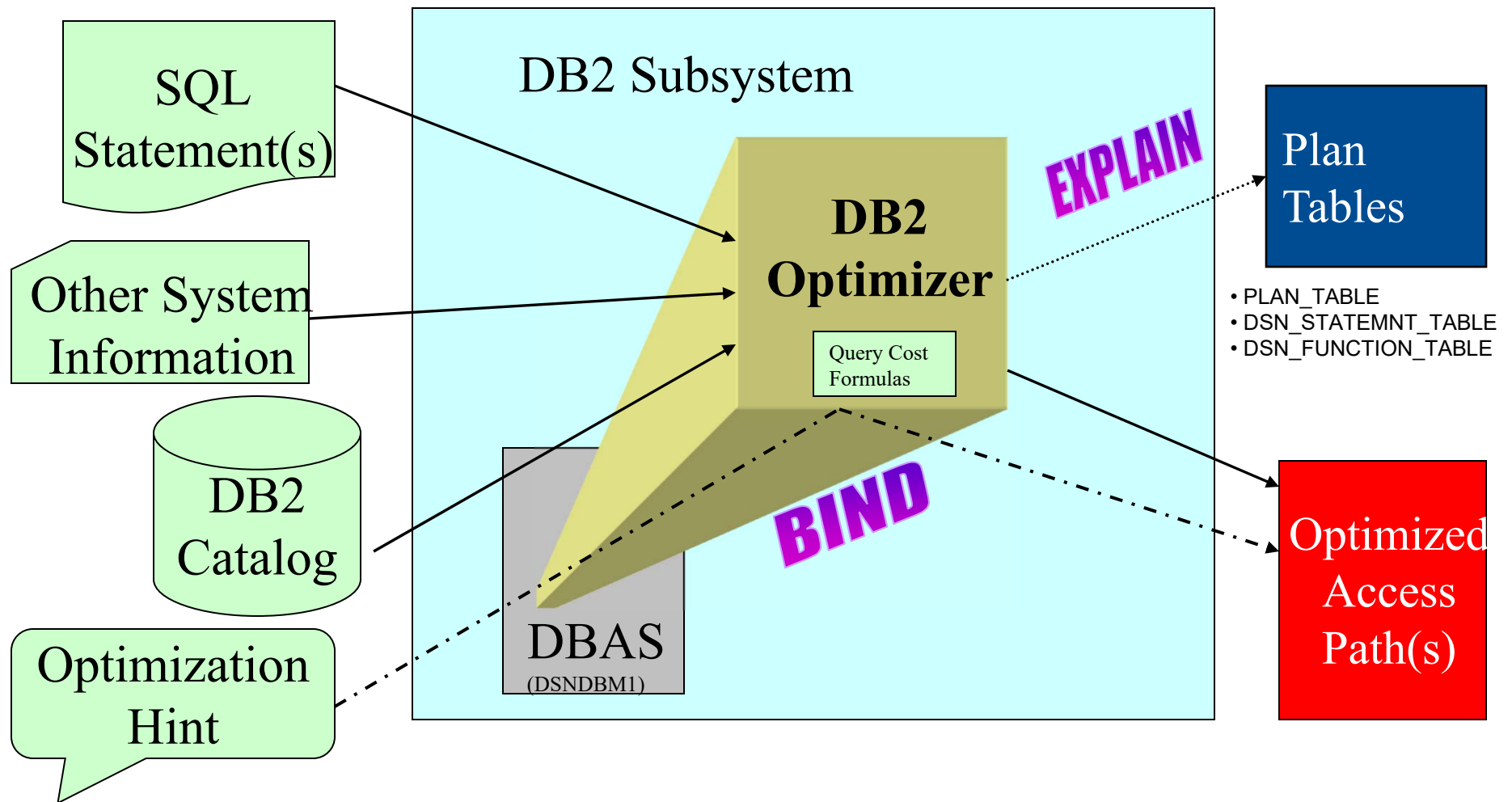


Application Performance

- Relational Optimization
 - *CPU Cost*
 - *I/O Cost*
 - *Database Statistics*
 - *Query Analysis*
 - *Access Path Choices*
 - Joining & Join Order
 - Table Scans
 - Indexed Access
 - Hashed Access
 - Parallel Access



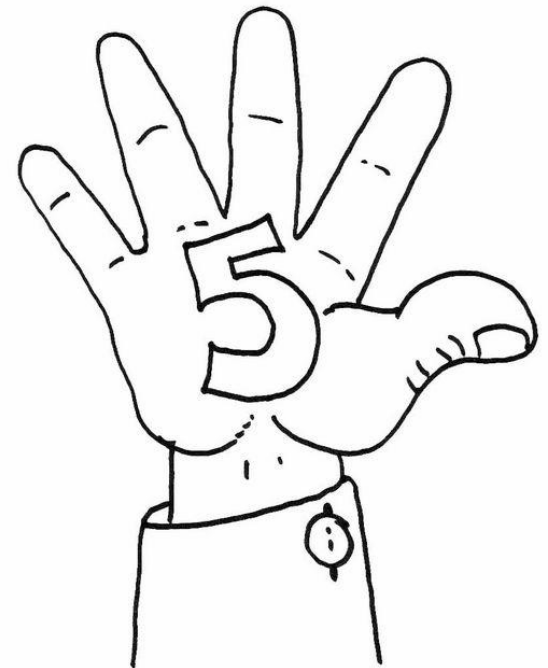
Example: Cost-Based Relational Optimization



The Five R's

When should programs be re-optimized (rebound)?

- The best approach is to perform regular REBINDs over time as your data changes.
- The shorthand for this approach is **The Five Rs**, consisting of the following steps:
 1. Real-time statistics (RTS) inspection
 2. REORG
 3. RUNSTATS
 4. REBIND
 5. Review the results



Data Integrity

◆ Database structure integrity

- Consistency Options
- Database Checking
- Memory Usage
- Additional Options

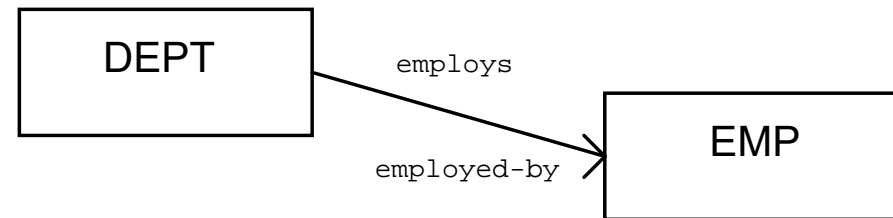
◆ Semantic data integrity

- Entity Integrity
- Referential Integrity
 - User- vs. System-Managed RI
- Unique Constraints
- Data Types
- Default Values
- Check Constraints
- Triggers



Referential Integrity

- Primary Key
- Foreign Key
- DELETE RULE
 - Restrict
 - Cascade
 - Set NULL
- Managing Referential Sets
 - Backup & Recovery
 - LOAD
 - CHECK
- Database-enforced vs. User- or Program-enforced



Database Security *in a Nutshell*

- Authentication
 - Who is it?
- Authorization
 - Who can do it?
- Encryption
 - Who can see it?
- Audit
 - Who did it?



Database Security

- Granting and Revoking Authority
 - *Types of Privileges*
 - Granting Table Privileges
 - Granting Database Object Privileges
 - Granting System Privileges
 - Granting Program and Procedure Privileges
 - *Granting to Public*
 - *Revoking Privileges*
 - Cascading Revokes
 - Chronology and Revokes
 - *Security Reporting*



Additional Database Security

- Authorization Roles and Groups
 - Roles
 - Groups
 - Limit the Number of SA Users
 - Group-Level Security and Cascading Revokes
- Trusted Context
- Other Database Security Mechanisms
 - Using Views for Security
 - Using Stored Procedures for Security
- Auditing
 - Internal and external approaches
 - Regulatory requirements
- External Security
 - Job Scheduling and Security
 - External OS and File Security



Database Backup

◆ Image Copy Backups

- Full Versus Incremental Image Copy Backups
- Merging Incremental Copies
- Database Objects and Backups
 - ◆ *Copying Indexes*
- DBMS Control
- Backup Consistency
 - ◆ *When to Create a Point of Consistency*
- Log Archival and Backup
- Determining the Backup Schedule
- DBMS Instance Backup
- Designing the DBMS Environment for Recovery
- Database Object Definition Backups

◆ Recovery Time Objectives (RTO) = recovery SLAs

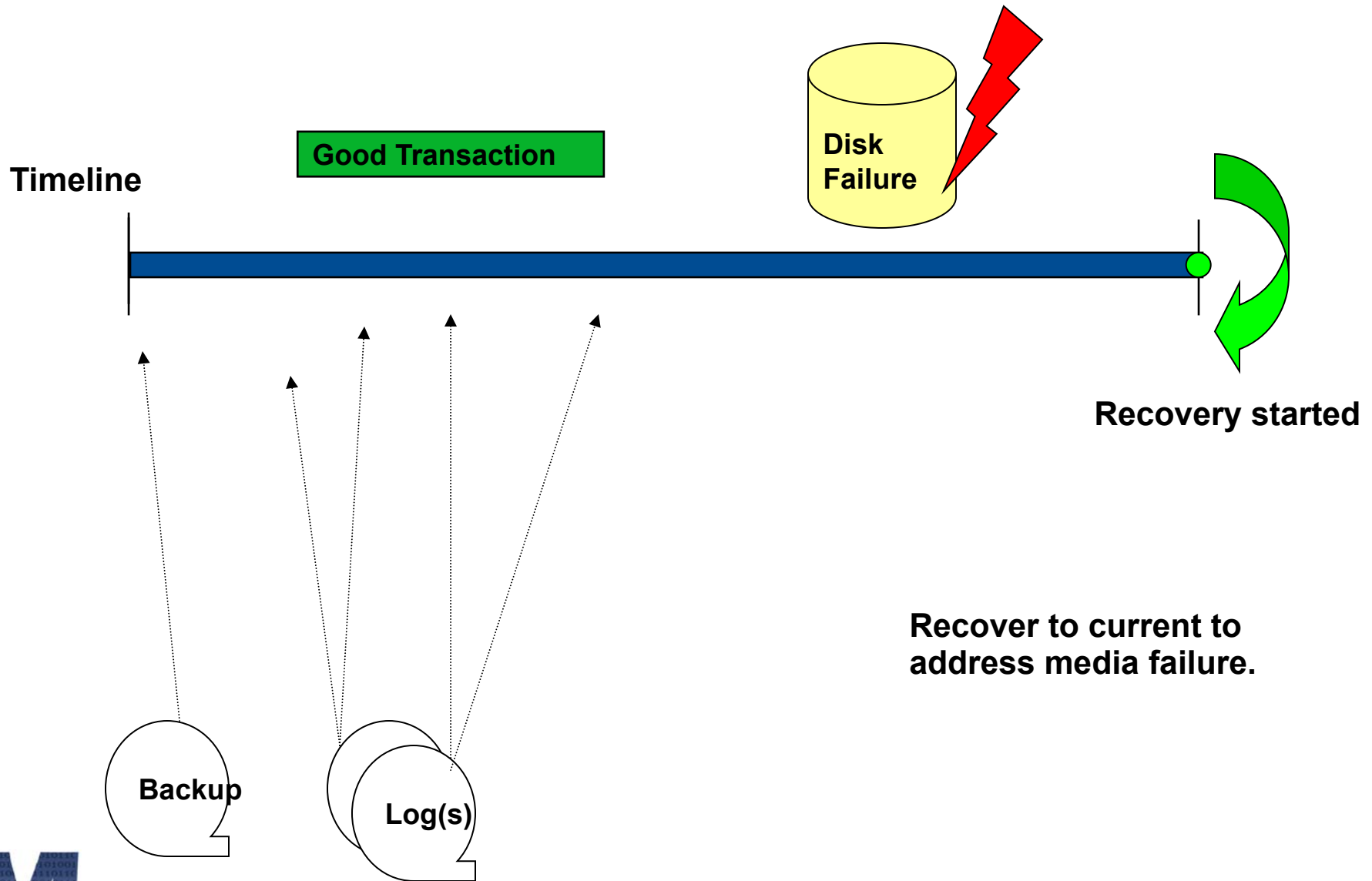


Database Recovery

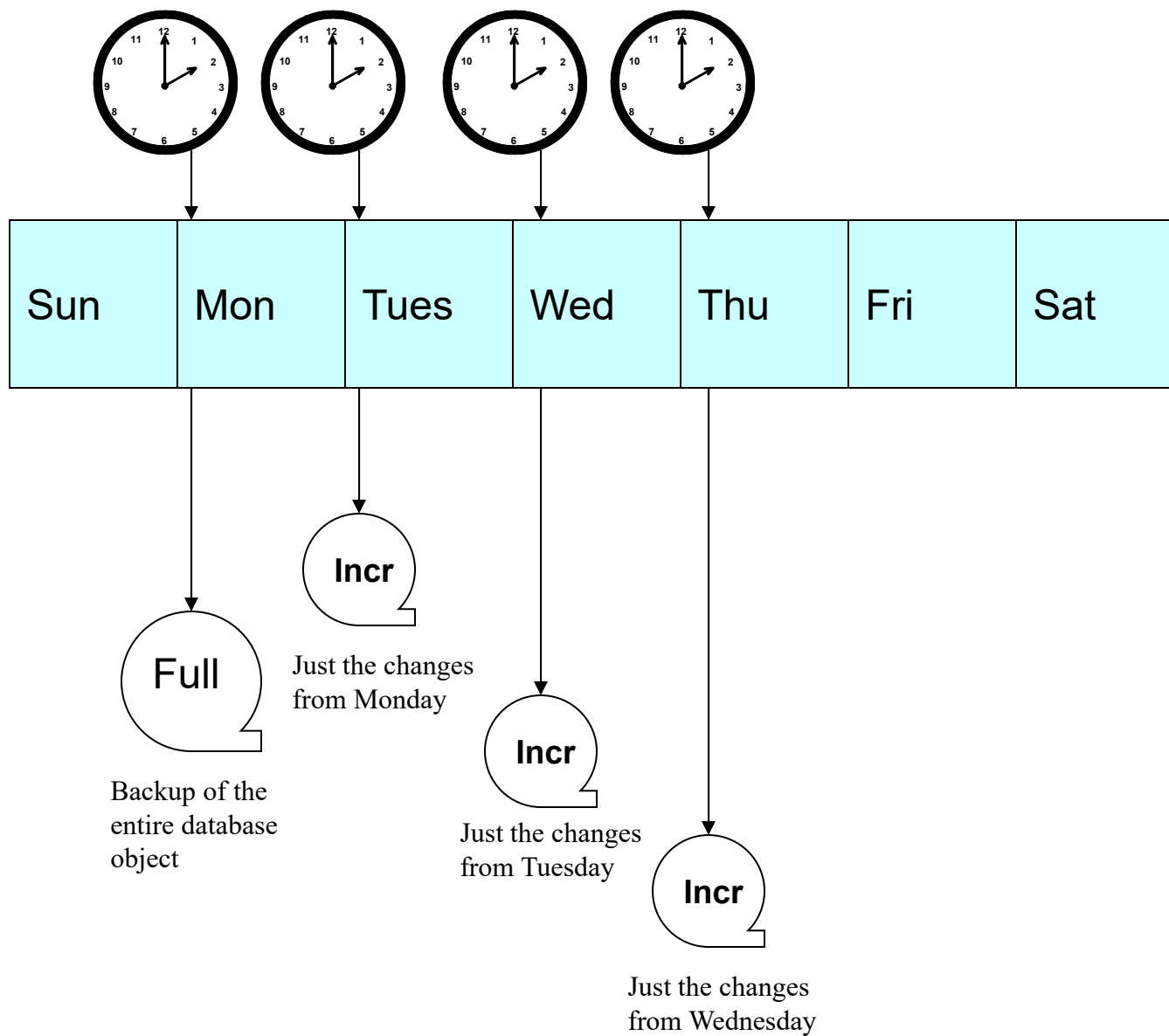
- ◆ **Recovering from Image Copy Backups**
 - **Determining Recovery Options**
 - ◆ Error Analysis
 - ◆ Image Copy Analysis
 - **Types of Recovery**
 - ◆ Recover to Current
 - ◆ Point in Time Recovery
 - ◆ SQL Based Recovery
 - **Index Recovery**
 - **Testing Your Recovery Plan**
 - **Recovering a Dropped Database Object**
 - **Recovery of Broken Blocks and Pages**



A Recovery Scenario



Another Example

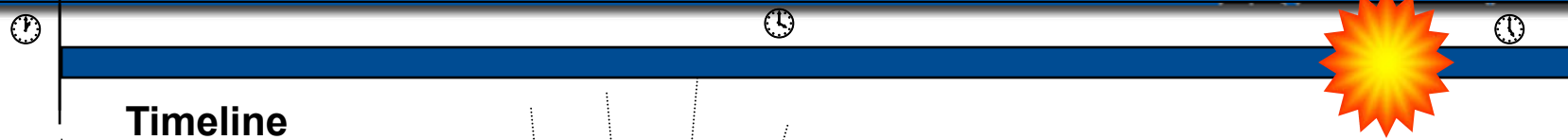


Disaster Planning

- Disaster: any unplanned, extended loss of critical business applications due to lack of computer processing capabilities for more than a 48-hour period.
 - *Sungard Recovery Services*
- DBAs must integrate database recovery into the corporate disaster recovery plan
- DBAs must test the disaster plan
- DBAs must work with the application owners/sponsors to accurately gauge the criticality of each piece of data to create a valid database disaster plan



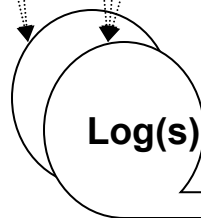
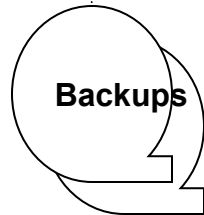
A Disaster Recovery Scenario



Timeline

Disaster occurs taking down the local site.

Database modifications applied and logged at the local site.



Replication / Mirroring is also an option (e.g. GDPS)

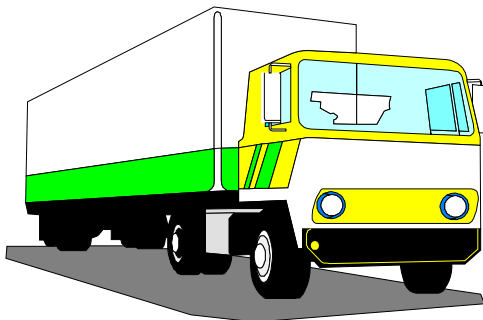


Image copy backups are taken and one is sent off-site to the remote location.



Data & Storage Management

◆ Files and Data Sets

- File Placement on Disk

◆ Space Management

- Data Page Layouts
 - Allocation Pages
 - Data Record Layouts
 - Calculating Table Size
- Index Page Layouts
 - Calculating Index Size
- Transaction Logs

◆ Storage Options

- RAID
- JBOD
- SANs
- Network Attached Storage (NAS)
- DAFS

◆ Planning for the Future

- Capacity Planning



Distributed Database Management

- Setting up a Distributed Environment
- Data Distribution Standards
- Accessing Distributed Data
- Two-Phase Commit
- Distributed Performance Problems



D
R
D
A



Data Warehouse Administration

- Data Warehouse Design
- Data Movement
 - ETL
- Data Cleansing
- Data Warehouse Scalability
- Data Warehouse Performance
 - Automated Summary Tables
- Data Freshness
- Data Content
- Data Usage
- Query Design: OLAP versus OLTP
- Financial Chargeback
- Backup and Recovery
- Integration with Production Systems That Feed the Data Warehouse



Other DBA Issues

◆ DBA Tools

- Evaluation, budget, installation, training, usage...

◆ Database Utility Management

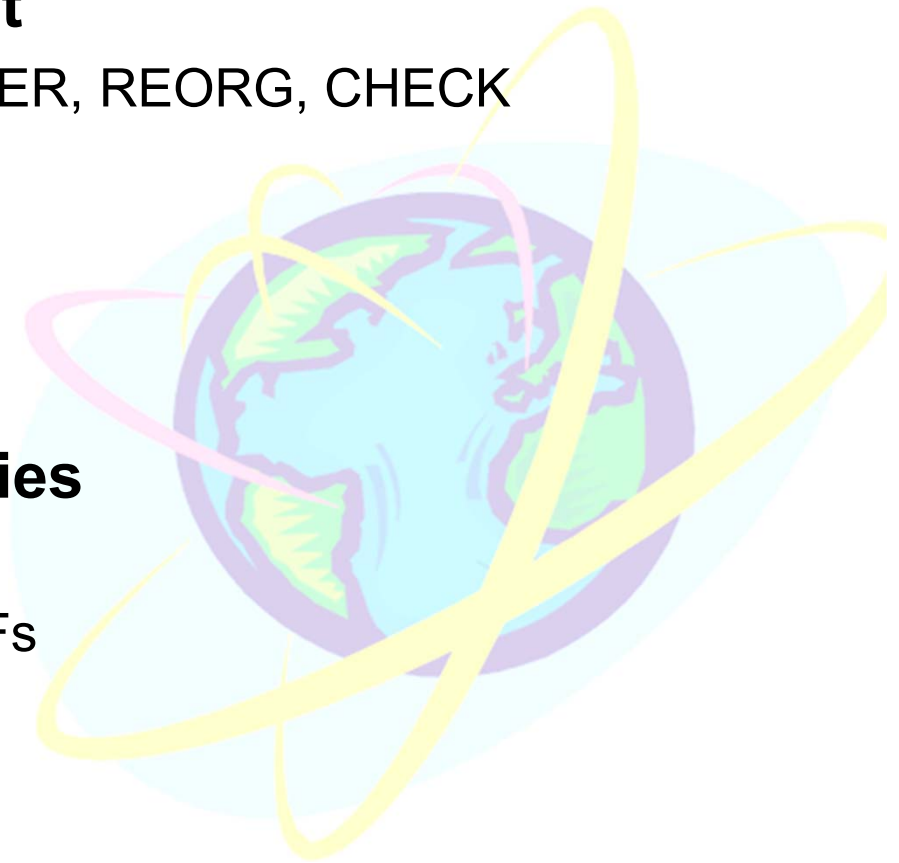
- LOAD, UNLOAD, COPY, RECOVER, REORG, CHECK

◆ Database Connectivity

- Client/Server
- Internet
- XML

◆ Coping With New Technologies

- Internet/Java/XML
- Triggers, Stored Procedures, UDFs
- BYOD: Smartphones, iPads, etc.
- NoSQL, Big Data, Analytics

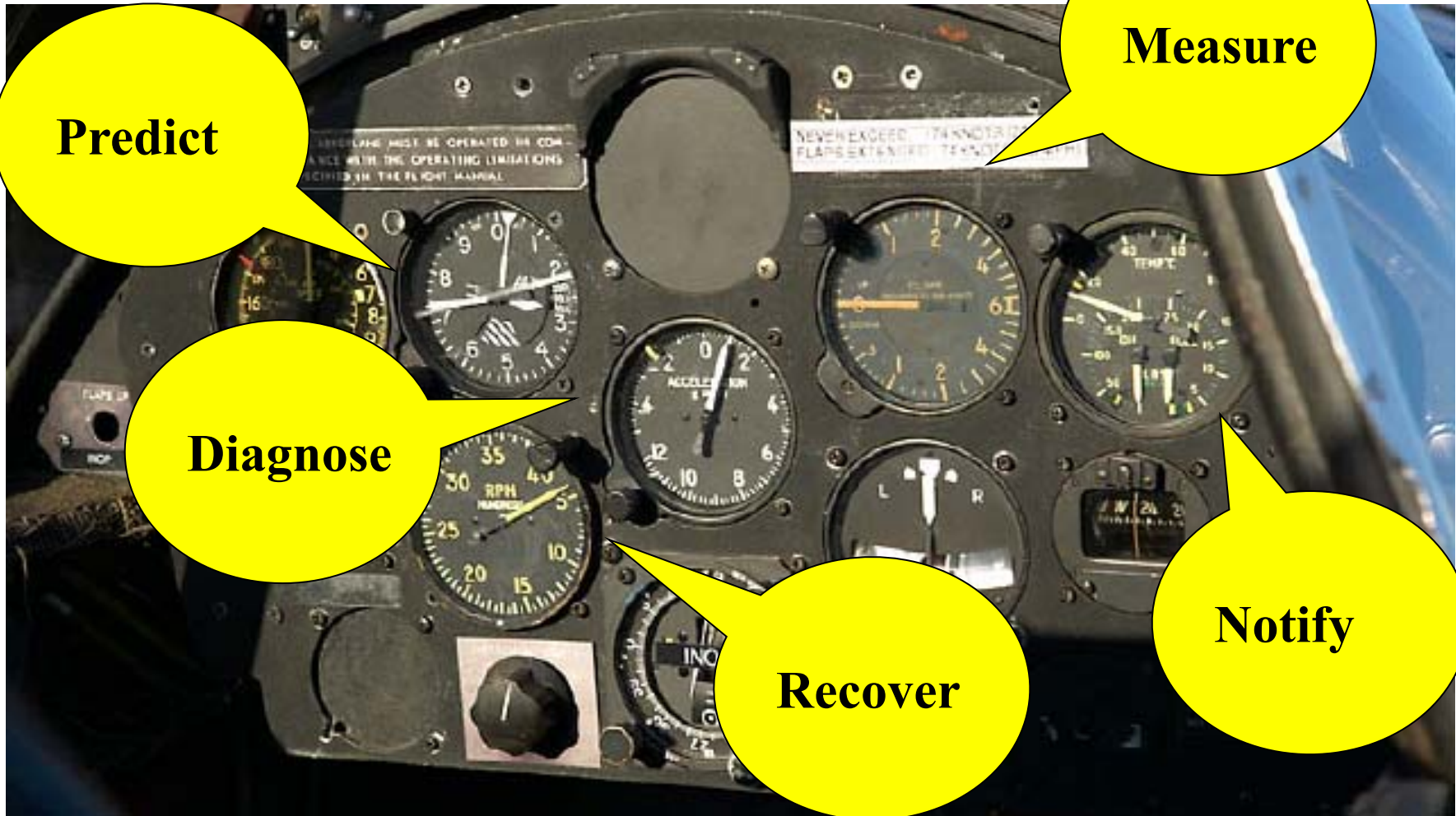


Soft Skills

- Communication
- Collaboration
- Calm disposition and demeanor
- Documentation
- Educator
- Packrat
- Thirst for knowledge



DBA Automation



Summary

- ◆ DBAs require many different types of skills
 - not just technology
 - business knowledge is important
- ◆ DBA is not an easy job
- ◆ Cooperation is required
- ◆ Communication is essential
- ◆ Technology-wise
 - need to know where to find answers
 - don't need to know all the answers off the top of your head
- ◆ No DBA is an “island”



Additional Information



<http://www.toadworld.com/>

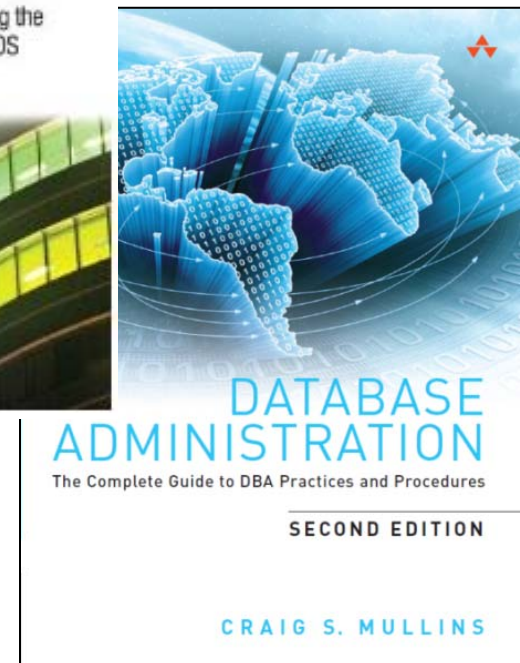
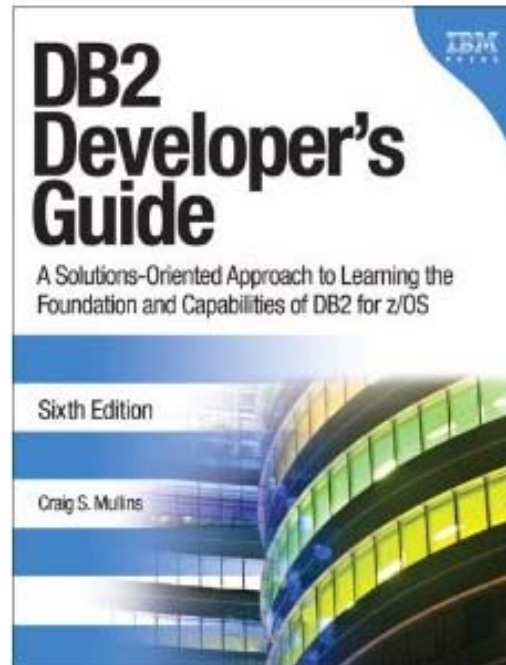
Mullins Consulting, Inc.
15 Coventry Ct
Sugar Land, TX 77479

E-mail: craig@mullinsconsultinginc.com

Web: www.mullinsconsultinginc.com



Books by Craig S. Mullins



<http://craigsmullins.com/cm-book.htm>

http://craigsmullins.com/dba_book.htm

The Management Discipline of Database Administration: *DBA Best Practices.*

Craig S. Mullins

Mullins Consulting, Inc.

craig@craigsmullins.com

<http://www.mullinsconsultinginc.com>



Sponsored by 