SPATIAL DATABASE SYSTEMS LECTURE 1

FT OKYERE

REQUIREMENTS

- HANDOUT
- PG ADMIN
- POSTGRESQL
- QGIS

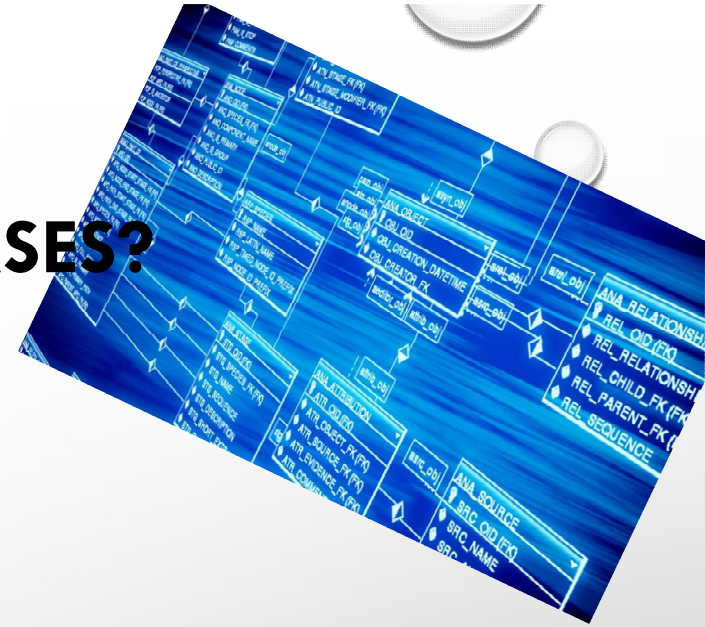
WHAT AND WHY DATABASES?

- **WHAT AND WHY DATABASES?**
- A DATABASE IS A COLLECTION OF INFORMATION THAT IS ORGANIZED
- SO THAT IT CAN EASILY BE
- ACCESSED,
- MANAGED,
- AND UPDATED..

WHAT AND WHY DATABASES?

- IN ONE VIEW, DATABASES CAN BE CLASSIFIED ACCORDING TO TYPES OF CONTENT: BIBLIOGRAPHIC,
- FULL-TEXT,
- NUMERIC, AND
- IMAGES.
- DATABASES ARE AN INTEGRAL PART OF A SPECIAL SYSTEM KNOWN AS A DATABASE MANAGEMENT SYSTEM!!

WHAT AND WHY DATABASES?



- **A DATABASE MANAGEMENT SYSTEM**
- MANAGES DATA EFFICIENTLY AND
- ALLOWS USERS TO PERFORM MULTIPLE TASKS WITH EASE.
- A DATABASE MANAGEMENT SYSTEM STORES, ORGANIZES AND MANAGES A **LARGE** AMOUNT OF INFORMATION WITHIN A SINGLE SOFTWARE APPLICATION.
- USE OF THIS SYSTEM INCREASES EFFICIENCY OF BUSINESS OPERATIONS AND REDUCES OVERALL COSTS.

WHAT AND WHY DATABASES?

- DATABASE MANAGEMENT SYSTEMS ARE IMPORTANT TO BUSINESSES AND ORGANIZATIONS BECAUSE THEY PROVIDE A HIGHLY EFFICIENT METHOD FOR HANDLING MULTIPLE TYPES OF DATA.
- SOME OF THE DATA THAT ARE EASILY MANAGED WITH THIS TYPE OF SYSTEM INCLUDE: EMPLOYEE RECORDS,
- STUDENT INFORMATION,
- PAYROLL,
- ACCOUNTING, PROJECT MANAGEMENT, INVENTORY AND LIBRARY BOOKS. THESE SYSTEMS ARE BUILT TO BE EXTREMELY VERSATILE.

WHY DATABASES?

- WITHOUT DATABASE MANAGEMENT,
- TASKS HAVE TO BE DONE MANUALLY AND
- TAKE MORE TIME.
- DATA CAN BE CATEGORIZED AND STRUCTURED TO SUIT THE NEEDS OF THE COMPANY OR ORGANIZATION.
- DATA IS ENTERED INTO THE SYSTEM AND ACCESSED ON A ROUTINE BASIS BY ASSIGNED USERS.
- EACH USER MAY HAVE AN ASSIGNED **PASSWORD** TO GAIN **ACCESS** TO THEIR PART OF THE SYSTEM. MULTIPLE USERS CAN USE THE SYSTEM AT THE SAME TIME IN DIFFERENT WAYS.

WHY DATABASES?

- FOR EXAMPLE, A COMPANY'S HUMAN RESOURCES DEPARTMENT USES THE DATABASE TO MANAGE EMPLOYEE RECORDS, DISTRIBUTE LEGAL INFORMATION TO EMPLOYEES AND CREATE UPDATED HIRING REPORTS.
- A MANUFACTURER MIGHT USE THIS TYPE OF SYSTEM TO KEEP TRACK OF PRODUCTION, INVENTORY AND DISTRIBUTION.
- IN BOTH SCENARIOS, THE DATABASE MANAGEMENT SYSTEM OPERATES TO CREATE A SMOOTHER AND MORE ORGANIZED WORKING ENVIRONMENT.

WHAT 'R' DATABASES?

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28

- A SIMPLE DATABASE HAS A SINGLE TABLE WITH
- **R**OWS FOR THE DATA AND
- COLUMNS THAT DEFINE THE DATA ELEMENTS.
- FOR AN ADDRESS BOOK, THE TABLE COLUMNS DEFINE DATA ELEMENTS SUCH AS NAME, ADDRESS, CITY, STATE AND PHONE NUMBER,
- WHILE A TABLE ROW, OR RECORD, CONTAINS DATA FOR EACH PERSON IN THE BOOK.
- THE QUERY LANGUAGE PROVIDES A WAY TO FIND SPECIFIC TYPES OF DATA IN EACH RECORD AND RETURN RESULTS THAT MATCH THE CRITERIA. THESE RESULTS DISPLAY IN A FORM THAT USES THE DEFINED DATA ELEMENTS BUT ONLY SHOWS RECORDS THAT MEET THE CRITERIA. THESE THREE COMPONENTS MAKE UP ALMOST EVERY TYPE OF DATABASE.

WHAT 'R' DATABASES?

- **RELATIONAL DATABASES** USE MULTIPLE TABLES AND DEFINE RELATIONSHIPS BETWEEN THEM USING A SCHEMA IN ADDITION TO DATA ELEMENTS.
- RECORDS AND DATA ELEMENTS FROM EACH TABLE MERGE, BASED ON THE QUERY, AND DISPLAY IN THE FORM.
- ROUTINELY USED QUERIES OFTEN BECOME REPORTS. A REPORT USES THE SAME QUERY BUT REPORTS ON CHANGES IN DATA OVER TIME.

WHAT ‘R’ DATABASES?

- THERE ARE FIVE MAJOR COMPONENTS IN A DATABASE ENVIRONMENT:
- DATA,
- HARDWARE,
- SOFTWARE,
- PEOPLE AND
- PROCEDURES.

WHAT 'R' DATABASES?

- THE **DATA** IS A COLLECTION OF FACTS, TYPICALLY RELATED.
- THE **HARDWARE** IS THE PHYSICAL DEVICES IN THE DATABASE ENVIRONMENT.
- OPERATING SYSTEMS, DATABASE MANAGEMENT SYSTEMS AND APPLICATIONS MAKE UP THE **SOFTWARE**.
- EXAMPLES OF **PEOPLE** IN THE DATABASE ENVIRONMENT ARE THE SYSTEM ADMINISTRATOR, PROGRAMMERS AND END USERS.
- **PROCEDURES** ARE THE INSTRUCTIONS AND RULES FOR THE DATABASE.

DIFFERENCE BETWEEN SPATIAL DATABASE AND NON-SPATIAL DATABASE

- BASICALLY, A DATABASE CONSISTS OF AN ORGANIZED COLLECTION OF DATA FOR ONE OR MORE USES, TYPICALLY IN DIGITAL FORM.

  SPATIAL DATABASE: HAS THE ABILITY TO STORE AND ACCESS BOTH LOCATION/SPATIAL INFORMATION AND ATTRIBUTES/NON-SPATIAL INFORMATION.

  NON-SPATIAL DATABASE: HAS THE ABILITY TO STORE AND ACCESS ONLY ATTRIBUTES/NON-SPATIAL INFORMATION.

DIFFERENCE BETWEEN SPATIAL DATABASE AND NON-SPATIAL DATABASE

- BASICALLY, THE DIFFERENCES ARE IN THE STORAGE, FUNCTION, AND QUERY CAPABILITIES BETWEEN THE TWO.
- A SPATIAL DATABASE SUPPORTS SPECIAL **DATA TYPES** FOR GEOMETRIC OBJECTS AND ALLOWS YOU TO STORE **GEOMETRIC DATA** (USUALLY OF A GEOGRAPHIC NATURE**) IN TABLES** WHILE A NON-SPATIAL DATABASE DOESN'T SUPPORT SUCH.
- A SPATIAL DATABASE PROVIDES **SPECIAL FUNCTIONS AND INDEXES** FOR **QUERYING AND MANIPULATING GEOSPATIAL DATA** USING SOMETHING LIKE STRUCTURED QUERY LANGUAGE (SQL) WHILE NON-SPATIAL DATABASE DOESN'T PROVIDE SUCH FUNCTIONS AND INDEXES.

DIFFERENCE BETWEEN SPATIAL DATABASE AND NON-SPATIAL DATABASE

- A SPATIAL DATABASE IS OFTEN USED AS **A STORAGE CONTAINER** FOR GEOSPATIAL DATA, BUT IT CAN DO MUCH MORE THAN THAT. WHILE NON-SPATIAL DATABASE IS OFTEN USED AS A STORAGE CONTAINER FOR NON-SPATIAL DATA.
- A SPATIAL DATABASE USES **SPATIAL QUERY IN GEOMETRIC FUNCTIONS** TO **ANSWER QUESTIONS ABOUT SPACE** AND OBJECTS IN SPACE. WHILE NON-SPATIAL DATABASE DON'T SUPPORT SPATIAL QUERIES.

  IN ADDITION TO BEING ABLE TO **ANSWER QUESTIONS ABOUT THE USE OF SPACE,** SPATIAL DATABASE FUNCTIONS ALLOW YOU TO CREATE AND MODIFY OBJECTS IN SPACE. THIS PORTION OF SPATIAL ANALYSIS IS OFTEN REFERRED TO AS GEOMETRIC OR SPATIAL PROCESSING.

  A SPATIALLY ENABLED DATABASE CAN INTRINSICALLY WORK WITH DATA TYPES LIKE RIVERS (MODELED AS **LINESTRINGS**), LAND PARCELS (MODELED AS **POLYGONS**), AND TREES (MODELED AS **POINTS**). WHILE NON-SPATIAL DATABASE CAN'T WORK WITH THESE FORMS OF MODELS.

EXAMPLES OF SPATIAL AND NON-SPATIAL DATABASES

- THERE ARE SEVERAL EXAMPLES OF SPATIAL AND NON-SPATIAL DATABASE SERVERS AVAILABLE, BUT THE TOP COMMON ONCE INCLUDE:

❖ **SPATIALITE** – SQLITE

❖ **ORACLE SPATIAL/LOCATOR** – ORACLE

❖ **MYSQL SPATIAL** – MYSQL

❖ **POSTGIS** – POSTGRESQL

❖ **DB2 SPATIAL** - IBM DB2/INFORMIX

❖ MSSQL SPATIAL - MSSQL SERVER

❖ SPATIAL HADOOP – HADOOP

❖ MS ACCESS - THIS IS NON-SPATIAL DATABASE, IT HAS NO SUPPORT FOR SPATIAL DATABASE AT THE TIME OF WRITING.

DATA ADMINISTRATION AND THE DATABASE ADMINISTRATOR

- **DATA ADMINISTRATION** OR **DATA RESOURCE MANAGEMENT** IS AN ORGANIZATIONAL FUNCTION THAT PLANS, ORGANIZES, DESCRIBES AND CONTROLS DATA RESOURCES.
- DATA RESOURCES ARE USUALLY AS STORED IN DATABASES UNDER A DATABASE MANAGEMENT SYSTEM OR OTHER SOFTWARE SUCH AS ELECTRONIC SPREADSHEETS.
- IN MANY SMALLER ORGANIZATIONS, DATA ADMINISTRATION IS PERFORMED OCCASIONALLY, OR IS A SMALL COMPONENT OF THE DATABASE ADMINISTRATOR'S WORK.
- IN THE CONTEXT OF INFORMATION SYSTEMS DEVELOPMENT, DATA ADMINISTRATION IDEALLY BEGINS AT SYSTEM CONCEPTION, ENSURING THERE IS A DATA DICTIONARY TO HELP MAINTAIN CONSISTENCY, AVOID REDUNDANCY, AND MODEL THE DATABASE SO AS TO MAKE IT LOGICAL AND USABLE, BY MEANS OF DATA MODELING, INCLUDING DATABASE NORMALIZATION TECHNIQUES.

DATA RESOURCE MANAGEMENT

- DATA RESOURCE MANAGEMENT IS "THE DEVELOPMENT AND EXECUTION OF **ARCHITECTURES**, **POLICIES**, **PRACTICES** AND **PROCEDURES** THAT PROPERLY MANAGE THE FULL DATA LIFECYCLE NEEDS OF AN ENTERPRISE".
- DATA RESOURCE MANAGEMENT MAY BE THOUGHT OF AS A MANAGERIAL ACTIVITY THAT APPLIES INFORMATION SYSTEM AND OTHER DATA MANAGEMENT TOOLS TO THE TASK OF MANAGING AN ORGANIZATION'S DATA RESOURCE TO MEET A COMPANY'S BUSINESS NEEDS, AND THE INFORMATION THEY PROVIDE TO THEIR SHAREHOLDERS.
- FROM THE PERSPECTIVE OF DATABASE DESIGN, IT REFERS TO THE DEVELOPMENT AND MAINTENANCE OF DATA MODELS TO FACILITATE DATA SHARING BETWEEN DIFFERENT SYSTEMS, PARTICULARLY IN A CORPORATE CONTEXT.
- DATA RESOURCE MANAGEMENT IS ALSO CONCERNED WITH BOTH DATA QUALITY AND COMPATIBILITY BETWEEN DATA MODELS.

DATA RESOURCE MANAGEMENT

- SINCE THE **BEGINNING OF THE INFORMATION AGE,** BUSINESSES NEED ALL TYPES OF DATA ON THEIR BUSINESS ACTIVITY. WITH EACH DATA CREATED,
- ***WHEN A BUSINESS TRANSACTION IS MADE, NEW DATA IS CREATED !!!***
- WITH THESE DATA, NEW DIRECTION IS NEEDED THAT FOCUSES ON MANAGING DATA AS A CRITICAL RESOURCE OF THE ORGANIZATION TO DIRECTLY SUPPORT ITS BUSINESS ACTIVITIES.
- THE DATA RESOURCE MUST BE MANAGED WITH THE SAME INTENSITY AND FORMALITY THAT OTHER CRITICAL RESOURCES ARE MANAGED.
- ORGANIZATIONS MUST EMPHASIZE THE INFORMATION ASPECT OF INFORMATION TECHNOLOGY, DETERMINE THE DATA NEEDED TO SUPPORT THE BUSINESS, AND THEN USE APPROPRIATE TECHNOLOGY TO BUILD AND MAINTAIN A HIGH-QUALITY DATA RESOURCE THAT PROVIDES THAT SUPPORT.

DATA RESOURCE MANAGEMENT

- DATA RESOURCE QUALITY IS A MEASURE OF **HOW WELL THE ORGANIZATION'S DATA RESOURCE SUPPORTS THE CURRENT** AND THE FUTURE BUSINESS INFORMATION DEMAND OF THE ORGANIZATION.
- IT MUST SUPPORT BOTH THE CURRENT AND THE FUTURE BUSINESS INFORMATION DEMAND.
- THE ULTIMATE DATA RESOURCE QUALITY IS STABILITY ACROSS CHANGING BUSINESS NEEDS AND CHANGING TECHNOLOGY.

DATA RESOURCE MANAGEMENT

- A CORPORATE DATA RESOURCE MUST BE DEVELOPED WITHIN SINGLE, ORGANIZATION-WIDE **COMMON DATA ARCHITECTURE**. A DATA ARCHITECTURE IS THE SCIENCE AND METHOD OF DESIGNING AND CONSTRUCTING A DATA RESOURCE THAT IS BUSINESS DRIVEN, BASED ON REAL-WORLD OBJECTS AND EVENTS AS PERCEIVED BY THE ORGANIZATION, AND IMPLEMENTED INTO APPROPRIATE OPERATING ENVIRONMENTS.
- IT IS THE OVERALL STRUCTURE OF A DATA RESOURCE THAT PROVIDES A CONSISTENT FOUNDATION ACROSS ORGANIZATIONAL BOUNDARIES TO PROVIDE EASILY IDENTIFIABLE, READILY AVAILABLE, HIGH-QUALITY DATA TO SUPPORT **THE BUSINESS INFORMATION DEMAND.**
- THE COMMON DATA ARCHITECTURE IS A FORMAL, COMPREHENSIVE DATA ARCHITECTURE THAT PROVIDES A COMMON CONTEXT WITHIN WHICH ALL DATA AT AN ORGANIZATION'S DISPOSAL ARE UNDERSTOOD AND INTEGRATED.

DATABASE ADMINISTRATION AND AUTOMATION

- **DATABASE ADMINISTRATION** IS THE FUNCTION OF MANAGING AND MAINTAINING DATABASE MANAGEMENT SYSTEMS (DBMS) SOFTWARE.
- MAINSTREAM DBMS SOFTWARE SUCH AS ORACLE, IBM DB2 AND MICROSOFT SQL SERVER NEED ONGOING MANAGEMENT.
- AS SUCH, CORPORATIONS THAT USE DBMS SOFTWARE OFTEN HIRE SPECIALIZED IT (INFORMATION TECHNOLOGY) PERSONNEL CALLED DATABASE ADMINISTRATORS OR DBAS.

DBA RESPONSIBILITIES

- INSTALLATION,
- CONFIGURATION AND
- UPGRADING OF DATABASE SERVER SOFTWARE AND RELATED PRODUCTS.
- EVALUATE DATABASE FEATURES AND DATABASE RELATED PRODUCTS.
- ESTABLISH AND MAINTAIN SOUND BACKUP AND RECOVERY POLICIES AND PROCEDURES.
- TAKE CARE OF THE DATABASE DESIGN AND IMPLEMENTATION.

DBA RESPONSIBILITIES

- IMPLEMENT AND MAINTAIN **DATABASE SECURITY** (CREATE AND MAINTAIN USERS AND ROLES, ASSIGN **PRIVILEGES**).
- DATABASE TUNING AND PERFORMANCE MONITORING.
- APPLICATION TUNING AND PERFORMANCE MONITORING.
- SETUP AND MAINTAIN DOCUMENTATION AND **STANDARDS**.
- PLAN **GROWTH** AND CHANGES (CAPACITY PLANNING).
- WORK AS PART OF A **TEAM** AND PROVIDE 24X7 SUPPORT WHEN REQUIRED.
- DO GENERAL TECHNICAL **TROUBLESHOOTING** AND GIVE CONS.
- DATABASE RECOVERY.

TYPES OF DATABASE ADMINISTRATION

- **SYSTEMS DBAS** (ALSO REFERRED TO AS PHYSICAL DBAS, OPERATIONS DBAS OR PRODUCTION SUPPORT DBAS): FOCUS ON THE **PHYSICAL ASPECTS** OF DATABASE ADMINISTRATION SUCH AS DBMS INSTALLATION, CONFIGURATION, PATCHING, UPGRADES, BACKUPS, RESTORES, REFRESHES, PERFORMANCE OPTIMIZATION, MAINTENANCE AND DISASTER RECOVERY.
- **DEVELOPMENT DBAS**: FOCUS ON THE LOGICAL **AND DEVELOPMENT ASPECTS OF DATABASE** ADMINISTRATION SUCH AS DATA MODEL DESIGN AND MAINTENANCE, DDL (DATA DEFINITION LANGUAGE) GENERATION, SQL WRITING AND TUNING, CODING STORED PROCEDURES, COLLABORATING WITH DEVELOPERS TO HELP CHOOSE THE MOST APPROPRIATE DBMS FEATURE/FUNCTIONALITY AND OTHER PRE-PRODUCTION ACTIVITIES.
- **APPLICATION DBAS**: USUALLY FOUND IN ORGANIZATIONS THAT HAVE PURCHASED 3RD PARTY APPLICATION SOFTWARE SUCH AS **ERP** (ENTERPRISE RESOURCE PLANNING) AND CRM (CUSTOMER RELATIONSHIP MANAGEMENT) SYSTEMS.

TYPES OF DATABASE ADMINISTRATION

- THEY USUALLY MANAGE ALL THE APPLICATION COMPONENTS THAT INTERACT WITH THE DATABASE AND CARRY OUT ACTIVITIES SUCH AS APPLICATION INSTALLATION AND PATCHING, APPLICATION UPGRADES, DATABASE CLONING, BUILDING AND RUNNING DATA CLEANUP ROUTINES, DATA LOAD PROCESS MANAGEMENT, ETC.
- WHILE **INDIVIDUALS USUALLY SPECIALIZE IN ONE TYPE OF DATABASE ADMINISTRATION**, IN SMALLER ORGANIZATIONS, IT IS NOT UNCOMMON TO FIND A SINGLE INDIVIDUAL OR GROUP PERFORMING MORE THAN ONE TYPE OF DATABASE ADMINISTRATION.

NATURE OF DATABASE ADMINISTRATION

- THE DEGREE TO WHICH THE ADMINISTRATION OF A DATABASE IS AUTOMATED DICTATES **THE SKILLS AND PERSONNEL** REQUIRED TO MANAGE DATABASES.
- ALTERNATIVELY AN ORGANIZATION MIGHT CHOOSE TO **AUTOMATE A SIGNIFICANT AMOUNT OF THE WORK** THAT COULD BE DONE MANUALLY THEREFORE REDUCING THE SKILLS REQUIRED TO PERFORM TASKS. AS AUTOMATION INCREASES,
- **DATABASE ADMINISTRATION WORK IS COMPLEX**, REPETITIVE, TIME-CONSUMING AND REQUIRES SIGNIFICANT **TRAINING**. SINCE DATABASES HOLD VALUABLE AND MISSION-CRITICAL DATA, COMPANIES USUALLY LOOK FOR CANDIDATES WITH MULTIPLE YEARS OF EXPERIENCE. DATABASE ADMINISTRATION OFTEN REQUIRES DBAS TO PUT IN WORK DURING OFF-HOURS

NATURE OF DATABASE ADMINISTRATION

- ONE KEY SKILL REQUIRED AND OFTEN OVERLOOKED WHEN SELECTING A DBA IS **DATABASE RECOVERY** (UNDER DISASTER RECOVERY). IT IS NOT A CASE OF "IF" BUT A **CASE OF "WHEN" A DATABASE SUFFERS A FAILURE,** RANGING FROM **A SIMPLE FAILURE** TO A FULL CATASTROPHIC FAILURE. THE FAILURE MAY BE DATA CORRUPTION, MEDIA FAILURE, OR USER INDUCED ERRORS.
- IN EITHER SITUATION THE DBA MUST HAVE **THE SKILLS TO RECOVER THE DATABASE TO A GIVEN POINT IN TIME** TO PREVENT A LOSS OF DATA. A HIGHLY SKILLED DBA CAN SPEND A FEW MINUTES OR EXCEEDINGLY LONG HOURS TO GET THE DATABASE BACK TO THE OPERATIONAL POINT.

DATABASE ADMINISTRATION TOOLS

- OFTEN, THE DBMS SOFTWARE COMES WITH CERTAIN TOOLS TO HELP DBAS MANAGE THE DBMS. SUCH TOOLS ARE CALLED NATIVE TOOLS.
- FOR EXAMPLE, MICROSOFT SQL SERVER COMES WITH **SQL SERVER MANAGEMENT STUDIO AND ORACLE HAS TOOLS SUCH AS SQL*PLUS AND ORACLE ENTERPRISE MANAGER/GRID CONTROL.** IN ADDITION, **3RD PARTIES** SUCH AS BMC, QUEST SOFTWARE, EMBARCADERO TECHNOLOGIES OFFER GUI TOOLS TO MONITOR THE DBMS AND HELP DBAS CARRY OUT CERTAIN FUNCTIONS INSIDE THE DATABASE MORE EASILY. **PGSL COMES WITH PGADMIN**
- THE PROCESS OF CREATING A NEW DATABASE CAN CONSIST OF HUNDREDS OR **THOUSANDS OF UNIQUE STEPS** FROM SATISFYING PREREQUISITES TO CONFIGURING BACKUPS WHERE EACH STEP MUST BE SUCCESSFUL BEFORE THE NEXT CAN START.
- A **HUMAN** CANNOT BE EXPECTED TO COMPLETE THIS PROCEDURE IN THE SAME EXACT WAY TIME AFTER TIME – EXACTLY THE GOAL WHEN MULTIPLE DATABASES EXIST.
- AS THE NUMBER OF DBAS GROWS, **WITHOUT AUTOMATION** THE NUMBER OF UNIQUE **CONFIGURATIONS FREQUENTLY GROWS TO BE COSTLY/DIFFICULT TO SUPPORT**. ALL OF THESE COMPLICATED PROCEDURES CAN BE MODELED BY THE BEST DBAS INTO DATABASE AUTOMATION SOFTWARE AND EXECUTED BY THE STANDARD DBAS.