EXCEL Homework 2

# Popular Terms

1. Data warehouse - A data warehouse essentially combines information from several sources into one comprehensive database, used to guide management decisions. For example, in the business world, a data warehouse might incorporate customer information from a company's point-of-sale systems (the cash registers), its website, its mailing lists and its comment cards.
2. Transactional , or Real Time database - A **transactional database** is a database management system (DBMS) that has the capability to roll back or undo a database transaction or operation if it is not completed appropriately. Since data integrity and consistency are critical to proper database operation, most transactions use SQL or a SQL-like language to conduct operations using the following pattern:

Step 1 – Initiate the transaction.

Step 2 – Execute the provided set of data queries or manipulations.

Step 3 – Commit the transaction and complete it if there are no errors.

Step 4 – If an error occurs in the transaction, roll it back and end the operation.

For example a person’s ATM withdrawal.

A **Real Time database** is a database system which uses real-time processing to handle workloads whose state is constantly changing. Real-time processing means that a transaction is processed fast enough for the result to come back and be acted on right away. They use timing constraints that represent a certain range of values for which the data are valid. This range is called temporal validity. For example, a stock market changes very rapidly and is dynamic, yet a database has to keep track of current values for all of the markets of the New York Stock Exchange.

1. Analytical database - An analytical database, is a read-only system that stores historical data on business. Business analysts, corporate executives and other workers can run queries and reports against an analytic database. An analytic database is specifically designed to support business intelligence and analytic applications. This differentiates it from an operational, transactional database, which is used for transaction processing. For example, a business might store metrics related to sales performance and inventory levels.
2. Time series database - A software system that is optimized for handling time series data, arrays of numbers indexed by time (a datetime or a datetime range). In short, it is any data that has a timestamp.  A time series database is optimized to meet the challenges of handling massive amounts of data from thousands or more devices. The ability to scale out, up, and down as data grows is a basic requirement for time series use cases. For example tracking temperature or tide levels over time.
3. Flat file database - A flat file database is a database that stores data in a plain text file. Each line of the text file holds one record, with fields separated by delimiters, such as commas or tabs. It is basically a giant collection of data in which the tables and records have no relation between any other tables. For example the information in an Excel spreadsheet for say a CD collection.
4. Spatial database - A database system that offers spatial data types (data that defines a geometric space) in its data model and query language and supports spatial data types in its implementation, providing at least spatial indexing and spatial join methods. Spatial database systems offer the underlying database technology for geographic information systems.
5. Cloud database - A database management system that is hosted by a third-party service provider on a remote server and accessed over the Internet. It runs on a cloud provider’s platform and data can only be stored or accessed when there is an Internet connection.Examples include Amazon EC2, Rackspace, and Microsoft Azure.

# Popular Database Products

1. [Oracle](https://en.wikipedia.org/wiki/Oracle_Corporation) – closed source, solid product, can run on all major platforms, provide full support, online backup and recovery. Basically the first relational database server, too pricey to convert now. Used by MetLife Insurance, Osh Kosh, Pandora, GM, Mazda, Kellogg’s, many Fortune 500 companies, hospitals, government probably because of full and complete customer support.
2. [MySQL](https://en.wikipedia.org/wiki/MySQL) – open source, scalable, can be customized, all major operating systems: Linux, UNIX, Windows, Mac. Speed. MySQL Enterprise would offer full software support. Used by US Navy and Symantech because reliable, robust and low-cost.
3. [Microsoft SQL Server](https://en.wikipedia.org/wiki/Microsoft_SQL_Server) – closed source. Seamlessly compatible with Microsoft Office. Feature rich which may cost in speed. Resistant to data corruption. Only runs on Windows. Lower license cost. Used by Eagle Creek Software Services, probably because of ease of use.
4. [PostgreSQL](https://en.wikipedia.org/wiki/PostgreSQL) – open source, prides itself in standards compliance, all major operating systems: Linux, UNIX, Windows and Mac, customizable. Sponsored by Red Hat Inc., HP, Skype, Google probably because of the open source availability and adaptability.
5. DB2 – (IBM product) Lower licensing cost than Oracle, faster deployment, 5x better compression rates than Oracle, simpler administration (less complexity). Used by Yazaki Europe Ltd (Automotive Parts Supplier) selected for speed of delivery and report availability.
6. Vertica – (by Hewlett Packard) Column storage, compression and clustering. Sorts data for faster queries. Ideal for read-intensive workloads. Used by Facebook, Wayfair.com, Twitter for faster queries.
7. SAP Sybase ASE - To streamline processes, reliable (20 yrs). Used by Adidas Group, AAA for predictive analytics.
8. Teradata – Easy to use business analytics. Used by Ford Motor Company, The Coca Cola Company.

# The class focuses on learning SQL using Oracle. All of the databases above use SQL. The instructor postulates what you learn in this class is 95-97% transferrable to any of the products listed that in a typical job setting is. Is that true? Or, is the instructor full of beans and gravy.

# Founded in 1918, ANSI (American National Standards Institute) is a private, non-profit organization that administers and coordinates the U.S. voluntary standards and conformity assessment system. Periodically ANSI updates and republishes the “Database Languages – SQL” standards. Each manufacturer claims ANIS compliance with these standards.

# Below is a table showing syntax differences between Oracle SQL and Microsoft SQL Server SQL. Although you may intuit meaning from their names, we have not covered any of these and you are not expected to understand them. Three of the columns are blank. Fill in these three columns.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Description**  **SQL Functions: Description and Syntax** | **Oracle** | **MS SQL Server** | **PostgreSQL** | **MySQL** | **Vertica** |
| Find smallest integer >= n | CEIL | CEILING | CEIL | CEIL | CEIL |
| Modulus | MOD | % | % | % | % |
| Truncate number | TRUNC | <none> | TRUNC | TRUNCATE | TRUNC |
| Translate NULL to n | NVL | ISNULL | ISNULL | ISNULL | ISNULL |
| Return NULL if two values are equal | DECODE | NULLIF | NULLIF | NULLIF | NULLIF |
| String concatenation | CONCAT(str1,str2) | str1 + str2 | CONCAT | CONCAT | CONCAT |
| Capitalize first letters of words | INITCAP | <none> | INITCAP |  | INITCAP |
| Find string in string | INSTR | CHARINDEX | POSITION | INSTR | INSTR |
| Find pattern in string | INSTR | PATINDEX |  | INSTR | INSTR, PATTERN\_ID |
| String length | LENGTH | DATALENGTH | LENGTH | LENGTH | LENGTH |
| Pad string with blanks | RPAD, RPAD | <none> | LPAD, RPAD | LPAD, RPAD | LPAD, RPAD |
| Trim leading or trailing chars other than blanks | LTRIM, RTRIM, TRIM | <none> | LTRIM, RTRIM, TRIM | LTRIM, RTRIM, TRIM | LTRIM, RTRIM, TRIM |
| Replace chars in string | REPLACE | STUFF | REPLACE | REPLACE | REPLACE |
| Convert number to string | TO\_CHAR | STR, CAST | TO\_CHAR | CHAR | CHR |
| Convert string to number | TO\_NUMBER | CAST | TO\_NUMBER | CAST | TO\_NUMBER |
| Get substring from string | SUBSTR | SUBSTRING | SUBSTR | SUBSTR | SUBSTR |
| Date addition | ADD\_MONTH or + | DATEADD | + | ADDDATE |  |
| Date subtraction | MONTHS\_BETWEEN or - | DATEDIFF | AGE or - | DATEDIFF, DATE\_SUB | DATEDIFF |
| Last day of month | LAST\_DAY | <none> | LASTDAY | LAST\_DAY | LAST\_DAY |
| Time zone conversion | NEW\_TIME | <none> | AT TIME ZONE | CONVERT\_TZ | TIMESTAMPTZ |
| Next specified weekday after date | NEXT\_DAY | <none> |  |  | NEXT\_DAY |
| Convert date to string | TO\_CHAR | DATENAME, CONVERT | TO\_CHAR | DATE\_FORMAT | TO\_CHAR |
| Convert string to date | TO\_DATE | CAST | TO\_DATE | STR\_TO\_DATE | TO\_DATE |
| Convert date to number | TO\_NUMBER(TO\_CHAR(d)) | DATEPART | DATE\_PART |  | DATE\_PART |
| Date round | ROUND | CONVERT |  |  | ROUND |
| Date truncate | TRUNC | CONVERT | DATE\_TRUNC |  | DATE\_TRUNC, TRUNC |
| Current date | SYSDATE | GETDATE | CURRENT\_DATE | CUR\_DATE | GETDATE |
| If statement in an expression | CASE, DECODE, COALESCE | CASE, COALESCE | IF, CASE | IF, CASE | CASE |
| Current user | USER | USER | CURRENT\_USER | USER | CURRENT\_USER |