**Expressions for BI's & Data Analytics| BASIC**

Data base is a toll to manage large amount of data on computer in the hard drive. It gives a quick access to the data and allows to query it. Now days, the 'principle of tables' is the most common principle for data maintenance and managing in tables.

* All data stored in tables
* Each table storing data about essence (customers, suppliers and more)
* Each table composed of columns, rows and fields
* Every field/column as his own data type

**A. Tbl, ERD, DB Normalization process:**

This is a process where all tables set into logical order, no duplicates, easy maintenance, completeness and simplified. During the process we will make sure that:

* Each field contain single value

**NF1**

* Each field contain unique value
* Same data type for each column
* NO duplicates
* NO aggregate fields

**NF2**

* There is functional hierarchy to the primary key for each row,

If not, we will separate

* All columns can be determined only by the primary key column,

Ask yourself "does all columns of table depend on PK?"

**NF3**

All NF before the third are existing and valid

**Methodological Data normalizations:**

1. Check for duplicates (SELECT COUNT(\*) source and destination)
   1. Rowes
   2. Columns
   3. Tables
2. Anonymous Tables/Columns
   1. Give a name
3. Columns order scientifically | Id’s > Attributes > More details
4. Tables with at least 1 column only
5. NULLS check and replacement

**How to check connection cardinality:**

SELECT COUNT(\*) FROM X GB ID HAVING Y > 1

* + Choose leading Tbl > On both Tbls and check who has more results > then It is the many to 1

**B. SQL | Structured Query Language:**

It is a code language using to deal with Data bases Management and analyte.

Through SQL we can create, update, insert, modify, manipulating the data base and the tables in it. Min Syntax, Easy, and intuitive.

In SQL we have different command sets –

**DML** – Data manipulation language, which helps to query and modify objects in data base

**DDL** – Data define language, set of commands to modify, create and managing data bases

**DCL** – Data control language, for Dba use usually, controlling permissions and users

**T**-**SQL** – transact Structured Query Language, extending the set of commands and gives more tolls to deal with data manipulation and query.

**SSMS** – Sql server management studio. The interface for Data management through SQL.

**C. I would like to explain about some of the frequent expressions we use in SQL:**

**Primary key (PK)** – Is unique identifier for tables. Uses for relative row in specific table, prevents duplicates and helps to process data by getting to it faster.

**Foreign key** **(FK)** – Is a primary key imported to other table for establishing a connection between 2 or more tables, keeps data normalized and integrity (ex: preventing duplicates values)

**Surrogate key –** Alternative key for any tbl chooses from multiple of options including: PK, Numerate ext.

**Table** – Set of fields, columns, rows and constraints that together as rational order and this set composing a table.

**SELECT command** – For drawing out certain data from specific table

**FROM –** Defines which table or tables are going to be used in query

**WHERE clause** – Uses for filtering results by some condition

**GROUP BY** **clause** – aggregate data

**HAVING** **clause** – filter layer which relevant for aggregate columns

**D. I would like to present common data types:**

**INT –** Full number

**FLOAT** – Full number with 2 places after "."

**DECIMAL** – Full number with defining option for places after and before "."

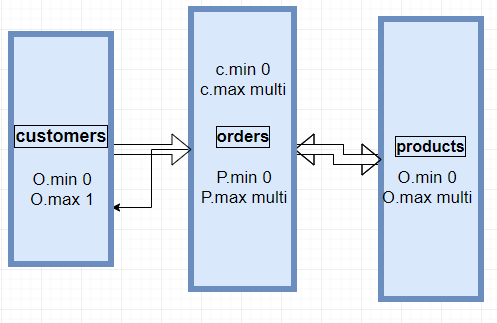
**BIT** – 0/1 | TRUE/FALSE

**DATE** – date format XX-XX-XXXX

**DATETIME** – date and time format XX-XX-XXXX XX: XX:XX

**STRINGS** **TYPES** (ex: varchar)

We can create, alter, and cast data types using DDL, DML, TSQL languages.

**More In General:**

**ERD** – Entity Relationship Data Diagram Model.

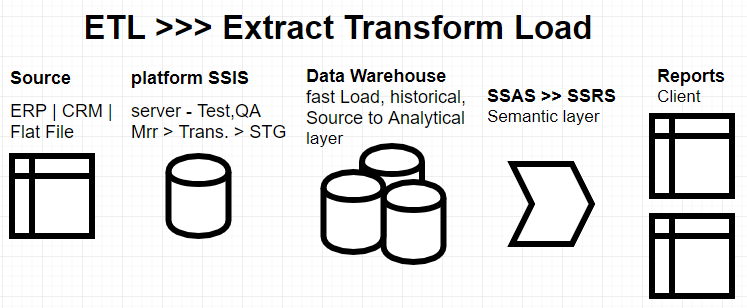
Objects (tbl) and their Relationship type. There are several types of connections:

1. 1-1 Single to single (ex: Student to beg)

2. 1-∞ One to Many (ex: One class to many student)

3. ∞-∞ Many to Many (ex: many Student to Many lectures)

**ETL –** Any transportation of data from source to target, Extract > Transform > Load

We can use SSIS platform for Tis process (Sql Server Integration Studio)

**DateWareHouse DWH DW –** A place for all data, historical and updated.

Dim – Attribute of facts**. Central Data repository** one Organization True: Completeness, integrity, consistency

Fact **- Facts** of Transactions Described by attribute column and identity column

**Mirror (Mrr) –** Temp tbl exactly as the source uses for manipulations and transformations before importing to stg tbl.

**Staging –** The final stage before loading the Data. This is the last check of valid info importing onto the destination.

**Production Server –** Live business info, CRM, ERP

**API –** Application Programming interface, which helps us to connect between different systems.

**Transforms –**

Cleansing Irrelevant columns and Rows we don’t need in des.

Cleansing 'Null's'

Translation and Splits – Data types and Constraints

Joins Tbl and Data – Bringing data and keeping the integrity

Aggregations – calculations

**ETL Project management and accomplishment process –**

**Stg. 1 Mapping business**

business environment investigating -

1. gather info about activity and workflow and methods practice

2. gather any detail of info for requirements and need

3. be creative Search for match points in business Logic and Max productivity

4. Search for Problems and weakness in Data flow

**Stg. 2 Mapping Operational Databases**

business Data source and Destination understanding -

1. where it stored and generated

2. how it generated

3. who rolls it

4. maintenance methods

5. which formats are coming

6. where should it be load

7. transformation requirements

8. constraints and needs

**Stg. 3 Planning & Qualifications**

Project plan – flexibility and easy to maintain and update as possible

1. thinking about all options possible for failure

2. taking in consideration easy maintenance and changing model

3. building a process map, workFlow, stages and responsibilities

4. Rolls and Technical characterization for each Atom part in the process

5. QA for the program Third party

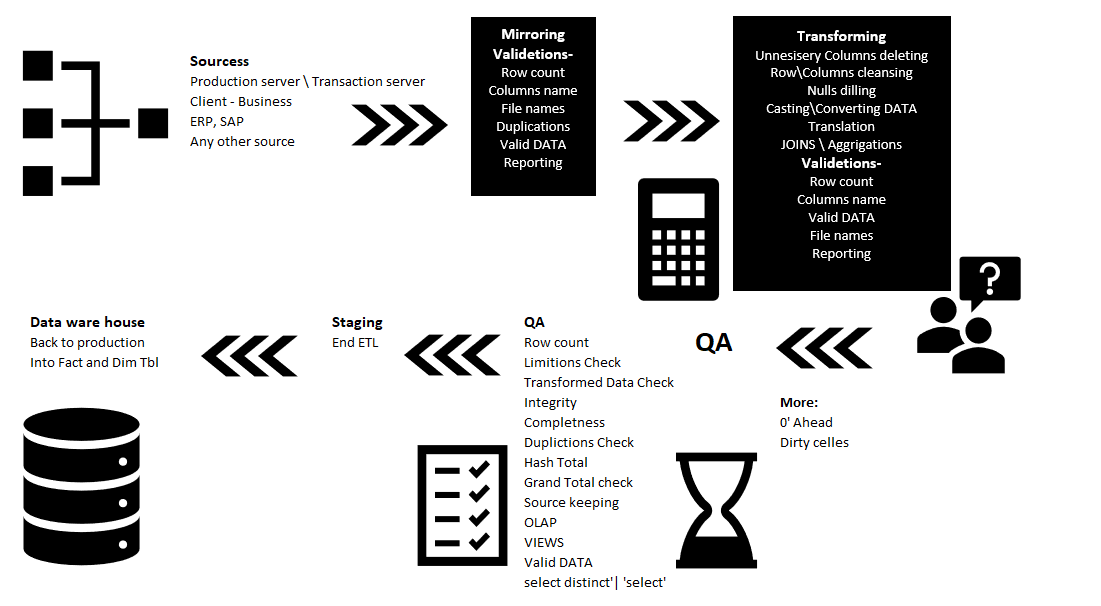
6. Ignition Document, Start-End Dates, targets-Goals, KPI’s for Successes

**Stg. 4 Developing SSIS/Informatica Solutions**

Developing SSIS solutions according to requirements and planning Stage

Transforms To be Done: Nulls, ASCII, Cleansing Column and Records

Planting Data > CDC Activating > SCD Defining > Automation and reports Sys. > QA



**Stg. 5 SSAS semantic layer**

**Tabular Projects Construction –**

Tables, ERD's, KPI's, measures, OLAP (bringing data to multi-dimensional Analytical queries fast)

**Stg. 6 Front end Tolls**

**Developing** **end2end solution -**

Power BI

QLikSense

Tableau

Excel

Models

**Statistic Terms For Data Analytic -**

**Statistics** **Theory** – Gather, Organized, Processing Data Into visualizations

**Raw** **Score** – Non processed Data gathered (ציונים גולמיים)

**משתנה** **איכותי** – משתנה המתאר ערכים בצורה איכותית כגון מין, צבע, מוצא. מתואר לא מספרית אלא מילולית

**משתנה** **כמותי** – משתנה שערכיו מתארים מספרית

**משתנה** **בדיד** – בעל ערך סופי

**משתנה** **רציף** – בעל אינסוף ערכים אשר מתארים אותו

**משתנה מקרי –** פונקציה המתאימה לכל אירוע אפשרי במרחב הסתברות לערך מספרי

**שכיחות** – מספר הפעמים אשר משתנה מופיע

**היסטוגרמה** – הצגה חזותית של שכיחויות

**Average** – Sum off all Values split to 'n' events

**ממוצע** **משוקלל** – ממוצע המבוסס על סכום האירועים כפול שכיחות האירועים לחלק לכמות האירועים

**Standard deviation** - how values could be spread

**Sctergram** – X, Y scattering

**Correlations** – how strong the connections between X, Y

**Linear Regression** – Statistic method for check linear connections between parameters Most common In Science and data Projects

**KPI'****s -** kpi used for measuring performance in achieving key Goals. Some Ex. for Kpi's in two different Departments in Business:

**How to Define the right KPI's TOP 5:**

1. company mission definition

2. what are the values?

3. define business Goals

4. Set Kpi's for each business core and department

5. Review And adjust goals and KPI's

**MORE SQL** -

**Intersect** – Bring me what’s is common only

**Except** – Bring me what’s exist unlikely in first Table only from both Tables

**Union** – Bring Distinct values only and Order

**Union** **all** – Bring all also duplicates and do not order

**MORE ETL-**

**CDC –** Changing data Capture, Incremental load into mirr Stage, only for data changes, saves time, efficiency, resources. Operating on DB and Tbls Level. Capture DML commands

**SCD –** Slowly changing dimension, Takes Data changes into DIM STG Tbls before updating DWH DIM, by certain Type (0,1,2,3,4,6 [NO 5]) Google for each Typ. According to Business needs.

**SQL AGENT / JOBS – Google it**