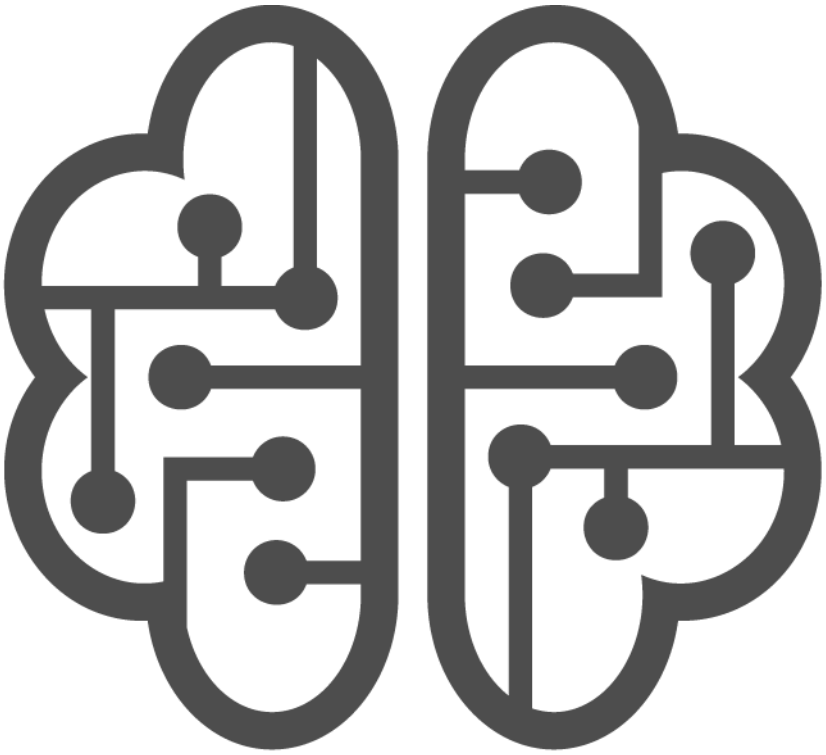

Conception methodology

December 2018

Contents

Introduction:

This deliverable is dedicated to explain all the norms adopted to realize the conception activity. In our project we have several phases to do the conception, each phase is consecrate to a part of the final deliverable product.

First things first we will conceive the development of our apps using the agile software development, we have chosen Scrum method. Obviously to follow this process we have a set of diagrams and models to achieve, that's why we need to use UML (Unified Modeling Language). After we will pass to choose the multidimensional schema in order to model our Data warehouse. The schema that we will use in our project is snowflake schema. And finally we will use the KDD process (Knowledge Discovery in Databases) in order to implement the Data mining.

Scrum Method:

Scrum is an agile, iterative and incremental method for managing software development. A key principle of Scrum is the dual recognition that customers will change their minds about what they want or need (often called *requirements volatility*) and that there will be unpredictable challenges, that's why we found in this approach the basic unit called "Sprint" which represents the iterations of the development, restricted to a specific duration.

Here is an image describing the methodology, for more detail ready this article[1]

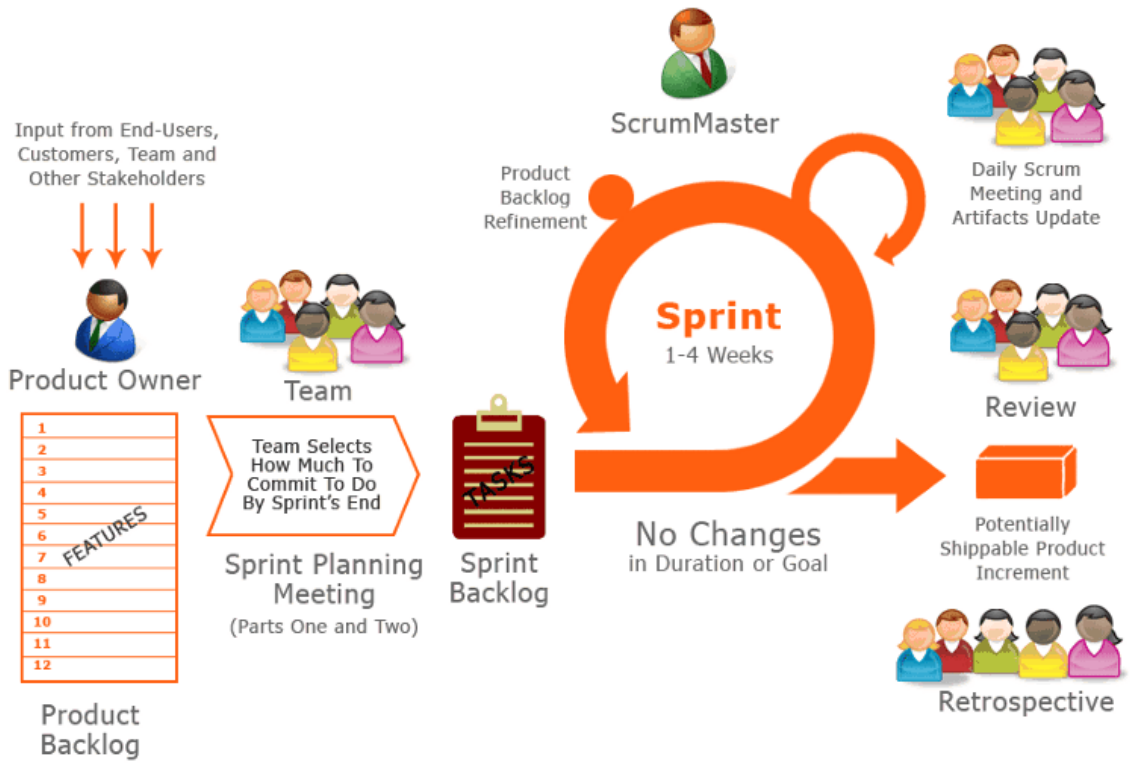

Figure 1 Scrum Approach

[1] https://medium.com/chingu/a-short-introduction-to-the-scrum-methodology-7a23431b9f17

UML:

As any software development process we need to do the modeling and the conception of our apps, therefore we will use UML to produce the different diagrams.

Here is some of the diagram we will be using:

- Use case diagram (to show the actors and the interaction).
- Activity diagram (describe the steps performed in a UML use case).
- Class diagram (to show the structure).
- Sequence diagram (describes how and in what order a group of objects works together).

Here is an image containing the different diagrams in UML, for more detail read this article [2]

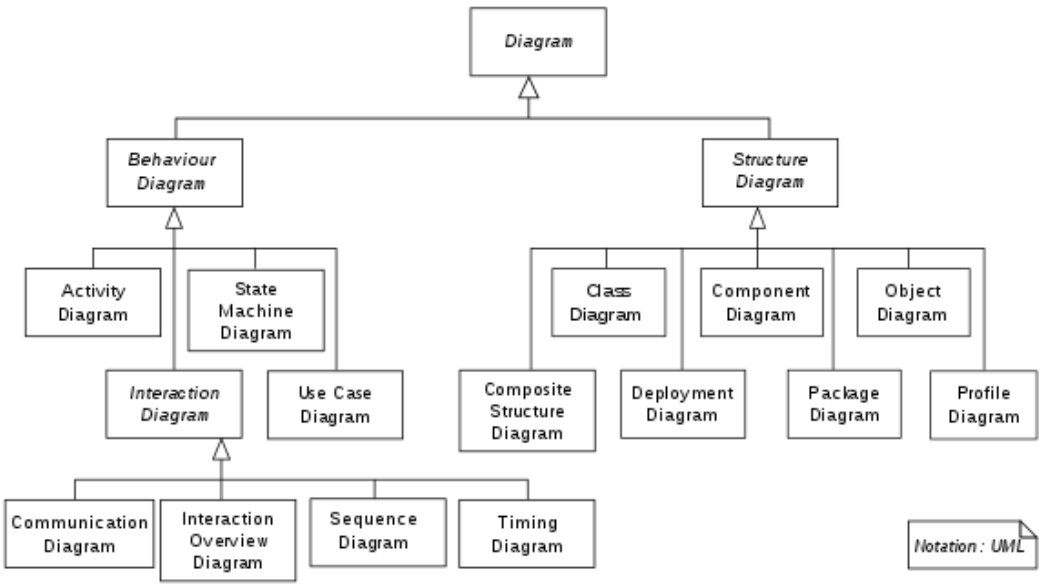

Figure 2 UML Diagrams

[2] https://medium.com/federicohaag/uml-unified-modeling-language-5a2a0c2fb973

Snowflake schema:

Snowflake schema is a logical arrangement of tables in a multidimensional database, it is an extension of a Star Schema, the main difference is that the snowflake dimensions are normalized by removing low cardinality attributes and forming separate tables which reduce redundancy and benefit to us a smaller disk space use. This schema is represented by centralizing one fact tables in the middle, connected to multiple dimensions and also the schema is designed to address the unique needs of very large databases designed for the analytical purpose (OLAP).

The following image describe general structure of snowflake schema. And like always for more detail visit this website [3]

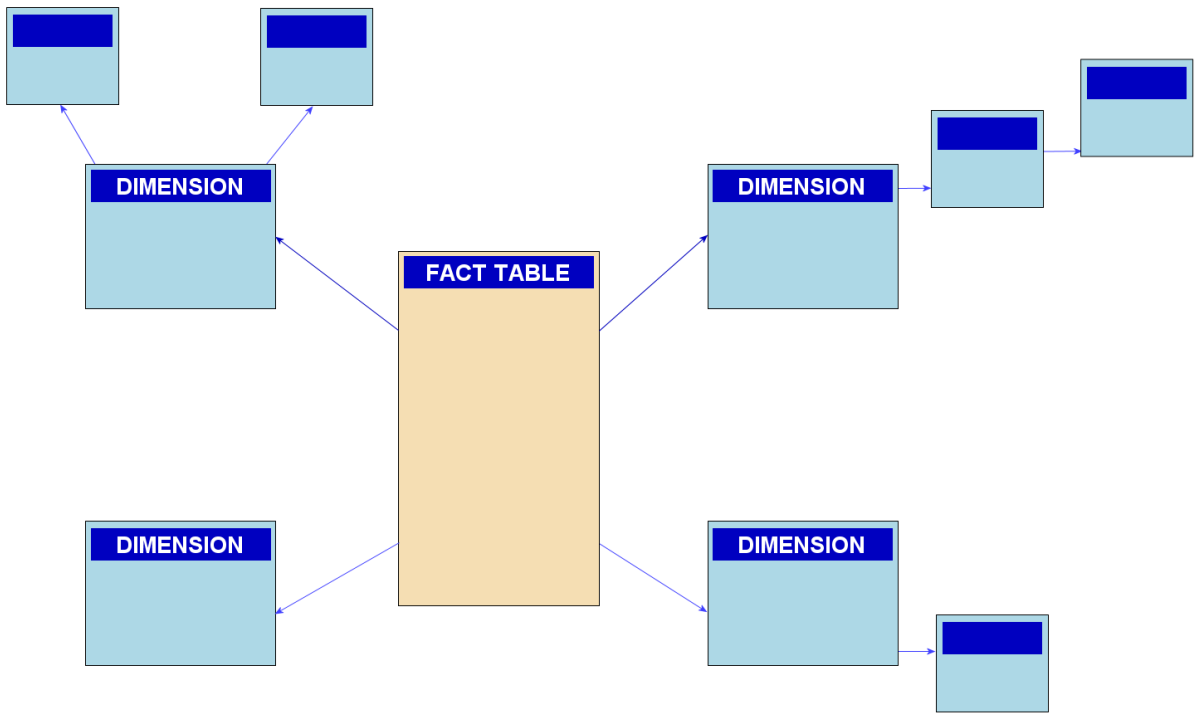

Figure 3 Generic Snowflake schema

[3] https://www.guru99.com/star-snowflake-data-warehousing.html

KDD (Knowledge Discovery in Databases):

KDD, refers to the broad process of finding knowledge in data, and emphasizes the "high-level" application of particular data mining methods. The unifying goal of the KDD process is to extract knowledge from data in the context of large databases. To achieve our objective using this process, we have to go through a sequence of steps.

Here is the steps[4]:

- Identify the goal of the KDD process from the customer's perspective.
- Understand application domains involved and the knowledge that's required
- Select a target data set or subset of data samples on which discovery is be performed.
- Cleanse and preprocess data by deciding strategies to handle missing fields and alter the data as per the requirements.
- Simplify the data sets by removing unwanted variables. Then, analyze useful features that can be used to represent the data, depending on the goal or task.
- Match KDD goals with data mining methods to suggest hidden patterns.
- Choose data mining algorithms to discover hidden patterns. This process includes deciding which models and parameters might be appropriate for the overall KDD process.

[4] https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd

- Search for patterns of interest in a particular representational form, which include classification rules or trees, regression and clustering.
- Interpret essential knowledge from the mined patterns.
- Use the knowledge and incorporate it into another system for further action.
- Document it and make reports for interested parties.

Here is another image describes the steps, and for more details visit the linked bellow website [5]

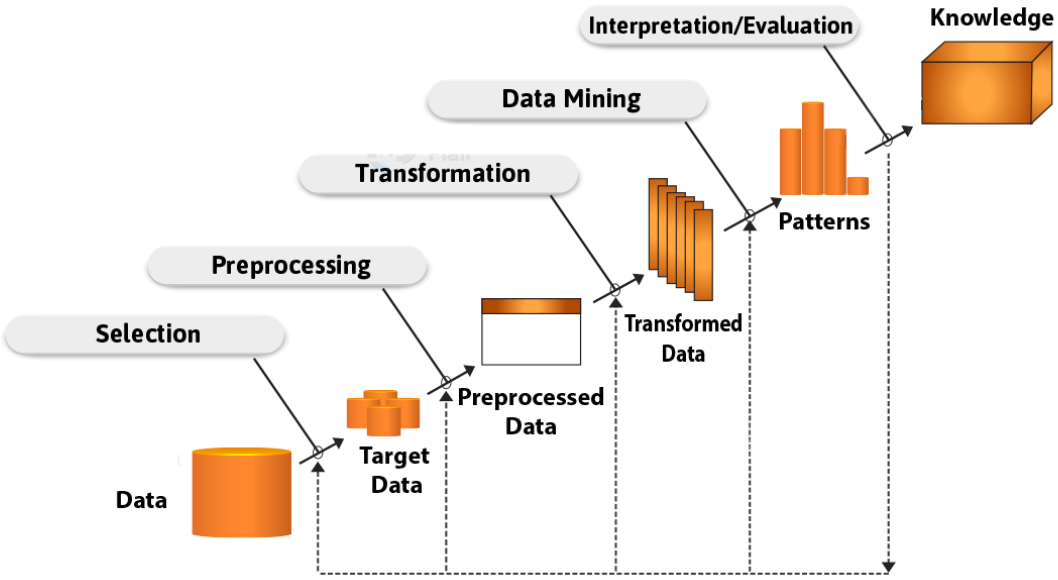

Figure 4 KDD Process

Conclusion:

At the end we want to mention that this techniques and methodologies norm are widely used in all BI projects around the world for achieving the best product quality.

Thanks for choosing Data brains and for changing you vision to Data.

[5] https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd