

## OBJETIVO/RESTRIÇÕES:

1. A nossa disciplina é focada em preparação, orquestração e fluxo de dados, não fazendo escopo das aulas de laboratórios: *criação e visualização dos dados em dashboards, técnicas que exijam conhecimentos estatísticos avançados e Machine Learning.*
2. O Objetivo desse laboratório (estudo de caso) é que você possa criar conjuntos de dados, conforme requisitos de negócios apresentados, assim como, ampliar o conhecimento em novas ferramentas/linguagens.

## PROJETO: TRATA CHURN\*

A “MAIS SAÚDE” é uma empresa brasileira que atua na área de saúde, sendo uma das maiores operadoras de planos de saúde do país **cuja área de atuação é na região nordeste**.

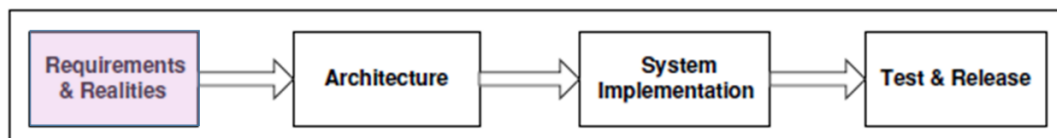
A “MAIS SAÚDE” deseja acompanhar e compreender a taxa de churn afim de manter uma base de clientes saudável e sustentável a longo prazo.

O escopo de medição da taxa de churn englobará somente os clientes que contrataram planos na **modalidade coparticipativa**.

O projeto foi **contratado em etapas** e foi demandado por diferentes áreas da empresa (marketing, vendas, auditoria etc.). Você fará parte da equipe de ETL e deverá entregar conjuntos de dados para atender de forma peculiar cada demanda.

Sabe-se que a origem dos dados será o sistema **SIS\_CRM**, onde todos os clientes são identificados como ativos ou inativos, porém para nosso escopo somente clientes inativos são alvo de análise neste projeto. Clientes identificados diferentes de inativos nos dados enviados deverão ser enviados para área de auditoria, assim como duplicidades de clientes na base deverão ser também analisados a priori.

Os dados do SIS\_CRM serão entregues todo dia primeiro de cada mês, referenciando o mês anterior. EX: Churns entregues no dia 01/07 – referem-se à posição da base do mês de junho (06)



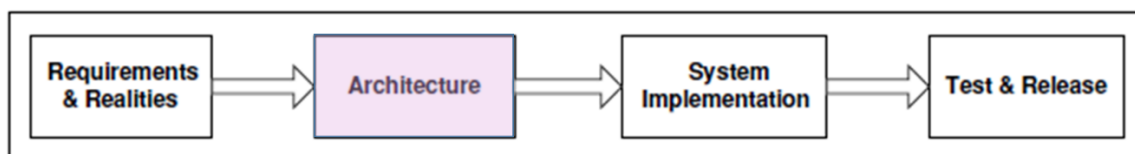
**Figure Intro-1** The Planning and Design Thread.

### Definições:

(\*) "Churn" é um termo amplamente utilizado no mundo dos negócios e se refere à taxa de perda de clientes ou usuários ao longo de um determinado período de tempo.

## ARQUITETURA

- Fontes: flat files (txt,csv), tabelas (SGBD: SQLSERVER), JSON
- Saídas: Tabelas, flat files (txt,csv) e planilhas Excel
- Agendamento: todo dia primeiro de cada mês e período da noite, a partir da 02:00h.
- Em caso de impedimento de execução no dia 01, reprocessar o quanto antes dentro do mês corrente, uma vez que os dados lidos se referem ao mês anterior.
- Ferramentas/Linguagens: PDI/SQL/Python



**Figure Intro-1** The Planning and Design Thread.