

# H:DATA

## PROCESSO SELETIVO Maio/2024

**VAGA:** Analista de Engenharia de Dados (Júnior)

**ETAPA 2:** Resolução de Cases

### INTRODUÇÃO

#### DESAFIOS DA ÁREA:

O dia a dia do Analista de Dados (ênfase em Engenharia de Dados) é repleto de fluxos de ELT's e ETL's: extrações de dados de fontes distintas, manipulações e cargas em outras bases de dados, permitindo a correta e completa ingestão dos dados nos ambientes RAW, REFINED e TRUSTED.

De modo geral, após definição de fontes/origens dos dados, os objetivos das atividades do time de Engenharia de Dados podem ser:

1. Coleta, análise (quantitativa e qualitativa) e limpeza de dados;
2. Enriquecimento da base de dados recebida a partir da aplicação de regras de negócio estabelecidas (regras de negócio ou normalizações);
3. Aderência dos dados recebidos (após enriquecimento) a arquitetura de dados já definidas como padrão para uso interno (exemplo: ingestão de dados para a arquitetura de DW da H:Data).

Para facilitar a interface entre os times das áreas de CS ou Negócio com o time de Analistas de Dados, na etapa de levantamento de requisitos e refinamento, uma das atividades-chaves principalmente em demandas de Aderência é o Mapeamento de Aderência. É nesta etapa que, de posse da estrutura de dados recebida e da estrutura de dados de uso da H:Data (layouts de dados), são informadas as regras a serem aplicadas e como “encaixar” os dados recebidos na estrutura de destino. O papel Analista de Dados é analisar os dados recebidos e implementar este “encaixe”, documentando todo o processo e o otimizando, facilitando futuras cargas incrementais e a reutilização dos métodos utilizados na ETL desenvolvida.

## CASE 1 – Enriquecimento de Dados Financeiros

### SITUAÇÃO:

Sob nova direção desde agosto/2023, o Hospital Estrela (HE) procurou a H:Data para apoiar em seu processo de estruturação financeira. O time de Negócios da H:Data já possui aplicações de BI e painéis integrados ao DW da H:Data, porém precisa da ajuda do time de Analistas de Dados para que estes dados sejam de fácil consumo.

Para facilitar o fluxo de trabalho, o time de Negócios já confeccionou o Mapa de Aderência referente ao dataset recebido pelo HE.

### CASE:

A partir do material do Case 1:

1. Colete as informações da base de dados recebida;
2. A partir do Layout de Dados e das tabelas-bases da H:Data (cadastros-bases, a partir dos quais serão realizados de/para) e do Mapa de Aderência, enriqueça os dados coletados;
3. Exporte os dados enriquecidos em uma base única (tabelão);
4. Exporte os dados enriquecidos na estrutura do DW (dica: utilize as tabelas e campos citados no Mapa de Aderência para modelar as tabelas de destino).

### BÔNUS:

Desenvolva um método ou uma função que interprete o arquivo de Layout de Dados de Fluxo de Caixa e, a partir da estrutura de Mapa de Aderência apresentada, se adeque a outros mapas desenvolvidos pelo time de Negócios. Desta forma, o processo de ingestão de dados de outras bases para o mesmo fim (Fluxo de Caixa) será facilitado, necessitando que o Analista de Dados entre apenas com customizações simples como caminho do arquivo de leitura e de destino.

## CASE 2 – Aderência dados de Pronto Atendimento

### SITUAÇÃO:

Em acréscimo à análise da situação financeira do Hospital Estrela (Case 1), foi demandado ao Time H:Data a análise também dos dados da operação de Pronto Atendimento.

Para este fim, a equipe de TI do HE exportou uma base contendo os atendimentos da unidade: atendimentos, algumas informações categóricas e tempos de algumas etapas do atendimento.

De informações adicionais, o time de TI apenas informou que:

- IE\_TIPO\_ATENDIMENTO = 3 corresponde a atendimentos de Pronto Atendimento;
- IE\_CLINICA corresponde à especialidade do atendimento, sendo:
  - 1: Clínica Médica;
  - 4: Pediatria;
  - 10: Cirurgia Geral.

Como neste caso a base foi enviada diretamente para os Analistas de Dados (não passaram pelo time de Negócios), é esperado que os analistas analisem a base e apresentem os principais pontos identificados ao outro time.

### CASE:

A partir do material do Case 2:

1. Faça uma interpretação dos dados recebidos, bem como da qualidade dos campos identificados como principais;
2. Relacione os dados recebidos, de modo a qualificar a informação da base principal a partir de informações complementares;
3. Um dos objetivos do time de Negócios é compreender a Jornada do Paciente, portanto:

# H:DATA

- a. Identifique quais etapas podem ser monitoradas por meio da base de dados enviada;
  - b. Identifique a ordem prioritária do fluxo de atendimento na unidade, bem como possíveis desvios de fluxo do atendimento e possíveis motivos;
  - c. Tangibilize a Jornada do Paciente, explicitando tempos de esperas entre as diversas etapas envolvidas no fluxo de atendimento (dica: observar possíveis dados nulos, inversões de etapas e durações negativas).
4. Exporte os dados enriquecidos em uma base única (tabelão).

## **BÔNUS:**

O HE já sinalizou que gravou em seu banco as views correspondentes aos datasets exportados e formalizou sua intenção em enviar periodicamente estes dados atualizados. Portanto, é fundamental que seja preparada uma estrutura que não apenas seja capaz de fazer a ingestão dos dados recebidos, mas que também possa ser utilizada periodicamente. Diante deste cenário:

- Para melhor gerenciamento dos dados e otimização da nossa estrutura de banco, sugira uma arquitetura de banco que faça uso de modelos otimizados para DW (dica: Star Schema);
- Prepare seu algoritmo de modo que ele seja capaz de ingerir cargas complementares, garantindo a não duplicidade de itens já de nossa posse e update de dados referentes a registros anteriormente inseridos.