**CIÊNCIA DE DADOS - BLUE EDTECH**

**Módulo 5 - Projeto 1**

**Modelagem e Estruturação de Dados**

Nomes: Henrique Grandi Baldo e Antonio Duarte Marcos Junior

1. **Introdução**

Este trabalho apresenta um projeto de banco de dados para o setor de seleção e recrutamento de uma empresa fictícia chamada BlueRH. O banco de dados deve registrar as informações do processo seletivo e servir para que a equipe de recrutamento estime o LEAD TIME médio das vagas disponibilizadas

Para a construção de um banco de dados existem duas estruturas principais o: SQL e NoSQL. Os bancos SQL são estruturas relacionais enquanto os NoSQL utilizam estruturas não relacionais. Um banco de dados relacional tem um formato estruturado rígido baseado em tabelas que se relacionam de acordo com as regras de negócio da empresa (4LINUX). Alguns dos principais bancos relacionais existentes hoje são: MySQL, PostgreSQL, Oracle Database e SQL Server da Microsoft. Todos estes bancos citados são baseados em SQL (*Structured Query Language*) o que facilita a migração entre um banco e outro, já que a maioria dos recursos e comandos são semelhantes.

Nos bancos não relacionados os dados são documentados e permitem o registro de informações diferentes de tabelas como por exemplo: fotos, vídeos, áudios (4LINUX). Exemplos de bancos de dados não relacionais são: MongoDB, ElasticSearch e Redis. Em um banco NoSQL é possível adicionar dados sem ter que definir um schema de banco de dados, diferentemente do SQL.

Os bancos de dados podem ser ainda estruturados como OLAP e OLTP. Os OLTP (*Online Transaction Processing* que traduzindo fica Processamento de Transações Online) são os sistemas operacionais com dados transacionais. Fornecem assistência para as atividades na visão operacional, os OLTP processam informações dos diversos sistemas com informações variadas da empresa (BORGES, 2020).

Os OLAP (*Online Analytical Processing* que traduzindo fica Processamento Analítico Online) são executados dentro do Data Warehouse e realizam análises em grandes volumes de dados. Contribuem com as atividades de BI para visualização das informações e dão suporte às funções de análises do negócio (BORGES, 2020).

O processo de modelagem de dados passa ainda por três etapas: modelo conceitual, modelo lógico e modelo físico (LEHMKUHI e EGER). A primeira etapa é a elaboração do modelo conceitual, nela são descritas, de forma abstrata, os fatos da empresa de forma natural, suas propriedades e relacionamentos. O modelo conceitual é a forma de entender as relações entre entidades, validar conceitos e mapear o ambiente. Após a construção do modelo conceitual é criado o modelo lógico. O modelo lógico visa representar as estruturas que irão armazenar os dados dentro de um Banco de Dados, nele são definidas as entidades e seus atributos. Por fim, é criado o modelo físico. O modelo físico é a concretização do modelo lógico. Nele é definido qual o SGBD será utilizado, tipo e tamanho dos campos, relacionamentos, restrições, etc. Nele são descritas as estruturas físicas de armazenamento: tabelas, índices, gatilhos, funções, visões, etc.

**1.1. Objetivos**

**1.1.1 Objetivo geral**

Levando em consideração o contexto apresentado da introdução deste trabalho o objetivo geral será: implementar um Banco de Dados para o setor de seleção e recrutamento da BlueRH. Considerando a estrutura da empresa (apresentada na seção análise de negócio) será utilizado um modelo relacional utilizando o MySQL como SGBD e será estruturado para ser utilizado como um sistema OLAP.

**1.1.2 Objetivos específicos:**

* Construir os modelos: conceitual, lógico e físico do Banco de Dados;
* Definir entidades e relacionamentos;
* Estabelecer os atributos de cada entidade;
* Estabelecer as relações entre entidades;
* Construir o schema do Banco de Dados;
* Construir as tabela para armazenamento dos dados de cada entidade;
* Implementar o script para construção do Banco de Dados.

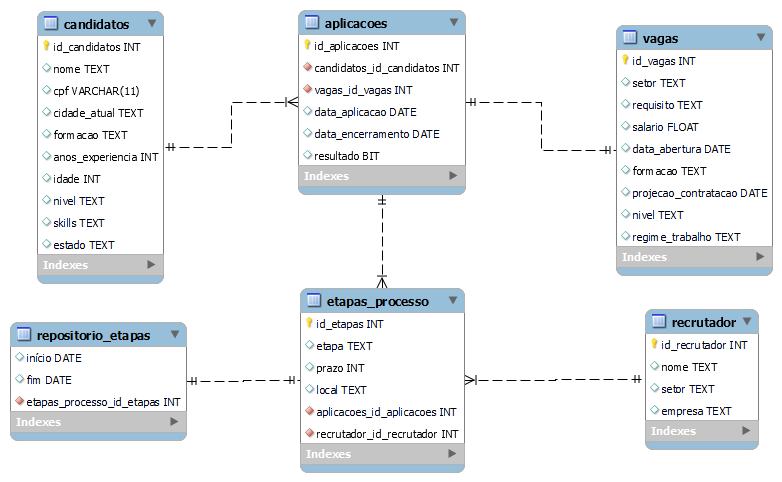
1. **Fase de análise de negócios**

Uma grande empresa entrou em contato com o grupo BLUERH para assumir o processo de seleção e recrutamento. A empresa por ter muitos processos simultâneos, de diversos setores e vagas, estava passando por dificuldades em analisar o LEAD TIME do processo seletivo. A gestão de recrutadores não estava conseguindo atender as demandas de contratação, fazendo com que outros setores ficassem sem funcionários, acarretando atrasos nas entregas.

Assim o grupo BLUERH formou uma comissão especial formada pelos analistas de RH, Henrique Grandi Baldo e Antonio Duarte Marcos Junior, para atender as demandas da empresa e conseguir organizar o fluxo de contratação. Desta maneira a empresa não iria mais ter problemas com prazos, nem na hora de contratar, nem na hora das entregas para seus clientes.

1. **Discussão do tema**

Para discutir sobre o tema, os dois se reuniram para conversar, entender o fluxo passado pela empresa e montar um diagrama explicativo. Durante este encontro foram discutidas as possíveis soluções para o problema em questão, chegando no resultado abaixo:

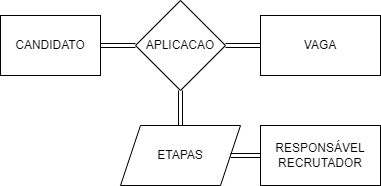


**Figura 1** - Estrutura do Banco de Dados criado.

O fluxo foi dividido em 6 tabelas para que ficassem mais claras as divisões de tarefas de cada um e que fosse mais fácil de acompanhar as datas durante os processos seletivos. Abaixo iremos descrever o processo para chegar ao resultado final.

**3.1. Modelo conceitual**

Como dito na introdução do trabalho, o modelo conceitual consiste em identificar as principais relações do problema. Na Figura 2 é apresentado o modelo conceitual criado. Neste modelo o CANDIDATO realiza a APLICAÇÃO para uma VAGA (o candidato pode realizar aplicação mais de uma VAGA). Após a aplicação são iniciadas as ETAPAS de seleção. Cada uma das ETAPAS de seleção têm um agente RESPONSÁVEL o RECRUTADOR. Em todas as etapas do processo é necessário realizar o registro de datas de início e fim para que seja possível realizar o cálculo do LEAD TIME no final do processo.

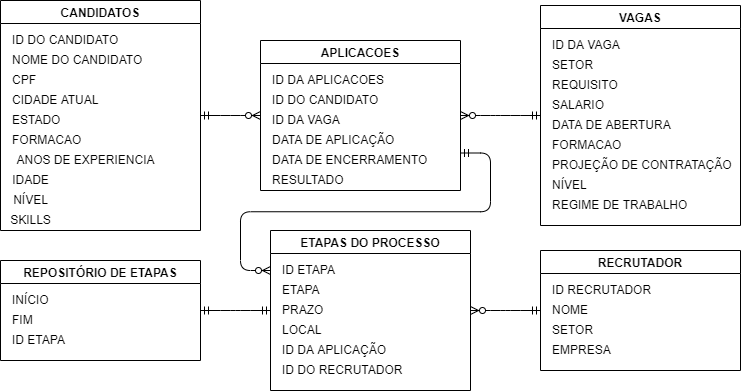
****

**Figura 2 -** Modelo conceitual do sistema de recrutamento da BLUERH.

**3.2. Modelo lógico**

A partir do modelo conceitual foi então criado o modelo lógico apresentado na Figura 3. No modelo lógico já são definidas as tabelas para armazenamento dos dados, seus atributos e como se relacionam. Nesta etapa foi identificada a necessidade de se criar uma tabela auxiliar “REPOSITÓRIO DE ETAPAS” para melhor organização dos registros e recuperação dos dados necessários para o LEAD TIME.

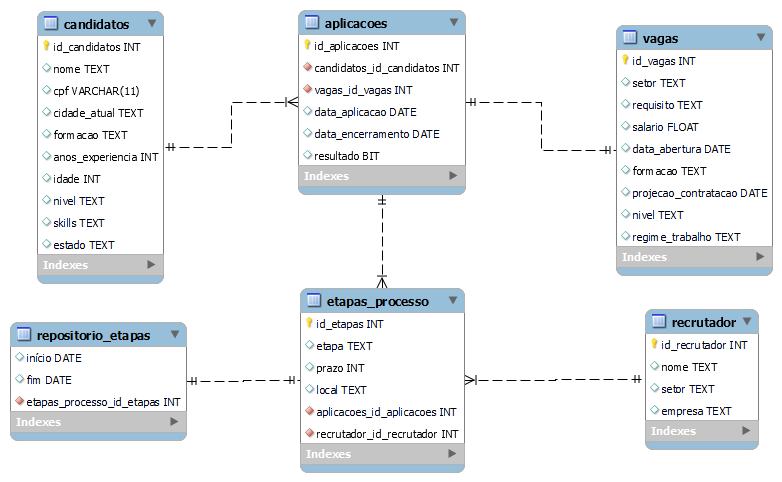
A relação entre as tabelas CANDIDATOS e APLICAÇÕES é de um para muitos pois cada candidato pode realizar mais de uma aplicação. A relação entre APLICAÇÕES e VAGAS é de muitos para um pois muitas aplicações podem ser realizadas para uma mesma vaga. A relação entre APLICAÇÕES e ETAPAS DO PROCESSO é de um para muitos já que uma mesma aplicação passa por diferentes etapas até que seja finalizada. RECRUTADOR têm uma relação de um para muitos com ETAPAS DO PROCESSO pois um mesmo recrutador pode ser responsável por diferentes etapas do processo de seleção.

****

**Figura 3** - Modelo lógico do sistema de recrutamento da BLUERH.

**3.3. Modelo Físico**

Após a aprovação do modelo lógico foi feita a construção do modelo físico descrito na Figura 4.



**Figura 4 -** Modelo físico do sistema de recrutamento da BLUERH.

A seguir é apresentada a descrição detalhada de cada uma das tabelas do modelo físico, seus atributos, tipos de dados, descrições e chaves das tabelas.

**Tabela 1** - Candidatos

| candidatos | | |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| id\_candidato | Integer | identificador único gerado conforme a inscrição do candidato. Valor inicial igual a 1. |
| nome | Text | Nome do candidato |
| cpf | VarChar(11) | Número do documento (CPF) |
| cidade\_atual | Text | Cidade de residência do candidato. |
| estado | Text | Estado de residência do candidato. |
| formacao | Text | A formação escolar do candidato |
| anos\_experiencia | Integer | Anos de experiência na área |
| idade | Integer | Idade do candidato em anos |
| nivel | Text | O grau do candidato (estagiário, júnior, pleno, sênior). |
| skills | Text | Habilidades principais do candidato. |
| Chaves da tabela | | |
| Chave primária | id\_candidato | |

**Tabela 2** - Vagas

| vagas | | |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| id\_vagas | Integer | Identificador único gerado conforme o cadastro de vagas no sistema, iniciando pelo id '1'. |
| setor | Text | Setor da empresa para o qual a vaga pertence. |
| requisito | Text | O requisito mínimo para o candidato participar do processo seletivo. |
| salario | Float | Salário médio que a vaga oferecerá. |
| data\_abertura | Date | Data de abertura da vaga |
| formacao | Text | Formação mínima para a vaga. |
| nivel | Text | Nível hierárquico da vaga (estagiário, júnior, pleno, sênior, …) |
| projecao\_contratacao | Date | Data projetada para o início de trabalho do candidato. |
| regime\_trabalho | Text | Tipo de trabalho (remoto, presencial, híbrido) |
| Chaves da tabela | | |
| Chave primária | id\_vagas | |

**Tabela 3** - Aplicações

| aplicacoes | | |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| id\_aplicacoes | Integer | Identificador único gerado conforme as aplicações dos candidatos no sistema, iniciando pelo id '1'. |
| candidatos\_id\_candidatos | Integer | Identificador referenciando o candidato. |
| vagas\_id\_vagas | Integer | Identificador referenciado a vaga. |
| data\_aplicacao | Date | Data em que o candidato fez a aplicação. |
| data\_encerramento | Date | Data em que foi encerrada a aplicação. |
| resultado | Bit | 1 se o candidato for aprovado; 0 se o candidato for aprovado |
| Chaves da tabela | | |
| Chave primária | id\_aplicacoes | |
| Chaves estrangeiras | candidatos\_id\_candidatos  vagas\_id\_vagas | |

**Tabela 4** - Etapas do processo

| etapas\_processo | | |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| id\_etapas | Integer | Identificador único gerado conforme a ordem das etapas, iniciando pelo id '1'. |
| etapas | Text | Nome das etapas (cadastro, entrevista com RH, entrevista com gestor, proposta, contratação). |
| prazo | Integer | Tempo (em dias) previsto para a conclusão da etapa. |
| local | Text | Onde ou como será a etapa (presencial, online, dinâmica de grupo…). |
| aplicacoes\_id\_aplicacoes | Integer | ID referenciando a aplicação para vaga. |
| recrutador\_id\_recrutador | Integer | ID referenciando os recrutadores. |
| Chaves da tabela | | |
| Chave primária | id\_etapas | |
| Chaves estrangeiras | aplicacoes\_id\_aplicacoes  recrutador\_id\_recrutador | |

**Tabela 5** - Recrutador

| recrutador | | |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| id\_recrutador | Integer | Identificador único gerado no cadastro do recrutador. |
| nome | Text | Nome do recrutador |
| setor | Text | Setor/função do recrutador |
| empresa | Text | Empresa a qual pertence (pode ser terceirizada) |
| Chaves da tabela | | |
| Chave primária | id\_recrutador | |

**Tabela 6** - Repositório das vagas

| repositorio | | |
| --- | --- | --- |
| Atributo | Tipo | Descrição |
| inicio | Date | Data de Início da aplicação. |
| fim | Date | Data de encerramento da aplicação. |
| etapas\_processo\_id\_etapas | Integer | Identificação referenciando a etapa do processo. |
| Chaves da tabela | | |
| Chave estrangeira | etapas\_processo\_id\_etapas | |

Após a finalização e apresentação da modelagem para os gestores do BLUERH o próximo passo foi a montagem da estrutura de dados. Esta estrutura será a base para que na próxima etapa possamos alimentar com os dados referentes aos processos de seleção e recrutamento.

Para esta etapa usamos o Scripts SQL, o mesmo estará em anexo a este documento. Após finalizar a modelagem e a documentação da estrutura de dados necessária para atender o cliente, os analistas do grupo BLUERH terminaram o trabalho inicial. Agora a equipe de RH da empresa está apta para realizar as contratações nas datas programadas, fazendo com que a empresa não atrase mais entregas.

O grupo BLUERH continuará fazendo o acompanhamento para que os processos seletivos ocorram da melhor maneira possível.

1. **Considerações finais**

O trabalho exigiu conhecimento de modelagem de dados e de MySQL, o que foi visto nestas duas primeiras semanas do módulo 5 do curso Ciência de Dados da Blueedtech. Neste trabalho foi proposto que os alunos ajudassem uma grande empresa que estava passando por problemas para identificar o LEAD TIME no processo de seleção e recrutamento.

Então no trabalho foi feita a modelagem dos dados, considerando um processo de seleção de candidatos a determinadas vagas. Nesta modelagem foram criadas 6 tabelas que depois se transformaram em linguagem SQL para que fossem estruturadas no script para receber os dados provenientes dos processos.

Foi utilizada a ferramenta MySQL Workbench para o desenvolvimento deste projeto. A modelagem construída foi capaz de atingir todos os objetivos apresentados na introdução do trabalho. Os modelos conceitual, lógico e físico do sistema de recrutamento foram construídos. Em todas as etapas do processo de seleção foram criados campos para registro e fim de cada etapa o que permitirá a obtenção do LEAD TIME de cada vaga. A estrutura também permitirá identificar quais etapas são mais rápidas ou demoradas.

A principal limitação do trabalho foi a ausência de contato com a equipe de recrutamento e seleção. Em um ambiente corporativo a equipe de recrutamento deveria ser consultada para alinhar todos os requisitos necessários para a formulação do Banco de Dados.

O projeto pode ser melhorado com a contribuição da equipe de recrutamento. A equipe poderá fornecer informações que contribuirão para a construção de um Banco de Dados mais eficiente para a demanda do setor.

Por fim, gostaríamos de ressaltar o empenho dos dois colegas presentes neste trabalho, um projeto feito em parceria com contribuição mútua. Um trabalho quando feito em grupo geralmente tende a ser um trabalho bem feito, ambos podem agregar para o desenvolvimento, assim alcançando os objetivos finais.

**Referências Bibliográficas**

**Diferença entre banco de dados relacional nosql**. **4Linux**, [s.d.]. Disponível em: <https://4linux.com.br/diferenca-banco-dados-relacional-nosql/>. Acesso em: 6 ago. 2022

BORGES, D. **Diferença entre OLTP e OLAP**. **Medium**, 5 jul. 2020. Disponível em: <https://medium.com/@david.borges/diferen%C3%A7a-entre-oltp-e-olap-b6dab6f924b8>. Acesso em: 6 ago. 2022

LEHMKUHI, D. e EGER, D. R. **Princípios de Banco de Dados**. Uniasselvi. Disponível em: <https://uniasselvi-digital.web.app/>. Acesso em: 6 ago. 2022.