*O Papel dos Bancos de Dados SQL e NoSQL na Engenharia de Dados*

**Contexto geral sobre o uso de Banco de dados na atualidade**

**“**Entender o conceito de cada um desses bancos e o propósito para qual cada um foi pensado”

**O que é SQL e NoSQL**

- O que é um Banco de Dados Relacional (SQL)?

É um tipo de Banco em que desejamos armazenar informações que tenham relação entre si.

Informações Normalizadas  
Consistência e Integridade  
Esquema de Dados Rígido (Sabemos exatamente o que queremos armazenar neste Banco)  
 Ex.: Sistemas Transacionais como E-commerce (cliente, nome, endereço, produto, descrição etc.)

- O que é um Banco de Dados Não Relacional (NoSQL)?

Surgiu da necessidade de haver uma alternativa aos bancos relacionais, no que diz respeito à Escalabilidade, grande volume e manipulação de dados não estruturados e/ou com estruturas diferentes entre si

**O NoSQL não foi criado para substituir o SQL**

- Por que dizer que o NoSQL foi criado para substituir o SQL é um equívoco?

“O NoSQL veio para atender o que a gente não conseguia fazer no relacional”, “São altamente complementares”

“Devido ao alto volume de dados na internet! O Big data necessita de formas mais flexíveis de armazenamento de dados” – Daniel Canavarro

**Como se consulta um dado armazenado no NoSQL?**

*“Sobre o NoSQL, como encontrar algo armazenado quando não se sabe a estrutura? E como isso pode performar?”*

É importante definir antes de armazenar os dados(ou criar o banco de dados) uma definida maneira mínima como eles serão consultados. Do contrário teremos um grande volume de dados sem sequer saber como consultá-los, se torna uma bagunça.  
Como a criação de uma chave que deve ser respeitada para encontrar a informação, e pensando nas características do banco.

“A aplicação tem que saber lidar com a falta de previsibilidade de um *schema*”

**Conhecer um SGBD de cada tipo (PostgreSQL e MongoDB, por exemplo) é o suficiente para iniciar?**

É importante conhecer um pouco ou, ao menos, o conceito de cada Banco de Dados. Pois ajuda quando for necessário decidir qual o melhor tipo/mais adequado para o contexto em que estiver trabalhando.  
É importante se apegar aos conceitos e não à uma tecnologia específica, estrita.

*“Se você direcionou o Banco correto para o seu problema, dificilmente você vai ter problema”*

**CONCEITO > TECNOLOGIA > APROFUNDAMENTO TÉCNICO**

**A flexibilidade do NoSQL frente ao SQL**

- “O SQL foi feito para os OLTP que precisam do ACID. Já o NoSQL veio para perverter um pouco isso e gerar mais flexibilidade”

**No Banco Carrefour, quais os SGBD’s mais utilizados?**

- “A infra do Banco está toda em nuvem, on-premise ou híbrido? Tem muito problema com manutenção?”

Oracle;

**Evolução da arquitetura de sistemas e transições de estruturas**

É preciso sempre levar em conta os casos de uso, estratégias e custos de criação, transição etc.  
É importante considerar que operações de transição tem seus custos e essencialmente devem ser feitos de maneira suave.  
Plataformas Multi-cloud são bem-vindas em muitos ambientes

**Sobre DataLake e Databricks**

-

**Quais os maiores desafios na hora de realizar o ETL?**

“Usamos uma ferramenta específica para ‘schedular’, monitorar, acompanhar os mais de mil processos diários”.

**O tipo de banco de dados influencia na complexidade de ingestão, transformação e processamento de dados?**

A princípio, não.  
A parte mais complexa se dá quando estamos construindo a arquitetura a usar dentro daquele banco, quais as formas de consulta, quais as chaves que têm de ser previstas, quais índices colocar entre outros. Um erro neste processo pode trazer dificuldades num futuro.

**Como é gerada a demanda dos dados e quem define quais dados serão coletados?**

Times distribuídos. Squads. Tribos.   
Times multi-disciplinares  
Times de engenharia muito próximos do time de negócios.

“Os nossos engenheiros acabam tendo uma grande proximidade do negócio, também para conseguir aportar, além do conhecimento técnico, entender de fato o porquê estamos fazendo aquilo e conseguir contribuir mais”

**Engenheiro e Cientista de Dados**

O papel do engenheiro está mais ligado à preparação da informação. Exige mais conhecimento técnico para coletar, preparar, armazenar e deixar a informação disponível para que os Cientistas consigam utilizar.

O Cientista está mais ligado à criação de modelos IA, Machine Learning entre outros. Aplicar um processamento na informação gerada pelo engenheiro. O que vai gerar, normalmente, algum valor associado ao negócio. Insights, comportamentos do Cliente, Predição de Negócios.

**Deficiência em pessoas com skill em Estatística**

Matemática e Estatística são skills apreciadas no contexto de Ciência de Dados e Modelagem de ML. Skills importantes assim como outras do seguimento.