



robbyson

Projeto **Nexus**

Índice

01

**Motivações e
objetivos do projeto**

02

Hipóteses



Dores e objetivos



Principal dor

Acompanhar o crescimento

- Dentro do cenário de crescimento anual de 22% a AeC está preocupada com a escalabilidade do processamento de dados junto à plataforma Robbyson.

Com essa média de crescimento, a AeC precisa garantir que não terá problemas com integração de dados, pois a Robbyson é extremamente necessária em sua operação. Além disso, enxergar escalabilidade e qualidade nos processamentos.



Objetivos do projeto Nexus

- **Processamento**

- Objetivo: Otimizar o tempo e escalabilidade do processamento de dados entre AeC e Robbyson.

- **Capacidade**

- Objetivo: Avaliar alternativas para maior capacidade em servidor (Aumentar a capacidade do servidor).

- **Consistência de Dados**

- Objetivo: Mitigar ausências de dados após reprocessamentos, e precisão de resultados informados.



Evolução
necessária
também
para a
Robbyson.

O Projeto Nexus possui como objetivo trazer um processo mais rápido, assertivo, confiável, flexível e escalável, o que se alinha com o que o time de engenharia e dados vêm pensando, para trazer escalabilidade ao que se refere a dados dentro da Robbyson.

BAVIA

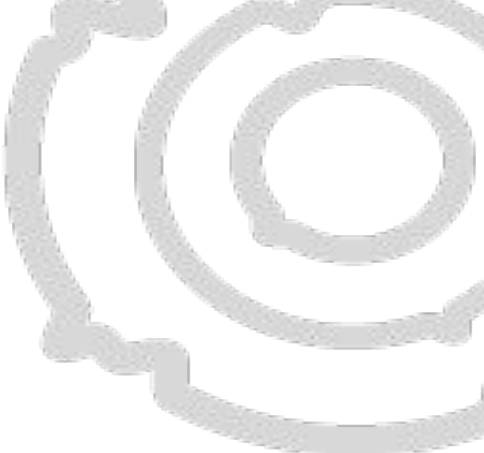


Quais perguntas precisamos responder para atingir os objetivos do projeto?

Capacidade de Processamento: Quais otimizações podemos fazer para melhorar a capacidade de processar grandes volumes de dados de forma ágil? Existe algum gargalo atual?

Parametrização: Podemos garantir que apenas as linhas alteradas sejam processadas, reduzindo a carga?

Tempo de Execução: Existe algum processo específico que está prolongando os períodos de processamento? O que podemos otimizar aqui?



A partir dessas perguntas, tentamos unir opções que podem atender a necessidade da AeC e tornar a Robbyson mais preparada para desafios como esse.



Hipóteses de soluções



1. Processamento paralelo e distribuído

De forma complementar, o **processamento paralelo** é uma forma de executar várias tarefas **simultaneamente em um único sistema**, e o processamento distribuído divide as tarefas em **vários sistemas** (chamados de “nós”).

Essas formas de processamentos - paralelo e distribuído - **são amplamente utilizados no mercado**, especialmente em grandes volumes de dados.

Empresas que lidam com **big data**, como redes sociais, plataformas de streaming e bancos, usam essas técnicas para acelerar o processamento e **melhorar a eficiência ao lidar com grandes quantidades** de informações em tempo real.



Benefícios

- **Escalabilidade:** Ele pode ser facilmente escalado de forma horizontal, atendendo o aumento de demanda com facilidade.
- **Desempenho:** Permite que as tarefas sejam executadas simultaneamente, ajudando a reduzir o tempo de processamento.
- **Tolerância à falha:** Se torna um sistema mais robusto e mais resiliente, pois a falha em uma das máquinas não prejudica o processamento total.
- **Gerenciamento de grandes volumes de dados:** Ao distribuir os dados em várias máquinas o gerenciamento de grandes volumes se torna mais eficiente, facilitando o processamento de grandes quantidades de dados.
- **Flexibilidade:** Permite atender as diversas necessidades e pode integrar inúmeras tecnologias para melhorar o processo como um todo.

Pontos de atenção



- **Necessário fazer um discovery da solução**
- **Alto custo (se mal implementado)**
- **Complexidade de implementação (nova arquitetura)**

2. Não processar todos os indicadores

Sabemos que a **AeC** possui **muitos indicadores** que não são mais utilizados ou **não possuem** uma atualização recorrente.

Se pudéssemos **definir em algum local** quais indicadores devem ser atualizados naquele momento, poderíamos **reduzir consideravelmente** o volume, pois atualizaríamos somente o necessário.

Um outro caso que pode se encaixar aqui são indicadores que não tiveram nenhuma alteração.

De que maneira poderíamos trabalhar juntos para identifica-los?



Benefícios

Desempenho: Reduzimos os dados que não precisam ser processados e, conseqüentemente, aumentamos a velocidade.

Flexível e parametrizável: A decisão de incluir ou remover um indicador do processamento poderia ser mais simples.

Pontos de atenção



Devido às várias possibilidades de alteração, teríamos que estudar maneiras seguras de fazer isso junto a AeC, demandaria um discovery mais longo.

3. Quantidade de colunas no relatório e no retorno

Nem todas as **colunas disponíveis no relatório** e no retorno são utilizadas pela AeC.

A ideia nesse caso, seria poder **parametrizar individualmente** (por contratante), removendo algumas colunas **que não são utilizadas** em determinados relatórios.



Benefícios

Redução no tempo de processamento: Sabemos que processar todas as colunas é um fator que contribui no tempo de execução. Ao tratarmos somente com as necessárias, ganharíamos um tempo bastante considerável. Lembrando que nem todas as colunas serão parametrizáveis pois algumas são obrigatórias.

Pontos de atenção



Por enquanto, não mensuramos nenhum ponto de atenção acerca dessa melhoria, mas elas podem surgir durante o processo de viabilidade.

4. Refatorar o processamento

Otimizar o **código atual** de processamento de dados. Pois o código atual foi feito quando o volume de dados era menor, além da tecnologia ter **evoluído bastante**.



Benefícios

Utilizar tecnologias atuais: Introduzir um código mais evoluído, aumentando a qualidade do processamento e diminuindo o tempo de execução.

Pontos de atenção



Demandaria discovery para confirmar sua efetividade para o caso.

5. Utilizar API

Trocar a forma de integração de banco para banco, para API.

Conforme já debatido em outros momentos, entendemos ser interessante citar essa possibilidade novamente, pois pode ser um ponto de melhora que se alinha com as outras sugestões descritas nesta apresentação.



Benefícios

Possibilidade de rodar apenas indicadores específicos: A API permite essa seleção com facilidade.

Maior segurança: Método de integração mais seguro que o atual.

Acompanhar o padrão Robbyson: Hoje na Robbyson somente a AeC utiliza o método banco para banco

Pontos de atenção



Seria necessário entender se essa mudança atenderia o propósito levantado.



Muito

Obrigado

