

GUIA DE BOAS PRÁTICAS
PLATAFORMA DE ANALYTICS IIOT

Sumário

Introdução	3
Gerenciamento de Dados.....	4
Estes dados geram valor?	4
Por quanto tempo preciso deste dado?	5
Qual o intervalo ideal de coleta deste dado?	6
O quão críticos são esses dados?	6
Dimensionamento de Recursos	7
Criação de Consultas	8
Diminua as possibilidades antes de fazer buscas	8
Conheça seus dados	8
Utilize os índices	9
Seja claro.....	9
Criação de Visualizações	10
Tenha apenas os dados necessários	10
Utilize filtros	10
Evite consultas complexas.....	10
Entenda a necessidade do usuário.....	11

Introdução

Este guia tem a finalidade de reunir boas práticas disponíveis em cada uma das frentes selecionadas, para que você possa consumir os dados gerados pela plataforma com o melhor desempenho possível.

A não observância a estas orientações não necessariamente se refletirá em perda de performance de forma imediata, tão pouco é este guia um limitador para as possibilidades observados pelos times do campo, porém é essencial que todos os colaboradores tenham ciência que diversos passos rumo a soluções menos performáticas, no médio e longo prazo, sacrificam a performance geral, e gera penalizações de desempenho para todos os que buscam consumir estes recursos.

Desejamos a todos a melhor experiência possível, e nos colocamos a disposição para entender os problemas particulares de cada área, para atender aos requisitos da melhor forma possível, visando sempre gerar o menor impacto possível nos recursos disponibilizados a todos.

Gerenciamento de Dados

Quando nos referimos a gerenciamento de dados, é importante que tenhamos em mente algumas perguntas essenciais, que são:

- Estes dados geram valor?
- Por quanto tempo preciso deste dado?
- Qual o intervalo ideal de coleta deste dado?
- O quão críticos são esses dados?

Responder estas perguntas é essencial para que possamos manter o consumo dos recursos no geral, saudáveis, bem como restringir o universo de dados disponível ao que realmente gera valor para a companhia, evitando desperdício de recursos. Abordamos então cada questão de maneira individual, explicitando as razões de cada uma, e como isso contribui para o gerenciamento de dados, e conseqüentemente, para todas as demais etapas subsequentes.

Estes dados geram valor?

Apesar de parecer uma pergunta inocente e de óbvia resposta, é essencial entender que todo dado tem um ciclo de vida, e que seu uso pressupõem a solução de um problema causador de prejuízos a execução das rotinas ou mesmo, de qualidade em etapas posteriores. Logo, se coletamos um dado meramente para “uso futuro”, esta coleta está equivocada e não deveria ser realizada nesse momento, assim como coletas que acontecem simplesmente para atividades corriqueiras, tal qual preenchimento de planilhas, e que poderiam ser obtidas a partir de outros sistemas automatizados.

Logo, quando pensamos em coleta de dados industriais, devido ao altíssimo volume que se pode alcançar caso queiramos obter todos os dados do parque industrial de uma unidade, temos que primordialmente pensar na coleta destes dados para solução de um problema maior, por exemplo, alimentação de um sistema de execução de manufatura (MES), que caso não seja automatizado, gera um consumo excessivo de mão-de-obra para tarefas triviais. Outro caso

interessante, por exemplo, seria a coleta de parâmetros utilizados pelo laboratório de qualidade, sem necessidade de extração da amostra, para análise prévia.

Resumidamente, pensemos sempre em qual problema queremos resolver com a coleta automatizada dos dados de IIOT, antes de presumir que “iremos precisar”, pois desta maneira além de diminuir a carga daqueles que farão o cadastro da coleta, ainda é diminuído o volume de dados que terá que ser gerenciado e processado, diminuindo assim o custo de manutenção e de infraestrutura.

Por quanto tempo preciso deste dado?

Quando entramos no escopo de IIOT, é necessário entender que estamos em uma escala de volume de dados muito diferente dos sistemas transacionais e, portanto, a manutenção de um período muito extenso sem a devida agregação, implica em volumes altíssimos de dados e consequentemente, na necessidade de um maior volume de armazenamento e poder de processamento. Assim, considerando este ambiente e o uso habitual, é comum que o dado em sua menor granularidade, ou seja, no menor intervalo de tempo possível, seja coerente dentro de um mesmo ciclo de processo e fora do tempo deste ciclo, podemos aplicar agregações.

Por exemplo, se um determinado processo, de seu início ao fim, dura dois dias, é essencial que possamos acompanhar as variações ao longo desse período, mas após este período, podemos apenas sumarizar os dados e eliminar os anteriores, liberando espaço e economizando poder de computação.

Qual o intervalo ideal de coleta deste dado?

Esta pergunta define o tempo de coleta do dado, que pode variar de milissegundos até a mudanças apenas quando existem alterações. Idealmente, coletamos dados apenas quando necessário, e por isso, coletasse o dado a cada mudança em seu estado. Porém, em alguns casos, aquele dado só é necessário em períodos fixos, que geram intervalos bem maiores entre as coletas, do que se coletássemos os dados de acordo com a ocorrência de suas alterações, e nessa situação, o ideal é que a coleta seja feita com a periodicidade que lhe seja coerente para a finalidade escolhida.

O quão críticos são esses dados?

A criticidade dos dados é definida pelo seu uso, desta forma, se temos um dado que irá simplesmente alimentar um *dashboard* referente a coisas triviais, com pouco ou nenhum impacto real na rotina, esse dado não é crítico e, portanto, não precisa de uma atenção especial no monitoramento. Tenha em mente que dados críticos são aqueles essenciais ao funcionamento dos processos industriais e sua coleta deve ser monitorada de perto, tanto para assegurar o funcionamento geral, quanto para entender possíveis variações atípicas.

Dimensionamento de Recursos

O dimensionamento de recursos deve ser feito sempre que possível em observância às condições a que será submetida a sua instância, e como nos testes realizados durante este trabalho, uma instância inicial com a configuração demonstrada na Tabela 1, é suficiente para cargas baixas a moderadas:

Recurso	Unidade	Quant
CPU	Cores	2
Memória RAM	GB	2
Disco	GB	500

Tabela 1 - Recursos mínimos

Como regra, todas as instâncias instaladas devem ter o monitoramento do consumo de seu *hardware*, e uma vez que seja ultrapassado o limiar saudável de maneira consistente, o *hardware* deve ser redimensionado. Também pela necessidade de escalar as capacidades das instâncias que executam cada fluxo de dados, foi selecionada a opção gerenciada em nuvem, de modo que pode ser configurado o *autoscaling*, que diminui a necessidade de acompanhamento.

Por fim, para fins de testes e acompanhamentos em ambiente pré-produtivo, recomenda-se a utilização da ferramenta *timescaledb-tune*, disponível em <https://github.com/timescale/timescaledb-tune>, para que o dimensionamento seja o mais próximo possível do ideal, dentro dos parâmetros estabelecidos pela equipe desenvolvedora do Timescale.

Criação de Consultas

A criação de consultas utilizando o TimescaleDB utiliza como padrão a linguagem SQL e, portanto, compartilha com as demais aplicações da linguagem o conjunto de boas práticas para seu uso. Por se tratar de uma linguagem amplamente difundida, o conjunto de boas práticas é igualmente amplo e por isso iremos resumir apenas alguns pontos principais que impactarão de forma específica o TimescaleDB, pelo fato do mesmo ser um banco de dados Time Series.

Diminua as possibilidades antes de fazer buscas

O TimescaleDB é um banco Time Series relacional, e isso oferece possibilidades de performar consultas em universos muito maiores de dados do que um banco time series regular, porém, todos os bancos de dados criam planos de execução para buscar os dados requisitados e isso requer que o mapeamento dos recursos envolvidos seja adequado. Desta forma, se eu desejo buscar um dado em uma tabela específica que possui algumas condições conhecidas, aplica-se a condição direto na query contendo a tabela principal apenas, evitando joins desnecessários que gerariam diminuição de performance.

Conheça seus dados

Antes de realizar consultas, busque se familiarizar com os dados, de forma que você entenda como aquela determinada entidade foi modelada, o que ela representa e as variações que ela pode apresentar devido a interações com o usuário. Diversas aplicações modificam os dados de maneira não otimizada, que fogem das boas práticas, e o desconhecimento deste tipo de interação inevitavelmente entregará dados que aparentemente não estão corretos, mas que sofreram modificações dentro de um escopo previsto.

Utilize os índices

A utilização de índices aumenta a performance das pesquisas, uma vez que oferece ao banco de dados uma estrutura organizada para identificar a localização do dado requisitado. Esta boa prática se relaciona com a anterior, uma vez que para poder realizar consultas de forma otimizada é ideal que você conheça o dado e, portanto, saiba como ele está organizado dentro do banco de dados.

Seja claro

A criação de queries específicas é muito bem-vinda em qualquer companhia, pois muitas vezes ela sumariza o conhecimento empírico dos seus colaboradores, e para que esse conhecimento se torne realmente um ativo para a companhia, é importante que os demais colaboradores consigam entender o racional por trás das consultas e que elas sejam claras. A Figura ilustra os cenários mencionados:

```
1
2  -- EVITE ESTE TIPO DE QUERY
3  SELECT a.funcionario_id, b.nome
4  FROM atividade AS a
5  LEFT JOIN funcionario as b
6  ON a._funcionario_id = b.id
7
8
9  -- PREFIRA SER EXPLICITO
10 SELECT atividade.funcionario_id, func.nome
11 FROM execucao_atividade AS atividade
12 LEFT JOIN funcionario as func
13 ON atividade._funcionario_id = func.id
```

Figura 1 - Escrita de queries

Criação de Visualizações

Tenha apenas os dados necessários

A carga de dados necessários irá causar um aumento considerável no tempo de carregamento dos visuais, causando perda de performance notável para os usuários finais. Há que se lembrar que os usuários finais tem menos conhecimento das causas das demoras, e são menos tolerantes, pois as pessoas são expostas a soluções altamente performáticas em seu cotidiano, o que causa desconforto e estranheza quando se deparam com soluções lentas e que são, ao menos aparentemente, inadequadas.

Utilize filtros

A utilização adequada de filtros está correlacionada com a etapa anterior, de forma que se a pessoa que irá construir as visualizações considerar a utilização de filtros de forma adequada, um volume ainda menor de dados irá ser exposto e consumido.

Vale lembrar, que não é a quantidade de dados em uma visualização que gera o valor para o negócio, e sim, o quão objetiva e rica é a informação que aquela determinada construção está entregando.

Evite consultas complexas

Consultas complexas implicam em maior tempo de processamento, e consequentemente, em um maior tempo de resposta. Tenha em mente qual o objetivo do usuário, quais são os requerimentos para que essa entrega realmente atenda aos critérios que foram estabelecidos.

Caso seja inviável evitar uma consulta complexa, é importante ao menos escrever essa consulta de maneira clara e com o menor número possível de operações que penalizem a performance, para que futuros mantenedores

possam desempenhar adequadamente as manutenções necessários ou futuras atualizações deste código.

Entenda a necessidade do usuário

Todas as construções feitas utilizando os dados devem necessariamente atender a uma necessidade do negócio, de forma que se essa premissa não foi observada, não existe motivo para dar seguimento. Uma vez observada essa premissa inicial, é importante levantar os requisitos básicos da iniciativa, tais como: qual a frequência de atualização é necessária, quais os indicadores que preciso acompanhar, a quais áreas irei atender com esta solução e qual o impacto que terei caso a solução venha a falhar.

O usuário está sempre no centro de todas as interações e os requisitos estabelecidos devem ser estudados, validado e enriquecidos em conjunto com os mesmos para que as visualizações entregues de fato entreguem o valor esperado.