

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

**Tecnologias de Banco de Dados (TI-BD)**

**Bases de Dados das Coisas (DoT - *Database of Things*)**

**Arthur Alves de Matos 19/0010495**

Brasília, DF

2023

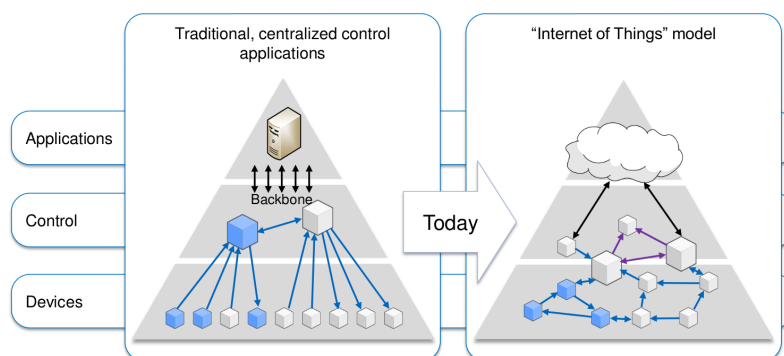
## DoT (Database of Things) - Definição e Objetivos.

Bases de Dados das Coisas, ou Database of Things (DoT) é um conceito que surge a partir da problemática da gestão de dados gerados por dispositivos da Internet das Coisas, ou Internet of Things (IoT). Seu objetivo principal é fazer com que os dados coletados a partir dos diversos dispositivos conectados por IoT, sejam devidamente armazenados e gerenciados, de maneira a fornecer um banco centralizado para permitir melhores análises e por consequência permitir melhores tomadas de decisão.

A adoção cada vez maior de dispositivos IoT acarreta, atualmente, em uma geração diária cada vez maior de dados. Tais, dados obtidos, podem ser bastante diversos, podendo estar relacionados tanto aos próprios dispositivos, com o armazenamento de informações das condições de uso dos dispositivos, quanto aos usuários, obtendo referências relacionadas ao uso e preferências. A gestão desses dados permite a empresas um melhor entendimento sobre os dispositivos e seu uso e por consequência a proposição de melhorias mais alinhadas a tais aspectos.

Devido a característica da interligação entre múltiplos dispositivos, a implementação de um DoT pode ser desafiadora. Além do fato de que os diversos dispositivos possuem diferentes fabricantes utilizando diferentes formatos de dados, o aspecto da segurança dos dados obtidos é uma preocupação constante para a implementação de tais sistemas.

Figura 1 - Modelos comparativos: Tradicional x IoT



Fonte: Db.in.tum.de

## DoT e Big Data - Diferenciação.

Figura 2 - DoT e Big Data



Fonte: FCamara

Database of Things (DoT) e Big Data são conceitos relacionados, porém podem se notar algumas diferenças entre ambos os conceitos.

O conceito de Big Data diz respeito a grandes volumes de dados, gerados rapidamente e em diversos formatos que impedem a utilização de tecnologias tradicionais para o gerenciamento. De acordo com HWAIIYU, Geng (2016, pg. 273, tradução nossa):

"Big data é sobre o uso de grandes conjuntos de dados para prever e revelar padrões comportamentais humanos. É aqui que a localização interna e o posicionamento podem ajudar desde os profissionais de marketing até mantenedores da paz da ONU a entender melhor situações e reagir de acordo".

Enquanto isso, o conceito de DoT está focado na problemática do armazenamento dos dados gerados pelos dispositivos de IoT, podendo ser considerado uma aplicação mais específica de Big Data. Em resumo, ambos os conceitos são interligados, sendo Big Data relacionado ao gerenciamento de grandes conjuntos de dados de diversas fontes enquanto que DoT se refere especificamente a problemática dos dados gerados por dispositivos IoT, sendo que ambos envolvem técnicas avançadas de análise para obter informações úteis para tomada de decisões.

## **DoT e Bancos Relacionais - Principais vantagens.**

Em relação ao gerenciamento de dados gerados por dispositivos IoT, as Bases de Dados das Coisas tem algumas vantagens sobre Bancos de Dados Relacionais (RDBs) comuns.

O DoT tem maior facilidade em trabalhar com grandes volumes de dados não estruturados, diferentemente dos bancos de dados relacionais tradicionais. Isso implica em 2 vantagens sobre os bancos relacionais. Primeiramente, a flexibilidade de armazenamento de dados de diferentes formatos, visto que os RDBs são projetados para trabalhar com conjuntos estruturados de dados enquanto o DoT não tem a limitação de formatos ou esquemas de dados pré definidos. Em segundo, a escalabilidade de DoT é facilitada, visto que sua estrutura apresenta suporte ao armazenamento de dados distribuído entre diversos servidores, facilitando a integração de mais dispositivos sem afetar necessariamente a performance do sistema e evitando o gargalo de processamento em um único servidor.

Outra vantagem sobre os RDBs, é que as Bases de Dados das Coisas apresentam otimização para o processamento de dados em tempo real, podendo trazer informações imediatas para a tomada de decisão. Além disso, trazem consigo técnicas otimizadas de indexação para o contexto de armazenamento em grande escala, o que resulta em tempos de resposta cada vez menores.

Por fim, ainda pode-se citar o suporte a integração de diversas ferramentas de análise avançada, utilizando de tecnologias como aprendizado de máquina, análise preditiva, entre outras, para facilitar a tomada de decisão tanto para empresas quanto para usuários finais.

## **DoT e Bancos Relacionais - Principais desvantagens.**

Apesar do DoT ter muitas vantagens em relação aos Bancos de Dados Relacionais tradicionais, ainda existem algumas desvantagens a serem consideradas.

O DoT é uma tecnologia relativamente nova em relação aos RDBs. Isso implica que há menos informação e conhecimento sobre implementação e gerenciamento do uso da tecnologia. Esse aspecto traz a implementação e a manutenção do DoT mais complexidade para as equipes de desenvolvimento.

O DoT também não segue as regras tradicionais de normalização de dados, que são utilizadas nos Bancos de Dados Relacionais. Isso torna ainda mais complexa a tarefa de garantir a consistência dos dados de diferentes servidores DoT. Esse aspecto traz outra desvantagem sobre os Bancos de Dados Relacionais. As Bases de Dados das Coisas são otimizadas para lidar com os grandes volumes de dados não estruturados gerados por dispositivos IoT. Isso significa que pode não ser a melhor opção para lidar com dados estruturados ou dados gerados por outros sistemas. Isso pode limitar sua capacidade de integração com outros sistemas e plataformas.

Por fim, pode-se apontar que DoT pode apresentar riscos de segurança. As informações geradas pelos dispositivos IoT podem ser altamente sensíveis. Por isso, o conhecimento sobre a segurança será denotado pelas empresas que optam por utilizar a tecnologia. É crucial para essas, garantir que as medidas de segurança adequadas sejam implementadas para proteger os dados armazenados em suas Bases de Dados das Coisas.

## **Exemplos de DoT utilizados no mercado:**

Dentro do contexto das Bases de dados da Coisas existem diversas opções no mercado, dentre elas, são apresentadas 2 exemplos de bases de código aberto difundidas no mercado:

### **InfluxDB**

O InfluxDB é um banco de dados DoT de código aberto que é projetado para lidar com dados de séries temporais. Isso torna o InfluxDB adequado para o contexto da IoT. Muitos dos dados gerados por dispositivos IoT muitas vezes são dados de séries temporais, que consistem em uma conjunto de pontos de dados que são gerados ao longo do tempo.

O InfluxDB é otimizado para lidar com grandes volumes de dados de séries temporais e tem recursos para processar dados em tempo real. Ele suporta o armazenamento de dados em massa e oferece recursos para consultas rápidas e eficientes de dados de séries temporais. O InfluxDB também oferece uma linguagem de consulta semelhante a estrutura da linguagem SQL, chamada InfluxQL, que é projetada para consultar os dados das séries temporais.

Por fim, o InfluxDB ainda possui recursos para trabalhar com dados não estruturados e permite que os usuários armazenem e consultem dados em vários formatos, como JSON, CSV e outros formatos de dados de séries temporais.

### **Amazon DynamoDB**

O Amazon DynamoDB é um banco de dados NoSQL criado e gerenciado pela AWS. O DynamoDB é projetado para lidar com grandes volumes de dados devido à sua escalabilidade horizontal automática. Isso permite que ele receba dados gerados por diversos dispositivos IoT, sem que seja afetada sua performance, adaptando sua capacidade para suprir sua demanda. O

DynamoDB, também é projetado para performar de maneira rápida e consistente, ainda que com um grandes volumes de dados e alta velocidade de gravação e leitura. Esse aspecto é crucial para lidar com dados gerados em tempo real.

Além disso, o DynamoDB oferece vários modelos de armazenamento de dados, tais como documentos, chaves e valores e grafos. Isso permite que a empresa escolha os modelos mais adequados para os dados advindos dos diversos dispositivos IoT.

Por fim, o DynamoDB está integrado aos outros serviços na AWS, permitindo uma integração do mesmo a um ambiente único de gerenciamento não somente do DoT, mas também de toda a arquitetura necessária para o sistema.

## **Case de sucesso InfluxDB - Volvo.**

A Volvo utilizou uma solução de monitoramento DevOps por mais de quinze anos. Com o tempo, a mesma se tornou obsoleta e não mais as informações que os desenvolvedores precisavam. Quando problemas de TI começaram a influenciar a produção, a equipe de capacitação DevOps da Volvo usou do InfluxDB e do Grafana para criar novos painéis e ferramentas que forneciam inteligência acionável. Após a implantação, ao invés de reagir a problemas existentes, os desenvolvedores da Volvo puderam detectar proativamente problemas e resolvê-los mais rapidamente, evitando atrasos ou paralisações na produção.

## **Problema**

A equipe de capacitação DevOps da Volvo atua como provedora interna de métricas para desenvolvedores em toda a empresa. No entanto, eles enfrentaram problemas com uma solução de monitoramento legado ineficaz que monitorava 99,8% dos componentes de TI da empresa, mas não produzia dados efetivos para os desenvolvedores. Mesmo que tivessem muitos dados,

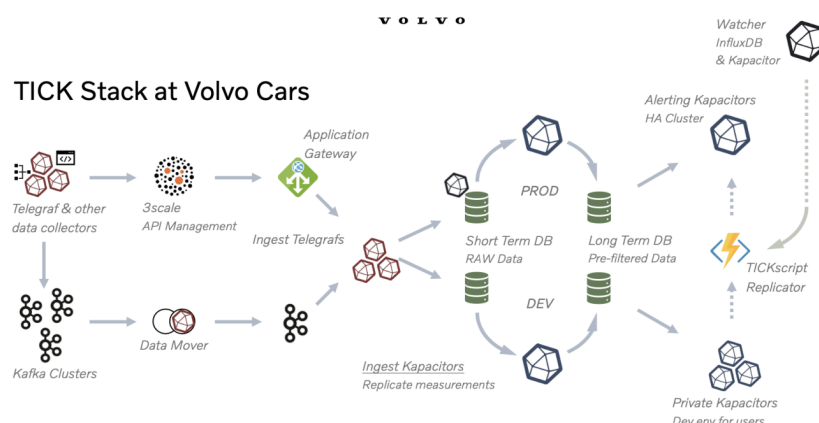
os mesmos não eram úteis e não atendiam às necessidades dos desenvolvedores.

Os desenvolvedores da Volvo não tinham como entender quando um problema atingia um nível crítico antes que fosse tarde demais, pois os limites estavam configurados incorretamente e os membros da equipe negligenciaram a correção. Isso resultou em um número crescente de incidentes e problemas de TI que afetaram as operações de produção no chão de fábrica. A falta de transparência foi outro problema, já que apenas uma pequena equipe tinha acesso a todos os dados gerados pelo sistema legado e era responsável por comunicar as informações críticas às pessoas certas no momento certo.

Para resolver esses problemas, a equipe de capacitação DevOps da Volvo queria fornecer um repositório centralizado que desenvolvedores em toda a empresa pudessem acessar e usar para acompanhar suas próprias métricas. Eles buscavam fornecer a infraestrutura e as ferramentas que permitissem que os desenvolvedores definissem seus próprios limites e gerenciassem as necessidades em constante mudança sem precisar buscar soluções externas.

## Solução

Figura 3 - Arquitetura solução volvo



Fonte: Volvo Case de sucesso



A equipe de capacitação DevOps da Volvo criou um painel do Grafana que visualizou o estado de todos os servidores da empresa. No entanto, apenas saber o estado das coisas não resolveu o problema. A equipe teve dificuldade em descobrir por onde começar a resolver esses problemas e isso exigiu muito tempo e esforço. De acordo com o aumento do número de problemas, o tempo de resolução também aumentava, o que tornava as coisas ainda piores antes que a equipe pudesse começar a resolvê-las. Isso criou um ciclo reativo de resposta e a equipe de capacitação DevOps sabia que algo precisava mudar.

A equipe então criou um novo sistema de monitoramento com o InfluxDB no núcleo. Este sistema utiliza o Telegraf e outros scripts personalizados para extração de dados. Eles usam principalmente o 3scale para gerenciamento de aplicativos e integração, e o Kafka para dados de telemetria.

Independentemente da rota, os dados chegam aos servidores de ingestão do Telegraf, que replicam os dados em ambos os clusters de desenvolvimento e produção. Isso é um ponto-chave, pois ter dados idênticos em ambos os locais permite que a equipe da Volvo desenvolva e configure os TICKscripts de forma mais eficiente. Os desenvolvedores podem testar as alterações no ambiente de desenvolvimento com os mesmos dados que estão no ambiente de produção e ver como essas atualizações afetariam os dados.

Cada cluster possui dois bancos de dados, um de curto prazo para dados brutos e um de longo prazo para dados limpos. A partir daí, a equipe configurou dois clusters do Kapacitor para notificações. Novamente, um opera na produção enquanto o outro é para desenvolvimento. Eles configuraram esses clusters para que as mudanças no ambiente de desenvolvimento possam ser rapidamente empurradas para o ambiente de produção. Os desenvolvedores simplesmente precisam mudar as informações meta no script de desenvolvimento e o replicador de TICKscript pega essas mudanças e as empurra para a produção.

O novo sistema de monitoramento de TI na Volvo, oferece transparência, autoatendimento e uma abordagem orientada a dados para a gestão de TI. A equipe DevOps utilizou dados do novo sistema para alimentar um serviço Grafana, fornecendo acesso amplo e transparente aos dados do

sistema para todos na empresa. Equipes também podem modificar alertas e outras configurações no sistema de maneira autônoma, permitindo um processo de diagnóstico e resolução mais eficiente. A implementação deste novo sistema resultou em uma mudança positiva na abordagem da empresa em relação ao monitoramento orientado a dados.

## Referências Bibliográficas

1. MATOS, David. Database of Things (DoT) – Banco de Dados das Coisas. [S. l.], 3 dez. 2015. Disponível em: <https://www.cienciaedados.com/database-of-things-dot-banco-de-dados-das-coisas/>. Acesso em: 15 abr. 2023.
2. “Blockseminar: The “Internet of Things” for Industrial Applications - Chair for Database Systems.” Disponível em: [Db.in.tum.de](http://db.in.tum.de), [db.in.tum.de/teaching/ws1314/industrialIoT/?lang=en](http://db.in.tum.de/teaching/ws1314/industrialIoT/?lang=en). Acesso em: 15 Abr. 2023.
3. HWAIIYU, Geng. Internet of Things and Data Analytics Handbook. 1. ed. New York: Wiley, 2016.
4. Volvo. Disponível em: [<https://www.influxdata.com/customer/volvo/>](https://www.influxdata.com/customer/volvo/). Acesso em: 16 abr. 2023.
5. Amazon DynamoDB. Disponível em: [<https://aws.amazon.com/pt/dynamodb/>](https://aws.amazon.com/pt/dynamodb/). Acesso em: 16 abr. 2023.