**Reconhecer cenários de aplicação de Big Data considerando dados semi e não estruturados**

Um gerente de uma grande empresa multinacional deseja utilizar técnicas de Big Data em sua companhia. Segundo seu relato, a empresa possui três áreas de atuação distintas, sendo elas: setor elétrico, setor logístico e por fim um setor de atendimento ao cliente, com as seguintes características:

* **Setor Elétrico**. A companhia é pioneira na utilização de sensores inteligentes para medição de consumo energético. Sendo que, cada sensor inteligente é instalado no local do consumidor, seja uma residência ou uma empresa. Cada sensor elétrico envia a cada minuto a informação sobre o consumo energético do cliente para uma central. O setor elétrico da companhia gerencia toda a região sul do país e atualmente conta com mais de 10 milhões de clientes.;
* **Setor Logístico.** A companhia também possui um setor de entrega de encomendas para compras on-line. Este setor permite que outras empresas utilizem o setor logístico da companhia de maneira terceirizada. Atualmente este setor possui 100 empresas de grande porte, totalizando mais de 1 milhão de entregas mensalmente. Este setor gerencia 1000 caminhões da própria companhia;
* **Setor de Atendimento ao Cliente.** A companhia também possui um setor de atendimento ao cliente. Este setor é responsável pelo atendimento dos clientes do setor logístico e elétrico. Este setor atende em média 100 clientes diariamente.

Considerando a descrição de cada um dos setores acima, e segundo o seu conhecimento sobre cenários de Big Data, para cada uma das necessidades do gerente detalhadas a seguir: (i) determine se é um cenário de Big Data, justifique sua resposta considerando os 5v’s; e (ii) determine, se o cenário for relativo a Big Data, quais dados são estruturados, não estruturados ou semi estruturados

1. *Análise em tempo real do consumo energético dos clientes do setor elétrico, a fim de reconhecer possíveis fraudes no consumo*;

**Resposta:**

Volume: O volume de dados seria alto.

Velocidade: Constante, limitada a razão de tempo.

Variedade: Não muito variável.

Veracidade: 100%.

Valor: Identificação de fraudes.

Não é um cenário de Big Data.

1. *Armazenamento da gravação dos atendimentos do último mês do setor de atendimento ao cliente;*

**Resposta:**

Volume: Não há um grande volume sendo gerado, ele é modificado todos os dias sem aumento no espaço ocupado.

Velocidade: Constante.

Variedade: Não muito variado.

Veracidade: 100%.

Valor: Validação de atendimento.

Não é um cenário de Big Data.

1. *Rastreamento da localização em tempo real dos caminhões da empresa;*

**Resposta:**

Volume: Limitado ao número de caminhões, não muito alto.

Velocidade: Constante.

Variedade: Não muito variável.

Veracidade: 100%.

Valor: Identificação de falhas ou desvios.

É um cenário de Big Data, e seria mais bem representado de forma estruturada.

1. *Análise do sentimento dos clientes do setor elétrico através do monitoramento de redes sociais;*

**Resposta:**

Volume: O volume é alto.

Velocidade: Indefinida, possivelmente alta .

Variedade: Diferentes tipos de escrita de texto, variado.

Veracidade: Questionável.

Valor: Identificação de sentimento dos clientes.

É um cenário de Big Data, e se mais bem representado de forma não-estruturada.

1. *Armazenamento das medições dos sensores elétricos do último mês;*

**Resposta:**

Volume: O volume de dados seria fixo.

Velocidade: Constante.

Variedade: Não muito variável.

Veracidade: 100%.

Valor: Identificação de irregularidades.

Não é um cenário de Big Data.