#### **1. Introdução**

* **Nome do Projeto:**  Análise de Big Data para E-commerce
* **Versão do Documento: 1.0**
* **Data: 24/10**
* **Autores:** Lucas Mello Muller de Oliveira , Nicolas da Silva Santos, Pedro Henrique Renner, Eduardo Varela Smyk, Iago Vicente Salvador.
* **Objetivo:**
  + Este projeto visa implementar uma solução de análise de Big Data para um sistema de e-commerce, utilizando Scala para processar grandes volumes de dados, identificar tendências de vendas e comportamento de clientes, e fornecer insights para apoiar a tomada de decisões estratégicas.

#### **2. Visão Geral**

* **Descrição do Sistema:**
  + O sistema de análise de Big Data será capaz de processar dados de transações, comportamento de navegação dos clientes, taxas de conversão, entre outros. Os dados processados serão utilizados para gerar relatórios e dashboards que ajudem a entender melhor o desempenho das vendas, a eficiência das campanhas de marketing e a satisfação do cliente.  
    O foco será na utilização de ferramentas como Apache Spark com Scala para processar grandes conjuntos de dados de forma eficiente.
* **Stakeholders:**
  + **Gerente de Marketing:** Utiliza os insights gerados para ajustar campanhas e estratégias de marketing.
  + **Gerente de Vendas:** Acompanha os dados de vendas e estoque em tempo real para tomar decisões operacionais.
  + **Equipe de TI:** Garante a infraestrutura adequada para o processamento e armazenamento dos dados.
  + **Equipe de Data Science:** Desenvolve os modelos de análise, utilizando Scala para processar e transformar os dados.
  + **Equipe de Suporte ao Cliente:** Analisa dados relacionados à satisfação do cliente e o tempo de resposta nas interações.

#### **3. Requisitos Funcionais**

| **ID** | **Descrição** | **Prioridade** | **Dependências** | **Critério de Aceitação** |
| --- | --- | --- | --- | --- |
| RF-01 | O sistema deve processar dados de transações de vendas diariamente. | Alta |  | Dados de vendas são processados e agregados diariamente sem falhas.. |
| RF-02 | O sistema deve gerar relatórios de desempenho de vendas por categoria de produto. | Média | RF-01 | Relatórios mostram corretamente as vendas por categoria dentro do período especificado. |
| RF-02 | O sistema deve permitir análise do comportamento de navegação dos clientes. | Alta | RF-01 | Relatórios mostram padrões de navegação como páginas mais acessadas e tempo médio no site. |
| RF-02 | O sistema deve identificar padrões de compra e sugerir produtos populares. | Média | RF-01 | Sugestões de produtos populares baseadas em padrões de compra são geradas corretamente |

#### **4. Requisitos Não Funcionais**

| **ID** | **Descrição** | **Prioridade** | **Dependências** | | **Critério de Aceitação** |
| --- | --- | --- | --- | --- | --- |
| RNF-01 | O sistema deve ser capaz de processar grandes volumes de dados (> 10TB) diariamente. | Alta | - | Testes de carga mostram que o sistema processa volumes de dados superiores a 10TB diariamente. | |
| RNF-02 | O tempo de resposta para consultas em grandes conjuntos de dados não deve exceder 3 segundos. | Média | RNF-01 | Testes de desempenho mostram tempo de resposta médio abaixo de 3 segundos para consultas. | |
| RNF-03 | O sistema deve ser capaz de realizar processamento de dados em lote (batch processing) e em tempo real (real-time processing). | Alta |  | Testes mostram que o sistema consegue processar grandes volumes de dados históricos (batch) e dados em tempo real (ex.: dados de comportamento de usuários) sem comprometer o desempenho. | |
| RNF-04 | O sistema deve garantir a segurança dos dados dos clientes, atendendo às exigências da LGPD. | Alta |  | Auditorias de segurança mostram conformidade com a LGPD e que os dados são manipulados de forma segura. | |

### 

### **5. Casos de Uso**

**Caso de Uso 1:** Analisar Vendas por Produto  
**Descrição:** O gerente de vendas utiliza o sistema para analisar as vendas por categoria de produto em um determinado período.  
**Ator(es):** Gerente de Vendas  
**Pré-condição:** O sistema deve estar integrado à base de dados de transações do e-commerce.  
**Fluxo Principal:**

1. O gerente acessa o sistema e escolhe o período de análise.
2. O sistema processa os dados de vendas para o período selecionado.
3. O sistema exibe os relatórios com a quantidade de vendas por produto e categoria.  
   **Fluxo Alternativo:**Se não houver dados para o período selecionado, o sistema notifica o usuário.

**Caso de Uso 2:** Analisar Comportamento de Clientes  
**Descrição:** O gerente de marketing acessa o sistema para analisar os padrões de comportamento de navegação dos clientes.  
**Ator(es):** Gerente de Marketing  
**Pré-condição:** O sistema deve ter acesso aos dados de navegação dos clientes.  
**Fluxo Principal:**

1. O gerente seleciona o tipo de comportamento a ser analisado (ex.: páginas mais acessadas, tempo médio no site).
2. O sistema processa os dados de navegação e gera os relatórios correspondentes.  
   **Fluxo Alternativo:**Se os dados de navegação não estiverem disponíveis, o sistema exibe uma mensagem de erro.

**Caso de Uso 3:** Identificar Produtos Populares  
**Descrição:** O sistema analisa os padrões de compra para identificar produtos mais populares e fornecer recomendações.  
**Ator(es):** Sistema (processamento automatizado)  
**Pré-condição:** Os dados de vendas e comportamento de cliente devem estar disponíveis.  
**Fluxo Principal:**

1. O sistema analisa as transações para identificar produtos mais comprados.
2. Sugestões de produtos populares são geradas automaticamente e exibidas nos relatórios.  
   **Fluxo Alternativo:**Se não houver dados suficientes para a análise, o sistema notifica o usuário.

### **6. Restrições**

* O sistema deve estar disponível 24/7 para processar os dados e gerar relatórios de forma contínua.
* O processamento dos dados deve ser feito de acordo com as regulamentações de proteção de dados (LGPD).
* O sistema deve estar integrado a uma infraestrutura de Big Data que suporte escalabilidade.

### **7. Tecnologias Utilizadas**

* **Linguagem de Programação:** Scala
* **Plataforma de Big Data:** Apache Spark para processamento de grandes volumes de dados.
* **Banco de Dados:** Hadoop HDFS para armazenamento distribuído dos dados.
* **Ferramentas de Visualização:** Tableau, Power BI para visualização dos dados analisados.
* **Infraestrutura:** Amazon Web Services (AWS) para processamento em nuvem e escalabilidade.
* **Segurança:** Implementação de autenticação OAuth e criptografia SSL para proteção de dados.