

****template - Arquitetando o Universo Big Data da Melhores Compras****

**Patrícia Maura Angelini**

**Versão 1**

**HISTÓRICO DE VERSÕES**

|  |  |  |  |
| --- | --- | --- | --- |
| Versão | Data | Responsável | Descrição |
| 1 | 18/06/2024 | Patrícia Maura Angelini | Versão Inicial Template PBL Fase 6 - CAP 01 - ARQUITETANDO O UNIVERSO BIG DATA DA MELHORES COMPRAS |
|  | 27/06/2024 | Rita de Cássia Rodrigues | Revisão acadêmica |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

FICHA CATALOGRÁFICA   
**[NÃO PREENCHER - PARA USO DO DEPTO DE EAD E BIBLIOTECA]**

A000a Sobrenome, Nome

Título [livro eletrônico] / Nome Sobrenome. -- São Paulo : Fiap, 2016.

x MB ; ePUB

Bibliografia.

ISBN 000-00-00000-00-0

Categoria. 2. Subcategoria. S., Nome. II. Título.

CDU 000.000.00

RESUMO

Template para atividade de PBL fase 6 1º ano TSC.

Palavras-chave: PBL. FASE 6. TEMPLATE

LISTA DE Figuras

[Figura 1 – Backlog e Atividades no Quadro Kanban 13](#_Toc169645381)

[Figura 2 - Detalhe item de backlog 13](#_Toc169645382)

[Figura 3 – Big Data Pipeline Architecture 16](#_Toc169645383)

[Figura 4 – Big Data Pipeline, diversos vendors 17](#_Toc169645384)

[Figura 5 – Big Data Pipeline, Open Source Stack 18](#_Toc169645385)

LISTA DE QUADROS

[Quadro 1 – Quadro resumo das tarefas do PBL 12](#_Toc169645325)

LISTA DE TABELAS

[Tabela 1 – Origem de Dados 14](#_Toc169645329)

[Tabela 2 - Dicionário de dados de colunas de tabelas 15](#_Toc169645330)

LISTA DE CÓDIGOS-FONTE

No table of figures entries found.LISTA DE ComandoS de prompt do sistema operacional

No table of figures entries found.

Sumário

[Potencializando o desempenho com NoSQL 11](#_Toc169645297)

[1 iNSTRUÇÕES PARA USO DO TEMPLATE 11](#_Toc169645298)

[1.1 Template 11](#_Toc169645299)

[1.1 Instruções 11](#_Toc169645300)

[2 desafio(s) enfrentado(s) pela melhores compras 12](#_Toc169645301)

[2.1 Contextualização do Problema 12](#_Toc169645302)

[3 Planejamento das atividades 12](#_Toc169645303)

[3.1 Kanban das Atividades 12](#_Toc169645304)

[3.2 Detalhamento das Atividades 13](#_Toc169645305)

[3.3 Link do Kanban 14](#_Toc169645306)

[4 origem dos dados 14](#_Toc169645307)

[4.1 Panorama geral das fontes de dados 14](#_Toc169645308)

[4.2 Justificativas das Fontes de Dados 14](#_Toc169645309)

[4.3 Detalhamento das fontes de dados 15](#_Toc169645310)

[5 Arquitetura de Solução Big Data / Pipeline de Dados 16](#_Toc169645311)

[5.1 Desenho da Arquitetura 19](#_Toc169645312)

[5.2 Justificativa da Arquitetura 19](#_Toc169645313)

[5.3 Detalhamento da Arquitetura 19](#_Toc169645314)

[REFERÊNCIAS 20](#_Toc169645315)

[GLOSSÁRIO 20](#_Toc169645316)

# ****Potencializando o desempenho com NoSQL****

# 1 iNSTRUÇÕES PARA USO DO TEMPLATE

## 1.1 Template

Um template é um modelo predefinido que simplifica e padroniza a criação de documentos, páginas da web ou outros elementos de design, permitindo que usuários preencham apenas as partes específicas, economizando tempo e garantindo consistência visual ou estrutural.

Para seu uso leia as instruções nesse e use o template a partir da seção 2

## 1.1 Instruções

O template está dividido por seções que representam pedidos feitos para o PBL. No template iremos encontrar exemplos ilustrativos de entregas, cujo contexto não tem relação ao que está sendo solicitado no PBL.

Ao usar o template, apague por completo a seção 1 de instruções para que o seu trabalho fique mais organizado.

Não se esqueça de salvar o template em .pdf para a entrega.

A seguir um Quadro Resumo das tarefas:

|  |
| --- |
| Para um ou mais problemas a serem resolvidos (ou crie um).   1. Abandono do carrinho de compras em larga escala. 2. Falta de conhecimento do perfil do usuário anônimo e do cliente. 3. Diminuição da taxa de conversão (usuário anônimo para cliente). 4. Estoque ocorrendo rupturas. 5. Aumento considerável no tempo de entrega e crescente aumento de entregas indevidas. 6. Aumento do número de reclamações no SAC. 7. Alto consumo de tempo (inviável) para processamento dos dados.   A partir deles desenvolva os passos:   * Contextualização do problema * Planejamento de atividades   + Kanban contendo todas as atividades   + Detalhamento da atividade   + Link para navegação no Kanban * Origem de Dados   + Panorama Geral das Fontes de Dados   + Justificativas das Fontes de Dados   + Detalhamento das Fontes de Dados * Arquitetura de Solução Big Data / Pipeline de Dados   + Desenho da Arquitetura (com link opcional)   + Justificativa da Arquitetura   + Detalhamento da Arquitetura |

Quadro 1 – Quadro resumo das tarefas do PBL

Fonte: Elaborado pelo autora (2024)

# 2 desafio(s) enfrentado(s) pela melhores compras

## 2.1 Contextualização do Problema

Cite o(s) problema(s) a escolhidos para serem resolvidos e explique por que resolver esses problemas são prioridade para a Melhores Compras.

# 3 Planejamento das atividades

Planeje seu projeto organizando suas atividades por meio de um Kanban Board. Utilize ferramentas como o Trello (https://trello.com/), ou o Azure Boards (https://dev.azure.com/), ou qualquer outra de sua preferência, como no exemplo. Não se esqueça de mostrar os detalhes das principais atividades

Se quiser, além das imagens, disponibilize o link.

### 3.1 Kanban das Atividades

Coloque a imagem (print+screen) do Quadro Kanban como no exemplo:

A screenshot of a computer

Description automatically generated

Figura 1 – Backlog e Atividades no Quadro Kanban

Fonte: Elaborado pela autora (2023)

.

### 3.2 Detalhamento das Atividades

Para cada atividade, coloque a imagem (print + screen) do seu detalhamento como no exemplo:

A screenshot of a computer

Description automatically generated

Figura 2 - Detalhe item de backlog

Fonte – Elaborado pela autora (2023)

### 3.3 Link do Kanban

Forneça o link para navegação no Kanban. Não esqueça de deixar o Kanban público.

# 4 origem dos dados

Defina quais origens de dados serão utilizadas, bem como seu formato, velocidade, volume, horário de coleta, localização e o proprietário dessas fontes de dados. Caso essa fonte de dados seja estruturada, você pode detalhar ainda mais, criando metadados de alta qualidade.

## 4.1 Panorama geral das fontes de dados

Coloque aqui o quadro geral das fontes de dados

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Origem | Formato | Velocidade | Volume | Horário Coleta | Localização | Proprietario |
| Banco de Dados Oracle | Tabela | Coletado em Batch diariamente | Pequeno: aproximadamente 1.000 registros dias | Todos dias as 20:30hs | Oracle Cloud | Diretor de desenvolvimento de produto (P&D), único autorizado a incluir ou alterar dados de produtos |
| Servidor Apache | Logs de Servidor | Coletado em Batch  duas vezes ao dia  (uma de manhã e uma a tarde) | Um volume pequeno aproximadamente 1.200 registros dia | Manhã: 8h  Tarde: 18h |  | Diretor de Desempenho da plataforma Melhores Compras |

Tabela 1 – Origem de Dados

Fonte -Elaborado pela autora (2024)

## 4.2 Justificativas das Fontes de Dados

Justifique as escolhas das fontes de dados. Observe o exemplo a seguir:

**Banco de Dados Oracle:** A empresa já possui, de maneira estruturada, um vasto conjunto de dados que abrange diversas áreas cruciais para o seu funcionamento. Estes dados incluem informações detalhadas sobre clientes, produtos, pedidos, pagamentos e inventário. Esses dados estruturados serão fundamentais para o pipeline de processamento. A empresa poderá otimizar processos, melhorar a experiência do cliente e maximizar a eficiência nas operações diárias.

**Logs de Servidor:** Importante para que possamos entender e descobrir, onde temos um maior volume de acessos, dessa forma podemos observar se os recursos de rede estão gerando maiores delays dentro da plataforma e quais páginas estão sendo mais procuradas para que sejam mais bem otimizadas.

## 4.3 Detalhamento das fontes de dados

Detalhe as fontes de dados.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tabela | Apelido | Descrição | Interessados | Dono da informação | Retenção |
| T\_SIC\_Produto | Prod | Retém dados de produtos comercializados pela empresa. | Vendedores, gerentes e diretores de vendas, para consulta do catálogo de produtos | Diretor de desenvolvimento de produto (P&D), único autorizado a incluir ou alterar dados de produtos | Somente produtos que tenham vendas nos últimos 300 dias são mantidos no sistema. Os demais são considerados descontinuados e podem ser eliminados dos registros. Backup diário full. |

Tabela 2 - Dicionário de dados de colunas de tabelas

Fonte - Material da fase – Elaborado pelo autor

# 5 Arquitetura de Solução Big Data / Pipeline de Dados

O objetivo agora é fazer o desenho da sua solução de big data. Inicie o desenho criando a camada de Data Source representando as origens de dados definidas. Descreva os passos da sua arquitetura

Observe com atenção os modelos de Arquitetura de Big Data para base / inspiração, e desenhe a sua arquitetura de dados, justificando cada passo

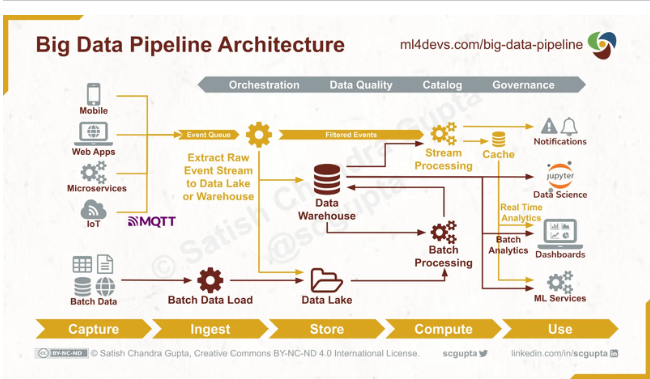
****

Figura 3 – Big Data Pipeline Architecture

Fonte -http://scgupta.link/big-data-pipeline

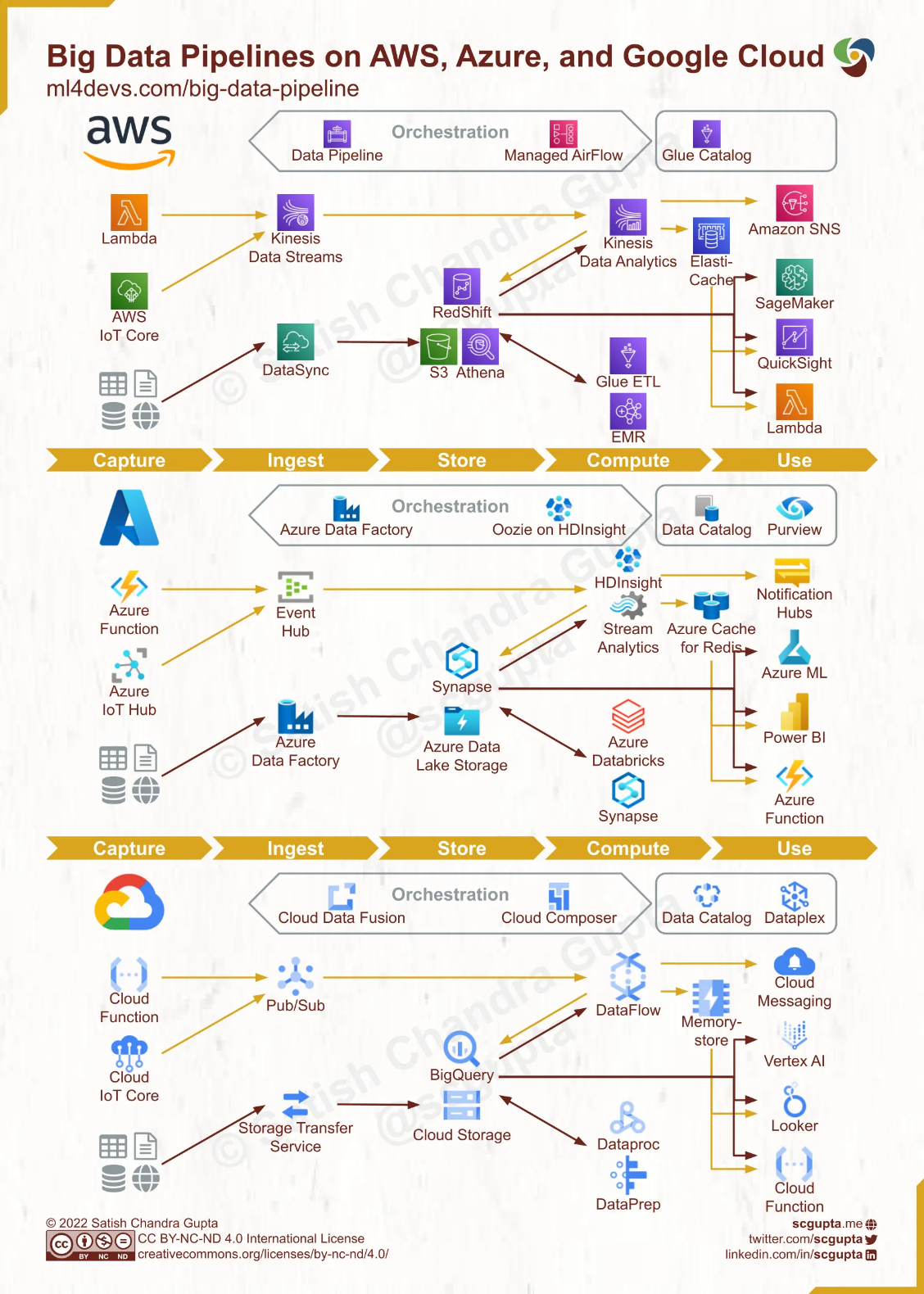


Figura 4 – Big Data Pipeline, diversos vendors

Fonte -http://scgupta.link/big-data-pipeline

A screenshot of a computer

Description automatically generated

Figura 5 – Big Data Pipeline, Open Source Stack

Fonte -http://scgupta.link/big-data-pipeline

## 5.1 Desenho da Arquitetura

Faça o seu desenho.

a) Complemente seu desenho de solução com a camada Data Lake e respectivos formatos de armazenamento, utilizando como premissa as informações da camada de Data Source. A camada Data Lake é responsável por adquirir dados das fontes e, se necessário, convertê-los para um formato adequado à maneira como os dados devem ser analisados. E o melhor, essa camada pode ser distribuída, ou seja, estar localizada em diversos provedores de cloud ou mesmo ser implementada em um modelo híbrido (Cloud e On premise).

b) Incorpore a camada de análise em seu desenho de solução: essa camada lê os dados do Data Lake, aplica algoritmos de análise em busca de insights para melhorar a tomada de decisão.

c) Incorpore a camada de consumo em seu desenho de solução: essa camada tem por objetivo consumir a saída fornecida pela camada de análise nos mais diversos formatos (visualização, KPIs operacionais, alertas de negócio, disponibilizar dados higienizados para machine learning, entre outros).

Se quiser disponibilize o link

Para mais detalhes:

Link: https://www.canva.com/design/xxxxxx

## 5.2 Justificativa da Arquitetura

Justifique porque está escolhendo esta arquitetura.

## 5.3 Detalhamento da Arquitetura

Explique cada uma das camadas e componentes da sua arquitetura.

A camada de captura de dados considera.....

Para o processamento de dados estruturado escolhemos o componente.....

A camada de ingestão trata....

(continue para as outras camadas)

# REFERÊNCIAS

SOBRENOME, Nome do autor abreviado. **Título do livro.** Local da edição: Editora, ano.

# GLOSSÁRIO

|  |  |
| --- | --- |
| **Termo** | Explicação. |
| **Termo** | Explicação. |