

Especificação de Requisitos e Mapeamento de Dados – SIPD Fase 1

1. Introdução

Este documento tem como objetivo detalhar os requisitos funcionais da fase 1 do projeto SIPD, abordando a automação da integração entre o Data Lake e o Power BI, além de fornecer um mapeamento detalhado dos dados armazenados no MongoDB que serão utilizados nos dashboards do Power BI.

2. Visão Geral da Solução

A solução proposta tem como finalidade otimizar o fluxo de dados entre o MongoDB, o Data Lake e o Power BI, garantindo uma extração eficiente, armazenamento seguro e disponibilização adequada das informações para análise e tomada de decisão.

O fluxo de dados será estruturado da seguinte forma:

1. Extração de Dados:

- a. O sistema realizará a extração dos dados diretamente do MongoDB por meio de consultas estruturadas, garantindo a obtenção das informações necessárias sem impactar o desempenho da base de produção.
- b. A extração será realizada em intervalos configuráveis, assegurando a atualização contínua dos dados.

2. Armazenamento no Data Lake:

- a. Os dados extraídos do MongoDB serão transformados e armazenados no Azure Data Lake, seguindo um modelo de particionamento que facilite a organização e otimização de consultas.
- b. O armazenamento será estruturado por categorias, tais como alertas, rondas e verificações, permitindo um acesso eficiente aos diferentes tipos de informações

3. Processamento e Transformação:

- a. O sistema realizará a padronização e normalização dos dados, garantindo consistência entre os diferentes registros.
- b. Serão aplicadas validações para identificação e tratamento de inconsistências nos dados antes de serem disponibilizados para consumo.

4. Integração com Power BI:

- a. O Power BI se conectará diretamente ao Data Lake para consumo dos dados processados.
- b. A disponibilização das informações no Power BI seguirá um modelo otimizado, utilizando tabelas dimensionais e de fatos para garantir a eficiência na análise e visualização.

Com essa abordagem, a solução visa eliminar a necessidade de processos manuais de extração e transformação de dados, proporcionando maior confiabilidade e eficiência na análise operacional e de desempenho do SIPD.

3. Requisitos Funcionais

ID	Requisito	Descrição
RF01	Extração de Alertas	O sistema deve extrair dados de alertas gerados no SIPD e armazená-los no Data Lake. Essa extração deve ocorrer em intervalos configuráveis e incluir dados de geolocalização, timestamps e classificação de alerta.
RF02	Extração de Rondas	O sistema deve consolidar informações sobre as rondas realizadas, incluindo status, tempo de execução, equipe responsável e rota percorrida. Deve permitir a recuperação de históricos e correlacionar com alertas próximos.
RF03	Extração de Verificações	Dados de verificação devem ser extraídos e consolidados para análise de eficiência das equipes. Deve incluir detalhes como motivo da verificação, status final, ações tomadas e tempo médio para resolução.
RF04	Integração com Power BI	Os dados armazenados no Data Lake devem estar disponíveis para análise no Power BI. A integração deve ser feita via pipelines automatizados, garantindo a atualização periódica dos dados.
RF05	Transformação e Padronização	O sistema deve realizar transformações nos dados extraídos para garantir consistência, como padronização de timestamps, normalização de status e unificação de registros duplicados.

4. Mapeamento de Dados (MongoDB x Power BI)

A tabela abaixo apresenta o mapeamento de onde os dados utilizados no Power BI estão localizados no MongoDB.

Indicador no Power BI	Collection no MongoDB	Campo(s) Relacionado(s)
Filtro de Sistemas	CCPD-REGISTRATIONS	<ul style="list-style-type: none">• Ccpd-registrations.gasducts• Ccpd-registratrions.states
Filtro de Período da Patrulha	CCPD-PATROL	<ul style="list-style-type: none">• Ccpd-patrol.patrol_repetition• Ccpd-patrol.patrols• Ccpd-patrol.patrolteams
Filtro Tipo de alerta	CCPD-ALERTS	<ul style="list-style-type: none">• Ccpd-alerts.alerts
Filtro Status de Alerta	CCPD-ALERTS	<ul style="list-style-type: none">• Ccpd-alerts.alerts
Gráfico de barras: Qtd alertas por sistemas	CCPD-REGISTRATIONS	<ul style="list-style-type: none">• Ccpd-registrations.opticalProcessingSystem• Ccpd-registrations.gasducts• Ccpd-registrations.inspection_points
Gráfico de barras: Tipos de Alerta (qtd)	CCPD-REGISTRATIONS	<ul style="list-style-type: none">• Ccpd-registrations.procedures
Gráfico de rosca: Status (verificados)	Ccpd-alerts	<ul style="list-style-type: none">• Ccpd-alerts.alerts
Gráfico de barras: KP mais alarmados	CCPD-EVENT-VERIFICATION	<ul style="list-style-type: none">• Ccpd-event-verification.verifications

Gráfico barras: Quantidade de Alertas	CCPD-ALERTS	<ul style="list-style-type: none"> • Ccpd-alerts.alerts
Gráfico de barras: Quantidade de Patrulhas por período	CCPD-PATROLS	<ul style="list-style-type: none"> • Ccpd-patrol.patrol_repetition • Ccpd-patrol.patrols • Ccpd-patrol.patrolteams
Card: Tempo Médio de ruído	CCPD-EVENT- VERIFICATION e CCPD-ALERTS	<ul style="list-style-type: none"> • Ccpd-event-verification.events • Ccpd-event-verification.events_verification
Card: Tempo médio por alerta	CCPD-EVENT- VERIFICATION	<ul style="list-style-type: none"> • Ccpd-event-verification.events
Card: Qtd de alertas	CCPD-ALERTS	<ul style="list-style-type: none"> • Ccpd-alerts.alerts • Ccpd-alerts.sequence
Canais mais alarmados	CCPD – ALERTS, CCPD-EVENT- VERIFICATION e CCPD- REGISTRATIONS	<ul style="list-style-type: none"> • Ccpd-alerts.alerts.event-verification • Ccpd-event-verification.event_verifications • Ccpd-registrations.channels • Ccpd-registrations.channels • Ccpd-registrations.companies • Ccpd-registrations.companies
TACCPD: tempo de atendimento entre o início do alerta e a solicitação para	Database: CCPD-EVENT- VERIFICATIONS	<ul style="list-style-type: none"> • taccpd = sentDate - eventCreatedAt

verificação/inspeção PD.	Collection: verifications	
TACS: tempo entre a criação do registro e a solicitação para verificação/inspeção PD.	Database: CCPD-EVENT- VERIFICATIONS Collection: verifications	<ul style="list-style-type: none"> tacs = sentDate - createdAt
TAPD: tempo de atendimento para verificação/inspeção PD.	Database: CCPD-EVENT- VERIFICATIONS Collection: verifications	<ul style="list-style-type: none"> tapd = dateSendEndVerification - sentDate
TTA: tempo total de atendimento.	Database: CCPD-EVENT- VERIFICATIONS Collection: verifications	<ul style="list-style-type: none"> tta = dateSendEndVerification - eventCreatedAt
TERF: tempo de entrega do relatório final.	Database: CCPD-EVENT- VERIFICATIONS Collection: verifications	<ul style="list-style-type: none"> terf = dateSendEndVerification - dateManagementReport

TVI: tempo de verificação/inspeção.	Database: CCPD-EVENT-VERIFICATIONS Collection: verifications	<ul style="list-style-type: none"> • $tv_i = \text{endDateInspection} - \text{startDateInspection}$
TSI: tempo entre solicitação e início da Inspeção.	Database: CCPD-EVENT-VERIFICATIONS Collection: verifications	<ul style="list-style-type: none"> • $ts_i = \text{startDateInspection} - \text{sentDate}$
TRIVI: tempo de resposta para início da verificação/inspeção.	Database: CCPD-EVENT-VERIFICATIONS Collection: verifications	<ul style="list-style-type: none"> • $trivi = \text{startDateMovement} - \text{sentDate}$
Card: Qtd de verificação	CCPPD-EVENT-VERIFICATION	<ul style="list-style-type: none"> • Ccpd-event-verifications.events • Ccpd-event-verifications.event_verification • Ccpd-event-verifications.event_main_actions

Resumo dos locais a serem extraídos os dados

Banco de Dados: CCPD-REGISTRATIONS

- Ccpd-registrations.gasducts
- Ccpd-registrations.states
- Ccpd-registrations.opticalProcessingSystem
- Ccpd-registrations.inspection_points
- Ccpd-registrations.procedures
- Ccpd-registrations.channels
- Ccpd-registrations.companies

Banco de Dados: CCPD-PATROL

- Ccpd-patrol.patrol_repetition
- Ccpd-patrol.patrols
- Ccpd-patrol.patrolteams

Banco de Dados: CCPD-ALERTS

- Ccpd-alerts.alerts
- Ccpd-alerts.sequence

Banco de Dados: CCPD-EVENT-VERIFICATION

- Ccpd-event-verification.verifications
- Ccpd-event-verification.events
- Ccpd-event-verification.events_verification
- Ccpd-event-verification.event_verifications
- Ccpd-event-verification.event_main_actions

5. Arquitetura e Fluxo de Dados

- Os dados são extraídos do MongoDB por meio de consultas estruturadas nas collections listadas acima.
- Após a extração, os dados são armazenados no Azure Data Lake, organizados por data e categoria.
- O Power BI se conecta ao Data Lake para acessar os dados processados e gerar dashboards.

6. Considerações Finais

Este documento servirá como referência para o DBA responsável pela extração de dados e também para a equipe de BI na definição de relatórios e dashboards. Ajustes adicionais podem ser realizados conforme a necessidade do projeto.