APENSO II - CATÁLOGO DE SERVIÇOS

(Referente aos itens 2 e 3 do TR)

1. PREMISSAS DO MODELO DE OPERAÇÃO ASSISTIDA

Diferente do suporte técnico, a Operação Assistida é proativa. Ela assume a responsabilidade pela criação, manutenção e evolução dos ativos de dados.

Natureza Proativa: Diferente do suporte técnico reativo, a Operação Assistida assume a responsabilidade pela criação, manutenção e evolução contínua dos ativos de dados.

Modelo de Mensuração: HST (Horas de Serviço Técnica agrupadas por "Produto de Dados". A prestação de serviços ocorrerá sob demanda, mensurada por HST, baseada estritamente na entrega de produtos e resultados, sendo expressamente vedada a mera locação de mão de obra.

Transferência de Conhecimento: Todas as entregas devem ser acompanhadas de execução de treinamentos, *workshops* técnicos e criação de documentação detalhada (Wikis, Manuais), assegurando que o conhecimento intelectual permaneça no MPMG.

Ambiente: Cluster TDP Kubernetes (foco em elasticidade, microsserviços e containers).

2. INTRODUÇÃO

As atividades elencadas no escopo típico de cada item de serviço constituem o núcleo essencial desses itens. Todavia, tais atividades não se esgotam em si mesmas, podendo ser necessárias outras ações complementares para o pleno atendimento da demanda prevista em cada Ordem de Serviço. Nessas situações, deverá ser observada a estrita correlação entre a natureza do item contratado e todas as atividades a ele associadas.

3. ITENS, ATIVIDADES E COMPLEXIDADE

GRUPO 1: OPERAÇÃO ASSISTIDA ESPECIALIZADA EM ENGENHARIA DE DADOS

Objetivo: Garantir a saúde, eficiência computacional da plataforma e a ingestão automatizada e segura dos dados para o Datalake.

Item 1.1 - Desenvolvimento e Sustentação de Pipelines de Ingestão (Data Ops)

Descrição das Atividades:

- Apoio à Migração do Legado (Hadoop/HDP): Execução e apoio na migração de workloads do ambiente legado para a nova arquitetura cloud-nativa (TDP Kubernetes), com adequação e refatoração de pipelines para minimizar riscos de descontinuidade.
- Construção de APIs de Dados: Desenvolvimento de Application Programming Interfaces (APIs), criando conjuntos de regras e definições seguras para que diferentes sistemas de investigação e softwares do MPMG consumam os dados do Data Lake de forma padronizada.
- Implementação de CDC (Change Data Capture): Configuração de ferramentas para capturar alterações em bases transacionais em tempo real, enviando os dados para o Data Lake e evitando cargas full pesadas.
- Gestão de Backfill e Reprocessamento Histórico: Execução de rotinas controladas para reimportar grandes volumes de dados passados sem derrubar o ambiente produtivo ou duplicar registros.
- Evolução de Schema (Schema Evolution): Implementação de lógica para lidar automaticamente com mudanças estruturais nas bases de origem, garantindo que o pipeline não quebre.
- Observabilidade e Monitoramento Proativo: Implementação de métricas, logs e tracing para pipelines e serviços. Criação de dashboards operacionais, monitoramento de falhas, degradação de performance e apoio direto à análise de causa raiz (RCA) nas esteiras de dados.
- Gestão de Segredos e Credenciais: Configuração segura de Kubernetes Secrets ou integração com Vault para que os pipelines acessem bancos de dados em conformidade com as políticas de segurança.

Entregáveis:

a. Documentação Técnica da Solução:

Documento contendo o desenho da arquitetura do fluxo de dados (origem, destino, transformações).

Mapeamento de dependências e pré-requisitos para execução das rotinas.
Dicionário de dados atualizado, descrevendo os novos campos, tipos de dados e regras de negócio aplicadas.

b. Scripts e Rotinas de Automação:

Códigos-fonte das rotinas de extração, transformação e carga (ETL/ELT), versionados no repositório oficial do Órgão.

Arquivos de configuração dos conectores e orquestradores de fluxo.

Scripts de criação e alteração de estruturas de banco de dados (DDL).

c. Evidências de Execução e Monitoramento:

Relatório de carga demonstrando o sucesso da ingestão (volumetria processada, tempo de execução e taxa de sucesso).

Logs de execução comprovando o tratamento de erros e exceções.

Evidência de validação da consistência dos dados entre a origem e o destino.

d. Relatório de Segurança e Acesso:

Documentação das políticas de mascaramento, anonimização e controle de acesso implementadas, em conformidade com a LGPD.

e. Relatório de Transferência de Conhecimento:

Materiais didáticos, gravações de treinamentos ou workshops realizados com a equipe técnica do órgão.

Wiki ou Base de Conhecimento atualizada com os procedimentos operacionais padrão (POP).

Complexidade/HST:

Baixa (<40h):

Sustentação e evolução incremental de pipelines de ingestão (batch/streaming) em ambiente TDP Kubernetes, com execução controlada de backfill/reprocessamentos, ajustes pontuais para lidar com alterações simples de esquema (schema evolution) e produção de documentação técnica e evidências básicas de execução (relatório de carga, logs e validação de consistência origem–destino).

Média (de 40h até 80h):

Desenvolvimento e refatorização de pipelines de ingestão, incluindo implementação/configuração de CDC (Change Data Capture), construção de APIs de dados para consumo padronizado por sistemas do MPMG, gestão segura de segredos/credenciais (Kubernetes Secrets ou integração com Vault) e reforço de observabilidade (métricas, logs, tracing e dashboards operacionais), com entrega de scripts/rotinas versionadas e relatórios de execução, segurança e acesso.

Alta (>80h):

Execução end-to-end da modernização e sustentação de pipelines de ingestão com apoio à migração do legado (Hadoop/HDP) para arquitectura cloud-native em TDP Kubernetes, abrangendo CDC em tempo real, reprocessamentos históricos de grande volume (backfill) com controlo de risco, lógica robusta de schema evolution para evitar quebras, monitorização proactiva com apoio à análise de causa raiz (RCA) e reforço de conformidade de segurança (segredos/credenciais e políticas de acesso/mascaramento/anonimização), incluindo documentação completa e transferência de conhecimento (materiais, workshops e wiki/POP).

Item 1.2 - Governança Técnica e Qualidade de Dados (Data Quality)

Descrição das Atividades:

- Gestão de Expurgo Seguro (LGPD - Direito ao Esquecimento): Criação de rotinas automatizadas para deletar ou anonimizar dados de cidadãos que solicitaram exclusão, varrendo todas as camadas do Datalake (Raw, Trusted, Refined).
- Monitoramento de Linhagem (Data Lineage) Ponta a Ponta: Mapeamento visual no Apache Atlas desde a origem (banco SQL) até o painel final (Dashboard), permitindo análise de impacto ("Se eu mudar essa tabela, quais painéis param de funcionar?").
- Master Data Management (MDM) Básico: Implementação de regras de "De-Para" e unificação de cadastros duplicados (ex: unificar "João Silva" e "Joao da Silva" baseado no CPF).
- Auditoria de Acessos: Configuração de relatórios mensais do Apache Ranger para identificar quem acessou dados sensíveis e tentativas de acesso negado.

Entregáveis:

a. **Entregável Principal: Relatório de Conformidade de Expurgo (Disposal Certificate).**

O que contém: Um documento log (auditável) listando os IDs anônimos que foram processados, a data da execução e o *status* da remoção em cada camada (ex: "ID 12345: Removido da Raw [OK], Removido da Trusted [OK], Anonimizado na Refined [OK]").

Evidência Técnica: Log de execução do *Job* de expurgo demonstrando a redução volumétrica ou a alteração dos registros (print do antes/depois amostral).

Complexidade/HST:

Baixa (<40h):

Execução assistida e validada de rotina de expurgo ou anonimização de dados em conjuntos de dados delimitados, com geração de relatório auditável contendo o registo dos IDs processados, data da execução e estado da remoção nas camadas do Data Lake.

Média (de 40h até 80h):
Implementação e execução estruturada de processos de expurgo seguro, com varrimento controlado das camadas Raw, Trusted e Refined, produção de relatório de conformidade detalhado e recolha de evidências técnicas demonstrando a efectividade da remoção ou anonimização dos dados.

Alta (>80h):
Desenho, execução e validação completa de processos de expurgo seguro em conformidade com exigências de governação e protecção de dados, incluindo rastreabilidade integral dos identificadores processados, evidências técnicas robustas e documentação auditável apta a fiscalização interna e externa.

→ **Para a Atividade: Monitoramento de Linhagem (Data Lineage)**

b. Entregável Principal: Mapa de Rastreabilidade e Análise de Impacto.

O que contém: exportação gráfica ou documental do fluxo de dados crítico (do Sistema de Origem até o Dashboard do Promotor). Deve incluir uma **Matriz de Dependência** (lista das quais painéis/relatórios dependem de quais tabelas-fonte).

Evidência Técnica: Atualização do catálogo no Apache Atlas refletindo as novas tabelas e transformações criadas no período.

Complexidade/HST:

Baixa (<40h):

Execução de rotinas padronizadas e previamente estabelecidas de governança técnica e qualidade de dados, incluindo a aplicação de expurgo seguro ou anonimização em conjuntos de dados específicos, validação básica de linhagem de dados já cadastrada e geração de relatórios operacionais e de conformidade.

Escopo típico:

- Operação assistida de job de expurgo seguro (LGPD) em bases delimitadas.
- Conferência da remoção/anonimização nas camadas Raw, Trusted e Refined.
- Atualização pontual do catálogo de dados (Apache Atlas) para novas tabelas simples.

- Emissão do Relatório de Conformidade de Expurgo com logs de execução e evidência amostral.

Média (de 40h até 80h):

Implementação, ajuste e documentação de mecanismos de governança técnica, abrangendo automação de rotinas de expurgo seguro, mapeamento de linhagem ponta a ponta em fluxos críticos de dados, aplicação de regras básicas de Master Data Management (MDM) e consolidação de relatórios de auditoria de acessos.

Escopo típico:

- Desenvolvimento ou adaptação de rotinas automatizadas de expurgo/anonimização em múltiplas camadas do Datalake.
- Mapeamento visual de Data Lineage desde a origem (banco relacional) até dashboards analíticos.
- Construção de matriz de dependência entre tabelas-fonte e painéis/relatórios.
- Implementação de regras de "De-Para" e unificação de cadastros duplicados.
- Atualização estruturada do catálogo no Apache Atlas e geração de evidências auditáveis.

Alta (>80h):

Desenho, implementação e governança avançada de processos integrados de qualidade e conformidade de dados, com enfoque em LGPD, rastreabilidade ponta a ponta, análise de impacto sistêmico e auditoria contínua de acessos a dados sensíveis, envolvendo múltiplos sistemas, equipes e domínios de dados.

Escopo típico:

- Arquitetura completa de expurgo seguro com varredura transversal em ambientes heterogêneos e grandes volumes de dados.
- Implementação consolidada de Data Lineage ponta a ponta com análise avançada de impacto ("change impact analysis").
- Definição e governança de MDM básico transversal, com regras de unificação baseadas em identificadores sensíveis (ex.: CPF).
- Configuração e análise recorrente de relatórios de auditoria do Apache Ranger para dados sensíveis e acessos negados.
- Produção de Relatórios de Conformidade robustos, com evidências técnicas completas, logs auditáveis e documentação institucional.

→ **Para a Atividade: Master Data Management (MDM) Básico**

c. Entregável Principal: Relatório de Saneamento e Unificação de Entidades.

O que contém: Estatísticas de qualidade demonstrando a eficácia da higienização. Exemplo: "*Foram identificados 5.000 cadastros duplicados; após o processamento, restaram 3.200 Registros Dourados (Golden Records) únicos.*"

Evidência Técnica: Disponibilização da Tabela de "De-Para" (Tabela de Referência) no Datalake, documentando quais IDs antigos apontam para qual ID unificado novo.

Complexidade/HST:

Baixa (<40h):

Execução controlada de saneamento básico de cadastros, com identificação de duplicidades evidentes e aplicação de regras simples de unificação previamente definidas, resultando na geração assistida de registros consolidados e tabela de referência "De-Para".

Escopo típico:

- Identificação de registros duplicados com base em chaves diretas e previamente validadas.
- Aplicação de regras simples de unificação (ex.: correspondência direta por CPF).
- Geração inicial de estatísticas descritivas de qualidade dos dados.
- Disponibilização da Tabela "De-Para" no Datalake como evidência técnica.
- Emissão do Relatório de Saneamento e Unificação com enfoque operacional.

Média (de 40h até 80h):

Implementação estruturada de processos de Master Data Management em nível básico, incluindo higienização de dados, aplicação de regras de correspondência (matching), consolidação de múltiplas fontes e produção de indicadores de qualidade demonstrando a efetividade do saneamento.

Escopo típico:

- Identificação massiva de cadastros duplicados oriundos de diferentes sistemas ou bases.
- Definição e aplicação de regras de "De-Para" e critérios de unificação mais elaborados.
- Geração de Registros Dourados (Golden Records) com rastreabilidade completa.
- Produção de estatísticas comparativas pré e pós-processamento (ex.: cadastros identificados versus consolidados).
- Publicação da Tabela de Referência no Datalake, documentando o mapeamento entre IDs antigos e IDs unificados.

Alta (>80h):
Desenho e execução de um processo abrangente de Master Data Management básico, com saneamento em larga escala, unificação de entidades críticas, rastreabilidade integral dos identificadores e produção de evidências auditáveis para suporte à governança e à qualidade de dados institucionais.

Escopo típico:

- Consolidação de grandes volumes de cadastros duplicados, heterogêneos e historicamente inconsistentes.
- Definição de golden records com aplicação de regras hierarquizadas e resolução de conflitos entre atributos.
- Produção de métricas robustas de qualidade, demonstrando a efetividade do saneamento de dados.
- Disponibilização estruturada e governada da Tabela "De-Para" no Datalake como ativo institucional.
- Elaboração de Relatório de Saneamento e Unificação com caráter técnico-gerencial e apto à auditoria e fiscalização.

→ **Para a Atividade: Auditoria de Acessos**

d. **Entregável Principal: Relatório Gerencial de Auditoria de Acessos**.

O que contém: Um resumo executivo mensal destacando:

- Quem acessou os dados mais sensíveis/sigilosos.
- Tentativas de acesso bloqueadas (quem tentou ver o que não podia).
- Usuários com permissões excessivas (recomendação de revogação).

Complexidade/HST:

Baixa (<40h):
Consolidação mensal de informações de auditoria de acessos a dados sensíveis, a partir de logs existentes, com foco na identificação descritiva de acessos realizados e tentativas de acesso bloqueadas, resultando na emissão de relatório gerencial de caráter informativo.

Média (de 40h até 80h):
Análise estruturada e recorrente dos acessos a dados sensíveis e sigilosos, incluindo a identificação de tentativas de acesso indevido e de usuários com permissões excessivas, com

elaboração de relatório gerencial contendo recomendações técnicas para ajuste ou revogação de acessos.

Alta (>80h):

Implantação e operação de um processo contínuo de auditoria e governança de acessos, com análise aprofundada de padrões de uso, detecção de riscos associados a privilégios excessivos e produção de relatórios estratégicos para apoio à gestão de segurança da informação e conformidade institucional.

Item 1.3 - Otimização de Performance em Kubernetes (FinOps/Tuning)

Descrição das Atividades:

- **Gestão de "Small Files" (Arquivos Pequenos):** Rotinas de manutenção para compactar milhões de arquivos pequenos (que matam a performance do HDFS/Ozone) em arquivos maiores e otimizados.
- **Definição de Resource Quotas por Área:** Configuração de *Namespaces* no Kubernetes para garantir que a "Secretaria de Saúde" não consuma toda a memória do cluster, deixando a "Segurança Pública" sem recursos.
- **Estratégia de Cache Distribuído:** Configuração de cache em memória (Alluxio ou Spark Cache) para datasets muito acessados, reduzindo o I/O em disco e acelerando painéis.
- **Análise de Custo por Workload:** Relatório mensal de atribuição de custo ("O pipeline X consumiu 30% da CPU do cluster este mês").

Entregáveis:

→ **Para a Atividade: Gestão de "Small Files" (Arquivos Pequenos)**

a. Entregável Principal: Relatório de Sanidade do Armazenamento (Storage Health Check).

O que contém: Um comparativo mensal de "Antes e Depois" da compactação. Ex: "*Redução de 5 milhões de arquivos pequenos para 50 mil arquivos otimizados; Redução da pressão de memória no NameNode em 20%.*"

Evidência Técnica: Gráficos de monitoramento demonstrando a estabilidade da RPC (Remote Procedure Call) do sistema de arquivos e tempo médio de leitura dos blocos de dados.

→ **Para a Atividade: Definição de Resource Quotas por Área**

b. **Entregável Principal: Mapa de Governança de Recursos Computacionais (Resource Policy Map).**

O que contém: Documento listando a distribuição aprovada de vCPU e RAM por Secretaria/Departamento (Namespace). Deve detalhar os limites (Hard Limits) e as garantias (Requests) configuradas.

Evidência Técnica: Exportação da configuração dos *Namespaces* do Kubernetes demonstrando as quotas aplicadas e relatório de incidentes onde "pods" foram barrados por tentarem exceder a cota (prova de funcionamento do bloqueio).

→ **Para a Atividade: Estratégia de Cache Distribuído**

c. Entregável Principal: Relatório de Aceleração de Dados (Data Acceleration Report).

O que contém: Lista dos *datasets* críticos que foram colocados em memória (Hot Data) e o ganho de performance obtido. Ex: "*Painel de Monitoramento de Leitos: Tempo de carga reduzido de 40s para 3s após implementação de cache Alluxio.*"

Evidência Técnica: Estatísticas de *Cache Hit Ratio* (taxa de acerto do cache) comprovando que a maioria das consultas foi respondida pela memória e não pelo disco.

→ **Para a Atividade: Análise de Custo por Workload (FinOps)**

d. Entregável Principal: Demonstrativo Gerencial de Alocação de Custos (Showback Report).

O que contém: Gráficos de pizza e tabelas atribuindo o consumo do cluster aos centros de custo ou projetos. Deve incluir recomendações de economia (ex: "*O processo X está superdimensionado, sugerimos reduzir em 50% a memória alocada*").

Evidência Técnica: Relatório extraído da ferramenta de monitoramento (Prometheus/Grafana ou Kubecost) detalhando o consumo de *Core-Hours* e *Memory-GiB-Hours* por *Label* de projeto.

Complexidade/HST:

Baixa (<40h):

Execução de ajustes pontuais de performance e análise operacional do ambiente Kubernetes, com foco em melhorias incrementais de armazenamento, controle básico de recursos e geração de relatórios descritivos de utilização e desempenho.

Escopo típico:

- Execução assistida de rotinas de compactação de *small files* em ambientes HDFS/Ozone.
- Geração de comparativos simples de "antes e depois" da compactação.
- Extração de métricas básicas de desempenho do sistema de arquivos.
- Consolidação de informações iniciais de consumo de recursos por workload.
- Elaboração de relatórios gerenciais com caráter informativo.

Média (de 40h até 80h):

Implementação estruturada de ações de otimização de performance e governança de recursos em Kubernetes, integrando armazenamento, cache distribuído e análise de consumo computacional, com foco em eficiência operacional e suporte à decisão gerencial.

Escopo típico:

- Implementação sistemática de políticas de compactação e sanidade de armazenamento (*Storage Health Check*).
- Definição e configuração de *Resource Quotas* por área/namespace, com limites e garantias de vCPU e RAM.
- Configuração de estratégias de cache distribuído (Alluxio ou Spark Cache) para datasets críticos.
- Geração de relatórios de aceleração de dados e ganhos de performance observados.
- Emissão de demonstrativos de alocação de custos (*Showback*) com análise comparativa mensal.

Alta (>80h):

Desenho, implementação e operação contínua de uma estratégia avançada de otimização de performance e FinOps em ambientes Kubernetes, integrando governança de recursos, aceleração de dados, controle de custos e evidências técnicas robustas para gestão e auditoria.

Escopo típico:

- Redesenho estrutural da estratégia de armazenamento, com redução massiva de *small files* e mitigação de impactos no NameNode.
- Implantação completa de governança de recursos computacionais por área, com monitoramento de violações de cota e comprovação de bloqueios efetivos.

- Implementação avançada de cache distribuído, garantindo tempos de resposta significativamente reduzidos para painéis críticos.
- Análise detalhada e recorrente de consumo de CPU e memória por workload, projeto ou centro de custo.
- Produção de relatórios gerenciais estratégicos de FinOps, com recomendações explícitas de economia, redimensionamento e otimização de recursos.

Item 1.4 - Administração de Sistemas, Infraestrutura e Segurança RHEL (Legacy & Foundation)

Foco na base do cluster: Sistema Operacional, Rede, Armazenamento e Runtimes de Contêineres.

Descrição das Atividades:

- Gestão do Sistema Operacional RHEL: Instalação, configuração, manutenção e aplicação de hardening no Red Hat Enterprise Linux (RHEL) conforme boas práticas e políticas institucionais. Inclui o gerenciamento profundo de usuários, permissões, processos e serviços via systemd.
- Serviços de Infraestrutura Legada e Rede: Instalação e configuração de serviços nativos (Apache, Samba, NFS, FTP, DHCP, OpenLDAP, MySQL, Postfix). Engloba a solução de falhas de serviços, análise avançada de logs (journalctl, rsyslog), configuração de rede e homologação de novas soluções integradas ao ambiente.
- Segurança Avançada e Conformidade: Aplicação contínua de atualizações (releases) e patches de segurança para mitigação de vulnerabilidades (CVEs). Gerenciamento restrito de segurança via SELinux (troubleshooting e tuning) e configuração de auditoria avançada (auditd) para rastreabilidade, apoiando a conformidade com a LGPD e políticas internas.
- Administração de Armazenamento: Configuração de acesso a armazenamento em rede (SAN - Storage Area Network), gestão avançada de volumes lógicos (LVM) e sistemas de arquivos corporativos.
- Sustentação de Contêineres no Nível do SO: Suporte técnico à operação de nós Kubernetes sobre RHEL, com diagnóstico de problemas entre o sistema operacional e os workloads de dados. Inclui troubleshooting de container runtimes (CRI-O, Podman), análise de consumo de recursos por pods e integrações de orquestração.
- Automação Base e Troubleshooting Avançado: Desenvolvimento e manutenção de scripts (bash, etc.) para automação de rotinas de infraestrutura. Execução de troubleshooting de inicialização e análise avançada de incidentes (RCA de infraestrutura).

- Alta Disponibilidade e Tuning: Configuração de recursos de alta disponibilidade do sistema operacional. Identificação de gargalos no ambiente base, realizando mapeamento e ajuste fino (tuning) de performance dos serviços executados sobre RHEL.

Principais Entregáveis:

a. **Relatório de Hardening e Conformidade:** Documento evidenciando a aplicação de travas de segurança e correção de vulnerabilidades no SO.
b. **Mapa de Configuração de Infraestrutura:** Documentação atualizada de serviços de rede, regras de firewall local e topologia de discos (LVM/SAN).
c. **Laudo de RCA (Root Cause Analysis):** Relatório de análise de causa raiz para incidentes críticos que tenham origem na camada de infraestrutura.

Complexidade/HST:

Baixa (<40h):

Execução de rotinas operacionais de administração de sistemas RHEL, incluindo manutenção básica do sistema operativo, suporte a serviços de infraestrutura legada, aplicação de configurações padrão de segurança e produção de documentação técnica atualizada.

Escopo típico:

- Instalação e configuração básica do Red Hat Enterprise Linux (RHEL).
- Manutenção de serviços nativos de infraestrutura e rede em ambiente controlado.
- Aplicação pontual de hardening e correções de segurança previamente definidas.
- Suporte operacional a nós Kubernetes no nível do sistema operativo.
- Elaboração de mapas de configuração de infraestrutura e relatórios informativos.

Média (de 40h até 80h):

Administração estruturada da camada base de infraestrutura, com foco em segurança, disponibilidade e desempenho do sistema operativo RHEL, incluindo troubleshooting de serviços críticos, automação de rotinas e geração de relatórios técnicos de conformidade.

Escopo típico:

- Gestão contínua do sistema operativo RHEL com ênfase em usuários, permissões e serviços (systemd).
- Configuração e suporte avançado a serviços de infraestrutura e rede.
- Aplicação sistemática de patches, mitigação de vulnerabilidades (CVEs) e configuração de SELinux.
- Administração de armazenamento corporativo (SAN, LVM e sistemas de arquivos).

- Desenvolvimento de scripts de automação e execução de análises de causa raiz (RCA).
- Emissão de Relatórios de Hardening, Conformidade e Laudos de RCA.

Alta (>80h):

Desenho, operação e sustentação avançada da base de infraestrutura RHEL que suporta o cluster de dados, integrando segurança, alta disponibilidade, performance e conformidade, com atuação direta em incidentes críticos e otimização contínua do ambiente.

Escopo típico:

- Administração profunda do sistema operativo RHEL em ambientes críticos de produção.
- Configuração avançada de segurança (SELinux, auditd) com rastreabilidade completa e aderência à LGPD e políticas institucionais.
- Sustentação especializada de nós Kubernetes no nível do SO, incluindo troubleshooting de container runtimes (CRI-O, Podman).
- Implementação de alta disponibilidade e tuning fino de performance da infraestrutura base.
- Realização de análises completas de causa raiz para incidentes críticos de infraestrutura.
- Produção de documentação técnica robusta, auditável e orientada à gestão estratégica da infraestrutura.

GRUPO 2: OPERAÇÃO ASSISTIDA ESPECIALIZADA EM CIÊNCIA DE DADOS

Item 2.1 - Modelagem Analítica e Estatística (Data Science)

Foco na descoberta de padrões e respostas de negócio.

Descrição das Atividades:

- Formular e prototipar algoritmos e modelos de Machine Learning e Deep Learning;
- Criar modelos para Reconhecimento de Padrões em Big Data;
- Executar tarefas de Processamento de Linguagem Natural (NLP);
- Coletar, preparar e persistir os dados necessários para as análises.

Entregáveis:

a. Notebooks (Jupyter/Zeppelin) e/ou scripts em linguagem Python executáveis;

b. Relatório de performance do modelo, apresentando as métricas de assertividade do modelo (acurácia, precisão, erro médio) obtidas nos testes.;

c. Documentação da metodologia de treinamento descrevendo as técnicas e features utilizadas, orientações para retreinamento e para uso dos modelos.**Complexidade/HST:**

Baixa (<40h):

Desenvolvimento de análises estatísticas exploratórias e modelos analíticos simples, com foco na compreensão descritiva dos dados e validação inicial de hipóteses, apoiando a geração de evidências para decisões de negócio.

Escopo típico:

- Planeamento estatístico básico de análises ou testes comparativos simples.
- Construção de notebooks executáveis (Jupyter/Zeppelin) com análises descritivas ou modelos de regressão simples.
- Elaboração de documentação metodológica sucinta.
- Geração de relatório básico de performance do modelo com métricas padrão.

Média (de 40h até 80h):

Implementação estruturada de modelos analíticos e estatísticos para identificação de padrões, incluindo testes controlados, engenharia de variáveis e seleção sistemática de modelos, com avaliação comparativa de desempenho.

Escopo típico:

- Desenho e execução de Testes A/B para avaliação de políticas públicas ou intervenções.
- Engenharia de *features* mais elaboradas, incluindo variáveis temporais e agregações móveis.
- Utilização de técnicas de AutoML para seleção de modelos *champion/challenger*.
- Construção de notebooks reprodutíveis e documentação detalhada da metodologia.
- Emissão de relatórios de performance com métricas como Acurácia e F1-Score.

Alta (>80h):

Desenvolvimento de soluções avançadas de ciência de dados para descoberta profunda de padrões e suporte a decisões estratégicas, integrando engenharia complexa de variáveis, privacidade por design e avaliação rigorosa de modelos.

Escopo típico:

- Planejamento e execução de análises estatísticas e modelos avançados de aprendizagem automática.

- Geração de dados sintéticos para preservação da privacidade e conformidade com princípios de *privacy by design*.
- Engenharia complexa de *features*, incluindo grafos e séries temporais avançadas.
- Seleção e comparação extensiva de modelos através de AutoML.
- Produção completa de notebooks executáveis, documentação metodológica robusta e relatórios detalhados de performance do modelo.

Item 2.2 - MLOps e Produtização de Modelos (Deploy)

Descrição das Atividades:

- Gestão de Feature Store: Criação de um repositório central de variáveis (Features) para que diferentes cientistas não precisem recalcular a mesma métrica (ex: "Renda Presumida") repetidamente.
- Versionamento de Modelos (Model Registry)**:** Implementação de controle de versão (MLflow) para garantir que seja possível reverter (rollback) para a versão anterior do modelo em caso de falha.
- Testes de Carga na API de Inferência: Simulação de milhares de requisições simultâneas na API do modelo para garantir que o Kubernetes escale os Pods corretamente antes de liberar para o público.
- Monitoramento de Viés (Bias): Auditoria algorítmica para verificar se o modelo não está discriminando grupos específicos (ex: gênero, raça ou região geográfica).

Entregáveis:

a. Pacote de software contendo o modelo treinado pronto para uso (API ou rotina Batch).
b. Manual de uso ou integração para que outros sistemas possam consumir o resultado do modelo.
c. Monitoramento de saúde do modelo.

Complexidade/HST:

Baixa (<40h):

Produtização básica de modelos analíticos, com disponibilização de endpoint de inferência em ambiente controlado, versionamento inicial do modelo e monitoramento operacional simples de disponibilidade e saúde.

Escopo típico:

- Publicação inicial do modelo como API de inferência em ambiente Kubernetes.
- Registro básico do modelo em *Model Registry* para controle de versão.
- Monitoramento elementar de disponibilidade do endpoint (*health check*).
- Geração de evidências funcionais de deploy do modelo.

Média (de 40h até 80h):

Implantação estruturada de práticas de MLOps, abrangendo organização de variáveis reutilizáveis, controle de versões de modelos e validação de escalabilidade da infraestrutura de inferência.

Escopo típico:

- Implementação de *Feature Store* para reutilização padronizada de variáveis.
- Gestão de versionamento de modelos com possibilidade de *rollback*.
- Execução de testes de carga na API de inferência para validação de escalabilidade dos pods no Kubernetes.
- Monitoramento de saúde do modelo e do serviço de inferência.

Alta (>80h):

Desenho e operação de um pipeline completo de MLOps para modelos em produção, integrando governança de features, versionamento robusto, escalabilidade, monitoramento avançado e auditoria algorítmica.

Escopo típico:

- Estruturação avançada de *Feature Store* compartilhada entre múltiplos modelos e cientistas de dados.
- Gestão completa de ciclo de vida de modelos com *Model Registry* e estratégias *champion/challenger*.
- Testes intensivos de carga e resiliência da API de inferência em ambiente Kubernetes.
- Monitoramento contínuo de viés (*bias*) e comportamento do modelo em produção.
- Disponibilização de API de modelo em produção com monitoramento ativo de saúde, desempenho e conformidade.

Item 2.3 - Visualização de Dados e Storytelling

Descrição das Atividades:

- Construção de Painéis Gerenciais e Dashboards nas soluções indicadas pela Contratante.
- Criação de visões semânticas (camada de consumo) para facilitar o *Self-Service BI*.
- Adições focadas na experiência do usuário (UX):
- Otimização de Queries SQL para Dashboards: Refatoração de consultas pesadas que alimentam os painéis, criando, por exemplo, tabelas agregadas.

- Desenvolvimento de Alertas Inteligentes: Configuração de gatilhos (Webhooks/E-mail) que avisam quando um indicador sai da normalidade
- Adaptação para Dispositivos Móveis: Ajuste de layout responsivo para que os painéis gerenciais possam ser consumidos em tablets e celulares.

Entregáveis:

1. Construção de Painéis Gerenciais e Dashboards
2. Desenvolvimento de Alertas Inteligentes
3. Criação de visões semânticas e/ou otimização de queries SQL

Complexidade/HST:

Baixa (<40h):

Resumo da atividade:
Construção de painéis gerenciais e dashboards a partir de dados já disponíveis, com criação de visões semânticas básicas para consumo analítico, ajustes pontuais de performance e disponibilização de artefatos documentais de metodologia e resultados.

Escopo típico:

- Construção de Painéis Gerenciais e Dashboards nas soluções indicadas pela Contratante.
- Construção de visões semânticas simples para facilitar o consumo dos dados.
- Ajustes pontuais de consultas SQL para melhoria de tempo de resposta.
- Elaboração da memória de cálculo e metodologia aplicada.
- Geração de relatório básico de performance do modelo, quando aplicável.

Média (de 40h até 80h):

Desenvolvimento estruturado de soluções analíticas com foco em self-service BI e experiência do usuário, incluindo otimização sistemática de consultas, criação de camadas de consumo, configuração de alertas inteligentes e análise comparativa de desempenho.

Escopo típico:

- Desenvolvimento de dashboards gerenciais complexos, com múltiplas métricas e filtros.
- Criação estruturada de camada semântica para uso autônomo por usuários finais.
- Refatoração de consultas SQL pesadas, com uso de tabelas agregadas para garantir carregamento rápido dos painéis.
- Implementação de alertas automáticos (e-mail/webhook) para desvios de indicadores.

- Elaboração de relatórios de análise exploratória e de performance, com comparação entre versões do modelo ou solução.

Alta (>80h):

Resumo da atividade:
Implementação completa de soluções analíticas produtizadas, com forte orientação à experiência do usuário, alto desempenho, automação de alertas e disponibilização de modelos estatísticos ou algorítmicos prontos para integração sistêmica.

Escopo típico:

- Arquitetura e desenvolvimento avançado de painéis gerenciais e dashboards críticos.
- Criação robusta de camada semântica para self-service BI em larga escala.
- Otimização avançada de performance, garantindo tempos de resposta inferiores a 3 segundos.
- Implementação de alertas inteligentes acionados por desvios de comportamento dos indicadores.
- Adaptação responsiva dos painéis para dispositivos móveis utilizados por gestores e autoridades.
- Entrega de artefatos de produtização, incluindo pacotes de software (API ou batch) e manuais de uso ou integração.
- Produção completa de documentação metodológica, relatórios de performance e evidências de evolução da solução.