APENSO I - CADERNO DE ESPECIFICAÇÕES DETALHADAS DO OBJETO

1. OBJETIVO DO CADERNO DE ESPECIFICAÇÃO TÉCNICA

Este Caderno de Especificações Técnicas descreve os requisitos técnicos mínimos e obrigatórios da solução a ser contratada, bem como os critérios objetivos de aceitação, operação e suporte, aplicáveis a todos os itens do objeto, no que couber.

As disposições aqui contidas possuem efeito normativo vinculante para fins de julgamento das propostas, execução contratual, fiscalização, medição, glosa e aceite.

O objeto desta contratação é constituído por dois eixos estruturantes: a Plataforma Tecnológica (Software) e os Serviços Especializados (Operação Assistida), que juntos compõem a solução de modernização do ambiente de dados do MPMG:

Item	Descrição Resumida	Unidade	Estimativa (12 meses)
01	Subscrição TDP Kubernetes. Direito de Uso, atualização e suporte para clusters de Big Data de até 10 nós de processamento.	ANO	1
02	Operação Assistida - Engenharia de Dados. Serviços especializados para sustentação, evolução e modernização do ecossistema de dados da organização, abrangendo ingestão, processamento e governança de dados, desenvolvimento e otimização de pipelines, além da integração, migração e adoção de tecnologias correlatas que suportam o ciclo de vida completo dos dados.	HST	5.389
03	Operação Assistida - Análise e Ciência de Dados. Serviços especializados para exploração, análise e modelagem de dados, abrangendo desenvolvimento e implantação de modelos estatísticos, analíticos, preditivos e de inteligência artificial, criação de dashboards e visualizações, bem como evolução contínua de soluções analíticas para apoio à tomada de decisão.	HST	4.769

2. ITEM 1: PLATAFORMA DE DADOS CORPORATIVA (SOFTWARE)

Fornecimento de subscrição de software da solução padronizada Tecnisys Data Platform (TDP), na edição Enterprise, baseada mandatoriamente em arquitetura Cloud Native (contêineres), orquestrada por Kubernetes, contemplando direito de uso, garantia de atualização tecnológica e suporte técnico nível 3.

A plataforma deve atender aos seguintes requisitos funcionais e não funcionais mandatórios:

2.1 Arquitetura e Orquestração Geral Cloud-Native e Princípios Arquiteturais Mandatórios

- A solução deverá consistir em Plataforma de Dados Corporativa baseada em arquitetura cloud-native, executada integralmente em contêineres, com orquestração obrigatória por Kubernetes, admitindo implantação.
- Orquestração Nativa: A solução deve operar sobre containers Docker orquestrados por Kubernetes, permitindo a gestão automatizada do ciclo de vida dos serviços (instalação, atualização e otimização).
- Desacoplamento de Recursos: Suporte nativo à arquitetura de Computação e Armazenamento Desagregados (Disaggregated Compute and Storage), permitindo ao MPMG escalar a capacidade de processamento (vCPUs) independentemente da capacidade de armazenamento (TB).
- Soberania e Infraestrutura: Capacidade de instalação Full Stack em infraestrutura local (On-Premise) ou Nuvem Privada do MPMG, garantindo que o plano de controle e os dados residam sob jurisdição brasileira.
- Contêineres + Kubernetes: Todos os serviços do ecossistema executam como pods em namespaces segregados (produção, homologação, desenvolvimento), com RBAC, quotas e limites de recursos para isolamento e previsibilidade.
- Automação por Operators/Deployments: Provisionamento, upgrade, auto-healing, backup/restore e configuração dos componentes realizados por Operators e Deployments, reduzindo intervenção manual e risco operacional.
- Elasticidade: Autoscaling em nível de workload (HPA/VPA), com possibilidade de expansão controlada de nós (cluster) conforme demanda institucional.
- Desacoplamento compute-storage: Separação entre processamento e armazenamento utilizando Object Storage compatível com S3 (ex.: Apache Ozone/MinIO), reduzindo ociosidade e TCO.

- Rede e desempenho: CNI compatível (Calico/Cilium), afinidade/anti-afinidade, taints/tolerations, priority classes e QoS para isolar cargas críticas (ex.: investigações sensíveis, janelas de ingestão massiva).
- Observabilidade e SRE: Monitoramento, Logs, métricas , traces integrados e auditoria do ambiente (ex.: EFK/ELK, Prometheus/Grafana, OpenTelemetry), com SLOs e alertas para incidentes de disponibilidade e performance.

2.2 Stack tecnológica suportada (componentes mandatórios suportados)

A subscrição da plataforma deverá integrar e suportar oficialmente os seguintes componentes Open Source, fornecendo suporte oficial Enterprise, com correções, atualizações e atendimento especializado, para os seguintes componentes mínimos:

- Armazenamento distribuído: Apache Ozone — armazenamento de objetos nativo para ambientes containerizados, alta escalabilidade e throughput - elevado e eficiência no gerenciamento de grande volume de arquivos.
- Computação distribuída: Apache Spark executando sobre Kubernetes, sem dependência de YARN, com escalonamento e isolamento nativos do cluster para workloads de ETL/ELT, aprendizado de máquina (machine learning) e processamento de dados em larga escala.
- Arquitetura Lakehouse: Apache Iceberg para gerenciamento de tabelas no data lake com propriedades transacionais, evolução de schema, particionamento avançado, time travel e operações ACID, garantindo consistência e interoperabilidade entre múltiplos motores.
- Consultas federadas e virtualização de dados: Trino (MPP) para federação de consultas sobre fontes heterogêneas (data lake, bancos relacionais, NoSQL e arquivos), com otimizações de execução, predicate/aggregation pushdown e conectores extensíveis, reduzindo a necessidade de replicação e movimentação de dados.
- Análise de dados em tempo real: ClickHouse (banco colunar OLAP) para altíssima performance em consultas analíticas e agregações com baixa latência e alta concorrência, suportando particionamento, replicação e distribuição de dados. ClickHouse como banco analítico colunar para agregações e consultas de alta concorrência com latência muito baixa, habilitando painéis interativos e exploração analítica em grandes volumes, com mecanismos de particionamento e distribuição escalável.
- Streaming e mensageria: Apache Kafka como barramento de eventos para publish/subscribe, retenção e reprocessamento, promovendo desacoplamento

entre produtores e consumidores, e sustentando pipelines real-time e integrações assíncronas.

- Barramento de dados: Apache Kafka como barramento de eventos para pub/sub, retenção, reprocessamento e desacoplamento entre produtores e consumidores, sustentando pipelines de streaming em larga escala.
- Gestão de fluxos de dados: Apache NiFi para orquestração, transformação e governança de fluxos de dados (coleta, roteamento, priorização, enriquecimento e rastreabilidade), promovendo integração confiável entre sistemas heterogêneos. Apache NiFi para desenho e operação de fluxos com roteamento, transformação, priorização, enriquecimento, auditoria e rastreabilidade, facilitando integrações robustas entre sistemas e fontes diversas.
- Orquestração de pipelines: Apache Airflow para agendamento e automação de pipelines, com DAGs declarativos, dependências explícitas, retentativas, observabilidade e integração com múltiplos sistemas para operações de dados. Apache Airflow para coordenar pipelines de dados, dependências explícitas, retentativas e observabilidade, padronizando execução e governança operacional.
- BI e visualização de dados: Apache Superset para dashboards e exploração por demanda (ad hoc), com conectividade abrangente, governança de acesso por perfis e capacidade de democratizar análises para áreas de negócio. Apache Superset para dashboards e análises ad hoc, com controle de acesso por perfis e ampla compatibilidade com engines SQL, acelerando a democratização de dados com governança.
- Ambiente interativo e colaborativo para exploração e desenvolvimento: JupyterHub para notebooks multiusuário, promovendo experimentação, prototipação e colaboração em ciência de dados e engenharia, com controle de acesso e gestão de ambientes de execução. JupyterHub como plataforma multiusuário de notebooks para experimentação, exploração, prototipação e colaboração, com gestão de ambientes e controle de acesso.
- Segurança avançada: Apache Ranger para gestão de políticas de segurança (RBAC/ABAC), auditoria e mascaramento dinâmico de dados (dynamic data masking), apoiando requisitos de conformidade com a LGPD. Apache Ranger para políticas centralizadas (RBAC/ABAC), trilhas de auditoria, mascaramento dinâmico de dados sensíveis em tempo de consulta e controles de autorização aplicados de forma consistente nas camadas de acesso aos dados.
- Governança de dados e compliance: OpenMetadata para catálogo de dados, glossário/taxonomias, classificação de ativos, linhagem (data lineage) e

observabilidade de dados (qualidade, profiling e SLAs), centralizando metadados técnicos e de negócio. OpenMetadata para catálogo, glossário, classificação, linhagem e observabilidade de dados, consolidando metadados técnicos e de negócio para rastreabilidade e confiança no dado.

Comprovação de capacidade técnica de suporte:
Para fins de habilitação e garantia da qualidade do suporte (aceitabilidade):

- A licitante deverá apresentar declaração formal de que possui estrutura de suporte técnico no Brasil, em língua portuguesa (Brasil) falada e escrita.
- Caso a licitante não seja a própria fabricante do software, deverá apresentar a Carta de Credenciamento Oficial, comprovando que é parceira autorizada e que possui contrato de suporte "Back-to-Back" ativo com o fabricante, assegurando que o MPMG terá acesso à atualização de versões e ao suporte de engenharia (N3) oficial.

2.3 Segurança da informação, governança e soberania

- Soberania e topologia: Execução on-premises ou em nuvem soberana nacional, com opção de operação air-gapped (isolada), assegurando que dados e metadados sensíveis não trafeguem para jurisdições estrangeiras.
- Criptografia: Em trânsito (mTLS/TLS 1.2+) e em repouso (criptografia no storage e em bancos de metadados). Integração com KMS/HSM para gerenciamento de chaves.
- Identidade e acesso: SSO via SAML/OIDC, MFA, RBAC/ABAC no Ranger, IAM centralizado (integração com AD/LDAP).
- Políticas e auditoria: Trilha de auditoria fim a fim (acesso, alteração de políticas, operações de dados e de plataforma) com retenção aderente a normas internas.
- Conformidade: Atende a LGPD (Lei nº 13.709/2018) quanto a minimização, finalidade, segurança e prestação de contas, e às diretrizes de gestão de riscos (art. 103, Lei 14.133/2021).

2.4 Resiliência, continuidade e recuperação de desastre (dr)

- Alta disponibilidade: múltiplas réplicas/partições para serviços críticos (Kafka, Trino, ClickHouse, metastore de catálogos etc).

- Backup e restauração: políticas formais de backup de componentes de estado (metadados, catálogos, configurações), com testes de restauração periódicos.
- DR multi-site: estratégia ativo-passivo (RPO/RTO definidos em norma interna), com replicação de metadados e dados críticos conforme classificação da informação.

2.5 Governança e qualidade de dados

- Catálogo corporativo (OpenMetadata), glossário e taxonomias para padronização semântica.
- Linhagem de dados (data lineage): rastreia a origem e o consumo de dados, suportando prestação de contas e auditorias.
- Qualidade e confiabilidade: políticas de SLA de dados, profiling, alertas de drift e validações nos pipelines (ex.: regras de consistência, completude e unicidade).
- Privacidade por design: pseudonimização, mascaramento dinâmico, minimização e separação de ambientes para dados sensíveis.

2.6 Interoperabilidade e preservação do legado (sem replatforming)

- Formatos abertos: Parquet/Avro e Iceberg como padrão de tabelas transacionais do data lakehouse; evita bloqueio tecnológico.
- Compatibilidade: reuso de scripts, DAGs e jobs já existentes (HDP/TDP), reduzindo risco e custo de migração.
- Conectividade: conectores para RDBMS, NoSQL, APIs, filas, streams, ampliando integrações com sistemas institucionais (áreas-fim e áreas-meio).

2.7 Entregáveis e critérios de aceitação da subscrição (síntese)

- Ambiente Kubernetes operacional com namespaces segregados, RBAC e observabilidade implantados.
- Componentes mandatórios instalados, integrados e registrados no catálogo de serviços (inventário).
- Políticas de segurança e governança configuradas (Ranger, criptografia, auditoria).
- Pipelines de referência em produção (ex.: ingestão batch e streaming), com SLA de execução e alertas. A integração / migração de, ao menos, 1 (um) pipeline

crítico do ambiente legado, proporcionando testes de aprovação com dados reais, é requisito mandatório.
- Documentação (arquitetura, runbooks, padrões) e treinamentos realizados.
- Teste de restauração (backup/DR) aprovado e check de conformidade (LGPD e normas internas).

2.8 Racional técnico-jurídico (síntese)

- Eficiência e economicidade: elasticidade por workload, desacoplamento compute-storage e subscrição previsível em R$, reduzindo TCO ao longo do ciclo de vida (art. 34, Lei 14.133/2021).
- Segurança e soberania: operação on-prem/soberana e opção air-gapped, atendendo LGPD e gestão de riscos (art. 103).
- Padronização e continuidade: preserva o legado (HDP/TDP), evitando replatforming e mitigando risco de interrupção de serviços, em linha com o princípio da padronização e da continuidade.
- Competitividade: solução oferecida via rede de revendas, permitindo disputa em preço sem perda de padronização tecnológica.

2.9 Licenciamento, dimensionamento e faixas

Modelo de subscrição por Faixa de Nós de Processamento, com preço fixo em moeda nacional e previsibilidade orçamentária:

Faixa única: até 10 nós

Definição de Nó (unidade de licenciamento): servidor físico (bare metal) ou máquina virtual, agnóstico de hypervisor (ex.: VMware, Nutanix) e sem distinção tarifária entre físico/virtual.

Ambientes: A subscrição deve cobrir produção, homologação e desenvolvimento, garantindo consistência de versões e testes seguros antes do go-live.

2.10 Dos requisitos de suporte técnico, manutenção e atualização do TDP

A subscrição da Plataforma de Dados (**TDP Kubernetes**) deve incluir, obrigatoriamente, a prestação de serviços de Suporte Técnico Especializado e Garantia de Evolução Tecnológica (*Software Assurance*), a serem prestados diretamente pelo fabricante ou por canal certificado com acesso comprovado à engenharia do produto, atendendo aos seguintes requisitos mandatórios:

- **Sustentação do Ecossistema:** Resolução de incidentes, bugs e falhas de configuração em todos os componentes da plataforma.
- **Garantia de Evolução:** Direito de atualização para todas as novas versões (Major e Minor releases) lançadas pelo fabricante durante a vigência contratual, incluindo patches de segurança (Hotfixes) para correção de vulnerabilidades (CVEs), sem custos adicionais de licenciamento.
- **Acesso à Engenharia (Nível 3)**: O suporte não deve limitar-se ao atendimento de primeiro nível (N1). A Contratada deve garantir o escalonamento de incidentes complexo diretamente para a equipe de Engenharia de Produto do Fabricante (Product Engineering) para análise de código-fonte e correção de defeitos no core da plataforma.
- A subscrição da plataforma TDP Kubernetes deverá incluir suporte técnico Enterprise 24x7, em língua portuguesa, com Upgrade Assurance (majors/minors/patches, inclusive CVEs) e acesso a engenharia N3 (com correções no core da plataforma quando necessário). O escopo abrange:

    a. Sustentação do ecossistema: incidentes, problemas, mudanças, performance, recomendações de arquitetura;
    b. Componentes suportados (stack Open Source integrada): Spark, Trino, Kafka, Iceberg, Ozone, NiFi, Airflow, ClickHouse, Ranger, OpenMetadata, Superset, JupyterHub.
- **Ambientes:** produção, homologação e desenvolvimento, com segregação por namespaces e RBAC.
- **Atendimento Nacional:** Visando a segurança da informação e a celeridade na comunicação, todo o atendimento técnico, desde a abertura do chamado até a interação com a engenharia (Nível 3), deve ser realizado em **Língua Portuguesa (Brasil)**.
- **Fuso Horário:** A equipe de suporte deve operar em fuso horário compatível com o horário oficial de Brasília, garantindo a disponibilidade de especialistas durante o horário de expediente forense do MPMG.

Da garantia de atualização tecnológica (Cláusula de Modernização)

Fica resguardado ao MPMG (Contratante) o direito à atualização tecnológica contínua e ao suporte técnico especializado da solução contratada. É imperativo destacar que o valor estipulado para a subscrição da Plataforma de Dados (TDP) já contempla, de forma indissociável, o direito de uso, o suporte técnico de engenharia (Nível 3) e todas as atualizações tecnológicas (minor e major releases, correções de bugs, patches de segurança e evoluções de componentes).

Dada a natureza contínua e estruturante do objeto, o contrato decorrente deste Termo de Referência poderá ser sucessivamente renovado por um período de até 15 (quinze) anos, nos termos da legislação vigente. Contudo, o modelo de licenciamento adotado é o de Subscrição Anual. Dessa forma, para que a plataforma permaneça ativa e o MPMG faça jus ao recebimento contínuo de correções, melhorias de segurança e suporte da engenharia do fabricante, o valor da subscrição deverá ser pago a cada ciclo anual, sendo a prestação destes serviços condicionada à efetiva renovação da subscrição a cada ano.

Ademais, caso a plataforma Tecnisys Data Platform (TDP) sofra evoluções arquiteturais disruptivas durante o período de vigência contratual, inicialmente fixado em 12 (doze) meses, e durante eventuais prorrogações, observado o limite máximo de até 15 (quinze) anos, ou surja no mercado um novo paradigma de gestão de dados que venha a substituir o atual modelo (Kubernetes) e seja adotado pelo fabricante como padrão de mercado (state of the art), a Contratada deverá, mediante alinhamento e autorização do MPMG, disponibilizar a atualização e/ou migração do ambiente para a versão ou tecnologia superior correspondente.

Tal salto arquitetural ou evolução tecnológica deverá ocorrer sem qualquer ônus financeiro adicional de licenciamento para a Administração. Manter-se-á o equilíbrio econômico-financeiro e os valores de subscrição anuais originalmente pactuados (observados apenas os reajustes inflacionários legais). Essa garantia assegura que, ao manter a sua subscrição anual ativa, o Ministério Público usufrua sempre da tecnologia mais moderna disponível no portfólio do fabricante ao longo de até 15 (quinze) anos, eliminando definitivamente o risco de obsolescência tecnológica do parque de dados institucional.

3. ITENS 2 e 3 - OPERAÇÃO ASSISTIDA – ESCOPO CONSOLIDADO

Prestação de serviços técnicos continuados especializados em Ciência de Dados e IA (Analytics & AI) e em Engenharia de Dados, incluindo a infraestrutura acessória (Red Hat), executados na modalidade de Operação Assistida, mensurados por entregáveis previamente definidos e Horas de Serviço Técnico (HST) atreladas a esses entregáveis, vedada a mera alocação de mão de obra.

O objetivo destes serviços é operacionalizar a plataforma descrita no Item 1, transformando a infraestrutura em produtos de inteligência de dados para o MPMG.

3.1 Engenharia de Dados (Data Engineering & Ops)

- Os serviços deverão abranger todo o ciclo de vida dos dados, incluindo ingestão, processamento, armazenamento, governança, disponibilização e observabilidade, garantindo a sustentação, evolução e modernização contínua do ecossistema de dados institucional;

- Escopo: Desenvolvimento de pipelines de ingestão (Batch/Streaming), implementação de CDC (Change Data Capture), saneamento e padronização de dados, gestão de processamentos com rastreabilidade, e otimização de serviços para melhor aproveitamento dos recursos computacionais do cluster. Entregáveis: Códigos versionados (Git), relatórios técnicos e pipelines em produção.
- Desenvolvimento de Pipelines: Construção e manutenção de rotinas automatizadas de extração, transformação e carga (ETL/ELT);
- Construção de API's (Application Programming Interface): desenvolvimento de conjuntos de regras e definições que permitam que dois softwares conversem entre si, mesmo que tenham sido criados de formas diferentes;
- Ingestão em Tempo Real (CDC): Implementação de capturas de mudanças de dados (Change Data Capture) de bases transacionais para o Data Lake;
- Qualidade de Dados: Implementação de regras de validação, limpeza e de duplicação de registros (Saneamento);
- Otimização de Serviços: Monitoramento proativo de falhas e degradação de performance e ajuste fino (tuning) de componentes para garantir eficiência computacional. Implementação de métricas, logs e tracing de pipelines e serviços, criação de dashboards operacionais e apoio à análise de causa raiz (RCA);
- Apoio à operação de workloads de dados em contêineres, configuração e ajuste de deployments, jobs e operadores, troubleshooting em ambiente Kubernetes aplicado a dados e apoio à evolução da arquitetura baseada em microsserviços;
- Apoio à migração de workloads do ambiente legado (Hadoop/HDP), com adequação de pipelines para arquitetura cloud-nativa, evolução tecnológica contínua do ecossistema e minimização de riscos de descontinuidade;
- Implementação e manutenção de políticas de governança de dados com controle de acesso, auditoria e rastreabilidade (data lineage), garantia de qualidade e consistência dos dados;
- Diagnóstico de problemas entre sistema operacional e workloads de dados;
- Apoio técnico às equipes internas (GSI/STI), documentação de pipelines, arquiteturas e procedimentos, disseminação de boas práticas e suporte à autonomia progressiva da equipe interna;
- Atividades técnicas especializadas na plataforma Red Hat Enterprise Linux, compreendendo:
    a. Instalação e configuração de serviços nativos do Red Hat Enterprise Linux, como: Apache, Samba, NFS, FTP, DHCP, OpenLDAP, MySQL, Postfix,

incluindo troubleshooting de inicialização, falhas de serviços, análise de logs (journalctl, rsyslog), e problemas de rede, coleta e análise de métricas, garantindo visibilidade e suporte à operação contínua do ambiente;

b. Instalação, configuração, hardening e manutenção do sistema operacional RHEL;
c. Configuração de acesso a SAN - Storage Area Network;
d. Suporte a Ambientes Containerizados e Kubernetes, apoio à operação de nós Kubernetes em RHEL e troubleshooting de problemas de container runtime (CRI-O, Podman) tal como análise de consumo de recursos por pods e workloads e integração entre sistema operacional e orquestração;
e. Troubleshooting Avançado e Análise de Incidentes;
f. Configuração dos recursos de Alta Disponibilidade;
g. Desenvolvimento e manutenção de scripts (bash, etc.);
h. Configuração dos recursos de Gerenciamento e Administração incluindo gerenciamento de usuários, permissões, serviços (systemd), processos, armazenamento (LVM) e sistemas de arquivos;
i. Orientação em caso de dúvidas e dificuldades na utilização do ambiente, elaboração e atualização de documentação técnica, realização do registro de procedimentos operacionais e o apoio à capacitação da equipe interna.;
j. Apoio a homologação de novas soluções ou de mudanças de infraestrutura, com análise do impacto no ambiente;
k. Atualização de novos releases, analisando os impactos relativos à instalação de novos releases nos ambientes produtivos e sugerindo estratégias para contornar possíveis problemas;
l. Identificação de problemas no ambiente e tunning dos serviços nele executados, mediante análise e mapeamento dos problemas encontrados, propondo melhores práticas de administração dos serviços executados;
m. Atuação, sob demanda do MPMG, em atividades referentes ao ambiente Red Hat Enterprise Linux, incluindo definições de arquitetura, configurações, tuning, integrações, implementações de recursos, e quaisquer produtos participantes da subscrição ou dos repositórios oficiais;
n. Apoio à sustentação, diagnóstico e otimização de ambientes distribuídos e de alta criticidade, incluindo aqueles baseados em plataformas de Big Data e Analytics executadas sobre infraestrutura RHE
o. Aplicação de hardening conforme boas práticas e políticas institucionais, gerenciamento de SELinux (troubleshooting e tuning), aplicação de patches de segurança e mitigação de vulnerabilidades (CVEs), configuração de

auditoria (auditd), rastreabilidade e apoio à conformidade com LGPD e políticas internas.L e Kubernetes;

3.2 Ciência de Dados e IA (Analytics & AI)

- **Escopo:** Mineração de dados, Processamento de Linguagem Natural, desenvolvimento de modelos de IA / Machine Learning / Deep Learning e criação de dashboards. **Entregáveis:** Modelos treinados, APIs e Dashboards publicados e relatórios técnicos, além de todos artefatos que embasaram os desenvolvimentos
- Modelagem Avançada: Desenvolvimento de modelos de Machine Learning (Supervisionado e Não Supervisionado) para agrupamento, classificação, detecção de padrões e análise preditiva, gestão do ciclo de vida de modelos (MLOps) dentro do ambiente Kubernetes;
- Análise exploratória de dados estruturados e não estruturados, identificação de padrões, correlações, anomalias e tendências, apoio à compreensão de bases complexas oriundas de múltiplas fontes e preparação de datasets para uso analítico;
- 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.Limpeza, normalização e transformação de dados, tratamento de dados inconsistentes, ausentes ou duplicados, feature engineering para modelos analíticos e construção de datasets analíticos (camadas curated/gold);
- Implantação de modelos em ambiente produtivo, integração com pipelines de dados e sistemas existentes, versionamento de modelos, monitoramento de performance e degradação (model drift);
- Definição de métricas de qualidade de dados e saneamento de bases legadas, catalogação de ativos de dados e manutenção da linhagem de dados no OpenMetadata ou outra solução de gestão de metadados indicada pela Contratante;
- Criação de dashboards complexos e intuitivos em soluções de BI como Tableau, Power BI ou outra indicada pela Contratante. Desenvolvimento de painéis gerenciais (dashboards interativos), criação de relatórios analíticos, construção de visualizações para apoio à tomada de decisão da alta administração e uso de ferramentas de BI e visualização, realização de análises exploratórias de dados para identificação de padrões de fraude ou anomalias. Desenvolvimento de notebooks colaborativos via JupyterHub ou outra solução indicada pela Contratante, para atendimento as demandas do negócio;

- Processamento de Linguagem Natural (NLP): Criação de algoritmos para leitura, classificação e extração de entidades e relacionamentos em documentos não estruturados (PDFs, DOCX, TXT, etc);
- MLOps (Produtização): Empacotamento, versão e implantação de modelos de IA em produção (API Serving) dentro do ambiente Kubernetes ou em outro indicado pela Contratante;
- 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.
- Apoio às equipes internas na interpretação de resultados, disseminação de boas práticas em ciência de dados, documentação de modelos, análises e pipelines e capacitação técnica das equipes do MPMG.

3.3 A Operação Assistida compreende ainda as atividades:

- Consultoria especializada em plataformas correlatas, existente no ambiente do MPMG, compreendendo: instalação, configuração de recursos de alta disponibilidade, configuração de serviçosentre outros que fazem parte do ambiente de tecnologia atual;
- Atualização de novas releases, com análise de impactos nos ambientes produtivos, sugerindo estratégias para contornar possíveis problemas;
- Mapeamento dos problemas encontrados com proposição de melhores práticas para serem implementadas;
- Elaboração de documentação técnica;

4. TRANSFERÊNCIA DE CONHECIMENTO (REQUISITO OBRIGATÓRIO)

Execução de treinamentos e workshops técnicos e criação de documentação detalhada (Wikis e Manuais) das atividades técnicas realizadas, assegurando que o conhecimento intelectual permaneça no MPMG ao final do contrato.

Transição Contratual e Transferência de Conhecimento

Ao término contratual, a contratada deverá assegurar a completa transferência de conhecimento, incluindo:

(i) documentação técnica detalhada e atualizada;
(ii) entrega de scripts, pipelines, configurações e artefatos;
(iii) realização de treinamentos e sessões de handover;

(iv) disponibilização de repositórios organizados e estruturados;
(v) não retenção de quaisquer artefatos ou informações relacionadas ao contrato.

É expressamente vedada a retenção de conhecimento técnico ou dependência operacional que comprometa a continuidade dos serviços.

5. SUPORTE TÉCNICO E NÍVEIS DE SERVIÇO (SLA ÚNICO)

Prestar suporte técnico especializado em Língua Portuguesa, em regime 24x7 para incidentes de qualquer severidade.

Manter centro de engenharia no Brasil capaz de realizar correções de código (Nível 3) sem dependência de matriz estrangeira (Soberania).

Incidentes críticos (Sev. 1): atendimento 24x7 por portal, celular e e-mail; escalonamento direto a N3 quando caracterizada indisponibilidade ou perda de dados.

Mudanças planejadas: realizadas em janelas acordadas, com comunicação prévia e plano de reversão, podendo ocorrer em períodos que não sejam dias ou horas úteis, a fim de atender as especificidades da instituição.

→ **Severidades, tempos e prazos (SLA)**

Severidade	Descrição sintética	Tempo para Primeiro Atendimento (TMIA)	Tempo p/ Sol de Contorno (TMSO)	Tempo para Solução Definitiva (TMSD)
Crítica (S1)	Indisponibilidade total, perda de dados, impacto severo em produção	**15 min (24x7)**	**4 h (24x7)**	**24 h (24x7) h corrida**
Alta (S2)	Operação degradada com impacto relevante	**30 min (24x7)**	**8 h (24x7)**	**48 h (24x7) h corrida**
Média (S3)	Impacto parcial/não crítico, operação mantida	**2 h (24x7)**	**24 h (24x7)**	**60 h (24x7) h corrida**
Baixa (S4)	Dúvidas/solicitação sem impacto	**8 h úteis**	**48 h úteis**	**72 h úteis**

Observações operacionais

a) Chamados de S1/S2 devem possuir ponte técnica (war room) e atualização de status contínua;
b) MTTR de componentes críticos (Kafka/Trino/Metastore) deve permanecer inferior ao TMSD;
c) Mudanças emergenciais só com anuência GSI/STI e plano de reversão validado.

→ **SLOs/Indicadores de serviço (mensais)**

Disponibilidade da plataforma (produção) ≥ 99,5% (excluídas janelas autorizadas).

Sucesso de jobs críticos (Airflow/Spark) ≥ 98% no mês (com retentativas).

Tempo p95 de consultas analíticas (Trino/ClickHouse) ≤ 5s para datasets de referência definidos.

Latência p95 de ingestões (NiFi/Kafka) ≤ 60s para fluxos de referência.

Prazo de aplicação de patches críticos (CVE high/critical) ≤ 15 dias corridos após disponibilização do fabricante.

Precisão do catálogo e lineage (OpenMetadata) ≥ 95% dos ativos críticos mapeados e atualizados.

→ **Relatórios e auditoria**

Relatório mensal de SLA: volume e severidade de chamados, tempos TMIA/TMSO/TMSD, mudanças, disponibilidade, CVEs tratadas, recomendações.

Evidências: registros do ITSM, artefatos de mudança, logs/metrics/traces, atas do comitê técnico.

→ **Penalidades e créditos de serviço**

Glosas por descumprimento de metas de SLA/SLO, calculadas proporcionalmente à criticidade e recorrência, com possibilidade de creditar horas técnicas para mitigação, sem prejuízo de sanções administrativas (Lei 14.133/2021).

Reincidência em S1/S2 fora de SLA autoriza plano de ação formal com prazos mandatórios.

→ **Continuidade e recuperação de desastres (DR)**

RPO/RTO definidos em norma interna; testes de backup/restore e DR semestrais, com relatório e ações corretivas.

Topologias: suporte a ativo-passivo e replicação de metadados/dados críticos conforme classificação da informação.

→ **Canais de Atendimento**

→ A Contratada deverá disponibilizar múltiplos canais oficiais de suporte, redundantes e complementares, assegurando comunicação clara, registro rastreável e disponibilidade contínua, especialmente para incidentes de alta criticidade. Os canais deverão operar de acordo com as capacidades e regimes de atendimento previstos no SLA.

i) Portal Web de Suporte (Service Desk / ITSM) - Obrigatório

Canal primário para registro, acompanhamento e gestão integral dos chamados. O portal deverá permitir:

- abertura de incidentes, requisições e problemas;
- upload de logs, arquivos de diagnóstico e evidências;
- consulta a base de conhecimento (KB) e histórico;
- visualização em tempo real de SLA, severidade e status;
- interação com atendentes e engenharia;
- autenticação segura (SSO/OAuth2/SAML).

Requisito técnico: A solução deverá **ser acessível via HTTPS e operar 24x7.**

ii) Linha telefônica direta (Telefone Celular ou fixo nacional) - Obrigatória

Canal de **resposta imediata** para incidentes **Severidade 1 (Críticos)**, com suporte **24x7**.

Requisitos:

- atendimento humano imediato (sem URA em casos críticos);
- prioridade para ativação de *bridge room* (sala de crise/war room);
- conferência com engenheiros N2/N3 quando necessário.

Objetivo: assegurar tempo mínimo de reação e comunicação síncrona durante falhas graves.

iii) E-mail Corporativo - Obrigatório

Canal para demandas não urgentes, orientações administrativas, anexos extensos ou comunicação formal com o suporte.

Requisitos:

- caixa monitorada dentro do horário de expediente ampliado;

- confirmação automática de recebimento;
- vinculação ao ITSM para rastreamento.

iv) Aplicativo de Mensageria Segura (opcional recomendável)

A contratada poderá disponibilizar canal complementar via **mensageria corporativa segura** (ex.: Microsoft Teams, Rocket. Chat corporativo ou equivalente), para:

- interação rápida com equipes N1/N2;
- acompanhamento de incidentes em tempo real;
- abertura de *war rooms* e videoconferências emergenciais.

Obs.: não substitui o registro formal no ITSM, mas acelera comunicação em situações críticas.

v) Canal de videoconferência para War Rooms (Obrigatório para S1 / S2)

A contratada deve operar **salas virtuais de emergência** para incidentes de severidade alta ou crítica.
Requisitos mínimos:

- acesso imediato;
- gravação opcional;
- compartilhamento de tela;
- histórico de chat;
- convidados ilimitados da equipe GSI/STI.

vi) API de Integração com o ITSM (Recomendável)

A contratada deverá fornecer, quando disponível:

- **API REST** para integração com sistemas internos (ex.: GLPI, ServiceNow, OTRS);
- sincronização de status e comentários;
- exportação estruturada de chamados e métricas.

CANAIS DE ATENDIMENTO (CONSOLIDAÇÃO)

Canal	Tipo	Obrigatório?	Finalidade Principal	Disponibilidade

Portal Web (ITSM)	Assíncrono	**Sim**	Registro e gestão formal de chamados	24x7
Telefone Celular ou Fixo Nacional	Síncrono	**Sim**	Incidentes críticos (S1)	24x7
E-mail Corporativo	Assíncrono	**Sim**	Comunicação administrativa e técnica não urgente	Horário ampliado
Mensageria Segura	Síncrono/Assíncrono	Recomendável	Agilidade de comunicação e acompanhamento	Horário estendido / conforme acordo
Videoconferência (War Room)	Síncrono	**Sim para S1/S2**	Gestão de crises e falhas críticas	Sob demanda imediata
API de Integração	Automação	Recomendável	Integração com ITSM do MPMG	24x7

6. MATRIZ DE RESPONSABILIDADE (RACI)

A Matriz utiliza a metodologia internacional de atribuição de responsabilidades:

[R] Responsável (Responsible): Quem executa a tarefa

[A] Autoridade/Aprovador (Accountable): Quem tem a palavra final e assina a entrega. Gestor e Fiscais do Contrato (MPMG).

[C] Consultado (Consulted): Quem possui expertise e deve ser ouvido antes ou durante a tarefa.

[I] Informado (Informed): Quem deve ser avisado após a conclusão da tarefa.

A. EIXO: INFRAESTRUTURA E PLATAFORMA (KUBERNETES/TDP)

Atividade / Processo	**Eng. de Dados (Contratada)**	**Ciência de Dados (Contratada)**	**Fiscalização (MPMG)**

Instalação/Upgrade do TDP no Cluster	**R**	**I**	**A**
Monitoramento de disponibilidade 24x7	**R**	**I**	**I**
Tuning de performance do Apache Spark	**R**	**C**	**A**
Gestão de Segurança (Apache Ranger)	**R**	**C**	**A**
Backup e Disaster Recovery (Metadados)	**R**	**-**	**A**

B. EIXO: ENGENHARIA E GOVERNANÇA DE DADOS

Atividade / Processo	**Eng. de Dados (Contratada)**	**Ciência de Dados (Contratada)**	**Fiscalização (MPMG)**	**Gestores de Negócio**
Construção de Pipelines de Ingestão	**R**	**C**	**A**	**I**
Definição de Catálogo (OpenMetadata)	**R**	**R**	**A**	**C**
Saneamento e Limpeza de Bases Legadas	**R**	**C**	**I**	**-**
Implementação de camadas (Lakehouse)	**R**	**C**	**A**	**-**

C. EIXO: CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL

Atividade / Processo	**Eng. de Dados (Contratada)**	**Ciência de Dados (Contratada)**	**Fiscalização (MPMG)**	**Gestores de Negócio**

Treinamento de Modelos de IA / NLP	**S** (Suporte)	**R**	**A**	**C**
Validação de Acurácia dos Modelos	**-**	**R**	**A**	**C**
Deploy de Modelos em Produção (MLOps)	**R**	**R**	**A**	**I**
Criação de Dashboards (Superset)	**C**	**R**	**A**	**R** (Valida)

D. EIXO: GESTÃO CONTRATUAL E CONFORMIDADE

Atividade / Processo	**Preposto (Contratada)**	**Equipe Técnica (Contratada)**	**Fiscalização (MPMG)**	**Gestão de Pagamento**
Emissão de Relatório Mensal (RMT)	**R**	**S**	**A**	**I**
Transferência de Conhecimento (Wiki)	**I**	**R**	**A**	**-**
Assinatura de Termos de Sigilo	**R**	**R**	**R**	**-**
Solicitação de Glosa por baixo IMR	**I**	**I**	**R/A**	**I**

CONSIDERAÇÕES VINCULADAS À GESTÃO ESTRATÉGICA DO MPMG:

Segregação de Funções: Note que a **Engenharia de Dados** atua como suporte [S] para a **Ciência de Dados** no deploy de modelos. Isso garante que o cientista foque na lógica do algoritmo, enquanto o engenheiro garante que o Kubernetes suporte a carga.

Soberania Decisória: Em todas as linhas, o MPMG retém o papel de **[A] Autoridade**. A contratada executa, mas o MPMG decide o que entra em produção.

Propriedade Intelectual: A coluna de "Equipe Técnica" deve estar ciente de que o resultado da sua responsabilidade [R] gera um ativo que pertence ao MPMG, conforme cláusula já existente no TR.

7. DIRETRIZES DE EXECUÇÃO

Propriedade Intelectual: Conforme o Termo de Referência, todos os scripts de automação, códigos-fonte de modelos de IA e estruturas de pipelines desenvolvidos nestas atividades são de propriedade exclusiva do MPMG.

Transferência de Conhecimento: É obrigação de ambos os perfis a documentação técnica atualizada no Wiki/Repositório do MPMG e a realização de reuniões de repasse técnico mensais.

Local de Prestação: Os serviços serão realizados preferencialmente de forma remota, com acesso via VPN segura, salvo necessidade de intervenção presencial determinada pelo GSI/DINI e STI.

8. REQUISITOS DE SEGURANÇA DA INFORMAÇÃO E CONFORMIDADE COM A LGPD (LEI 13.709/2018)

A plataforma ofertada deve possuir uma arquitetura de segurança unificada e "Zero Trust", garantindo a confidencialidade, integridade e disponibilidade dos dados sensíveis e sigilosos.

A solução deve prover ferramentas nativas para operacionalizar os direitos dos titulares de dados, incluindo:

Direito ao Esquecimento (Expurgo Lógico): Suporte a transações ACID (via Delta Lake, Iceberg ou Hive ACID) para permitir a atualização e deleção pontual de registros de pessoas físicas em grandes *Data Lakes*, sem a necessidade de reescrever petabytes de dados.

Relatório de Impacto: Capacidade de gerar relatórios de linhagem via **OpenMetadata**, demonstrando visualmente onde os dados sensíveis são originados, como são transformados e por quem são consumidos.

9. GOVERNANÇA E CONTROLE DE ACESSO CENTRALIZADO

A solução deve fornecer um módulo de segurança (baseado no padrão **Apache Ranger** ou funcionalmente equivalente) capaz de aplicar políticas de acesso

granulares de forma transversal na camada de acesso à dados, com as seguintes capacidades:

Granularidade Fina: Capacidade de definir permissões de acesso não apenas a nível de banco de dados ou tabela, mas especificamente a nível de **Coluna** e **Linha**.

- *Exemplo de uso:* Um analista pode ver a tabela "Investigados", mas não pode ver as linhas classificadas como "Sigilo Nível 5".

Controle de acesso por papéis (RBAC): Definição de permissões por funções e perfis (ex.: *Data Engineer*, *Analista*, *Auditor*), mapeadas a usuários e grupos, com concessão de acesso padronizada e auditável a recursos (bancos, schemas, tabelas, colunas e operações), reduzindo permissões *ad hoc* e facilitando governança e segregação de funções.

Mascaramento Dinâmico de Dados (*Dynamic Data Masking*): Capacidade de ocultar, mascarar (ex: ***-45) ou anonimizar dados sensíveis (CPF, CNPJ, Nomes) em tempo de execução da consulta (*Query Time*), baseando-se no perfil do usuário, sem a necessidade de duplicar o dado ou alterar o armazenamento físico.

10. AUTENTICAÇÃO E GESTÃO DE IDENTIDADE

Integração Corporativa: A plataforma deve integrar-se nativamente com o Diretório Corporativo do MPMG (Active Directory/LDAP) para provisionamento e sincronização de usuários e grupos.

11. CRIPTOGRAFIA E PROTEÇÃO DO DADO

Criptografia em Trânsito: Obrigatoriedade de uso de protocolo TLS 1.2 (ou superior) para toda a comunicação entre os nós do cluster e conexões de clientes externos (JDBC/ODBC/REST).

12. AUDITORIA E SOBERANIA (TRAVA DE SEGURANÇA NACIONAL)

Auditoria Imutável: A solução deve gerar logs de auditoria detalhados de todos os eventos de acesso (sucesso e negação) e operações administrativas, garantindo a rastreabilidade forense de quem acessou o quê e quando.

Soberania dos Logs e Metadados:

- Os logs de auditoria e os metadados do catálogo (informações sobre esquemas, usuários e políticas) devem ser armazenados e processados

exclusivamente na infraestrutura designada pelo MPMG (*Local* ou *Nuvem Privada Nacional*).

- Vedação: É vedada a utilização de soluções que exijam o envio obrigatório de telemetria, metadados ou logs de auditoria para planos de controle situados em nuvens públicas estrangeiras (SaaS exógeno), visando mitigar riscos de espionagem industrial ou acesso indevido por força de leis estrangeiras (ex: *Cloud Act*).

Operação *Air-Gapped*: A plataforma deve possuir a capacidade técnica comprovada de operar em ambientes totalmente isolados da Internet (*Air-Gap*), sem prejuízo de suas funcionalidades críticas, garantindo a máxima segurança para redes de inteligência.

13. DOS REQUISITOS LEGAIS, DE CONFORMIDADE E GOVERNANÇA

A presente contratação rege-se pelos princípios da Legalidade, Impessoalidade, Moralidade, Publicidade e Eficiência, devendo a **CONTRATADA** e a **Solução Tecnológica** ofertada atenderem estritamente aos seguintes dispositivos legais e normativos:

13.1 DA CONFORMIDADE COM A LEI GERAL DE PROTEÇÃO DE DADOS (LEI Nº 13.709/2018)

A Plataforma de Dados (Software) deve possuir arquitetura orientada à privacidade (*Privacy by Design*), oferecendo nativamente funcionalidades que permitam ao MPMG operacionalizar os direitos dos titulares previstos no Art. 18 da LGPD, especificamente:

Anonimização e Bloqueio: Ferramentas para mascaramento de dados sensíveis em ambientes de desenvolvimento e teste.

Relatório de Impacto: Capacidade de rastreabilidade (*Data Lineage*) para geração de Relatórios de Impacto à Proteção de Dados Pessoais (RIPD). A Contratada obriga-se a manter o sigilo absoluto sobre quaisquer dados processados, vedada a utilização para fins distintos do objeto contratual, sob pena de responsabilização civil, penal e administrativa.

13.2 DA VEDAÇÃO À VARIAÇÃO CAMBIAL (LEI Nº 10.192/2001 E JURISPRUDÊNCIA TCU)

Em estrita observância ao Art. 2º da Lei nº 10.192/2001 (Plano Real) e à jurisprudência consolidada do Tribunal de Contas da União (Acórdão TCU nº 1.253/2012-Plenário), é **vedada a estipulação de pagamento em moeda estrangeira** ou a indexação direta

de reajustes à variação cambial (Dólar/Euro) para serviços prestados e softwares licenciados em território nacional. A proposta comercial deve ser apresentada em **Moeda Corrente Nacional (Real)**, devendo a licitante absorver os riscos de flutuação cambial de eventuais insumos importados, garantindo a **previsibilidade orçamentária** (Art. 18 da Lei 14.133/21).

14. DA PROPRIEDADE INTELECTUAL E INEXISTÊNCIA DE VÍNCULO (SÚMULA 331/TST)

Propriedade dos Artefatos: Todos os **Produtos de Dados** desenvolvidos durante a execução dos serviços de Operação Assistida (Modelos de IA, Pipelines ETL/ELT, Dashboards e automações) serão de **propriedade exclusiva e perpétua do Ministério Público de Minas Gerais**, devendo ser entregues em repositório Git corporativo, livres de ofuscação.

Inexistência de Subordinação: Os serviços técnicos (Itens 2 e 3) serão executados na modalidade de **Operação Assistida por Produto/Resultado**, não havendo qualquer relação de subordinação jurídica, hierárquica ou pessoalidade entre os servidores do MPMG e os funcionários da Contratada, nos termos da Súmula nº 331 do Tribunal Superior do Trabalho (TST), sendo vedada a ingerência do MPMG na gestão interna da equipe terceirizada.