

Chamada de orçamento para Prestação de Serviço Implantação de Data Lake na Unioeste, Integração e Disponibilização de Dados das Universidades Paranaenses através de Data Hub (SETI).

PROJETO

Projeto de Implantação de Data Lake na Unioeste, Integração e Disponibilização de Dados das Universidades Paranaenses através de Data Hub (SETI).

CONTEXTO

As sete Instituições Estaduais de Ensino Superior (IEES) do Paraná, enfrentam um cenário desafiador em relação à gestão eficiente dos dados acadêmicos e administrativos. O aumento exponencial na geração de informações aliada à fragmentação dos sistemas existentes tem gerado dificuldades significativas para a monitorização do desempenho acadêmico dos alunos, o que compromete a capacidade das instituições em identificar situações de evasão e implementar estratégias pedagógicas eficazes.

O Sistema Estadual de Ciência, Tecnologia e Ensino Superior do Paraná é composto pelas Universidades: Universidade Estadual de Londrina (UEL); Universidade Estadual de Maringá (UEM); Universidade Estadual de Ponta Grossa (UEPG); Universidade Estadual do Oeste do Paraná (Unioeste); Universidade Estadual do Centro-Oeste (Unicentro); Universidade Estadual do Norte do Paraná (UENP); Universidade Estadual do Paraná (Unespar). Em 2022, as sete instituições estaduais de ensino superior somam 73 mil estudantes regularmente matriculados em 381 cursos de graduação, 304 cursos de especialização, 196 cursos de mestrado e 92 cursos de doutorado. Nas universidades atuam 7.720 docentes (90% com Mestrado e Doutorado) e 8.847 agentes universitários.

A ausência de uma infraestrutura integrada para a centralização, armazenamento e análise de dados resulta em limitações na formulação de políticas educacionais baseadas em evidências, prejudicando tanto as ações internas das universidades quanto as diretrizes mais amplas estabelecidas pela Secretaria de Estado da Ciência, Tecnologia e Inovação (SETI). Essa lacuna impõe um desafio considerável às IEES, que necessitam de dados consolidados e acessíveis para fundamentar suas decisões estratégicas.

Além disso, a falta de interoperabilidade entre as diversas instituições do estado impede a troca de informações cruciais, essencial para uma gestão colaborativa e sinérgica dos recursos

educacionais, impactando diretamente a qualidade do ensino oferecido e a adequação às necessidades dos estudantes. A personalização do ensino conforme os desafios apresentados é limitada, resultando em um ensino que muitas vezes não atende às especificidades da comunidade acadêmica.

Assim, torna-se imprescindível abordar essa problemática por meio de uma solução tecnológica inovadora, com o objetivo de viabilizar a implementação de Data Lakes interoperáveis nas IEES. Essa solução não apenas responderia à demanda latente por um gerenciamento eficiente e integrado dos dados, mas também promoveria a inovação e a sustentabilidade nos processos educacionais, alinhando-se aos objetivos estratégicos da Política Estadual de Ciência, Tecnologia e Inovação (PECTI-PR 2024-2030) e aos Objetivos de Desenvolvimento Sustentável (ODS), especialmente ao ODS 4, que visa garantir educação inclusiva e equitativa de qualidade, e ao ODS 9, que propõe construir infraestruturas resilientes e promover a inovação.

Portanto, a descrição dessa necessidade revela um problema que transcende o âmbito interno das instituições, afetando diretamente o interesse público, ao comprometer a qualidade da educação e a eficácia das políticas públicas. É necessário um empenho conjunto para superar esses desafios, buscando sempre fomentar a melhoria contínua no processo educativo e na formação de cidadãos capacitados e críticos.

Para tanto, os recursos disponíveis para esta contratação são oriundos do contrato GMS 9134/2025, celebrado entre a FUNDEP e a UNIOESTE, cujo objeto o apoio com gerenciamento de recursos oriundos da EG: 05/2025 e TED 136/2025, projeto protocolado sob o no 24.911.003.

OBJETO

Implementação de projeto de Data Lake e Data Hub com infra on-premise na Universidade Estadual do Oeste do Paraná, atendendo o Data Lake e a sumarização no Data Hub para a SETI. O projeto deverá ser implementado seguindo rigorosamente o anexo já contratado e construído por empresa especializada. Este rigor é para garantir a integração eficiente e à gestão analítica avançada dos dados acadêmicos e administrativos, contribuindo para a inovação na gestão universitária e para a formulação de políticas públicas baseadas em evidências, conforme descrição abaixo.

Meta:

Implementar integralmente o projeto executivo de dados contratado com a empresa **DataEx**, estabelecendo duas plataformas cruciais para a gestão e compartilhamento de dados no Paraná:

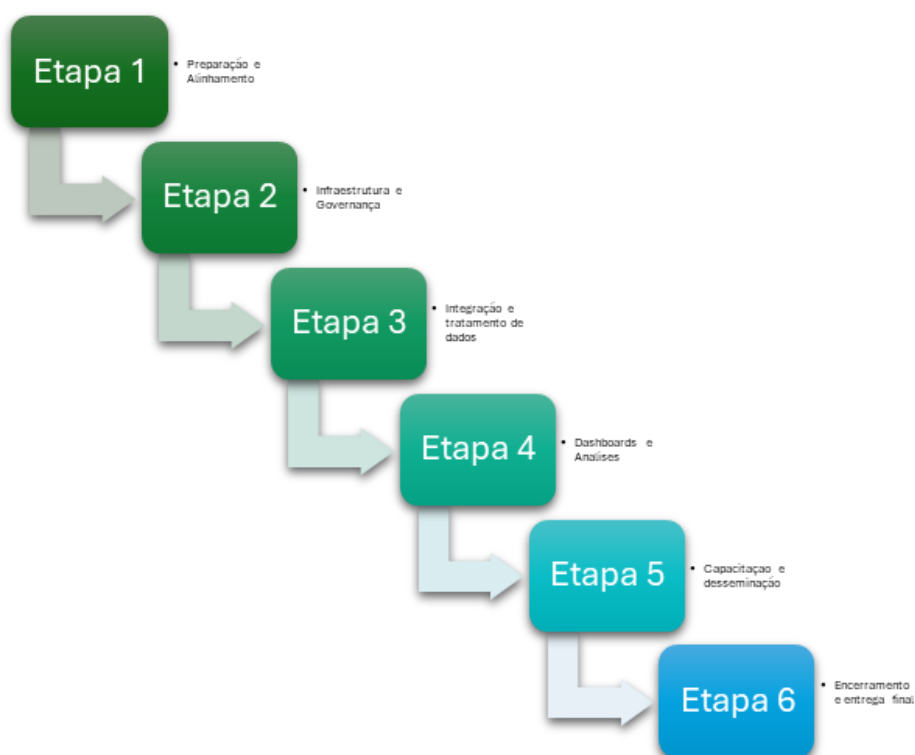
1. **Data Lake da Unioeste:** Criação e operacionalização do ambiente de **Data Lake** na **Universidade Estadual do Oeste do Paraná (Unioeste)**.

2. **DataHub da SETI:** Criação e operacionalização do **DataHub** para a **Secretaria da Ciência, Tecnologia e Ensino Superior do Paraná (SETI)**.

A implantação deve ser realizada em **conformidade rigorosa** com:

- **Especificações Técnicas** detalhadas no projeto.
- **Padrões de Interoperabilidade** que garantam a comunicação e troca eficiente de dados entre sistemas.
- **Requisitos de Segurança da Informação, Governança de Dados e Privacidade de Dados**, em observância às regulamentações vigentes (incluindo a LGPD).

Roadmap:



Etapas:

Etapa 1 – Preparação e Alinhamento

Objetivo: Garantir que todos os envolvidos entendam o escopo, metas e responsabilidades do projeto.

Atividades principais:

- Reunião de kick-off com todas as partes interessadas.
- Revisão e confirmação do escopo do projeto.
- Definição das responsabilidades por equipe e indivíduo (Matriz RACI).
- Aprovação do cronograma macro e metodologia de trabalho.

- Priorização das áreas de negócio para ciclo de desenvolvimento:
 - Refinamento de indicadores (KPI's);
 - Priorização de Integração e tratamento de dados (ETL) por área;
 - Priorização de desenvolvimento de Modelos Semânticos por área;
 - Priorização de desenvolvimento de Indicadores do Escopo da área;
 - Priorização de desenvolvimento dos dashboards e Análise por área.

Entregáveis:

- Documento de escopo validado.
- Cronograma macro aprovado.
- Mapa de responsabilidades (RACI).
- Plano de comunicação do Projeto

Etapa 2 – Infraestrutura e Governança

Objetivo: Criar a base tecnológica e organizacional para o tratamento e uso seguro dos dados.

Atividades principais:

- Configuração e homologação da infraestrutura tecnológica;
- Implementação da governança de dados (políticas, fluxos, papéis);
- Definição e aplicação de políticas de acesso e segurança;
- Validação e testes dos ambientes criados.

Entregáveis:

- Ambientes de trabalho configurados e homologados.
- Documento de política de governança de dados.
- Plano de segurança e controle de acessos.

Etapa 3 – Integração e Tratamento de Dados (ETL)

Objetivo: Garantir que todas as bases de dados estejam integradas, tratadas e aptas para análises, priorizando as áreas de negócio mais estratégicas.

Atividades principais:

- Priorização das áreas de negócio para o ciclo de desenvolvimento.
- Desenvolvimento de Integração e tratamento de dados (ETL).
- Desenvolvimento de modelos semânticos para padronização e consistência das informações.
- Monitoramento contínuo da qualidade e consistência dos dados.
- Desenvolvimento de Automação dos processos de ingestão e atualização de dados (Jobs).

- Desenvolvimento de CICD (se aplicável)

Entregáveis:

- Bases de dados integradas e tratadas.
- Modelos semânticos documentados e aprovados.
- Relatório de qualidade de dados.
- Scripts e rotinas automatizadas de atualização.

Etapa 4 – Dashboards e Análises

Objetivo: Disponibilizar visualizações e análises que atendam às necessidades de monitoramento e tomada de decisão, com foco nos indicadores definidos no escopo.

Atividades principais:

- Criação dos indicadores previstos no escopo de cada área.
- Desenvolvimento dos dashboards e relatórios analíticos.
- Análises exploratórias para validação e ajustes.
- Testes de performance e usabilidade junto aos usuários-chave.
- Ajustes de layout, métricas e filtros conforme feedback.

Entregáveis:

- Indicadores criados e documentados.
- Dashboards em produção.
- Documento de especificações das métricas.
- Registro de feedback e ajustes aplicados.

Etapa 5 – Capacitação e Disseminação

Objetivo: Garantir que todos os usuários consigam operar e extrair valor das ferramentas criadas.

Atividades principais:

- Treinamento presencial e/ou remoto para os usuários finais.
- Produção de materiais de apoio (manuais, tutoriais).
- Comunicação institucional para engajar o uso das soluções.

Entregáveis:

- Plano de treinamento executado.
- Material didático distribuído.
- Registro de participação e avaliação dos treinamentos.

Etapa 6 – Encerramento e Entrega Final

Objetivo: Formalizar o encerramento do projeto, garantindo aceite formal e registro de lições aprendidas.

Atividades principais:

- Validação final dos entregáveis com stakeholders.
- Obtenção do aceite formal de entrega.
- Encerramento administrativo do projeto.
- Registro e compartilhamento das lições aprendidas.

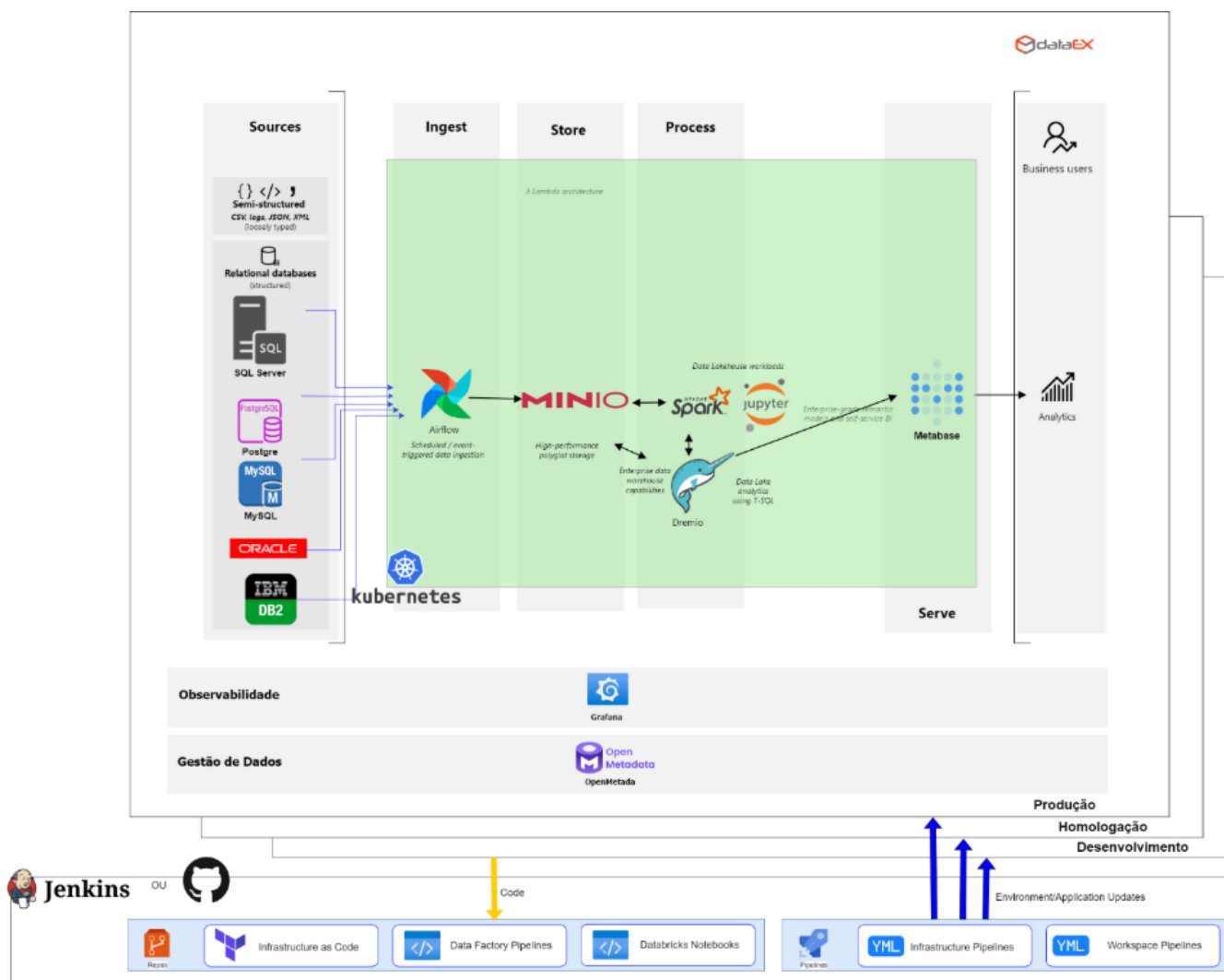
Entregáveis:

- Termo de aceite assinado.
- Relatório final do projeto.
- Documento de lições aprendidas.

Ferramentas tecnológicas propostas:

- **Sources (Fontes de Dados):** SQL Server / PostgreSQL / MySQL
- **Semi-estruturados:** CSV, JSON, XML, PDF (possivelmente via filesystem) / Acesso via pipelines Airflow + agentes customizados (ETL Python)
- **Ingest:** Apache Airflow
- **Store – MinIO**
- **Process:** Spark, Dremio, Jupyter
- **Serve:** Metabase
- **Observabilidade:** Grafana
- **Governança:** OpenMetadata
- **CI/CD:** Jenkins ou GitHub Actions

Documento Técnico – Arquitetura Onpremises (Unidata)



Visão Geral da Arquitetura

Esta arquitetura define uma stack **100% on-premises e open source**, com foco em governança, flexibilidade, escalabilidade e controle total dos dados. Adota o padrão **Lambda Architecture**, separando ingestão, armazenamento, processamento e consumo analítico.

Componentes por Camada

1. Sources (Fontes de Dados)

Sistemas de origem on-premises:

- **Bancos relacionais:**
 - SQL Server
 - PostgreSQL
 - MySQL
- **Semi-estruturados:**
 - CSV, JSON, XML, PDF (possivelmente via filesystem)
- **Acesso via pipelines Airflow + agentes customizados (ETL Python)**

2. Ingest – Airflow

- **Apache Airflow:**
 - Orquestração de jobs de ingestão
 - Agendamento de DAGs para extração de bases relacionais e arquivos
 - Integração com Kubernetes para escalar workers
 - Pipelines versionadas via Git

3. Store – MinIO

- **MinIO:**
 - Object storage compatível com S3
 - Armazena dados brutos e tratados em formato Parquet/CSV
 - Alta performance, replicação e versionamento
 - Montado como o “OneLake” da stack on-premises

4. Process – Spark, Dremio, Jupyter

- **Apache Spark:**
 - Transformações distribuídas
 - ETL em grande volume
 - Integração com Jupyter para ciência de dados
- **Jupyter Notebooks:**
 - Desenvolvimento de notebooks analíticos

- Treinamento de modelos e análises exploratórias
- **Dremio:**
- SQL Lakehouse engine
- Conecta no MinIO para leitura em Parquet/Delta
- Linhagem, aceleração com reflections e governança SQL

5. Serve – Metabase

- **Metabase:**
- Plataforma de BI open source
- Self-service para usuários de negócio
- Conecta diretamente ao Dremio
- Dashboards interativos e RBAC nativo

6. Observabilidade – Grafana

- Dashboards de monitoramento para:
- Pipelines Airflow
- Jobs Spark
- Performance do cluster Kubernetes

7. Governança – OpenMetadata

- Catálogo de dados
- Linhagem de pipelines Airflow, dados MinIO e dashboards Metabase
- Termos de glossário, classificações e owners
- Integração com CI/CD (Jenkins ou GitHub)

8. CI/CD – Jenkins ou GitHub Actions

- Infraestrutura como código (IaC) via YAML e Terraform
- Integrações com:
- Pipelines do Airflow
- Deploy de notebooks
- Provisionamento de ambientes
- Ambientes: **Desenvolvimento, Homologação, Produção**

DA PROPOSTA

A proposta deve conter obrigatoriamente:

- Descrição completa do Serviço
- Valor global
- Validade da proposta
- Dados da empresa
- Contato para dúvidas

Demais itens esperados na proposta:

- Cronograma referencial para implementação até dezembro/26
- Garantia de validar e alinhar o projeto com todas as IEES e com a SETI, por meio de encontros técnicos presenciais e/ou remotos ao longo da execução;
- Implantação/validação da Arquitetura de Referência;
- Planejamento de Governança e Segurança;
- Implantação da Infraestrutura;
- Desenvolvimento dos Pipelines e Modelos de Dados;
- Elaboração dos dashboards e indicadores/análises;
- Validação e Testes;
- Workshop/Transferência de Conhecimento para usuários e administradores;
- Encerramento e Suporte Inicial no período de garantia.

Lista 50 Indicadores Principais

Indicador	Descrição breve
Taxa de evasão (com recortes)	Percentual de estudantes que evadem no período.
Taxa de retenção	Percentual de estudantes que permanecem matriculados após período.
Taxa de conclusão (prazo/tempo ampliado)	Percentual de formandos que concluem no prazo ou em tempo ampliado.
Tempo médio de formação	Tempo médio que estudantes levam para concluir o curso.
Índice de rendimento acadêmico (IRA)	Média de notas/créditos obtidos.
Índice de aproveitamento	Aprovações sobre disciplinas cursadas.
Índice de reprovação por disciplina	Percentual de reprovação em disciplinas específicas.
Frequência média dos estudantes	Percentual médio de presença dos estudantes em aulas.
Número de alunos por turma	Média de estudantes por turma/disciplina.
Satisfação acadêmica dos estudantes	Grau de satisfação dos estudantes.

Conceito Enade	Desempenho dos concluintes no Enade.
IDD	Valor agregado entre ingressantes e concluintes.
CPC	Avaliação multidimensional do curso.
IGC	Qualidade institucional agregada.
Participação dos concluintes no Enade	Percentual de concluintes que realizaram o exame.
Índice de empregabilidade de egressos	Percentual de egressos empregados.
Participação em atividades extracurriculares	Proporção de estudantes em IC, extensão, monitoria etc.
Proporção de estudantes com apoio pedagógico/assistência financeira	Percentual de alunos com suporte institucional.
Participação relativa por segmentos	Distribuição percentual do corpo discente por segmentos.
Taxa de resposta da busca ativa	Percentual de alunos que responderam às ações de contato/apoio.
Índice de execução orçamentária (macro)	Executado sobre orçado.
Percentual de execução orçamentária (detalhe)	Executado/orçado por centro/área.
Total de receita orçamentária (LOA)	Volume anual aprovado na LOA.
Receita própria	Recursos gerados pela universidade.
Repasse do Tesouro Estadual	Recursos transferidos pelo Estado.
Despesas totais realizadas	Gastos efetivos por área.
Grau de dependência de repasse estatal	Proporção Tesouro/receita total.
% do orçamento destinado à pesquisa e extensão	Despesa em pesquisa/extensão sobre orçamento total.
Investimento em infraestrutura	Aplicações em obras, laboratórios, TI.
Restos a pagar acumulados	Compromissos empenhados e não pagos.
Tempo médio de atendimento a processos	Tempo entre abertura e conclusão.
Taxa de atendimento digital	Serviços administrativos disponíveis digitalmente.
Índice de transparência (ativa)	Disponibilidade de informações ao público.
Taxa de licitações com sucesso	Licitações concluídas sobre iniciadas.
Índice de eficiência licitatória	Proporção de dispensa/inexigibilidade sobre total.
Índice de cumprimento de prazos legais	Processos concluídos dentro dos prazos.
Índice de sustentabilidade administrativa	Práticas sustentáveis em recursos/processos/infra.
Taxa de renovação normativa	Total de regulamentos/resoluções aprovados/alterados.
Índice de publicações científicas	Contagem por tipo/indexação.
Índices de colaboração científica	Coautorias e parcerias nacionais/internacionais.

Total de citações	Soma de citações recebidas.
FWCI (impacto ponderado por área)	Citações recebidas ÷ esperadas.
h-index (institucional e por docente)	Produtividade/impacto.
Publicações com acesso aberto (%)	Percentual de publicações em open access.
Projetos financiados (% e total)	Quantidade e proporção de projetos com fomento.
Receita por projetos financiados	Total de recursos obtidos via P&D.
Carga horária docente (total e por eixo)	Carga total e distribuída em ensino/pesquisa/extensão.
Taxa de absenteísmo (geral e doença)	Dias de ausência sobre dias trabalhados.
Qualificação do corpo (titulação)	% de docentes/técnicos com pós-graduação.
People analytics (satisfação, avaliação, turnover, retenção, capacitação)	Pacote de acompanhamento de clima e desenvolvimento.

Descrição dos Requisitos da Contratação

A seguir apresentam-se os requisitos mínimos e indispensáveis que a futura contratação deverá observar, conforme boas práticas recomendadas pelo TCU e diretrizes internas da Unioeste/SETI-PR. Estes requisitos serão transpostos para o Termo de Referência com o objetivo de permitir a seleção da proposta mais vantajosa, assegurando qualidade, sustentabilidade e conformidade legal.

1. Natureza do Serviço

- **Tipo:** Prestação de serviço técnico especializado **não continuado** (produto) de consultoria em arquitetura de dados, governança e capacitação.
- **Prazo Máximo de Execução:** 12 (doze) meses contados da assinatura do contrato.
- **Forma de Entrega e pagamento:** Por produto, com pagamentos vinculados à aceitação dos seis marcos: (i) Preparação e Alinhamento, (ii) Infraestrutura e Governança, (iii) Integração e Tratamento de Dados (ETL), (IV) Dashboards e Análises, (V) Capacitação e Disseminação e (VI) Encerramento e Entrega Final.

2. Padrões Mínimos de Qualidade

a) Equipe Profissional

- Coordenador de projeto com experiência ≥ 3 anos em projetos de Data Lake/Data Hub e certificação PMP ou Prince2, ou equivalente.
- Arquiteto de dados com comprovação de atuação em, no mínimo, 3 projetos de arquitetura de dados em ambiente de nuvem híbrida.
- Cientista de dados com portfólio de modelos em produção com comprovação de atuação em, no mínimo, 3 projetos.

b) Metodologia

- Aplicação de práticas ágeis (*Scrum/Kanban*) com *sprint reviews* quinzenais e *dailies* documentadas.

- Documentação de requisitos conforme padrões internacionais como BABOK v3, modelagem de processos em BPMN 2.0 e artefatos técnicos em padrão ArchiMate 3.2.

c) Entregáveis

- Relatórios técnicos, diagramas/ código-fonte (ex.: *scripts* ETL) em repositório Git privado da Unioeste.
- Arquitetura de referência aderente a padrões abertos (Apache Iceberg, Parquet, Trino, Spark e/ou outras).
- Plano de Governança compatível com LGPD – contendo política de classificação, máscara e retenção de dados.
- Material didático de capacitação.

d) Indicadores de Desempenho (SLAs)

- Cumprimento de prazos $\leq 100\%$ do cronograma.
- Satisfação dos participantes na capacitação $\geq 80\%$ (nota média $\geq 8/10$).

3. Sustentabilidade e Responsabilidade Socioambiental

Adoção do princípio *digital-first*: todos os documentos devem ser eletrônicos, assinados digitalmente e armazenados em repositório de baixo consumo energético.

4. Conformidade Legal e de Segurança

- Total aderência à Lei 14.133/2021, LGPD (Lei 13.709/2018) e Normas ABNT ISO/IEC 27001 e 27701.
- Assinatura de Acordo de Confidencialidade (NDA) cobrindo dados sensíveis das IEES.
- Garantia de não *lock-in*: todos os artefatos devem ser entregues em formatos abertos ou amplamente adotados.

5. Obrigações da Contratada

- Disponibilizar gerente de projeto como *single point of contact* durante todo o contrato.
- Transferir conhecimento (*hand-over*) abrangente para a equipe técnica das Unioeste e SETI
- Manter cópias de segurança de todos os artefatos produzidos até 30 dias após a entrega final.
- Responder a eventuais auditorias do Estado ou órgãos de controle.

6. Habilitação Jurídica, Fiscal e Trabalhista:

- a. Comprovante de CNPJ.
- b. Certidões negativas com as Fazendas Federal, Estadual e Municipal;
- c. Certidão negativa de débitos trabalhistas;

d. Certidão negativa do FGTS.

PRAZO E CONDIÇÕES PARA PARTICIPAÇÃO

Empresas interessadas devem enviar a proposta e os documentos de habilitação Jurídica e regularidade Fiscal e Trabalhista até o dia 09.01.2026 para os e-mails fundep@fundeppr.com.br e/ou fundep.pr@gmail.com.

Observações Finais Para mais informações, entre em contato pelos telefones 45 3220 3211 e 45 99944 0137, ou e-mails fundep@fundeppr.com.br e/ou fundep.pr@gmail.com.

Cascavel, 19 de dezembro de 2025.

Neucir Szinwelski
Diretor Presidente

Maria Inês Presrlak
Gestora Sênior