



GOVERNO DO ESTADO DO ESPÍRITO SANTO
SECRETARIA DE ESTADO DE CONTROLE E TRANSPARÊNCIA

1. Dados da Reunião		
NOTA DE REUNIÃO Nº 004/2022 - CTSP	Convocação Secretário de Estado de Controle e Transparência	
Local Gabinete do Secretário	Data 05/07/2022	Horário 16h00 as 18h00

2. Participantes (Membros)			
Edmar Moreira Camata (Secretário)	X	Alexandre Del Santo Falcão (SUBINT)	X
Marcelo Campos Antunes (SUBCONT)	X	Patrícia Bravim Melotti (GTA)	X
Fabiano da Rocha Louzada (SUBTRAN)	X	Rodolfo Pereira Netto (coordenador)	X
Helmut Mutiz D`Auvila (COGES)	X	Emerson Couto de Moura (NTSI)	X

3. Participantes (convidados)			
Giselli Cardoso Alencastre Ballestrassi (SUBCONT)	X	Valber Pinheiro Padilha (SUBCONT)	X

4. Objetivo
Apresentar e deliberar sobre os assuntos referentes ao Comitê Gestor de Tecnologia, Segurança da Informação e Proteção de Dados Pessoais (CTSP)

5. Assuntos Discutidos
<p>Os assuntos discutidos na reunião foram apresentados pelo servidor da SECONT lotado no setor de suporte em TI Emerson Couto de Moura:</p> <ol style="list-style-type: none">1. APRESENTAÇÃO DO ESTUDO TÉCNICO PRELIMINAR para aquisição de uma Solução Integrada de Inteligência e Processamento Analítico conforme os projetos, demandas e necessidades identificadas na Secretaria de Controle e Transparência (SECONT) (Documento Anexo)2. Em razão da verificação do custo de investimento e custeio que seria demandado para a realização do projeto foi deliberado pela Comissão o agendamento de reunião com a Coordenadoria de Inovação e Desenvolvimento Tecnológico da Secretaria de Estado de Governo em conjunto com os representantes da Empresa SAS para discutir sobre a possibilidade de contratação conjunta por mais de um órgão do Estado, com hospedagem e custeio anual no PRODEST

6. Encaminhamentos da reunião		
Tarefa	Responsável	Prazo
Agendar Reunião com CIDT e SAS	Secretário	Próxima reunião

7. Encaminhamentos da reunião anterior que continuam PENDENTES		
Tarefa	Responsável	Prazo

8. Fechamento da Nota de Reunião	
Data 29/07/2022	Nome do Relator Rodolfo Pereira Netto



ESTUDO TÉCNICO PRELIMINAR

Utilização do Registro de Preços 005/2022, mantido pela Secretaria de Segurança Pública e Defesa Social (SESP), para aquisição de uma Solução Integrada de Inteligência e Processamento Analítico conforme os projetos, demandas e necessidades identificadas na Secretaria de Controle e Transparência (SECONT).

Vitória, Espírito Santo
30 de junho de 2022



SUMÁRIO

1. IDENTIFICAÇÃO.....	3
2. EQUIPE DE PLANEJAMENTO DA CONTRATAÇÃO.....	3
3. OBJETIVO DO ESTUDO.....	3
4. DEFINIÇÃO DAS NECESSIDADES DE NEGÓCIO.....	3
5. DEFINIÇÃO DAS NECESSIDADES TECNOLÓGICAS.....	6
6. ANÁLISE COMPARATIVA DE SOLUÇÕES.....	15
7. MODELO DE PRESTAÇÃO DE SERVIÇO E PROJETOS CANDIDATOS.....	19
8. ESTIMATIVA DE CUSTO TOTAL DA CONTRATAÇÃO.....	26
9. NECESSIDADE DE ADEQUAÇÃO DO AMBIENTE.....	27
10. CONTINUIDADE DO NEGÓCIO DURANTE E APÓS A EXECUÇÃO CONTRATUAL.....	27
11. RISCOS DA CONTRATAÇÃO.....	29
12. CONSTRUÇÃO DA CENTRAL ANALÍTICA SECONT.....	31
13. POSSIBILIDADE DE COMPARTILHAMENTO DE RECURSOS COM OUTROS ÓRGÃOS.....	32
14. DECLARAÇÃO DE VIABILIDADE DA CONTRATAÇÃO.....	34



1. IDENTIFICAÇÃO.

Requisitante:	Edmar Moreira Camata Secretário de Controle e Transparência
Responsável técnico:	Emerson Couto de Moura Analista de Tecnologia da Informação / Assessor Especial
Descrição resumida do estudo:	Avaliação da adequação técnica e econômica da utilização do Registro de Preços 005/2022, mantido pela Secretaria de Segurança Pública e Defesa Social (SESP), para aquisição de uma Solução Integrada de Inteligência e Processamento Analítico conforme os projetos, demandas e necessidades identificadas na Secretaria de Controle e Transparência (SECONT).
Descrição resumida da solução em análise:	Solução Integrada de Inteligência e Processamento Analítico capaz de coletar, armazenar, processar e analisar grandes volumes de dados para apoiar as atividades finalísticas dos diversos setores da Secretaria de Controle e Transparência (SECONT).

2. EQUIPE DE PLANEJAMENTO DA CONTRATAÇÃO.

Integrante responsável:	Emerson Couto de Moura
Setor:	Suporte Técnico
Matrícula:	2765403
E-mail:	emerson.moura@secont.es.gov.br
Telefone:	27 3636-5369

3. OBJETIVO DO ESTUDO.

O presente estudo tem por objetivo demonstrar a viabilidade técnica e econômica da utilização do **Registro de Preços 05/2022**, da Secretaria de Segurança Pública e Defesa Social, para a aquisição de uma Solução Integrada de Inteligência e Processamento Analítico, bem como fornecer informações necessárias para subsidiar o respectivo processo de adesão, conforme prevê o Inciso IX do Art. 6º da Lei nº 8.666/93 e a Norma de Procedimento SCL 015 - Adesão a Ata de Registro de Preços da Administração Estadual (Portaria nº 36-R/2019).

4. DEFINIÇÃO DAS NECESSIDADES DE NEGÓCIO.

Os avanços permitidos pela tecnologia da informação apresentam novas possibilidades e desafios, podendo alterar de modo significativo os processos de trabalho de uma organização. Um dos desafios do Controle Interno é justamente acompanhar tais avanços tecnológicos para realizar da melhor maneira possível suas diversas atividades, tais como: conformidade, suporte a decisão da gestão e detecção de fraudes.

Um argumento recorrente sobre as atividades de controle interno é que as verificações de conformidade consomem boa parte dos recursos disponíveis, não permitindo um aprofundamento em aspectos capazes de agregar mais valor às atividades. Uma forma de liberar tais recursos seria automatizar e tornar mais efetiva, por meio da ciência de dados, as atividades de conformidade, otimizando o tempo de trabalho dos envolvidos e maximizando os resultados.



A utilização intensiva de tecnologia também permite indicar, através de métodos preditivos, tendências que permitam à organização adaptar seu modelo de gestão em nível macro ou mesmo nos processos, tornando-os mais efetivos em termos de resultados.

Técnicas de ciência de dados também são aplicadas na detecção de fraudes, já que estas surgem com diferentes padrões e diferentes intensidades sendo difícil para o analista detectá-las e relacioná-las com efetiva precisão. Desta forma, os sistemas de detecção de fraudes são auto adaptáveis, em tempo real, para verificarem sinais comuns de alterações de comportamento no momento em que ocorrem, sugerindo ao analista eventos potencialmente sensíveis e com a garantia do menor tempo de latência para a atuação do controle interno.

Portanto, uma tendência nas organizações é investir cada vez mais nas análises e visualizações disponibilizadas pela ciência de dados, de modo a vislumbrar mais oportunidades de negócio e aumentar a eficiência e eficácia operacionais. **Da mesma forma, órgãos públicos, em especial os de controle, estão coletando grandes quantidades de dados de seus parceiros e auditados, além do feedback da sociedade nas redes sociais.**

Como evidência podemos relacionar várias técnicas disponibilizadas pela Ciência de Dados com aplicações associadas ao Controle Interno, conforme observamos nos trabalhos acadêmicos da tabela abaixo:

Técnicas de Ciência de Dados	Aplicações associadas ao Controle Interno
Aprendizado de Máquina (<i>Machine Learning</i>) e Aprendizado Profundo (<i>Deep Learning</i>) – <i>Grid Search, Feature Selection, Ensembled Methods, BBN, DNN.</i>	<ul style="list-style-type: none">• Suporte a decisão do auditor comparando resultados de <i>Ensembled Methods</i> e <i>Bayesian Belief Networks</i>, para detecção de fraudes em declarações financeiras (HAJEK e HENRIQUES, 2017);• Comparação de métodos de aprendizagem com classificação (<i>K2, Naive Bayesian, Tree Augmented Naive Bayes, Decision Tree</i>) para detecção de fraudes em cartão de crédito (YEE, SAGADEVAN e MALIM, 2018);• Comparação de métodos supervisionados e não-supervisionados (<i>SVM, Logistic Regression</i>) na detecção de fraudes na área médica (BAUDER e KHOSHFOGTAAR, 2017);• Comparação de técnica econométricas (<i>logit</i>) e de redes neurais artificiais utilizando dados de auditoria para detectar problemas financeiros latentes (SANCHEZ, MONELLOS e LOPEZ, 2012).
Mineração de Texto (<i>Text Mining</i>), <i>Keyword Extraction, Sentiment Analysis, Processamento da Linguagem Natural.</i>	<ul style="list-style-type: none">• Avalia qualidade de relatórios de controle interno usando técnicas de MT (<i>Vector Space Model, Feature Selection, Linear Regression, Principal Component Analysis</i>) (BOSKOU, KIRKOS e SPATHIS, 2018);• Aplicação de Processamento da Linguagem Natural (NLP) para avaliar relatórios anteriores para obtenção de <i>insights</i> (FISHER, HUGHES e GARNSEY, 2016).
Agrupamento (<i>Clustering</i>), <i>K-Means, Hierárquico, Medidas de Distâncias.</i>	<ul style="list-style-type: none">• Comparação entre métodos de agrupamento em auditorias (<i>k-means, Ward</i>) (BYRNES, 2015);• Exemplifica como o Big Data aplicado em outras áreas pode ser aproveitado pela auditoria, incluindo técnicas de agrupamento e mineração de texto (CAO, CHYCHYLA e STEWART, 2015).
Classificação – <i>Árvore de Decisão, SVM, ANN, Naive Bayes, Random Forest.</i>	<ul style="list-style-type: none">• Comparação entre <i>Árvore de Decisão (Random Forest), SVM e ANN</i> para priorização na etapa de Planejamento de Auditoria (NASCIMENTO, SANTOS, et al., 2018);• Comparação de métodos (<i>Bagging, Gradient Boost, Naive Bayes, Random Forest</i>) para detecção de fraudes (PACHECO JR., 2019).
Regressão – <i>Regressão Linear e Logarítmica, Séries Temporais.</i>	<ul style="list-style-type: none">• Comparação entre modelos de predição aplicados a auditoria (<i>ARIMA e Least Square</i>) (GABER e LUSK, 2017);• Apresenta aplicações de séries temporais em <i>Big Data</i>, incluindo auditoria contínua (REZAEI, DORESTANI e ALIABADI, 2017) e fraudes (REZAEI, DORESTANI e ALIABADI, 2018).



Da mesma forma, podemos relacionar diversos projetos envolvendo Ciência de Dados e Inteligência Artificial em operação e produzindo resultados no poder público, conforme foi destacado no 7º SEMINÁRIO INTERNACIONAL SOBRE ANÁLISE DE DADOS NA ADMINISTRAÇÃO PÚBLICA (2021):

Projetos	Profissionais e órgãos responsáveis
Uso de modelos preditivos de <i>machine learning</i> para seleção de amostra de auditoria de processos de pedidos de transposição de servidores oriundos dos territórios federais do AP, RO e RR.	Leonardo Amato Loriato, auditor federal de controle externo e assessor para análise de dados no Tribunal de Contas da União (TCU).
Ferramenta de Análise de Risco de Ouvidoria: processamento de linguagem natural para automatizar a análise de denúncias.	Fernando Sola Pereira, cientista de dados na Coordenação-Geral de Inteligência de Dados na Controladoria-Geral da União (CGU).
Automatização da análise textual e registro de acordãos do TCU.	Victor Lahiri Hart, auditor federal de controle externo do Tribunal de Contas da União (TCU).
Deteção de sobrepreço de compras públicas a partir de notas fiscais.	Krerley Oliveira, coordenador do Laboratório de Estatística e Ciência dos Dados da Universidade Federal de Alagoas (UFAL).
Uso de clusterização para identificação de similaridade em produtos adquiridos pela administração pública, visando a deteção de anomalias para subsidiar a atuação dos órgãos de controle.	Rogério Oliveira de Jesus, Coordenador do Núcleo de Informações Estratégicas do Tribunal de Contas do Espírito Santo (TCE-ES).
Aplicação de inteligência artificial na priorização de análise de atos de pessoal da Administração Pública Federal.	Aloisio Dourado Neto, auditor federal de controle externo do Tribunal de Contas da União (TCU).
Utilização de análise de séries temporais na previsão do percentual de gastos com pessoal.	Isabella de Albuquerque Ceravolo, assessora no Núcleo de Gestão de Dados do Tribunal de Contas do Espírito Santo (TCE-ES).
Uso de XAI (Inteligência Artificial Explicável) na construção de políticas públicas baseadas em evidências.	Maxwell Sarmento de Carvalho, servidor na Agência Nacional de Energia Elétrica (ANEEL).

A Solução Integrada de Inteligência e Processamento Analítico proposta neste estudo busca permitir através de diversas técnicas de ciências de dados analisar uma maior quantidade de informação e em maior frequência, provendo auditorias e outros procedimentos de controle mais abrangentes e em menor espaço de tempo.

Com a utilização de maior quantidade de dados, o processo de previsão dos riscos se torna mais preciso e possibilita automatizar processos repetitivos, deixando-os mais ágeis e liberando os auditores e demais servidores de processos manuais que consomem muita força de trabalho. **Conforme preconiza o IIA (2017), além de trazer mais eficiência e maior garantia nas auditorias, a ciência de dados permite a entrega de insights estratégicos para agregar valor à organização, auxiliando a gestão.**

Os requisitos e funcionalidades de cada projeto envolvendo a solução proposta será descrito oportunamente neste estudo, mas em linhas gerais é prioritário que a Solução Integrada de Inteligência e Processamento Analítico permita à SECONT:

- Criar um *Data Lake* (Central de Dados) para armazenar e integrar as bases de dados estruturadas e não estruturadas mantidas e custodiadas pela SECONT.
- Automatizar, via APIs ou conexões nativas aos SGBDs, a ingestão de dados em bases de análise a partir de fontes relacionais diversas ou de arquivos, tais como: bases de dados dos sistemas corporativos estaduais (SIGA, SIGEFES, SIARHES, etc.) e bases disponíveis através dos contratos de parceria de dados da Junta Comercial e da Receita Federal; bem como de fontes não relacionais, como o Portal da Transparência e o sistema E-DOCS, não se limitando aos aqui citados.



- Criar mecanismos de apoio às atividades de controle interno, transparência, integridade e correição com o uso de inteligência artificial para análise de informações em bases de dados ou e em processos administrativos buscando a identificação de correlações e anomalias, bem como o desenho de redes de relacionamentos que enriqueçam as análises efetuadas.
- Criar painéis dinâmicos e publicáveis na Internet (*dashboards*) que permitam a visualização em níveis dos relacionamentos, correlações, modelos multidimensionais e demais funcionalidades necessárias às atividades da SECONT.
- Criar modelos multidimensionais a partir das bases de dados disponíveis no *Data Lake* ou consultadas via API, tais como: SIGA; SIARHES; SIGEFES; IBGE; dados de pessoas físicas da Receita Federal; CFM; CAU; CFO; SUSEP; CADASTUR; CFC; ANTT; CVM; IBAMA; renda estimada; Bolsa Família; IRPF; seguro desemprego; CNEP; CEIS; BACEN; COAF; CEAFF; PGFN/DAU; Razão Social e CNPJ (Cadastro Nacional da Pessoa Jurídica) da Receita Federal; CNAE (Classificação Nacional de Atividades Econômicas); Cadastro Fiscal (Simples Nacional, Sintegra, ISS); SICAF - Sistema de Cadastramento Unificado de Fornecedores.

5. DEFINIÇÃO DAS NECESSIDADES TECNOLÓGICAS.

5.1. Objeto.

Aquisição de uma Solução Integrada de Inteligência e Processamento Analítico composta de: a) fornecimento de licenças de software para análise preditiva, ciência de dados e aprendizado de máquina; b) serviços de implantação, suporte, treinamento e desenvolvimento de soluções analíticas.

O objeto é composto pelos itens a seguir, conforme as condições e descrições que constam na especificação detalhada presente nos requisitos tecnológicos:

ITEM 1 – Licenças perpétuas de uma Solução Integrada de Inteligência e Processamento Analítico que permita: manipulação de grandes massas de dados; detecção e gestão de casos de fraudes; inteligência artificial; previsões e tendências analíticas; produção de relatórios, gráficos, tabelas e painéis de análise preditiva e estatística; capacidade de traçar cenários e encontrar padrões entre ocorrências definidas. Deve possuir garantia de atualização e suporte técnico do fabricante pelo prazo de 12 (doze) meses.

ITEM 2 – Serviço de suporte técnico, manutenção e direito de atualização de versão da Solução do **ITEM 1** durante o período de 12 (doze) meses, para o segundo ano do contrato.

ITEM 3 – Subscrição de plataforma de computação distribuída, baseado em tecnologia **Hadoop** durante o período de 12 (doze) meses.

ITEM 4 – Instalação, configuração, customização, testes, documentação da implantação, transferência de conhecimento e conexão inicial às fontes internas e externas de dados que se encontrem disponíveis.

ITEM 5 – Serviço de treinamento na administração da solução contratada.

ITEM 6 – Serviço de treinamento em **Big Data** e Ciência de Dados aplicada na solução contratada.

ITEM 7 – Contratação sob demanda de **57.518** (cinquenta e sete mil quinhentos e dezoito) “Unidades de Serviços Técnicos” (UST) para atendimento *on-site* (no local) durante o período de 12 (doze) meses. As USTs serão consumidas pelos serviços de instalação, configuração, customização, testes, documentação da implantação e transferência de conhecimento das soluções previstas no ITEM 1 e ITEM 3, bem como pelos serviços de treinamento na Plataforma Analítica e nos conceitos de Ciência de Dados e **Big Data**.



5.2. Requisitos tecnológicos.

Seguem os requisitos tecnológico gerais e específicos que a Solução Integrada de Inteligência e Processamento Analítico deve oferecer.

5.2.1. Requisitos gerais.

- 5.2.1.1. A solução deve possuir arquitetura segmentada em pelo menos 2 camadas, sendo uma camada cliente e uma ou mais camadas do tipo servidor. As camadas do tipo servidor podem ser instaladas em um único servidor ou cada camada em um servidor específico.
- 5.2.1.2. Todos os componentes das camadas de servidor da solução devem ser passíveis de serem instalados em servidores de rede físicos e virtuais e compatíveis com pelo menos um dos seguintes sistemas operacionais:
 - a. Windows Server 2012 R2 ou superior;
 - b. Red Hat Enterprise Linux 6.3 ou superior; ou
 - c. Oracle Linux 6.6 ou superior.
- 5.2.1.3. Os módulos cliente, se necessários, devem executar em estações de trabalho com sistema operacional Windows 10 e versões superiores e suportar as seguintes opções de regionalização:
 - a. Inglês Americano (EN-US);
 - b. Português Brasileiro (PT-BR).
- 5.2.1.4. A interface WEB, quando presente, deve executar em estações de trabalho com sistema operacional Windows 10 e suportar os seguintes navegadores:
 - a. Mozilla Firefox (versões 24 ou superiores);
 - b. Google Chrome (versões 34 ou superiores);
 - c. Safari (versões 5 ou superiores).
- 5.2.1.5. A solução deve possuir mecanismos que permitam monitorar a execução dos serviços.
- 5.2.1.6. A solução deve possuir mecanismos para agendar execução de modelos em modo *batch* e monitorar e controlar sua execução com, no mínimo, as seguintes opções:
 - a. Agendar;
 - b. Iniciar;
 - c. Interromper;
 - d. Reiniciar;
 - e. Visualizar log de execução e de erros.
- 5.2.1.7. Todas as operações listadas neste estudo como requisitos da solução devem ser realizadas por meio de operações diretas em interface gráfica. Todavia, será permitida a utilização de linguagem de programação para customização e otimização avançada dos modelos estatísticos e de mineração de dados gerados por meio da interface gráfica.
- 5.2.1.8. A solução deve possibilitar a incorporação e execução de scripts desenvolvidos em linguagem de programação estatística R.
- 5.2.1.9. Os requisitos listados neste estudo devem ser atendidos, preferencialmente, de forma nativa pela solução. Caso algum requisito não seja atendido nativamente, será permitido a sua implementação



por meio de customização em linguagem R, limitado a 15% do total de requisitos. As funcionalidades implementadas via customização devem ser incorporadas de forma transparente na solução, de forma que fiquem disponíveis por meio da interface gráfica. A verificação da quantidade e qualidade dos requisitos implementados por meio de customização será feita na oportunidade da avaliação da amostra.

- 5.2.1.10. Os módulos componentes da solução devem funcionar de forma integrada e as informações sobre as bases de dados e modelos estatísticos e de mineração devem ser armazenados em um repositório de metadados comum, de modo que as informações e variáveis sobre determinada base criada em determinado módulo fiquem disponíveis nativamente para os demais módulos da solução.

5.2.2. Requisitos de segurança.

- 5.2.2.1. Possuir controle de usuários por autenticação e autorização.
- 5.2.2.2. Prover autenticação de usuários integrada com o serviço de diretório *Microsoft Active Directory (AD)*.
- 5.2.2.3. Possibilitar a criação de contas de usuário no ambiente analítico com diferentes níveis de acesso de administração e operação.
- 5.2.2.4. Possuir arquitetura própria de segurança, independentemente da utilização de soluções de segurança de terceiros.
- 5.2.2.5. Possibilitar controle centralizado de segurança para atribuição de privilégios a níveis de perfis de usuários, grupos e projetos.
- 5.2.2.6. Possibilitar restrição de funcionalidades da solução para um determinado usuário, grupo, projeto ou perfil.
- 5.2.2.7. Possibilitar restrição de acesso a objetos criados pelos desenvolvedores para um determinado usuário, grupo ou projeto.
- 5.2.2.8. Possibilitar o controle de acesso de acordo com o perfil do usuário.
- 5.2.2.9. Possuir controle de autorização de acesso a dados no grão de tabela ou visão.
- 5.2.2.10. Possuir controle de autorização de execução de tarefas no grão de projeto.
- 5.2.2.11. Possuir controle de autorização de edição de fluxos no grão de projeto, com: funcionalidades habilitadas ou desabilitadas por perfil ou usuário.
- 5.2.2.12. Possuir mecanismos de monitoração de tarefas e funções executadas (*logs*).

5.2.3. Processo para o desenvolvimento de modelos.

- 5.2.3.1. Permitir a orientação do desenvolvimento de soluções por projetos.
- 5.2.3.2. Permitir trabalho colaborativo.



- 5.2.3.3. Permitir importação e exportação de modelos estatísticos e de mineração de dados em formato *Predictive Model Markup Language* (PMML 4.2).
- 5.2.3.4. A solução deve possuir mecanismos próprios de controle de versão dos modelos estatísticos e de mineração de dados ou permitir a realização do controle de versão dos modelos analíticos usando ferramentas especializadas como CVS, Git ou SVN.
- 5.2.3.5. A solução deve possibilitar o monitoramento da performance dos modelos analíticos, incluindo comparação entre diferentes modelos por meio de métricas como KS e Gini, *Deviation Index*.

5.2.4. Acesso e armazenamento de dados.

- 5.2.4.1. A solução deve permitir o armazenamento e recuperação de dados em:
 - a. Sistema de Gerenciamento de Banco de Dados (SGBD) incluindo, no mínimo, os seguintes conectores:
 - 1. Oracle;
 - 2. SQL Server;
 - 3. Conector genérico ODBC ou JDBC.
 - b. Em arquivos texto, .csv, XML.
 - c. Em arquivos Microsoft Office xls,xlsx e mdb.
 - d. Armazenados em tecnologia de dados massivos Hadoop (HDFS).
- 5.2.4.2. A solução deve permitir que múltiplos arquivos de dados possam ser carregados no software.

5.2.5. Preparação de dados.

- 5.2.5.1. Deve permitir importação e exportação de dados.
- 5.2.5.2. Deve permitir a realização de operações de *merge* e *append*.
- 5.2.5.3. Deve permitir a eliminação automática de valores raros para variáveis de classificação cujas ocorrências sejam menores do que um número especificado.
- 5.2.5.4. Deve permitir o tratamento automático de valores faltantes com, no mínimo, as seguintes opções:
 - a. Substituição por valor default especificado pelo usuário;
 - b. Substituição pela média;
 - c. Substituição pela mediana;
 - d. Substituição pelo máximo;
 - e. Substituição pelo mínimo;
 - f. Eliminação da linha.
- 5.2.5.5. Redução de dimensão:
 - a. Seleção das variáveis a serem utilizadas no modelo, pelos métodos de coeficiente de correlação;
 - b. Remoção das variáveis com alta proporção de valores ausentes (*missing*);
 - c. Categorização de variáveis: em intervalos equiespaçados (*bucketed*), intervalos com frequências iguais (*quantile*) e categorização das variáveis levando em consideração o relacionamento com a variável *target*.
 - d. Eliminação automática de valores extremos utilizando:
 - 1. Número de desvios em torno da mediana;



2. Número de desvios em torno da média;
3. Distância em relação à moda central;
4. Percentis extremos;
5. Eliminação manual.

5.2.5.6. Transformações:

- a. Logarítmica;
- b. Raiz quadrada;
- c. Inversa;
- d. Quadrática;
- e. Exponencial;
- f. Padronizada.

5.2.6. Análise descritiva e exploração de dados.

5.2.6.1. Distribuições de frequência simples.

5.2.6.2. Medidas de posição: média, mediana, moda, quartis, decis, percentis, máximo e mínimo.

5.2.6.3. Medidas de dispersão: desvio padrão, variância.

5.2.6.4. Quantidade (n).

5.2.6.5. Medida de curtose.

5.2.6.6. Medida de assimetria (*Skewness*).

5.2.6.7. Histograma.

5.2.6.8. Ramos e folhas.

5.2.6.9. *Box Plot*.

5.2.6.10. Teste de hipóteses: teste T, teste F.

5.2.6.11. Intervalos de confiança/cálculo do *p-value*.

5.2.6.12. Coeficiente de Correlação de Pearson.

5.2.6.13. Testes para Correlação de Pearson.

5.2.6.14. QQ Plots.

5.2.7. Gráficos e visualizações.

5.2.7.1. Elaboração de gráficos de forma interativa/assistente (por menu ou comando) para os seguintes tipos:

- a. Setores (Pizza);
- b. Linhas;
- c. Colunas (Barras);



- d. *Boxplot* (diagrama de caixa);
- e. *Scatterplot* (diagrama de dispersão);
- f. Pareto;
- g. Gráfico de Bolha;
- h. Histograma;
- i. *Treemap* (mapa de árvore);
- j. *Heatmap* (mapa de calor);
- k. Superfície.

5.2.7.2. Opções entre diferentes estilos de cor.

5.2.7.3. Configurar títulos e notas de rodapé.

5.2.7.4. Gerar e exportar gráficos para outros aplicativos nos seguintes formatos:

- a. JPEG ou PNG;
- b. PDF.

5.2.8. Amostragem e filtros.

5.2.8.1. Seleção de registros com utilização de variáveis, operadores lógicos e funções.

5.2.8.2. Seleção de Amostra de Treinamento e Teste com parametrização, a partir da partição dos dados originais, utilizando:

- a. Amostragem aleatória simples;
- b. Amostragem estratificada;
- c. Amostragem balanceada (Ponderada);
- d. Primeiras N observações do conjunto de dados.

5.2.9. Modelagem estatística.

5.2.9.1. Modelo de Regressão:

- a. Regressão Simples;
- b. Regressão Múltipla;
- c. Regressão Não-Linear;
- d. Regressão Multinomial;
- e. Regressão Ordinal;
- f. Regressão *Probit*;
- g. Regressão *Logit*;
- h. As rotinas devem contemplar, quando necessário, opção para métodos de seleção de variáveis do tipo *Stepwise: Forward e Backward*;
- i. Os relatórios de saída contêm opções para exibir/imprimir informações estatísticas para avaliação do modelo quanto a:
 - 1. Diagnóstico de multicolinearidade;
 - 2. Análise de variância/Ajuste de modelo;
 - 3. Análise dos resíduos.

5.2.9.2. Análise de Cluster:

- a. Construção de um número pré-definido de clusters e segmentação automática (definição automática do número ideal de clusters);



- b. Utilização de métodos hierárquicos;
- c. Utilização do algoritmo K-MEANS.

5.2.9.3. Análise de Variância:

- a. Modelo de Efeito Fixo e Aleatório:
 - 1. Análise de Variância Univariada;
 - 2. Análise de Variância com dois fatores.
- b. Utilização dos métodos ANCOVA/MANCOVA.
- c. Utilização dos métodos ANOVA/MANOVA.

5.2.9.4. Análise Multivariada:

- a. Análise de componentes principais;
- b. Análise fatorial;
- c. Análise discriminante;
- d. Correlação Canônica;
- e. Análise de Regressão multivariada.

5.2.9.5. Análise de Dados Categóricos:

- a. Modelos Log-lineares;
- b. Métodos de *Mantel-Haenszel*;
- c. Modelos de Decisão Discreta.

5.2.9.6. Estatística não-paramétrica:

- a. Teste *Kolmogorov-Smirnov* para uma amostra independente;
- b. Comparação de duas ou mais amostras independentes;
- c. Tabelas 2x2 - Teste Quiquadrado;
- d. Testes de correlação: *Kendall* e *Spearman*.

5.2.9.7. Análise de Sobrevivência:

- a. Regressão de Cox;
- b. *Life Table*;
- c. *Kaplan-Meier*.

5.2.9.8. Análise Psicométricas:

- a. Escalonamento multidimensional;
- b. Análise de correspondência;
- c. Análise conjunta.

5.2.10. Análise de séries temporais.

5.2.10.1. Alisamento exponencial.

5.2.10.2. ARIMA (*Box-Jenkins*) e seus derivados.

5.2.10.3. Decomposição de séries temporais e ajustamento sazonal.

5.2.10.4. Previsão pontual e intervalar.

5.2.10.5. Dados de Painel.



- 5.2.10.6. Análise Espectral.
- 5.2.10.7. Regressão Polinomial.
- 5.2.10.8. Estatísticas para identificação do melhor modelo:
 - a. R²; Erro absoluto médio;
 - b. Estatística do ajuste;
 - c. Função de auto correlação;
 - d. Função de auto correlação parcial;
 - e. Estimação dos parâmetros.

5.2.11. Mineração de dados.

- 5.2.11.1. Permitir a geração de modelos preditivos Suporte a variáveis de entrada "input" e variáveis resposta "target": binárias, nominais, ordinais, intervalares e contínuas.
- 5.2.11.2. Permitir a criação e edição visual de fluxos de mineração de dados.
- 5.2.11.3. Permitir a criação e gerenciamento de projetos em mineração de dados com processos do tipo batch e de tempo real a partir da gravação de arquivos de saída em banco de dados e pela geração de código fonte do modelo em JAVA ou SQL.
- 5.2.11.4. Permitir a comparação dos modelos produzidos – Avaliação de múltiplos modelos em um único framework para todas as amostras (treinamento e teste), selecionando automaticamente o melhor modelo baseado nos critérios: Curvas ROC e Matriz de confusão.
- 5.2.11.5. Árvores de Decisão:
 - a. Árvores de regressão e classificação, tipo *chaid*, C&RT e regressora;
 - b. Métodos avançados para seleção da árvore baseado em medidas de ganho;
 - c. Critérios de divisão: testes de probabilidade de qui-quadrado e F, Gini, Entropia, redução da variância;
 - d. Crescimento/poda interativa;
 - e. Diagrama de árvore editável, imprimível e exportável.
- 5.2.11.6. Redes Neurais:
 - a. Possuir arquitetura de rede flexível com:
 - 1. Parametrização de funções de ativação (mínimo Linear e Sigmóide);
 - 2. Padronização automática das entradas;
 - 3. Treinamento interativo;
 - 4. Técnica que utiliza o algoritmo do vizinho mais próximo "*k-nearest neighbour*" para categorização ou predição de observações.
 - b. Arquiteturas de redes MLP - *Multi Layer Perceptron* e *Radial Basis Function*;
 - c. Definição automática da construção da rede neural *Perceptrons* em várias camadas com *backpropagation* para a configuração ótima;
 - d. Redes *Kohonen*.
- 5.2.11.7. Regras de Associação:
 - a. Algoritmos de descoberta de regras de associação;
 - b. Algoritmos de descoberta de sequência (análise sequencial);
 - c. Estatísticas das regras: acurácia (*lift*), confiança (*confidence*) e suporte (*support*) das regras.



- 5.2.11.8. Outros Algoritmos de *Machine Learning*:
- SVM - Support Vector Machine;
 - K-Nearest neighbors;
 - Random Forrest;
 - Gradient Boosting;
 - Bayesian Network.
- 5.2.11.9. Séries Temporais:
- Calculo automático de previsões a partir de múltiplos modelos efetuando a comparação entre eles e identificando o melhor modelo de forma parametrizada e automática, sem necessidade de programação.
- 5.2.11.10. Geração Automática de Modelos:
- Possuir técnicas automatizadas de modelagem e preparação de dados para acelerar o processo de criação de modelos contemplando particionamento dos dados de entrada entre treinamento e validação, substituição de *missing values* e seleção de variáveis;
 - Permitir comparar de forma simples os modelos gerados para que seja possível eleger o modelo mais adequado.

5.2.12. Análise de redes de relacionamentos.

- 5.2.12.1. Permitir a representação visual interativa e dinâmica de redes de relacionamento.
- 5.2.12.2. Possibilitar a identificação de relações de parentesco por associação de sobrenomes utilizando algoritmos fonéticos.
- 5.2.12.3. Possibilitar a identificação de relações diretas e indiretas entre entidades, como por exemplo pessoas físicas e jurídicas (exemplo: pessoa A tem participação na empresa B que tem participação na empresa C e que tem como sócio a pessoa D).
- 5.2.12.4. Permitir ao usuário registrar comentários sobre os elementos da rede.
- 5.2.12.5. Analisar os dados e cria um modelo de dados de todas as entidades e atribui sua ligação fundamental.
- 5.2.12.6. Permitir a combinação de entidades altamente ligadas, mesmo que os dados possuam má qualidade e não sejam relacionados diretamente.
- 5.2.12.7. Suportar atualização incremental nas redes com adição de novos dados.
- 5.2.12.8. Interface de visualização de rede permitir que os auditores vejam as conexões da rede para que eles possam descobrir relações anteriormente desconhecidas e conduzir investigações mais eficazes e eficientes.
- 5.2.12.9. Possibilitar descobrir relações por meio de geolocalização (por exemplo, endereços próximos).
- 5.2.12.10. Geração automática de diagramas de rede, permitindo que os auditores vejam padrões e características que podem levar à implementação de controles melhorados e novas técnicas de monitoramento.



- 5.2.12.11. Permitir aplicar cores, ícones ou outros elementos de diferenciação para representar propriedades de uma entidade.
- 5.2.12.12. Permitir agrupar visualmente itens com propriedades comuns.
- 5.2.12.13. Permitir correlacionar de maneira gráfica e manual elementos da rede, apenas ligando uma entidade a outra, de forma que esse conhecimento seja considerado para o processo de criação de novas redes.
- 5.2.12.14. Prover algoritmos e técnicas de *Social Networking Analysis* (SNA), incluindo:
 - a. Detecção de laços fortes e fracos (*strong/weak ties*);
 - b. Homofilia e afiliação;
 - c. Análise de *links*, com detecção de *hubs* e *authorities* (*PageRank*);
 - d. *Power laws*;
 - e. Modelagem de *small worlds*.
- 5.2.12.15. Possui as seguintes métricas de centralidade de rede:
 - a. Intermediação (*betweness*);
 - b. Proximidade (*closeness*);
 - c. Grau (*degree*);
 - d. *Eigenvectors*.

5.2.13. Exportação.

- 5.2.13.1. Permite a exportação dos resultados dos processamentos analíticos nos seguintes formatos:
 - a. Bancos de dados;
 - b. *Portable Document Format* (.pdf);
 - c. Arquivos texto (.csv).

6. ANÁLISE COMPARATIVA DE SOLUÇÕES.

6.1. Identificação de soluções.

Em 2019 a Secretaria de Estado do Governo (SEG) realizou uma Consulta Pública para a contratação de uma Solução Integrada de Inteligência e Processamento Analítico através da Coordenadoria de Inovação e Desenvolvimento Tecnológico (CIDT), na oportunidade o objetivo era definir uma solução para apoiar a melhoria da gestão do Governo do Espírito Santo.

Para tanto foi utilizada a consultoria do *Gartner Group* no intuito de definir os requisitos mínimos que a solução pretendida deveria oferecer, bem como identificar os potenciais fornecedores da solução de Inteligência Analítica pretendida. O *Gartner* é um dos líderes mundiais em pesquisa e consultoria em tecnologia da informação, sua missão consiste em gerar informações, métricas e análises a respeito de soluções tecnológicas para que seus clientes tomem decisões estratégicas de aquisição.

O *Gartner* publica regularmente estudos técnicos chamados de **Quadrantes Mágicos**, que avaliam provedores e soluções de tecnologia para as mais diversas aplicações. Os quadrantes mágicos apresentam uma visão de alto nível dos principais provedores e suas soluções tecnológicas, o estudo demonstra como as soluções estão posicionadas para atender uma organização a longo prazo. Para isto, o *Gartner* aplica um conjunto uniforme de



critérios de avaliação que consideram a visão e capacidade de execução dos competidores, e os posicionam relativamente em um gráfico nas seguintes posições: *Líderes; Visionários; Fornecedores de nicho; Desafiantes.*

À época da Consulta Pública realizada pela SEG / CIDT diversas empresas ofereciam soluções para o mercado de **Plataformas de Ciência de Dados e Aprendizado de Máquina**, tais como: **TIBCO, IBM, Microsoft, SAS, Alteryx, RapidMiner, Databricks**, dentre outras. Os requisitos técnicos definidos pela SEG / CIDT foram submetidos ao mercado e ao fim da Consulta Pública não houve qualquer questionamento que indicasse a existência de algum item que porventura estivesse afastando algum fabricante da possibilidade de participar do certame licitatório.

Todos os fornecedores identificados possuem casos de sucesso no Brasil e no mundo, e ainda agregam uma infraestrutura de parceiros prestadores de serviços técnicos especializados ampla e variada, possibilitando um atendimento com níveis de serviços satisfatórios e competitivos. Percebe-se, ainda, que a relação de empresas não se alterou muito desde a realização da Consulta Pública, conforme observamos no Quadrante Mágico atualizado na figura abaixo.

Figure 1: Magic Quadrant for Data Science and Machine Learning Platforms



Source: Gartner (March 2021)

Pelo exposto, o presente estudo reaproveita os requisitos estabelecidos pela SEG / CIDT e posteriormente pela SESP, atualizando alguns aspectos tecnológicos que não demandariam a realização de uma nova Consulta Pública ou pesquisa de mercado para verificação da sua aceitação. **Foram identificadas soluções Open Source que atenderiam nativamente a alguns requisitos técnicos definidos, mas que demandariam um grande volume de programação para atender os demais requisitos necessários ao atendimento das necessidades da SECONT.**

Em especial, considerou-se a utilização do ecossistema Apache para prover os recursos de análise e armazenamento de dados de forma distribuída, mas tal abordagem demandaria uma curva de aprendizado considerável e praticamente inviável em função da força de trabalho disponível atualmente no órgão.



Dessa forma, a decisão pela utilização ou não de soluções *Open Source* levou em consideração os seguintes aspectos:

1. Aspectos positivos.

- a. **Custo da licença:** sendo composto por códigos abertos, o custo de aquisição pode ser gratuito ou não. A maioria das licenças “*enterprise*” que teriam as qualificações mínimas para atender um cliente do porte da SECONT, são onerosas, o que mitiga o impacto deste aspecto positivo;
- b. **Custo do suporte:** da mesma forma, o uso de licenças “*enterprise*” significará dispêndios para suporte e garantia de versão, também mitigando o impacto positivo deste aspecto;
- c. **Base de conhecimento:** como premissa, há a necessidade de conhecimento de linguagem de programação, principalmente R ou Python. Detendo esse conhecimento é possível criar novas funcionalidades e compartilhar livremente com a comunidade de usuários e desenvolvedores;
- d. **Compartilhamento de informações:** soluções *Open Source* são largamente utilizadas em meios acadêmicos e por usuários não corporativos, fazendo com que a base de usuários existentes seja bem razoável, possibilitando o compartilhamento de informações entre os usuários das soluções e promovendo fóruns de discussão com ajuda mútua.

2. Aspectos Negativos.

- a. **Resiliência de implementações estruturadas:** ainda que os usuários possam criar diversas aplicações com a utilização de *Open Source* mesmo para processos críticos é difícil manter a governança e garantir o devido funcionamento da plataforma em ambiente de produção;
- b. **Custos de implementação:** por não contar com funcionalidades de apoio ao usuário final, baseando-se em sua maior parte em código, a implementação demora mais tempo e será mais custosa que a implementação de softwares de fornecedores que já detêm essa capacidade. Não há possibilidade de utilizar a solução sem conhecimento de R ou Python, devendo este aspecto ser considerado como custo de implementação, bem como qualquer custo de consultoria ou de suporte necessários para a implementação da solução;
- c. **Sustentação:** após o processo de implementação, manutenções realizadas em código *Open Source* podem se tornar complexas pela dificuldade de interpretação de técnicas de desenvolvimento anteriores e por atualizações de versões que não garantem a compatibilidade com versões antigas do software;
- d. **Suporte:** em caso de falhas ou erros gerados devido a bibliotecas mal implementadas ou códigos fonte não confiáveis, não há suporte de fabricante para solicitar apoio;
- e. **Acoplamento:** é a dependência tecnológica entre a organização e a tecnologia utilizada por ela. Há um severo risco de dependência do usuário envolvido com uso da programação específica para a solução. Este acoplamento é reduzido quando há possibilidade de interação por meio de interface gráfica que provê facilidade cognitiva. Devido às constantes inovações em TI e, conseqüentemente, das ferramentas *Open Source*, estas podem ser descontinuadas pelo surgimento de novas ferramentas mais modernas;
- f. **Escalabilidade:** devido à difícil aplicação destes softwares em ambientes de produção, o uso massivo de grandes volumes de dados ou de processamento geram dificuldades de manutenção e dificultam a escalabilidade.

Considerando os aspectos gerais positivos e negativos acima expostos e ainda a situação da SECONT que possui uma estrutura insuficiente de servidores do órgão com conhecimento em Tecnologia da Informação, a adoção de softwares livres (*Open Source*) se mostrou inviável para os objetivos almejados. Ainda que uma consultoria especializada fosse contratada para construir com softwares livres toda a solução pretendida pela SECONT, o tempo necessário e os riscos envolvidos tornariam esta opção inaceitável.

Portanto, identificamos as seguintes empresas e soluções capazes de atender aos requisitos técnicos presentes neste estudo e necessários aos projetos e necessidades da SECONT:



Empresas	Soluções
Databricks	<ul style="list-style-type: none">• <i>Databricks for Data Science</i>• <i>Databricks Lakehouse Platform</i>
IBM	<ul style="list-style-type: none">• <i>IBM Data Science and Business Analytics Platform</i>• <i>IBM Watson</i>
RapidMiner	<ul style="list-style-type: none">• <i>RapidMiner Platform</i>
TIBCO	<ul style="list-style-type: none">• <i>TIBCO Data Science</i>• <i>TIBCO Data Science - Team Studio</i>• <i>TIBCO Spotfire</i>
Alteryx	<ul style="list-style-type: none">• <i>Alteryx Machine Learning</i>• <i>Alteryx Intelligence Suite</i>
SAS	<ul style="list-style-type: none">• <i>SAS Visual Data Mining and Machine Learning</i>• <i>SAS Visual Analytics</i>• <i>SAS Visual Forecasting</i>• <i>SAS Detection and Investigation</i>
Microsoft	<ul style="list-style-type: none">• <i>Azure Synapse Analytics</i>• <i>Azure Machine Learning</i>• <i>Azure Cognitive Services</i>

6.2. Disponibilidade de soluções similares em outro órgão do Poder Executivo estadual.

Na amplitude necessária e com todas as funcionalidades pretendidas não foi possível localizar soluções similares em outros órgãos do Poder Executivo estadual. Todavia, identificamos a existência de soluções SAS na PGE, SEGER, SEFAZ e DETRAN, além da contratação da SESP que originou este estudo.

6.3. Disponibilidade de softwares livre conforme descrito na Portaria SGD/ME nº 3/2019.

Em pesquisa no Catálogo de Software Público Brasileiro (<https://www.gov.br/governodigital/pt-br/software-publico/catalogo/catalogo>) não foi possível encontrar soluções livres capazes de atender, no todo ou em parte, o objeto descrito no presente estudo.

6.4. Aquisição da solução na forma de licenças “perpétuas” ou subscrição.

O modelo de aquisição de software conhecido como “licença perpétua” garante sua classificação orçamentária como uma despesa de investimento, já o modelo de aquisição de software como serviço ou subscrição, por sua vez, exige a classificação como despesa corrente ou de custeio.

Historicamente, o orçamento de custeio da SECONT é consideravelmente baixo e alterá-lo para absorver o pagamento periódico das despesas com a Solução Integrada de Inteligência e Processamento Analítico que se pretende adquirir dependeria de uma série de ajustes financeiros e orçamentários. Se considerarmos os constantes cenários de instabilidade orçamentária, os riscos de manter a solução pretendida no modelo de subscrição a sujeitaria a descontinuidade em função de algum contingenciamento.

Em função disso, a opção pelo licenciamento perpétuo com contratos anuais de manutenção e evolução tecnológica se mostra mais razoável diante de um cenário orçamentário incerto, bem como evita que uma solução estratégica da SECONT sofra interrupções em seu funcionamento caso o Estado passe por alguma restrição orçamentária.



6.5. Escolha da solução tecnológica definida no PE 003/2022 da SESP.

O processo E-DOCS 2021-TXF91, da Secretaria de Segurança Pública e Defesa Social, instruiu a aquisição de uma Solução Integrada de Inteligência e Processamento Analítico que se realizou através do Pregão Eletrônico 003/2022, ocorrido em 10/06/2022. A empresa VERT SOLUÇÕES EM INFORMÁTICA LTDA foi declarada vencedora do referido pregão oferecendo as soluções tecnológicas da empresa SAS.

A empresa SAS é líder mundial na área de processamento analítico, ela está presente em 150 países e tem suas soluções instaladas em mais de 80.000 empresas, governos e universidades. Das 100 primeiras empresas listadas no índice Fortune 500, 88 usam soluções SAS. Seu portfólio de soluções possui um vasto repertório de aplicações: detecção de fraudes, aprendizado de máquina, inteligência artificial, inteligência analítica, mineração de dados, mineração de textos, etc.

O modelo de negócio estabelecido pela SESP em sua contratação define uma forma de aquisição que permite ao cliente utilizar a totalidade das soluções disponíveis no produto *SAS Detection and Investigation for Government*, a depender da necessidade do negócio deste cliente. A precificação do licenciamento é feita por cada conjunto de "32 cores" (32 núcleos de processamento), sendo que os serviços de implantação, customização, treinamento e repasse de conhecimento são cobrados em separado.

O modelo definido pela SESP é vantajoso quando as demandas de negócios exigem a utilização de várias soluções distintas, em conjunto ou isoladamente, pois não será necessário contratar e licenciar diversas soluções para realizar o mesmo serviço. Da mesma forma, novos projetos podem se beneficiar da possibilidade de utilização de recursos não previstos nos projetos iniciais, algo muito comum com demandas de análise de dados constantemente mutáveis.

Cabe destacar, como será demonstrado ao longo deste estudo, que os projetos previstos pela SECONT cobrem uma parte considerável do portfólio de soluções da SAS, bem como existem demandas reprimidas e ainda não devidamente mapeadas que exigirão novos conjuntos de funcionalidades. Importante ressaltar que a SECONT utiliza soluções SAS desde 2014 em seu Laboratório de Dados, realizando cruzamento de informações que auxiliam diversos processos finalísticos do órgão.

Pelo exposto, e considerando que os softwares SAS em uso atualmente pela SECONT estão defasados, descontinuados e não podem mais ser renovados, a solução tecnológica proposta pela SESP através do Pregão 003/2022 atende todos os requisitos presentes no presente estudo técnico, com a vantagem de aproveitar o conhecimento técnico da equipe de auditores responsável pelas análises.

7. MODELO DE PRESTAÇÃO DE SERVIÇO E PROJETOS CANDIDATOS.

7.1. Justificativa de utilização da métrica UST e estimativa de projetos.

Há uma constante confusão quanto aos termos e métodos que devem ser utilizados em projetos quando tratamos de temas como ciências de dados, *big data*, *analytics* e *business intelligence*. Projetos de desenvolvimento analítico têm características diferentes daquelas encontradas em desenvolvimento de software clássico, demandando abordagens, processos, conhecimento e métricas próprias.

O desenvolvimento analítico envolve o uso de raciocínio lógico indutivo e dedutivo, e de pensamento crítico no exame dos dados e conceitos de negócio envolvidos. Isso implica na análise de evidências dos dados para a criação de generalizações daquilo que foi observado. Ainda, da identificação e compreensão de problemas, do desenvolvimento de teorias, testes de hipóteses e comunicação de resultados.



O método para decidir a melhor caracterização das relações do problema em um modelo analítico é o que se chama de desenvolvimento do modelo. Dentre os vários métodos disponíveis, o mais conhecido e utilizado é o **Cross Industry Standard Process for Data Mining (CRISP-DM)**.

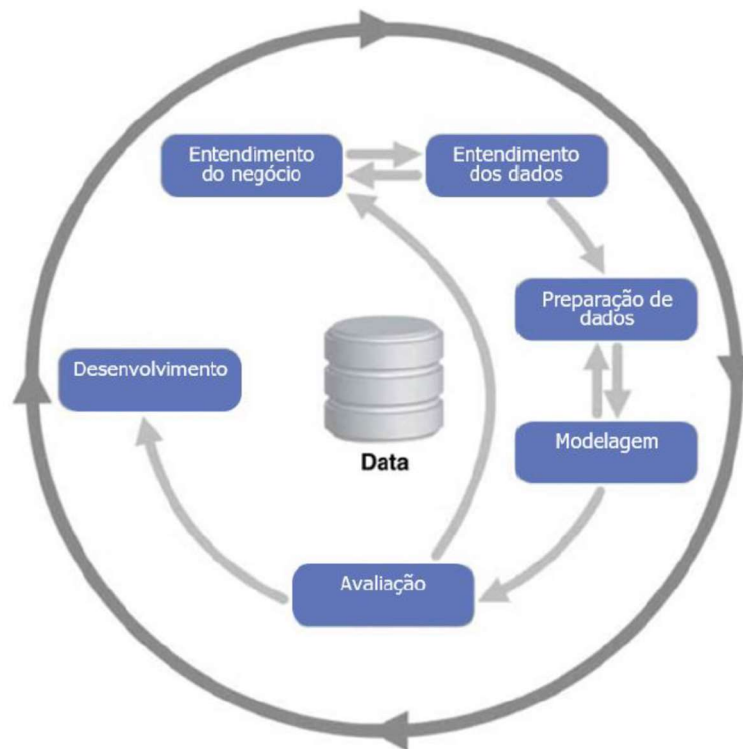


Figura 1 - Cross Industry Standard Process for Data Mining (CRISP-DM).

As fases envolvidas nas atividades analíticas não são compostas de passos rígidos e obrigatoriamente sequenciais. É normal e, esperado que, à medida que se explore os dados, a própria definição do problema em análise venha a ser refinada e as perguntas de negócios venham a ser realinhadas. Pressupostos iniciais podem ser trocados ou abandonados no processo e pode ser possível que o escrutínio dos dados mostre que a informação necessária não está disponível, obrigando a repensar a abordagem do problema de forma mais ampla.

Daí que a natureza de desenvolvimento analítico é distinta daquela encontrada no desenvolvimento de software clássico, aplicado em fábricas de software. Os conceitos e métodos aplicados neste último, não se adequam bem a *Analytics*, de forma que **as técnicas de mensuração de esforço baseadas em Pontos de Função não podem ser aplicadas de maneira realista ou adequada ao desenvolvimento de soluções analíticas.**

Dessa forma, se faz necessário buscar outras métricas de mensuração de esforço que façam maior sentido no contexto investigativo de *Analytics*. **O método mais adequado apresentado até o momento para projetos e desenvolvimento de produtos dessa natureza se baseia na mensuração por UST (Unidades de Serviço Técnico),** que deve ser definida necessariamente atrelada a um catálogo de serviços, com as atividades que fazem parte do ciclo de vida do desenvolvimento analítico estabelecidas e enumeradas, bem como, do estabelecimento do grau de complexidade dessas atividades.

A contratação de serviços utilizando a modalidade de UST, difere da contratação de mão de obra e possibilita a indexação do valor dos serviços em termos de alocação de esforço profissional, de competências e experiências variadas, conforme a complexidade das atividades. A estimativa deve partir de um levantamento



de esforços prévio à contratação e, precedido de um maior detalhamento dos serviços quando da abertura das Ordens de Serviços, **sendo o mesmo critério sido utilizado pelo STJ e aprovado pelo TCU, conforme TC 028.577/2011-6.**

Neste sentido, efetuamos um levantamento prévio de todos os projetos que serão desenvolvidos durante a vigência contratual que será prevista em um eventual processo de contratação. Aprovada a aquisição, estes projetos serão indicados no Termo de Referência que solicitará a adesão da SECONT ao Registro de Preços da SESP, sendo que o levantamento efetuado será a base para a estimativa do esforço que será necessário para o desenvolvimento e implantação destes projetos.

Cabe destacar que **no relatório do TCU – TC 002.116/2015-4 fica demonstrado que não há obrigatoriedade de utilização da métrica de contagem por pontos de função**, fato consolidado na Súmula-TCU 269. **No supracitado relatório são citados casos com mensurações por UST e USTIBB**, sendo a USTIBB uma derivação da UST, aplicada pelo Banco do Brasil, **citada como Boa Prática de mensuração e, sem contrariedade à jurisprudência daquele Órgão**. O relatório versa que **“Tais fatos permitem concluir que a obrigação é de que sejam usados critérios objetivos e baseados em resultados, não exclusivamente a Análise de Pontos de Função”**.

Sendo aprovada a presente contratação, será adotada a **METODOLOGIA PARA DESENVOLVIMENTO ANALÍTICO** desenvolvida pela SESP no Pregão 003/2022 (Anexo IV), tal metodologia será utilizada para a construção do **Catálogo de Serviços** que relacionará as atividades que deverão ser executadas em cada fase do desenvolvimento de projetos analíticos e sua respectiva complexidade, evidenciando claramente o esforço esperado para a execução do projeto e os resultados previstos em cada etapa.

A referida METODOLOGIA atende aos requisitos recomendados pelo TCU, pois nela consta:

- A definição de todas as atividades que poderão ser demandadas ao longo da contratação;
- A descrição das atividades e sua complexidade, indicando o tempo exigido para sua execução e a qualificação técnica do profissional responsável por sua execução;
- Os resultados esperados e os entregáveis de cada etapa do ciclo de desenvolvimento analítico;
- Os padrões de qualidade exigidos em cada etapa do ciclo de desenvolvimento analítico;
- A descrição dos procedimentos necessários para execução das atividades.

A partir da METODOLOGIA proposta pela SESP foi possível estimar **preliminarmente** o esforço necessário para o desenvolvimento dos projetos candidatos da SECONT, conforme o quadro abaixo:

Projeto	ESTIMATIVA PRELIMINAR DE ESFORÇO DOS PROJETOS						Esforço Total
	Esforço por fase do Ciclo de Vida do Projeto (UST)						
	Definição do problema	Exploração de Dados	Desenvolvimento Analítico	Operação	Gestão	Infraestrutura	
Reformulação do Laboratório de Dados	325	268	387	201	87	173	1.441
Data Lake SECONT	4.913	4.050	5.860	3.042	1.311	2.617	21.793
Serviços de Detecção de Fraude e Erro	2.470	3.706	1.853	741	1.023	954	10.747
Serviços de Segunda Linha	4.688	3.483	6.762	3.826	1.537	3.241	23.537
						Total Geral	57.518

Dois projetos foram previstos, mas não precificados: **“Serviços de Terceira Linha”** e **“Serviços de Monitoria Correicional”**. Estes projetos serão atendidos por esforço interno após a capacitação dos auditores envolvidos.



7.2. Projetos candidatos.

Utilizaremos a metodologia **5W2H** para descrever as características gerais dos projetos candidatos à execução através da Solução Integrada de Inteligência e Processamento Analítico definida no presente estudo. Aprovada a contratação, os projetos candidatos serão formalmente descritos no Termo de Referência que solicitará a adesão da SECONT ao Registro de Preços da SESP.

Reformulação do Laboratório de Dados (LabData).

What (O que será feito?)	Why (Porque será feito?)
<ul style="list-style-type: none"> Consultoria para aplicação de técnicas e modelos de Inteligência Analítica nas bases de dados disponíveis no Data Lake SECONT, bem como definição de padrões para a gestão de dados. Criação, com repasse de conhecimento, de um modelo multidimensional ou score capaz de aprimorar as consultas e cruzamentos nas bases de dados estruturadas ou não, de interesse da SECONT. Criação de um novo padrão arquitetural para os projetos desenvolvidos pelos auditores do Lab.Data. 	<ul style="list-style-type: none"> Necessidade de atualização das ferramentas de análise de dados em uso pela SECONT e capacitação dos auditores e servidores envolvidos com seu uso. Necessidade de um modelo de dados capaz de responder com mais eficiência e eficácia as diversas demandas de consultas e análise das áreas fim do órgão. Necessidade de uma arquitetura capaz de automatizar a ingestão de dados e permitir o compartilhamento de informações.
Where (Onde será feito?)	Why (Quando será feito?)
<ul style="list-style-type: none"> Os serviços técnicos serão executados na infraestrutura da SECONT no PRODEST, as capacitações serão ministradas localmente. 	<ul style="list-style-type: none"> Este projeto será o primeiro a ser executado após a contratação. Início previsto para até 60 dias após a assinatura do contrato.
Who (Por quem será feito?)	How (Como será feito?)
<ul style="list-style-type: none"> Será realizado pela equipe técnica da contratada com acompanhamento dos auditores do LabData e de servidores do Suporte Técnico. 	<ul style="list-style-type: none"> Intervenções presenciais realizadas pela equipe técnica da contratada com apoio da equipe técnica da SECONT.
How much (Quanto custará?)	
<ul style="list-style-type: none"> 1.441 UST – R\$ 108.075,00. 	

Data Lake SECONT.

What (O que será feito?)	Why (Porque será feito?)
<ul style="list-style-type: none"> Criação, com repasse de conhecimento, do Data Lake SECONT através da integração das bases de dados estruturadas dos sistemas corporativos estaduais, bases estruturadas de outros poderes e da carga de dados não estruturados diversos. Criação de modelos automatizados de ingestão onde for possível. 	<ul style="list-style-type: none"> Necessidade de um repositório de dados único e capaz de receber atualizações periódicas dos órgãos fornecedores e dados. Necessidade de fornecer as fontes de dados para os diversos projetos de análise em curso na SECONT.
Where (Onde será feito?)	Why (Quando será feito?)
<ul style="list-style-type: none"> Os serviços técnicos serão executados na infraestrutura da SECONT no PRODEST, as capacitações serão ministradas localmente. 	<ul style="list-style-type: none"> Este projeto deverá ser executado paralelamente ao projeto de Reformulação do Laboratório de Dados em função de dependências técnicas.
Who (Por quem será feito?)	How (Como será feito?)
<ul style="list-style-type: none"> Será realizado pela equipe técnica da contratada com acompanhamento dos auditores do LabData e de servidores do Suporte Técnico. 	<ul style="list-style-type: none"> Intervenções presenciais realizadas pela equipe técnica da contratada com apoio da equipe técnica da SECONT.
How much (Quanto custará?)	
<ul style="list-style-type: none"> 21.793 UST – R\$ 1.634.475,00. 	



Serviços de Detecção de Fraude e Erro.

What (O que será feito?)	Why (Porque será feito?)
<ul style="list-style-type: none">• Construção, com repasse de conhecimento, de modelos analíticos para apoio às investigações realizadas nos processos administrativos de responsabilização, processos administrativos disciplinares e nos procedimentos que requeiram cruzamentos e correlações em diversas bases de dados relacionais ou não.• Desenho de redes de relacionamentos centradas nos investigados a partir dos parâmetros necessários em cada tipo de investigação.• Definição de scores por CNPJ, CPF e demais itens, bem como criação de dashboards diversos, publicados e internos.	<ul style="list-style-type: none">• Necessidade de apoiar as detecções de fraude e erro realizadas pelos diversos setores da SECONT, bem como aprimorar a obtenção das informações necessárias para cada procedimento.• Necessidade de agilizar os procedimentos de investigação e torná-los mais eficazes.• Criar rotinas que permitam aos usuários realizarem suas próprias consultas sem depender da intervenção de algum auditor do LabData ou servidor com formação técnica.• Ampliar o escopo das investigações para além das bases de dados existentes.
Where (Onde será feito?)	Why (Quando será feito?)
<ul style="list-style-type: none">• Os serviços técnicos serão executados na infraestrutura da SECONT, as capacitações serão ministradas localmente.	<ul style="list-style-type: none">• Este projeto deverá ser executado após a criação do Data Lake SECONT. Início previsto em até 120 dias após a assinatura do contrato.
Who (Por quem será feito?)	How (Como será feito?)
<ul style="list-style-type: none">• Será realizado pela equipe técnica da contratada com acompanhamento dos auditores do LabData e de servidores do Suporte Técnico.	<ul style="list-style-type: none">• Intervenções presenciais realizadas pela equipe técnica da contratada com apoio da equipe técnica da SECONT.
How much (Quanto custará?)	
<ul style="list-style-type: none">• 10.747 UST – R\$ 806.025,00.	

Serviços de Segunda Linha.

What (O que será feito?)	Why (Porque será feito?)
<ul style="list-style-type: none">• Construção, com repasse de conhecimento, de modelos analíticos para apoio às atividades de segunda linha, tais como: obtenção e cruzamento de dados em bases de interesse, relacionais ou não; registros de eventos e alertas a partir das bases do SIARHES e SIGEFES, avaliação de risco, etc.• Construção, com repasse de conhecimento, de painéis diversos com dados do RELUCI, PCA, segregação de função e outros, com alertas e eventos.• Fornecimento de dados para sistemas administrativos via API nativa da solução.	<ul style="list-style-type: none">• Necessidade de apoiar as atividades de segunda linha, bem como aprimorar a obtenção das informações necessárias ao desenvolvimento dessas atividades.• Necessidade de agilizar e automatizar alguns procedimentos de segunda linha e torná-los mais eficazes.• Criar rotinas que permitam aos usuários realizarem suas próprias consultas sem depender da intervenção de algum auditor do LabData ou servidor com formação técnica.
Where (Onde será feito?)	Why (Quando será feito?)
<ul style="list-style-type: none">• Os serviços técnicos serão executados na infraestrutura da SECONT, as capacitações serão ministradas localmente.	<ul style="list-style-type: none">• Este projeto deverá ser executado após a criação do Data Lake SECONT. Início previsto em até 120 dias após a assinatura do contrato.
Who (Por quem será feito?)	How (Como será feito?)
<ul style="list-style-type: none">• Será realizado pela equipe técnica da contratada com acompanhamento dos auditores do LabData e de servidores do Suporte Técnico.	<ul style="list-style-type: none">• Intervenções presenciais realizadas pela equipe técnica da contratada com apoio da equipe técnica da SECONT.
How much (Quanto custará?)	
<ul style="list-style-type: none">• 23.537 UST – R\$ 1.765.275,00.	



Serviços de Terceira Linha.

What (O que será feito?) <ul style="list-style-type: none">• Construção, com repasse de conhecimento, de modelos analíticos para apoio às atividades de terceira linha, tais como: obtenção e cruzamento de dados em bases de interesse, relacionais ou não; obtenção de evidências em auditorias; identificação de anomalias e distorções relevantes em registros contábeis; etc.• Construção, com repasse de conhecimento, de painéis diversos com dados de bases corporativas com alertas e eventos.• Fornecimento de dados para sistemas administrativos via API nativa da solução.	Why (Porque será feito?) <ul style="list-style-type: none">• Necessidade de apoiar as atividades de terceira linha, bem como aprimorar a obtenção das informações necessárias ao desenvolvimento dessas atividades.• Necessidade de agilizar e automatizar alguns procedimentos de terceira linha e torná-los mais eficazes.• Criar rotinas que permitam aos usuários realizarem suas próprias consultas sem depender da intervenção de algum auditor do LabData ou servidor com formação técnica.
Where (Onde será feito?) <ul style="list-style-type: none">• Os serviços técnicos serão executados na infraestrutura da SECONT, as capacitações serão ministradas localmente.	Why (Quando será feito?) <ul style="list-style-type: none">• Este projeto deverá ser executado após a criação do Data Lake SECONT. Início previsto em até 180 dias após a assinatura do contrato.
Who (Por quem será feito?) <ul style="list-style-type: none">• Será realizado pela equipe técnica da contratada com acompanhamento dos auditores do LabData e de servidores do Suporte Técnico.	How (Como será feito?) <ul style="list-style-type: none">• Intervenções presenciais realizadas pela equipe técnica da contratada com apoio da equipe técnica da SECONT.
How much (Quanto custará?) <ul style="list-style-type: none">• 0 UST – R\$ 0,00.	

Serviços de Monitoria Correicional.

What (O que será feito?) <ul style="list-style-type: none">• Construção, com repasse de conhecimento, de modelos analíticos para apoio às atividades correicionais, tais como: obtenção e cruzamento de dados em bases de interesse, relacionais ou não; obtenção de evidências em processos administrativos disciplinares; identificação de anomalias em registros funcionais; etc.• Construção, com repasse de conhecimento, de painéis diversos com dados de bases corporativas com alertas e eventos, em especial para acúmulo de cargo e evolução patrimonial.• Fornecimento de dados para sistemas administrativos via API nativa da solução.	Why (Porque será feito?) <ul style="list-style-type: none">• Necessidade de apoiar as atividades correicionais, bem como aprimorar a obtenção das informações necessárias ao desenvolvimento dessas atividades.• Necessidade de agilizar e automatizar alguns procedimentos correicionais e torná-los mais eficazes.• Criar rotinas que permitam aos usuários realizarem suas próprias consultas sem depender da intervenção de algum auditor do LabData ou servidor com formação técnica.
Where (Onde será feito?) <ul style="list-style-type: none">• Os serviços técnicos serão executados na infraestrutura da SECONT, as capacitações serão ministradas localmente.	Why (Quando será feito?) <ul style="list-style-type: none">• Este projeto deverá ser executado após a criação do Data Lake SECONT. Início previsto em até 180 dias após a assinatura do contrato.
Who (Por quem será feito?) <ul style="list-style-type: none">• Será realizado pela equipe técnica da contratada com acompanhamento dos auditores do LabData e de servidores do Suporte Técnico.	How (Como será feito?) <ul style="list-style-type: none">• Intervenções presenciais realizadas pela equipe técnica da contratada com apoio da equipe técnica da SECONT.
How much (Quanto custará?) <ul style="list-style-type: none">• 0 UST – R\$ 0,00.	



7.3. Serviços por demanda X Mão de obra residente.

Como observado anteriormente, os projetos de desenvolvimento analítico possuem características e requisitos diferentes daquelas encontradas no desenvolvimento de software, demandando abordagens, processos, conhecimentos e métricas próprias. Para o desenvolvimento dos projetos definidos pela SECONT neste estudo será necessário a utilização de perfis profissionais especializados, algo não disponível atualmente no órgão.

Os principais perfis profissionais necessários aos projetos previstos neste estudo são os seguintes:

a) Cientista de Dados.

Atividades: Coletar grandes quantidades de dados estruturados ou não estruturados, dados sem tratamento ou qualificação, selecioná-los, transformá-los e prepará-los com a finalidade de realizar abordagens analíticas que solucionem problemas de negócio do órgão, com a aplicação de métodos e técnicas analíticas descritivas e preditivas, uso de diversas linguagens de programação, especialmente, R e Python, ter uma sólida experiência em modelagem estatística, incluindo regressões e métodos de aprendizagem de máquina, aprendizagem profunda e processamento de linguagem natural, sendo capaz de se comunicar bem e colaborar com times de TI e com as áreas de negócios.

Perfil Profissional: Profissional com formação em nível superior, com o mínimo de 3 (anos) anos de experiência profissional em ciências de dados e uso de ferramentas analíticas em todo o ciclo de vida.

b) Engenheiro de Dados.

Atividades: Definição de arquitetura, estrutura e modelagem de dados, focado no *design*, construção e manutenção de ambientes de dados (Bancos, *Data Warehouse* e *Data Lake*).

Perfil Profissional: Profissionais com formação em nível superior, com o mínimo de 3 (anos) anos de experiência profissional em arquitetura e gerenciamento de dados, com aplicação de ferramentas de modelagem, manipulação e consulta a dados estruturados ou não estruturados.

A dinâmica dos projetos exige a atuação pontual dos profissionais conforme a etapa do ciclo de vida analítico do projeto em desenvolvimento (**Definição do Problema; Exploração de Dados; Desenvolvimento Analítico; Operação; Gestão; Infraestrutura**). Por isso, optou-se pelo modelo de contratação por demanda mediante Ordens de Serviço que permitirá a alocação dinâmica dos profissionais requeridos em cada momento do projeto.

Estes perfis profissionais, por sua especificidade e especialização, não se confundem com os perfis usualmente alocados em contratos de terceirização de mão de obra que possuem maior foco no desenvolvimento de sistemas tradicionais e na sustentação da infraestrutura.

Apesar da contratação terceirizada desta mão de obra, entendemos que é fundamental que o conhecimento técnico necessário para a continuidade dos projetos de Inteligência Analítica e Modelagem Estatística, após o término da vigência contratual, seja internalizado, de forma a permitir que os servidores da SECONT assumam a sustentação dos modelos analíticos e rotinas desenvolvidos pela consultoria externa.

Com este objetivo, utilizaremos os treinamentos específicos no uso da solução em tempo de projeto e atividades de transferência de conhecimento para garantir a continuidade dos serviços técnicos especializados, mas executados pelos próprios servidores da SECONT.



8. ESTIMATIVA DE CUSTO TOTAL DA CONTRATAÇÃO.

A precificação considerou o resultado do Pregão SESP 003/2022, ocorrido em 10/06/2022, na qual foi declarada vencedora a empresa VERT SOLUÇÕES EM INFORMÁTICA LTDA que concorreu oferecendo soluções de *analytics* da empresa SAS. Caso a contratação definida neste estudo seja autorizada, uma nova pesquisa de preços será realizada conforme a Norma de Procedimento SCL 015 – Adesão a Ata de Registro de Preços da Administração Estadual (Portaria nº 36-R/2019).

Cabe destacar que a nova pesquisa de preços será realizada com base na descrição dos objetos definidos no Pregão SESP 003/2022, bem como nas soluções de *analytics* da empresa SAS. Destaca-se também que a aquisição de uma solução de *hardware baremetal*, convergente ou hiperconvergente pode não ser necessária caso o PRODEST absorva a carga de trabalho necessária em sua infraestrutura virtualizada.

O custo total da contratação pretendida é demonstrado na tabela abaixo:

SERVIÇOS / MATERIAIS / LICENÇAS DE SOFTWARE	MÉTRICA	VALOR UNITÁRIO	QTD	VALOR TOTAL
Aquisição de licenças perpétuas de software de Solução Analítica Integrada de Processamento Preditivo, Estatística e Mineração de Dados com garantia de atualização e suporte técnico do fabricante por 12 meses (ITEM 1).	Conjunto de 32 núcleos de processamento	R\$ 3.900.000,00	1	R\$ 3.900.000,00
Subscrição de licenças de software de plataforma para armazenamento e processamento de dados em computação distribuída (Hadoop) durante o período de 12 (doze) meses (ITEM 3)	Conjunto de 5 nós de processamento	R\$ 864.097,50	1	R\$ 864.097,50
Serviço de Instalação, configuração, customização, testes, documentação da implantação, transferência de conhecimento e conexão inicial às fontes internas e externas de dados que se encontrem disponíveis (ITEM 4).	Unidade	R\$ 138.815,00	1	R\$ 138.815,00
Serviço de treinamento em administração do Software de Solução Analítica Integrada de Processamento Preditivo, Estatística e Mineração de Dados (ITEM 5).	Turma de 10 pessoas	R\$ 53.998,00	1	R\$ 53.998,00
Serviço de treinamento em <i>Big Data</i> e Ciência de Dados aplicada ao Software de Solução Analítica Integrada de Processamento Preditivo, Estatística e Mineração de Dados (ITEM 6).	Turma de 20 pessoas	R\$ 194.000,00	1	R\$ 194.000,00
Serviços técnicos: Projeto de Reformulação do Laboratório de Dados (ITEM 7).	UST	R\$ 75,00	1.441	R\$ 108.075,00
Serviços técnicos: Projeto Data Lake SECONT (ITEM 7).	UST	R\$ 75,00	21.793	R\$ 1.634.475,00
Serviços técnicos: Projeto Serviços de Detecção de Fraude e Erro (ITEM 7).	UST	R\$ 75,00	10.747	R\$ 806.025,00
Serviços técnicos: Projeto Serviços de Segunda Linha (ITEM 7).	UST	R\$ 75,00	23.537	R\$ 1.765.275,00
SECONT Solução de hardware <i>baremetal</i> , convergente ou hiperconvergente necessária para absorver a carga de trabalho do Software de Solução Analítica Integrada de Processamento Preditivo, Estatística e Mineração de Dados.	Unidade	R\$ 288.000,00	1	R\$ 288.000,00
VALOR TOTAL				R\$ 9.752.760,50



9. NECESSIDADE DE ADEQUAÇÃO DO AMBIENTE.

9.1. Infraestrutura tecnológica.

Todo *backend* da infraestrutura tecnológica da SECONT opera no DATACENTER governamental, no PRODEST. As soluções de software descritas no presente estudo serão implantadas em estruturas técnicas deste DATACENTER.

Todavia, em função da necessidade de processamento necessária às soluções descritas, considera-se também a aquisição de uma solução de *hardware* para absorvê-la, podendo tal solução ser *baremetal*, convergente ou hiperconvergente. De qualquer forma, caso esta solução de hardware seja, de fato, adquirida, ela será implantada no DATACENTER governamental.

9.2. Infraestrutura elétrica.

Não se aplica, pois será utilizada a infraestrutura do DATACENTER governamental.

9.3. Logística.

Será provida pelo PRODEST conforme as regras de utilização do DATACENTER governamental.

9.4. Espaço físico.

Não se aplica, pois será utilizada a infraestrutura do DATACENTER governamental.

10. CONTINUIDADE DO NEGÓCIO DURANTE E APÓS A EXECUÇÃO CONTRATUAL.

10.1. Modelo de licenciamento.

O modelo de negócio estabelecido pela SESP em sua contratação define uma forma de aquisição que permite ao cliente utilizar a totalidade das soluções disponíveis no produto *SAS Detection and Investigation for Government*, a depender da necessidade do negócio deste cliente. O licenciamento é perpétuo com suporte por 12 (doze) meses e sua precificação é feita por cada conjunto de “32 *cores*” (32 núcleos de processamento), sendo que os serviços de implantação, customização, treinamento e repasse de conhecimento são cobrados em separado.

10.2. Recursos materiais.

Acerca dos recursos materiais, a SECONT será responsável por prover os elementos físicos necessários para o desempenho das atividades dentro do órgão, quando for o caso, compreendendo:

Recursos de infraestrutura.

- Equipamentos servidores;
- Estações de trabalho;



- Toda a infraestrutura física e lógica para instalação da solução;
- Toda a infraestrutura física e lógica para acomodação dos profissionais da contratada para a instalação e configuração da solução.

Recursos de software.

- Softwares necessários ao funcionamento do ambiente computacional da SECONT.

10.3. Recursos humanos.

A SECONT providenciará servidores para assumir os seguintes papéis, consoante com a legislação aplicável:

- **Fiscal do Contrato:** servidor em exercício e com conhecimento técnico do assunto, indicado pela autoridade do órgão para ser encarregado pelo acompanhamento da execução, ateste das faturas e notas fiscais e pela conferência dos serviços prestados pela contratada, desde o início até o término da vigência do contrato.
- **Gestor do Contrato:** servidor em exercício na área de compras e contratos, que será incumbido da gestão administrativa do contrato, desde o início até o término da vigência do contrato.

No que se refere à empresa contratada, esta nomeará um Preposto para representá-la, e este executará suas atividades nas dependências da SECONT ou remotamente, cabendo a ele assumir a função de gestor de relacionamento e executar suas atividades conforme o planejamento apresentado pela contratada. Caberá, ao preposto, a gestão de pessoal dos funcionários da contratada.

Não haverá ingerência da SECONT em qualquer das atribuições de gestão de equipe da contratada, exceto aqueles casos que digam respeito eminente e exclusivamente a aspectos técnicos. O preposto deverá ter a sua indicação formalizada junto ao órgão e contar com a anuência deste. A CONTRATADA deverá dimensionar todos os recursos necessários à prestação dos serviços.

10.4. Estratégia de continuidade contratual.

Será exigido que ao término do contrato, seja por decurso de vigência ou por rescisão antecipada, a contratada promoverá a transição contratual adotando um plano de transição com transferência de tecnologia, sem perda de informações e capacitando, se solicitado, os auditores e demais servidores da SECONT e/ou os da nova pessoa jurídica que continuará a execução dos serviços.

As responsabilidades serão definidas conforme o quadro abaixo:

ID	AÇÃO DE CONTIGÊNCIA	RESPONSÁVEL
1	Apresentação do projeto de gerenciamento de serviços de transição contratual final.	SECONT e empresa contratada
2	Reunião de alinhamento de expectativas contratuais.	SECONT e empresa contratada
3	Execução	Empresa contratada
4	Término do projeto de gerenciamento de serviços de transição contratual final.	Empresa contratada

10.5. Limite máximo de vigência contratual.

O estabelecimento do limite de vigência legal máximo permitido para o contrato deverá observar o seguinte:



- a) O ITEM 1 do objeto refere-se à contratação de licenças perpétuas de software, equivalendo à aquisição de um bem, não havendo que se falar em duração máxima do contrato.
- b) O ITEM 2 do objeto são serviços acessórios contratados juntamente com a licença perpétua do software, devendo respeitar o inciso II do art. 57 da Lei nº 8666/93 por referirem-se a serviços de natureza contínua, estando limitados ao prazo máximo de 60 (sessenta) meses de vigência contratual.
- c) O ITEM 3 do objeto refere-se à contratação de subscrição de software caracterizando uma prestação de serviço de licenciamento temporário de software. Por essa razão, deve respeitar o inciso IV do art. 57 da Lei nº 8666/93, estando limitado ao prazo máximo de 48 (quarenta e oito) meses de vigência contratual.
- d) Os ITENS 4, 5 e 6 são de natureza de execução pontual, não havendo que se falar em duração máxima do contrato nesses casos.
- e) O ITEM 7 do objeto refere-se à contratação de serviços técnicos especializados. Em razão disso, este ESTUDO TÉCNICO PRELIMINAR prevê, na descrição do objeto da contratação, a quantidade estimada desses serviços e o prazo máximo para execução, conforme abaixo:

ITEM 7 - Contratação sob demanda de 57.518 (cinquenta e sete mil quinhentos e dezoito) “Unidades de Serviços Técnicos” (UST) para atendimento on-site (no local) durante o período de 12 (doze) meses. As USTs serão consumidas pelos serviços de instalação, configuração, customização, testes, documentação da implantação e transferência de conhecimento das soluções previstas no ITEM 1 e ITEM 3, bem como pelos serviços de treinamento na Plataforma Analítica e nos conceitos de Ciência de Dados e Big Data.

11. RISCOS DA CONTRATAÇÃO.

11.1. Identificação dos riscos tecnológicos e de gestão.

Não seria exagero afirmar que em breve o Controle Interno não será viável sem o apoio constante das técnicas e ferramentas da Ciência de Dados. O volume, a variedade e a velocidade dos dados que precisam ser considerados nos trabalhos de avaliação e consultoria não permitem mais apenas a análise humana, sob pena de prejudicar o julgamento e o zelo profissional dos auditores.

Dessa forma, dotar a SECONT de uma solução analítica capaz de apoiar suas atividades finalísticas nos parece um caminho natural, mas alguns riscos tecnológicos e de gestão precisam ser considerados pela Administração, tais como:

- a) Ausência de auditores capacitados ou com interesse nos temas relacionados à Ciência de Dados.
- b) Dificuldade de engajamento para novas capacitações em função do excesso de atribuições e escassez de pessoal.
- c) Pouca discussão sobre a real aplicação da Ciência de Dados nas atividades fim da SECONT, o que resultou em projetos de dados baseados em percepções generalistas e excessivamente subjetivas.
- d) Ausência de um plano de dados ou de conhecimento em gestão de dados que possa nortear os projetos elegíveis para construção através da solução analítica pretendida.



- e) Subutilização da solução pretendida em função da ausência de complexidade dos projetos e do modelo de licenciamento adotado pela SESP.
- f) Necessidade constante de atualização do suporte do fabricante para evitar a descontinuidade de serviços ou funcionalidades, mesmo no modelo de licenciamento perpétuo.
- g) Dificuldade em conseguir as bases de dados necessárias para implementação plena dos projetos elencados no presente estudo.
- h) Baixa performance das análises em função da infraestrutura tecnológica disponibilizada pelo PRODEST, fato que pode demandar a aquisição de equipamentos pela SECONT.

11.2. Identificação dos riscos financeiros.

a) Modelo de licenciamento adotado.

Um dos riscos financeiros da aquisição pretendida decorre do modelo de licenciamento adotado pela SESP no ITEM 1, que permite ao cliente utilizar a totalidade dos produtos e serviços disponíveis no portfólio da solução/plataforma **SAS Detection and Investigation**, a depender das necessidades do negócio deste cliente.

A solução *SAS Detection and Investigation* conta com os seguintes produtos e serviços:

- Base SAS; SAS/SHARE; SAS/ETS; SAS/OR; SAS/QC; SAS/STAT; SAS Enterprise Guide; SAS Visual Investigator; SAS Social Network Analysis Server; SAS Contextual Analysis Server; SAS Financial Crimes Monitor Server; Credit Scoring for SAS Enterprise Miner; SAS Model Manager Server; SAS Data Management Standard Server; SAS Data Quality Standard Server; SAS Enterprise BI Server (including SAS OLAP Server); SAS Forecast Server; SAS Enterprise Miner Server; SAS Text Miner Server; SAS Quality Knowledge Base; SAS Visual Process Orchestration Server; SAS High-Performance Entity and Network Generation Server; SAS Enterprise Content Categorization Server; SAS Crawler Server; SAS Metadata Server; SAS Metadata Bridges; SAS Access to Hadoop; SAS Access to Oracle; SAS Access to ODBC; SAS Access to PC Files; SAS Viya; SAS Visual Analytics (on SAS Viya); SAS Decision Manager (on SAS Viya); SAS Model Manager (on SAS Viya); SAS Visual Statistics (on SAS Viya); SAS Data Preparation (on SAS Viya); SAS Visual Data Mining and Machine Learning (on SAS Viya); SAS Visual Forecasting (on SAS Viya); SAS Visual Text Analytics (on SAS Viya).

Ocorre que não foi possível localizar no processo de aquisição realizado pela SESP um comparativo que indique a economicidade do licenciamento integral da solução *SAS Detection and Investigation*, comparando-o com o licenciamento apenas dos produtos necessários aos projetos elencados.

Cabe destacar, todavia, que projetos na área de Ciência de Dados são consideravelmente mutáveis e abrangentes em função dos requisitos que surgem a partir das demandas de negócio, mesmo em tempo de execução. Considere-se ainda a própria estrutura de funcionamento das soluções SAS, que criam redes de dependência entre seus produtos que tornam muito difícil o mapeamento de quais deles são efetivamente necessários em cada projeto.

b) Custo elevado de manutenção.

Outro risco financeiro que deve ser destacado decorre do elevado custo do serviço de suporte técnico, manutenção e direito de atualização de versão por 12 meses da solução *SAS Detection and Investigation*, representado pelo ITEM 2 na aquisição da SESP.



Os valores praticados para os serviços de renovação por 12 meses descritos no ITEM 2 equivalem a **27,56%** do valor de aquisição da solução indicada no ITEM 1 (respectivamente **R\$ 1.075.140,00** e **R\$ 3.900.000,00**). Mesmo se tratando de licenças perpétuas, o custo para manter a solução atualizada e com suporte do fabricante é excessivo e configuraria um compromisso financeiro contínuo. Por outro lado, abrir mão de tal suporte significa aceitar o risco de interrupção de serviços por desatualização ou por problemas que a equipe interna da SECONT não seja capaz de resolver.

O mesmo ocorre com a manutenção anual do ITEM 3, que trata da subscrição do software *Hadoop* por 12 meses. O valor definido no pregão foi de **R\$ 864.097,50** e trata-se de um componente essencial da solução (*Data Lake*). O custo das obrigações anuais, portanto, alcança **R\$ 1.939.237,50** e excede em muito os valores usualmente alocados pela SECONT para o custeio de sua infraestrutura tecnológica.

11.3. Análise e avaliação dos riscos.

Não é objetivo deste estudo realizar uma análise e avaliação formal dos riscos identificados, mas tão somente relacioná-los para auxiliar a subsidiar a decisão pela viabilidade ou não da contratação. Considerando que os principais riscos envolvem a gestão de pessoas e o planejamento dos serviços, entendemos que os gestores envolvidos no processo decisório poderão considerar os riscos elencados e trata-los conforme sua relevância.

12. CONSTRUÇÃO DA CENTRAL ANALÍTICA SECONT.

Apesar da SECONT possuir uma coordenação específica para análise de dados, o *Lab.Data*, sua composição e recursos não se mostram capazes de atender a crescente demanda de serviços de dados do órgão. Os objetos descritos neste Estudo Técnico permitirão a criação de uma estrutura conceitual que chamamos de “Central Analítica”, sua composição se apresenta conforme a figura abaixo:

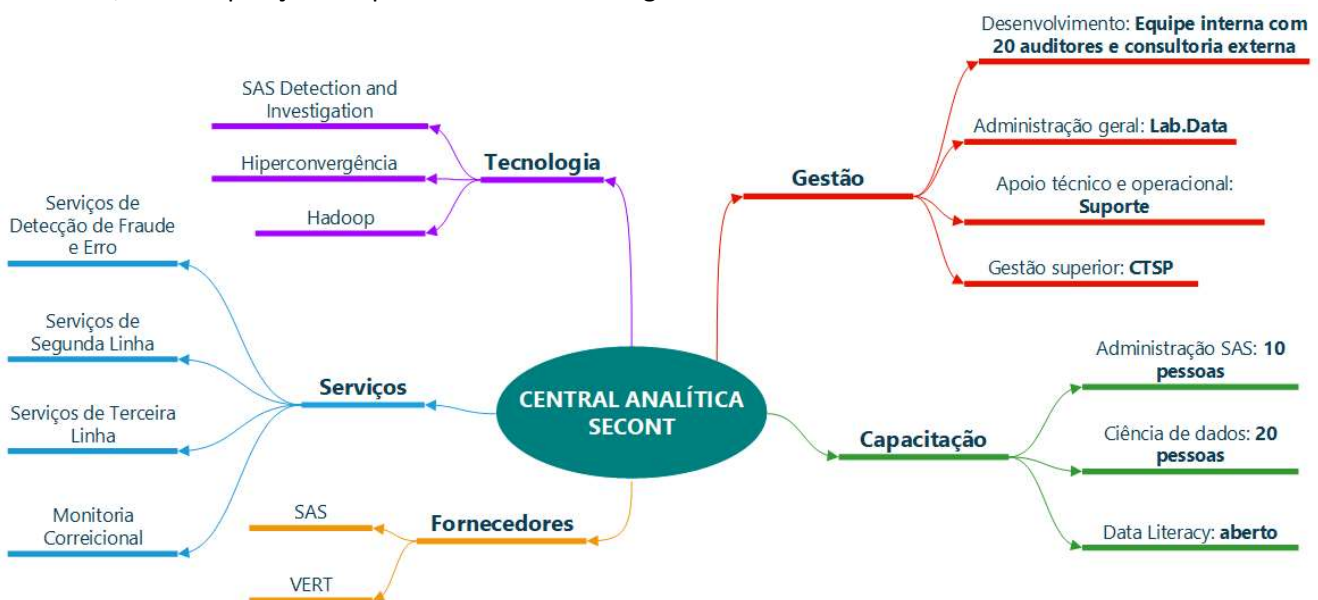


Figura 2 - Central Analítica SECONT.

O modelo funcional e o posicionamento da Central Analítica na estrutura organizacional da SECONT ainda serão definidos em futuras deliberações do CTSP (Comitê Gestor de Tecnologia, Segurança da Informação e Proteção de Dados Pessoais).



Eis uma breve descrição dos componentes da Central Analítica proposta:

TECNOLOGIA: a Central Analítica será baseada no **SAS Detection and Investigation** como solução de inteligência analítica, no **Hadoop** como solução de armazenamento de dados (*Data Lake*) e em soluções de *hardware* convergentes ou hiperconvergentes instalados no PRODEST para sustentação física.

GESTÃO: O modelo de gestão proposto prevê o seguinte:

- 20 (vinte) auditores/servidores na equipe de desenvolvimento interno de soluções.
- Uma empresa especializada prestando consultoria externa.
- Administração geral sob a responsabilidade do Lab.Data e com até 10 (dez) auditores/servidores capacitados nas rotinas e atividades de manutenção.
- Apoio operacional do Suporte Técnico.
- Gestão superior exercida pelo CTSP (Comitê Gestor de Tecnologia, Segurança da Informação e Proteção de Dados Pessoais).

CAPACITAÇÃO: As capacitações necessárias para operação da Central analítica são as seguintes:

- 20 (vinte) auditores/servidores treinados em Ciência de Dados e aptos a criar modelos e solução analíticas.
- 10 (dez) auditores/servidores capacitados na administração da ferramenta.
- Curso aberto de *Data Literacy* (alfabetização de dados).

FORNECEDORES: Os fornecedores de softwares e serviços serão os seguintes:

- SAS – Soluções de inteligência analítica.
- VERT – Serviços de consultoria em ciência de dados.

SERVIÇOS: A Central Analítica deverá prever serviços de análise de dados que permitam o uso pelas áreas clientes, sem intermédio dos auditores do Lab.Data ou dos auditores alocados na equipe de desenvolvimento de soluções. Os serviços devem prever consultas já parametrizadas ou a construção de consultas por demanda, bem como a utilização de painéis (*dashboards*) e alertas.

Entende-se com **Fator Crítico de Sucesso** da Central Analítica proposta o engajamento e dedicação da equipe de desenvolvimento interno de soluções, pois dela dependerá a continuidade, melhoria e ampliação dos serviços prestados.

13. POSSIBILIDADE DE COMPARTILHAMENTO DE RECURSOS COM OUTROS ÓRGÃOS.

Apesar de não ser premissa da ARP SESP 005/2022, vislumbra-se inúmeras possibilidades de compartilhamento de recursos computacionais entre os órgãos para reduzir custos em função de aquisições duplicadas das soluções propostas neste estudo.

Ainda que tal compartilhamento implique em preocupações extras com segurança, desempenho e gestão, a economia proporcionada justifica uma avaliação detalhada desta possibilidade pelos órgãos que pretendam aderir ao referido Registro de Preços. Deve ser considerado que o governo estadual possui um DATACENTER capaz de agregar toda a carga de trabalho necessária e permitir o acesso simultâneo aos serviços mantidos.

Além da economia, o compartilhamento de recursos poderá permitir também o compartilhamento de dados, ajudando a ampliar a cultura necessária para avançar a maturidade de gestão de dados nos órgãos envolvidos, bem como facilitar projetos e ideias. Um dos possíveis cenários de compartilhamento seria a criação de um *Data Lake* corporativo, conforme observamos na Figura 3.

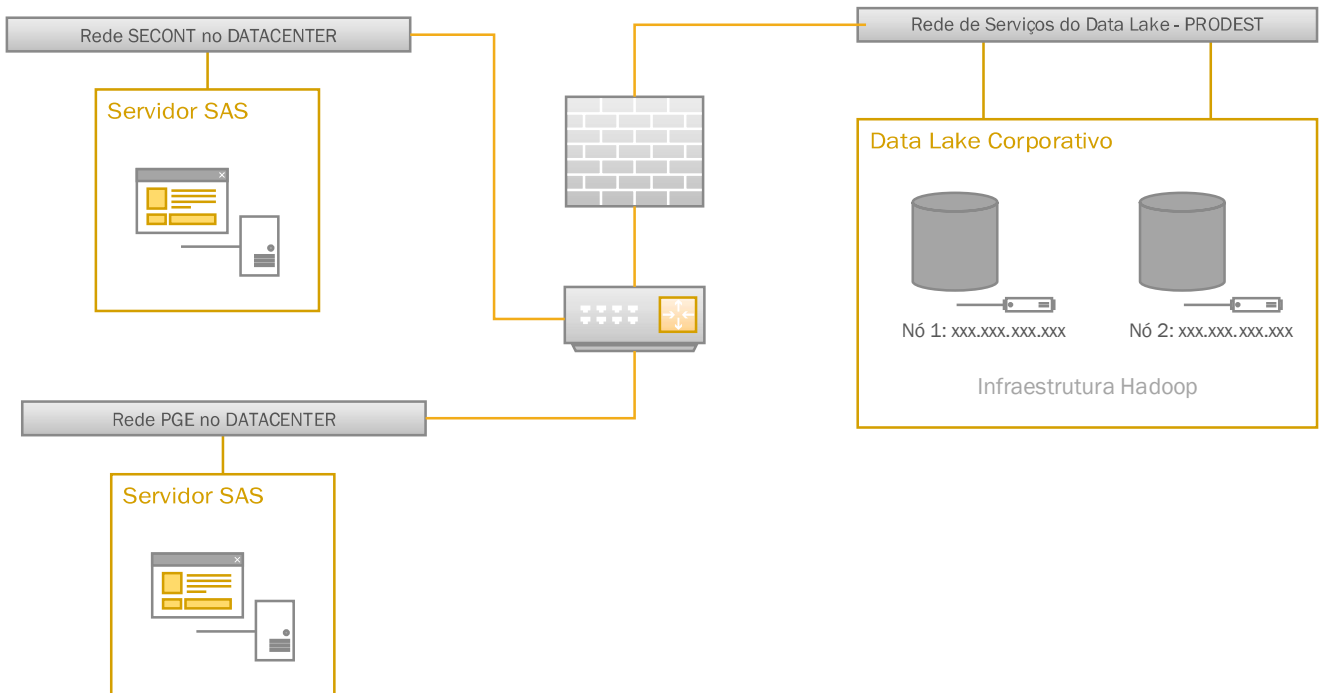


Figura 3 - Exemplo de compartilhamento do Data Lake entre dois órgãos.

Na figura acima temos um exemplo fictício de um *Data Lake* corporativo compartilhado por dois órgãos, SECONT e PGE. Tal estrutura evitaria a aquisição duplicada do ITEM 3 da ARP SESP 005/2022, evitando um custo anual de **R\$ 864.097,50** por órgão. Da mesma forma, os recursos de hardware seriam compartilhados proporcionando mais racionalidade nas aquisições.

Tal estrutura também facilitaria os processos de ingestão de dados nas bases armazenadas, tornando-as aptas ao processamento analítico pelos órgãos com permissões de acesso. Cabe destacar que o PRODEST é capaz de prover a segurança na camada de rede, e a infraestrutura *Hadoop* se encarregaria da segurança na camada de dados, cabendo aos órgãos a definição de um modelo de gestão que viabilize o compartilhamento.

Mas as possibilidades de compartilhamento não se restringiriam ao armazenamento e ingestão de dados em um *Data Lake* corporativo, as capacitações também poderiam ser compartilhadas a partir de acordos entre órgãos participantes com demandas semelhantes, em especial no ITEM 5.

O investimento no licenciamento da solução de Inteligência Analítica proposta é vultuoso, se consideramos apenas os ITENS 1 e 2 teríamos **R\$ 4.975.140,00** por órgão, sendo que o valor **R\$ 1.075.140,00** (ITEM 2) deve ser tratado como despesa contínua anual. Evitar, portanto, a duplicação desnecessária destes valores onde possível deve ser obrigação dos gestores públicos envolvidos.

Ainda que seja possível realizar estudos de custo/benefício que justifiquem todas as despesas citadas, os riscos de alguns dos projetos citados não atingirem plenamente seus objetivos são reais. Portanto, a SECONT deve buscar formas de reduzir os riscos identificados através de ações de mitigação e redução, tanto nos aspectos de gestão quanto financeiros.

Por fim, deve-se discutir com a empresa vencedora do Pregão Eletrônico SESP 003/2022 inclusive a possibilidade técnica e contratual de órgãos distintos compartilharem uma instância do *SAS Detection and Investigation*, que seria provida como serviço pelo PRODEST a partir do DATACENTER.



14. DECLARAÇÃO DE VIABILIDADE DA CONTRATAÇÃO.

Declaramos que o presente **ESTUDO TÉCNICO PRELIMINAR** para avaliação da aquisição de uma **Solução Integrada de Inteligência e Processamento Analítico** foi elaborado em conformidade com os requisitos necessários aos projetos e demandas da Secretaria de Controle e Transparência, e que os riscos identificados refletem adequadamente a situação atual do órgão.

Por tudo exposto, entendemos pela:

- Aprovação da aquisição proposta e instrução do processo de adesão.
 Reprovação da aquisição proposta.

Responsável Técnico:

[ASSINADO ELETRONICAMENTE]

EMERSON COUTO DE MOURA

Analista de Tecnologia da Informação

Avaliação:

[ASSINADO ELETRONICAMENTE]

MARCELO CAMPUS ANTUNES

Subsecretário de Controle

[ASSINADO ELETRONICAMENTE]

HELMUT MUTIZ D'AUVILA

Corregedor Geral do Estado

[ASSINADO ELETRONICAMENTE]

ALEXANDRE DEL SANTO FALCÃO

Subsecretário de Integridade Governamental e
Empresarial

[ASSINADO ELETRONICAMENTE]

EDMAR MOREIRA CAMATA

Secretário de Controle e Transparência

[ASSINADO ELETRONICAMENTE]

FABIANO DA ROCHA LOUZADA

Subsecretário de Transparência

[ASSINADO ELETRONICAMENTE]

PATRICIA BRAVIM MELOTTI

Gerente Técnica Administrativa

ASSINATURAS (8)

Documento original assinado eletronicamente, conforme MP 2200-2/2001, art. 10, § 2º, por:

RODOLFO PEREIRA NETTO
FUNCAO GRATIFICADA - AGE/SECONT
ASSTEC - SECONT - GOVES
assinado em 29/07/2022 17:04:56 -03:00

HELMUT MUTIZ D AUVILA
CORREGEDOR GERAL DO ESTADO QCE-01
COGES - SECONT - GOVES
assinado em 29/07/2022 17:51:49 -03:00

ALEXANDRE DEL SANTO FALCÃO
SUBSECRETARIO ESTADO QCE-01
SUBINT - SECONT - GOVES
assinado em 01/08/2022 11:12:05 -03:00

EMERSON COUTO DE MOURA
ASSESSOR ESPECIAL NIVEL IV QCE-03
GABSEC - SECONT - GOVES
assinado em 29/07/2022 17:43:40 -03:00

EDMAR MOREIRA CAMATA
SECRETARIO DE ESTADO
SECONT - SECONT - GOVES
assinado em 02/08/2022 17:40:40 -03:00

MARCELO CAMPOS ANTUNES
SUBSECRETARIO ESTADO QCE-01
SUBCONT - SECONT - GOVES
assinado em 02/08/2022 09:59:52 -03:00

PATRICIA BRAVIM MELOTTI
GERENTE FG-GE
GTA - SECONT - GOVES
assinado em 04/08/2022 11:29:55 -03:00

FABIANO DA ROCHA LOUZADA
SUBSECRETARIO ESTADO QCE-01
SUBTRAN - SECONT - GOVES
assinado em 01/08/2022 12:11:43 -03:00



INFORMAÇÕES DO DOCUMENTO

Documento capturado em 04/08/2022 11:29:55 (HORÁRIO DE BRASÍLIA - UTC-3)
por RODOLFO PEREIRA NETTO (FUNCAO GRATIFICADA - AGE/SECONT - ASSTEC - SECONT - GOVES)
Valor Legal: ORIGINAL | Natureza: DOCUMENTO NATO-DIGITAL

A disponibilidade do documento pode ser conferida pelo link: <https://e-docs.es.gov.br/d/2022-G27VCX>