

Termo de Referência COGTI 01280/2015

Consulta pública para contratação de solução de appliance para processamento de inteligência analítica.

A consulta pública será acompanhada pelos empregados:

- Charles Morais Magalhães lotado na COGTI/CIPOA, telefone (61) 2021-7259, e-mail: charles.magalhaes@SERPRO.gov.br.

- Alexandre Carlos Domingues da Silva lotado na COGTI/CIPOA/CIPCD, ramal #61 7867, e-mail: alexandre.silva@serpro.gov.br.

1.0 Objeto

1.1. Consulta pública para contratação de solução de appliance para processamento de inteligência analítica, compreendendo funcionalidades de análise estatística, mineração de dados e de textos.

2.0 Especificação do objeto a ser contratado

2.1. A solução de appliance para processamento de inteligência analítica, compreendendo funcionalidades de análise estatística, mineração de dados e de textos com cessão de direito de uso permanente, para usuários ilimitados, deverá ser disponibilizado em um produto único, homologado e certificado sob responsabilidade do mesmo fabricante, provendo alta disponibilidade, redundância e tolerância a falhas.

2.1.1. A solução de appliance para processamento de inteligência analítica deve ser entregue totalmente funcional de fábrica, com todos os seus componentes de hardware e software pré-instalados e pré-configurados e licenciados.

2.1.2. A solução de appliance para processamento de inteligência analítica deve prover a execução dos algoritmos e funções da solução analítica em modo in-database.

2.1.3. Os componentes da solução de appliance para inteligência analítica deverão ser disponibilizados para serem usados em qualquer área de interesse do negócio, não restringindo a um determinado setor, tal como combate a fraude financeira ou segurança pública.

2.1.4. As funcionalidades e algoritmos dos requisitos técnicos, devem ser disponibilizados de forma nativa pela solução de inteligência analítica, isto é, não devem ser executadas através de bibliotecas de terceiros ou linguagens de script em ambientes embutidos (modo embedded).

2.1.5. Técnicas de mineração de dados

2.1.5.1. Pré-processamento

2.1.5.1.1. Possuir técnicas para seleção de amostras representativas.

2.1.5.1.2. Prover algoritmos para seleção de atributos e redução de dimensionalidade, incluindo:

2.1.5.1.2.1. Seleção baseada em correlação de variáveis;

2.1.5.1.2.2. Seleção baseada em informação mútua;

2.1.5.1.2.3. Seleção baseada em análise de componentes principais (PCA);

2.1.5.1.2.4. Redução usando mapas auto-organizáveis de Kohonen (SOM); e

2.1.5.1.2.5. Redução usando algoritmo EM (Expectation Maximization).

- 2.1.5.1.3. Prover algoritmos para normalização, incluindo:
 - 2.1.5.1.3.1. Escala ordinal (substituição do valor pelo seu rank);
 - 2.1.5.1.3.2. Escala intervalar (ajuste de escala dos valores para um intervalo determinado);
 - 2.1.5.1.3.3. Escala aditiva (subtração de cada valor do atributo pela média dos valores); e
 - 2.1.5.1.3.4. Escala absoluta (mapeamento de valores).
- 2.1.5.1.4. Prover algoritmos para discretização.
- 2.1.5.1.5. Prover as seguintes técnicas de amostragem:
 - 2.1.5.1.5.1. Aleatória simples;
 - 2.1.5.1.5.2. Estratificada;
 - 2.1.5.1.5.3. Balanceada (ponderada);
 - 2.1.5.1.5.4. Conglomerados;
 - 2.1.5.1.5.5. Sistemática;
 - 2.1.5.1.5.6. Amostragem de grafos; e
 - 2.1.5.1.5.7. Cálculo de representatividade de amostras.
- 2.1.5.1.6. Possibilitar manipulação de diferentes tipos de atributos, em particular: numéricos (discretos e contínuos), categóricos e ordinais.
- 2.1.5.2. Estatística básica e avançada
 - 2.1.5.2.1. Prover as seguintes funções estatísticas:
 - 2.1.5.2.1.1. Análise descritiva e estatísticas sumárias dos dados;
 - 2.1.5.2.1.2. Análise de distribuições;
 - 2.1.5.2.1.5. Correlações; e
 - 2.1.5.2.1.5. Frequências.
 - 2.1.5.2.2. Prover os seguintes testes estatísticos:
 - 2.1.5.2.2.1. Testes para normalidade;
 - 2.1.5.2.2.2. Comparação de distribuições;
 - 2.1.5.2.2.3. Testes para uma ou duas amostras;
 - 2.1.5.2.2.4. Comparação entre duas amostras independentes;
 - 2.1.5.2.2.5. Comparação entre duas amostras relacionadas; e
 - 2.1.5.2.2.6. Medidas de correlação e testes associados.
 - 2.1.5.2.3. Prover análise inferencial através de:
 - 2.1.5.2.3.1. Regressões;
 - 2.1.5.2.3.2. Análise de variância;
 - 2.1.5.2.3.3. Testes Qui-Quadrado;
 - 2.1.5.2.3.4. Teste T; e
 - 2.1.5.2.3.5. Tabelas cruzadas.
 - 2.1.5.2.4. Prover análise de dados categóricos através de:
 - 2.1.5.2.4.1. Modelos log-lineares;
 - 2.1.5.2.4.2. Métodos de Mantel-Haenszel; e
 - 2.1.5.2.4.3. Modelos de decisão discreta.

- 2.1.5.2.5. Prover análise de correspondência múltipla através de:
 - 2.1.5.2.5.1. Análise discriminante;
 - 2.1.5.2.5.2. Análise de cluster;
 - 2.1.5.2.5.3. Correlação canônica;
 - 2.1.5.2.5.4. Análise de componentes principais (PCA);
 - 2.1.5.2.5.5. Análise fatorial; e
 - 2.1.5.2.5.6. Análise de correspondência.
- 2.1.5.2.6. Prover escalonamento multidimensional (Multidimensional Scaling - MDS).
- 2.1.5.2.7. Prover medidas de proximidade e distância.
- 2.1.5.2.8. Prover correlações canônicas e bivariadas.
- 2.1.5.2.9. Prover análise de confiabilidade.
- 2.1.5.2.10. Prover as seguintes técnicas de regressão:
 - 2.1.5.2.10.1. Linear;
 - 2.1.5.2.10.2. Logística;
 - 2.1.5.2.10.3. Não linear;
 - 2.1.5.2.10.4. Modelos lineares generalizados; e
 - 2.1.5.2.10.5. Multinomial.
- 2.1.5.2.11. Prover as seguintes funções de transformação:
 - 2.1.5.2.11.1. Logarítmica;
 - 2.1.5.2.11.2. Raiz quadrada;
 - 2.1.5.2.11.3. Inversa;
 - 2.1.5.2.11.4. Quadrática;
 - 2.1.5.2.11.5. Exponencial;
 - 2.1.5.2.11.6. Padronizada;
 - 2.1.5.2.11.7. Amplitude;
 - 2.1.5.2.11.8. Quantil;
 - 2.1.5.2.11.9. Agrupamento ótimo maximizando relacionamento com variáveis dependentes;
 - 2.1.5.2.11.10. Definida pelo usuário;
 - 2.1.5.2.11.11. Maximização da normalidade; e
 - 2.1.5.2.11.12. Maximização da correlação com variáveis dependentes.
- 2.1.5.2.12. Prover análise de sobrevivência.
- 2.1.5.2.13. Prover análises psicométricas.
- 2.1.5.2.14. Prover previsões de séries temporais usando:
 - 2.1.5.2.14.1. Modelagens econométricas;
 - 2.1.5.2.14.2. Modelo auto-regressivo integrado de média móvel (AutoRegressive Integrated Moving Average - ARIMA);
 - 2.1.5.2.14.3. Suavização exponencial;
 - 2.1.5.2.14.4. Dados de painel - combinação de informação sobre múltiplas causas (componente interseccional ou transversal) e informação sobre casos no decorrer do tempo (componente longitudinal ou de série temporal);

- 2.1.5.2.14.5. Análise espectral;
- 2.1.5.2.14.6. Interferência; e
- 2.1.5.2.14.7. Regressão polinomial.

2.1.5.3. Mineração de dados

2.1.5.3.1. Possuir os seguintes algoritmos para cada uma dos problemas de mineração de dados:

2.1.5.3.1.1. Classificação:

2.1.5.3.1.1.1.1. Máquina de vetores de suporte (Support Vector Machines - SVM);

2.1.5.3.1.1.1.2. Regressão logística;

2.1.5.3.1.1.1.3. Redes neurais artificiais;

2.1.5.3.1.1.1.4. Árvores de decisão;

2.1.5.3.1.1.1.5. Descoberta de regras; e

2.1.5.3.1.1.1.6. Redes bayesianas.

2.1.5.3.1.2. Regressão:

2.1.5.3.1.2.1. Regressão linear;

2.1.5.3.1.2.2. Interpolação e extrapolação;

2.1.5.3.1.2.3. Regressão não-linear;

2.1.5.3.1.2.4. Métodos bayesianos;

2.1.5.3.1.2.5. Redes neurais artificiais; e

2.1.5.3.1.2.6. Árvores de decisão para regressão.

2.1.5.3.1.3. Agrupamento (clustering):

2.1.5.3.1.3.1. Hierárquico;

2.1.5.3.1.3.2. K-means;

2.1.5.3.1.3.3. EM (Expectation Maximization); e

2.1.5.3.1.3.4. Mapas auto-organizáveis de Kohonen (Self-Organizing Maps - SOM).

2.1.5.3.1.4. Regras de associação:

2.1.5.3.1.4.1. Apriori e ou variações do algoritmo.

2.1.5.3.1.5. Detecção de anomalias (outliers):

2.1.5.3.1.5.1. Baseada em densidade (ex: k-nearest neighbor e local outlier factor);

2.1.5.3.1.5.2. Baseada em subespaço e correlação para dados multidimensionais; e

2.1.5.3.1.5.3. Máquinas de vetor de suporte de classe única.

2.1.5.3.2. Prover recursos para treinamento, teste e validação de modelos.

2.1.5.3.3. Possibilitar exploração e análise dos dados de forma interativa através de:

2.1.5.3.3.1. Análise de distribuições univariadas;

2.1.5.3.3.2. Investigação de distribuições multivariadas;

2.1.5.3.3.3. Análise de correlações;

2.1.5.3.3.4. Ajuste de modelos explicativos, incluindo:

2.1.5.3.3.4.1. Análise de variância;

2.1.5.3.3.4.2. Regressão; e

- 2.1.5.3.3.4.3. Modelos lineares generalizados.
- 2.1.5.3.3.5. Análise de associação (ex: market basket analysis); e
- 2.1.5.3.3.6. Análise de sequências.
- 2.1.5.3.4. Prover técnicas para recuperação de valores ausentes.
- 2.1.5.3.5. Prover filtragem de valores extremos (outliers), incluindo:
 - 2.1.5.3.5.1. Eliminação automática de valores extremos;
 - 2.1.5.3.5.2. Eliminação automática de valores raros para variáveis cujas ocorrências sejam menores do que um número especificado;
 - 2.1.5.3.5.3. Eliminação valores extremos, utilizando:
 - 2.1.5.3.5.3.1. Número de desvios em torno da mediana;
 - 2.1.5.3.5.3.2. Número de desvios em torno da média;
 - 2.1.5.3.5.3.3. Distância em relação à moda central; e
 - 2.1.5.3.5.3.4. Percentis extremos.
 - 2.1.5.3.5.4. Eliminação manual.
- 2.1.5.3.6. Prover funcionalidade de reposição automática de valores para:
 - 2.1.5.3.6.1. Variáveis intervalares, utilizando:
 - 2.1.5.3.6.1.1. Média;
 - 2.1.5.3.6.1.2. Mediana;
 - 2.1.5.3.6.1.3. Amplitude média;
 - 2.1.5.3.6.1.4. Baseado na distribuição;
 - 2.1.5.3.6.1.5. Imputação de dados utilizando árvore de decisão;
 - 2.1.5.3.6.1.6. Imputação de dados utilizando árvore de decisão levando em consideração os valores faltantes, substituindo a regra de imputação que os contenha por uma que não os contenha;
 - 2.1.5.3.6.1.7. Estimadores M; e
 - 2.1.5.3.6.1.8. Constantes definidas.
 - 2.1.5.3.6.2. Variáveis categóricas, utilizando:
 - 2.1.5.3.6.2.1. Valor que possui maior frequência;
 - 2.1.5.3.6.2.2. Baseado na distribuição;
 - 2.1.5.3.6.2.3. Imputação por árvore de decisão;
 - 2.1.5.3.6.2.4. Imputação de dados utilizando árvore de decisão levando em consideração os valores faltantes, substituindo a regra de imputação que os contenha por uma que não os contenha; e
 - 2.1.5.3.6.2.5. Constantes definidas.
- 2.1.5.3.7. Permitir seleção de variáveis, em especial:
 - 2.1.5.3.7.1. Remoção de variáveis não relacionadas à variável dependente utilizando Qui-Quadrado e R²;
 - 2.1.5.3.7.2. Remoção de variáveis que possuam relacionamento hierárquico a fim de evitar informação redundante;
 - 2.1.5.3.7.3. Remoção de variáveis que possuem muitos valores faltantes; e
 - 2.1.5.3.7.4. Categorização de variáveis intervalares para identificação de relacionamento não-

linear com variável dependente.

2.1.5.3.8. Prover as seguintes técnicas de modelagem preditiva:

2.1.5.3.8.1. Regressão, incluindo:

2.1.5.3.8.1.1. Regressão linear e logística; e

2.1.5.3.8.1.2. Seleções "forward", "backward" e "stepwise".

2.1.5.3.8.2. Árvores de decisão, incluindo:

2.1.5.3.8.2.1. Chi-squared Automatic Interaction Detector (CHAID);

2.1.5.3.8.2.2. Classification And Regression Tree (CART);

2.1.5.3.8.2.3. Árvore regressora; e

2.1.5.3.8.2.4. Crescimento e poda interativa de árvores.

2.1.5.3.8.3. Redes neurais, incluindo:

2.1.5.3.8.3.1. Arquitetura de redes flexível;

2.1.5.3.8.3.2. Funções de combinação;

2.1.5.3.8.3.3. Funções de ativação;

2.1.5.3.8.3.4. Técnicas de treinamento através dos algoritmos de Levenberg-Marquardt (LM), de Retro-propagação (Backpropagation) e método do gradiente descendente;

2.1.5.3.8.3.5. Otimização preliminar;

2.1.5.3.8.3.6. Padronização automática das entradas; e

2.1.5.3.8.3.7. Treinamento interativo.

2.1.5.3.8.4. Modelo MBR (Memory-Based-Reasoning), técnica que utiliza o algoritmo do vizinho mais próximo (k-Nearest Neighbour - KNN) para categorização ou predição de observações;

2.1.5.3.8.5. Geração de modelos preditivos definidos pelo usuário, permitindo a inserção de modelos que foram gerados a partir de técnicas criadas pelo mesmo;

2.1.5.3.8.6. Geração de modelos de dois estágios, com a finalidade de estimar uma variável intervalar baseada em classes de uma determinada variável categórica; e

2.1.5.3.8.7. Geração de modelos que permitam a união de técnicas de modelagem, para criação de um novo modelo, estimando as probabilidades posteriores (para variáveis dependentes categóricas) e valores preditos (para variáveis dependentes intervalares) provenientes de múltiplos modelos (redes neurais, regressões, MBR, árvore de decisão).

2.1.5.3.9. Prover funcionalidade de avaliação através de:

2.1.5.3.9.1. Curvas "lift" para avaliação simples e de múltiplos modelos;

2.1.5.3.9.2. Curvas ROC (Receiver Operating Characteristic);

2.1.5.3.9.3. Curvas de lucro, baseadas em uma matriz de custo/benefício;

2.1.5.3.9.4. Impacto marginal; e

2.1.5.3.9.5. Matriz de confusão.

2.1.5.3.10. Permitir a realização de processamento de variáveis categóricas em grupo, para análise de múltiplas variáveis alvo (variáveis dependentes).

2.1.5.3.11. Prover particionamento dos dados, incluindo:

2.1.5.3.11.1. Criação automática de amostras, utilizando técnicas estatísticas de amostragem para treinamento, validação e teste de modelos; e

2.1.5.3.11.2. Comparação de modelos.

2.1.5.3.12. Permitir alterar os atributos dos dados em quaisquer etapas da mineração de dados.

2.1.5.4. Mineração de textos

2.1.5.4.1. Prover técnicas e algoritmos para mineração de texto (text mining).

2.1.5.4.2. Possuir funções para mineração de textos que descubram padrões e relacionamentos não explícitos, ocultos em gigantescas massas de dados.

2.1.5.4.3. Permitir que métodos analíticos e funções preditivas possam ser usados no processamento de textos, integrando recursos estatísticos e de mineração de dados.

2.1.5.4.4. Prover os seguintes processos de análise textual:

2.1.5.4.4.1. Recuperação de informação proveniente de diversas fontes textuais; e

2.1.5.4.4.2. Processamento de linguagem natural (NLP - Natural Language Processing).

2.1.5.4.5. Prover as seguintes tarefas de mineração de textos:

2.1.5.4.5.1. Categorização de textos;

2.1.5.4.5.2. Agrupamento de textos, incluindo a técnica LDA (Latent Dirichlet Allocation);

2.1.5.4.5.3. Extração de conceitos e entidades;

2.1.5.4.5.4. Análise de sentimentos; e

2.1.5.4.5.5. Análise textual quantitativa (contagem e posicionamento de palavras).

2.1.5.4.6. Possuir dicionários de palavras para tratamento semântico durante a análise textual nos seguintes idiomas:

2.1.5.4.6.1. Inglês Americano (EN-US); e

2.1.5.4.6.2. Português Brasileiro (PT-BR).

2.1.5.5. Especializações

2.1.5.5.1. Possuir suíte de programas e regras predefinidas sob responsabilidade do mesmo fabricante visando proteção pró-ativa contra fraudes e crimes financeiros.

2.1.5.5.2. Permitir a exploração de padrões de comportamento de entidades baseada no seu histórico para descobrir ou identificar situações de risco e oportunidade de análises, em particular:

2.1.5.5.2.1. Detecção por regras, anomalias, relacionamentos, etc;

2.1.5.5.2.2. Capacidade de aprender e prever situações baseadas no comportamento das entidades e análises humanas realizadas; e

2.1.5.5.2.3. Disponibilização de formas de visualização como redes, histogramas e dispersão.

2.1.5.5.3. Prover construção de representações visuais interativas e dinâmicas de redes de relacionamento.

2.1.5.5.4. Possuir solução para coleta de dados provenientes de redes sociais e feeds de notícias, dentre as quais:

2.1.5.5.4.1. Twitter;

2.1.5.5.4.2. Facebook; e

2.1.5.5.4.3. Padrão RSS (Really Simple Syndication).

2.1.5.5.5. Prover algoritmos e técnicas de Social Networking Analysis (SNA), incluindo:

2.1.5.5.5.1. Detecção de laços fortes e fracos (strong/weak ties);

2.1.5.5.5.2. Homofilia e afiliação;

2.1.5.5.5.3. Análise de links, com detecção de hubs e authorities (PageRank);

2.1.5.5.5.4. Power laws;

2.1.5.5.5.5. Modelagem de small worlds;

- 2.1.5.5.5.6. Integração com redes sociais online; e
- 2.1.5.5.5.7. Métricas de centralidade de rede:
 - 2.1.5.5.5.7.1. Intermediação (betweness);
 - 2.1.5.5.5.7.2. Proximidade (closeness);
 - 2.1.5.5.5.7.3. Grau (degree); e
 - 2.1.5.5.5.7.4. Eigenvectors.
- 2.1.5.5.6. Prover funcionalidades voltadas à mineração de grafos, incluindo:
 - 2.1.5.5.6.1. Algoritmos de grafos esparsos (graph sparsifiers);
 - 2.1.5.5.6.2. Particionamento de grafos (graph partitioning), particularmente através de modularidade e detecção de comunidades;
 - 2.1.5.5.6.3. Subgrafos densos;
 - 2.1.5.5.6.4. Menor caminho;
 - 2.1.5.5.6.5. Árvore geradora mínima; e
 - 2.1.5.5.6.6. Detecção de quasi-clique.
- 2.1.5.5.7. Prover algoritmos e técnicas de mineração de informações geográficas, incluindo:
 - 2.1.5.5.7.1. Segmentação espacial (agrupamento e classificação);
 - 2.1.5.5.7.2. Dependência espacial (regras de associação espaciais);
 - 2.1.5.5.7.3. Detecção espacial de tendências; e
 - 2.1.5.5.7.4. Caracterização e generalização geográfica.

2.1.6. Acesso a dados

2.1.6.1. Arquivos

- 2.1.6.1.1. Prover importação de dados a partir dos seguintes formatos de arquivos:
 - 2.1.6.1.1.1. Microsoft Access (.mdb);
 - 2.1.6.1.1.2. Microsoft Excel (.xls e .xlsx);
 - 2.1.6.1.1.3. OpenDocument Spreadsheet (.ods);
 - 2.1.6.1.1.4. Arquivos de texto (.txt);
 - 2.1.6.1.1.5. Arquivos de texto separados por caracteres delimitadores (.csv e .tsv); e
 - 2.1.6.1.1.6. eXtensible Markup Language (.xml).
- 2.1.6.1.2. Prover capacidade de extrair conteúdo textual a partir dos formatos de arquivo indicados anteriormente, além dos seguintes formatos:
 - 2.1.6.1.2.1. HyperText Markup Language (.html);
 - 2.1.6.1.2.2. Portable Document Format (.pdf);
 - 2.1.6.1.2.3. Rich Text Format (.rtf);
 - 2.1.6.1.2.4. Microsoft Word (.doc e .docx); e
 - 2.1.6.1.2.5. OpenDocument Writer (.odt).
- 2.1.6.1.3. Prover manipulação de arquivos de texto em várias codificações (encodings), em especial ASCII, LATIN1 (ISO-8859-1) e Unicode (UTF-8).

2.1.6.2. Bancos de dados

- 2.1.6.2.1. Prover manipulação de dados, incluindo operações de leitura e escrita, nos seguintes

sistemas gerenciadores de bancos de dados relacionais:

2.1.6.2.1.1. IBM DB2 (versões 9 e superiores);

2.1.6.2.1.2. Oracle Database (versões 10 e superiores);

2.1.6.2.1.3. Microsoft SQL Server (versões 2005 e superiores);

2.1.6.2.1.5. Teradata Database (versões 12 e superiores);

2.1.6.2.1.5. PostgreSQL (versões 8.4 e superiores); e

2.1.6.2.1.6. MySQL (versões 5.5 e superiores).

2.1.6.2.2. Possibilitar acesso aos bancos de dados compatíveis através de tecnologias ODBC e JDBC.

2.1.6.2.3. Prover capacidade de visualizar os dados de forma limitada, determinando a quantidade máxima de registros a serem apresentados ao usuário.

2.1.6.2.4. Prover manipulação de dados massivos armazenados nas tecnologias:

2.1.6.2.4.1. Hadoop (HDFS); e

2.1.6.2.4.2. NoSQL, incluindo MongoDB.

2.1.6.3. Serviços web

2.1.6.3.1. Prover importação de dados através de chamadas a serviços web (webservices), incluindo os padrões:

2.1.6.3.1.1. SOAP (baseado no formato XML); e

2.1.6.3.1.2. REST (baseado no formato JSON).

2.1.7. Interface

2.1.7.1. Cliente

2.1.7.1.1. Possuir cliente desktop para a plataforma Windows (versões 7 e superiores).

2.1.7.1.2. Possuir cliente web compatível com os seguintes navegadores:

2.1.7.1.2.1. Mozilla Firefox (versões 24 ou superiores);

2.1.7.1.2.2. Internet Explorer (versões 9 ou superiores);

2.1.7.1.2.3. Google Chrome (versões 34 ou superiores); e

2.1.7.1.2.4. Safari (versões 5 ou superiores).

2.1.7.1.3. Possuir regionalização nos clientes nos seguintes idiomas:

2.1.7.1.3.1. Inglês Americano (EN-US); e

2.1.7.1.3.2. Português Brasileiro (PT-BR).

2.1.7.1.4. Possuir documentação do usuário com suporte a todas as técnicas disponíveis no idioma Português Brasileiro (PT-BR).

2.1.7.2. Desenvolvimento

2.1.7.2.1. Prover importação e exportação de modelos analíticos em formato Predictive Model Markup Language (PMML).

2.1.7.2.2. Permitir a programação visual de fluxos de mineração de dados.

2.1.7.2.3. Prover geração da documentação dos modelos analíticos criados na ferramenta.

2.1.7.2.4. Possuir um construtor ou editor de consultas que permita acessar e manipular os dados visualmente, com:

2.1.7.2.4.1. Habilidade de realizar junções várias tabelas simultaneamente;

- 2.1.7.2.4.2. Capacidade de agrupar e ordenar dados utilizando multiprocessamento;
- 2.1.7.2.4.3. Habilidade de criar visualmente colunas calculadas a partir de uma lista de funcionalidades preexistentes;
- 2.1.7.2.4.4. Habilidade de alterar a ordem das colunas de acordo com as necessidades;
- 2.1.7.2.4.5. Habilidade de realizar transposição de tabelas;
- 2.1.7.2.4.6. Capacidade analítica para realizar consultas e análises estatísticas descritivas dos dados, com todos os módulos de estatística tradicional e parte gráfica; e
- 2.1.7.2.4.7. Habilidade de criar tabelas com os dados por meio do editor de consultas, utilizando linguagem SQL ou linguagem nativa da ferramenta.
- 2.1.7.2.5. Possuir uma interface amigável, do tipo "point-and-click" e ou "drag-and-drop", com diagrama de fluxo para visualização gráfica dos processos e 2.1.7.2.6. possibilidade de inserir códigos de programação gerados manualmente na linguagem da ferramenta.
- 2.1.7.2.7. Prover as seguintes funcionalidades de gráficos e visualização:
 - 2.1.7.2.7.1. Geração de gráficos em modo batch (agendado); e
 - 2.1.7.2.7.2. Geração de gráficos dinâmicos e interativos (dashboards).
- 2.1.7.2.8. Possuir interface para execução de algoritmos na linguagem R.
- 2.1.7.2.9. Permitir desenvolvimento e inclusão de novos algoritmos via linguagem visual ou de scripting.
- 2.1.7.2.10. Permitir incorporação de programas na linguagem utilizada em todos os produtos da família do software permitindo a criação de modelos customizados e transformações complexas.
- 2.1.7.2.11. Prover suporte a desenvolvimento de novas funções em Map Reduce.
- 2.1.7.2.12. Prover execução de funções Map Reduce através de linguagem amigável, baseada na estrutura SQL (Structured Query Language).
- 2.1.7.2.13. Permitir publicação das visualizações no padrão de portlets e portais (JSR-268).
- 2.1.7.2.14. Prover API (interface de programação) na forma de SDK (kit de desenvolvimento) para as linguagens Java, C++, Python e R.
- 2.1.7.3. Visualização
 - 2.1.7.3.1. Prover técnicas de visualização de dados.
 - 2.1.7.3.2. Permitir plotagem em interface gráfica própria dos resultados dos processos analíticos.
 - 2.1.7.3.3. Prover a criação dos seguintes tipos de gráficos:
 - 2.1.7.3.3.1. Setores (Pizza);
 - 2.1.7.3.3.2. Linhas;
 - 2.1.7.3.3.3. Colunas (Barras);
 - 2.1.7.3.3.4. Boxplot (diagrama de caixa);
 - 2.1.7.3.3.5. Scatterplot (diagrama de dispersão);
 - 2.1.7.3.3.6. Pareto;
 - 2.1.7.3.3.7. Controle;
 - 2.1.7.3.3.8. Histograma;
 - 2.1.7.3.3.9. Treemap (mapa de árvore);
 - 2.1.7.3.3.10. Heatmap (mapa de calor);
 - 2.1.7.3.3.11. Superfície;
 - 2.1.7.3.3.12. Redes (grafos); e

2.1.7.3.3.13. Mapas geográficos, incluindo:

2.1.7.3.3.13.1. Coropletos (ex: mapas temáticos);

2.1.7.3.3.13.2. Símbolos proporcionais (ex: círculos graduados); e

2.1.7.3.3.13.3. Pontos (ex: marcadores).

2.1.7.3.4. Possibilitar ao usuário examinar o relacionamento entre as variáveis de forma visual, e com isso auxiliar a:

2.1.7.3.4.1. Detecção de valores extremos;

2.1.7.3.4.2. Descoberta de padrões, tendências e riscos; e

2.1.7.3.4.3. Geração de histogramas para variáveis discretas e contínuas.

2.1.7.3.5. Deve ser capaz de realizar técnicas de descoberta de dados e exploração visual, visando o desenvolvimento e refinamento de visões, análises e pesquisas de dados estruturados e não estruturados.

2.1.7.4. Exportação

2.1.7.4.1. Prover a exportação dos resultados dos processamentos analíticos nos seguintes formatos de arquivos:

2.1.7.4.1.1. HyperText Markup Language (.html);

2.1.7.4.1.2. Portable Document Format (.pdf);

2.1.7.4.1.3. Comma Separated Value (.csv);

2.1.7.4.1.4. Tab Separated Value (.tsv); e

2.1.7.4.1.5. eXtensible Markup Language (.xml).

2.1.7.5. Gerenciamento

2.1.7.5.1. Possuir administração centralizada por console único de gerenciamento.

2.1.7.5.2. Possuir monitor visual de alocação de recursos para os processos da ferramenta.

2.1.7.5.3. Permitir criação de alertas, capazes de monitorar eventos computacionais, como estouro de memória ou escassez de espaço disponível em disco.

2.1.7.5.4. Possuir mecanismo de monitoração (logs).

2.1.7.5.5. Permitir separação dos ambientes de desenvolvimento e produção.

2.1.8. Processo

2.1.8.1. Ambientes

2.1.8.1.1. Permitir separar ambientes de experimentação e produção.

2.1.8.1.2. Prover geração automática de documentação do fluxo de processos em HTML com todas as configurações selecionadas.

2.1.8.1.3. Possibilitar a implantação (deployment) automatizada de componentes (exemplo: novos algoritmos).

2.1.8.1.4. Permitir, dentro do fluxo de processos de um diagrama, a utilização de pontos de controle para reduzir o número de conexões entre os possíveis procedimentos.

2.1.8.1.5. Possibilitar criar subdiagramas com a finalidade de agrupar um conjunto de procedimentos em um único ponto de controle, com a finalidade de simplificar os diagramas dos processos.

2.1.8.1.6. Permitir a integração da solução de estatística e mineração de dados com sistemas transacionais de maneira automatizada, isto é, sem intervenção do usuário final.

2.1.8.2. Projetos

2.1.8.2.1. Prover a orientação do desenvolvimento de soluções por projetos.

2.1.8.2.2. Permitir trabalho colaborativo.

2.1.8.2.3. Permitir a orientação das funções por algum processo de mineração de dados (ex: CRISP-DM).

2.1.8.2.4. Prover o mapeamento dos usuários autenticados e suas permissões para atores em um processo de mineração de dados.

2.1.8.2.5. Prover integração com as seguintes soluções de ETL (extração, transformação e carga de dados):

2.1.8.2.5.1. Kettle - Pentaho Data Integrator (PDI); e

2.1.8.2.5.2. Informatica Power Center.

2.1.8.2.6. Permitir a realização do controle de versão dos modelos analíticos usando ferramentas especializadas como CVS, Git ou SVN.

2.1.8.2.7. Prover repositório único para toda a solução, armazenando as informações de usuários, restrições de acesso, informações de camada semântica e 2.1.8.2.8. armazenamento de relatórios desenvolvidos pelos usuários.

2.1.9. Infraestrutura

2.1.9.1. Prover arquitetura cliente-servidor, de forma que todos os cálculos sejam executados remotamente no servidor, sem que dados massivos precisem ser trafegados para os clientes.

2.1.9.2. Possibilitar que a execução das tarefas analíticas seja realizada no servidor de forma:

2.1.9.2.1. Instantânea, mediante requisição do usuário; e

2.1.9.2.2. Agendada, mediante definição de data e hora.

2.1.9.3. Realizar processamento distribuído de processos (nuvem, cluster).

2.1.9.4. Realizar processamento paralelo de algoritmos.

2.1.9.5. Prover execução de algoritmos na abordagem in-memory.

2.1.9.6. Prover capacidade de balanceamento e ou clusterização dos serviços executados pela própria ferramenta.

2.1.9.7. Prover publicação de funcionalidades no modelo:

2.1.9.7.1. SaaS (Software as a Service); e

2.1.9.7.2. PaaS (Platform as a Service) - através de API.

2.1.9.8. Prover monitoração de recursos computacionais alocados.

2.1.9.9. Permitir alocação automática e dinâmica de recursos computacionais.

2.1.9.10. Permitir o controle de custo computacional dos processos.

2.1.9.11. Prover manipulação de objetos em memória visando desempenho no acesso aos dados.

2.1.10. Segurança

2.1.10.1. Possuir controle de usuários por autenticação e autorização.

2.1.10.2. Prover autenticação de usuários integrada com os seguintes serviços de diretório:

2.1.10.2.1. LDAP (Lightweight Directory Access Protocol); e

2.1.10.2.2. Microsoft Active Directory (AD).

- 2.1.10.3. Permitir autenticação de usuários usando tecnologia de certificação digital padrão ICP Brasil.
- 2.1.10.4. Possibilitar a criação de contas de usuário no ambiente analítico com diferentes níveis de acesso de administração e operação.
- 2.1.10.5. Possuir arquitetura própria de segurança, independente da utilização de soluções de segurança de terceiros.
- 2.1.10.6. Possibilitar controle centralizado de segurança para atribuição de privilégios a níveis e perfis de usuários, grupos e projetos.
- 2.1.10.7. Possibilitar a restrição de acesso a um conjunto de dados específico, seja por registros, colunas ou ambos.
- 2.1.10.8. Possibilitar restrição de funcionalidades da solução para um determinado usuário, grupo, projeto ou perfil.
- 2.1.10.9. Possibilitar restrição de acesso a objetos criados pelos desenvolvedores para um determinado usuário, grupo ou projeto.
- 2.1.10.10. Possibilitar o controle de acesso de acordo com o perfil do usuário.
- 2.1.10.11. Possuir controle de autorização de acesso a dados no grão de tabela ou visão.
- 2.1.10.12. Possuir controle de autorização de execução de tarefas no grão de projeto.
- 2.1.10.13. Possuir controle de autorização de edição de fluxos no grão de projeto, com funcionalidades habilitadas ou desabilitadas por perfil ou usuário.
- 2.1.10.14. Possuir mecanismos de monitoração de tarefas e funções executadas (logs).

3.0 Níveis de serviço e sancionamentos

3.1 Níveis de Serviço Aplicáveis a Solução de Appliance de DW

3.1.1. Suporte técnico ao(s) hardware(s) ofertado(s):

- 3.1.1.1. Possuir suporte técnico para o(s) hardware(s) ofertado(s), durante o período de vigência do contrato, assegurando prazos de atendimento compatíveis com a instalação, ou seja, 24 (vinte e quatro) horas por dia e sete (7) dias por semana (à exceção dos chamados de Severidade 4).
- 3.1.1.2. O atendimento aos chamados deverá obedecer a seguinte classificação quanto ao nível de severidade:

Severidade	Descrição		Tempo de Atendimento	Tempo de Solução	Observações
1 - Crítica	Chamados referentes a situações de emergência ou problema crítico, caracterizados pela existência de ambiente paralisado	On-site	No máximo 2 (duas) horas após a abertura do chamado, incluindo percurso do técnico até as instalações do SERPRO	No máximo 6 (seis) horas após a abertura do chamado	O atendimento não poderá ser interrompido até o completo restabelecimento do produto envolvido, mesmo que se estenda por períodos noturnos e dias não úteis.
2 - Alta	Chamados associados a situações de alto impacto, incluindo os casos de degradação severa de desempenho	On-site	No máximo 2 (duas) horas após a abertura do chamado, incluindo percurso do técnico até as instalações do SERPRO	No máximo 8 (oito) horas após a abertura do chamado	O atendimento não poderá ser interrompido até o completo restabelecimento do produto envolvido, mesmo que se estenda por períodos noturnos e dias não úteis.
3 - Média	Chamados referentes a situações de baixo impacto ou para aqueles problemas que se apresentem de forma intermitente, incluindo os casos em que haja a necessidade de substituição de componente(s) que possua(m) redundância	Remoto, com exceção das situações em que seja necessária intervenção física	No máximo 4 (quatro) horas após a abertura do chamado	No máximo 10 (dez) horas após a abertura do chamado	Caso o problema não possa ser resolvido remotamente, a CONTRATADA deverá colocar à disposição do SERPRO, um especialista devidamente habilitado e credenciado que trabalhará o tempo que for necessário para a solução do problema, sendo que o ônus financeiro de tal providência será da CONTRATADA.
	Chamados com objetivo de solicitar acompanhamento técnico presencial para o desligamento e posterior ligação do(s) equipamento(s), em virtude de atividade programada	On-site	No máximo 4 (quatro) horas após a abertura do chamado	Conforme agendamento	O atendimento deverá ser realizado conforme o agendamento, mesmo que contemple períodos noturnos e dias não úteis.
4 - Baixa	Chamados com objetivo de sanar dúvidas quanto ao uso ou à implementação do produto	Remoto	No máximo 24 (vinte e quatro) horas após a abertura do chamado	No máximo 72 (setenta e duas) horas após a abertura do chamado	Os chamados classificados com Severidade 4 serão atendidos em horário comercial, ou seja, das 08:00 h. às 18:00 h., de segunda-feira a sexta-feira, horário de Brasília

3.1.1.3 Será aberto um chamado técnico para cada problema reportado, sendo iniciada a contagem do tempo de atendimento a partir da hora de acionamento.

3.1.1.4 Por necessidade de serviço, o SERPRO poderá solicitar a escalonamento de chamado para níveis superiores de severidade. Os prazos dos chamados escalados passam a contar novamente do início.

3.1.2 Suporte técnico ao(s) software(s) ofertado(s):

3.1.2.1 Possuir suporte técnico remoto para o(s) software(s) ofertado(s), durante o período de vigência do contrato, assegurando prazos de atendimento(s) compatíveis com a instalação, ou seja, 24 (vinte e quatro) horas por dia e sete (7) dias por semana (à exceção dos chamados de Severidade 3 e 4).

3.1.2.2 O atendimento aos chamados deverá obedecer à seguinte classificação quanto ao nível de severidade:

Severidade	Descrição	Tipo de Atendimento	Tempo de Atendimento	Tempo de Solução	Observações
1 - Crítica	Chamados referentes a situações de emergência ou problema crítico, caracterizados pela existência de ambiente paralisado	Remoto	No máximo 2 (duas) horas após a abertura do chamado	No máximo 6 (seis) horas após a abertura do chamado	Caso o problema não possa ser resolvido remotamente, a CONTRATADA deverá colocar à disposição do SERPRO, um especialista devidamente habilitado e credenciado que trabalhará o tempo que for necessário para a solução do problema, sendo que o ônus financeiro de tal providência será da CONTRATADA. Se após 3 (três) horas de iniciado o atendimento remoto ao chamado, o ambiente afetado não tiver sido restabelecido, o atendimento on-site deverá ser iniciado em até 4 (quatro) horas do início do atendimento remoto. O atendimento não poderá ser interrompido até o completo restabelecimento do produto envolvido, mesmo que se estenda por períodos noturnos e dias não úteis.
2 - Alta	Chamados associados a situações de alto impacto, incluindo os casos de degradação severa de desempenho	Remoto	No máximo 2 (duas) horas após a abertura do chamado	No máximo 8 (oito) horas após a abertura do chamado	Caso o problema não possa ser resolvido remotamente, a CONTRATADA deverá colocar à disposição do SERPRO, um especialista devidamente habilitado e credenciado que trabalhará o tempo que for necessário para a solução do problema, sendo que o ônus financeiro de tal providência será da CONTRATADA. Se após 4 (quatro) horas de iniciado o atendimento remoto ao chamado, o ambiente afetado não tiver sido restabelecido, o atendimento on-site deverá ser iniciado em até 5 (cinco) horas do início do atendimento remoto. O atendimento não poderá ser interrompido até o completo restabelecimento do produto envolvido, mesmo que se estenda por períodos noturnos e dias não úteis.
3 - Média	Chamados referentes a situações de baixo impacto ou para aqueles problemas que se apresentem de forma intermitente	Remoto	No máximo 4 (quatro) horas após a abertura do chamado	No máximo 72 (setenta e duas) horas após a abertura do chamado	Caso o problema não possa ser resolvido remotamente, a CONTRATADA deverá colocar à disposição do SERPRO, um especialista devidamente habilitado e credenciado que trabalhará o tempo que for necessário para a solução do problema, sendo que o ônus financeiro de tal providência será da CONTRATADA. Os chamados classificados com Severidade 3 serão atendidos em horário comercial, ou seja, das 08:00 h. às 18:00 h., de segunda-feira a sexta-feira, horário de Brasília.
4 - Baixa	Chamados com objetivo de sanar dúvidas quanto ao uso ou à implementação do produto	Remoto	No máximo 24 (vinte e quatro) horas após a abertura do chamado	No máximo 120 (cento e vinte) horas após a abertura do chamado	Os chamados classificados com Severidade 4 serão atendidos em horário comercial, ou seja, das 08:00 h. às 18:00 h., de segunda-feira a sexta-feira, horário de Brasília.

3.1.2.3 Será aberto um chamado técnico para cada problema reportado, sendo iniciada a contagem do tempo de atendimento a partir da hora de acionamento.

3.1.2.4 Por necessidade de serviço, o SERPRO poderá solicitar a escalonamento de chamado para níveis superiores de severidade. Os prazos dos chamados escalados passam a contar novamente do início.

3.1.2.5 Em quaisquer casos e quando necessário, a CONTRATADA deverá assistir remotamente na instalação e uso dos software(s) ofertado(s), fornecendo orientações para diagnóstico de problemas e ajuda na interpretação de traces, dumps e logs. Nos casos de defeitos não conhecidos, as documentações enviadas pelo SERPRO (tais como: traces, dumps e logs) deverão ser encaminhadas ao laboratório da solução a fim de que sejam fornecidas as devidas correções.

3.1.2.6 Em quaisquer casos e quando necessário, a CONTRATADA deverá fornecer informações sobre as correções a serem aplicadas ou a própria correção.

3.1.3 Da Manutenção.

3.1.3.1 Para os equipamentos ofertados, a CONTRATADA deverá realizar manutenção preventiva, tanto do hardware quanto do firmware e software instalados, sendo de responsabilidade da

CONTRATADA prover todas as correções e atualizações necessárias, de forma sistemática e programada, de acordo com a periodicidade e os procedimentos especificados no(s) manual(is) do fabricante.

3.1.3.1.1 Caso não haja recomendação específica quanto à periodicidade, a CONTRATADA deverá apresentar, em até 60 (sessenta) dias após a assinatura do contrato, um cronograma de manutenções preventivas, que deverá ser aprovado pelo SERPRO.

3.1.3.1.2 A CONTRATADA deverá entregar, a cada manutenção preventiva e corretiva realizada, relatório técnico contendo os procedimentos executados.

3.1.3.1.3 Todas as peças de reposição deverão ser novas e sem uso.

3.1.3.2 No caso de manutenções, preventivas ou corretivas, em que haja risco de indisponibilidade total ou parcial dos equipamentos, o SERPRO deverá ser previamente notificado para que se proceda a aprovação e o agendamento da manutenção em horário conveniente ao SERPRO.

3.1.3.3 Para os equipamentos ofertados, a CONTRATADA poderá ser acionada pelo SERPRO até 3 (três) vezes, para efetuar o serviço de desligar e religar os equipamentos, durante a vigência do contrato.

3.1.4 Canais de atendimento.

3.1.4.1 Atendimento através de canal telefônico gratuito 0800, 24 horas por dia, 7 dias por semana, tanto para o hardware quanto para o software.

3.1.4.2 Chamado técnico através de site na Internet, 24 horas por dia, 7 dias por semana e/ou canal telefônico gratuito 0800.

3.1.4.3 Acionamento automático da CONTRATADA no caso de falha de quaisquer dos componentes do(s) equipamento(s) ofertado(s).

3.1.5. Penalidades:

3.1.5.1 O não atendimento dentro do prazo estabelecido para o chamado ensejará aplicação de multa à CONTRATADA, conforme o nível de severidade do mesmo:

Severidade 1 – 1,0% (um por cento) do valor mensal dos serviços de manutenção, atualização e suporte técnico constante no contrato, por hora ou fração de hora de atraso.

Severidade 2 – 0,8% (oito décimos por cento) do valor mensal dos serviços de manutenção, atualização e suporte técnico constante no contrato, por hora ou fração de hora de atraso.

Severidade 3 – 0,6% (seis décimos por cento) do valor mensal dos serviços de manutenção, atualização e suporte técnico constante no contrato, por hora ou fração de hora de atraso.

Severidade 4 – 0,4% (quatro décimo por cento) do valor mensal dos serviços de manutenção, atualização e suporte técnico constante no contrato, por hora ou fração de hora de atraso.

3.1.6. Relatórios sobre a Prestação dos Serviços

3.1.6.1. A CONTRATADA deverá emitir mensalmente um relatório constando os acionamentos técnicos efetuados no período, por regional, com o mínimo de informações de número de acionamento, descrição da ocorrência, severidade, nome do responsável do SERPRO pela abertura do chamado, data e hora de abertura do chamado, data e hora do início do atendimento, data e hora de solução ou contorno e descrição da resolução adotada. O relatório deverá ser entregue mesmo quando não houver chamados no período.