



# **Consulta Pública Eletrônica para Contratação de Solução de Virtualização de Dados.**

<b>1.Objeto.....</b>	<b>4</b>
2.Especificação do Objeto .....	4
<b>3.Arquitetura Conceitual de Dados .....</b>	<b>17</b>
<b>4.Publicação .....</b>	<b>18</b>
<b>5.Período .....</b>	<b>18</b>
<b>6.Responsáveis.....</b>	<b>18</b>

Às empresas interessadas devem responder a consulta pública com as seguintes informações:

**1. Contatos:**

- 1.1. Nome completo do responsável pelas respostas desta Consulta Pública.
- 1.2. Cargo, telefones e endereço de e-mail.

**2. Identificação da Empresa:**

- 2.1. Nome completo e fantasia.
- 2.2. CNPJ.
- 2.3. Endereço completo.
- 2.4. Site WEB (www).

**3. Solução**

- 3.1. Nome da solução oferecida, objeto desta consulta pública.
- 3.2. Site WEB do fabricante da solução (www).
- 3.3. Descrição detalhada da solução e seus componentes (Documentos/datasheet, etc).
- 3.4. Forma de licenciamento da solução e seus componentes (Licença perpétua /subscrição anual, e outras ), conforme exemplo abaixo:

Part Number	Descrição da Solução	Licença de uso perpétuo / subscrição anual	Unidade / Métrica	Faixa / Quantidade	Estudo de referência do Valor Unitário (R\$)

- 3.5. Forma e condições de pagamento da solução e seus componentes (Licença perpétua /subscrição anual, etc ).

**4. Base de Clientes:**

- 4.1. Quantidade de clientes no Brasil.
- 4.2. Nomes dos entes públicos que já adquiriram a solução.

**5. Experiência e Suporte:**

- 5.1. Possui equipe de suporte técnico para atendimento fora do horário comercial e em dias não úteis.
- 5.2. O suporte é prestado pelo fabricante ou parceiro?
- 5.3. Quais os níveis de serviços ofertados para a solução (Tempo de atendimento, tempo de
- 5.4. Informar a forma de repasse de conhecimento, resumos das grades e carga horária.

# 1.Objeto

O objetivo desta consulta é subsidiar decisão futura sobre contratação de solução de virtualização de dados, a ser inserida na arquitetura de dados do Serpro, de modo a sustentar a disponibilidade, consumo e bilhetagem de dados alinhados com os objetivos empresariais. As origens dos dados a serem virtualizados são bases hospedadas nos ambientes: data lake, data warehouse, cloud platform, SGBDs tradicionais, Mainframes mantidas pelo Serpro. E os destinos são os diversos órgãos governamentais.

## 2.Especificação do Objeto

### 2.1. Tabela de requisitos funcionais

ID	Requisitos Funcionais	Detalhamento dos Requisitos	Forma de Atendimento	
			Nativo (N) / Nativo por Parametrização (NP)	Customizado (C) / Não Atende (NA)
1	Fonte de Dados (datasources)	Conectividade à fonte de dados relacionais, no mínimo Teradata, Oracle Exadata, Oracle, Sql Server, PostgreSQL, IBM DB2, My SQL e Maria DB		
2		Conectividade à fonte de dados mainframe, no mínimo Adabas e DB2		
3		Conectividade a fonte de dados não relacionais (NoSQL), no mínimo Elasticsearch, MongoDB, Cassandra, Redis, CouchDB, MarkLogic e Neo4J		
4		Conectividade a fonte de dados em Web Service, SOAP, REST (XML, RSS, ATOM, JSON)		

5		Conectividade à fonte de dados de arquivos semiestruturados como: XML, JSON, TSV, CSV		
6		Conectividade à fonte de dados de arquivo não estruturados como: TXT, PDF, MS Word, MS Excel		
7		Conectividade à fontes de dados em Data Lake Storage, no mínimo S3, Azure Data Lake Storage, Google Cloud Storage, Parquet, Avro e ORC		
8		Conectividade à fontes de dados disponíveis nas tecnologias e/ou protocolos JDBC, ODBC, OData, NFS, SAN, HTTP/HTTPS, FTP/SFTP/FTPS, Open API, GraphQL, OAuth, SAML, SSL		
9		Conectividade e integração a protocolos de serviços de diretórios: LDAP e AD		
10		Conectividade à infraestrutura de mensageria, incluindo aquelas providas por middleware de integração, no mínimo Kafka, RabbitMQ, IBM MQ, TIBCO EMS, JMS e MQTT, MQSeries		
11		Conectividade a ambientes baseados em Apache Hadoop (Big Data com as seguintes tecnologias Hive, Impala, Spark, Kafka, Kudu, Solr, Hadoop HDFS, HBase, HCatalog e Databricks.		
12		Conectividade a séries temporais e dados geoespaciais residentes em sistemas como GIS, ESRI e outros.		

13		Conectividade com soluções proprietárias, de provedores de nuvem: Amazon RDS, Amazon Redshift, Amazon Athena, Amazon DynamoDB, Amazon Aurora, Amazon Glacier, Azure SQL, Azure Cosmos DB, Azure Synapse Analytics, Snowflake, Google Cloud SQL, Google BigQuery, e MongoDB Atlas		
14		Permitir consultas externas aos dados (exposição dos dados) dos virtuais databases (VDBs) via protocolo JDBC e ODBC		
15		Permitir conectividade com ferramentas analíticas e de integração para consumo de dados: MS PowerBI, Tableau, Qlik Sense, Visual insight, Talend, Pentaho.		
16	User Interface (UI)	A solução deve apresentar interface de gerenciamento completo a ser acessado em navegadores web de versões mais recentes (Firefox e Chrome) para sistema operacional Linux e Windows		
17		A interface gráfica não deve ser dependente das tecnologias Java ou Flash por questões de compatibilidade com alguns navegadores		
18		Como alternativa, em caso de falha da interface por navegador web, a solução deve permitir a execução de todas as operações por outros meios de acesso, como linha de comandos ou interface desktop nas plataformas Windows e Linux		
19		Interface administrativa para gerenciar implantações, bem como		

		administrar a infraestrutura in cloud		
20		Ferramenta de desenvolvimento (studio) deve ser simples e de fácil entendimento, intuitiva, com possibilidade da utilização de recursos "drag-and-drop" and "low-code" para o desenvolvimento dos Virtual Databases (VDB), assim como deve possuir ambiente único para realização de tais tarefas		
21		Deve apresentar visualização facilitada para exibir estatísticas de uso operacional, limites de capacidade de armazenamento e gráficos de performance.		
22		A solução deve ser distribuída na modalidade <i>on-premise</i> para execução em ambiente dedicado (bare metal) ou virtualizado		
23		Permitir a execução da solução nos seguintes sistemas operacionais, preferencialmente livres: Server side - (Microsoft 32-bit / 64-bit plataformas): Windows Server 2019 ou superior (Linux 32-bit / 64-bit plataformas): Red Hat Enterprise Linux RHEL 7 ou superior, CentOS 7 ou superior Client side - (Microsoft 32-bit / 64-bit plataformas): Windows 10 ou superior (Linux 32-bit / 64-bit plataformas): Ubuntu 18.04 LTS ou superior		
24	Monitoramento	Interface centralizada para monitoramento, auditoria, e troubleshooting		



25		Possui recursos de monitoramento da saúde do ambiente produtivo em "real time". E as disponibiliza de forma gráfica, de modo a facilitar a gestão		
26		Deve apresentar visualização facilitada para exibir estatísticas de uso operacional, limites de capacidade de armazenamento e gráficos de performance		
27		Possuir mecanismo de monitorar qualquer alteração estrutural nas origens já mapeadas nos VDBs		
28		Possui recursos de alertas por email, sms e etc na eventualidade de ocorrer problemas críticos no ambiente, erros de configuração, indisponibilidade		
29		Possuir recurso de monitoramento e diagnóstico visando entregar as informações de monitoramento para análises históricas e em tempo real		
30		Permitir monitoração compatíveis com os padrões SNMP e JMX, sendo possível integrar com ferramentas de mercado: HP OpenView, Nagios, IBM Tivoli, MS WinRM, Zabbix		
31		Ter métricas para verificação de consumo de CPU, Memória, volume de tráfego de rede, frequência de execução de tarefa (job/fluxo/pipeline). Tais métricas devem ser aferidas na origem, na própria ferramenta		
32	Segurança	Possibilidade de utilizar SSL/TLS (1-way) ou (2-way), criptografando o canal de comunicação com uso de chave pública, bem como certificado		

		digital (segurança a nível do canal de comunicação)		
33		A solução deve estar aderente ao Regulamento Geral sobre a Proteção de Dados 2016/679 (General Data Protection Regulation – GDPR) da União Europeia e à Lei Geral de Proteção de Dados (LGPD), Lei nº 13.709/2018		
34		O acesso à interface gráfica deve poder ser restrito a usuários (perfis) definidos pela área técnica		
35		Possibilidade de restrição de acesso a bases a nível de linha e coluna (política de segurança)		
36		Possibilidade de criar uma blacklist de tabelas que não poderão integrar VDBs, bem como, se possível descrever a descrição da restrição ou normativo		
37		Deve ser capaz de configurar perfis de usuários (roles e grupos) de modo a permitir operações distintas (a todos os bancos de dados) por diferentes tipos de usuários		
38		Possuir interface gráfica para administração da distribuição e mascaramento integrados e centralizados		
39		Suporte a mascaramento dinâmico de dados, em tempo real, quando um usuário acessa dados, com base em privilégios de acesso		
40	Autenticação / Autorização	A plataforma deve ser capaz de se integrar nativamente ao diretórios corporativos AD / LDAP para		

		permitir a autenticação centralizada dos usuários		
41		A plataforma deve ser capaz de se integrar OAuth e SAML		
42		A plataforma deve ser capaz de se integrar com Kerberos		
43		Possuir interface gráfica para gestão dos usuários cadastrados nos diretórios corporativos AD / LDAP		
44	Catálogo de Metadados	Permitir recurso de gestão de metadados para consumo interno e externo, possibilitando, ainda, consulta com diferentes perfis de acesso.		
45		Permitir exploração de dados e data discovery dos dados e metadados		
46		Permitir o cadastramento de categorias de negócios e tags de negócios		
47		Possuir capacidade de realizar busca simples e avançada nos metadados		
48		Permitir exibição de dados de conexão das bases virtualizadas na interface de metadados		
49		Permitir exportação de dados das coleções de metadados para diversos formatos, tais como: CSV, XLX, Tableau data		
50		Possuir mecanismos de linhagem de dados		
51		Sincronização unidirecional e/ou bidirecional metadados, no mínimo, descrição de tabela e descrição de colunas		
52		Identificar de forma automática o domínio de dados de cada atributo		

		da base virtualizada		
53		Possibilidade de marcar dados sensíveis consoantes com a LGPD		
54		Funcionalidade para importar e exportar metadados, com interface RESTful para armazenar e recuperar metadados		
55		Permitir recursos de rastreabilidade para garantir análise de impacto da mudança, contendo árvore de dependência		
56		API para publicar metadados, incluindo informações de linhagem para ferramenta de governança de dados		
57		Permitir integração com outras ferramentas de catálogos de dados		
58		Permitir a geração das logs em arquivos (arquivos de logs) e em banco de dados		
59	Auditoria	Criar Logs de eventos que registrem todas as operações no ambiente, sendo possível realizar buscas por: consultas realizada, identificação do usuário, IP do usuário, resultado da consulta (success, cancel, error), data/hora de início e fim da operação, duração, quantidade de linhas retornadas		
60		Permitir identificar nos logs registros de erros, warning e info referentes as etapas de configuração de ambiente, datasources e demais configurações		
61		Permitir controle de bilhetagem de acordo com as linhas consumidas		

62	Agendamento (Schedule)	Possuir interface administrativa de agendamento de jobs, permitindo as ações de habilitar, desabilitar, parar, executar, remover, adicionar, filtrar, acompanhar o status de execução, periodicidade		
63	Data Quality	Possuir biblioteca, extensão ou módulo de transformação, filtro, funções de matching e regras de qualidade para validar, limpar, enriquecer, padronizar, combinar e mesclar dados		
64		Possuir recurso para realização de data profiling ou possuir integração com soluções de data quality.		
65	Modelagem de Dados	A plataforma deve ser capaz de se integrar com soluções externas de modelagem de dados, pelo menos uma das listadas: ER/Studio Data Architect, ERWin Data Modeler, IBM Data Architect, Power Designer		
66	Escalabilidade / desempenho	Possuir mecanismo de aceleração de consultas para análises (otimização de query - agregações, consolidações)		
67		Permitir a visualização do plano de execução de consultas no ambiente do virtualizador, bem como a configuração para melhoria de desempenho		
68		Deve suportar mecanismos de escalabilidade horizontal e vertical, bem como alta disponibilidade		
69		Possuir integração de processamento paralelo massivo MPP (Impala, Presto, Spark, Databricks) visando otimizar consultas e cache		

70	Operação	Possibilidade de criar cópias de um database virtual (VDBs) para cada cliente		
71		Permitir a utilização de filtros por meio de queries, seja de forma gráfica ou em editores SQL		
72		Possibilidade de utilizar functions ou linguagens de programação específicas de SGBDs		
73		Possuir compatibilidade com o padrão SQL ANSI para construções de filtros (queries)		
74		Possibilidade de realizar cruzamento de dados entre databases diferentes de uma mesma origem, bem como de origens e tecnologias diferentes		
75		Possibilidade de realizar operação de leitura, bem como leitura e escrita diretamente na origem.		
76		Conectividade para consumo de informações por meio dos protocolos padrões: JDBC, ODBC, OData e Rest		
77		Contemplar recursos de Continuous Integration/Continuous Development (CI/CD), bem como integração repositório de controle de versão Git		
78		Permitir a criação de uma base de dados virtualizada para atender vários clientes distintos, e ainda assim, conseguir distinguir o processo de bilhetagem, auditoria de cada um deles		
79		Permitir configuração/personalização da base virtualizada com a possibilidade de definir limites de		

		caracteres para campos do tipo CLOB		
80		Permitir configuração/personalização da base virtualizada com a possibilidade de definir limites de consumo, como por exemplo: 20 milhões de registros / mês. Na hipótese de extrapolar o limite, o acesso à base deve ser bloqueado.		
81		A solução deve ser capaz de construir e consumir dados dos Virtuais Databases (VDBs) de diversas fontes de origem sem ter que, obrigatoriamente, realizar cópias e movimentação de dados		
82		Permitir a realização de transformações de dados considerando operações de conversão de tipo de dados, formatação e consolidação de dados		
83		Permitir configuração/personalização da base virtualizada com a possibilidade de definir limites de usuários logados.		
84		Permitir exportação de dados das bases virtualizadas para arquivos abertos, mínimamente: CSV, TXT		
85	Mascaramento	Funcionalidade para Identificar dados confidenciais em colunas, em arquivos de formato não estruturados (PDF, arquivos do Excel, arquivos de texto, arquivos de log etc.) e conteúdos semi estruturados (XML, JSON etc.)		

86		Funcionalidade para proteger os dados confidenciais sem prejudicar o projeto, por meio da remoção, mascaramento ou simplificação dos dados considerados confidenciais		
87		Funcionalidade para mascarar dados confidenciais com diversas formas e tipos, como nomes, data/hora, números, documentos pessoais, endereços de e-mail, CEP, cartão de crédito e tipo sanguíneo		
88		Funcionalidade para incluir ou customizar conjunto de regras que definem os campos que serão mascarados e a função de mascaramento que será usada		
89		Funcionalidade para permitir criação de formatos de mascaramento definidos pelo usuário		

## 2.2. Tabela de requisitos não funcionais

ID	Requisitos Funcionais	Detalhamento dos Requisitos	Forma de Atendimento	
			Nativo (N) / Nativo por Parametriza ção (NP)	Customizado (C) / Não Atende (NA)
1	Interface com o usuário	Interfaces de usuário compatíveis com distribuições Ubuntu, Windows e MacOS em estações de trabalhos, navegadores Google Chrome, Firefox, Apple Safari e Microsoft Edge. A interface com o usuário em navegadores não		



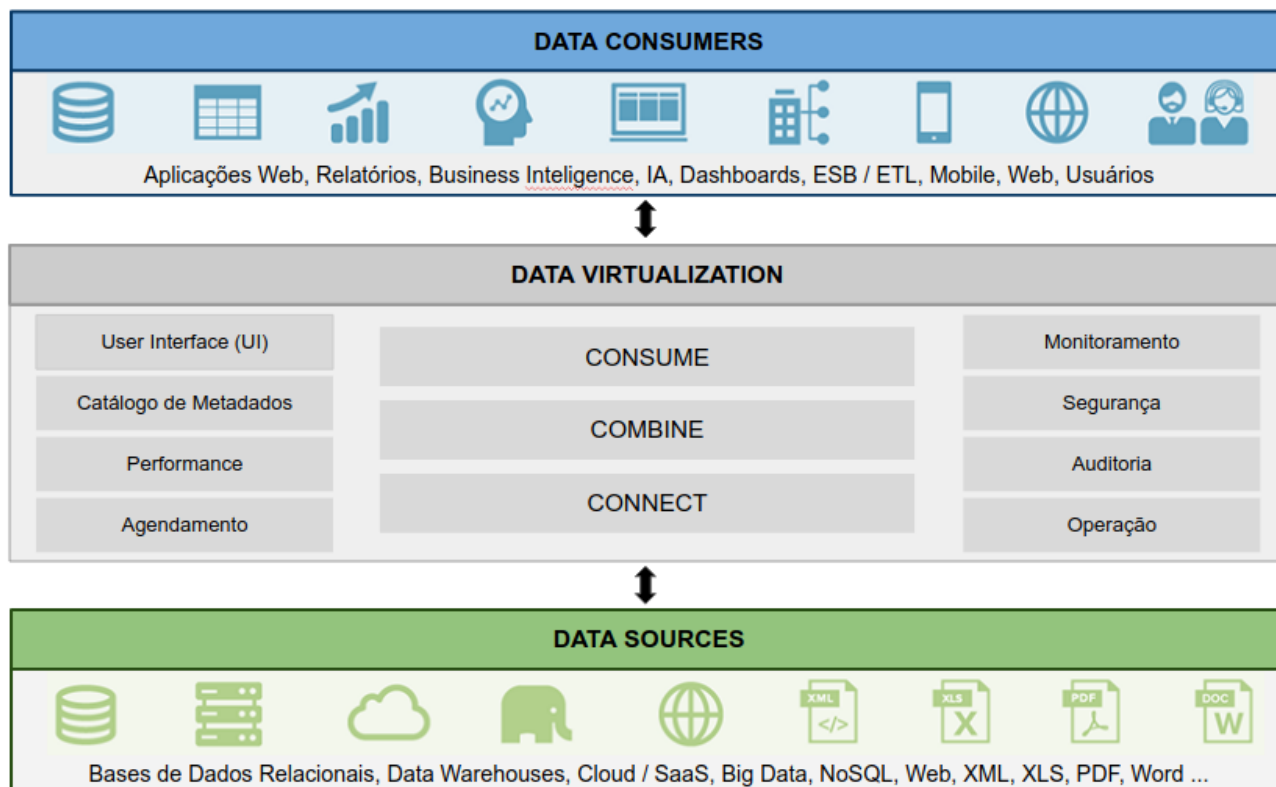
		poderá fazer uso de Applets Flash e Java.		
2		Interface gráfica (GUI – Graphical User Interface) e interface de linha de comando (CLI – Command Line Interface). A interface gráfica deve conter, no mínimo, Organização hierárquica por projetos, Possibilidade de reutilização de scripts, Interface gráfica adequada para uso por perfis de usuário com menor qualificação técnica ou restrição de acesso a módulos administrativos, como analistas de negócio, permitindo que os mesmos desenvolvam seus processos de forma self-service com independência da equipe de TI, incluindo interface a ser apresentada no idioma de escolha do usuário com documentação on-line.		
3		Possibilidade de customização de painéis para monitoramento de execução de tarefas e problemas.		
4		A solução deve incluir em seu licenciamento todos os componentes, complementos, extensões e serviços necessários ao seu pleno funcionamento durante toda a vigência do contrato. A solução deve também, conter todos os componentes necessários para o atendimento aos requisitos apresentados sem depender de aquisições adicionais.		

5	Integrações	Fornecimento de bibliotecas de desenvolvimento de software (SDK) e permita que os usuários criem conectores adicionais para fontes de dados exclusivas.		
6	Segurança	Log e auditoria, permitindo identificar toda operação realizada no produto, contendo as informações: responsável, data e hora da operação, consulta realizada, IP de origem e destino, identificador da conexão, eventos da conexão, quantidade de linhas retornadas, nome da base virtual utilizada, etc.		
7		Segurança granular baseada em perfis, permitindo associar funcionalidades da ferramenta a determinados perfis.		
8	Performance	Processamento de consultas que permita retornar, no mínimo, 15 mil linhas por segundo.		
9		Capacidade de processar, no mínimo, 140 mil consultas por dia		
10		Gerenciar requisições simultâneas de, no mínimo, 1000 por minuto		
11		Capacidade mínima de extração de 600 milhões de linhas por dia		

### 3.Arquitetura Conceitual de Dados

Com esta arquitetura, o Serpro deseja modernizar seu gerenciamento de dados e padrões arquitetônicos que, quando combinados, possam suportar cargas de trabalho cada vez mais complexas, diversas e distribuídas e possam permitir a integração, o compartilhamento e o controle dos dados, além de garantir alinhamento da abordagem com os requisitos de negócios, combinando as características e os casos de uso comuns de cada uma dessas estruturas data lake, data warehouse,

cloud platform, cache e data set.



## 4.Publicação

4.1. Consulta pública com fulcro no Art. 31, da Lei nº 9.784/1999, objetivando esclarecimentos sobre produtos, processos, soluções e tecnologias junto ao mercado.

## 5.Período

5.1. A consulta pública eletrônica ficará publicada pelo período de 15 (quinze) corridos.

## 6.Responsáveis

6.1 A Consulta Pública Eletrônica será acompanhada pelos empregados:

6.1.1. Leandro Dutra e Silva, matrícula 2107216-7, lotado na DIOPE/SUPEC/ECTAN/ECTPA, Telefone: (61) 2021-8726, e-mail: leandro.silva@serpro.gov.br.

6.1.2. Bruno Pacheco de Assis, matrícula: 01065246, lotado na DIOPE/SUPEC/ECTAN/ECTPA, Telefone: (11) 2173-1312, e-mail: bruno.pacheco@serpro.gov.br

6.1.3. Ricardo Silva Teixeira, matrícula 2107889-0, lotado na DIDES/SUPAI/AIGSB/AIPSA, e-mail: ricardo.teixeira@serpro.gov.br

6.2 Toda comunicação será realizada através do email cpvirtualizacaodedados@serpro.gov.br