

Estudo Técnico Preliminar 16/2021

1. Informações Básicas

Número do processo: 10265.285554/2021-52

2. Descrição da necessidade

2.1 Um componente crítico para o cumprimento dos objetivos estratégicos da RFB é a habilidade de obter conhecimento a partir de toda a informação disponível.

2.2 Nos últimos anos, o volume de dados disponível aumentou substancialmente principalmente devido ao grande volume de documentos e escriturações do Sped (Sistema Público de Escrituração Digital) recebidos.

2.3 Para atender às necessidades da RFB de extrair conhecimento desse grande volume de dados, e melhorar sua capacidade analítica foi contratada junto a um prestador de serviço de TI a solução Receita Data (RD), um datalake que utiliza tecnologia de Big Data.

2.4 O RD permite aos usuários, por meio de um conjunto de ferramentas analíticas, gerar relatórios a partir de cruzamentos de dados complexos, gerar painéis gráficos com dados agregados de fácil visualização e aplicar técnicas de inteligência artificial utilizando as linguagens R e Python. Além disso, o contrato possibilita que uma aplicação desenvolvida pela RFB, chamada ContÁgil, possa acessar os dados do RD e realizar o processamento de grandes volumes de dados.

2.5 O RD vem trazendo inúmeros benefícios para o processo de análise dos dados na RFB, contudo, existem algumas deficiências e problemas no RD que precisam ser apresentados:

- Valores dispendidos são significativos, necessidade de buscar maior eficiência com mais serviços e menor dispêndio de recursos.
- O espaço contratado de 1.43PB já está próximo de ser atingido;
- Somente o ContÁgil tem acesso aos dados armazenados no RD. Necessidade de nova proposta comercial para possibilitar que qualquer outro software desenvolvido pela RFB ou adquirido pela RFB tenha acesso aos dados do RD;
- Ausência de uma ferramenta gráfica de preparação e de mineração de dados que dispense o conhecimento de uma linguagem de programação pelo usuário;
- Diversos recursos do RD são oferecidos com grandes limitações;
- O RD costuma apresentar instabilidades que podem durar algumas horas ou até dias, prejudicando a utilização do ambiente durante estes períodos.
- O armazenamento de dados que não são atualizados tem o mesmo custo de dados que estão constantemente sendo atualizados.

2.6 Pelas razões expostas acima, o ambiente analítico da RFB não está alcançando todo o potencial que poderia. Existe a necessidade de buscar alternativas que mitiguem os problemas levantados e apresentem melhores custos para a RFB.

2.7 Além disso, a RFB nos últimos anos tem desenvolvido sistemas internamente e utilizado um data center da RFB ou contratado junto aos prestadores de serviço de TI a infraestrutura necessária para hospedar a produção destes sistemas.

2.8 Os valores despendidos com o serviço de hospedagem são significativos e existe a necessidade de buscar maior eficiência com maiores serviços e menor dispêndio de recursos.

3. Área requisitante

Área Requisitante	Responsável
Coordenação-Geral de Tecnologia e Segurança da Informação	Juliano Brito da Justa Neves

4. Necessidades de Negócio

4.1 As seguintes necessidades de negócio devem ser atendidas:

- Redução de custos.
- Aumentar o espaço disponível para armazenar dados de interesse econômico fiscal (de 1.43 PB para 2PB).
- Melhorar a performance e a disponibilidade do ambiente analítico.
- Acessar os dados a partir de aplicativos desenvolvidos na RFB ou softwares livres ou comerciais homologados pela RFB.
- Garantir a autonomia da RFB sobre os seus dados, sem as limitações de acesso impostas pelo prestador de serviço.
- Reduzir o intervalo de tempo entre a disponibilização de novidades tecnológicas pelo mercado e a sua efetiva utilização pela RFB, principalmente aquelas relacionadas a inteligência artificial.

4.2 Importante destacar que os dados são propriedade da RFB, desta forma, havendo qualquer condição que implique no encerramento do contrato, a RFB poderá transferir os dados para outro provedor de nuvem ou mantê-los no mesmo provedor e utilizar outro cloud broker.

5. Necessidades Tecnológicas

5.1 As seguintes necessidades tecnológicas devem ser atendidas:

- Possibilitar uma maior agilidade em aumentar ou reduzir a capacidade de processamento de acordo com as demandas de negócio, reduzindo a ociosidade do ambiente contratado e conseqüentemente dos custos associados;
- Hospedar a produção dos sistemas desenvolvidos pela RFB em infraestrutura capaz de prover alta capacidade de processamento paralelo, especialmente para treinamento supervisionado e não supervisionado de modelos preditivos.

6. Demais requisitos necessários e suficientes à escolha da solução de TIC

6.1 O prazo de vigência da contratação será de 12 (doze) meses, a contar da data de sua assinatura, prorrogável por iguais e sucessivos períodos, produzindo efeitos a partir da data da publicação do contrato. Justifica-se a possibilidade de contratação como serviço continuado pela necessidade permanente do órgão na utilização de um ambiente analítico e dos sistemas desenvolvidos internamente e em operação.

6.2 Os serviços prestados pelo Contratado deverão pautar-se sempre no uso racional de recursos e equipamentos, de forma a evitar e prevenir o desperdício de insumos e materiais consumidos, bem como a geração excessiva de resíduos, a fim de atender às diretrizes de responsabilidade ambiental adotadas pelo Contratante.

6.3 O Contratado deverá instruir os seus empregados quanto à necessidade de racionalização de recursos no desempenho de suas atribuições, bem como das diretrizes de responsabilidade ambiental adotadas pelo Contratante. A contratada deverá, ainda, observar, no que couber, durante a execução contratual, os critérios de sustentabilidade ambiental, previstos na Instrução Normativa nº 1, de 19 de janeiro de 2010, da Secretaria de Logística e Tecnologia da Informação, do então Ministério do Planejamento, Orçamento e Gestão.

7. Estimativa da demanda - quantidade de bens e serviços

7.1 A quantidade de bens e serviços a ser contratada dependerá diretamente da tecnologia da solução escolhida. Em termos gerais, esta quantidade deverá ser suficiente para oferecer uma performance igual ou superior a performance do Receita Data, com o aumento do volume de dados armazenado de 1.43PB para 2PB.

7.2 Além disso, deverá ser possível hospedar a produção dos sistemas desenvolvidos pela RFB em infraestrutura capaz de prover alta capacidade de processamento paralelo, especialmente para treinamento supervisionado e não supervisionado de modelos preditivos conforme necessidade dos sistemas relacionados a seguir:

--	--	--	--	--	--	--	--	--	--

	VMs	vCPUs	GPUs	RAM GB	Disco HD	Disco complementar de alta performance
i-Processo	5	40	2	160	20 TB	2TB com 6000 IOPS + 3TB com base de 6000 IOPS e 25000 IOPS por 1/5 do tempo.
FAPE	1	4	0	16	1 TB	1 TB com 6000 IOPS
ChatBots	11	124	5	104	20 TB	1TB com 6000 IOPS + 2TB com 12000 IOPS
VIVII	2	32	0	64	5 TB	1 TB com 6000 IOPS
ANITA/sisam/Ágata	2	36	0	128	4 TB	2 TB com 6000 IOPS
AJNA/Laudos/API Recintos	6	36	5	360	10 TB	2 TB com 6000 IOPS
SGBD MySQL	2	36	0	128	6 TB	6 TB com 12000 IOPS
Núcleo do SISAM	7	448	0	1.792	9 TB	7 TB (12000 IOPS durante horário comercial e 6000 nos demais horários) + 2 TB com 25000 IOPS por 1/10 do tempo)
Totais	36	756	12	2752	75	29

7.3 Por fim, haverá necessidade de prestação dos seguintes serviços:

- Serviço de Sustentação com capacidade de 400 chamadas por ano, 24hx7 com prazo máximo de atendimento de 2 horas.
- 7.3 Por fim, haverá necessidade de prestação dos seguintes serviços:
- Serviço de consultoria com especialista de até 500 horas por ano.

8. Levantamento de soluções

8.1 Após a análise das necessidades tecnológicas e de negócio, e considerando as diferentes abordagens em contratações realizadas pela RFB e pelos órgãos federais além das diversas alternativas oferecidas pelo mercado é possível identificar três possíveis soluções:

Número	Solução	Descrição
1	Hospedagem em datacenter próprio	Aquisição de infraestrutura necessária para a disponibilização de um Data Center Tear IV, ou seja, capaz de tolerar falhas de equipamento individual e interrupção de energia sem afetar as operações do site. Esta solução inclui a aquisição uma sala cofre, de nobreaks, e ar condicionados redundantes, servidores com alta capacidade de processamento e virtualização, unidades de armazenamento e backup, firewalls e a contratação de circuitos dedicados redundantes de alta capacidade.
2	Renovação do contrato com o Serpro	Renovação dos serviços analíticos e de hospedagem contratados com o Serpro com aumento da capacidade de armazenamento de 1.43PB para 2PB e inclusão das hospedagens para os sistemas relacionados no item 7.2.
3	Contratação de um Cloud Broker	Contratação de um Cloud Broker e utilização dos serviços do tipo PaaS para suportar as necessidades analíticas e dos serviços IaaS para suportar a hospedagem dos sistemas desenvolvidos pela RFB ou softwares homologados pela RFB

8.2 Solução 1 - Hospedagem em data center próprio

8.2.1 Segundo o Gartner, com os métodos de computação mais distribuídos, a exemplo do conceito de Edge-Computing, os data center ainda são comuns, mas estão se tornando menos comuns. O termo continua a ser usado para se referir à área que é responsável por esses sistemas, não importa quão dispersos eles sejam. As tendências do mercado e do setor estão mudando a maneira como as empresas abordam suas estratégias de data center. Vários fatores estão levando as empresas a olhar e rever as estruturas tradicionais de infraestrutura de tecnologia com vistas a racionalizá-las e otimizá-las.

8.2.2 A aquisição dos seguintes equipamentos é necessária para a implantação da solução 1:

- Sala cofre certificados de acordo com a NBR 15247.
- Ar condicionado redundante.
- Nobreaks
- Servidores com alta capacidade de processamento e virtualização
- Unidades de armazenamento e backup
- Firewalls

8.2.3 Importante destacar que haverá necessidade de alocar mão de obra especializada para operar estes equipamentos e administrar, configurar e gerenciar os softwares instalados nos servidores adquiridos.

8.3 Solução 2 - Renovação do contrato com o Serpro

8.3.1 Esta solução compreende a renovação dos serviços analíticos e de hospedagem contratados com o Serpro com aumento da capacidade de armazenamento de 1.43PB para 2PB e inclusão das hospedagens para os sistemas relacionados no item 7.2.

8.3.2 A solução atualmente contratada vem trazendo inúmeros benefícios para o processo de análise dos dados na RFB, contudo, existem algumas deficiências que não seriam resolvidos com o aumento da capacidade de armazenamento e inclusão de novas hospedagens:

- Alto custo, tanto para sua manutenção quanto para a inclusão de novas bases;
- Somente o ContÁgil tem acesso aos dados armazenados no RD. Necessidade de nova proposta comercial para possibilitar que qualquer outro software desenvolvido pela RFB ou adquirido pela RFB tenha acesso aos dados do RD;
- Diversos recursos do RD são oferecidos com grandes limitações;
- O RD costuma apresentar instabilidades que podem durar algumas horas ou até dias, prejudicando a utilização do ambiente durante estes períodos.
- O armazenamento de dados que não são atualizados tem o mesmo custo de dados que estão constantemente sendo atualizados.
- Demora excessiva entre a disponibilização de novidades tecnológicas pelo mercado e a sua efetiva utilização pela RFB, principalmente aquelas relacionadas a inteligência artificial.

8.4 Solução 3 - Contratação de um Cloud-Broker

8.4.1 A prestação de serviços em nuvem segue o modelo arquitetural de referência proposto pelo NIST (National Institute of Standards and Technology) e citado na ISO 17.799:2005, conforme figura a seguir.

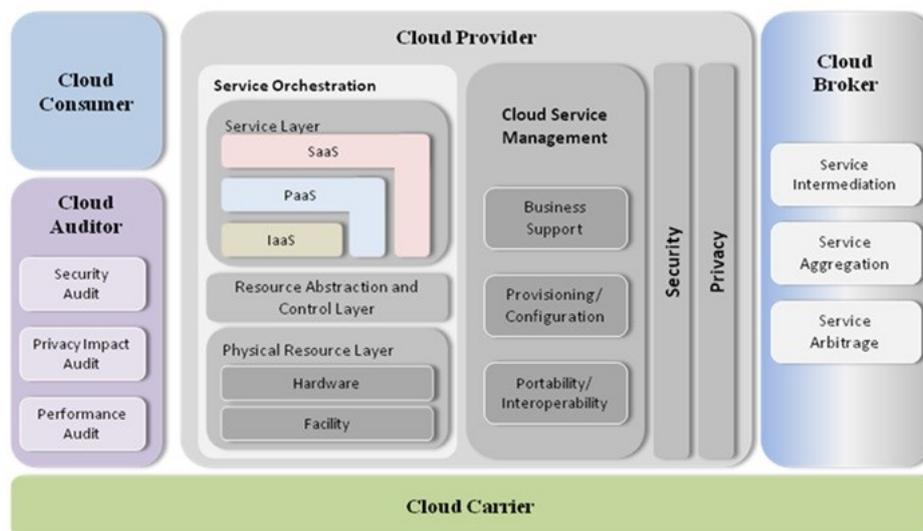


Figura 1 - Arquitetura de Referência para Computação em Nuvem

Fonte: The Cloud Computing Conceptual Reference Model - NIST

8.4.2 Conforme pode ser observado na figura 1, o modelo de prestação de serviços requer a contratação de um *broker* (integrador) que auxiliará na prestação dos serviços providos pelo provedor de nuvem. Segundo estudo realizado pela organização Open Data Center Alliance (ODCA), Usage Model: Cloud Service Brokering Ver. 1.0 (2014), à medida que a computação em nuvem se torna um aspecto cada vez mais importante das operações de TI corporativas, as complexidades de obter segurança, eficiência e os serviços de nuvem econômicos deram origem a uma nova entidade: o *Broker* de serviços em nuvem.

8.4.3 Esses serviços intermediários, posicionados entre o consumidor de nuvem e um ou mais provedores de nuvem, podem ajudar a RFB a alcançar seus objetivos de computação principalmente considerando a pouca maturidade que a RFB possui no gerenciamento e na configuração de uma nuvem.

8.4.4 Os serviços oferecidos por um Cloud Broker podem auxiliar a RFB a implantar as políticas de segurança, a implantar serviços em que a RFB tenha pouco ou nenhum conhecimento e a sanar eventuais problemas encontrados durante a operação do ambiente analítico.

8.4.5. Segundo estudo realizado pela consultoria Grand View Research, especializada no setor de Cloud Service Brokerage, o mercado global de serviço de Cloud Brokerage foi avaliado em US \$ 5.9 bilhões de dólares em 2020 e espera-se um crescimento de 16,8% no período de 2020 a 2025 (fonte: https://www.researchandmarkets.com/reports/5238097/cloud-services-brokerage-market-by-service-type?w=4&utm_source=BW&utm_medium=PressRelease&utm_code=svbbc8).

8.4.6. Importante destacar que outros órgãos públicos estão adotando soluções similares a esta. Neste sentido a SGD realizou em fevereiro de 2021 uma licitação centralizada de serviços de computação em nuvem. Segundo o TR desta licitação (processo nº 19973.100103/2020-51) a categoria nuvem apurada nos diferentes Planos Anuais de Contratação (PAC) dos órgãos da administração pública é a décima quarta maior iniciativa em termo de volume financeiro previsto para 2020, alcançando a ordem de R\$54 milhões, distribuídas em 64 ações.

8.4.7 Cloud Brokers

8.4.7.2 Inúmeras empresas oferecem serviços de Cloud Broker no mercado brasileiro, dentre elas se destacam duas empresas públicas: o Serviço Federal de Processamento de Dados (Serpro, uma empresa do Governo Federal com sede em Brasília-DF e a Prodesp, uma empresa do Governo do Estado de São Paulo com sede em Taboão da Serra - SP.

8.4.7.3 O Serpro provê recursos de computação em nuvem, sob demanda, incluindo serviços de avaliação de ambientes, definição de arquiteturas, gestão de topologias, manutenção e suporte técnico, e disponibilização continuada de recursos de infraestrutura. Os serviços são suportados por ambientes e infraestrutura de nuvem compostos por recursos próprios do SERPRO e por recursos de provedores parceiros, para os quais o SERPRO atua como integrador de serviços (cloud broker).

8.4.7.4 A Prodesp provê duas modalidades de serviços em nuvem pública: Serviço Gerenciado e Serviço Não Gerenciado. Em ambas as modalidades, a Prodesp agrega serviços como disponibilizar um link dedicado entre a RFB e o provedor de nuvem contratado e prover uma central de atendimento que possibilite que os usuários da RFB possam abrir chamados em casos de problemas e esclarecimento de dúvidas.

8.4.7.4 A diferença entre o serviço gerenciado e não gerenciado é que no primeiro a Prodesp faz o gerenciamento do uso da nuvem, inclusive recomendando alterações, análises de comportamentos atípicos dentre outros serviços de acompanhamento do uso da nuvem e no segundo a Prodesp intermedia a contratação entre o cliente e o provedor de nuvem e assim toda a gestão e gerenciamento do uso da nuvem é feita pelo cliente por meio das ferramentas que as próprias nuvens disponibilizam.

8.4.8 Provedores de Nuvem

8.4.8.1 O Decreto 10332/2020 estabeleceu, dentre as estratégias do governo digital, adotar tecnologia de processos e serviços governamentais em nuvem como parte da estrutura tecnológica dos serviços e setores da administração pública federal.

8.4.8.2 O quadrante mágico do Gartner para infraestrutura na nuvem como serviço possibilita a identificação de três empresas que mais se destacam nos quesitos “habilidade em execução” e “completude da visão”.



Figura 2 - Quadrante Mágico Gartner para Computação em Nuvem

8.4.8.3 Como pode ser observado na figura 2, as empresas que se destacam são:

- Amazon Web Services;
- Google;
- Microsoft.

8.4.8.4 Cabe destacar que a Amazon Web Services e a Microsoft possuem parceria com o Serpro e Prodesp enquanto que a Google possui parceria somente com a Prodesp.

8.4.8.5 Em dezembro de 2020 por intermédio da NT COTEC 58/2020 foi instituída uma prova de conceito (POC) com o objetivo de medir a performance de serviços ofertados pelas nuvens como IaaS (Infrastructure as a Service) e PaaS (Plataform as a Service) comparando-os com o Receita Data.

8.4.8.6 Em julho de 2021, por intermédio da NT Cotec 38/2021 foi instituída nova POC com objetivo de complementar os testes feitos na POC anterior, especialmente em relação a nova solução tecnológica lançada pela Microsoft em 2021.8.4.6.6 Os testes realizados das duas POC possibilitaram:

- Medir o tempo de execução de programa que simula o processo de transformação dos dados e geração de tabelas.
- Realizar consultas analíticas por meio de comandos SQL sobre as tabelas geradas, medindo o desempenho (duração da consulta) em diversas condições.
- Realizar o processamento de muitas expressões regulares sobre um fluxo contínuo de conteúdo XML, gerando um fluxo de saída com as indicações de quais expressões foram satisfeitas sobre quais documentos.
- Executar testes sobre performance de CPUs, GPUs, TPUs, memória e armazenamento disponíveis aplicando implementações próprias e de terceiros.
- Executar testes sobre bancos chave/valor, comparando os resultados de uma implementação própria (BTree implementada em Java) com as soluções ofertadas pelos provedores.
- Executar testes de segurança com o objetivo de avaliar os riscos do armazenamento de dados protegidos pelo sigilo fiscal em uma nuvem.

8.4.8.7 Os resultados dos testes demonstram que:

- As nuvens oferecem diversas opções de CPU, GPUs, SSD, HD e memória possibilitando a contratação gradual, podendo atender a vários sistemas que se encontram com demandas reprimidas para ambientes de produção com um custo inferior ao cobrado pelo Serpro nos serviços de Hosting.
- Os testes de acesso chave/valor, desempenho de CPU, memória e disco demonstraram que as nuvens são mais rápidas que o serviço hosting oferecido pelo Serpro.
- Os testes com GPU demonstraram que as nuvens oferecem de 1 a 8 GPUs por máquina com ótima performance e possibilidade de adições e reduções de acordo com nossa necessidade. Isto atende à crescente demanda da RFB por este tipo de processamento sem a necessidade de uma contratação inicial superdimensionada.

- Foram geradas três tabelas (a maior delas possuindo algo na ordem de grandeza de um trilhão de linhas e 600 colunas) que simulam o comportamento das tabelas NFe, Dimensão PJ e Exclusão. Estas tabelas ocuparam cerca de 65TB e foram geradas com a utilização de um programa que preenche seus os campos com conteúdo aleatório, mantendo a cardinalidade e a estrutura das tabelas originais. O tempo de execução deste programa nas três nuvens foi equivalente ao tempo de execução no Receita Data quando a mesma quantidade de CPU e Memória são empregadas.
- Foi criada uma aplicação SPARK responsável pela produção de um fluxo contínuo de 300 documentos XML por segundo e outra responsável pela aplicação de 1000 expressões regulares sobre cada conteúdo XML. Ficou demonstrado que todos as nuvens conseguem oferecer uma infraestrutura capaz de suportar este processamento.
- Foram feitas diversas considerações sobre os riscos de utilização de uma solução PaaS ou IaaS na nuvem e foram apresentadas algumas soluções que poderiam mitigar em parte estes riscos como criptografia homomórfica e o uso de enclaves, ficando claro que as discussões sobre segurança precisam ser aprofundadas antes da decisão sobre a contratação de PaaS ou IaaS na nuvem.
- Os testes de performance de SQL demonstraram que as consultas SQL executadas na nuvem igualam ou superam a performance das mesmas consultas no Receita Data.

8.4.8.8 Os relatórios com o detalhamento dos trabalhos e dos resultados alcançados nas POC constam das NT Cotec 7/2021 e da NT Cotec 54/2021 e foram anexadas no processo administrativo de origem.

8.4.8.9 Os riscos de utilização de uma solução PaaS ou IaaS na nuvem, identificados na POC, foram avaliados na NT COTEC 035 /2021 seguindo o Modelo de Gestão de Riscos da RFB (probabilidade, impacto, controles existentes, controles propostos e risco residual) com a classificação "Média". A referida NT concluiu que "se a RFB entende ser possível a contratação do Serpro para hospedagem dos seus dados, significa que aceita o nível de risco onde é possível a contratação de um prestador de serviços de computação em nuvem".

8.4.8.10 Por fim, ficou demonstrado que a Google, Microsoft e Amazon alcançaram as metas definidas, igualando ou superando a performance do Receita Data e dos serviços de hosting contratados com o Serpro. Demonstrando inequivocamente a viabilidade técnica de substituir o Receita Data por soluções em nuvem.

8.4.8.11 Cabe destacar que os provedores de nuvem possuem tecnologias proprietárias para os serviços PaaS e que eles conseguem atender aos requisitos apresentados pela RFB de maneira distinta, o que torna praticamente impossível definir um item padrão que consiga representar toda a complexidade das necessidades da RFB. Um exemplo desta diferença é que enquanto uns oferecem bancos de dados do tipo Warehouse com unidade de capacidade de processamento DBU, outros oferecem clusters de processamento distribuído, onde é possível definir o tipo de servidor e a quantidade. A unidade DBU também não segue um padrão específico, apresentando diferentes capacidades de processamento conforme o provedor.

8.4.8.12 O dimensionamento dos serviços PaaS de cada um dos provedores somente foi possível após a realização das Provas de Conceito. A contratação direta de empresa pública de TI possibilitaria a utilização dos dimensionamentos já encontrados e evitaria a necessidade de definir um item padrão que possa ser licitado.

9. Análise comparativa de soluções

9.1 Diante da identificação de diferentes modelos de negócio adotados por órgãos públicos e pelo mercado apresentadas na seção anterior, nesta seção será apresentada uma análise comparativa das soluções consideradas viáveis.

9.2 A solução 1 foi considerada inviável, conforme justificativa apresentada na seção 10.

9.2 A tabela a seguir apresenta um comparativo das soluções 2 e 3.

Necessidade	Solução 2 - Renovação do contrato com o Serpro	Solução 3 - Cloud Broker
Performance adequada	Sim	Sim
Possibilidade de alterar a capacidade de processamento conforme a demanda por serviço, reduzindo ociosidade do ambiente e consequentemente reduzindo os custos	Não. Infraestrutura dimensionada para suportar a maior demanda.	Sim, grande parte das soluções possuem escalabilidade automática. Demais soluções podem ser dimensionadas manualmente ou por API.
Flexibilidade para incluir novos serviços ou alterar a capacidade de processamento alocada.	Não. Necessário alteração contratual	Sim.

Nível de serviço	90% das 7 às 22 horas. Menos de 72% glosa de 6% do custo do serviço contratado.	99.9%, 24h. Menos de 95% glosa de 50% (Google) ou 100% (AWS e Azure) do custo do serviço contratado.
Acesso aos dados pelos sistemas desenvolvidos pela RFB	Somente o ContÁgil.	Qualquer sistema autorizado pela RFB.
Utilização de ferramenta de construção de painéis e visualização de dados.	Somente MicroStrategy em algumas bases de dados previamente contratadas.	Qualquer ferramenta autorizada pela RFB.
Acesso aos dados através de linguagens de programação como Java, Python e R.	Somente Jupyter Notebook e PIG disponibilizados e gerenciados pelo Serpro.	Qualquer linguagem de programação.
Utilização de banco de dados de baixa latência	Somente HBASE, somente o Serpro tem a administração do banco de dados.	Diversas opções de banco de dados, controle total da RFB.
Utilização de serviços de indexação	Somente o SOLR, a RFB não tem autorização para criar ou gerenciar novos índices.	Diversos serviços de indexação, controle total da RFB.
Utilização de serviço de fluxos contínuos de dados.	Somente o Kafka, a RFB não tem autorização para criar novas filas.	Diversos serviços, controle total da RFB.
Armazenamento de dados frios, ou seja, pouco ou nunca atualizados e pouco utilizado.	Somente dados quentes.	Diversos serviços de armazenamento com custos variando conforme a utilização e atualização dos dados.
Novas tecnologias e recursos.	Evolução lenta, apesar de ser um software livre, existe a necessidade do integrador Cloudera lançar uma nova versão.	Os provedores de nuvem estão constantemente lançando novos produtos.
Melhoria da infraestrutura tecnológica (servidores, firewalls, unidades de armazenamento,...)	Última renovação da infraestrutura do Receita Data ocorreu em março de 2020.	Parque tecnológico está sempre atualizado.

10. Registro de soluções consideradas inviáveis

10.1 A Instrução Normativa SGD/ME nº 01/2019 que versa sobre as contratações públicas de TIC do Governo Federal introduziu dispositivos que tornam a adoção do modelo baseado em nuvem como prioritário em relação ao investimento em infraestrutura própria, conforme pode-se verificar no item 4.1 do anexo desta IN.:

“4.1. Os órgãos e entidades que necessitem criar, ampliar ou renovar infraestrutura de centro de dados deverão fazê-lo por meio da contratação de serviços de computação em nuvem, salvo quando demonstrada a inviabilidade em estudo técnico preliminar da contratação.” IN. nº 01/2019 SGD/ME.

10.2 Considerando que a SGD já estabeleceu orientações de que a contratação de nuvem é mais vantajosa para a administração pública federal do que a aquisição de infraestrutura, a solução 1 será considerada inviável.

11. Análise comparativa de custos (TCO)

11.1 A presente seção registra comparação de Custos Totais de Propriedade (TCO) para as soluções técnica e funcionalmente viáveis, nos termos do inciso III do art. 11. da IN 01.2019 SGD/ME.

11.2 Solução 1 - Hospedagem em datacenter próprio

11.2.1 Esta solução não atendeu aos requisitos estabelecidos na IN SGD/MF nº1/2019, transcrito abaixo e foi considerada inviável.

“4.1. Os órgãos e entidades que necessitem criar, ampliar ou renovar infraestrutura de centro de dados deverão fazê-lo por meio da contratação de serviços de computação em nuvem, salvo quando demonstrada a inviabilidade em estudo técnico preliminar da contratação.” IN. nº 01/2019 SGD/ME.

11.3 Solução 2 - Renovação do contrato com o Serpro com aumento da capacidade de armazenamento de 1.43PB para 2PB e hosting do sistema SISAM e de 6 outros sistemas.

11.3.1 Primeiramente cabe realizar um levantamento dos custos atuais gastos no contrato como Serpro com o ambiente analítico e com a hospedagem de sistemas. Qualquer renovação de contrato com aumento de serviços representará um aumento do custo atual.

11.3.2 Os custos dos serviços do ambiente analítico por mês podem ser observados na tabela a seguir:

Código Contrato	Descrição do Serviço	Valor Contratado
10680	REPOSITÓRIO NACIONAL DE DADOS DO SIMPLES NACIONAL	R\$ 437.805,55
10685	API BACK-END -CONTÁGIL	R\$ 370.015,84
10695	FONTE COFIS - CTE	R\$ 168.882,35
10698	FONTE COANA-CONT	R\$ 132.742,31
10699	FONTE COPEES - ECF	R\$ 11.386,25
10700	FONTE COFIS - ICMS	R\$ 261.701,98
10701	FONTE COPEES - FIN	R\$ 1.042.966,12
10702	FONTE COPEES - FAC	R\$ 1.053,32
10703	FONTE COFIS - NFE	R\$ 1.810.201,65
10741	AMBIENTE INTEGRADO PROGRAMAÇÃO	R\$ 455.760,43
11325	Datalake RFB - Fonte Coana - DUE	R\$ 6.929,10
11461	Histórico DIRPF	R\$ 4.990,19
11592	Receita Data - Fonte Codac e-Social	R\$ 39.087,09
11706	DLK - Fonte IRPF	R\$ 23.703,41
11718	RECEITA DATA - Fonte RECEITANET	R\$ 4.470,24
11739	Receita Data - DUE - Fase 2	R\$ 12.036,97
11823	Receita Data - Fonte Cofis - Ação Fiscal	R\$ 3.742,64
11850	Análise de Dados via Receita Data Copes, fonte DOI	R\$ 13.573,96
11851	Análise de Dados via Receita Data Copes, fonte Dimob	R\$ 13.606,42
11852	Análise de Dados via Receita Data Copes, fonte DBF	R\$ 2.687,65
11853	Análise de Dados via Receita Data Copes, fonte DCP	R\$ 2.656,43
11854	Análise de Dados via Receita Data Copes, fonte Decred	R\$ 69.340,89
11855	Análise de Dados via Receita Data Copes, fonte Derc	R\$ 2.621,21
11892	Receita Data - Fonte COMAC - Webanálise	R\$ 12.475,48
11932	RECEITA DATA - Tema Central de Gestão	R\$ 236.303,34
11972	Receita Data - Fonte Cocad - CPF	R\$ 22.237,67
11976	Análise de Dados via Receita Data Cotec, Tabelas Institucionais	R\$ 444,64
12019	Receita Data - Chat RFB	R\$ 10.038,65
12099	RECEITA DATA - FONTE COANA - eDBV	R\$ 26.538,25
12121	RECEITA DATA - FONTE COANA - REMESSA	R\$ 1.786,00
12159	Suara - Auxílio	R\$ 3.948,51
12179	Cocad - CNIR	R\$ 13.353,64
12215	CORAT – DCTF Web	R\$ 41.052,97
	TOTAL:	R\$ 5.260.141,15

11.3.3 Importante ressaltar que os custos da solução 2, na maior parte dos casos, podem ser divididos em duas parcelas:

- Custo do processo de extração, transformação e carga (ETL).
- Custo da infraestrutura de processamento analítico, que representa os custos associados a manutenção e operação de um conjunto de servidores capazes de processar as consultas SQL executadas pelos usuários nos dados armazenados.

11.3.4 Caso a solução 1 ou 3 seja contratada, a parcela dos custos, que representa o processo de ETL, continuará sendo executado pelo Serpro de maneira similar a atual. Desta forma, é importante mensurar a parcela do custo relativa ao ETL de modo a excluir este custo da análise comparativa de custos (TCO) ora em análise.

11.3.5 O Serpro não detalha a formação de seus custos, contudo, é possível arbitrar o percentual dos custos de cada um dos serviços que representam o custo da plataforma de solução analítica (PSA), conforme a complexidade do processo de ETL.

11.3.6 Os serviços a seguir possuem processos de ETL com complexidade baixa que copiam arquivos XML ou TXT de um ambiente para outro, sem transformação. Apesar de existirem alguns dados de controle que também são copiados de outra base e depois armazenado juntamente, isso não prejudica a simplicidade do processo.

- FONTE COFIS - CTE
- FONTE COPEs - ECF
- FONTE COPEs - FIN
- FONTE COPEs - FAC
- FONTE COFIS - NFE

11.3.7 Os serviços a seguir possuem processos de ETL com complexidade baixa que copiam integralmente bases de dados armazenadas em banco de dados Oracle, Postgres ou Adabas para o Receita Data sem qualquer transformação:

- DLK - Fonte IRPF
- RECEITA DATA - Fonte RECEITANET
- Receita Data - Fonte Cofis - Ação Fiscal
- Análise de Dados via Receita Data Copes, fonte DOI
- Análise de Dados via Receita Data Copes, fonte Dimob
- Análise de Dados via Receita Data Copes, fonte DBF
- Análise de Dados via Receita Data Copes, fonte DCP
- Análise de Dados via Receita Data Copes, fonte Decred
- Análise de Dados via Receita Data Copes, fonte Derc
- RECEITA DATA - Tema Central de Gestão
- Receita Data - Fonte Cocad - CPF
- Cocad - CNIR
- CORAT – DCTF Web

11.3.8 Os serviços a seguir possuem processos de ETL com complexidade média, pois realizam agregações simples, mantendo, na maioria dos casos, as mesmas informações das tabelas originais.

- FONTE COANA-CONT
- FONTE COFIS - ICMS
- Datalake RFB - Fonte Coana - DUE
- Histórico DIRPF
- Receita Data - Fonte Codac e-Social
- Receita Data - DUE - Fase 2
- Receita Data - Fonte COMAC - Webanálise
- Receita Data - Chat RFB
- RECEITA DATA - FONTE COANA - eDBV
- RECEITA DATA - FONTE COANA - REMESSA
- Suara - Auxílio

11.3.9 O serviço "REPOSITÓRIO NACIONAL DE DADOS DO SIMPLES NACIONAL" possui processo de ETL com complexidade alta, pois realiza muitas agregações complexas e organiza as informações de modo a otimizar consultas OLAP.

11.3.10 A tabela a seguir arbitra o custo do processo de ETL considerando as complexidades pequena, média e alta e aplicando respectivamente os percentuais de 10%, 20% e 30%.

Complexidade do ETL	Custo
Pequena	R\$ 348.428,48
Média	R\$ 102.454,91
Alta	R\$ 131.341,67

TOTAL	R\$ 582.225,05
-------	----------------

11.3.11 O custo da solução 2 demonstrados em 11.3.2 (R\$ 5,260.141,15) descontando os custos associados ao processo de ETL (R\$ 582.225,05) representa o custo da infraestrutura de processamento analítico e é de R\$ 4.677.916,10.

11.3.12 Considerando que o RD dispõe hoje de 1.43PB e que deseja-se contratar 2PB, aplica-se uma regra de três ao custo da infraestrutura de processamento para se obter o custo previsto deste infraestrutura para 2PB, correspondendo a R\$6.542.540,00.

11.3.13 A memória de cálculo a seguir demonstra os cálculos feitos nos itens

(i)Custo do RD	R\$	5.260.141,15
(ii) Parcela ETL	R\$	582.225,05
(iii) Parcela Infraestrutura de Processamento Analítico	R\$	4.677.916,10
(iv) Custo processamento por PB -> (iii)/1,43PB	R\$	3.271.270,00
(v) Custo processamento para 2 PB -> (iv)*2PB	R\$	6.542.540,00

11.3.14 Ao custo do ambiente analítico (v), deve-se somar o custo do serviço de hosting do sistema SISAM atualmente contratado com o Serpro (código no contrato 07274) e que está abrangido no requisito deste ETP. O custo de hospedagem é de R\$ 621.982,09 por mês, resultando em um custo estimado total de R\$ 7.164.522,09.

11.3.15 Importante destacar que a hospedagem adicional de mais 7 sistemas não foi considerada na apuração do custo da solução 2, e que caso fosse, resultaria em um custo superior ao custo calculado.

11.4 Solução 3 - Contratação de um Cloud-Broker

11.4.1 A estimativa de preços da contratação foi elaborada mediante consulta aos custos dos serviços disponibilizados pelas provedoras nos seguintes sites:

PROVEDORA	CALCULADORA DE PREÇOS
Amazon AWS	https://aws.amazon.com/pt/pricing/
Google GCP	https://cloud.google.com/products/calculator?hl=pt-br
Microsoft Azure	https://azure.microsoft.com/pt-br/pricing/calculator/

11.4.2 Os custos foram contabilizados considerando as especificidades de cada um dos serviços oferecidos pelos provedores em prova de conceito realizada no quarto trimestre de 2020 e constante do relatório final, objeto da NT Cotec 7/2021.

11.4.3 Estimativa de custo da Google:

Categoria	Item	Quantidade	Unidade	Preço
Big Data	BigQuery Slots	24000	Slots	\$609.552,00
Big Data	Active Storage (Uncompressed)	4543829,333	GB	\$98.055,84
Big Data	Long-term Storage (Uncompressed)	4543829,333	GB	\$98.055,84
Cluster Spark	Spark vCPU	2000	vCPU/Hour	\$38.550,99

Cluster Spark	Spark Memory	4096	GB/Hour	\$7.254,20
Cluster Spark	Spark Local Storage Standard	10000	GB/Hour	\$498,00
Cluster MapReduce	Dataproc vCPU (120h)	2000	vCPU/Hour	\$6.337,15
Cluster MapReduce	Dataproc Memory (120h)	16384	GB/Hour	\$4.769,89
Cluster MapReduce	Dataproc Local Storage Standard (120h)	10000	GB/Hour	\$81,86
Cluster MapReduce	Dataproc Object Storage	5120	GB/Month	\$148,74
Cluster MapReduce	Classe A Operations - Storage	1000	per 10000 operations	\$41,50
Cluster MapReduce	Classe B Operations - Storage	1000	per 10000 operations	\$3,32
IaaS	IaaS vCPU	984	vCPU/Hour	\$18.967,09
IaaS	IaaS Memory	3648	GB/Hour	\$6.460,77
IaaS	IaaS Local Storage Standard	103424	GB/Hour	\$5.150,52
IaaS	IaaS Local Storage SSD	20480	GPU/Hour	\$1.291,88
IaaS	GPU T4	18	GPU/Hour	\$3.271,86

Tabela 4 – Estimativa de custos GCP – com BigQuery

11.4.3.1 Utilizando o dólar de R\$5,28 o valor estimado para a contratação da nuvem da Google é de R\$ 5.060.303,90.

11.4.4 Estimativa dos preços da Microsoft:

SERVIÇO	CUSTO
<p>Armazenamento – 1PB Quente</p> <p>Redundância de Data Lake Store Gen2, Padrão, LRS, Quente Camada de Acesso, Estrutura de arquivos Namespace hierárquico, Capacidade de 1.000 TB - PAGO CONFORME O USO, operações de gravação: 1 GB x 100.000 operações, 100.000 operações de listagem e criação de contêineres, operações de leitura: 4 MB x 100.000 operações, 100.000 operações de gravação iterativa, 100.000</p>	R\$ 128.858,67

outras operações. 1.000 GB Recuperação de Dados, 1.000 GB Gravação de Dados, 1.000 GB Armazenamento de metadados	
Armazenamento – 1PB Frio Redundância de Data Lake Store Gen2, Padrão, LRS, Fria Camada de Acesso, Estrutura de arquivos Namespace hierárquico, Capacidade de 1.000 TB - PAGO CONFORME O USO, operações de gravação: 1 GB x 100.000 operações, 100.000 operações de listagem e criação de contêineres, operações de leitura: 4 MB x 100.000 operações, 100.000 operações de gravação iterativa, 100.000 outras operações. 1.000 GB Recuperação de Dados, 1.000 GB Gravação de Dados	R\$ 88.162,21
DataBricks – Cluster SPARK Carga de trabalho Computação para Todas as Finalidades, camada Premium, 25 L80SV2 (80 vCPU (s), 640 GB de RAM) x 180 Horas, PAGO CONFORME O USO, 500 DBU x 180 Horas	R\$ 314.131,66
Cluster MapReduce Hadoop Componente: 2 E16V3 (16 núcleos, 128 GB RAM) de Cabeçalho nós x 120 Horas, 125 E16V3 (16 núcleos, 128 GB RAM) Região nós x 120 Horas, 0 E16V3 (16 núcleos, 128 GB RAM) Borda nós x 120 Horas	R\$ 80.961,91
VMs com GPU 5 NC24s v3 (24 vCPUs, 448 GB de RAM) x 730 Horas; Linux – Ubuntu; PAGO CONFORME O USO; 0 discos gerenciados do sistema operacional – P50; Tipo de transferência entre regiões, 5 GB transferência de dados de saída de Leste dos EUA para Ásia Oriental	R\$ 180.899,02
VMs 23 D32 v3 (32 vCPUs, 128 GB de RAM) x 730 Horas; Linux – Ubuntu; PAGO CONFORME O USO; 0 discos gerenciados do sistema operacional – E10, 100 unidades de transação; Tipo de transferência entre regiões, 5 GB transferência de dados de saída de Brazil South para Ásia Oriental	R\$172.954,46
Ultra Disk - 25.000 IOPS Discos gerenciados, Disco Ultra, 1 X 20 TiB Discos, 730 Horas, 25.000 IOPS, 2.000 Produtividade de MB/s	R\$ 46.103,95
Discos 1000 IOPS Discos gerenciados, SSD Premium, P50 Tipo de Disco 26 Discos, PAGO CONFORME O USO	R\$ 90.129,52
TOTAL	R\$ 1.102.201,39

11.4.4.1. O valor estimado para a contratação da nuvem Azure é de R\$1.102.201,39.

11.4.5 Estimativa dos preços da Amazon:

Descrição	Serviço	Resumo da Configuração	Preço (Dolar)

Big Data	Amazon Redshift	Nodes (32 instances of type ra3.16xlarge Reserved No Upfront 3yr), Data scanned (2000 TB)	\$221.159,80
S3 Big Data	S3 Standard	S3 Standard storage (1000 TB per month)	\$39.744,80
S3 Big Data	S3 Standard - Infrequent Access	S3 Infrequent Access storage (1000 TB per month)	\$26.634,00
Cluster Spark	Amazon EMR master node	Number of master EMR nodes (1), EC2 instance (c5.4xlarge), Utilization (100 %Utilized/Month)	\$124,10
Cluster Spark	Amazon EMR core node	Number of core EMR nodes (124), EC2 instance (c5.4xlarge), Utilization (100 %Utilized/Month)	\$15.388,40
Cluster Spark	Amazon EC2	Operating system (Linux), Quantity (125), Pricing strategy (EC2 Instance Savings Plans 3 Year No UpFront), Storage for each EC2 instance (General Purpose SSD (gp2)), Storage amount (10 GB), Instance type (c5.4xlarge)	\$41.526,25
Cluster Spark	S3 Standard	S3 Standard storage (5 TB per month)	\$214,92
Cluster MapReduce	Amazon EMR master node	Number of master EMR nodes (1), EC2 instance (r5.4xlarge), Utilization (4 Hours/Day)	\$30,66
Cluster MapReduce	Amazon EMR core node	Number of core EMR nodes (124), EC2 instance (r5.4xlarge), Utilization (4 Hours/Day)	\$3.801,84
Cluster MapReduce	Amazon EC2	Operating system (Linux), Storage for each EC2 instance (General Purpose SSD (gp2)), Storage amount (10 GB), Snapshot Frequency (No snapshot storage), DT Inbound: Not selected (0 TB per month), DT Outbound: Not selected (0 TB per month), DT Intra-Region: (0 TB per month), Workload (Daily, (Workload days: Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Baseline: 0, Peak: 125, Duration of peak: 4 Hr 0 Min)), Advance EC2 instance (r5.4xlarge), Pricing strategy (On-Demand)	\$24.486,67
Instâncias M5	Amazon EC2	Operating system (Linux), Quantity (28), Pricing strategy (Convertible Reserved Instances 3 Year No UpFront), Storage for each EC2 instance (General Purpose SSD (gp2)), Storage amount (50 GB), Instance type (m5a.8xlarge)	\$23.718,80
Storage Magnético	Amazon Elastic Block Store (EBS)	Number of instances (28), Average duration each instance runs (730 hours per month), Storage amount (3600 GB), Snapshot Frequency (Daily), Amount changed per snapshot (360 GB)	\$25.804,80
		Number of instances (28), Average duration each instance runs (730 hours per month), Storage amount (715 GB),	

Storage SSD	Amazon Elastic Block Store (EBS)	Snapshot Frequency (Daily), Amount changed per snapshot (72 GB), General Purpose SSD (gp3) - IOPS (1000), General Purpose SSD (gp3) - Throughput (250 MBps)	\$6.722,21
Instâncias GPU (8 placas)	Amazon EC2	Operating system (Linux), Quantity (2), Pricing strategy (Convertible Reserved Instances 3 Year No UpFront), Storage for each EC2 instance (General Purpose SSD (gp2)), Storage amount (100 GB), Instance type (g4dn.metal)	\$11.856,40
Instâncias GPU (1 placa)	Amazon EC2	Operating system (Linux), Quantity (2), Pricing strategy (Convertible Reserved Instances 3 Year No UpFront), Storage for each EC2 instance (General Purpose SSD (gp2)), Storage amount (100 GB), Instance type (g4dn.8xlarge)	\$3.319,78
		TOTAL	\$444.533,43

11.4.5.1 Utilizando o dólar de R\$5,28 o valor estimado para a contratação da nuvem da Amazon é de R\$ 2.347.136,51.

11.4.6. Diante do quadro resumo de preços, podemos concluir que o que o custo estimado mensal para fornecimento dos serviços utilizando a solução 3 (nuvem) varia entre R\$1.102.201,39 e 5.060.303,90.

Provedor de nuvem	Custo obtido pela Internet
Google GCP	R\$ 5.060.303,90
Microsoft Azure	R\$1.102.201,39
Amazon - AWS	R\$ 2.347.136,51

11.5 Comparação dos Custos

11.5.1 A tabela seguir resume os custos das soluções.

Número	Solução	Custo Mensal
1	Hospedagem em datacenter próprio	Solução inviável
2	Renovação do contrato com o Serpro com aumento da capacidade de 1.43 PB para 2PB e hospedagem do sistema SISAM (já contratado) e mais 7 sistemas	R\$ 7.164.522,09 + custo para a hospedagem adicional de mais 7 sistemas.
3	Contratação de um Cloud Broker	Entre R\$1.102.201,39 e 5.060.303,90.

11.5.2 Portanto, é possível concluir que a solução 3 (contratação de um Cloud Broker) apresenta um melhor custo que a solução 2, e representa uma economia superior a R\$ 2.104.218,19 por mês (considerando o maior custo possível para a solução 3, R\$5.060.303,90). Desta forma, a economia anual é superior a R\$ 25.250.618,28.

12. Descrição da solução de TIC a ser contratada

12.1 A tabela contendo o comparativo do atendimento de diversas necessidade constante do item 9, juntamente com a análise comparativa de custos (item 11), demonstram que a melhor escolha, tanto pelo aspecto financeiro como pelo aspecto funcional é a solução 3, ou seja, contratar um Cloud Broker.

12.3 A Prodesp e o Serpro oferecem através de seus parceiros dois tipos de serviço: Infraestrutura como Serviço (IaaS) para suportar a hospedagem de softwares homologados pela RFB e Plataforma como Serviço (PaaS) para suportar as necessidades analíticas.

12.4 De forma simplificada, é possível diferenciar os dois tipos de serviço da seguinte forma:

IaaS: Capacidade fornecida ao cliente para provisionar processamento, armazenamento, comunicação de rede e outros recursos de computação fundamentais nos quais o cliente pode instalar e executar softwares em geral, incluindo sistemas operacionais e aplicativos. O cliente não gerencia nem controla a infraestrutura na nuvem subjacente, mas tem controle sobre os sistemas operacionais, armazenamento e aplicativos instalados, e possivelmente um controle limitado de alguns componentes de rede.

PaaS: capacidade fornecida ao cliente para provisionar na infraestrutura de nuvem aplicações adquiridas ou criadas para o cliente, desenvolvidas com linguagens de programação, bibliotecas, serviços e ferramentas suportados pelo provedor de serviços em nuvem. O cliente não gerencia nem controla a infraestrutura na nuvem subjacente incluindo rede, servidores, sistema operacional ou armazenamento, mas tem controle sobre as aplicações instaladas e possivelmente sobre as configurações do ambiente de hospedagem de aplicações.

12.5 A solução a ser contratada deverá possuir serviços PaaS, serviços IaaS e serviços adicionais, conforme relação abaixo:

12.6 Plataforma como Serviço (PaaS):

12.6.1. Deverá ser previsto um serviço de armazenamento tabelas e dados brutos com capacidade total de armazenamento de 2 PB em dados quentes no formato Parquet. A capacidade hoje contratada no Receita Data é de 1.43PB, o aumento reflete a necessidade de armazenamento de novos dados, como o REINF, NFe de Serviço, NFe de consumidor final, dentre outras.

12.6.2. Deverá ser previsto um serviço de processamento de consultas a grandes volumes de dados, o serviço deverá ter performance superior a existente no Receita Data superando as metas abaixo:

Teste	COMANDO SQL	Meta (segundos)
1	Teste de volume: Mede a performance conforme o volume de dados aumenta. Faz uma agregação com cardinalidade de 280 elementos. Para cada um dos testes, executa 10 comandos em paralelo.	50 (1 ano) 206 (6 anos) 380 (11 anos)
2	Teste de junção: Compara a performance em três cenários diferentes: sem junção, com uma junção e com duas junções para o ano de 2010. Faz uma agregação com cardinalidade de 280 elementos. Para cada um dos testes, executa 30 comandos em paralelo.	65 (1 junção) 210 (2 junções) 260 (3 junções)
3	Teste de processamento em paralelo: Mede a performance de SQL executados em paralelo. O ano varia entre 2009 e 2020. Faz uma agregação com cardinalidade de 280 elementos. Executa 50 ou 100 comandos em paralelo, dependendo do teste.	100 (50 comandos) 160 (100 comandos)
4		245 (cardinalidade 280) 282 (cardinalidade 546)

	Teste de Agregação: Mede a performance para agrupamentos de diferentes cardinalidades. O ano varia entre 2009 e 2013. Para cada um dos testes, executa 30 comandos em paralelo.	275 (cardinalidade 1107)
5	Teste de criação de tabelas: Similar ao teste anterior, mas com uma cardinalidade maior. Os resultados são armazenados em novas tabelas. O ano varia entre 2015 a 2019. Executa 30 comandos em paralelo.	615
6	Teste de transformação: Mede a performance com a criação uma de tabela de 18 milhões de registros a partir da tabela NFe, com junção com a tabela dimensão e cancelamento. Não realiza agregação e filtra pelo ano 2010. Executa 30 comandos em paralelo.	245

12.6.3 Com o objetivo de reduzir a ociosidade do ambiente em momentos de pouca utilização, como horário fora do expediente e fins de semana, os recursos computacionais disponíveis para o processamento podem variar automaticamente ao longo do dia. O quadro abaixo apresenta a quantidade total de horas em um mês conforme o recurso de processamento alocado.

Recursos de processamento	Horas
100% dos recursos (performance suficiente para o atingimento das metas)	2 H
75% dos recursos (CPU e memória)	8 H
50% dos recursos (CPU e memória)	70 H
25% dos recursos (CPU e memória)	650 H

12.6.4. Deverá ser previsto um serviço de disponibilização de cluster de servidores capazes de executar aplicações do tipo MapReduce. Deverá ser dimensionado de forma a possibilitar o processamento diário durante 4 horas dos dados recebidos no SPED e para realizar reprocessamentos eventuais com o objetivo de evoluir a forma de armazenamento ou corrigir de eventuais problemas em um total de 1820 horas extras por ano. Para estes processamentos é necessário um cluster com 2000 VCPU e 8GB por core, totalizando 16TB.

12.6.5 Deverá ser previsto um serviço de disponibilização de cluster de servidores capazes de executar aplicações Spark com o objetivo de suportar as consultas e os scripts dos usuários da RFB executados por intermédio do ContAgil. O volume total de armazenamento disponível para o uso deverá ser de 5TB (armazenamento líquido com tolerância a falhas e criptografia). Deverão ser fornecidos 2000 VCPU com 2GB por core, totalizando 4TB. Todos os dias do ano, 24h.

12.6.6 Considerando os riscos de segurança inerentes ao armazenamento de informações, faz-se necessário que o provedor de nuvem adote altos padrões de segurança. Dentre as normas de segurança da informação existentes no mercado, tem-se:

NORMA	ABRANGÊNCIA	JUSTIFICATIVA
ISO/IEC 27001:201	Esta Norma especifica os requisitos para estabelecer, implementar, manter e melhorar continuamente um sistema de gestão da segurança da informação dentro do contexto da organização. Esta Norma também inclui requisitos para a avaliação e tratamento de riscos de segurança da informação voltados para as necessidades da organização.	A observância a esse normativo se faz necessária para assegurar a segurança da informação associada aos requisitos de documentação, divisões de responsabilidade, disponibilidade, controle de acesso, segurança, auditoria e medidas corretivas e preventivas.

ISO/IEC 27017:2015	Norma fornece diretrizes para os controles de segurança da informação aplicáveis à prestação e utilização de serviços em nuvem, fornecendo o seguinte: diretrizes adicionais para implementação de controles relevantes especificados na ISO/IEC 27002; controles adicionais com diretrizes de implementação que são relacionadas especificamente a serviços em nuvem.	A observância desse normativa busca assegurar a oferta de controles adicionais para lidar com ameaças e riscos de segurança de informações específicos da nuvem.
ISO/IEC 27018:204	Esta Norma estabelece objetivos de controle, controles e diretrizes comumente aceitos para implementação de medidas para proteger as Informações de Identificação Pessoal (PII) de acordo com os princípios de privacidade descritos na ISO/IEC 29100, para o ambiente de computação em nuvem pública.	A observância a esse normativo busca assegurar que os consumidores saibam onde os dados deles são armazenados, asseguram também que os dados não serão usados para fins de marketing ou publicidade sem seu consentimento explícito, entre outras garantias relacionadas a proteção individual dos dados.

12.6.7 Por fim, visando não somente a segurança quanto ao sigilo das informações, mas também resguardar a supremacia da legislação brasileira sobre os dados e informações, a IN nº 5, de 5 de agosto de 2021, publicada pelo Gabinete de Segurança Institucional da Presidência da República, autoridade para assuntos de segurança da informação para o Executivo Federal e o anexo à Portaria nº 20, de 14 de junho de 2016 da SGD (STI à época de sua publicação), que trata do assunto de computação em nuvem, determinam que os dados e informações de órgão do governo, contratante de serviços em nuvem, residam exclusivamente em território nacional.

12.7 Infraestrutura como Serviço (IaaS):

12.7.1 Deverá ser previsto um serviço de disponibilização de servidores com capacidade de processamento, memória, armazenamento de dados, e conectividade suficientes para suportar as necessidades dos sistemas abaixo:

	VMs	vCPUs	GPUs	RAM GB	Disco HD	Disco complementar de alta performance
i-Processo	5	40	2	160	20 TB	2TB com 6000 IOPS + 3TB com base de 6000 IOPS e 25000 IOPS por 1/5 do tempo.
FAPE	1	4	0	16	1 TB	1 TB com 6000 IOPS
ChatBots	11	124	5	104	20 TB	1TB com 6000 IOPS + 2TB com 12000 IOPS
VIVII	2	32	0	64	5 TB	1 TB com 6000 IOPS
ANIITA/sisam/Ágata	2	36	0	128	4 TB	2 TB com 6000 IOPS
AJNA/Laudos/API Recintos	6	36	5	360	10 TB	2 TB com 6000 IOPS
SGBD MySQL	2	36	0	128	6 TB	6 TB com 12000 IOPS
Núcleo do SISAM	7	448	0	1.792	9 TB	7 TB (12000 IOPS durante horário comercial e 6000 nos demais horários) + 2 TB com 25000 IOPS por 1/10 do tempo.
Totais	36	756	12	2752	75	29

12.8 Serviços Adicionais:

12.8.1 Com o objetivo de possibilitar a comunicação entre a RFB e o provedor de nuvem, deverá ser previsto uma conexão dedicada entre as partes com pelo menos 10Gbps.

12.8.2 Deverá ainda ser previsto um serviço de sustentação com capacidade de 400 chamadas por ano, 24hx7 com prazo máximo de atendimento de 2 horas.

12.8.3 Além disso, deverá ser previsto um serviço de consultoria com especialista de até 500 horas por ano.

12.8.4 Por fim, os seguintes serviços são necessários para o funcionamento adequado da nuvem: provisionamento de serviços em nuvem, controle de acesso e faturas unificadas, gestão unificada dos serviços, suporte aos clientes e gestão dos níveis de serviços.

13. Estimativa de custo total da contratação

Valor (R\$): 60.723.646,80

13.1. O demonstrativo do cálculo do custo total mensal para fornecimento dos serviços utilização a solução 3 (nuvem) consta do item 11.4 e varia entre R\$1.102.201,39 e 5.060.303,90. O custo anual portanto varia entre R\$ 13.226.416,68 e 60.723.646.80.

14. Justificativa técnica da escolha da solução

14.1 A tabela contendo o comparativo do atendimento de diversas necessidade constante do item 9 demonstra que a melhor escolha sob o aspecto técnico é a solução 3, ou seja, contratar um Cloud Broker.

15. Justificativa econômica da escolha da solução

15.1 A análise comparativa de custos (item 11) demonstra que a melhor escolha sob o aspecto econômico é a solução 3, ou seja, contratar um Cloud Broker.

15.2 Redução dos custos com a aquisição de infraestrutura e manutenção dos centros de dados do órgão central e das diversas superintendências da RFB.

16. Benefícios a serem alcançados com a contratação

16.1 O projeto nuvem possibilitará a utilização do serviço ofertados pelas principais empresas fornecedoras de nuvem. Os serviços que serão disponibilizados possibilitarão migrar todos os serviços existente no Receita Data para a nuvem, hospedar sistemas desenvolvidos pela RFB, além de possibilitar a utilização de tecnologias disruptivas de Inteligência Artificial.

16.2 Os seguintes benefícios são esperados no curto prazo:

- Aumentar o espaço disponível para armazenar dados de interesse econômico fiscal (de 1.43 PB para 2PB).

- Melhorar a performance e a disponibilidade do ambiente analítico.
- Acessar os dados a partir de aplicativos desenvolvidos na RFB ou ferramentas analíticas desenvolvidas por terceiros e homologadas pela RFB.
- Garantir a autonomia da RFB sobre os seus dados, sem as limitações de acesso impostas pelo prestador de serviço.
- Reduzir o intervalo de tempo entre a disponibilização de novidades tecnológicas pelo mercado e a sua efetiva utilização pela RFB, principalmente aquelas relacionadas a inteligência artificial.
- Possibilitar uma maior agilidade em aumentar ou reduzir a capacidade de processamento de acordo com as demandas de negócio, reduzindo a ociosidade do ambiente contratado e consequentemente dos custos associados;
- Eliminar a necessidade de utilização do DataCenter da RFB para os softwares migrados para a nuvem;
- Possibilitar a utilização de Infraestrutura como Serviço na nuvem para o atendimento das necessidades de diversos sistemas desenvolvidos e em desenvolvimento pela RFB.

16.3 Após a migração do RD e dos sistemas desenvolvidos internamente para a nuvem, é esperada uma redução dos custos, uma vez que se trata de uma tecnologia mais barata e altamente escalável. O dimensionamento da capacidade de processamento conforme a necessidade, evita períodos de ociosidade do ambiente, reduzindo ainda mais o custo total e evitando desperdícios de recurso público.

16.4 A estimativa de economia quando a solução estiver implantada e for capaz de substituir o Receita Data é de no mínimo R\$ 1.115.802,19 por mês.

17. Providências a serem Adotadas

17.1 Além da nomeação de equipe técnica para acompanhamento e fiscalização do contrato, serão necessárias as seguintes providências adicionais por parte da RFB no que concerne à adequação do ambiente da organização para que a contratação surta seus efeitos, nos termos da IN nº5, de 30 de agosto de 2021, publicada pelo Gabinete de Segurança Institucional:

- Editar um ato normativo sobre o uso seguro de computação em nuvem, assim como nomear uma equipe responsável por sua revisão;
- Realizar o gerenciamento de riscos, definindo as medidas de mitigação de riscos;
- Capacitar a equipe responsável pelo gerenciamento da nuvem.

18. Declaração de Viabilidade

Esta equipe de planejamento declara **viável** esta contratação.

18.1. Justificativa da Viabilidade

Ficou demonstrado neste documento que a solução escolhida possibilitará o atendimento a todas as necessidades relacionadas no item 2, além disso, pelo levantamento de mercado feito, tanto o Serpro, como a Prodesp possuem condições de fornecer os serviços de PaaS e IaaS ofertados pela Amazon, Microsoft e Google, além de fornecerem os serviços adicionais necessários para o suporte, administração e conexão aos dados e serviços das nuvens. Por fim, como demonstrado no item 12, haverá redução a médio e longo prazos dos custos relacionados a manutenção do ambiente analítico da RFB.

19. Responsáveis

Sou favorável

CLAUDIO VASCONCELOS BRAGA

Auditor-Fiscal da Receita Federal do Brasil

Sou favorável

FELIPE MENDES MORAES

Auditor-Fiscal da Receita Federal do Brasil - Integrante Requisitante

Sou favorável

JULIANO BRITO DA JUSTA NEVES

Auditor-Fiscal da Receita Federal do Brasil - Coordenador-Geral de Tecnologia e Segurança da Informação



PÁGINA DE AUTENTICAÇÃO

A Secretaria da Receita Federal do Brasil garante a integridade e a autenticidade deste documento nos termos do Art. 10, § 1º, da Medida Provisória nº 2.200-2, de 24 de agosto de 2001

Histórico de atividades sobre o documento:

Documento assinado digitalmente por:

FELIPE MENDES MORAES em 22/10/2021, CLAUDIO VASCONCELOS BRAGA em 21/10/2021, JULIANO BRITO DA JUSTA NEVES em 21/10/2021.

Confira o documento original pelo Smartphone conectado à Internet:



Dúvida? Acesse

<http://sadd.receita.fazenda.gov.br/sadd-internet/pages/qrcode.xhtml>

Confira o documento original pela Internet:

a) Acesse o endereço:

<http://sadd.receita.fazenda.gov.br/sadd-internet/pages/validadocumento.xhtml>

b) Digite o código abaixo:

AP22.1021.09136.7222

Código Hash obtido através do algoritmo SHA-256:

gizHKBNly3PxrV+SbcftTzuKnIyim3Kwh60e1Xr43M=



Ministério da Fazenda

PÁGINA DE AUTENTICAÇÃO

O Ministério da Fazenda garante a integridade e a autenticidade deste documento nos termos do Art. 10, § 1º, da Medida Provisória nº 2.200-2, de 24 de agosto de 2001 e da Lei nº 12.682, de 09 de julho de 2012.

Documento autenticado digitalmente conforme MP nº 2.200-2 de 24/08/2001. Corresponde à fé pública do servidor, referente à igualdade entre as imagens digitalizadas e os respectivos documentos ORIGINAIS.

Histórico de ações sobre o documento:

Documento juntado por CLAUDIO VASCONCELOS BRAGA em 01/12/2021 16:21:00.

Documento autenticado digitalmente por CLAUDIO VASCONCELOS BRAGA em 01/12/2021.

Esta cópia / impressão foi realizada por RECEITA FEDERAL - PARA USO DO SISTEMA em 06/12/2021.

Instrução para localizar e conferir eletronicamente este documento na Internet:

1) Acesse o endereço:

<https://cav.receita.fazenda.gov.br/eCAC/publico/login.aspx>

2) Entre no menu "Legislação e Processo".

3) Selecione a opção "e-AssinaRFB - Validar e Assinar Documentos Digitais".

4) Digite o código abaixo:

EP06.1221.13245.JLG4

5) O sistema apresentará a cópia do documento eletrônico armazenado nos servidores da Receita Federal do Brasil.

Código hash do documento, recebido pelo sistema e-Processo, obtido através do algoritmo sha2:

79D99725A3C7EEC2FB92B78E6D31FC12529F2ED88341B319F1B1FE6ADE84A94E