



Termo de Cooperação/Projeto:

**Termo de Cooperação Técnica  
FUB/FUNAPE e PGFN  
Pesquisas e inovação tecnológica  
aplicadas às temáticas da informação  
e das comunicações no domínio da  
Procuradoria-Geral da Fazenda  
Nacional – PGFN**

Desenvolvimento e inovação visam realizar um estudo de inovações tecnológicas, nas áreas de tecnologia da informação, com ênfase na interoperabilidade de sistemas, gestão corporativa, processos de gestão, gerência de redes e gestão de dados, para áreas estratégicas da Coordenação-Geral de Tecnologia da Informação (CGTI) da Procuradoria-Geral da Fazenda Nacional – PGFN.

Documento:

**RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial**

Data de Emissão:

**09/02/2023**

Elaborado por:

**Universidade de Brasília – UnB  
Laboratório de Tecnologias da Tomada de Decisão  
– LATITUDE.UnB**

Digital Object Identifier:

**10.5281/zenodo.7781710**

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	<b>Pág.2/45</b>
---------------	---------------------	---	-----------------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

**PROCURADORIA-GERAL DA FAZENDA NACIONAL –  
PGFN**

**UNIVERSIDADE DE BRASÍLIA**

**Ricardo Soriano de Alencar**  
Procurador-Geral da Fazenda Nacional

**Márcia Abrahão Moura**  
Reitora

**Daniel de Saboia Xavier**  
Procurador da Fazenda Nacional

**Prof.<sup>a</sup> Maria Emília Machado Telles Walter**  
Decana em exercício  
Decanato de Pesquisa e Inovação – DPI

**José Renato Fragoso Lobo**  
Coordenador-Geral de Gestão de Pessoas e Desenvolvimento Institucional

**Rafael Timóteo de Sousa Júnior**  
Coordenador do Laboratório de Tecnologias da Tomada de Decisão – LATITUDE

**Aleksey Lanter Cardoso**  
Diretor de Gestão Corporativa

**Rodrigo Otávio Pova Pullen Parente**  
Coordenador-Geral de TI

**EQUIPE TÉCNICA**

**EQUIPE TÉCNICA**

**Alan Zuanella Vila Nova**  
**Alex Aranda**  
**Fernando Maciel Lima e Sousa**  
**Hiago Pereira Kanashiro**  
**Julio Sergio Fernandes Alves**  
**Leandro Veloso Rodrigues**  
**Priscila Fatima Pinheiro de Siqueira**  
**Rosiane Aparecida Moreira**  
**Thiago Oliveira Hoerlle**  
Equipe da Algar TI Consultoria S/A

**Rafael Timóteo de Sousa Júnior**  
(Pesquisador Sênior)  
**Fábio Lúcio Lopes de Mendonça**  
(Pesquisador Sênior)  
**Georges Daniel Amvame Nze**  
(Pesquisador Sênior)  
**Robson de Oliveira Albuquerque**  
(Pesquisador Sênior)  
**William Ferreira Giozza**  
(Pesquisador Sênior)  
**Alex Mendes Pacheco da Silva**  
**Allan Filipe Almeida**  
**Ana Beatrice Neubauer de Moura**  
**Ana Paula Bernadi da Silva**  
**Ana Paula Morais Vale**  
**Ayra de Avila Almeida**  
**Caio Henrique Caetano**  
**Carlos Eduardo Ramalho de Souza**  
**Cleuber Santos Silva**  
**Diego Martins de Oliveira**  
**Felipe Barreto de Oliveira**  
**Flávio Garcia Praciano**  
**Gabriel Ribeiro de Araújo**  
**Guilherme Batista Meneses Alves**  
**Hiago Pereira Kanashiro**  
**Isaac Silva Martins**  
**Ismael Ithalo Barbosa Pinto**  
**João Paulo da Costa e Silva Garcia**  
**João Paulo Pimentel**  
**Johnan Nicholas Reed**  
**Klainer Mateus Estrela Gomes**  
**Kelly Santos de Oliveira Bezerra**  
**Leonardo Jorge França Morais**  
**Letícia Ramos Reinaldi**  
**Lucas da Silva Barbosa**  
**Luiz Augusto dos Santos Pires**  
**Marcos Tércio Ramos**  
**Maria Karoline Domingues**  
**Maria Tereza Correa Pacheco Alves**  
**Marília do Nascimento T. Valentim**

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.3/45
---------------	---------------------	---	----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

**Orlando Werbeth dos Santos Gomes**  
**Paulo Henrique Batista Rodrigues**  
**Paulo Lima Machado**  
**Phelipe Alan Almeida**  
**Priscila Batista Rodrigues**  
**Renato Jose da Silva Camoes**  
**Thiago Leite de Souza**  
**Valeria Simas Schultz**  
**Wellington Domingos Neves**

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	<b>Pág.4/45</b>
---------------	---------------------	---	-----------------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

## HISTÓRICO DE REVISÕES

Data	Versão	Descrição
03/03/2022	1.0	Versão inicial do documento para homologação.
12/05/2022	1.1	Atualização geral das informações do documento para homologação.
21/01/2023	1.2	Implementação do tópico do projeto Jucesp.
25/01/2023	1.3	Atualização da lista das figuras.
01/01/2023	1.4	Implementação do tópico do projeto Datajud.
08/02/2023	1.5	Correção dos parágrafos, figuras, sumários e correções gramaticais pontuais.
09/02/2023	1.6	Alteração de rodapé e ajustes pontuais.



Universidade de Brasília – UnB  
Campus Universitário Darcy Ribeiro – FT – ENE – Latitude  
CEP 70.910-900 – Brasília-DF  
Tel.: +55 61 3107-5598 – Fax: +55 61 3107-5590

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.5/45
---------------	---------------------	---	----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

PGFN	Procuradoria-Geral da Fazenda Nacional
TED	Termo de Execução Descentralizada
TIC	Tecnologias de informação e comunicação
RT	Relatórios técnicos
UnB	Universidade de Brasília
PPGEE	Programa de Pós-Graduação em Engenharia Elétrica
FUB	Fundação Universidade de Brasília
EAP	Estrutura Analítica do projeto
EAR	Estrutura Analítica de Risco
TI	Tecnologia da Informação
BPMN	Business Process Model and Notation
PMI	Project Management Institute
PMBOK	Project Management Body of Knowledge
PDI	Pentaho Data Integration
HDFS	Hadoop Distributed File System

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	<b>Pág.6/45</b>
---------------	---------------------	---	-----------------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
 É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.



## LISTA DE FIGURAS

Figura 1 – Arquitetura do projeto Datajud.....	14
Figura 2 – Estrutura do arquivo JSON.....	15
Figura 3 – Scripts para o Impala.....	17
Figura 4 – Processo ETL.....	17
Figura 5 – Arquitetura do projeto da Jucesp.....	19
Figura 6 – Tela de login do “Rundeck”.....	34
Figura 7 – Tela de projetos do “Rundeck”.....	35
Figura 8 – Criando novos projetos.....	36
Figura 9 – Deletando source padrão.....	36
Figura 10 – Criando novo source.....	37
Figura 11 – Ícone “comandos”.....	38
Figura 12 – Selecionando os “nós”.....	38
Figura 13 – Nó correspondente.....	38
Figura 14 – Exemplo de comando.....	38

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.7/45
---------------	---------------------	---	----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

1	INTRODUÇÃO	10
2	OBJETIVOS DO PROJETO	12
2.1	OBJETIVO GERAL	12
2.2	META GERAL	12
3	METODOLOGIA	13
4	ESTRUTURA E ATIVIDADES DO PROJETO (RESULTADOS)	14
4.1	DATAJUD	14
4.1.1	Problema	14
4.1.2	Arquitetura de projeto	14
4.1.3	Dados	15
4.1.4	Estrutura das tabelas	16
4.1.5	Processo ETL	17
4.1.6	Conclusão do projeto Datajud	18
4.2	JUCESP	18
4.2.1	Problema	18
4.2.2	Arquitetura do projeto	19
4.2.3	Máquinas virtuais	20
4.2.3.1	VM Buscador	20
4.2.3.2	VM Banco de dados (Logs)	20
4.2.3.3	VM Pentaho	20
4.2.3.4	VM Orquestrador – Rundeck	21
4.2.4	Implantação – VM Buscador	21
4.2.4.1	Habilitação de saída para a internet	22
4.2.4.2	Instalação do Python e bibliotecas	22
4.2.5	Implantação – VM Banco de dados	22
4.2.5.1	Criar usuários administrativo	22
4.2.5.2	Criar tabelas para receber os registros de Logs	23
4.2.6	Implantação – VM Banco de dados	25
4.2.6.1	Habilitação de acesso ao servidor Pentaho pelo SSH	25
4.2.6.2	<i>Acesso do usuário</i>	27
4.2.6.3	Arquivos de configuração	27
4.2.6.4	Ferramentas	28
	<b>4.2.6.4.1 LFTP</b>	28
	<b>4.2.6.4.2 Kerberos</b>	29

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.8/45
---------------	---------------------	---	----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.



<b>4.2.6.4.3 Pentaho Data Integration</b>	30
4.2.7 Implantação – Orquestrador - Rundeck	31
4.2.8 Teste de conexão	39
4.2.8.1 Teste de conectividade Pentaho PDI para FTPS	39
4.2.8.2 Teste de conectividade Pentaho PDI para Impala	40
4.2.8.3 Teste de conectividade SSH entre Orquestrador (rundeck) e Pentaho	40
4.2.8.4 Teste de conectividade SSH entre Orquestrador (rundeck) e Buscador	41
<b>5 CONCLUSÃO</b>	42
<b>REFERÊNCIAS</b>	44

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	<b>Pág.9/45</b>
---------------	---------------------	---	-----------------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
 É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

## 1 INTRODUÇÃO

Com base na Lei nº 2.642, de 9 de novembro de 1955, houve a criação da Procuradoria-Geral da Fazenda Nacional – PGFN, na forma atualmente conhecida, em substituição à Procuradoria-Geral da Fazenda Pública. Instituída como órgão de consultoria jurídica do Ministério da Fazenda, à PGFN era atribuída, principalmente, examinar e fiscalizar os contratos de interesse da União, apurar e inscrever a dívida ativa federal para fins de cobrança judicial e cooperar com o Ministério Público da União junto à justiça comum (art. 1º).

O Decreto-Lei nº 147, de 3 de fevereiro de 1967, estabeleceu a segunda lei orgânica da PGFN. Esse diploma legislativo fixou competências até hoje mantidas pelos demais atos normativos que o sucederam, na mesma direção do que previa a Lei nº 2.642, de 1955, estabelecendo o seguinte: a) a vinculação administrativa da PGFN como órgão do Ministério da Fazenda responsável pela prestação de serviços jurídicos da Pasta; b) a atribuição de apurar e inscrever, para fins de cobrança judicial, a dívida ativa da União, tributária ou de qualquer outra natureza; c) e sua atuação nacional por força da descentralização do órgão.

Com a promulgação da Constituição da República de 1988, houve uma mudança significativa da Procuradoria-Geral da Fazenda Nacional quanto a sua vinculação exclusiva ao Ministério da Fazenda. A PGFN passou a integrar a nascente Advocacia-Geral da União, órgão criado para defender, judicial ou extrajudicialmente, os interesses da União.

A Lei Complementar nº 73, de 10 de fevereiro de 1993, que institui a Lei Orgânica da Advocacia-Geral da União, previu, expressamente, a subordinação técnica e jurídica da PGFN ao Advogado-Geral da União, confirmando a finalidade do legislador constituinte em vincular a Procuradoria como órgão da PGFN responsável pela atuação na área fiscal. Com isso, a PGFN tornou-se órgão de direção superior da Advocacia-Geral da União e suas atribuições residem, principalmente, na representação da União em causas fiscais, na cobrança judicial e administrativa dos créditos tributários e não-tributários e no assessoramento e consultoria no âmbito do Ministério da Fazenda onde atualmente encontra-se no âmbito do Ministério da Economia.

De outro lado, temos o Laboratório de Tecnologias da Tomada de Decisão – LATITUDE, que foi criado em 2010 com recursos da Lei de Informática provenientes da DELL Computadores do Brasil, constituído como um ambiente de inovação e

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.10/45
---------------	---------------------	---	-----------

Confidencial.

desenvolvimento para projetos de pesquisa interdisciplinar entre as engenharias, a computação, a ciência da informação, bem como os demais domínios do conhecimento de interesse para a temática focal da tomada de decisão.

Criado no âmbito do Departamento de Engenharia Elétrica da Universidade de Brasília - UnB, onde tem seu espaço físico próprio, o Laboratório LATITUDE é vinculado ao Programa de Pós-Graduação em Engenharia Elétrica – PPGEE e ao Programa de Pós-Graduação Profissional em Engenharia Elétrica – Segurança Cibernética, contando também com professores, pesquisadores e estudantes de Programas de Pós-Graduação das áreas de Engenharia Civil, Engenharia Computação, Engenharia da Produção, Engenharia da Automação, Engenharia de Software, Ciência da Computação, Ciência da Informação, Administração, Direito, Educação, Psicologia, História, bem como dos respectivos cursos de graduação, nos Campus Darcy Ribeiro e Gama.

A PGFN vem buscando, gradativamente, desenvolver soluções para sanar parte dos problemas existentes nas áreas de tecnologia da informação, administração interna, gestão corporativa, pessoal e documental. Para isso foram realizados estudos em diversas áreas da PGFN, realizando um levantamento dos problemas e sua descrição.

No que se relaciona às tecnologias da informação e das comunicações, é necessário trazer inovações tecnológicas aos processos e sistemas administrativos que se integram aos sistemas estruturantes (Sistemas SIDA, SAJ, Flexa, SISPAR, Regularize, Dívida, FGTS) de forma que a PGFN consiga atender às suas demandas de forma rápida e eficiente, fornecendo uma plataforma inteligente que auxilia tanto o acesso às diversas pesquisas internas e externas realizadas como à tomada de decisão. Entende-se que esse aprimoramento envolve o sistema informacional, que organize a informação e banco de dados customizado, de acordo com necessidades das áreas de negócio envolvidas, procurando as seguintes resultantes: identificar, mitigar e tratar riscos, quantificar a ocorrência de fenômenos, mapear correspondentes processos. Além do acompanhamento contínuo das informações, com aprimoramento das análises administrativas e das defesas apresentadas. Ante o exposto, complementa-se que a informação é hoje um dos patrimônios mais importantes de uma organização, seja ela pública ou privada. As Tecnologias da Informação e Comunicações (TIC) se consolidam como ativo estratégico, onde integra recursos, processos, métodos, técnicas para obter, processar, armazenar, disseminar e fazer uso da informação. Sob essa ótica, a governança de TIC garante a boa e regular gestão dos serviços de TIC, que se desdobram ao encontro da estratégia

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.11/45
---------------	---------------------	---	-----------

Confidencial.

corporativa do órgão. Entretanto, esse alinhamento só é viável com a estruturação de um planejamento que reflita como a TIC contribuirá, através do alcance das suas metas e ações, para o alcance dos objetivos organizacionais.

Assim, é pertinente e relevante evoluir a maturidade das tecnologias utilizadas em tais sistemas de informação, assim como dos processos de gestão e governança associados. Em especial, coloca-se a necessidade de interoperação entre sistemas, assim como a preparação para a interoperação com outros sistemas que deverão ser concebidos, desenvolvidos e operacionalizados. Tal necessidade precisa de soluções inovadoras no que se refere à semântica da informação e à algorítmica de operação simultânea e paralela de módulos de sistemas interdependentes, sejam eles internos à PGFN, sejam sistemas externos que necessitem de serviços ou informações dos sistemas da PGFN. Colocam-se nesse contexto as questões de confidencialidade, integridade e disponibilidade das informações e serviços, o que implica no requisito de agregar uma abordagem de segurança da informação às atividades de gestão de sistemas de informação da PGFN. O próprio planejamento diretor desses sistemas e das respectivas tecnologias de suporte (armazenamento, processamento, rede, acesso, etc.) merece estudos que levem a seu aprimoramento e sua colocação em um processo de governança que contribuam para uma melhoria continuada da maturidade da PGFN em tais domínios de tecnologias da informação.

## 2 OBJETIVOS DO PROJETO

### 2.1 OBJETIVO GERAL

O presente projeto de pesquisa, desenvolvimento e inovação visa realizar um estudo de inovações tecnológicas, nas áreas de tecnologia da informação, com ênfase na interoperabilidade de sistemas, gestão corporativa, processos de gestão, gerência de redes e gestão de dados, para áreas estratégicas da Coordenação-Geral de Tecnologia da Informação (CGTI) da Procuradoria-Geral da Fazenda Nacional – PGFN.

### 2.2 META GERAL

O presente Projeto de pesquisa e desenvolvimento visa conceber e orientar a implementação de novas tecnologias de informação e comunicação vinculadas à PGFN,

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.12/45
---------------	---------------------	---	-----------

Confidencial.

por meio das atividades previstas na EAP, estabelecendo novos mecanismos de coleta e mineração de dados com o objetivo de gerar indicadores gerenciais.

Conseqüentemente, o Projeto busca contribuir para a modernização da Administração Pública, a otimização dos recursos, a racionalização dos custos e a melhoria da qualidade dos serviços prestados.

### 3 METODOLOGIA

No que diz respeito à gestão executiva do Projeto, será empregada a metodologia de gerenciamento de projetos baseada nos princípios do Project Management Institute (PMI) com adoção de ferramenta própria para suporte das atividades de gerenciamento sob a filosofia de *groupware* baseado em ferramentas de software livre.

A aplicação de metodologia de gestão de Projetos (gestão executiva) baseada na visão do PMI deverá criar um equilíbrio entre as demandas de escopo, tempo, custo, qualidade e bom relacionamento entre os diversos atores do Projeto. O sucesso dessa gestão estará relacionado ao alcance dos objetivos de: entrega dentro do prazo previsto, dentro do custo orçado, com nível de desempenho adequado, com plena aceitação pelo cliente e seus representantes (usuários finais), com atendimento de forma controlada às mudanças de escopo e em respeito à cultura da organização segundo o *Project Management Body of Knowledge (PMBOK) Guide* 5ª edição.

Cabe ressaltar que o gerenciamento de projetos, na visão do PMI, de acordo com o PMBOK, identifica e descreve as principais áreas de conhecimento e práticas. Cada uma destas áreas é descrita por meio de processos, e se refere a um aspecto a ser considerado dentro da gerência de projetos. As áreas de conhecimento de gerenciamento são: Gerenciamento de Integração do Projeto, Gerenciamento de Escopo do Projeto, Gerenciamento do Tempo do Projeto, Gerenciamento do Custo do Projeto, Gerenciamento da Qualidade do Projeto, Gerenciamento de Recursos Humanos do Projeto, Gerenciamento de Comunicação do Projeto, Gerenciamento do Risco do Projeto e Gerenciamento de Contratação do Projeto.

Outro ponto importante é que o uso destas metodologias, tanto de gestão executiva como de execução, já consolidadas na literatura e no mercado, permitirá o uso de artefatos padrões como marcos regulatórios e de gestão de cada um dos relatórios a serem entregues.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.13/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

Finalmente, ainda considerando os aspectos metodológicos, é importante ressaltar que visando atender os fatores de qualidade tanto na gestão executiva como no desenvolvimento/execução do Projeto, deverão ser:

- Organizados seminários e workshops periódicos, com a equipe UnB e a equipe da PGFN, para divulgação e acompanhamento das atividades relacionadas ao Projeto;
- Observada a qualidade pela contratação de pesquisadores qualificados para formação das equipes de trabalho.

## 4 ESTRUTURA E ATIVIDADES DO PROJETO (RESULTADOS)

### 4.1 DATAJUD

O Projeto Datajud (Dados do Poder Judiciário) consiste na elaboração de processos de *ETL* entre dados obtidos periodicamente pela PGFN, de formato semiestruturado (*JSON*), para que sejam transformados e normalizados para o formato tabular e sejam carregados em um servidor de arquivos de forma distribuída (*HDFS*) e, posteriormente convertidos para um formato otimizado (*Parquet*) e mapeados/consultados no Impala através de queries SQL ANSI.

#### 4.1.1 Problema

A fonte de dados é entregue à PGFN de tempos em tempos de forma semiestruturada e separada em diversos arquivos, de forma que dificulta a consulta e extração dos dados pelo órgão. Sendo assim, a equipe de pesquisa da UnB atua de forma a ler todos os arquivos recebidos, estruturá-los de forma tabular para que sejam melhores consultados e armazenados de forma otimizada para aproveitar da melhor forma o servidor de arquivos distribuído.

#### 4.1.2 Arquitetura de projeto

Para o planejamento dos trabalhos e implementação da arquitetura que se desse junto ao alinhamento das expectativas da Procuradoria-Geral da Fazenda Nacional (PGFN), as equipes de dados da Universidade de Brasília (UNB) e o órgão promoveram debates junto às reuniões semanais para que se pudesse alinhar o projeto para que fosse chegar em um modelo que atende-se ao problema.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.14/45
---------------	---------------------	---	-----------

Confidencial.

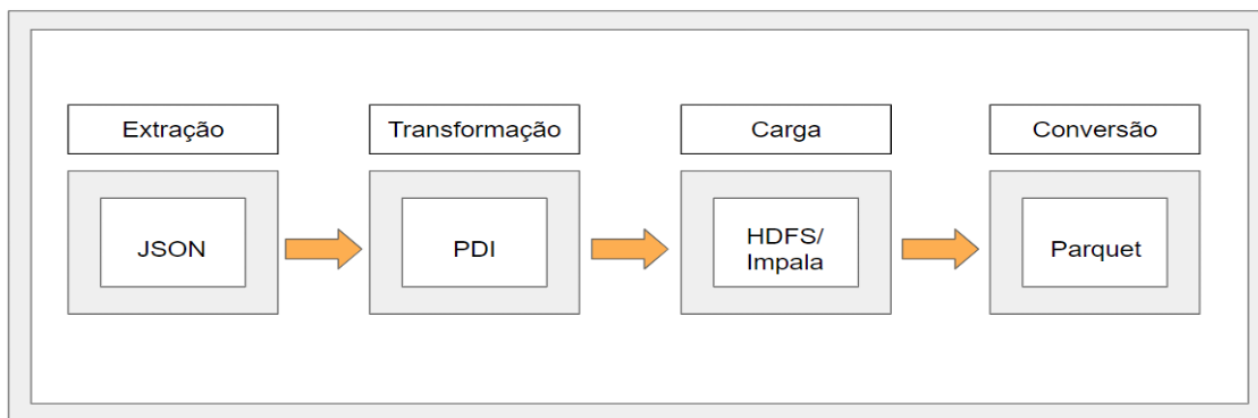


Figura 1 – Arquitetura do projeto Datajud

Fonte: Elaborada pelo autor

Na imagem acima, **figura 1**, podemos visualizar a arquitetura definida para o projeto, sendo os dados obtidos no formato *JSON* e dispersos em vários arquivos separados por estado e por tamanho. Já no segundo passo, temos as transformações realizadas pela ferramenta PDI (Pentaho Data Integration) para que os dados pudessem ser normalizados e inseridos no sistema de arquivos distribuído que, em nosso caso, o arquivo é enviado a um servidor *FTPS* disponibilizado pelo SERPRO. Este servidor é responsável por enviar automaticamente os arquivos para o *HDFS* e posteriormente distribuir os dados em tabelas do Impala. Após os dados serem carregados nas tabelas, eles são convertidos para o formato *Parquet* que é uma estrutura de arquivos em formato de colunas que proporciona consultas mais otimizadas no momento da exibição dos dados.

#### 4.1.3 Dados

Conforme a **figura 2**, verificamos um exemplo dos dados periodicamente recebidos pela PGFN e que são disponibilizados no formato *JSON* (JavaScript Object Notation), que veio para ser uma alternativa ao modelo *XML* e para dispor os dados na forma de atributo e valor. O formato é bastante utilizado por ser mais simples e leve para a comunicação entre aplicações e pelo modelo facilmente maleável, se comparado ao formato tabular. Porém, quando utilizamos esse formato para um grande volume de dados e em uma estrutura fixa, a recuperação dos dados acabam sendo impactadas pela redundância da quantidade de atributos e torna também a consulta bastante complexa e onerosa de ser feita, se comparada a uma query em dados estruturados em formato tabular.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.15/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN. É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

```

{
  "_index": "processos-tjto",
  "_type": "_doc",
  "_id": "TJTO_202_G2_33649_00308383520198270000",
  "_score": null,
  "_source": {
    "millisInsercao": 1595980458332,
    "dadosBasicos": {
      "valorCausa": 0.0,
      "numero": "00308383520198270000",
      "procEl": 1,
      "competencia": 2,
      "codigoLocalidade": "1",
      "grau": "G2",
      "assunto": [{
        "principal": true,
        "codigoNacional": 5000
      }, {
        "principal": false,
        "codigoNacional": 9418
      }
    ],
    "dscSistema": "8",
    "tamanhoProcesso": 320743,
    "dataAjuizamento": "20191024115558",
    "polo": [{
      "polo": "PA",
      "parte": [{
        "pessoa": {
          "tipoPessoa": "JURIDICA",
          "endereco": [{
            "cidade": "Palmas",
            "estado": "TO",
            "numero": "S/N",
            "logradouro": "QD. 104 NORTE, RUA NE 01, LOTE 18, SALA 03",
            "bairro": "PLANO DIRETOR NORTE",
            "pais": "BRASIL",
            "cep": "77006016"
          }, {
            "cidade": "Palmas",
            "estado": "TO"
          }
        ]
      }
    ]
  }
}

```

Figura 2 – Estrutura do arquivo JSON

Fonte: Elaborada pelo autor

#### 4.1.4 Estrutura das tabelas

Ao analisarmos os dados, conseguimos separar as informações em 4 grandes tabelas e 4 tabelas filhas para o armazenamento dos dados estruturados. Dessa forma as tabelas possuem duas devidas chaves compostas, sendo feitas através da combinação entre a linha do arquivo e nome do registro, de forma que é possível ligações entre as tabelas e consultas utilizando *Joins* entre elas. São elas:

- **Dados básicos;**
- **Assuntos;**
- **Movimentos;**
- **Movimentos nacionais;**
- **Movimentos locais;**
- **Polos;**

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.16/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.



- Polo parte;
- Polo parte advogado.

#### 4.1.5 Processo ETL

O processo de *ETL* (Extract, Transform e Load) de carga, consiste em gerar dados para uma tabela de contexto, contendo um hashid e data da carga para que seja possível recuperar as versões das cargas posteriormente. Em seguida, cada tabela possui sua etapa de extrair os dados do arquivo, mapear e estruturar cada coluna para o formato tabular e inseri-los em um servidor *FTPS* para que os dados sejam enviados para o *HDFS* e inseridos no Impala através de Scripts, conforme a **figura 3**. Assim, o Script que é executado no Impala cria uma tabela temporária para que os dados sejam armazenados no formato de texto, carrega eles na tabela através do mapeamento da pasta no *HDFS* e a tabela Parquet também é alimentada para que os dados sejam consumidos de forma mais rápida. Após o término da carga da tabela no formato otimizado, a tabela temporária é truncada para não armazenar os dados de forma desnecessária e excluída. Todo o processo ETL pode ser visualizado em sua íntegra na **figura 4**. O processo se inicia na caixa “Start”, executa todos os passos detalhados acima e se conclui na caixa denominada “Sucesso”.

```
CREATE EXTERNAL TABLE IF NOT EXISTS pgfn_ingestao_cgti.polos_hdfs_txt_teste (
  auto_id_registro_json bigint COMMENT 'Chave gerada sequencialmente no momento da carga da tabela',
  nome_arquivo string COMMENT 'Chave com o nome do arquivo que foi processado',
  id_polo bigint COMMENT 'Chave gerada sequencialmente com o id da linha do arquivo gerado',
  polo string COMMENT 'UF do polo',
  partes string COMMENT 'JSON com os dados das partes dos polos',
  id_carga string COMMENT 'Chave estrangeira da tabela (datajud_carga) com a chave da carga'
)
COMMENT 'tabela responsavel por armazenar os dados dos polos vindos do HDFS'
ROW FORMAT delimited fields terminated BY '|'
STORED AS textfile;

LOAD DATA inpath '/opt/approtinas/pgfn/s_pgfn_etl/polos.csv' OVERWRITE INTO TABLE pgfn_ingestao_cgti.polos_hdfs_txt_teste;

CREATE TABLE IF NOT EXISTS pgfn_ingestao_cgti.datajud_polos_teste LIKE pgfn_ingestao_cgti.polos_hdfs_txt_teste STORED AS parquet;

INSERT INTO pgfn_ingestao_cgti.datajud_polos_teste SELECT * FROM pgfn_ingestao_cgti.polos_hdfs_txt_teste;

TRUNCATE TABLE pgfn_ingestao_cgti.polos_hdfs_txt_teste;

DROP TABLE pgfn_ingestao_cgti.polos_hdfs_txt_teste;
```

Figura 3 – Scripts para o Impala

Fonte: Elaborada pelo autor

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.17/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

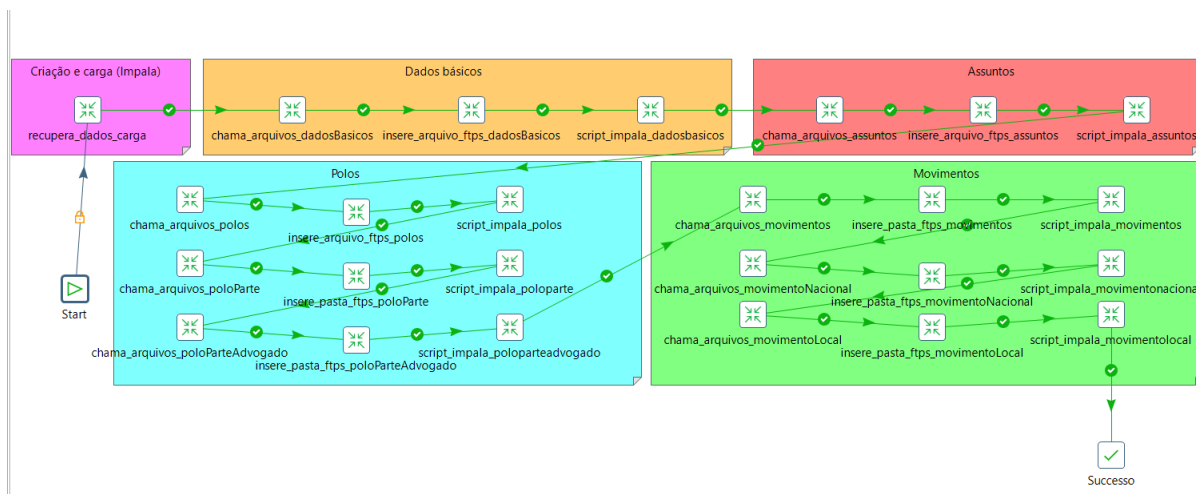


Figura 4 – Processo ETL

Fonte: Elaborada pelo autor

#### 4.1.6 Conclusão do projeto Datajud

Após a conclusão de todo o processo de ETL, estruturação das tabelas e pleno funcionamento do Job, a PGFN optou por colocar o projeto Datajud em processo de freezing devido à dificuldade de obtenção e confiança na fonte de dados recebidas periodicamente. Contudo, todo o processo foi válido para que o projeto da Jucesp conseguisse ser viabilizado, uma vez que, ele também usa o armazenamento dos dados em um servidor FTPS do SERPRO, inserido no HDFS e carregado em tabelas no Impala através de Scripts chamados pela ferramenta Pentaho Data Integration.

## 4.2 JUCESP

O projeto da Jucesp (Junta Comercial do Estado de São Paulo) consiste na implementação de um pipeline de dados completa, que vai desde a aquisição dos dados de empresas e sócios no *Web Service* da Jucesp até a carga destes dados no Data Lake Hadoop, uma estrutura de computação distribuída voltada para clusters e processamento de grandes volumes de dados, com atenção a tolerância a falhas.

### 4.2.1 Problema

Os dados de empresas e sócios de São Paulo são disponibilizados de segunda a sábado no *Web Service* da Jucesp em formato de texto. Em suma, a equipe de dados da PGFN necessitava de uma infraestrutura de Big Data para processamento e

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.18/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN. É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

armazenamento desses dados para centralizar a recuperação desses dados. Como outro requisito, todo o processo deveria ser cem por cento automatizado e os Logs de execução armazenados em um banco de dados.

#### 4.2.2 Arquitetura do projeto

Durante todo o processo de planejamento e execução da implementação da arquitetura que atendesse as necessidades do cliente, as equipes de dados da Universidade de Brasília (UNB) e da Procuradoria-Geral da Fazenda Nacional (PGFN) promoveram reuniões semanais e discussões de alinhamento do projeto que subsidiaram a concepção da seguinte arquitetura do projeto da Jucesp, conforme a **figura 5**.

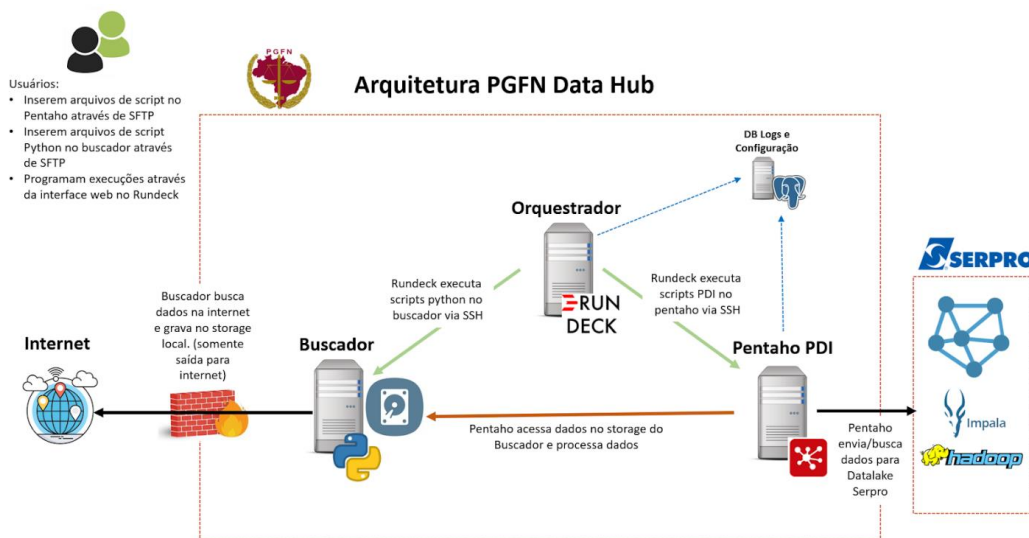


Figura 5 – Arquitetura do projeto da Jucesp

Fonte: Elaborada pelo autor

O processo de ETL - (Extrair, Transformar, Carregar) é coordenado pelo orquestrador, pelo uso da ferramenta **Rundeck**. No orquestrador são configurados os processos de ETL, que, em resumo, consistem dos passos:

1. Rundeck comanda o Buscador para executar Scripts Python que buscarão dados nas fontes necessárias (internet) e armazenam os mesmos, no seu formato original, em uma área de disco do próprio buscador, acessível pelo servidor Pentaho.
2. Rundeck comanda Pentaho para executar rotinas que acessam as informações obtidas pelo buscador, realizam transformações e fazem a ingestão dos dados no Data Lake.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.19/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN. É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

O **Pentaho Data Integration** é, portanto, a ferramenta principal para realização do processo de ETL. Para a etapa final de carga, os dados são inseridos no Data Lake em duas etapas:

1. Primeiramente são inseridos os dados no Data Lake por um servidor SFTP em formato.csv;
2. Após isso, são executados comandos Impala para transformação destes dados em tabelas.

### 4.2.3 Máquinas virtuais

#### 4.2.3.1 VM Buscador

Recursos	Detalhamento
Sistema Operacional	Linux CentOS 8
vCPUs	8
RAM	8 GB
Volumes	SO: 40GB volume/dados 500GB
Observações	O volume /datahub/data deve ser acessível pelo Buscador e também pelo servidor do Pentaho (leitura e escrita).

#### 4.2.3.2 VM Banco de dados (Logs)

Recursos	Detalhamento
Sistema Operacional	Linux CentOS 8
vCPUs	4
RAM	4 GB
Hostname	vmcentos7postgresqldatahub
Banco de dados	PostgreSQL 13
Volumes	SO: 40GB /pgdata/data 100GB e /pgdata/backup 50GB

#### 4.2.3.3 VM Pentaho

Recursos	Detalhamento
Sistema Operacional	Linux CentOS 8
vCPUs	8
RAM	16 GB
Volume	/aplicacao/ 200 GB
Servidor de Aplicação	JDK 8
Ferramenta de Integração de dados	Pentaho Data Integration
Ferramenta de acesso ao FTPS	LFTP – versão 4.8.4
Ferramenta de autenticação	Kerberos 5 – versão 1.18.2
Observações	Todas as portas e IPs devem estar liberados para que a conexão seja estabelecida corretamente e o arquivo de keytab deve estar salvo no servidor.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.20/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

#### 4.2.3.4 VM Orquestrador – Rundeck

O Orquestrador – Rundeck será implantado na infraestrutura de **Docker** da PGFN.

Observações pertinentes:

- VM Banco de Dados PostgreSQL: é utilizada somente para armazenamento dos Logs de execução das Jobs e Transformações.
- Jobs: As Jobs em questão são arquivos gerados a partir do Pentaho Data Integration e são usadas para disparar execuções de forma sequencial de ações e transformações, sendo possível identificá-las através da extensão .kjb.
- Transformações: Também são arquivos gerados a partir do Pentaho Data Integration, contudo a execução de suas etapas ocorre de forma paralela, sendo possível identificá-las através da extensão .ktr.
- VM Datahub: Servidor onde ficará hospedado e configurado Pentaho Data Integration Server, Kerberos e LFTP. Aplicação responsável pela execução dos processos (Jobs e Transformações) de carga de dados no LFTP e Scripts no Impala.

#### 4.2.4 Implantação – VM Buscador

Criação do usuário e pastas de trabalho:

1º – Como usuário root, criar o usuário “buscador” no sistema operacional.

```
# useradd -m -p "buscador@2022" buscador
```

2º – Criar diretório **/scripts**.

```
# mkdir /scripts
```

3º – Alterar dono (owner) **/scripts**.

```
# chown buscador:buscador /scripts
```

4º – Criar a partição **/dados**

- Esta pasta deve ser uma partição onde tanto o servidor Buscador tenha acesso de leitura e escrita, quanto o servidor Pentaho. Neste caso, os arquivos baixados pelo Buscador, estarão disponíveis para que os processos do Pentaho possam acessá-los.
- Esta pasta deve ter espaço configurado de pelo menos 500GB.

5º - Alterar dono (owner) **/dados**

```
# chown buscador:buscador /dados
```

Habilitar esse usuário para conexão SSH permitindo instalar bibliotecas Python.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.21/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

#### 4.2.4.1 **Habilitação de saída para a internet**

Configurar saída para internet sem filtro de conteúdo.

#### 4.2.4.2 **Instalação do Python e bibliotecas**

1º - Atualizar o gerenciador de pacotes do Centos 8.

```
$ sudo dnf update -y
```

2º - Instalar o Python 3.

```
$ sudo dnf install python3 -y
```

3º - Verificar se o Python foi instalado corretamente.

```
$ python3 -v
```

#### 4.2.5 **Implantação – VM Banco de dados**

Assume-se que já exista uma máquina, com acesso de administrador e o PostgreSQL versão 13 já instalado de forma padrão, denominada “VM Banco de Dados” neste documento.

##### 4.2.5.1 **Criar usuários administrativo**

Logar na VM Banco de Dados com o usuário “postgres” e executar os comandos abaixo. Nos campos <senha> criar a senha a ser utilizada para a conexão do Pentaho ao banco de dados:

```
$ createuser -s -i -d -r -l -w pentaho;
$ psql -c "alter role pentaho with password '<senha>';"
$ createuser user_pdi_log
$ psql -c "alter role user_pdi_log with password '<senha>';"
$ createuser owner_pdi_log
$ psql -c "alter role owner_pdi_log with password '<senha>';"
$ createdb --owner=owner_pdi_log pdi_log "Base para registro de log de execuções de rotinas ETL"
$ psql -c "create schema pdi_log" pdi_log
$ psql -c "grant connect on database pdi_log to user_pdi_log"
$ psql -c "grant usage on schema pdi_log to user_pdi_log" pdi_log
$ psql -c "grant insert,select on table pdi_log.tb_pdi_log to user_pdi_log" pdi_log
```

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.22/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

#### 4.2.5.2 Criar tabelas para receber os registros de Logs

As rotinas que serão executadas no servidor VM Datahub armazenará os Logs nesse banco de dados, para isso devemos criar as devidas tabelas conforme comandos abaixo.

Logar no banco de dados com o usuário owner\_pdi\_log e criar as tabelas conforme comandos abaixo.

```
CREATE TABLE pdi_log.tb_log_metrics(
ID_BATCH INTEGER
CHANNEL_ID VARCHAR(255)
, LOG_DATE TIMESTAMP
, METRICS_DATE TIMESTAMP
, METRICS_CODE VARCHAR(255)
, METRICS_DESCRIPTION VARCHAR(255)
, METRICS_SUBJECT VARCHAR(255)
, METRICS_TYPE VARCHAR(255)
, METRICS_VALUE BIGINT);
CREATE TABLE pdi_log.tb_log_step(
ID_BATCH INTEGER
, CHANNEL_ID VARCHAR(255)
, LOG_DATE TIMESTAMP
, TRANSNAME VARCHAR(255)
, STEPNAME VARCHAR(255)
, STEP_COPY SMALLINT
, LINES_READ BIGINT
, LINES_WRITTEN BIGINT
, LINES_UPDATED BIGINT
, LINES_INPUT BIGINT
, LINES_OUTPUT BIGINT
, LINES_REJECTED BIGINT
, ERRORS BIGINT);
CREATE INDEX IDX_tb_log_step_1 ON pdi_log.tb_log_step(TRANSNAME, LOG_DATE);
```

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.23/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

```

CREATE TABLE pdi_log.tb_log_channels(
ID_BATCH INTEGER
, CHANNEL_ID VARCHAR(255)
, LOG_DATE TIMESTAMP
, LOGGING_OBJECT_TYPE VARCHAR(255)
, OBJECT_NAME VARCHAR(255)
, OBJECT_COPY VARCHAR(255)
, REPOSITORY_DIRECTORY VARCHAR(255)
, FILENAME VARCHAR(255)
, OBJECT_ID VARCHAR(255)
, OBJECT_REVISION VARCHAR(255)
, PARENT_CHANNEL_ID VARCHAR(255)
, ROOT_CHANNEL_ID VARCHAR(255));

CREATE TABLE pdi_log.tb_log_perf(
ID_BATCH INTEGER
, SEQ_NR INTEGER
, LOGDATE TIMESTAMP
, TRANSNAME VARCHAR(255)
, STEPNAME VARCHAR(255)
, STEP_COPY INTEGER
, LINES_READ BIGINT
, LINES_WRITTEN BIGINT
, LINES_UPDATED BIGINT
, LINES_INPUT BIGINT
, LINES_OUTPUT BIGINT
, LINES_REJECTED BIGINT
, ERRORS BIGINT
, INPUT_BUFFER_ROWS BIGINT
, OUTPUT_BUFFER_ROWS BIGINT);

CREATE INDEX IDX_tb_log_perf_1 ON pdi_log.tb_log_perf(TRANSNAME, LOGDATE);

CREATE TABLE pdi_log.tb_log_trans(
ID_BATCH INTEGER
, CHANNEL_ID VARCHAR(255)
, TRANSNAME VARCHAR(255)

```

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.24/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.



```

, STATUS VARCHAR(15)
, LINES_READ BIGINT
, LINES_WRITTEN BIGINT
, LINES_UPDATED BIGINT
, LINES_INPUT BIGINT
, LINES_OUTPUT BIGINT
, LINES_REJECTED BIGINT
, ERRORS BIGINT
, STARTDATE TIMESTAMP
, ENDDATE TIMESTAMP
, LOGDATE TIMESTAMP
, DEPDATE TIMESTAMP
, REPLAYDATE TIMESTAMP
, LOG_FIELD TEXT);
CREATE INDEX IDX_tb_log_trans_1 ON pdi_log.tb_log_trans(ID_BATCH);
CREATE INDEX IDX_tb_log_trans_2 ON pdi_log.tb_log_trans(ERRORS, STATUS, TRANSNAME);
CREATE INDEX IDX_tb_log_trans_3 ON pdi_log.tb_log_trans(TRANSNAME, LOGDATE);
GRANT SELECT,INSERT,UPDATE,DELETE ON ALL TABLES IN SCHEMA pdi_log TO user_pdi_log;

```

## 4.2.6 Implantação – VM Banco de dados

### 4.2.6.1 **Habilitação de acesso ao servidor Pentaho pelo SSH**

Sobre a versão do sistema operacional Linux disponibilizado no servidor do Pentaho (Oracle Linux 9), foi identificada incompatibilidade de conectividade SSH implementado dentro do Pentaho Data Integration (PDI) e o servidor SFTP do SERPRO.

Observações: A implementação do ssh no Pentaho está baseada na biblioteca “trilead-ssh2”. Mesmo em sua última versão, esta biblioteca ainda requer algoritmos que foram descontinuados no RedHat 9 e seus derivados. Desta forma, os passos acima tiveram que ser aplicados no sistema operacional para permitir a conexão. O diretório **/aplicacao** e todos os arquivos e subdiretórios abaixo dele, deverão ter como dono e grupo o Pentaho para que ele possa:

Para resolver este problema, foram tomadas as seguintes ações:

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.25/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

### 1º - Cópia do arquivo original de configuração do SSH.

```
# cp /etc/crypto-policies/back-ends/opensshserver.config /etc/crypto-policies/back-ends/opensshserver.config.old
```

2º - Alterar o arquivo `/etc/crypto-policies/back-ends/opensshserver.config` adicionando o conteúdo abaixo:

```
Ciphers aes256-gcm@openssh.com,chacha20-poly1305@openssh.com,aes256-ctr,aes128-gcm@openssh.com,aes128-ctr
MACs hmac-sha2-256-etm@openssh.com,hmac-sha1-etm@openssh.com,umac-128-etm@openssh.com,hmac-sha2-512-etm@openssh.com,hmac-sha2-256,hmac-sha1,umac-128@openssh.com,hmac-sha2-512
GSSAPIKexAlgorithms gss-curve25519-sha256-,gss-nistp256-sha256-,gss-group14-sha256-,gss-group16-sha512-,gss-gex-sha1-,gss-group14-sha1-
KexAlgorithms curve25519-sha256,curve25519-sha256@libssh.org,ecdh-sha2-nistp256,ecdh-sha2-nistp384,ecdh-sha2-nistp521,diffie-hellman-group-exchange-sha256,diffie-hellman-group14-sha256,diffie-hellman-group16-sha512,diffie-hellman-group18-sha512,diffie-hellman-group-exchange-sha1,diffie-hellman-group14-sha1
HostKeyAlgorithms ecdsa-sha2-nistp256,ecdsa-sha2-nistp256-cert-v01@openssh.com,sk-ecdsa-sha2-nistp256@openssh.com,sk-ecdsa-sha2-nistp256-cert-v01@openssh.com,ecdsa-sha2-nistp384,ecdsa-sha2-nistp384-cert-v01@openssh.com,ecdsa-sha2-nistp521,ecdsa-sha2-nistp521-cert-v01@openssh.com,ssh-ed25519,ssh-ed25519-cert-v01@openssh.com,sk-ssh-ed25519@openssh.com,sk-ssh-ed25519-cert-v01@openssh.com,rsa-sha2-256,rsa-sha2-256-cert-v01@openssh.com,rsa-sha2-512,rsa-sha2-512-cert-v01@openssh.com,ssh-rsa,ssh-rsa-cert-v01@openssh.com
PubkeyAcceptedAlgorithms ecdsa-sha2-nistp256,ecdsa-sha2-nistp256-cert-v01@openssh.com,sk-ecdsa-sha2-nistp256@openssh.com,sk-ecdsa-sha2-nistp256-cert-v01@openssh.com,ecdsa-sha2-nistp384,ecdsa-sha2-nistp384-cert-v01@openssh.com,ecdsa-sha2-nistp521,ecdsa-sha2-nistp521-cert-v01@openssh.com,ssh-ed25519,ssh-ed25519-cert-v01@openssh.com,sk-ssh-ed25519@openssh.com,sk-ssh-ed25519-cert-v01@openssh.com,rsa-sha2-256,rsa-sha2-256-cert-v01@openssh.com,rsa-sha2-512,rsa-sha2-512-cert-v01@openssh.com,ssh-rsa,ssh-rsa-cert-v01@openssh.com
CASignatureAlgorithms ecdsa-sha2-nistp256,sk-ecdsa-sha2-nistp256@openssh.com,ecdsa-sha2-nistp384,ecdsa-sha2-nistp521,ssh-ed25519,sk-ssh-ed25519@openssh.com,rsa-sha2-256,rsa-sha2-512,ssh-rsa
```

### 3º - Atualização das políticas configuradas.

```
# update-crypto-policies --set DEFAULT:SHA1
```

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.26/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

#### 4.2.6.2 Acesso do usuário

Executar o seguinte comando como root.

1º - Criar o usuário “pentaho” no sistema operacional.

```
# useradd -m -p "pentaho@2022" pentaho
```

2º - Criar diretório /aplicacao.

```
# mkdir /aplicacao
```

**Observação:** Esta pasta só deve ser criada caso a mesma não tenha sido migrada ou copiada de outro servidor. Pois esta pasta contém toda a parte de processamento de arquivos pelo ETL (Pentaho Data Integration – PDI).

3º - Alterar owner /aplicação.

```
# chown -R pentaho:pentaho /aplicacao
```

**Observação:** A alteração do dono (owner) da pasta aplicação para o usuário “pentaho” se faz necessária para que este usuário, conectado local ou remotamente, consiga executar as rotinas de extração, transformação e carga, bem como manipular eventuais arquivos de dados.

#### 4.2.6.3 Arquivos de configuração

Para o armazenamento dos arquivos de configuração crie o diretório /config dentro do diretório principal da aplicação:

```
$ mkdir /aplicação/config
```

É necessário que os arquivos de keytab (s\_pgfn\_etl.keytab) e de configuração do Kerberos 5 (krb5.conf) enviados pelo SERPRO estejam no servidor da PGFN para que a conexão entre os servidores seja estabelecida da maneira correta (Consulte o responsável para adquirir os arquivos). Os dois arquivos (s\_pgfn\_etl.keytab e krb5.conf) devem ser salvos em /aplicação/config/. É necessário ainda que os IPs, DNS e portas estejam liberados no servidor da PGFN.

DNS/IP		Porta	
Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.27/45

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

dfcdsrv1280.srv.datalake.serpro	88 e 749
dfcdsrv1279.srv.datalake.serpro	88 e 749
dfcdsrvv4694.srv.cd.serpro	88 e 749
dfcdsrvv4564.srv.cd.serpro	88 e 749
spcdsrv1034.srv.datalake.serpro	88 e 749
spcdsrv1035.srv.datalake.serpro	88 e 749
10.30.143.32	10021

#### 4.2.6.4 Ferramentas

Para que tudo funcione como o esperado, precisamos instalar 3 ferramentas: LFTP, Kerberos e Pentaho Data Integration. Primeiro, instalaremos o LFTP através dos comandos listados abaixo.

##### 4.2.6.4.1 LFTP

1º - Atualizar os repositórios do SO.

```
$ sudo apt update
```

2º - Instalar a ferramenta LFTP.

```
$ sudo yum install lftp
```

3º - Verificar se o arquivo de keytab (**s\_pgnf\_eti.keytab**) está no servidor na pasta (**/aplicacao/config/**). Se sim, ele será mostrado ao executar o comando.

```
$ ls /aplicacao/config/
```

4º - Inserir na última linha do arquivo de configuração do LFTP (**/etc/lftp.conf**), caso não exista, o comando abaixo para a não verificação de certificado a cada acesso.

```
$ set ssl:verify-certificate no
```

5º - Criar um arquivo de teste para inserir no SFTP.

```
$ vi teste_HDFS.txt
```

6º - Inserir um texto.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.28/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

```
id nome
1 Jose
2 Maria
```

7º - Salvar o arquivo e sair.

8º - Armazenar a chave descryptografada em **base64** em uma variável.

```
$ export LFTP_PASSWORD="keytab:$(base64 -w 0 /aplicacao/config/s_pgfn_etl.keytab)"
```

9º - Acessar o servidor SFTP do Serpro (**10.30.143.32**) através do comando.

```
$ lftp --env-password -p 10021 -u s_pgfn_etl ftps://10.30.143.32
```

10º - Inserir o arquivo de teste criado no servidor.

```
$ put teste_HDFS.txt
```

11º - Verificar se o arquivo foi criado e seus devidos dados.

```
$ cat teste_HDFS.txt/teste_HDFS.txt
```

12º - Remover o arquivo de teste.

```
$ rm -rf teste_HDFS.txt/
```

#### 4.2.6.4.2 Kerberos

Uma vez que a conexão com o servidor SFTP está sendo estabelecida corretamente, deveremos instalar a ferramenta Kerberos 5 para que seja possível a utilização do ticket de autenticação. Conforme os passos abaixo.

1º - Instalar a ferramenta Kerberos.

```
$ sudo yum install krb5-workstation
```

2º - Fazer uma cópia de backup do arquivo de configuração do Kerberos.

```
$ sudo cp /etc/krb5.conf /etc/krb5.conf.old
```

3º - Substituir o arquivo de configuração do Kerberos pelo arquivo enviado pelo SERPRO (**krb5.conf**).

```
$ sudo cp /aplicacao/config/krb5.conf /etc/krb5.conf
```

4º - Iniciar o ticket Kerberos através do comando.

```
$ kinit -k -t /aplicacao/config/s_pgfn_etl.keytab s_pgfn_etl@DATA LAKE.SERPRO
```

5º - Verificar se o ticket Kerberos foi iniciado.

```
$ klist
```

6º - Ao finalizar o teste, invalidar o ticket.

```
$ kdestroy
```

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.29/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

#### 4.2.6.4.3 Pentaho Data Integration

Após a instalação do Kerberos 5, podemos instalar a ferramenta para os processos ETL e para a movimentação dos dados e inserção no Data Lake. A ferramenta escolhida é o Pentaho Data Integration na versão e para a instalação deverão ser executados os comandos abaixo.

1º - Instalação dos pacotes obrigatórios.

```
$ sudo yum install -y gtk3 redhat-lsb-core zip unzip wget dbus-x11 xauth webkitgtk3 mesa-dri-drivers
```

2º - Instalação do Java (JDK 8).

```
$ sudo yum install java
```

3º - Entrar na pasta onde ficará o PDI.

```
$ cd /aplicacao/
```

4º - Fazer o download do arquivo através do site do sourceforge em sua versão mais recente (<https://sourceforge.net/projects/pentaho/>). Usar o comando wget para recuperar o arquivo, ajustando para a versão mais recente, como abaixo:

```
$ wget https://sinalbr.dl.sourceforge.net/project/pentaho/Pentaho-9.3/client-tools/pdi-ce-9.3.0.0-428.zip
```

5º - Descompactar o arquivo.

```
$ unzip pdi-ce-9.3.0.0-428.zip
```

6º - Executar uma transformação de exemplo para verificar se o PDI está funcionando.

```
$/aplicacao/data-integration/pan.sh -file="/aplicacao/data-integration/samples/transformations/Memory Group By - simple example.ktr"
```

7º - Criar link para arquivo de propriedades.

```
$ touch ~/.kettle/kettle.properties
```

```
$ ln -s ~/.kettle/kettle.properties /aplicacao/data-integration
```

8º - Criar pastas auxiliares.

```
$ mkdir /aplicacao/jobs
```

```
$ mkdir /aplicacao/arquivos_a_processar
```

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.30/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

```
$ mkdir /aplicacao/arquivos_processados
```

9º - Configurar arquivo de propriedades. Adicionar as linhas abaixo no arquivo: kettle.properties (O arquivo kettle fica no home do usuário). No arquivo kettle abaixo, **<IP do servidor do banco de dados>** deve ser preenchido com o IP do ambiente em que o servidor do banco de dados está sendo executado, e **<senha>** com o valor da senha definida para o banco de dados.

```
# Banco de Dados
DB_LOG_HOST=<IP do servidor do banco de dados>
DB_LOG_PORT=5432
DB_LOG_DATABASE=pdi_log
DB_LOG_SCHEMA=pdi_log
DB_LOG_USER=user_pdi_log
DB_LOG_PASS=<senha>
CAMINHO_A_PROCESSAR=/aplicacao/arquivos_a_processar
# CAMINHOS PADRÃO
PATH_ETL=/aplicacao/etl
```

10º - Editar o arquivo **/aplicacao/data-integration/spoon.sh**, conforme indicações abaixo:

Alterar a variável abaixo:

```
De:
# Fix GTK 3 issues with SWT
export SWT_GTK3=0
Para:
# Fix GTK 3 issues with SWT
export SWT_GTK3=1
```

#### 4.2.7 Implantação – Orquestrador - Rundeck

Executar o seguinte comando como root.

1º - Criar o usuário “rundeck” no sistema operacional.

```
# useradd -m -p "rundeck@2022" rundeck
```

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.31/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN. É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

2º - Criar diretório **/scripts**.

```
# mkdir /scripts
```

3º - Alterar dono (owner) **/scripts**.

```
# chown rundeck:rundeck /scripts
```

Executar o seguinte comando como “rundeck”.

4º - Criar chave ssh.

```
# su - rundeck  
$ ssh-keygen -t rsa
```

**Observação:** Pressionar o ENTER para todas as perguntas emitidas pelo comando.

5º - Copiar a chave pública gerada para o servidor do Pentaho.

```
$ ssh-copy-id pentaho@<ip-servidor-pentaho>
```

**Observação:** Na primeira conexão é solicitada uma confirmação (Aceitar com “yes”). Em seguida, será solicitada a senha do usuário “pentaho” no servidor do Pentaho. Deve-se substituir <ip-servidor-pentaho> pelo IP do servidor Pentaho.

6º - Copiar a chave pública gerada para o servidor do Buscador.

```
$ ssh-copy-id buscador@<ip-servidor-buscador>
```

#### **Observações:**

- a) Na primeira conexão é solicitada uma confirmação (Aceitar com “yes”). Em seguida, será solicitada a senha do usuário “buscador” no servidor do Buscador.
- b) Deve-se substituir <ip-servidor-buscador> pelo IP do servidor Buscador.

7º - Configuração de Firewall para que o servidor Orquestrador consiga conectar ao servidor de banco de dados PostgreSQL (via porta do banco de dados, geralmente a 5432), onde se encontra o repositório do **Rundeck**.

8º - Baixar a imagem do “Rundeck”.

```
$ docker pull rundeck/rundeck:4.7.0
```

9º - Desabilitar o SELinux alterando o arquivo **/etc/selinux/config**:

```
SELINUX=disabled
```

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.32/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.



10º - Executar os comandos abaixo caso o Firewall esteja ativado:

```
# sudo iptables -t filter -F
# sudo iptables -t filter -X
# systemctl restart docker
# setenforce 0
```

11º - Executar o comando de start do container. Atenção aos campos parametrizados, que devem ser os utilizados no servidor Docker e na criação do BD PostgreSQL.

```
$ docker run -p <Ip docker>:4440:4440 -e RUNDECK_GRAILS_URL=http://<Ip docker>:4440 -e
RUNDECK_DATABASE_URL=jdbc:postgresql://<Ip PostgreSQL>:5432/rundeck -e
RUNDECK_DATABASE_DRIVER=org.postgresql.Driver -e
RUNDECK_DATABASE_USERNAME=<username> -e
RUNDECK_DATABASE_PASSWORD=<password> -e TZ=America/Sao_Paulo --name rundeck -t
rundeck/rundeck:4.7.0 &
```

12º - Conectar no container do rundeck e executar os comandos listados:

```
# docker exec -it rundeck bash
$ ssh-keygen -t ecdsa -m PEM
$ ssh-copy-id <ip-servidor-rundeck>
$ cat .ssh/id_ecdsa
$ exit
```

**<ip-servidor-rundeck> = Endereço IP do servidor rundeck.**

13º - Copiar e guardar o conteúdo da chave privada do rundeck mostrado no penúltimo comando executado acima.

14º - Criar um arquivo **/tmp/resources.xml** com o seguinte conteúdo:

```
<?xml version="1.0" encoding="UTF-8"?>
<project>
  <node name="rundeck" description="Servidor Rundeck" tags="NO1" hostname="<ip-servidor-
rundeck>" osArch="amd64" osFamily="unix" osName="Linux" ssh-key-storage-
path="keys/rundeck" ssh-authentication="privateKey" username="rundeck"/>
</project>
```

**<ip-servidor-rundeck> = Endereço IP do servidor rundeck.**

15º - Copiar o arquivo criado para dentro do container com o comando:

```
# docker cp /tmp/resources.xml rundeck:/home/rundeck/server/data/
```

16º - Conectar na interface do rundeck e logar como usuário "admin", conforme a **figura 6**:

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.33/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

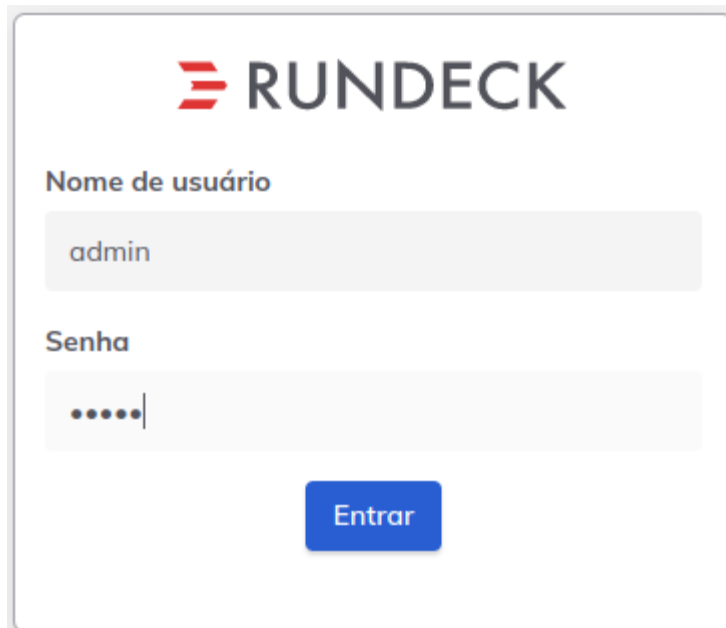


Figura 6 – Tela de login do “Rundeck”

Fonte: Elaborada pelo autor

17º - Clicar na engrenagem no canto superior direito e escolher "Key Storage" no menu, se seguir os passos:

- Clicar em "+ Add or Upload a Key"
- No "Key Type:" manter "Private Key"
- Em "Enter text:", cole a chave privada gerada anteriormente.

**Observação, a chave deve ser copiada inclusive com o texto inicial “-----BEGIN RSA PRIVATE KEY-----” e final “-----END RSA PRIVATE KEY-----”.**

- Em "Name:" digite "rundeck".
- Clique em "Save" para salvar a chave.

18º - Clique em "Projects" (**figura 7**) no combo no alto da tela e selecione "Create Project":

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.34/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

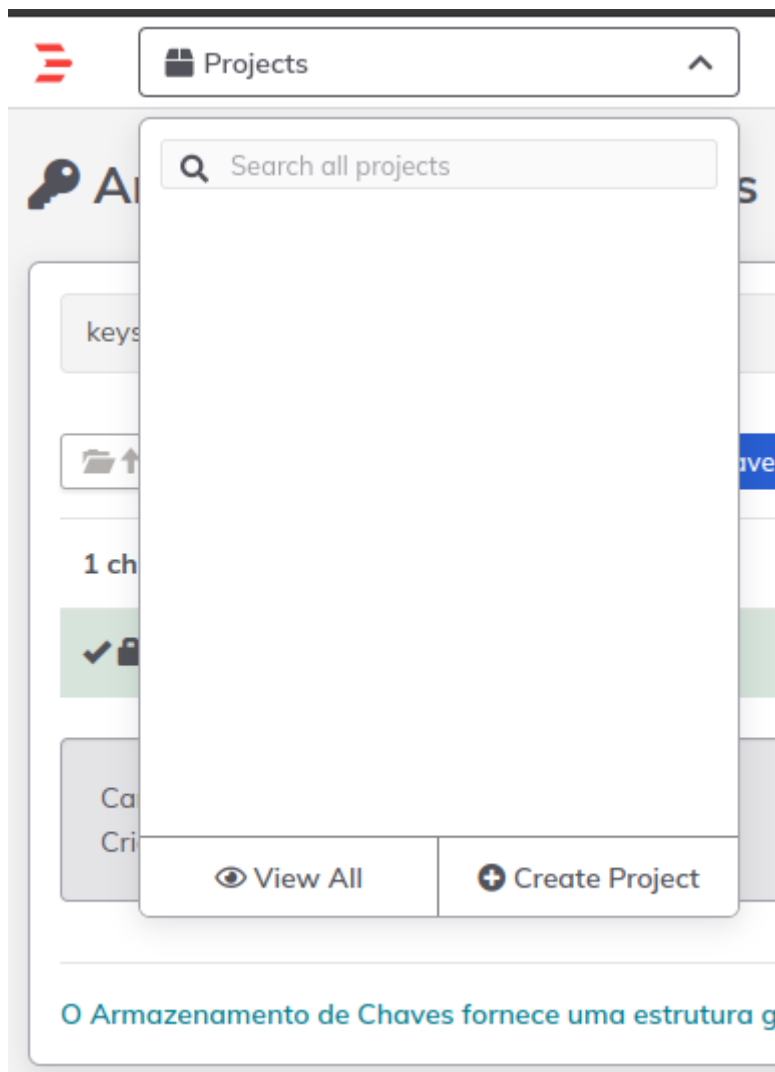


Figura 7 – Tela de projetos do “Rundeck”

Fonte: Elaborada pelo autor

19º - Na aba “Details”, dar um nome ao projeto: "PGFN". Clique em "Criar" para criar o projeto, como se pode ver na **figura 8**:

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.35/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN. É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

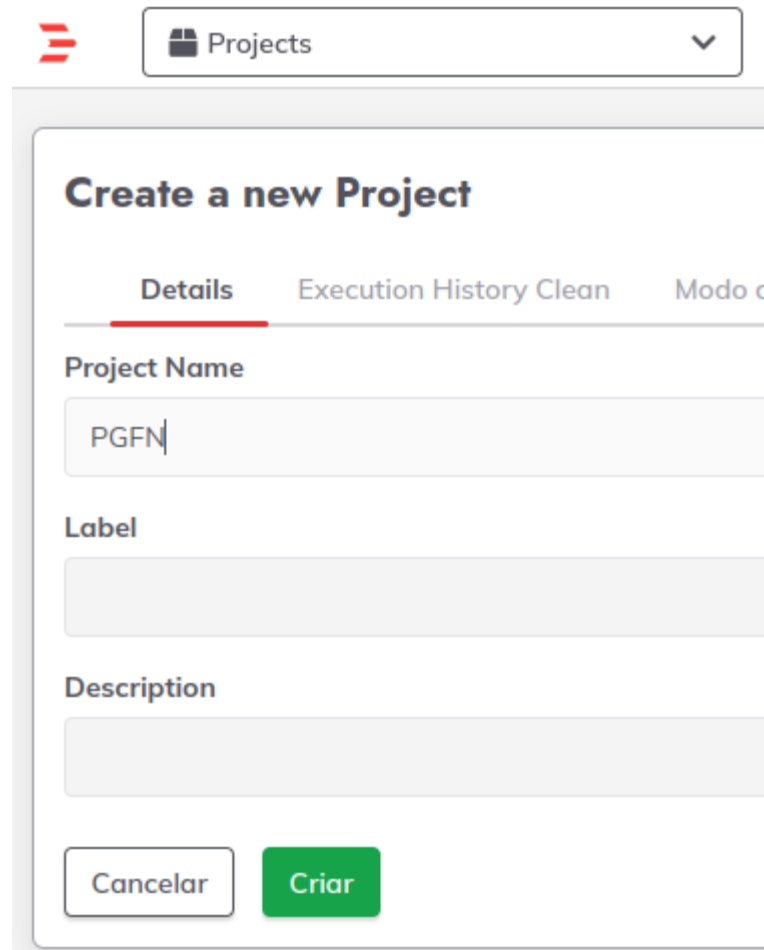


Figura 8 – Criando novos projetos

Fonte: Elaborada pelo autor

20º - Após criar o projeto, na aba “Sources” há um source chamado “Local”. Remova esse source clicando em “Delete” (Botão vermelho na **figura 9**).



Figura 9 – Deletando source padrão

Fonte: Elaborada pelo autor

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.36/45
---------------	---------------------	---	-----------

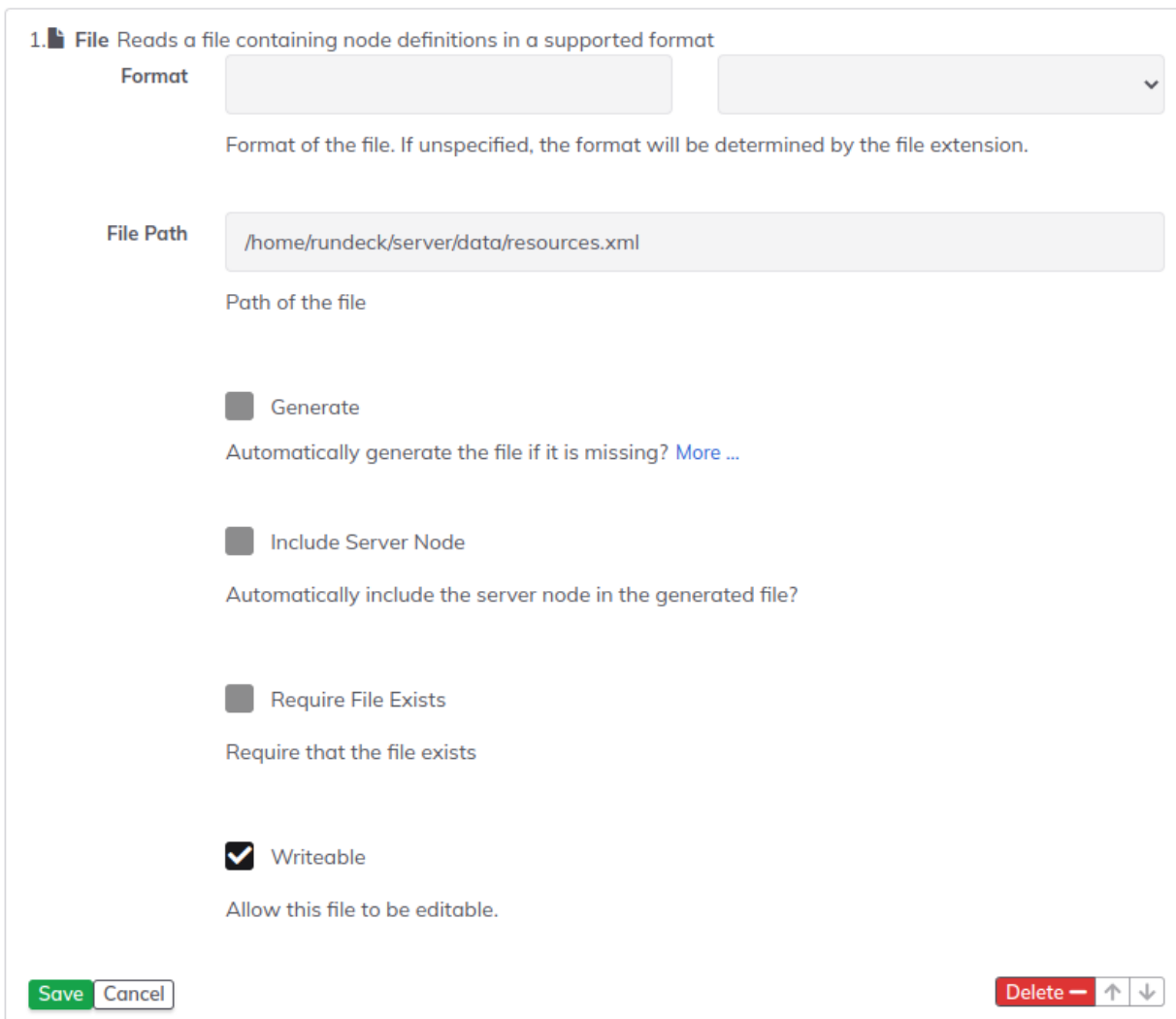
Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN. É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

Clique em “Save” para salvar a remoção do source.

21º - Clicar em “Add a new Node Source +” e escolha a opção “File”. Em seguida:

- Em “File Path”, preencher com “/home/rundeck/server/data/resources.xml” (sem aspas)
- Marque a opção “Writeable”.
- Clique em “Save” para salvar o resource.
- Clique em “Save” novamente para salvar o Nó (node), conforme a **figura 10**.



1. **File** Reads a file containing node definitions in a supported format

Format

Format of the file. If unspecified, the format will be determined by the file extension.

File Path

Path of the file

Generate  
Automatically generate the file if it is missing? [More ...](#)

Include Server Node  
Automatically include the server node in the generated file?

Require File Exists  
Require that the file exists

Writeable  
Allow this file to be editable.

Figura 10 – Criando novo source

Fonte: Elaborada pelo autor

22º - Para testar, clique no ícone “Comando” no menu à esquerda, conforme a **figura 11**.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.37/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.



Figura 11 – Ícone “comandos”

Fonte: Elaborada pelo autor

24º - No triângulo, ao lado de “Nós”, clique nele e depois em “\* Mostrar todos os nós”, conforme a **figura 12**.



Figura 12 – Selecionando os “nós”

Fonte: Elaborada pelo autor

25º - Verifique apenas um nó chamado “rundeck” foi listado, conforme a **figura 13**:

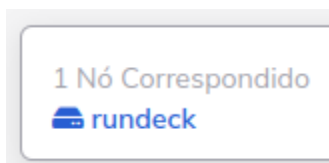


Figura 13 – Nó correspondente

Fonte: Elaborada pelo autor

26º - No campo “Digite um comando”, digite “hostname” (sem aspas), conforme a **figura 14**:

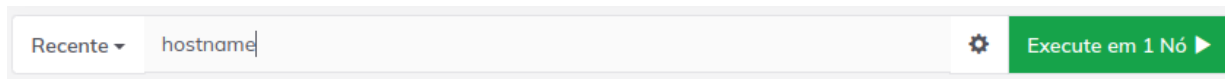


Figura 14 – Exemplo de comando

Fonte: Elaborada pelo autor

27º - Clique “Executar em 1 nó”. Em seguida, verifique se como resultado, foi mostrado o nome do host onde é executado o Docker do rundeck.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.38/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

## 4.2.8 Teste de conexão

As instruções abaixo devem ser executadas no servidor do Pentaho.

Com as ferramentas instaladas e funcionando corretamente, podemos testar as conexões entre o Pentaho Data Integration e os serviços do FTPS e do Impala. Para tanto, temos um pré-requisito que é ter os seguintes arquivos no caminho "/home/pentaho":

- teste\_inserir\_arquivo\_ftps\_SERPRO.ktr
- teste\_conexao\_impala\_SEPRO.ktr
- ImpalaJDBC41.jar

### 4.2.8.1 Teste de conectividade Pentaho PDI para FTPS

Primeiro, vamos testar a conexão do Pentaho Data Integration para o FTPS com os seguintes passos:

1º - Criar um arquivo de teste para ser inserido no servidor.

```
$ echo -e 'id nome\n1 Jose\n2 Maria' > /home/pentaho/teste2.txt
```

2º - Mover o arquivo para a pasta de jobs.

```
$ sudo mv /home/pentaho/teste2.txt /aplicacao/jobs/teste2.txt
```

3º - Executar a transformação de teste.

```
$/aplicacao/data-integration/pan.sh -file="/home/pentaho/teste_inserir_arquivo_ftps_SERPRO.ktr"
```

4º - Armazenar a keytab descriptografada em uma variável.

```
$ export LFTP_PASSWORD="keytab:$(base64 -w 0 /aplicacao/config/s_pgfn_etl.keytab)"
```

5º - Entrar no SFTP.

```
$ lftp --env-password -p 10021 -u s_pgfn_etl ftps://10.30.143.32
```

6º - Verificar se o arquivo foi criado dentro do SFTP.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.39/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

```
$ cat teste2.txt/teste2.txt
```

7º - Excluir o arquivo de teste.

```
$ rm -rf teste2.txt/
```

#### 4.2.8.2 Teste de conectividade Pentaho PDI para Impala

Após os testes de comunicação com o servidor FTPS, podemos testar a comunicação com o Impala através do Pentaho Data Integration. Para isso, seguimos os seguintes passos:

1º - Mover a biblioteca "ImpalaJDBC41.jar" para a pasta de libs do PDI.

```
$ mv /home/pentaho/ImpalaJDBC41.jar /aplicacao/data-integration/lib/ImpalaJDBC41.jar
```

2º - Gerar o ticket Kerberos.

```
$ kinit -c /tmp/krb5cc_0 -k -t /aplicacao/config/s_pgfn_etl.keytab s_pgfn_etl@DATA LAKE.SERPRO
```

3º - Verificar se o ticket foi gerado corretamente.

```
$ klist
```

4º - Executar a transformação.

```
$ sudo /aplicacao/data-integration/pan.sh -file="/home/pentaho/teste_conexao_impala_SEPRO.ktr"
```

5º - Verificar se o arquivo foi criado.

```
$ cat /home/pentaho/arquivo_gerado_teste.txt
```

6º - Remover o arquivo de teste.

```
$ rm arquivo_gerado_teste.txt
```

#### 4.2.8.3 Teste de conectividade SSH entre Orquestrador (rundeck) e Pentaho

1º - Logar com o usuário "rundeck" no servidor Orquestrador e executar o comando:

```
$ ssh pentaho@<ip-servidor-pentho>
```

**Observação:** Substituir <ip-servidor-pentaho> pelo IP do servidor Pentaho.

O comando acima deve conseguir conectar como usuário "pentaho" no servidor "Pentaho" se requerer senha.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.40/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.



#### 4.2.8.4 Teste de conectividade SSH entre Orquestrador (rundeck) e Buscador

1º - Logar com o usuário “rundeck” no servidor Orquestrador e executar o comando:

```
$ ssh buscador @<ip-servidor-buscador>
```

**Observação:** Substituir <ip-servidor-buscador> pelo IP do servidor Buscador.

O comando acima deve conseguir conectar como usuário “buscador” no servidor “Buscador” sem requerer senha.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.41/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

## 5 CONCLUSÃO

Através de um trabalho coordenado e interdependente entre as equipes da Procuradoria-Geral da Fazenda Nacional (PGFN) e da Universidade de Brasília (UnB), as atividades de elaboração deste RT foram planejadas, discutidas, executadas e documentadas.

A conclusão do RT é importante pois cumpre a etapa básica de inicialização, prevista na metodologia de gestão do projeto. O plano define a estratégia inicial adotada, servindo de referencial para adaptações dessa estratégia caso a monitoração do projeto assim o determine.

Além de conter as decisões iniciais da gestão do projeto, o presente relatório servirá então, durante as demais etapas, para registrar as alternativas e embasar a tomada de decisão para prosseguimento do projeto.

As atividades envolvidas nesta etapa observaram formalmente a execução dos passos da metodologia elencada para gestão do projeto, conforme definido pelo PMI em seu guia PMBOK.

A equipe da UnB considera que teve acesso a todas as informações necessárias à boa condução dos trabalhos e que a disponibilização dessas informações pela equipe do MDSA, assim como as atividades conjuntas de análise e discussão levaram à etapa do projeto a bom termo.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.42/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	<b>Pág.43/45</b>
---------------	---------------------	---	------------------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

## REFERÊNCIAS

APACHE HADOOP. 2006-2022. Disponível em: <<https://hadoop.apache.org/>>. Acesso em: 21 jan. 2023.

PENTAHO DATA INTEGRATION. Pentaho Data Integration. 2023. Disponível em: <[https://help.hitachivantara.com/Documentation/Pentaho/8.3/Products/Pentaho\\_Data\\_Integration](https://help.hitachivantara.com/Documentation/Pentaho/8.3/Products/Pentaho_Data_Integration)>. Acesso em: 21 jan. 2023.

PMBOK. 2023. Disponível em: <https://www.pmi.org/pmbok-guide-standards/foundational/pmbok>

RUNDECK. 2023. Disponível em: <https://www.rundeck.com/>

Projeto: PGFN	Emissão: 09/02/2023	Arquivo: 20230209 RT de Artefato da Prototipação de módulos de mobilidade, adaptação ao contexto, interoperação de serviços e dados - Parcial	Pág.44/45
---------------	---------------------	---	-----------

Confidencial.

Este documento foi elaborado pela Universidade de Brasília (UnB) para a PGFN.  
É vedada a cópia e a distribuição deste documento ou de suas partes sem o consentimento, por escrito, da PGFN.

Universidade de Brasília – UnB

FUNAPE - Fundação de Apoio à Pesquisa - UFG

Laboratório de Tecnologias da Tomada de Decisão – LATITUDE

[www.unb.br](http://www.unb.br) – <https://funape.org.br> – [www.latitude.unb.br](http://www.latitude.unb.br)

