

# 52<sup>ª</sup> COGEF

Comissão de Gestão Fazendária





# Ciência de Dados na SEF/MG

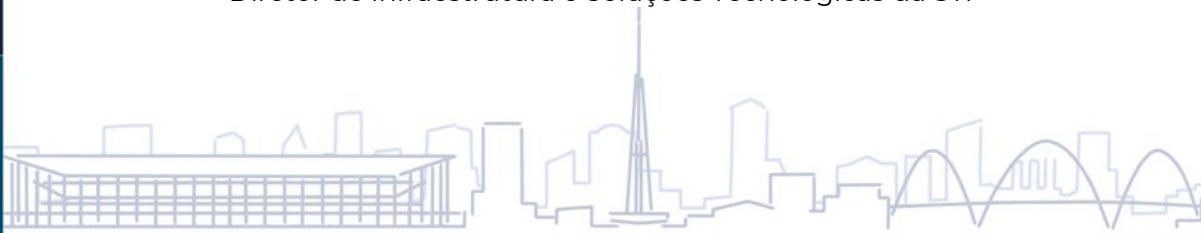
---

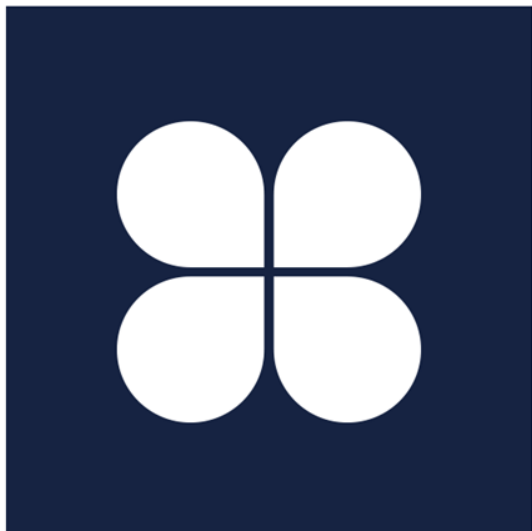
## **Bruno Meira Tenório D’Albuquerque**

Auditor Fiscal da Receita Estadual  
Gerente Núcleo de Soluções Tecnológicas da STI

## **Rogério Zupo Braga**

Auditor Fiscal da Receita Estadual  
Diretor de Infraestrutura e Soluções Tecnológicas da STI





## **Agenda**

- 1. Ambiente Analítico da SEF**
- 2. Arquitetura Lambda e Data Lake**
- 3. Papéis e fluxo de trabalho**
- 4. Projetos**

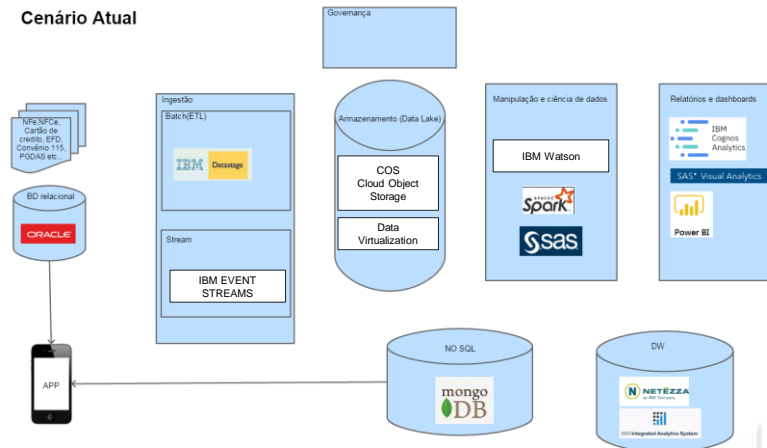




## **Ambiente Analítico da SEF**



### Cenário Atual



### Cenário Atual

- Solução IBM Datastage para os processos de ETL e integração de dados;
- Plataforma SAS para manipulação de dados e algumas funcionalidades de ciência de dados;
- IBM Cognos Analytics, Microsoft Power BI e o SAS Visual Analytics para visualização de dados;
- IBM/Netezza e o IBM/IAS para Data Warehouse.



# Inteligência Analítica



Infraestrutura



Armazenamento



Engenharia de Dados



Governança de Dados



Análise de Dados



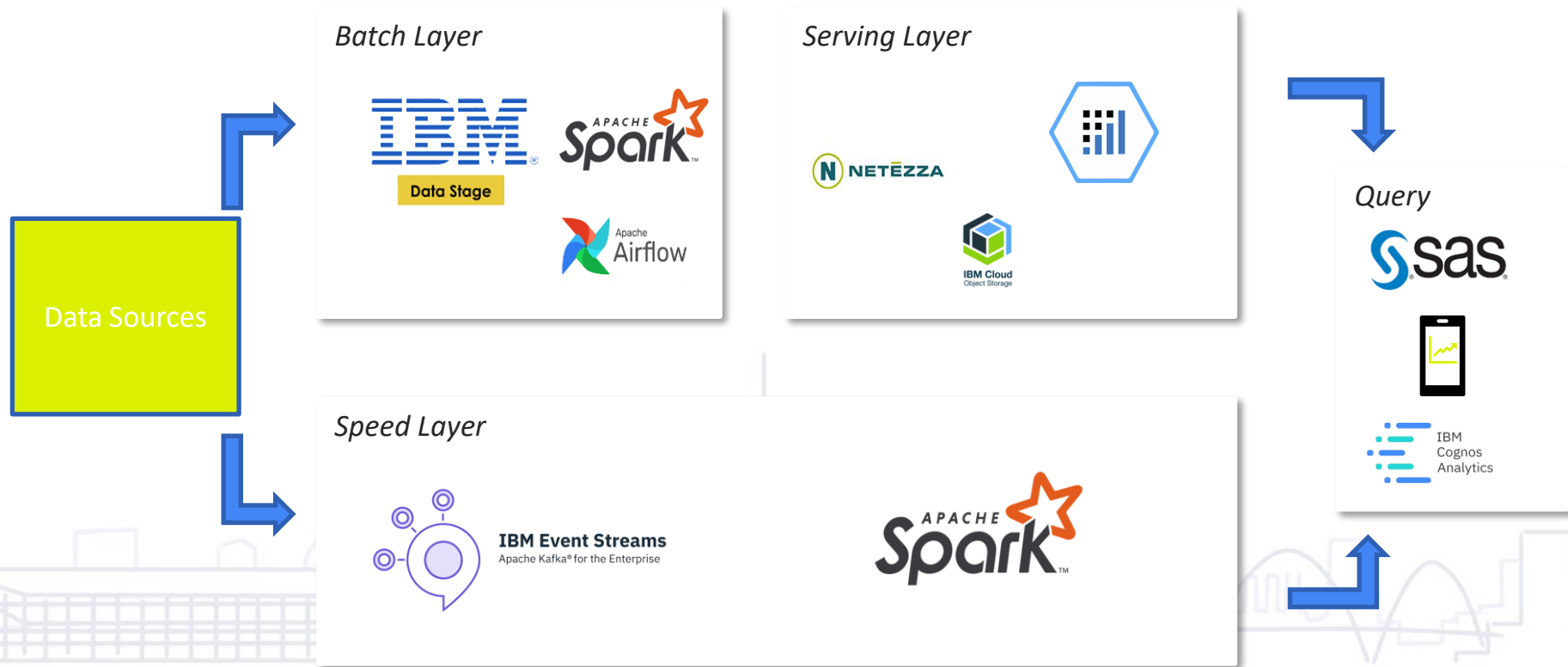
Ciência de Dados



## Arquitetura Lambda e Data Lake

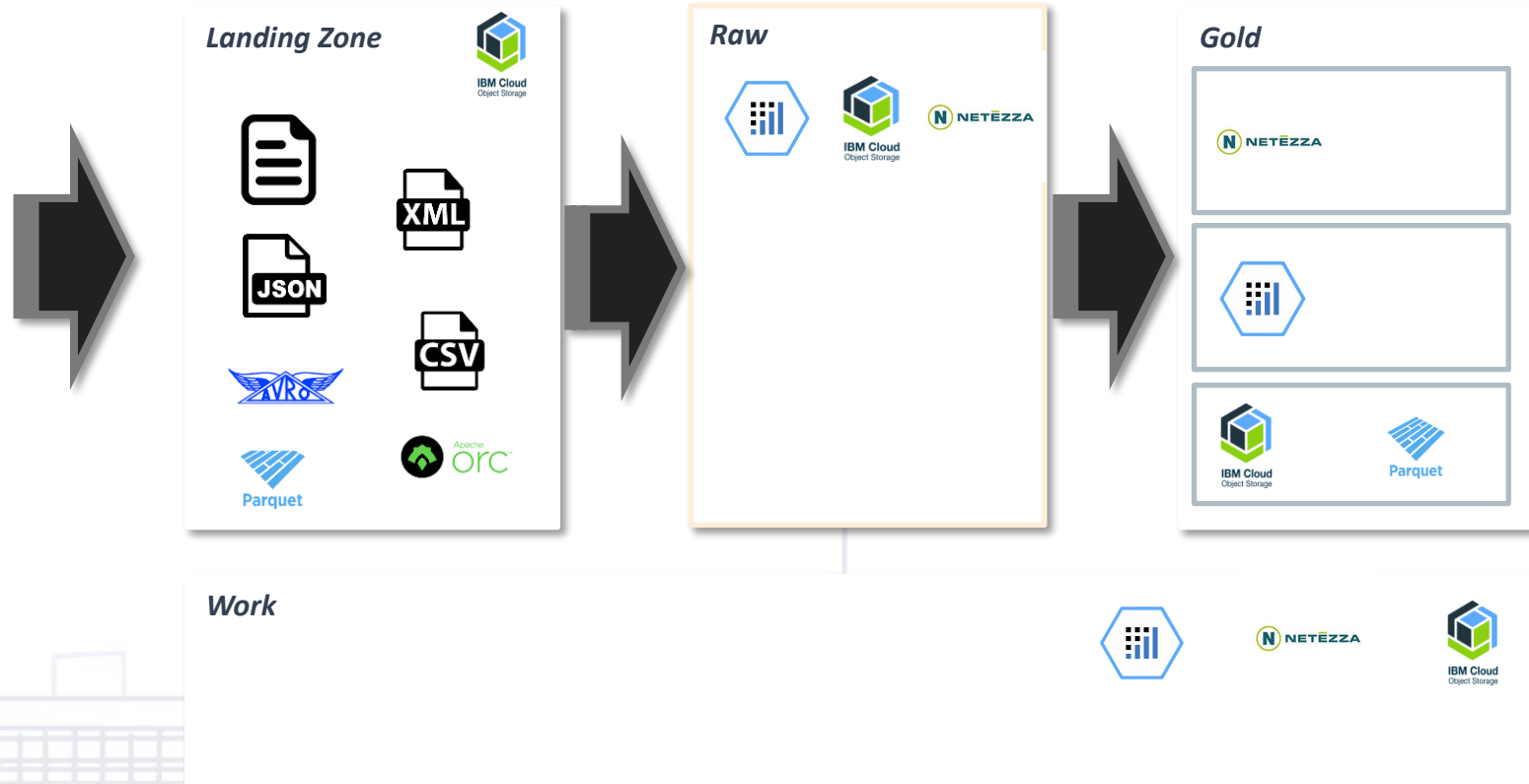


# Arquitetura Lambda





# Data Lake



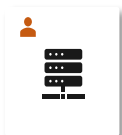


## Papéis e fluxo de trabalho



# Papéis

DIA/STI



## Administrador de Ambiente

- Administração dos Servidores
- Administração das Ferramentas
- DBA
- DevOps



## Engenheiro de Dados

- Arquitetura, processos e padrões
- Ingestão de Dados
- Operação
- Integração de dados – Analítico/Transacional

Negócio



## Desenvolvedor ETL

- Desenvolvimento de DW
- Criação de SAS Lib



## Analista de Dados/Negócio

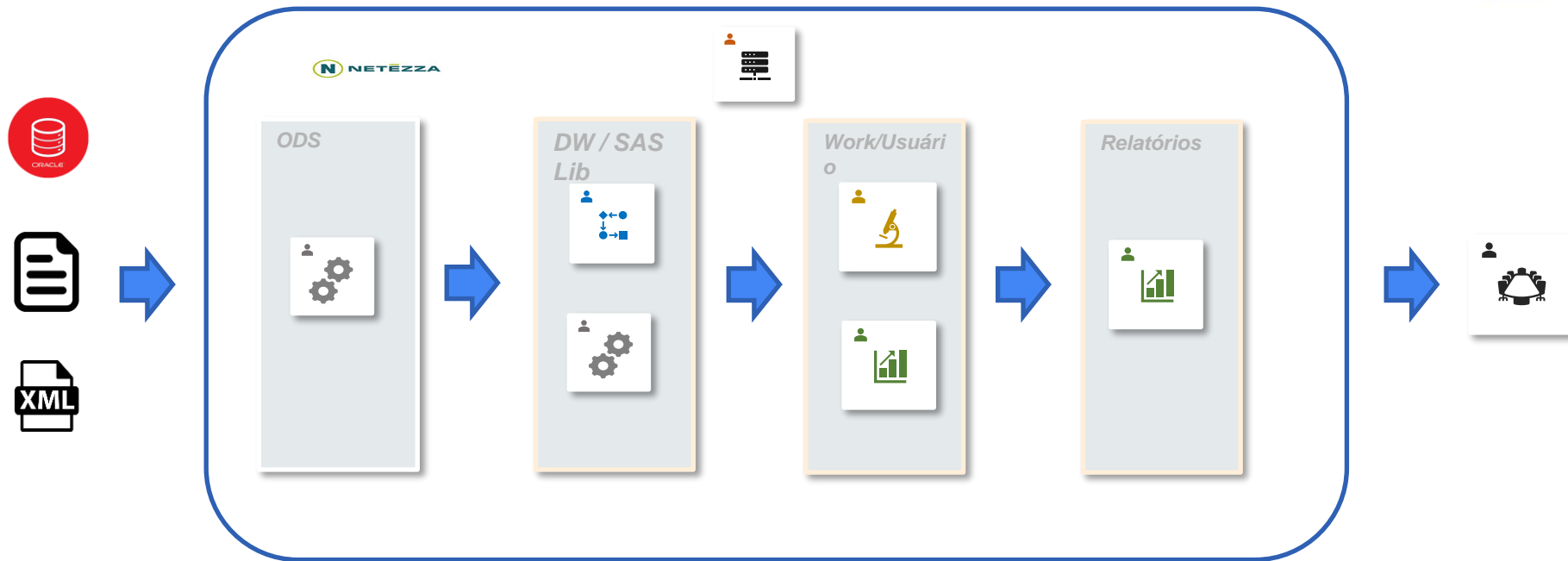
- Desenvolvimento de Relatórios
- Análises Ad/Hoc
- Fluxos/Malhas



## Cientista de Dados

- Análises Estatísticas
- Desenvolvimento de modelos
  - *Data Mining e Machine Learning*

# Fluxo de Trabalho



Engenheiro de Dados



Analista de Dados/Negócio



Administrador de Ambiente



Desenvolvedor ETL



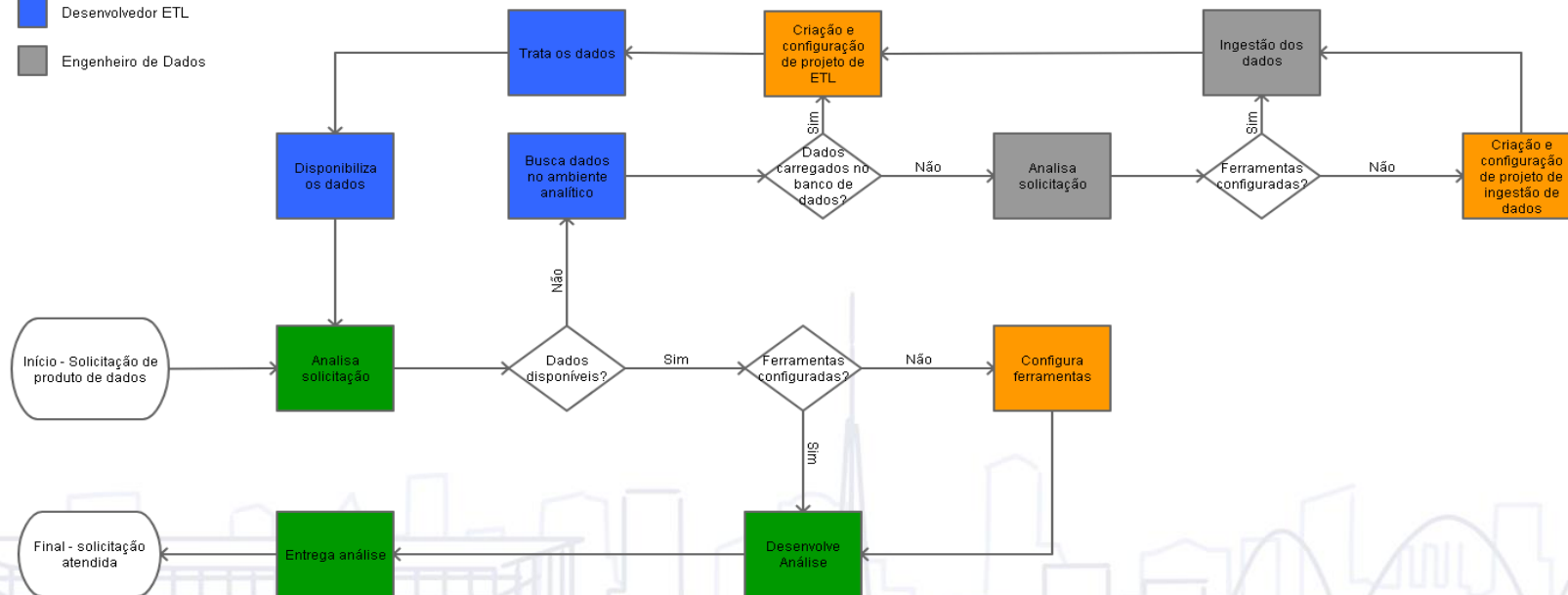
Cientista de Dados

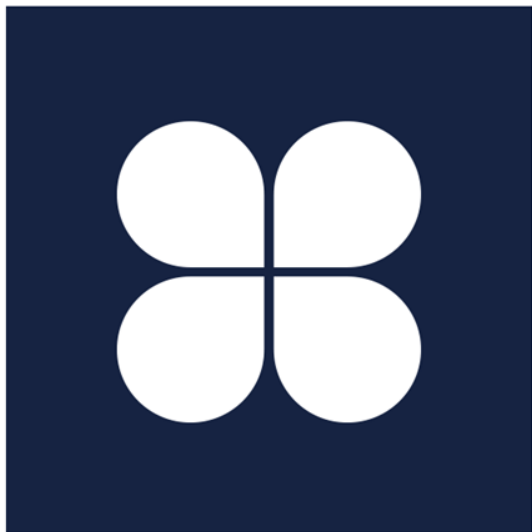


Consumidor

# Fluxo de trabalho

- Analista de dados/Cientista de dados
- Administrador de ambiente
- Desenvolvedor ETL
- Engenheiro de Dados





## Projetos



# Projetos

Engenharia de Dados	ETL	Análise de Dados	Ciência de Dados	Administração de Ambiente
<ul style="list-style-type: none"> <li>• Controle de Processamento</li> <li>• ODS</li> <li>• Bom Pagador</li> <li>• NFM (Em Andamento)</li> <li>• DTS (Em Andamento)</li> <li>• Integração Nfe</li> <li>• Integração Auditor Web DB2</li> <li>• Parsers</li> <li>• Migração Catálogo COS (Em Andamento)</li> </ul>	<ul style="list-style-type: none"> <li>• DW Receita CNPJ</li> <li>• DW Receita IE</li> <li>• DW Combustíveis</li> <li>• DW DAPI</li> <li>• DW EFD</li> <li>• DW Regime Especial</li> <li>• DW</li> <li>• Autorregularização</li> <li>• DW NFE Detalhada</li> <li>• DW GIAST</li> <li>• DW Simples Nacional</li> <li>• DW Parcelamento</li> <li>• DW Crédito Tributário (Em Andamento)</li> <li>• MDM</li> <li>• Lib DIMP</li> <li>• Lib IFISCO</li> <li>• Lib NFCe</li> <li>• Lib NFe</li> <li>• Lib CTe</li> <li>• Lib MDFe</li> <li>• Lib CTe_OS</li> <li>• Lib EFD</li> <li>• Lib SIAFI</li> <li>• Lib SEINFRA</li> <li>• Lib Conv. 115</li> <li>• DW SCANC (Parado)</li> </ul>	<ul style="list-style-type: none"> <li>• Relatórios de Receita – RMA</li> <li>• Radar GIAST</li> <li>• Radar Nfe</li> <li>• Painel de indicadores econômicos(COVID)</li> <li>• Painel de Receita</li> <li>• Relatório – Regime Especial</li> <li>• Painel</li> <li>• Autorregularização</li> <li>• Painel TTS</li> <li>• Painel IPI</li> <li>• Relatório – Controle de Omissos</li> <li>• Relatórios – SCANC (Parado)</li> </ul>	<ul style="list-style-type: none"> <li>• Classificador de combustíveis – NFM (Em Andamento)</li> </ul>	<ul style="list-style-type: none"> <li>• 26 VMs Linux</li> <li>• 4 VMs Windows</li> <li>• Datastage (2 instâncias)</li> <li>• IBM Cognos(3 instâncias)</li> <li>• SAS 9.4</li> <li>• SAS Viya 2 instâncias (Instalando)</li> <li>• Cluster Kubernetes</li> <li>• SPSS Server</li> <li>• DBA Netezza</li> <li>• DBA DB2</li> <li>• IBM COS</li> <li>• IBM Cloud Pak for Data System</li> <li>• IAS (2 instancias)</li> <li>• IBM Event Streams (Compartilhado)</li> <li>• Apache Spark</li> <li>• Apache Airflow (Previsto)</li> <li>• DevOps (Previsto)</li> </ul>



**Obrigado!**

[rogerio.zupo@fazenda.mg.gov.br](mailto:rogerio.zupo@fazenda.mg.gov.br)

[bruno.meira@fazenda.mg.gov.br](mailto:bruno.meira@fazenda.mg.gov.br)

