Projeto6-Plano

# Data Science Academy[¶](#Data-Science-Academy)

## Arquitetura de Plataforma de Dados e Modern Data Stack[¶](#X5edb4cd055596628df4b17a6078ac0bee655791)

## Projeto 6 - Plano[¶](#Projeto-6---Plano)

### Construção do Plano de Governança, Observabilidade, Qualidade e Segurança de Dados[¶](#X4091a48faaf367b4c91c18129b0100675d4754d)



# Definição do Projeto, Objetivo Principal e Importância[¶](#X6c4582c94984f858b11d85f35c8d19e3d1926c7)

## Definição do Projeto[¶](#Defini%C3%A7%C3%A3o-do-Projeto)

O Projeto 6 (que seria o projeto de criação e implementação do Plano de Governança) envolve a construção de um plano abrangente para estabelecer a governança, observabilidade, qualidade e segurança de dados em uma organização.

Esse plano visa garantir que os dados sejam gerenciados, monitorados e protegidos de maneira consistente e alinhada às melhores práticas e regulamentos vigentes. O projeto engloba o desenvolvimento de políticas, processos, métricas e ferramentas para monitorar e controlar a integridade, disponibilidade, confidencialidade e conformidade dos dados ao longo de seu ciclo de vida.

## Objetivo Principal[¶](#Objetivo-Principal)

O objetivo principal do Projeto é estruturar um plano eficaz e sustentável que permita à organização ter um maior controle sobre a governança, monitoramento, qualidade e segurança de seus dados. Isso inclui a implementação de práticas e tecnologias que permitam a identificação, rastreamento, monitoramento e proteção dos dados, além de assegurar que as informações estejam acessíveis, confiáveis e seguras.

## Importância[¶](#Import%C3%A2ncia)

A importância desse projeto reside na crescente dependência das empresas em dados para a tomada de decisões, inovação e eficiência operacional. Com o aumento no volume e diversidade de dados, há uma necessidade crítica de implementar uma governança sólida e práticas de segurança para evitar riscos, como violações de dados, perda de integridade e problemas de conformidade regulatória. Além disso, ao fortalecer a qualidade e observabilidade dos dados, a organização garante a capacidade de detectar problemas proativamente, assegurar dados precisos e confiáveis e, assim, otimizar processos e decisões baseadas em dados.

## Cenário de Negócio e Necessidades da Empresa[¶](#X7935ed533d92f03e82a4be84f631cb6ed45cac8)

A DSATech Solutions é uma empresa de tecnologia focada em oferecer soluções de Big Data e análise de dados para clientes de diferentes setores, como varejo, saúde, e-commerce e finanças. Com o crescimento rápido e a expansão para novos mercados, a empresa enfrenta desafios em relação ao volume e à diversidade dos dados coletados de seus clientes, exigindo um plano robusto de governança, observabilidade, qualidade e segurança de dados para garantir a confiabilidade e conformidade.

Necessidades da empresa: **Governança de Dados**: A DSATech Solutions precisa garantir que seus dados estejam bem documentados, categorizados e acessíveis de forma controlada, garantindo conformidade com regulamentações (LGPD, GDPR). Além disso, precisa definir papéis e responsabilidades sobre o ciclo de vida dos dados, como curadoria, controle de acesso e retenção de dados.

**Observabilidade**: Com o aumento da complexidade de seu ecossistema de dados, a empresa precisa de ferramentas que monitorem a infraestrutura de dados, detectem anomalias e antecipem problemas. Isso inclui o rastreamento de fluxos de dados, monitoramento em tempo real e criação de alertas em caso de falhas.

**Qualidade de Dados**: A DSATech Solutions trabalha com dados sensíveis e críticos, onde a qualidade é essencial para garantir a precisão das análises. A empresa precisa implementar mecanismos de validação de dados para garantir consistência, completude, atualidade e acurácia dos dados.

**Segurança de Dados**: A empresa lida com grandes volumes de dados confidenciais e qualquer vazamento ou mau uso pode comprometer tanto a reputação quanto a conformidade legal. É necessário implementar políticas rígidas de segurança, com criptografia, auditoria e controle de acesso, além de treinamentos contínuos para os colaboradores.

## Business Case[¶](#Business-Case)

**Resumo Executivo:**

O Projeto de Governança, Observabilidade, Qualidade e Segurança de Dados da DSATech Solutions visa criar uma estrutura para garantir a gestão eficiente e segura dos dados, atendendo a regulamentações e melhorando a tomada de decisão. O projeto proporcionará benefícios como conformidade legal, melhor qualidade de dados e maior segurança, além de redução de custos operacionais. Recomenda-se a implementação em fases para garantir controle e sucesso.

**Justificativa do Projeto:**

A DSATech Solutions está expandindo rapidamente, aumentando o volume e a diversidade de dados que gerencia. A ausência de um plano sólido de governança e segurança pode levar a erros, perda de dados e violações de conformidade, o que compromete a reputação e operação da empresa. O projeto é necessário para padronizar processos, aumentar a segurança e garantir a conformidade regulatória.

**Análise de Alternativas:**

Interna: Construção interna de uma solução personalizada para governança e segurança de dados.

* Prós: Controle total, personalização conforme as necessidades.
* Contras: Custos elevados, tempo de implementação longo, necessidade de expertise.

Soluções de Terceiros: Implementação de ferramentas de mercado para governança e segurança de dados.

* Prós: Soluções testadas, menor tempo de implementação.
* Contras: Custo de licenciamento contínuo, menos flexibilidade.

Manter Status Quo: Continuar com processos atuais e ajustes incrementais.

* Prós: Custo inicial baixo, sem mudanças bruscas.
* Contras: Alto risco de não conformidade e falhas de segurança, maior risco de perda de dados.

**Benefícios Esperados:**

* Tangíveis: Redução de custos com auditorias e multas, menor tempo gasto em gerenciamento manual de dados, aumento da eficiência operacional.
* Intangíveis: Melhor reputação no mercado, confiança dos clientes em relação à segurança e privacidade, cultura de dados consolidada.

**Custos Estimados:**

* Investimento Inicial: Aquisição de ferramentas de governança de dados, implementação de sistemas de segurança, treinamento de equipe.
* Custos Operacionais: Manutenção de sistemas, atualizações e suporte técnico.
* Manutenção: Acompanhamento contínuo de conformidade, ajustes em políticas de governança e segurança.

**Análise de Riscos:**

* Falhas na Implementação: Risco de atrasos ou falhas técnicas durante a implementação. Mitigação: Fases de teste e revisão em cada etapa.
* Conformidade Regulamentar: Mudanças nas regulamentações. Mitigação: Revisão contínua das políticas de conformidade.
* Resistência Interna: Resistência dos funcionários à adoção de novos processos. Mitigação: Treinamento e comunicação clara.

**Retorno sobre o Investimento (ROI):**

* Espera-se um retorno positivo em até 2 anos, com aumento de eficiência, redução de custos de compliance, menor risco de multas e perda de dados, além da melhoria na satisfação e retenção de clientes.

**Cronograma e Plano de Implementação:**

* Fase 1 - Planejamento e análise de requisitos (2 meses)
* Fase 2 - Seleção de Ferramentas e Infraestrutura (1 mês)
* Fase 3 - Implementação de Observabilidade e Qualidade de Dados (4 meses)
* Fase 4 - Implementação de Governança e Segurança de Dados (2 meses)
* Fase 5 - Testes Finais e Ajustes (1 mês)
* Conclusão - Rollout, Treinamento e Operacionalização (1 mês)

**Impactos no Negócio:**

* O projeto trará maior conformidade e segurança, com impacto positivo na operação ao reduzir riscos de violações de dados e aumentar a confiança dos clientes.
* No curto prazo, pode haver sobrecarga na equipe técnica durante a fase de implementação, o que pode exigir ajustes temporários.

**Indicadores de Sucesso:**

* Redução de incidentes de segurança: Número de incidentes e vazamentos de dados.
* Conformidade com regulamentações: Auditorias internas e externas.
* Qualidade de dados: Redução de inconsistências e erros nos dados.
* Satisfação dos clientes: Pesquisas e feedback sobre a confiança em relação à segurança de dados.
* Eficiência operacional: Redução do tempo gasto na gestão de dados.

## Planejamento e Definição de Requisitos[¶](#X2901e06fb4ac64a6c18e98c7238f7c9e097acf4)

**Objetivo Principal:**

Implementar um plano abrangente de governança, observabilidade, qualidade e segurança de dados na DSATech Solutions, garantindo conformidade regulatória, segurança dos dados e otimização de processos.

**Requisitos Funcionais:** Governança de Dados:

* Definir e implementar políticas de governança, com papéis e responsabilidades claros.
* Estruturação de um catálogo de dados para facilitar a acessibilidade e rastreabilidade.

Observabilidade:

* Implementar monitoramento em tempo real para fluxos de dados e detecção de anomalias.
* Criação de alertas e relatórios automáticos para identificação de falhas.

Qualidade de Dados:

* Estabelecer regras de validação e conformidade dos dados em cada estágio do pipeline.
* Automatizar auditorias de qualidade com relatórios periódicos.

Segurança de Dados:

* Implementar mascaramento de dados sensíveis.
* Controle de acesso baseado em funções (RBAC) para garantir a segurança dos dados sensíveis.
* Ferramentas de auditoria para rastreamento de atividades e acessos.

**Requisitos Técnicos:**

* Infraestrutura: Integração com as plataformas de cloud e big data já existentes (AWS, GCP, etc.).
* Ferramentas: Utilização de ferramentas como Apache Atlas para governança, Prometheus e Grafana para observabilidade e Great Expectations para controle de qualidade de dados.
* Segurança: Integração de sistemas de segurança com ferramentas como HashiCorp Vault para gerenciamento de credenciais e criptografia e Linguagem Python para mascaramento de dados sensíveis.

**Equipe e Papéis:**

* Gerente de Projeto: Responsável por coordenar a implementação e garantir o cumprimento dos prazos.
* Arquiteto de Dados: Responsável por definir políticas e estruturar o plano de gestão de dados.
* Engenheiro de Dados: Responsável pela implementação técnica das ferramentas de observabilidade e qualidade.
* Especialista em Segurança de Dados: Garantir que as medidas de segurança sejam seguidas e implementadas corretamente.
* Equipe de TI e Operações: Suporte à implementação e integração com os sistemas existentes.

**Recursos Necessários:**

* Ferramentas de software: Apache Atlas, Prometheus, Grafana, Great Expectations, HashiCorp Vault, Programação Python.
* Recursos humanos: Equipe dedicada para o desenvolvimento e implementação.
* Orçamento: Alocação de recursos financeiros para aquisição de ferramentas e treinamentos.

**Critérios de Aceitação:**

* Políticas de governança documentadas e implementadas.
* Sistemas de monitoramento e qualidade de dados operacionais.
* Segurança de dados atendendo aos padrões da indústria (LGPD, GDPR).
* Treinamento da equipe concluído e sistemas funcionando em produção.

## Análise de Situação Atual (AS-IS) - DSATech Solutions[¶](#X702824dbdd1d0bd44753bff8a6132045dbfc067)

**Governança de Dados:**

* Status Atual: A empresa não possui políticas de governança de dados formalizadas. Os dados estão distribuídos em diferentes sistemas, com pouca padronização e falta de documentação clara sobre a propriedade dos dados e seus ciclos de vida.
* Problemas: Falta de rastreabilidade, duplicação de dados e inconsistências nas fontes de dados, tornando difícil a gestão eficiente.

**Observabilidade:**

* Status Atual: Não há um sistema centralizado de monitoramento de fluxos de dados. O acompanhamento é feito manualmente ou reativo, sendo notados apenas após ocorrências de falhas, o que causa atrasos na identificação de problemas.
* Problemas: Ausência de monitoramento em tempo real e alertas automáticos, resultando em tempos de inatividade prolongados e dificuldades em prever falhas.

**Qualidade de Dados:**

* Status Atual: A qualidade dos dados não é monitorada de forma consistente. Erros nos dados são descobertos tardiamente e a correção é feita de maneira manual, sem processos automatizados de validação ou limpeza de dados.
* Problemas: Dados inconsistentes e incompletos, afetando a confiabilidade das análises e relatórios entregues aos clientes.

**Segurança de Dados:**

* Status Atual: Há medidas básicas de segurança, como controle de acesso e backup de dados. No entanto, a criptografia de dados e o monitoramento de acessos não são amplamente implementados. A empresa está exposta a riscos de violações de dados, especialmente dados sensíveis de clientes.
* Problemas: Vulnerabilidades de segurança, com potencial de perda ou vazamento de dados sensíveis, além de falhas de conformidade com regulações como LGPD e GDPR.

**Integração de Sistemas:**

* Status Atual: A integração entre diferentes sistemas de dados é limitada. Processos de extração e transformação de dados são feitos de forma semi-automatizada, com pouca interconectividade entre plataformas.
* Problemas: Ineficiências nos fluxos de trabalho, resultando em tempo elevado para consolidação de dados e geração de insights.

## Desenho do Estado Futuro (TO-BE) - DSATech Solutions[¶](#Xe5e6c8927241429fcefbdb7f5d3e3f7f4c92972)

**Governança de Dados:**

* Futuro Desejado: Políticas de governança de dados claramente definidas e implementadas, com um catálogo de dados centralizado, padronização das fontes e ciclos de vida bem documentados. Papéis e responsabilidades bem distribuídos, com processos de curadoria e gestão de dados automatizados.
* Benefícios: Maior controle e rastreabilidade dos dados, garantindo conformidade com regulamentações como LGPD e GDPR, além de facilitar a governança e o acesso adequado às informações.

**Observabilidade:**

* Futuro Desejado: Implementação de sistemas de monitoramento em tempo real para fluxos de dados, com alertas automáticos e dashboards centralizados. Ferramentas de observabilidade integradas que permitem rastrear e diagnosticar problemas de forma proativa.
* Benefícios: Detecção rápida de falhas, previsibilidade de problemas antes de impactarem o negócio, maior tempo de uptime e eficiência operacional.

**Qualidade de Dados:**

* Futuro Desejado: Implementação de ferramentas automatizadas de validação e monitoramento da qualidade de dados, com auditorias regulares e relatórios automáticos. Estabelecimento de regras e padrões de qualidade, como completude, consistência e acurácia.
* Benefícios: Redução de erros e inconsistências nos dados, aumento da confiabilidade das análises e relatórios, com decisões mais baseadas em dados precisos.

**Segurança de Dados:**

* Futuro Desejado: Mascaramento de dados sensíveis. Implementação de ferramentas de segurança robustas, como gerenciamento de credenciais e rastreamento de logs.
* Benefícios: Maior proteção contra violações de dados, conformidade garantida com regulamentações de privacidade e aumento da confiança dos clientes em relação à segurança dos dados.

**Integração de Sistemas:**

* Futuro Desejado: Integração completa entre sistemas de dados, com pipelines de ETL automatizados e interoperabilidade entre plataformas. Dados fluindo de forma contínua e sem fricção entre diferentes fontes, permitindo análises rápidas e eficientes.
* Benefícios: Processos mais ágeis, maior eficiência na manipulação de dados e capacidade de gerar insights em tempo real para apoiar decisões de negócio.

## Cronograma de Implementação do Projeto - DSATech Solutions[¶](#Xf6ec67b21b7713bd4240c122e2057ce73149803)

Duração Total: 10 meses

**Fase 1: Planejamento e Análise de Requisitos (2 meses)**

Mês 1:

* Definição dos objetivos específicos do projeto.
* Reunião com stakeholders para levantamento de necessidades e requisitos.
* Análise de requisitos técnicos e funcionais.
* Criação do plano detalhado de implementação.

Mês 2:

* Mapeamento de processos atuais (AS-IS).
* Desenho dos processos futuros (TO-BE).
* Identificação de ferramentas e tecnologias necessárias.
* Aprovação final do planejamento.

**Fase 2: Seleção de Ferramentas e Infraestrutura (1 mês)**

Mês 3:

* Avaliação e seleção de ferramentas para governança, observabilidade, qualidade e segurança de dados (Apache Atlas, Prometheus, Grafana, Great Expectations, HashiCorp Vault).
* Definição de arquitetura técnica para integração com os sistemas existentes.
* Aquisição de ferramentas e infraestrutura necessária (cloud, software, etc.).

**Fase 3: Implementação de Observabilidade e Qualidade de Dados (4 meses)**

Mês 4:

* Definição de regras de observabilidade.
* Definição de regras de qualidade de dados (completude, consistência, acurácia).

Mês 5:

* Implementação de ferramentas de observabilidade (Python).
* Implementação de ferramentas de validação e auditoria de qualidade de dados (Great Expectations em Python).

Mês 6:

* Configuração das ferramentas de monitoramento em tempo real (Prometheus, Grafana).
* Criação de dashboards para monitorar fluxos de dados e definir alertas automáticos.

Mês 7:

* Testes de monitoramento e qualidade de dados.

**Fase 4: Implementação de Governança e Segurança de Dados (2 meses)**

Mês 8:

* Configuração inicial das ferramentas de governança (catalogação de dados, definição de papéis e responsabilidades).
* Início da implementação de controles de acesso (RBAC), criptografia de dados e mascaramento.

Mês 9:

* Implementação das políticas de governança e documentação de ciclo de vida dos dados.
* Criação de mecanismos de auditoria para monitorar acessos e mudanças nos dados.
* Testes iniciais de segurança e conformidade.

**Fase 5: Testes Finais e Ajustes (1 mês)**

Mês 10:

* Execução de testes integrados para garantir que as ferramentas de governança, observabilidade, qualidade e segurança funcionem de maneira conjunta.
* Ajustes finais com base nos resultados dos testes.
* Treinamento da equipe em novos processos e ferramentas.

**Fase 6: Rollout, Treinamento e Operacionalização (1 mês)**

Mês 11:

* Capacitação de todos os times envolvidos (dados, segurança, operações) sobre os novos processos e ferramentas.
* Lançamento oficial do novo ambiente de governança, observabilidade, qualidade e segurança.
* Monitoramento inicial de performance e refinamento contínuo dos processos implementados.

Marcos Principais:

* Mês 2: Planejamento aprovado.
* Mês 3: Ferramentas e infraestrutura selecionadas.
* Mês 7: Observabilidade e qualidade de dados operacionais.
* Mês 9: Implementação de governança e segurança concluída.
* Mês 11: Projeto totalmente implementado e operacionalizado.

## Entregáveis do Projeto - DSATech Solutions[¶](#Xcc89e4a04a073e2cc51edd4d6128733dec179de)

**Políticas de Governança de Dados:**

* Documento formal com as políticas de governança de dados, definindo papéis, responsabilidades e ciclo de vida dos dados.
* Catálogo de dados centralizado com todas as fontes de dados mapeadas e documentadas.

**Sistema de Observabilidade:**

* Ferramentas configuradas para monitoramento em tempo real dos fluxos de dados (Prometheus, Grafana), com dashboards e alertas automáticos.
* Relatórios de observabilidade que detectam anomalias e previsões de falhas.

**Ferramentas de Qualidade de Dados:**

* Sistema automatizado de validação e auditoria de qualidade de dados (Great Expectations).
* Relatórios periódicos de qualidade, com métricas de consistência, completude e acurácia dos dados.

**Segurança de Dados:**

* Implementação de criptografia de dados em repouso e em trânsito.
* Mascaramento de dados sensíveis.
* Sistema de controle de acesso baseado em funções (RBAC), com auditorias de acessos e rastreamento de mudanças.
* Ferramenta de gerenciamento de credenciais e chaves de criptografia (HashiCorp Vault).

**Dashboards e Relatórios:**

* Dashboards operacionais para monitorar a saúde do sistema e os principais indicadores de performance (observabilidade, qualidade e segurança).
* Relatórios automatizados de conformidade e auditorias de segurança.

**Treinamento e Documentação:**

* Treinamento completo para as equipes de dados e operações sobre o uso das novas ferramentas e processos.
* Documentação técnica e de procedimentos para o uso contínuo das soluções implementadas.

**Ambiente Integrado e Operacional:**

* Ambiente completo com governança, observabilidade, qualidade e segurança de dados integrados e em operação.
* Plano de manutenção e evolução contínua das soluções implementadas.

## Estimativa de Custo para o Projeto - DSATech Solutions[¶](#X148b2bfd0c3a5b3048aaa80eabe5cb413f06c78)

Os valores abaixo são estimativas em dólar americano.

**1. Custos com Ferramentas e Licenças**

* Apache Atlas (governança de dados) – Ferramenta open-source: 0 (custo inicial, porém com custo de manutenção e suporte interno).
* Prometheus & Grafana (observabilidade e monitoramento) – Ferramentas open-source: 0 (custo inicial, porém com custo de implementação e manutenção).
* Great Expectations (qualidade de dados) – Ferramenta open-source: 0 (custo inicial, custo com integração e manutenção).
* HashiCorp Vault (gestão de credenciais e segurança) – Versão open-source: 0 (com custos de suporte e manutenção internos).
* Linguagem Python - Versão open-source.
* Total de Ferramentas e Licenças: 0

**2. Custos com Infraestrutura**

* Cloud (Amazon Web Services, AWS): Custo estimado com infraestrutura em nuvem para armazenar e processar dados, gerenciar pipelines e hospedagem das soluções.
* Custo mensal médio: 2,000
* Custo anual estimado: 24,000

**3. Custos com Consultoria e Implementação**

* Consultoria Especializada em Governança de Dados (3 meses): 15,000
* Implementação de Ferramentas de Observabilidade e Qualidade (4 meses): 25,000
* Configuração de Segurança e Criptografia (4 meses): 20,000
* Total de Consultoria e Implementação: 60,000

**4. Custos com Equipe Interna**

* Engenheiro de Dados (6 meses) – 10,000/mês: 60,000
* Arquiteto de Dados (6 meses) – 8,000/mês: 48,000
* Especialista em Segurança de Dados (6 meses) – 9,000/mês: 54,000
* Gerente de Projeto (6 meses) – 12,000/mês: 72,000
* Total de Custos com Equipe Interna: 234,000

**5. Treinamento e Capacitação**

* Treinamento para equipe técnica e operacional: 10,000
* Material e documentação: 5,000
* Total de Treinamento e Capacitação: 15,000

**6. Manutenção e Suporte**

* Manutenção e suporte técnico anual: 12,000
* Custos com atualizações de ferramentas: 3,000
* Total de Manutenção e Suporte: 15,000

**Custo Total do Projeto**

* Ferramentas e Licenças: 0
* Infraestrutura: 24,000
* Consultoria e Implementação: 60,000
* Equipe Interna: 234,000
* Treinamento e Capacitação: 15,000
* Manutenção e Suporte: 15,000

**Custo Final Estimado: 348,000**

# Fim[¶](#Fim)