

Gestão de Projeto de Ciências de Dados

Francisco Louzada Neto
ICMC/USP
louzada@icmc.usp.br



Objetivos da Disciplina de Gestão de Projeto de CD

- ✓ Fornecer ao aluno conhecimento básico de gestão de projeto de ciência de dados, mediante os diversos riscos e desafios inerentes da área, envolvendo as constantes mudanças tecnológicas, e a complexidade dos problemas e projetos.
- ✓ Discutir as tendências atuais gestão de projeto de ciência de dados, incluindo modelos de gerenciamento ágil de projetos e métodos híbridos.

O que veremos?

Gestão de Projeto de Ciência de Dados

- ✓ Motivação para implementação de um Projeto de Ciência de Dados
- ✓ Aspectos que devem ser analisados antes do início do Projeto
- ✓ Ciclo de Vida do desenvolvimento de produto de CD (Metodologia CRISP-DS)
- ✓ Incertezas e Gestão de Riscos
- ✓ Gestão Ágil de PCD
- ✓ Métodos Híbridos
- ✓ Exemplos de PCDs

Motivação

Gestão de Projeto de Ciência de Dados?

- ✓ O que é Gestão de Projeto?
- ✓ Por que precisamos de Gestão de Projetos?
- ✓ Porque Gestão de PCD?

Motivação para Gestão de um PCD

✓ O que é Gestão de Projeto?

- ✓ **Gestão de Projeto:** consiste em uma área da administração de empresas, a qual é responsável por definir os conhecimentos, habilidades e recursos necessários, incluindo custos e tempo, para a realização de um projeto com objetivos específicos.
- ✓ Consiste em um conjunto de práticas e competências utilizadas para planejar, executar, monitorar e controlar os projetos de uma organização, independentemente do tamanho ou da complexidade dos mesmos.

✓ Por que precisamos de Gestão de Projetos?

- ✓ Para manter o projeto sob controle!
- ✓ Para manter a direção do Projeto e fazer ajustes de percurso.

Motivação para Gestão de um PCD

- ✓ Porque Gestão de PCD?
- ✓ Qual a necessidade de utilizarmos Gestão de PCD?
- ✓ RESPOSTA DIRETA: PCDs são complexos
 - ✓ Alto valor de investimento
 - ✓ Projetos de médio e longo prazo (1 a 5 anos)
 - ✓ **Alto Risco de problema**
 - ✓ Não atingir os objetivos
 - ✓ Precisar de mais tempo para finalização
 - ✓ Custar mais do que o previsto
- ✓ **GESTÃO DE PCD:** aumenta as chances de sucesso!

Motivação para Gestão de um PCD

✓ Cenário simples de um PCD

PLANEJA/o

PROBLEMA
de
CIÊNCIA DE
DADOS

↑
Especificação
Necessidade
Business Case

Área de
Negócio

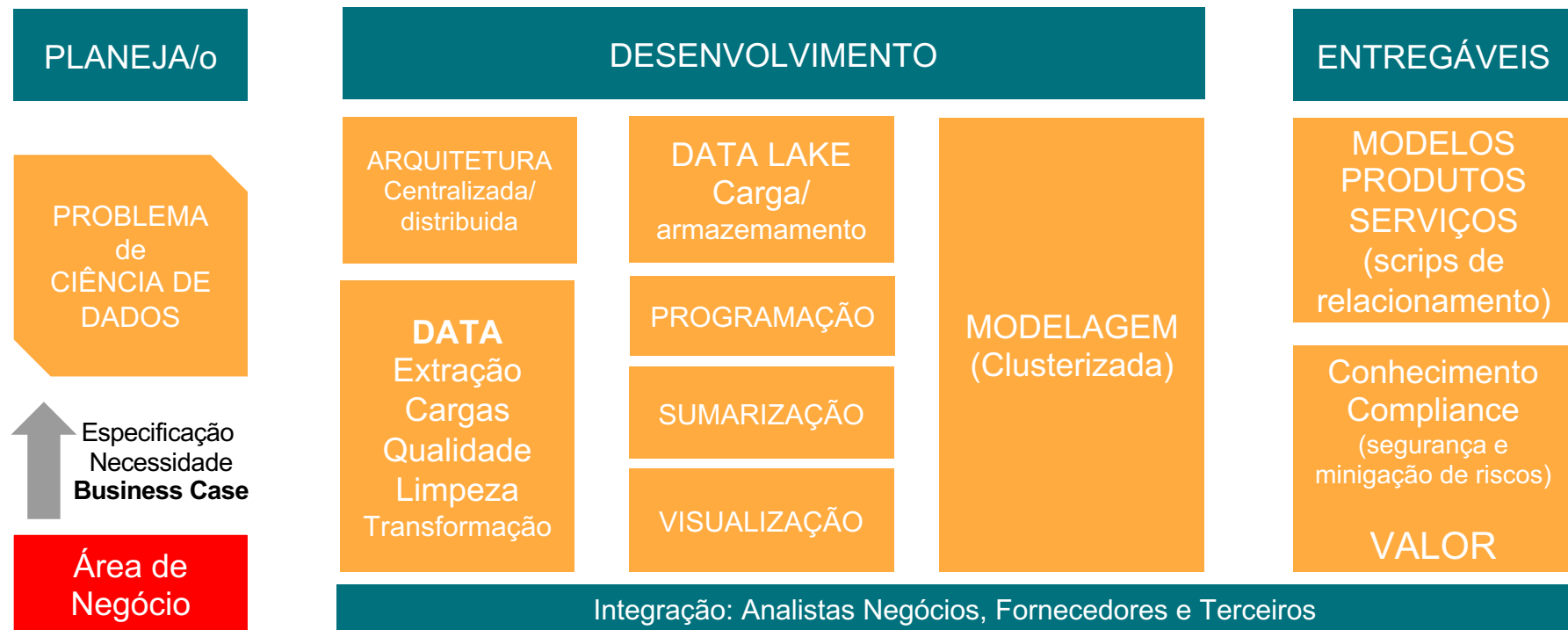
Motivação para Gestão de um PCD

✓ Cenário simples de um PCD



Motivação para Gestão de um PCD

✓ Cenário simples de um PCD



Motivação para Gestão de um PCD

- ✓ **Voltando:** Porque Gestão de PCD?
- ✓ **RESPOSTA DIRETA:** PCDs são complexos
 - ✓ Projetos de médio e longo prazo (1 a 5 anos)
 - ✓ Alto valor agregado (alto valor de investimento)
 - ✓ **Alto Risco**
 - ✓ Não atingir os objetivos
 - ✓ Precisar de mais tempo para finalização
 - ✓ Custar mais do que o previsto
- ✓ **GESTÃO DE PCD:** aumenta as chances de sucesso!

Motivação para Gestão de um PCD

- ✓ Mas complexos ainda...
- ✓ Produtos são virtuais
- ✓ Mudanças tecnológicas frequentes
- ✓ Difícil de se fazer mudanças com facilidade
- ✓ Ciência de Dados (com a conhecemos) é uma ciência nova:
 - ✓ Exemplos: Python (1989), R (1993), Hadoop (2012), Shiny (anos 2000)
- ✓ Fracassos: 20 a 25 % (minha experiência)
- ✓ Ou seja, **devemos fazer a gestão de um PCD!**

A situação atual



Estamos cercados por dados!

Responsáveis?



- ✓ Dados históricos coletados ao longo de muitos anos
- ✓ Processos de coleta automática de dados
- ✓ Instrumentação eletônica
- ✓ Transações on-line



- ✓ Sensores para captação de informações variadas
- ✓ Posts/Fotos/Vídeos nas redes sociais
- ✓ Transações de compras online
- ✓ Sinais de GPS de celulares
- ✓ *Internet das coisas (IoT)*

A situação atual

RECONHECER

O que fazer com esta grande quantidade de dados complexos?
Como analisar?



ENTREGAR

Como reverter estas informações em conhecimento útil que possa ser utilizado para tomada de decisão?



IDENTIFICAR

Como identificar informações escondidas nos dados?
Prever e se antecipar a possíveis problemas futuros?



A situação atual

Situação Atual – Obtenção de Grandes Bases de Dados

Presença de
Terabites/
exabites de
dados

Dados de
transmissão,
milissegundos
a segundos para
responder

Dados
estruturados,
não-estruturados
texto, multimídia

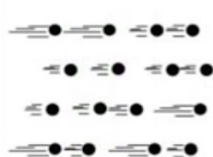
Inconsistências,
ambiguidades,
latência, dados
faltantes etc

Necessidade de
gerar valor com
os resultados
obtidos

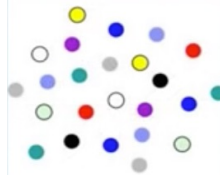
VOLUME



VELOCIDADE



VARIEDADE



VERACIDADE



VALOR



A situação atual

(Necessidade de Gestão:
Dados, Pessoas, Tecnologias e Requisitos)



Pessoas



Requisitos

Tecnologia



Dados



A situação atual

(Necessidade de Gestão:
Dados, Pessoas, Tecnologias e Requisitos)

✓ Temos dados com estrutura não convencional:

- ✓ O tamanho do banco de dados
- ✓ Dados contaminados
- ✓ Dados i i d
- ✓ Não estacionariedade
- ✓ No Covariáveis >>> no Observações
- ✓ Variáveis não numéricas
- ✓ Bases separadas fisicamente

A situação atual

(Necessidade de Gestão:
Dados, Pessoas, Tecnologias e Requisitos)

- ✓ Temos intervenção humana/Analistas Negócio:
 - ✓ Suporte no entendimento do problema (usuário e insider)
 - ✓ Definição dos requisitos técnicos e das funcionalidades da solução
 - ✓ Configuração do pré-processamento de dados
 - ✓ Desenvolvimento de sumarização, visualização, modelagem
 - ✓ Programação dos scripts de relacionamento
 - ✓ Definição as regras de negócio
 - ✓ Interpretação da legislação
 - ✓ Produção de relatórios

A situação atual

(Necessidade de Gestão:
Dados, Pessoas, Tecnologias e Requisitos)

- ✓ Temos necessidade de tecnologias específicas:
 - ✓ Hardware específico (Ex. Cluster Eucler, PLACA DE VIDEO NVIDIA 24GB RTX 3090)
 - ✓ Diferente tipos de softwares (Ex. SAS, Matlab, Python, R, React, Shiny)

A situação atual

(Necessidade de Gestão:
Dados, Pessoas, Tecnologias e Requisitos)

- ✓ Temos necessidade de atender a requisitos funcionais bem definidos:
 - ✓ A partir normas internas de necessidades gerenciais
 - ✓ A partir de leis de conformidade
 - ✓ A partir de regras de mercado

Motivação para Gestão de um PCD

✓ Ou seja

- ✓ **Necessidade de se fazer a Gestão de um PCD!**
- ✓ **Que pode trazer vários benefícios**

Benefícios da Gestão de um PCD

✓ Otimização do tempo

- ✓ Com base em um cronograma estruturado, o tempo gasto em cada uma das etapas pode ser estimado e depois comparado com o tempo realmente gasto.

✓ Redução de custo

- ✓ Com base em um orçamento estruturado, o custo de cada uma das etapas pode ser estimado e depois comparado com o custo real do projeto.

Benefícios da Gestão de um PCD

✓ Controle de Riscos

- ✓ Com base em um mapeamento de risco, podemos prever as ameaças ao projeto, identificando os indícios de que tais ameaças estão se concretizando.
- ✓ Também, podemos tentar mitigar os seus impactos negativos nos resultados do projeto e manter as metas.

✓ Impactos nos Resultados do Projeto

- ✓ Com base em uma gestão estruturada, os resultados do projeto se concretizarão e, em consequência, teremos maior satisfação do cliente.

O que veremos?

Gestão de Projeto de Ciência de Dados

- ✓ Motivação para implementação de um Projeto de Ciência de Dados
- ✓ **Aspectos que devem ser analisados antes do início do Projeto**
- ✓ CRISP-DS & Ciclo de Vida do desenvolvimento de produto de CD
- ✓ Incertezas e Gestão de Riscos
- ✓ Gestão Ágil de PCD
- ✓ Métodos Híbridos
- ✓ Exemplos de PCDs

✓ Antes do início do PCD

- ✓ Antes de iniciarmos um PCD, devemos considerar algumas questões no sentido de:
 - ✓ Minimizar riscos
 - ✓ Minimizar perda de tempo
 - ✓ Minimizar chance de fracasso do projeto

✓ Antes do início do Projeto

- ✓ Entender/determinar quais são os requisitos/objetivos
- ✓ O que pretendemos entregar de Valor
- ✓ Estipular critérios de sucesso do projeto

✓ Antes do início do Projeto

✓ Entender/determinar quais são os requisitos/objetivos

- ✓ Discutir com a direção da empresa ou solicitante sobre quais são as necessidades envolvidas na realização do projeto, qual é o business case em questão
 - ✓ Entender os requisitos/objetivos com clareza
 - ✓ Estar envolvido ou se envolver no problema

✓ Antes do início do Projeto

✓ O que pretendemos entregar de Valor

- ✓ Certificar de que o projeto vai estar focado em entregar valor para a organização ou ao negócio
 - ✓ Com foco nos requisitos/objetivos do projeto
 - ✓ Cuidando para que não se tenha desvios de finalidade

ENTREGÁVEIS

MODELOS
PRODUTOS
SERVIÇOS
(scrips de
relacionamento)

Conhecimento
Compliance
(segurança e
minigação de riscos)

VALOR

✓ Antes do início do Projeto

✓ Estipular critérios claros de sucesso

- ✓ Devem ser definidos conjuntamente com a empresa/instituição/solicitante/usuários: o que se quer entregar/produzir

SUCESSO

Projeto é sucesso?

Entregue no Tempo, Custo e Escopo estipulados inicialmente



✓ Antes do início do Projeto

✓ Estipular critérios claros de sucesso

- ✓ Devem ser definidos conjuntamente com a empresa/instituição/solicitante/usuários: o que se quer entregar/produzir
- ✓ Redução do custo operacional, da inadimplência, de fraude, de attrition, de perda de clientes
- ✓ Aumento de vendas, de cross selling, de up selling, de usuários satisfeitos, aumento da base de clientes, do lucro
- ✓ Se possível, mais de um único critério

SUCESSO

Projeto é sucesso?

Definir critérios claros de sucesso



✓ Antes do início do Projeto (Caso TCC)

✓ Estipular critérios claros de sucesso: Terminar!!!

- ✓ **TCC** – Focar no escopo e no cronograma
- ✓ Mitigar a possibilidade de atrasos
 - ✓ Discutir com o seu orientador
 - ✓ Pensar em alternativas. Ex. reduzir o escopo para conseguir fechar o **TCC** em um ano!

SUCESSO

Projeto é
sucesso?

Entregue no
Tempo,
Custo e escopo
iniciais



O que veremos na próxima aula?

Gestão de Projeto de Ciência de Dados

- ✓ Motivação para implementação de um Projeto de Ciência de Dados
- ✓ Aspectos que devem ser analisados antes do início do Projeto
- ✓ **CRISP-DS & Ciclo de Vida do desenvolvimento de produto de CD**
- ✓ **Exemplos de PCDs**
- ✓ Incertezas e Gestão de Riscos
- ✓ Gestão Ágil de PCD
- ✓ Métodos Híbridos

Dúvidas?

Gestão de Projeto de Ciência de Dados