

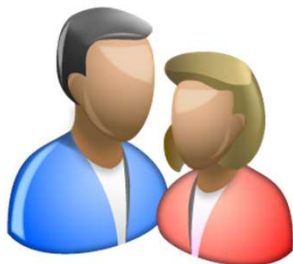
FIAP GRADUAÇÃO

# DATA SCIENCE

## DATA GOVERNANCE & DATA SECURITY MANAGEMENT

Prof. Dr. Renê de Ávila Mendes

# Você



Fale um pouco sobre você:

- Sua faixa etária
- Trabalha ou não na área
- Objetivo com o curso





## Renê

- Graduação em Sistemas de Informação (2007)
- Mestrado em Engenharia Elétrica (2017) – Arquiteturas para Big Data
- Doutorado em E.E. e Computação (2021) – Complexidade de Dados
- Administrador de Sistemas no Instituto Presbiteriano Mackenzie
- Professor de matérias de Big Data (Mineração de Dados, Coleta de Dados e Arquitetura para Big Data)
- [pf1752@fiap.com.br](mailto:pf1752@fiap.com.br)
- [www.linkedin.com/in/renemendes](https://www.linkedin.com/in/renemendes)

# Objetivos da disciplina

**DISCIPLINA:** Data Governance & Data Security Management

**OBJETIVOS:** Descubra como funciona um **projeto de banco de dados** dentro de um ambiente corporativo, aplicando **técnicas de levantamento e documentação de requisitos**, aderente aos projetos de bancos de dados e aprenda a representar esses requisitos em arquiteturas de solução tecnológica para Data distribution e Data integration, modelos de estruturas de dados e dicionários de dados buscando **Data quality**. Garanta a qualidade dos dados de uma empresa para prover os melhores subsídios à tomada de decisão de negócio, praticando **Data cleaning** para limpar, harmonizar, complementar e corrigir dados inconsistentes, incompletos ou incorretos. Compreenda como funciona o **ciclo de vida da informação** e as responsabilidades administrativas sobre os dados de negócio, buscando qualidade, segurança e compatibilidade com políticas de administração de informação corporativas auditáveis, aplicando práticas atuais de **Data profiling** e conhecendo os princípios de **Data auditing**, de forma a atender a **Lei Geral de Proteção de Dados (LGPD)**.

## **Assuntos – 1º Semestre**

- Introdução
- Estratégia Empresarial; Arquitetura Empresarial; TOGAF; Archimate e Archi
- Requisitos para projetos de bancos de dados
- SCRUM
- BPMN e Bizagi Modeler
- Qualidade em metadados; Oracle Data Modeler
- Arquiteturas de integração e distribuição física de banco de dados; Estudos de caso
- Master Data Management e Data Hub; Pentaho-DI



A Média Anual (MA), será obtida pela média das duas Médias Semestrais.

$$MA = (40\% \times \text{MMS 1º semestre} + 60\% \times \text{MMS 2º semestre})$$

Os critérios de aprovação baseiam-se na média semestral (para cursos semestrais) ou na média anual (para cursos anuais) obtida pelo aluno, conforme tabela abaixo:

MÉDIA FINAL (ANUAL OU SEMESTRAL)	SITUAÇÃO
0,0 a 3,9	Reprovado
4,0 a 5,9	Exame
6,0 a 10,0	Aprovado



# CHALLENGE

## PROJECT-BASED LEARNING

APRENDIZADO PRÁTICO PARA SOLUÇÕES DE PROBLEMAS REAIS

### 1a. Série 2023

- TEMA EM PROCESSO DE DEFINIÇÃO
- LIVE PARA O KICK OFF EM MARÇO/2023

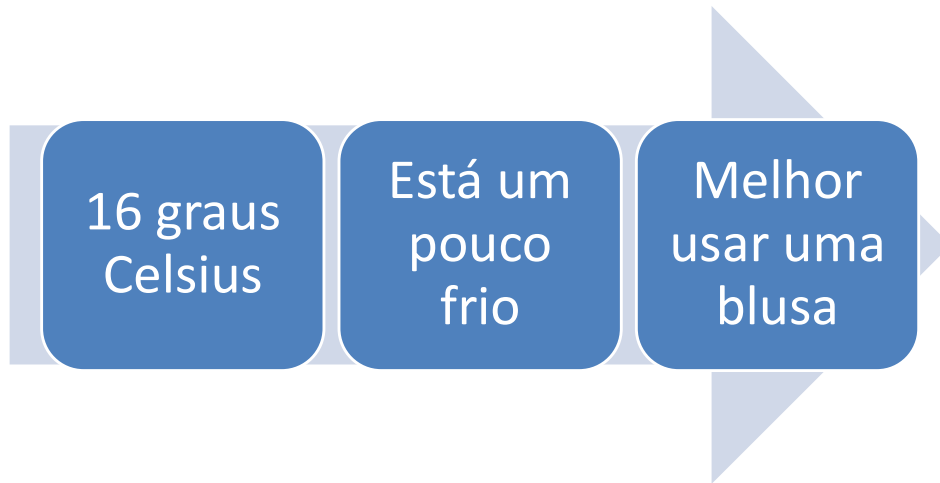


# Safra



# CONCEITOS FUNDAMENTAIS

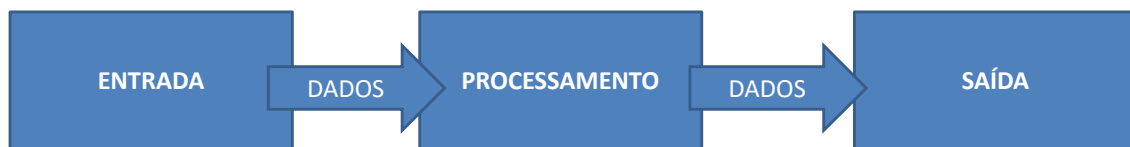
## Dados, informação, conhecimento



“O dado é um fato, um valor documentado ou um resultado de medição. Quando um sentido semântico ou um significado é atribuído aos dados, gera-se informação. Quando estes significados se tornam familiares, ou seja, quando um agente os aprende, este se torna consciente e capaz de tomar decisões a partir deles, e surge o conhecimento.”

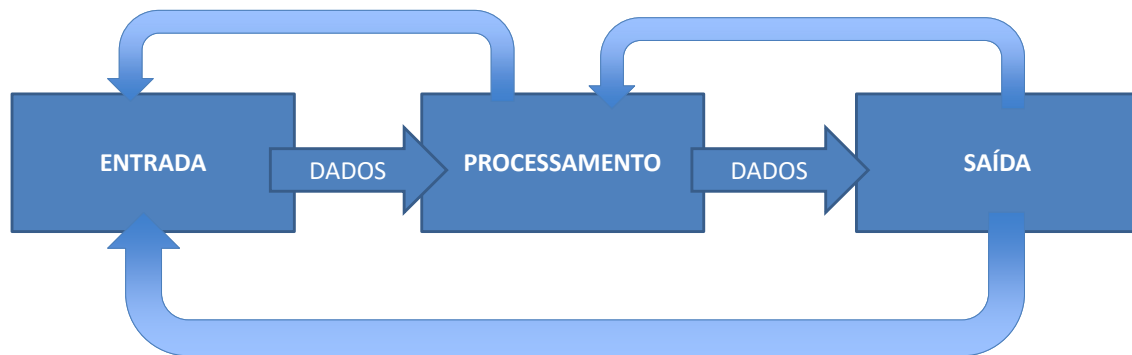
(DA SILVA, Leandro Augusto; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução à mineração de dados: com aplicações em R**. Elsevier Brasil, 2017, p6.)

## Sistema de negócio



Da mesma forma como na preparação de um café, utilizamos sistemas em outras atividades de nossa vida. Se escolher pedir uma refeição utilizando um aplicativo de seu celular, você utilizará um **sistema de negócio**, também chamado de **sistema de informação**. Nesse exemplo do aplicativo de encomenda de refeição, o sistema coleta dados (**entrada**), que são manipulados (**processamento**) e disseminados (**saída**) para alcançar o objetivo de pedir a refeição.

## Sistema de negócio



Os **dados** do seu pedido (tipo da refeição, quantidade, valor, restaurante), do atendimento do restaurante (duração da preparação, atendente) e da entrega (horário de saída, horário de entrega) serão armazenados e analisados para avaliar o lucro do fornecedor do aplicativo e para a melhoria do sistema de pedido de refeição (**realimentação do sistema**).

## Banco de dados

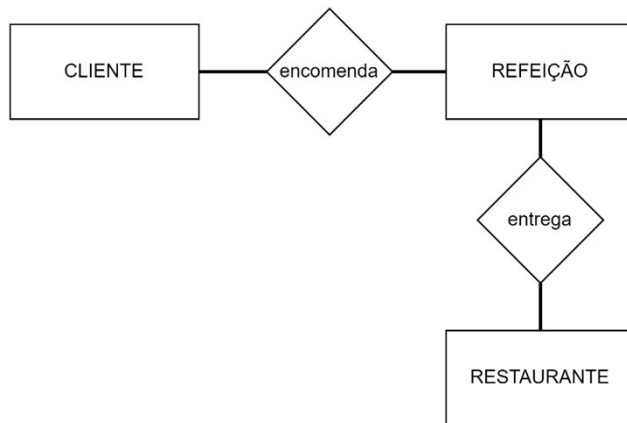
Uma coleção de dados bem **projetados**,  
**organizados** e cuidadosamente **gerenciados**.

## Banco de dados: projeto

Projetar um banco de dados significa utilizar um conjunto de técnicas, processos e notações que capturem os requisitos de dados, os requisitos estruturais para armazenamento dos dados e os requisitos físicos. Em essência, projetar um banco de dados envolve a criação de **modelos de dados**

Um **modelo** pode ser definido como “a representação abstrata e simplificada de um sistema real, com a qual se pode explicar ou testar o seu comportamento, em seu todo ou em partes” (COUGO, 2013, p.7). Em um projeto de banco de dados, o que se modela inicialmente são os **objetos** observados, **suas características e suas relações**, produzindo o que se chama de **modelo conceitual**.

## Banco de dados: projeto



Modelo  
conceitual

Um **modelo** pode ser definido como “a representação abstrata e simplificada de um sistema real, com a qual se pode explicar ou testar o seu comportamento, em seu todo ou em partes” (COUGO, 2013, p.7). Em um projeto de banco de dados, o que se modela inicialmente são os **objetos** observados, **suas características e suas relações**, produzindo o que se chama de **modelo conceitual**.

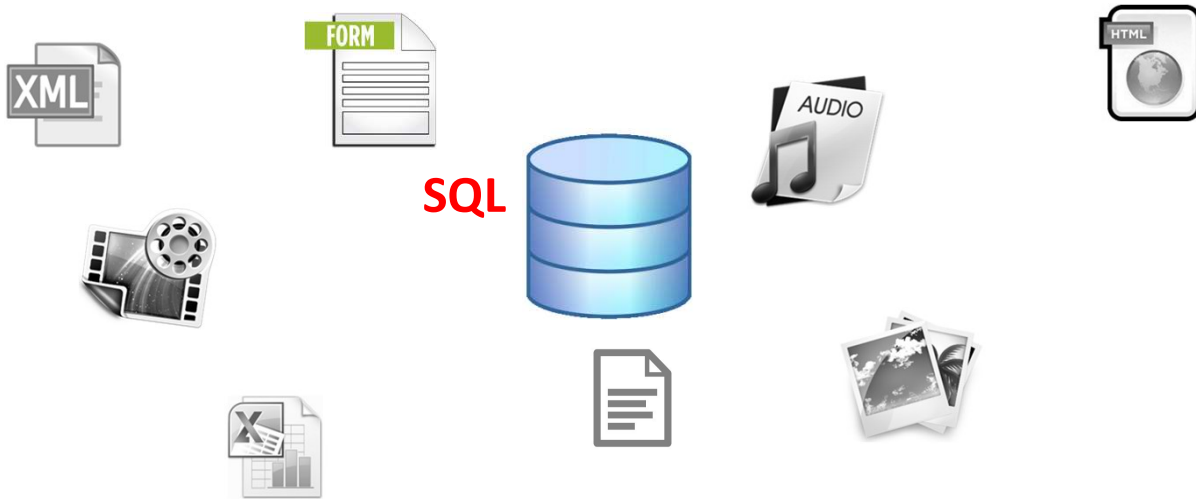


## Banco de dados: organização

Rotina diária	Sistema	Dado
Consultar mensagem no celular	Envio e recebimento de mensagens	Texto, imagem, ícone, vídeo, áudio
Ler notícias	Notificações, ou navegação web	Texto ou página HTML
Consultar redes sociais	Navegação web	Página HTML, imagem, vídeo, áudio
Pagar transporte público com cartão	Sistema de transações financeiras	Texto semiestruturado
Escutar música	Requisição e execução de áudio em fluxo, com ou sem salvamento	Áudio
Ler livro	Requisição, salvamento e apresentação de texto	Texto semiestruturado
Despachar pedidos, cadastrar e atender clientes	Sistema transacional	Texto, imagem
Ver filme	Requisição e execução de vídeo em fluxo, com ou sem salvamento	Vídeo

Além de bem planejado, um banco de dados que se proponha a representar satisfatoriamente os sistemas de negócio deve também ser **organizado**. Um arquiteto que prepara um bom projeto para seu cliente também pensará nos materiais e técnicas que usará para tornar seu projeto em algo real e funcional. No projeto de banco de dados os **requisitos estruturais de armazenamento** dos dados devem ser considerados.

# Dados estruturados



**DADOS ESTRUTURADOS** – dados que podem ser descritos por um **modelo de tipos de dados** e que podem ser armazenados, recuperados e processados com base neste modelo. Podem ser facilmente carregados, armazenados, recuperados e analisados. Geralmente são armazenados em bancos de dados relacionais. Exemplos: dados obtidos por formulários.

## Exemplo de dado estruturado

		$x_{ij}$						
ID		$i1$	$i2$	$i3$	$i4$	$i5$	$i6$	
$\vec{x}_i$		Matrícula	Nome	Sexo	Estado civil	Idade	Especialidade	Pretensão Salarial
	1	01	João	M	Viúvo	65	Garçom	R\$ 5000
	2	02	Maria	F	Solteiro	19	Cozinheiro	R\$ 2000
	...	...	...	...	...	...	...	...
	$n$	$n$	Pedro	M	casado	30	Entregador	R\$ 700

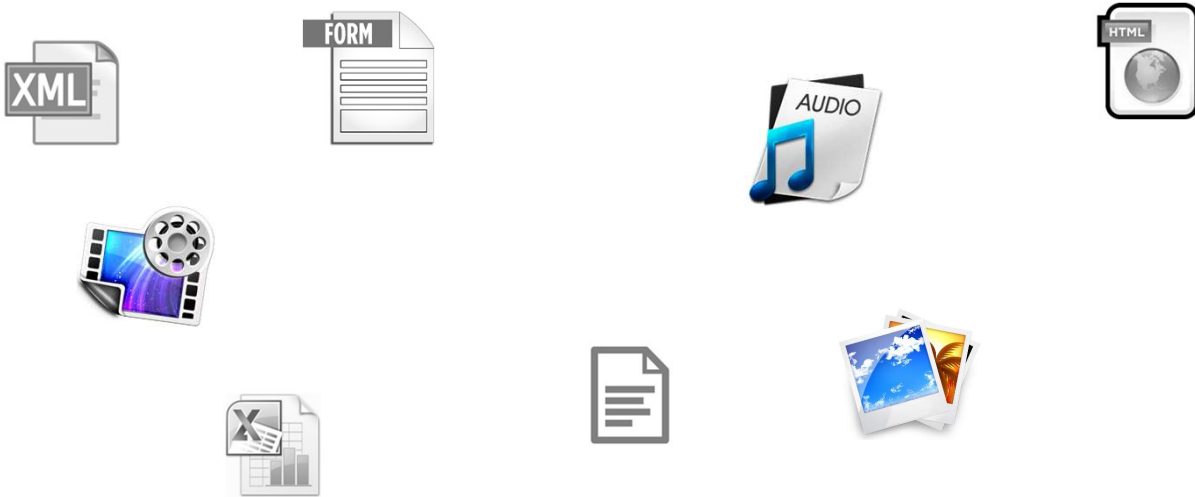
(DA SILVA, Leandro Augusto; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução à mineração de dados: com aplicações em R**. Elsevier Brasil, 2017, p10)

# Dados semiestruturados



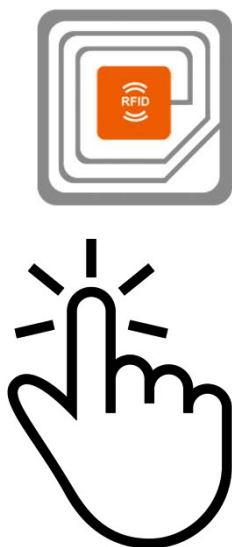
**DADOS SEMIESTRUTURADOS** – uma forma de dados estruturados que não se conforma à estrutura formal dos modelos de dados utilizados em bancos relacionais, **mas que contém tags ou outros marcadores que separam elementos semânticos e que garantem hierarquias de campos e registros dentro dos dados.** Geralmente armazenados em bancos de dados orientados a objetos. Exemplos: XML e JSON

## Dados desestruturados



**DADOS DESESTRUTURADOS** – dados que podem ser analisados mas que não se apresentam de forma estruturada. Podem ser transformados em dados estruturados por um processo que descreve os documentos desestruturados por suas características principais usando-se metadados.

## Eventos



```
hannsi21shop hanns tail -f -30 "/private/var/log/apache2/access_log"
127.0.0.1 - - [13/Jan/2017:15:13:41 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:13:46 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:13:46 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:13:56 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:13:56 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:13:57 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:13:57 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:13:58 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:13:58 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:26:02 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:02 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:26:03 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:03 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:05 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:09 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:09 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:26:09 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:09 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:26:09 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:26:09 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
127.0.0.1 - - [13/Jan/2017:15:26:29 +0000] "-" 405 -
127.0.0.1 - - [13/Jan/2017:15:27:17 +0000] "GET /admin/index.php?act=bill&op=show_status HTTP/1.1" 200 -
127.0.0.1 - - [13/Jan/2017:15:27:17 +0000] "GET /favicon.ico HTTP/1.1" 200 4286
```

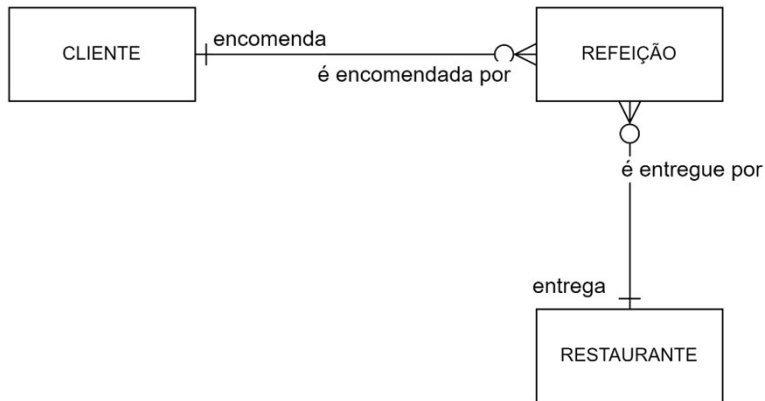
Acho que se você perguntasse a uma empresa 15 anos atrás quais dados ela possuía, ela descreveria seus dados transacionais, como usuários, produtos, pedidos e outros itens mantidos em tabelas em bancos de dados relacionais.

No entanto, nossa definição foi ampliada. Agora, a maioria das empresas também incluiria dados de eventos. Os dados de evento registram coisas que **acontecem** em vez de coisas que **existem**. Em sistemas da web, isso significa o registro da atividade do usuário, bem como os eventos e estatísticas no nível da máquina necessários para operar e monitorar de forma confiável o número de máquinas de um data center.

... Esses dados estão no cerne da Web moderna: a fortuna do Google, afinal, é gerada por um pipeline de relevância construído em cliques e impressões, ou seja, eventos.

KREPS, Jay. **I Heart Logs: Event Data, Stream Processing, and Data Integration.** "O'Reilly Media, Inc.", 2014, p. 13.

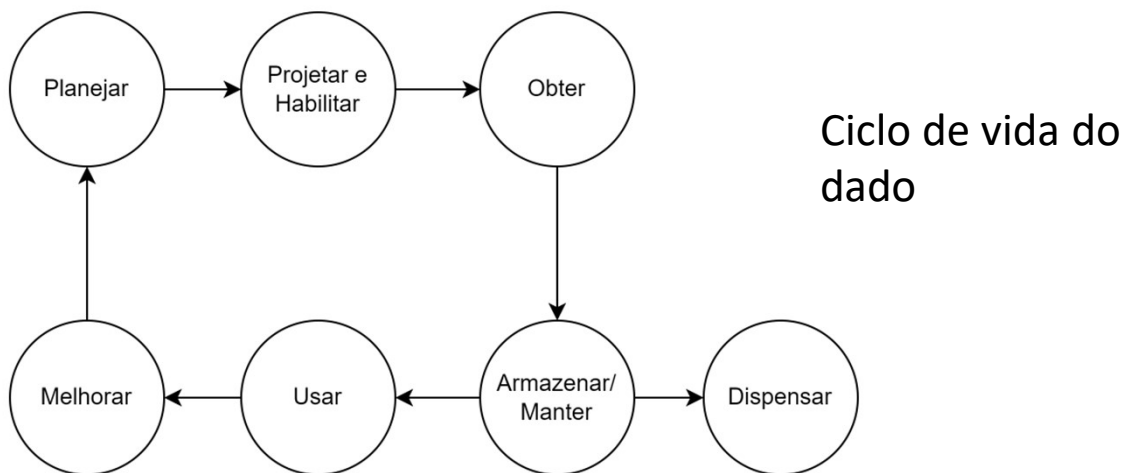
## Banco de dados: organização



Modelo lógico

Um **modelo lógico** pode ser definido como “aquele em que os objetos, suas características e relacionamentos têm a representação de acordo com as regras de implementação e limitantes impostos por algum tipo de tecnologia” (COUGO, 2013, p.29). O modelo lógico é obtido pela aplicação de regras de derivação a partir do modelo conceitual.

## Banco de dados: gerenciamento



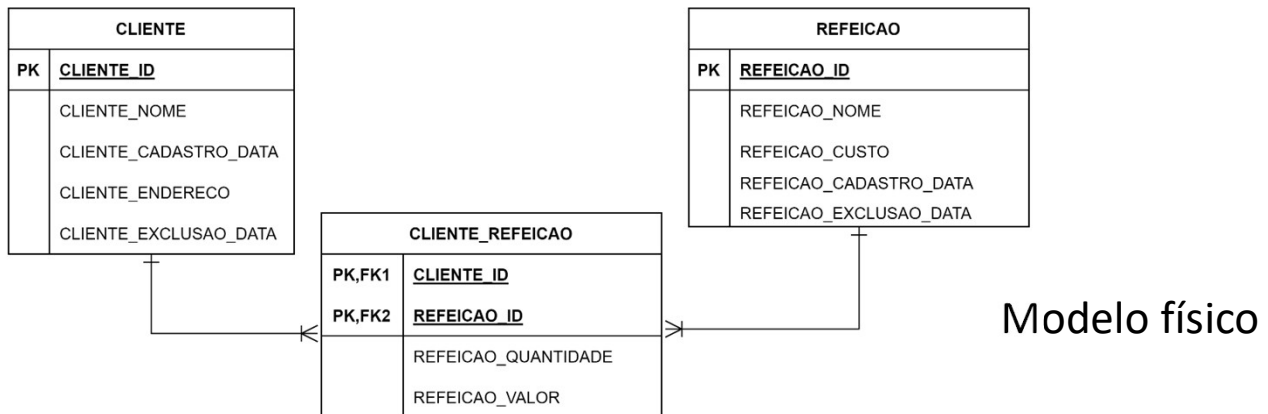
Assim como outros ativos de uma corporação, os dados têm um **ciclo de vida**, e por isso as corporações necessitam planejar usará os dados durante esse ciclo de vida.

Gerenciar um dado durante seu ciclo de vida significa ter processos bem definidos que **criem e obtenham** dados, outros processos que **movimentem, transformem, armazenem e compartilhem** os dados, outros processos que **usem** os dados, e, finalmente, processos para que os dados sejam **descartados** (DAMA, 2017, p.28). A **Governança de Dados** é a área de conhecimento que define e aplica as melhores práticas para o gerenciamento do ciclo de vida do



dado.

## Banco de dados: gerenciamento



Um **modelo físico** de dados pode ser definido como “aquele em que a representação dos objetos é feita sob o foco do nível físico de implementação das ocorrências, ou instâncias das entidades e seus relacionamentos” (COUGO, 2013, p.30). Produzir um modelo físico é representar os dados modelados considerando restrições da notação adotada na modelagem e restrições da tecnologia adotada no armazenamento dos dados.

## Elementos de sistemas de bancos de dados

- Hardware



O **hardware** pode ser definido como a parte física ou eletrônica de um computador ou outra parte de equipamento que não seja o **software** (HARDWARE, 2023). Nessa definição estão incluídos **dispositivos móveis pessoais** como celulares, relógios digitais e tablets, **desktops**, como notebooks e estações de trabalho, **servidores e clusters de servidores** (HENNESSY, 2019, p.5). É o **hardware** o responsável por

executar o processamento das operações de computação e o armazenamento dos dados. As atividades de processamento e armazenamento são executadas por componentes especializados dentro dos servidores: o processador, e os dispositivos de armazenamento (*hard disks* e *storages*).

## Elementos de sistemas de bancos de dados

- Hardware
- Software



Outro importante componente das soluções de sistema que armazenam os dados é o *software*. Em termos genéricos o software pode ser definido como a parte variável de um computador, oposta ao *hardware*, que é a parte física. Essa parte variável é composta por instruções, dados e programas que trabalham sobre o *hardware* para executar tarefas (SOFTWARE, 2023).

## Elementos de sistemas de bancos de dados

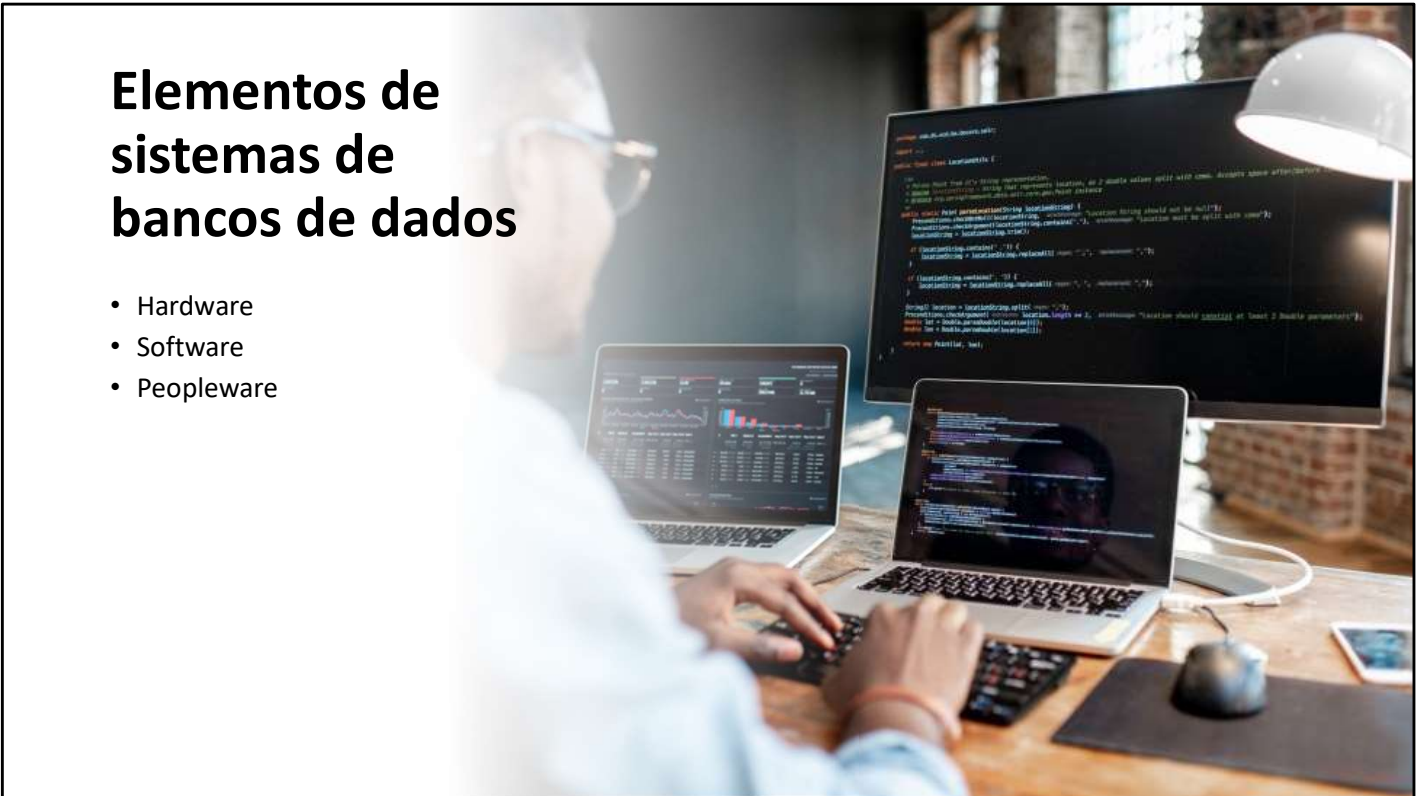
Softwares de Sistema	Softwares de Aplicação
Gerenciam processos e recursos básicos	Executam atividades específicas para o usuário
Escritos em linguagens de baixo nível, como Assembly	Escritos em linguagem de alto nível, como Python ou JavaScript
Interagem diretamente com o hardware e devem atender restrições impostas pelo hardware	Não interagem diretamente com o hardware
Instalados quando o Sistema Operacional é instalado	Instalados pelo administrador, de acordo com a necessidade
São iniciados quando o computador é ligado e permanecem ativos até que o computador seja desligado	São iniciados e finalizados pelo usuário
O usuário não os vê funcionando e nem interage com eles	O usuário interage diretamente com eles para executar tarefas específicas
Sua execução é independente	Dependem dos softwares de sistema
São necessários para que o computador seja iniciado	Não precisam estar instalados para que o computador funcione

Aos *softwares* que lidam diretamente com o *hardware* chamados de *softwares* de sistema. Esses *softwares* de sistema coordenam as atividades do *hardware* e fazem a comunicação do *hardware* com outros tipos de *software*. Acima da camada dos *softwares* de sistema operam outros tipos de *softwares* especializados que executam tarefas para atender a outros *softwares*, tais como *middlewares* e *drivers*, e também *softwares* que

executam funções para o atender ao usuário,  
chamados de *softwares* de aplicação.

## Elementos de sistemas de bancos de dados

- Hardware
- Software
- Peopleware



Embora seja um termo menos comum que os termos *hardware* e *software*, *peopleware* é o mais nobre entre os componentes das soluções de sistema que usam banco de dados. *Peopleware* corresponde ao componente humano e seus papéis na interação com o computador, tais como usuários finais, times de projeto, desenvolvedores, administradores de sistema, e outros (PEOPLEWARE, 2009).



## Camadas de uma aplicação de negócio



Em uma solução de aplicação que usa banco de dados, a camada de aplicação será a camada de *software* com a qual o usuário terá interface. Como definido anteriormente, aplicação é o termo mais simples para *software* de aplicação. No mundo corporativo as aplicações com as quais pessoas lidam dependerão da função exercida na empresa, ou do departamento em que se trabalha.

## Camadas de uma aplicação de negócio



O gerenciamento de um banco de dados armazenado em servidores centralizados ou distribuídos é feito com o apoio de um tipo específico de *software* de aplicação: o Sistema Gerenciador de Banco de Dados (SGDB). Os **Sistemas Gerenciadores de Bancos de Dados** proveem meios de **armazenar**, **manipular** e **recuperar** os dados, utilizando tecnologias e soluções que atendam aos usuários e aplicações

com um desempenho aceitável.

## Camadas de uma aplicação de negócio



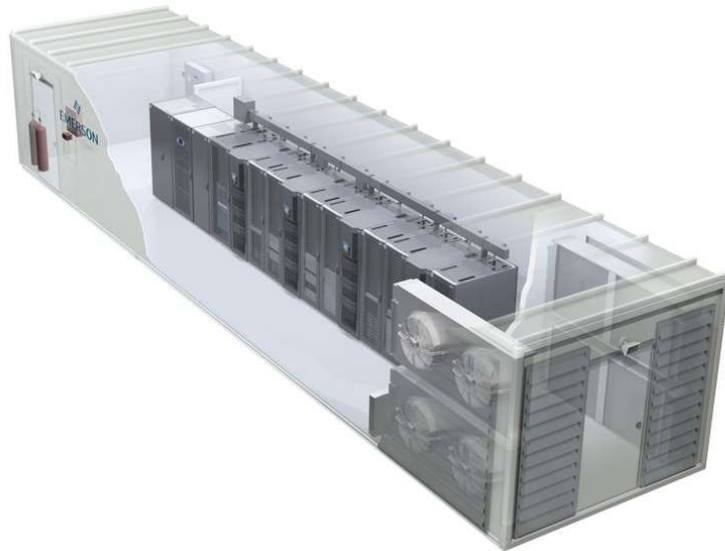
Abaixo dos *softwares* de aplicação está uma categoria de *software* de sistema conhecida como Sistema Operacional. Os **Sistemas Operacionais** exercem um papel essencial em um computador: eles abstraem a complexidade de lidar diretamente com o *hardware* do computador. Os sistemas operacionais são compostos por outros sistemas menores especializados que usam os componentes do *hardware* de maneira otimizada e que oferecem

aos *softwares* de aplicação uma maneira mais limpa e simples de gerenciar esses recursos computacionais (TANENBAUM, 2016, p.21).

## Camadas de uma aplicação de negócio



## Camadas de uma aplicação de negócio



Container Data Center



## Dispositivos de armazenamento de dados



## Dispositivos de armazenamento de dados

- Retêm os dados
- Direct Area Storage (DAS)
  - Conectados diretamente ao servidor
    - HDD, SDD, DVD, CD, BLU-RAY



A maneira como os dispositivos de armazenamento serão acessados pode ser categorizada entre acesso direto e acesso por rede. Dispositivos acessados diretamente são chamados de **Direct Area Storage (DAS)**. Os dispositivos DAS estão conectados diretamente ao computador que precisa acessá-los e em geral são acessados apenas pelo computador ao qual estão conectados. Exemplos desse tipo de dispositivo são: dispositivos

de disquete, de CD e de DVD, HDD (*hard disk drives*), dispositivos flash disco SSD (*solid-state drives*)

## Dispositivos de armazenamento de dados

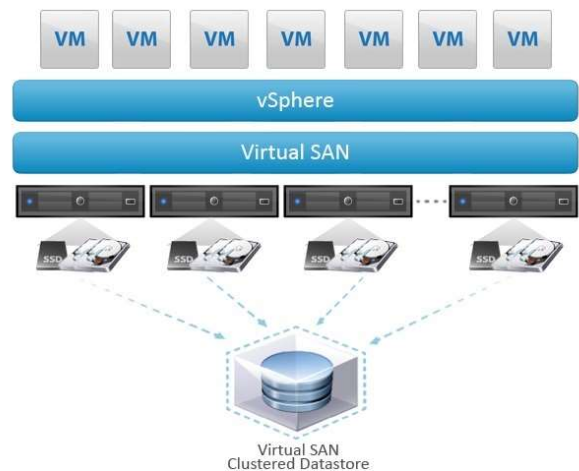
- Retêm os dados
- Direct Area Storage (DAS)
- Network-attached storage (NAS)
  - Acessível por rede
  - Disponível para vários servidores
  - Mantêm cópias redundantes dos dados



Na configuração NAS um dispositivo de armazenamento composto por múltiplos discos é disponibilizado para ser acessado por servidores por meio de uma rede. Nesse tipo de dispositivo os discos são organizados de forma a manterem cópias redundantes dos dados, evitando assim que a perda de um ou alguns desses discos impliquem a perda de todos os dados.

## Dispositivos de armazenamento de dados

- Retêm os dados
- Direct Area Storage (DAS)
- Network-attached storage (NAS)
- Storage Area Network (SAN)
  - Múltiplos dispositivos espalhados em uma rede
  - Vistos como um único storage



Fonte: <https://masteringvmware.com/vsan-what-is-vmware-vsan/>

Já na configuração SAN múltiplos dispositivos espalhados em uma rede são utilizados para armazenamento dos dados como se fossem um único dispositivo. Essa solução permite a combinação de diferentes tecnologias de armazenamento, tais como discos SSD e *flash*, *appliances* e até mesmo armazenamento em nuvem (WHAT IS DATA STORAGE, 2023). Um exemplo de uso desse tipo de armazenamento é o

VMWare

VSAN

(<https://www.vmware.com/br/products/vsan.html>).



# **Nuvem computacional e Tecnologia on-premises**

## Computação local

- On-premises
- Infraestrutura mantida e administrada pela corporação
  - Datacenter
  - Mão-de-obra
  - Manutenção e investimentos
- Custo fixo



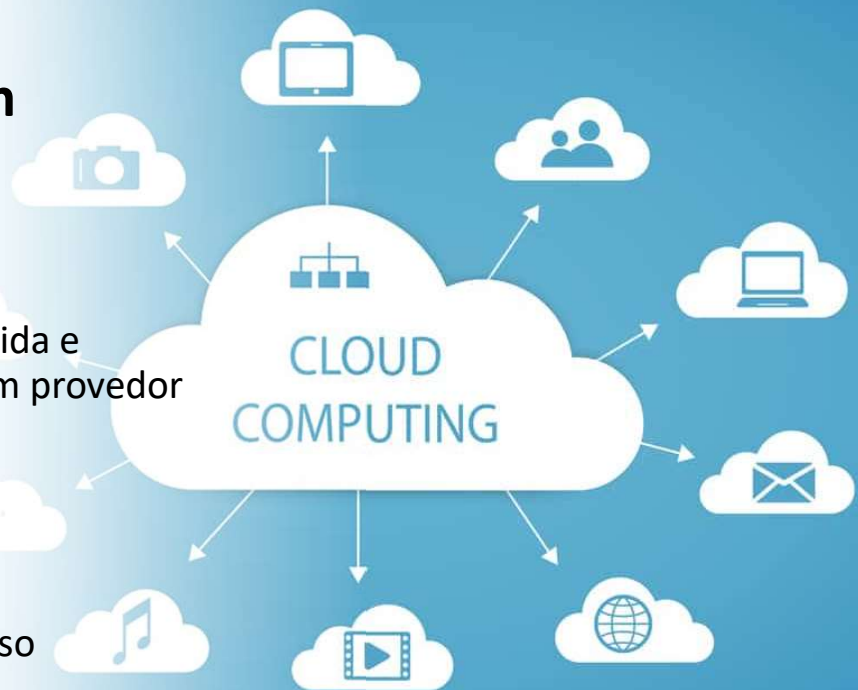
No conceito de **computação local**, também chamada de **computação *on-premises***, os recursos computacionais estão localizados em uma infraestrutura mantida e administrada pela corporação. Nesse modelo, a empresa que utiliza os recursos computacionais e de armazenamento é a responsável pela hospedagem dos servidores e pela sua administração, incluindo os custos de energia elétrica, de segurança das instalações, de

cópias dos dados, de manutenção dos servidores, dos profissionais que administram os servidores e seus aplicativos e por todos os investimentos necessários para manter e expandir a infraestrutura (WIKIPEDIA, 2023). No modelo de computação local, a corporação é responsável por administrar a falta e o excedente de recursos computacionais.



## Computação em nuvem

- Cloud computing
- Infraestrutura mantida e administrada por um provedor
  - Datacenter
  - Mão-de-obra
  - Manutenção e investimentos
- Custo conforme o uso



No modelo de **computação em nuvem** os recursos computacionais são fornecidos à corporação por um provedor de serviços em nuvem. Nesse modelo a responsabilidade pelos custos operacionais e pelos investimentos é do provedor de serviços, que cobra pelos recursos fornecidos aos clientes **conforme o uso**. Isso significa que a falta e o excedente de recursos computacionais serão gerenciados pelo provedor: na falta de

recursos computacionais, novos recursos são fornecidos sob demanda; se os recursos computacionais não forem mais necessários, os recursos serão liberados.

## **Tipos de computação em nuvem**

- **Infraestrutura como serviço (IaaS)**
  - Como se fosse a infra do cliente
  - Configurações e atualizações são do cliente
  - Responsabilidade pela infra é do provedor
- Exemplo: banco de dados instalado em um servidor Linux hospedado em nuvem

Na **Infraestrutura como serviço (IaaS)** os recursos computacionais e de armazenamento são disponibilizados sob demanda, e utilizados da mesma forma como se estivessem armazenados na infraestrutura do cliente. O cliente responsabiliza-se pelas configurações e pelas atualizações, e o provedor de serviço em nuvem responsabiliza-se pela infraestrutura física.

## Tipos de computação em nuvem

- Infraestrutura como serviço (IaaS)
- **Plataforma como serviço (PaaS)**
  - “Serviços totalmente gerenciados”
  - “Serverless”
  - Infra, SO, pacotes, atualizações são responsabilidade do provedor

Exemplos: banco de dados totalmente gerenciado (Aurora DB, MS SQL Server)

Na **Plataforma como serviço (PaaS)**, a gestão da infraestrutura subjacente ao serviço é de responsabilidade do provedor de serviços em nuvem. Nesse modelo, atividades como instalação e manutenção de recursos computacionais e de armazenamento são de responsabilidade do provedor. Os serviços PaaS também conhecidos como “serviços totalmente gerenciados” (pelo provedor).

## Tipos de computação em nuvem

- Infraestrutura como serviço (IaaS)
- Plataforma como serviço (PaaS)
- **Software como serviço (SaaS)**
  - Cliente usa a aplicação em nuvem

No modelo de **Software como serviço (SaaS)** o cliente preocupa-se apenas em usar o serviço provido, cabendo ao provedor manter a aplicação e todas as camadas subjacentes. Esse modelo aplica-se principalmente a aplicações para usuários finais (AWS, 2023).

A close-up photograph of a server rack. The image shows several vertical server units. Some units have glowing green lights, while others have glowing red lights. The background is dark and out of focus, emphasizing the lights and the structure of the rack.

# Governança de Dados

## Administração ativa do dado

- Dado visto como ativo
- Requer estratégia (escolhas e decisões tendo em vista um objetivo)

Os dados tornaram-se recursos tão importantes quanto os financeiros ou os humanos para os objetivos corporativos, necessitando ser **administrados ativamente** da mesma forma como pessoas e finanças precisam (DAMA, 2017).

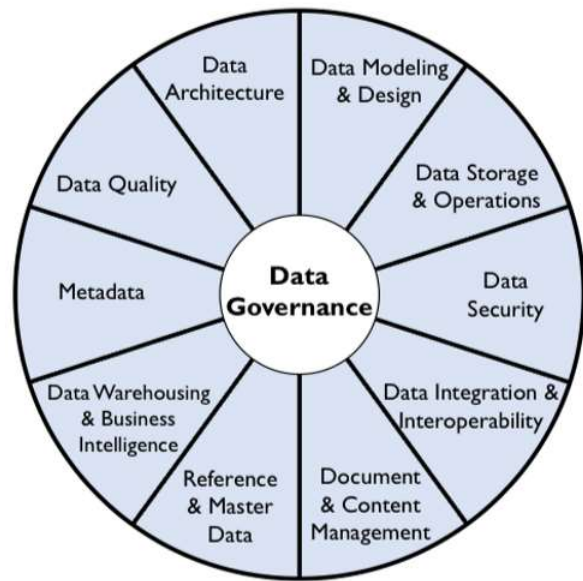
A administração ativa do dado inclui conhecer quais dados a corporação possui, o que pode ser alcançado com esses dados e como usar bem os dados para alcançar os objetivos da corporação. A

administração ativa do dado envolve uma **estratégia** para a administração do dado, isto é, **escolhas e decisões que traçam um curso de ação para a obtenção de um objetivo.**



## Administração ativa do dado

- Modelo de maturidade
  - Indica a maturidade em gestão do dado
  - Níveis de progressão
  - Exemplo: DAMA-DMBOK



Copyright © 2017 DAMA International

## Referências e leituras recomendadas

TEJADA, Zoiner. **Mastering Azure Analytics: Architecting in the Cloud with Azure Data Lake, HDInsight, and Spark.** O'Reilly Media, Inc., 2017.

<https://jornfranke.wordpress.com/2014/07/20/the-lambda-architecture-for-big-data-in-your-enterprise/>

<https://jornfranke.wordpress.com/2016/11/11/lambda-kappa-microservice-and-enterprise-architecture-for-big-data/>

MARZ, N.; WARREN, J. **Big Data: Principles and best practices of scalable realtime data systems**. [S.l.]: Manning Publications Co., 2015.

VANHOVE, T. et al. Managing the synchronization in the lambda architecture for optimized big data analysis. *IEICE Transactions on Communications*, The Institute of Electronics, Information and Communication Engineers, v. 99, n. 2, p. 297–306, 2016.

<https://blog.acolyer.org/2015/02/04/liquid-unifying-nearline-and-offline-big-data-integration/>

[http://lsds.doc.ic.ac.uk/sites/default/files/CIDR15\\_Paper25u.pdf](http://lsds.doc.ic.ac.uk/sites/default/files/CIDR15_Paper25u.pdf)



**Obrigado !**

renemendes@yahoo.com.br  
[www.linkedin.com/in/renemendes](http://www.linkedin.com/in/renemendes)