



Formação Engenheiro de Dados

Estruturas de Dados

Estruturas de Dados

- ◊ Dados Estruturados
- ◊ Dados Semiestruturados
- ◊ Dados não Estruturados



Dados Estruturados

- ◆ Armazenado em uma Mesma Estrutura
- ◆ Estrutura Rígida
- ◆ Exemplo: Tabela de SGBD, Planilha Eletrônica

CHEQUEESPECIAL	USO_CREDITO	HISTORICO_CREDITO	PROPOSITO	BALANCO_ATUAL	BALANCO_MEDIO_CREDITO	EMPREGADO	LOCAL
<0	6	crítica / outros de crédito existente	radio/tv	1169	nenhuma economia conhecida	>=7	4
0<=X<200	48	existente e pago	radio/tv	5951	<100	1<=X<4	2
sem cheque	12	crítica / outros de crédito existente	educação	2096	<100	4<=X<7	2
<0	42	existente e pago	mobilia/equipamento	7882	<100	4<=X<7	2
<0	24	atrasos anteriores	carro novo	4870	<100	1<=X<4	3
sem cheque	36	existente e pago	educação	9055	nenhuma economia conhecida	1<=X<4	2
sem cheque	24	existente e pago	mobilia/equipamento	2835	500<=X<1000	>=7	3
0<=X<200	36	existente e pago	carro usado	6948	<100	1<=X<4	2
sem cheque	12	existente e pago	radio/tv	3059	>=1000	4<=X<7	2
0<=X<200	30	crítica / outros de crédito existente	carro novo	5234	<100	desempregado	4
0<=X<200	12	existente e pago	carro novo	1295	<100	<1	3
<0	48	existente e pago	negócios	4308	<100	<1	3
0<=X<200	12	existente e pago	radio/tv	1567	<100	1<=X<4	1
<0	24	crítica / outros de crédito existente	carro novo	1199	<100	>=7	4

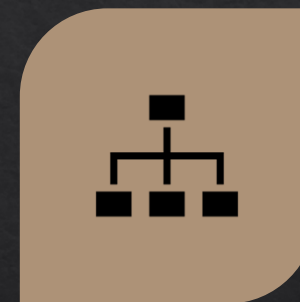
Semiestruturados



ESTRUTURA HETEROGÊNEA



NÃO ESTÃO COMPLETAMENTE
NÃO ESTRUTURADOS. EXISTE
UMA ESTRUTURA
REPRESENTADA



EVOLUTIVA: MAIS FÁCIL SUA
ESTRUTURA MUDAR

Semiestrutturados

- ◊ XML
- ◊ JSON
- ◊ RDF – Resource Description Framework
- ◊ OWL – Web Ontology Language

Lista de Alunos

```
{"Alunos": [
  { "nome": "Edson Sales Arantes", "notas": [ 8, 9, 5 ] },
  { "nome": "Luiz Livelli ", "notas": [ 8, 10, 7 ] },
  { "nome": "Caique Caicedo De Plata", "notas": [ 10, 10, 9 ] }
]}
```

Não Estruturados



Sem estrutura Definida, nem de forma implícita



Sem metadados ou descrição da estrutura



Maioria dos dados existentes (80 a 90%)



Formato que mais cresce

Não Estruturados

- ◊ Vídeos
- ◊ Imagens
- ◊ Páginas de Internet
- ◊ Recibos Eletrônicos
- ◊ E-mails
- ◊ Tweets
- ◊ Documentos em Geral



Barack Obama ✓

@BarackObama

 Seguir

Four more years. pic.twitter.com/bAJE6Vom

Formatos



CHAVE-VALOR



TEXTO: EX: CSV



ORC: ARQUIVO EM
COLUNAS OU
LINHAS, CHAVE-
VALOR



AVRO:
SERIALIZADO



PARQUET:
COLUNAS,
ANINHADAS

Ambiente
Transacional

Arquitetura de Dados

Ambiente
Analítico

Produção e
Guarda

Registrar
Fatos

Recuperar
Registros

Armazenamento
e Análise

Manter
Histórico

Analisar
Dados

Não há
produção

