

TECNOLOGIAS DE PROGRAMAÇÃO DE SISTEMAS DE
INFORMAÇÃO

FORMATOS DE DADOS

SISTEMAS GESTORES DE BASES DE DADOS II | Prof. Magno
Andrade | Prof. Jorge Louro

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Conteúdo da apresentação**
 - Formato de dados:
 - XML
 - JSON
 - BSON

FORMATO DE DADOS - XML

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- XML

- Introdução

- Significa *Extensible Markup Language*.
 - Representação de dados semi estruturados.
 - Derivado do *SGML (Standard Generalized Markup Language)*.
 - Os objectivos são:
 - Simplicidade, generalidade e usabilidade.
 - Tipo de extensão: **.xml, content type: text/xml*



UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- XML
 - Exemplo

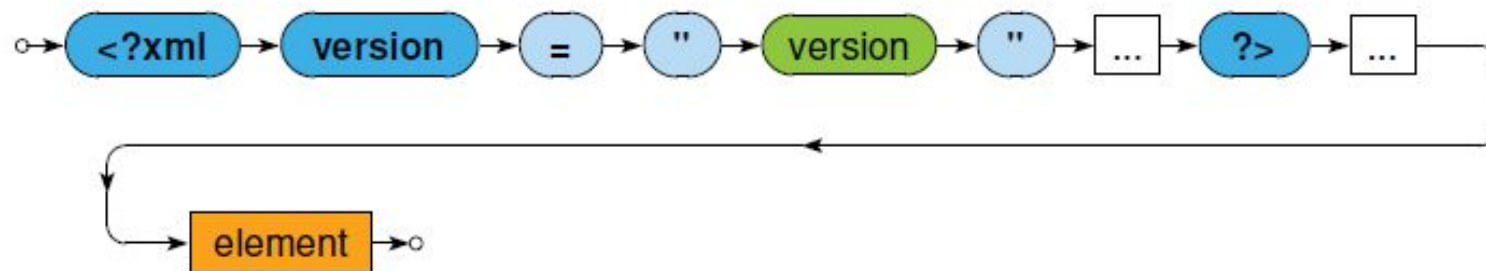
```
<?xml version="1.0" standalone="yes" ?>
- <shop location="Birmingham" size="Large">
- <food>
  <Name>Apple</Name>
  <type>fruit</type>
  <cost>15</cost>
</food>
- <food>
  <Name>Carrot</Name>
  <type>vegetable</type>
  <cost>10</cost>
</food>
</shop>
```


UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- XML

- Estrutura do documento

- Apenas um elemento raiz.
 - Contém um ou mais elementos “aninhados”/filhos.



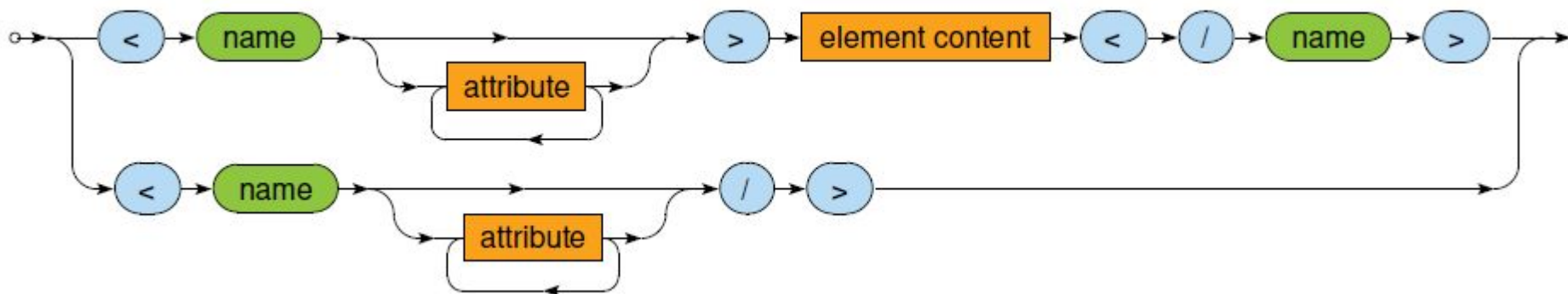
```
<?xml version="1.1" encoding="UTF-8"?>
<movie>
  ...
</movie>
```

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Construção**

- Elemento

- Feito com etiquetas de início e fim.
 - Cada elemento pode ter um **conjunto** de atributos.

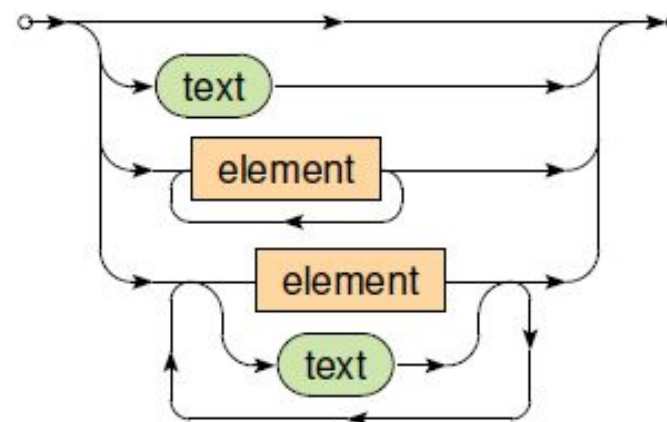


UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Construção**

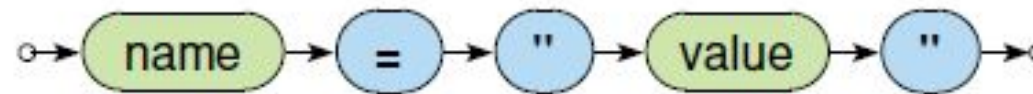
- Tipos de conteúdo de elementos

- Conteúdo vazio
- Conteúdo de texto.
- Conteúdo de elementos.
 - Uma sequência de elementos encaixados/"aninhados".
- Conteúdo misto.



UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Construção**
 - Atributos
 - Par nome-valor



- *Escaping sentences*
 - Utilizado nos valores dos atributos ou no texto do conteúdo de elementos.
 - Ex: *"* para " e *<* para <
- Exemplo
 - `<EscapeNeeded name="Pete "Maverick" Mitchell"/>`

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Conclusão**
 - Construção
 - Principais: elemento, atributo e texto
 - Esquemas de linguagem
 - *DTD, XSD (XML Schema), RELAX NG, Schematron*
 - Linguagens de pesquisa
 - *XPath, XQuery, XSLT*
 - Formato XML
 - *XSD, XSLT, XHTML, DocBook, SOAP, etc...*

FORMATO DE DADOS - JSON

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **JSON**

- Introdução

- Significa *JavaScript Object Notation*.
 - Padrão aberto para o transporte de dados.
 - Derivado do *JavaScript* mas independente.
 - Os objectivos são:
 - Simplicidade: baseado em texto, fácil de escrever e ler.
 - Universalidade: suportado pela maioria das linguagens de programação.
 - Tipo de extensão: **.json*, *content type: application/json*



UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- JSON
 - Exemplo

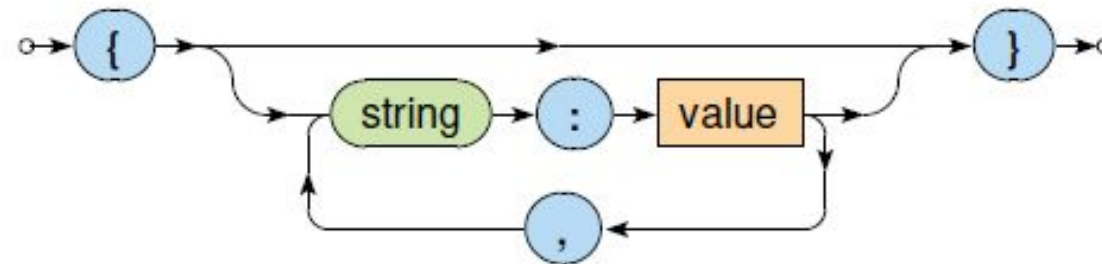
```
1  {  
2    "title" : "Rambo IV",  
3    "year" : 2008,  
4    "actors" : [  
5      {  
6        "firstname" : "Sylvester",  
7        "lastname" : "Stallone"  
8      },  
9      {  
10       "firstname" : "Julie",  
11       "lastname" : "Benz"  
12     }  
13   ],  
14   "director" : {  
15     "firstname" : "Sylvester",  
16     "lastname" : "Stallone"  
17   }  
18 }
```


UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Estrutura do documento**

- **Objecto**

- Colecção desordenada de pares nome:valor (propriedades)
 - Corresponde a estruturas, que podem ser objectos, *arrays*, listas, *arrays* associativos.
 - Os valores podem ser de tipos diferentes, mas os nomes devem ser únicos.



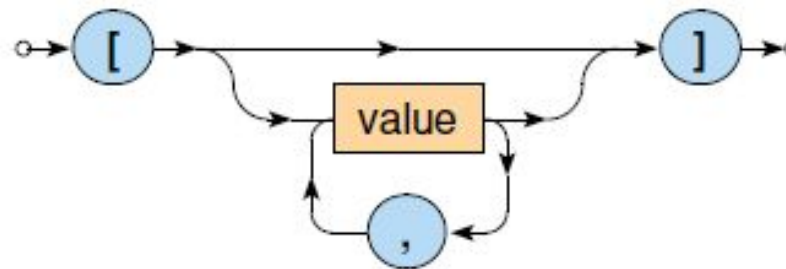
```
1 { "title" : "Rambo IV", "year" : 2008 }
```

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Estrutura do documento**

- *Array*

- Colecção ordenada de valores
 - Corresponde a estruturas que podem ser *arrays*, vectores, listas, sequências.
 - Os valores podem ser de tipos diferentes e valores duplicados são permitidos.



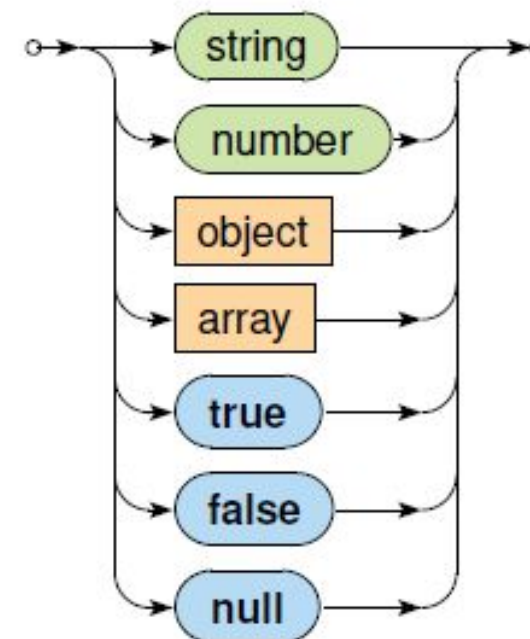
[2 , 7 , 10 , 12]
["Rambo" , 2008]

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Estrutura do documento**

- Valor

- *String Unicode*
 - Utilizado com aspas duplas
 - Exemplo: "John"
 - Número
 - Inteiro ou decimal
 - Exemplo: 1, 1.5
 - Objecto encaixado/"aninhado"
 - *Array* encaixado/"aninhado"
 - Valor booleano: *true*, *false*



UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Conclusão**
 - Construção
 - Colecções: *objecto*, *array*.
 - Valores escalares: *string*, número, booleano, *null*
 - Esquemas de linguagem
 - *JSON Schema*
 - Linguagens de pesquisa
 - *JSONiq*, *JMESPath*

FORMATO DE DADOS - BSON

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **BSON**

- Introdução

- Significa *Binary JSON*.
 - Codificação binária de documentos JSON.
 - Estende os tipos de dados que existem no JSON.
 - Utilizado pelo MongoDB.
 - As características são:
 - Leve, eficiente, fácil de percorrer.
 - Tipo de extensão: **.bson*

BSON is binary JSON



BSON is a JSON that has been serialized as a binary document.

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **BSON**
 - Exemplo

JSON

```
{  
  "title" : "Medvídek",  
  "year" : 2007  
}
```

BSON

```
24 00 00 00  
02 74 69 74 6C 65 00 0A 00 00 00 4D 65 64 76 C3 AD 64 65 6B 00  
10 79 65 61 72 00 D7 07 00 00  
00
```

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

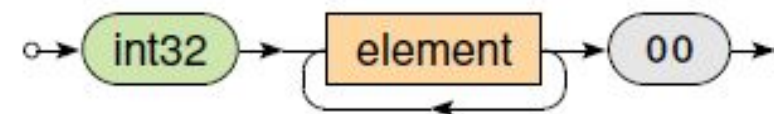
● Estrutura do documento

- Documento = serialização de um objecto JSON ou *array*
 - O objecto JSON é serializado directamente.
 - Os *arrays* em JSON são convertidos primeiro para um objecto.

■ `["Rambo", 2008] ->`
`{ "0" : "Rambo", "1" : "2008" }`

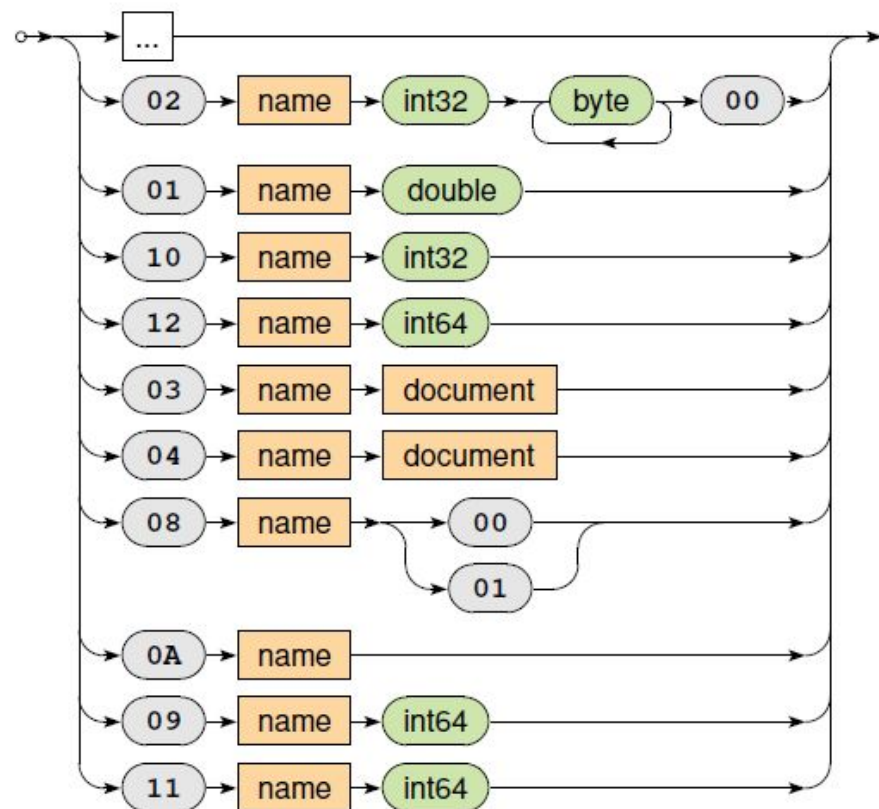
- Estrutura

- Tamanho do documento (em *bytes*)
- Sequência de elementos (codificados em propriedades JSON)
- Terminando em hexadecimal 00 *byte*.



UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Estrutura do documento**
 - Elemento = serialização de uma propriedade JSON



UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Elemento = serialização de uma propriedade JSON**

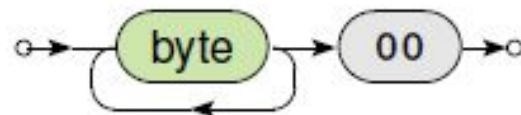
- Estrutura

- Selecção do tipo

- 02 (*string*)
- 01 (*double*), 10 (*32-bit integer*), 12 (*64-bit integer*)
- 03 (*object*), 04 (*array*)

- Nome da propriedade em *unicode*

- *String unicode* termina com 00



- Valor da propriedade

UNIDADE CURRICULAR : SISTEMAS GESTORES DE BASES DE DADOS II

- **Referências**

- Martin Svoboda, «Data Formats». 10-Out-2017.



CTeSP

CURSOS TÉCNICOS
SUPERIORES PROFISSIONAIS