Week 7 Assignment

Jiaxin Zheng

2024-10-19

```r
#create the data frame
data <- data.frame(
  Category = c("Electronics", "Electronics", "Electronics", "Electronics",
               "Home Appliances", "Home Appliances", "Home Appliances", "Home Appliances",
               "Clothing", "Clothing", "Clothing", "Clothing",
               "Books", "Books", "Books", "Books",
               "Sports Equipment", "Sports Equipment"),
  Item_Name = c("Smartphone", "Smartphone", "Laptop", "Laptop",
                "Refrigerator", "Refrigerator", "Washing Machine", "Washing Machine",
                "T-Shirt", "T-Shirt", "T-Shirt", "Jeans",
                "Fiction Novel", "Fiction Novel", "Non-Fiction Guide", "Non-Fiction Guide",
                "Basketball", "Tennis Racket"),
  Item_ID = c(101, 101, 102, 102, 201, 201, 202, 202, 301, 301, 301, 302, 401, 401, 402, 402, 501, 502)
  Brand = c("TechBrand", "TechBrand", "CompuBrand", "CompuBrand",
            "HomeCool", "HomeCool", "CleanTech", "CleanTech",
            "FashionCo", "FashionCo", "FashionCo", "DenimWorks",
            "-", "-", "-", "-",
            "SportsGear", "RacketPro"),
  Price = c(699.99, 699.99, 1099.99, 1099.99, 899.99, 899.99, 499.99, 499.99,
            19.99, 19.99, 19.99, 49.99, 14.99, 14.99, 24.99, 24.99, 29.99, 89.99),
  Variation_ID = c("101-A", "101-B", "102-A", "102-B",
                   "201-A", "201-B", "202-A", "202-B",
                   "301-A", "301-B", "301-C", "302-A",
                   "401-A", "401-B", "402-A", "402-B",
                   "501-A", "502-A"),
  Variation_Details = c("Color: Black, Storage: 64GB", "Color: White, Storage: 128GB",
                        "Color: Silver, Storage: 256GB", "Color: Space Gray, Storage: 512GB",
                        "Color: Stainless Steel, Capacity: 20 cu ft", "Color: White, Capacity: 18 cu ft"
                        "Type: Front Load, Capacity: 4.5 cu ft", "Type: Top Load, Capacity: 5.0 cu ft",
                        "Color: Blue, Size: S", "Color: Red, Size: M", "Color: Green, Size: L",
                        "Color: Dark Blue, Size: 32",
                        "Format: Hardcover, Language: English", "Format: Paperback, Language: Spanish",
                        "Format: eBook, Language: English", "Format: Paperback, Language: French",
                        "Size: Size 7, Color: Orange", "Material: Graphite, Color: Black")
)

library(jsonlite)


json_data <- toJSON(data, pretty = TRUE)
write(json_data, file = "json_data.json")
```

```python
# Save the json document to a file
print(json_data)
```

```
## [
##   {
##     "Category": "Electronics",
##     "Item_Name": "Smartphone",
##     "Item_ID": 101,
##     "Brand": "TechBrand",
##     "Price": 699.99,
##     "Variation_ID": "101-A",
##     "Variation_Details": "Color: Black, Storage: 64GB"
##   },
##   {
##     "Category": "Electronics",
##     "Item_Name": "Smartphone",
##     "Item_ID": 101,
##     "Brand": "TechBrand",
##     "Price": 699.99,
##     "Variation_ID": "101-B",
##     "Variation_Details": "Color: White, Storage: 128GB"
##   },
##   {
##     "Category": "Electronics",
##     "Item_Name": "Laptop",
##     "Item_ID": 102,
##     "Brand": "CompuBrand",
##     "Price": 1099.99,
##     "Variation_ID": "102-A",
##     "Variation_Details": "Color: Silver, Storage: 256GB"
##   },
##   {
##     "Category": "Electronics",
##     "Item_Name": "Laptop",
##     "Item_ID": 102,
##     "Brand": "CompuBrand",
##     "Price": 1099.99,
##     "Variation_ID": "102-B",
##     "Variation_Details": "Color: Space Gray, Storage: 512GB"
##   },
##   {
##     "Category": "Home Appliances",
##     "Item_Name": "Refrigerator",
##     "Item_ID": 201,
##     "Brand": "HomeCool",
##     "Price": 899.99,
##     "Variation_ID": "201-A",
##     "Variation_Details": "Color: Stainless Steel, Capacity: 20 cu ft"
##   },
##   {
##     "Category": "Home Appliances",
##     "Item_Name": "Refrigerator",
##     "Item_ID": 201,
```

```
##       "Brand": "HomeCool",
##       "Price": 899.99,
##       "Variation_ID": "201-B",
##       "Variation_Details": "Color: White, Capacity: 18 cu ft"
##     },
##     {
##       "Category": "Home Appliances",
##       "Item_Name": "Washing Machine",
##       "Item_ID": 202,
##       "Brand": "CleanTech",
##       "Price": 499.99,
##       "Variation_ID": "202-A",
##       "Variation_Details": "Type: Front Load, Capacity: 4.5 cu ft"
##     },
##     {
##       "Category": "Home Appliances",
##       "Item_Name": "Washing Machine",
##       "Item_ID": 202,
##       "Brand": "CleanTech",
##       "Price": 499.99,
##       "Variation_ID": "202-B",
##       "Variation_Details": "Type: Top Load, Capacity: 5.0 cu ft"
##     },
##     {
##       "Category": "Clothing",
##       "Item_Name": "T-Shirt",
##       "Item_ID": 301,
##       "Brand": "FashionCo",
##       "Price": 19.99,
##       "Variation_ID": "301-A",
##       "Variation_Details": "Color: Blue, Size: S"
##     },
##     {
##       "Category": "Clothing",
##       "Item_Name": "T-Shirt",
##       "Item_ID": 301,
##       "Brand": "FashionCo",
##       "Price": 19.99,
##       "Variation_ID": "301-B",
##       "Variation_Details": "Color: Red, Size: M"
##     },
##     {
##       "Category": "Clothing",
##       "Item_Name": "T-Shirt",
##       "Item_ID": 301,
##       "Brand": "FashionCo",
##       "Price": 19.99,
##       "Variation_ID": "301-C",
##       "Variation_Details": "Color: Green, Size: L"
##     },
##     {
##       "Category": "Clothing",
##       "Item_Name": "Jeans",
##       "Item_ID": 302,
```

```
##       "Brand": "DenimWorks",
##       "Price": 49.99,
##       "Variation_ID": "302-A",
##       "Variation_Details": "Color: Dark Blue, Size: 32"
##     },
##     {
##       "Category": "Books",
##       "Item_Name": "Fiction Novel",
##       "Item_ID": 401,
##       "Brand": "-",
##       "Price": 14.99,
##       "Variation_ID": "401-A",
##       "Variation_Details": "Format: Hardcover, Language: English"
##     },
##     {
##       "Category": "Books",
##       "Item_Name": "Fiction Novel",
##       "Item_ID": 401,
##       "Brand": "-",
##       "Price": 14.99,
##       "Variation_ID": "401-B",
##       "Variation_Details": "Format: Paperback, Language: Spanish"
##     },
##     {
##       "Category": "Books",
##       "Item_Name": "Non-Fiction Guide",
##       "Item_ID": 402,
##       "Brand": "-",
##       "Price": 24.99,
##       "Variation_ID": "402-A",
##       "Variation_Details": "Format: eBook, Language: English"
##     },
##     {
##       "Category": "Books",
##       "Item_Name": "Non-Fiction Guide",
##       "Item_ID": 402,
##       "Brand": "-",
##       "Price": 24.99,
##       "Variation_ID": "402-B",
##       "Variation_Details": "Format: Paperback, Language: French"
##     },
##     {
##       "Category": "Sports Equipment",
##       "Item_Name": "Basketball",
##       "Item_ID": 501,
##       "Brand": "SportsGear",
##       "Price": 29.99,
##       "Variation_ID": "501-A",
##       "Variation_Details": "Size: Size 7, Color: Orange"
##     },
##     {
##       "Category": "Sports Equipment",
##       "Item_Name": "Tennis Racket",
##       "Item_ID": 502,
```

```
##      "Brand": "RacketPro",
##      "Price": 89.99,
##      "Variation_ID": "502-A",
##      "Variation_Details": "Material: Graphite, Color: Black"
##    }
## ]
```

#JSON- * 1.Pros: Easy to read, and use. JSON support most of the modern programming languages.More compact than XML. * 2. Cons: Not optimized for large data, not as efficient when handling large datasets.

```r
library(knitr)
html_data <- kable(data, format = "html")

# Save the html document to a file
write(html_data, file = "html_data.html")

html_data
```

Category

Item_Name

Item_ID

Brand

Price

Variation_ID

Variation_Details

Electronics

Smartphone

101

TechBrand

699.99

101-A

Color: Black, Storage: 64GB

Electronics

Smartphone

101

TechBrand

699.99

101-B

Color: White, Storage: 128GB

Electronics

Laptop

102

CompuBrand

1099.99

102-A

Color: Silver, Storage: 256GB

Electronics

Laptop

102

CompuBrand

1099.99

102-B

Color: Space Gray, Storage: 512GB

Home Appliances

Refrigerator

201

HomeCool

899.99

201-A

Color: Stainless Steel, Capacity: 20 cu ft

Home Appliances

Refrigerator

201

HomeCool

899.99

201-B

Color: White, Capacity: 18 cu ft

Home Appliances

Washing Machine

202

CleanTech

499.99

202-A

Type: Front Load, Capacity: 4.5 cu ft

Home Appliances

Washing Machine

202

CleanTech

499.99

202-B

Type: Top Load, Capacity: 5.0 cu ft

Clothing

T-Shirt

301

FashionCo

19.99

301-A

Color: Blue, Size: S

Clothing

T-Shirt

301

FashionCo

19.99

301-B

Color: Red, Size: M

Clothing

T-Shirt

301

FashionCo

19.99

301-C

Color: Green, Size: L

Clothing

Jeans

302

DenimWorks

49.99

302-A

Color: Dark Blue, Size: 32

Books

Fiction Novel

401

- 14.99

  401-A

  Format: Hardcover, Language: English

  Books

  Fiction Novel

  401

  - 14.99

    401-B

    Format: Paperback, Language: Spanish

    Books

    Non-Fiction Guide

    402

    * 24.99

      402-A

      Format: eBook, Language: English

      Books

      Non-Fiction Guide

      402

      · 24.99

        402-B

        Format: Paperback, Language: French

        Sports Equipment

        Basketball

        501

        SportsGear

        29.99

        501-A

        Size: Size 7, Color: Orange

        Sports Equipment

        Tennis Racket

        502

        RacketPro

        89.99

        502-A

        Material: Graphite, Color: Black

        #HTML- * 1. Pros: Like JOSON, HTML is easy to read too. Can be interactive with CSS / JavaScript. Designed can be use to displayed in the web browsers.

      · 2. Cons: HTML don't have complex data type, and not efficient for data storage(not for big data)

```r
library("xtable")
html_data <- data
print(xtable(html_data), type="html", file="example.html")
```

```r
library(xml2)

# Create a root node
xml_root <- xml_new_root("inventory")
```

```r
# Loop through the data frame and create child nodes for each row
for (i in 1:nrow(data)) {

  item_node <- xml_add_child(xml_root, "item")

  xml_add_child(item_node, "Category", data$Category[i])
  xml_add_child(item_node, "Item_Name", data$Item_Name[i])
  xml_add_child(item_node, "Item_ID", as.character(data$Item_ID[i]))
  xml_add_child(item_node, "Brand", data$Brand[i])
  xml_add_child(item_node, "Price", as.character(data$Price[i]))
  xml_add_child(item_node, "Variation_ID", data$Variation_ID[i])
  xml_add_child(item_node, "Variation_Details", data$Variation_Details[i])
}

# Save the XML document to a file
write_xml(xml_root, file = "inventory_data_fixed.xml")

print(xml_root)
```

```
## {xml_document}
## <inventory>
##  [1] <item>\n  <Category>Electronics</Category>\n  <Item_Name>Smartphone</Ite ...
##  [2] <item>\n  <Category>Electronics</Category>\n  <Item_Name>Smartphone</Ite ...
##  [3] <item>\n  <Category>Electronics</Category>\n  <Item_Name>Laptop</Item_Na ...
##  [4] <item>\n  <Category>Electronics</Category>\n  <Item_Name>Laptop</Item_Na ...
##  [5] <item>\n  <Category>Home Appliances</Category>\n  <Item_Name>Refrigerato ...
##  [6] <item>\n  <Category>Home Appliances</Category>\n  <Item_Name>Refrigerato ...
##  [7] <item>\n  <Category>Home Appliances</Category>\n  <Item_Name>Washing Mac ...
##  [8] <item>\n  <Category>Home Appliances</Category>\n  <Item_Name>Washing Mac ...
##  [9] <item>\n  <Category>Clothing</Category>\n  <Item_Name>T-Shirt</Item_Name ...
## [10] <item>\n  <Category>Clothing</Category>\n  <Item_Name>T-Shirt</Item_Name ...
## [11] <item>\n  <Category>Clothing</Category>\n  <Item_Name>T-Shirt</Item_Name ...
## [12] <item>\n  <Category>Clothing</Category>\n  <Item_Name>Jeans</Item_Name>\ ...
## [13] <item>\n  <Category>Books</Category>\n  <Item_Name>Fiction Novel</Item_N ...
## [14] <item>\n  <Category>Books</Category>\n  <Item_Name>Fiction Novel</Item_N ...
## [15] <item>\n  <Category>Books</Category>\n  <Item_Name>Non-Fiction Guide</It ...
## [16] <item>\n  <Category>Books</Category>\n  <Item_Name>Non-Fiction Guide</It ...
## [17] <item>\n  <Category>Sports Equipment</Category>\n  <Item_Name>Basketball ...
## [18] <item>\n  <Category>Sports Equipment</Category>\n  <Item_Name>Tennis Rac ...
```

#XML- * 1. Pros: XML fellows a strict structure, each element has a clear meaning. * 2. Cons: Large file size, and slow parsing. Not for big data.

```r
library(arrow)
```

```
##
## Attaching package: 'arrow'

## The following object is masked from 'package:utils':
##
##     timestamp
```

```r
write_parquet(data, "inventory_data.parquet")

# To read the data back into R
parquet_data <- read_parquet("inventory_data.parquet")

print(parquet_data)
```

```
##              Category       Item_Name Item_ID      Brand   Price Variation_ID
## 1        Electronics      Smartphone     101  TechBrand  699.99        101-A
## 2        Electronics      Smartphone     101  TechBrand  699.99        101-B
## 3        Electronics          Laptop     102 CompuBrand 1099.99        102-A
## 4        Electronics          Laptop     102 CompuBrand 1099.99        102-B
## 5    Home Appliances    Refrigerator     201   HomeCool  899.99        201-A
## 6    Home Appliances    Refrigerator     201   HomeCool  899.99        201-B
## 7    Home Appliances  Washing Machine     202  CleanTech  499.99        202-A
## 8    Home Appliances  Washing Machine     202  CleanTech  499.99        202-B
## 9           Clothing         T-Shirt     301  FashionCo   19.99        301-A
## 10          Clothing         T-Shirt     301  FashionCo   19.99        301-B
## 11          Clothing         T-Shirt     301  FashionCo   19.99        301-C
## 12          Clothing            Jeans     302 DenimWorks   49.99        302-A
## 13             Books     Fiction Novel     401          -   14.99        401-A
## 14             Books     Fiction Novel     401          -   14.99        401-B
## 15             Books Non-Fiction Guide     402          -   24.99        402-A
## 16             Books Non-Fiction Guide     402          -   24.99        402-B
## 17 Sports Equipment       Basketball     501 SportsGear   29.99        501-A
## 18 Sports Equipment     Tennis Racket     502  RacketPro   89.99        502-A
##                         Variation_Details
## 1              Color: Black, Storage: 64GB
## 2             Color: White, Storage: 128GB
## 3            Color: Silver, Storage: 256GB
## 4         Color: Space Gray, Storage: 512GB
## 5   Color: Stainless Steel, Capacity: 20 cu ft
## 6            Color: White, Capacity: 18 cu ft
## 7        Type: Front Load, Capacity: 4.5 cu ft
## 8          Type: Top Load, Capacity: 5.0 cu ft
## 9                         Color: Blue, Size: S
## 10                         Color: Red, Size: M
## 11                       Color: Green, Size: L
## 12                    Color: Dark Blue, Size: 32
## 13        Format: Hardcover, Language: English
## 14        Format: Paperback, Language: Spanish
## 15             Format: eBook, Language: English
## 16         Format: Paperback, Language: French
## 17               Size: Size 7, Color: Orange
## 18           Material: Graphite, Color: Black
```

#Parquet - * 1. Pros: Good for big data procressing. Fast read time. * 2. Cons: Parquet may be overhead for small datasets, and its more complex, may require additional learning for a new user.