

Sheet Introduction to Big Data

Dr. Osama Ghoneim

2022-2023

احمد ابن امية صبرى محمود راسد

Sec1: Introduction MongoDB

Sec2: Query Document

Sec3: MongoDB Sort() & Limit()

Sec4: Projection

Sec5: indexing

Sec6: Aggregation

Sec7: Introduction talend

Sec8: reading file

Sec9: converting file

Sec10: AdvancedXml

Sec11: MultiSchema

Sec12: Extracting data from Excel files



Task 1

1) Define:

- Unstructured Data

-mongoDB

2) difference between SQL and nosql mongoDB

3) Install mongoDB in your device

Task 2

Given the following collection of books 1

```
{_id:ObjectId("5fb29ae15b99900c3fa24292"),
  title:"MongoDb Guide",
  tag:["mongodb","guide","database"],
  n:100,
  review_score:4.3,
  price:[{v: 19.99, c: "€", country: "IT"},
  {v: 18, c: "£", country:"UK"} ],
  author: {_id: 1, name:"Mario", surname: "Rossi"}
},
{_id:ObjectId("5fb29b175b99900c3fa24293" ,
  title:"Developing with Python",
  tag:["python","guide","programming"],
  n:352, review_score:4.6,
  price:[{v: 24.99, c: "€", country: "IT"}, {v: 19.49, c: "£", country:"UK"} ],
  author: {_id: 2, name:"John", surname: "Black"}
}
```

1. Find all the books with a number of pages greater than 250
2. Find all the books authored by Mario Rossi
3. Find all the books with a price less than 20 € for Italy (IT)

Task 3

Given the following collection

```
{_id:ObjectId("5fb29ae15b99900c3fa24292"),
  title:"MongoDb Guide",
  tag:["mongodb","guide","database"],
  n:100,
  review_score:4.3,
  price:[{v: 19.99, c: "€", country: "IT"},
  {v: 18, c: "£", country:"UK"} ],
  author: {_id: 1, name:"Mario", surname: "Rossi"}
},
{_id:ObjectId("5fb29b175b99900c3fa24293" ,
  title:"Developing with Python",
  tag:["python","guide","programming"],
  n:352, review_score:4.6,
  price:[{v: 24.99, c: "€", country: "IT"}, {v: 19.49, c: "£", country:"UK"} ],
  author: {_id: 2, name:"John", surname: "Black"}
}
```

-display the documents sorted by title in descending order.

- display the documents sorted by title in the Ascending order



Task 4



-display the next 5 restaurants after skipping the first 5 which are in the location cairo.

-display the first 5 restaurants which are in the location cairo.

احمد ابراهيم صبرى محمود راشد



Task 5

-display how to delete only the first record

-create index of the field (phone no) for ascending order

Task 6

```
{_id:ObjectId("5fb29ae15b99900c3fa24292"),
title:"MongoDb Guide",
tag:["mongodb","guide","database"],
n:100,
review_score:4.3,
price:[{v: 19.99, c: "€", country: "IT"},
{v: 18, c: "£", country:"UK"} ],
author: {_id: 1,
name:"Mario",
surname: "Rossi"}
},
{_id:ObjectId("5fb29b175b99900c3fa24293",
title:"Developing with Python",
tag:["python","guide","programming"],
n:352,
review_score:4.6,
price:[{v: 24.99, c: "€", country: "IT"},
{v: 19.49, c: "£", country:"UK"} ],
author: {_id: 2,
name:"John",
surname: "Black"}
},
```

-Find the maximum, the minimum and the average price of all the books with tag "database"

-Compute the number of books authored by Mario Rossi

-Insert the tag "NoSQL" for all the books with tag "mongodb"

Task 7

1) Define:

- Talend

-Data Integration

2) Install Talend Studio in your device



Task 8

How to read file in talend?

احمد ابراهيم صبرى محمود راشد



Task 9

How to convert files in talend?



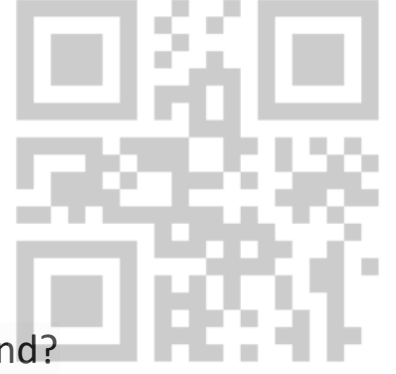
احمد ابراهيم صبرى محمود راشد



Task 10

How to Advanced XML output for complex XML in talend?

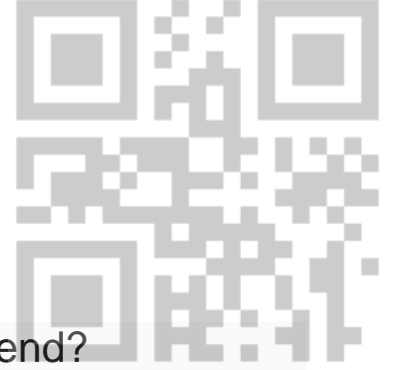
احمد ابراهيم صبرى محمود راشد



Task 11

How to Working with multi schema XML files in talend?

احمد ابراهيم صبرى محمود راشد



Task 12

How to Extracting data from Excel files in talend?



احمد ابراهيم صبرى محمود راشد

