IMAT3104: Database Management and Programming Assignment

Georgios Karseras
P2424630

Table of Contents

Questions:

Q1: Find the description and price of the product entitled “Kid's Astronaut Hat”. No other product details are required.

```
db.P2424630_products.aggregate([
    {$match: {gk_title: "Kid's Astronaut Hat"}},
    {$project: {_id: 0, description: "$gk_description",
        price: "$gk_price"}}])
```

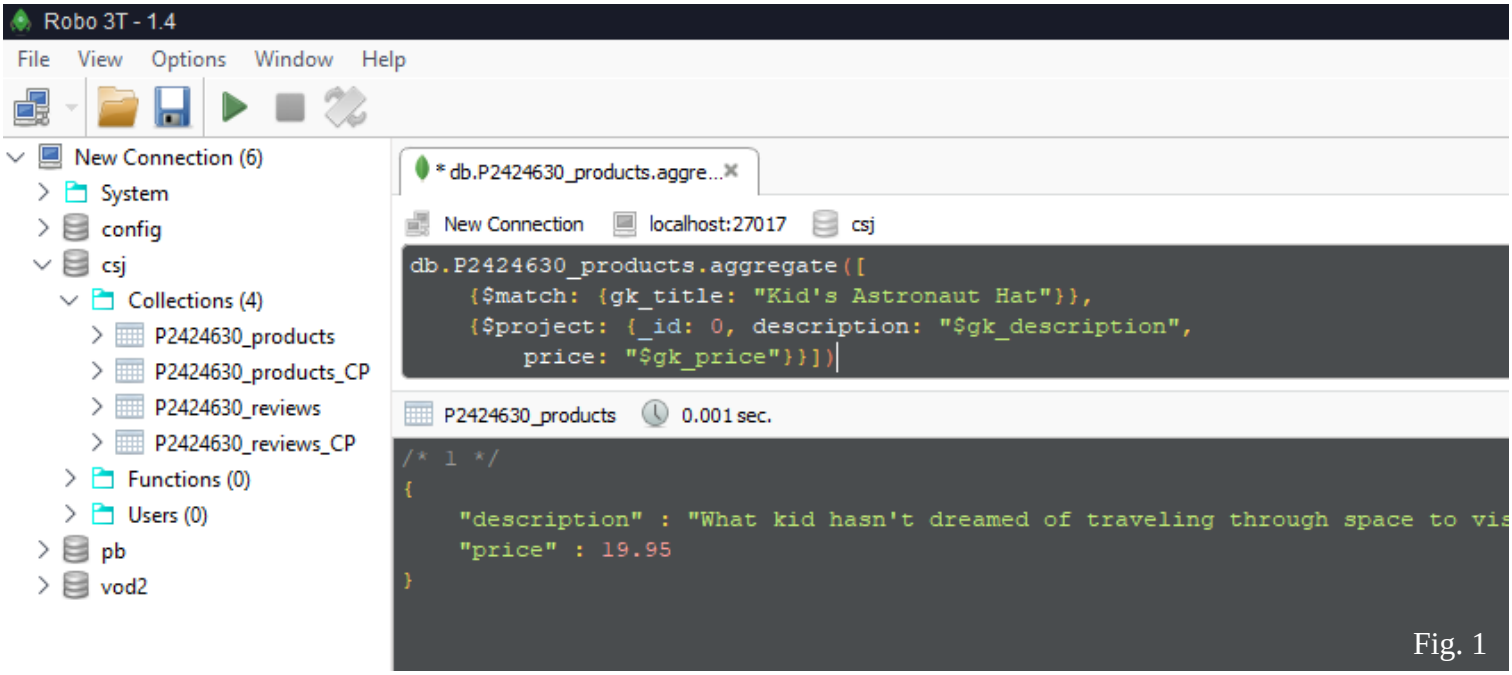

Fig. 1

Q2: Find all details of products that are priced at either $7.99 or $14.99.

```
db.P2424630_products.aggregate([
    {$match: {gk_price: {$in: [7.99, 14.99]}}}])
```

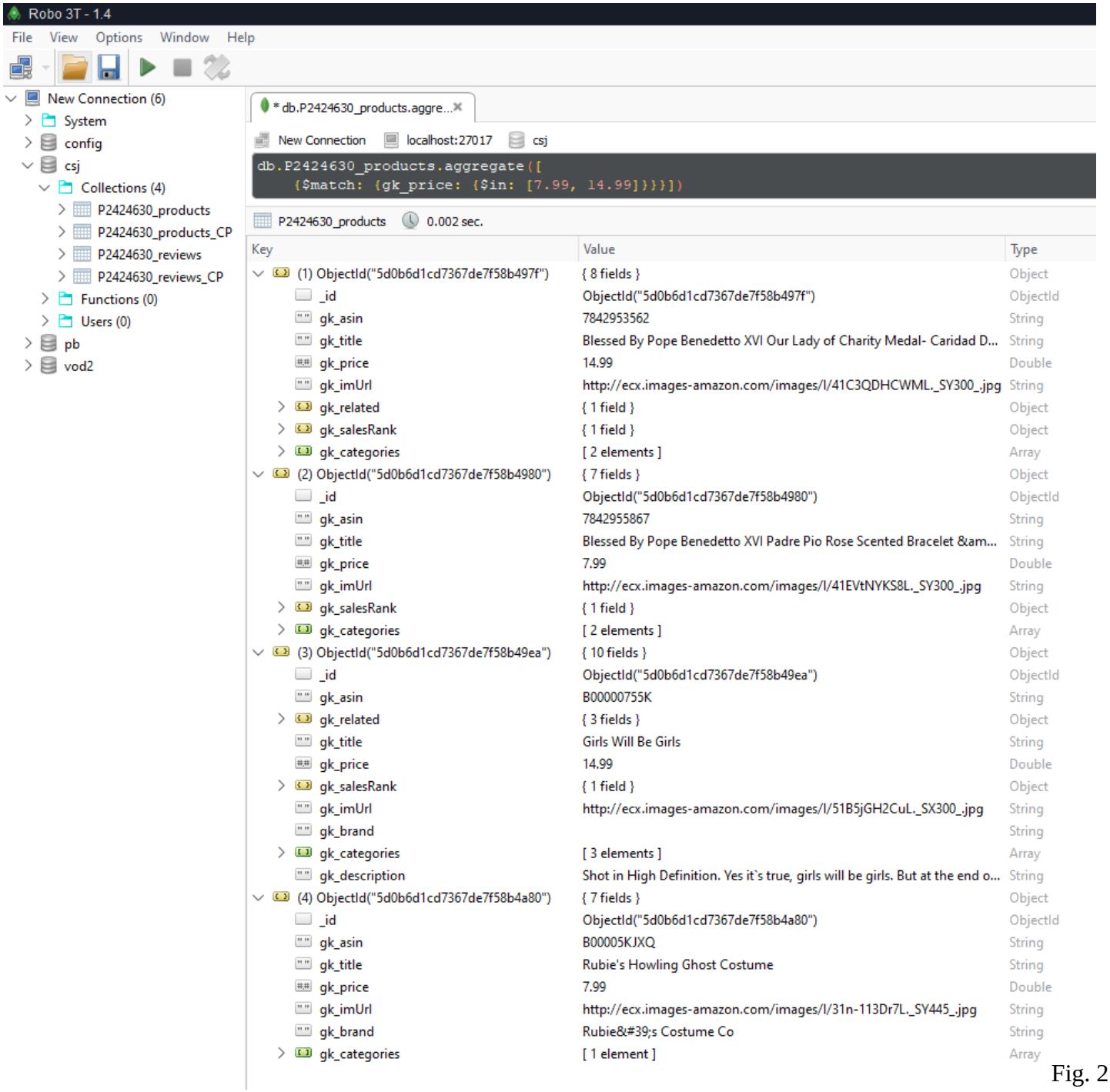

Fig. 2

Q3: Count the number of products that have a price but do not have a description. Clearly show the number you receive as your answer.

```
db.P2424630_products.aggregate([
    {$match: {gk_price: {$exists: true},
        gk_description: {$exists: false}}},
    {$count: "hasPriceAndNoDescription"}])
```

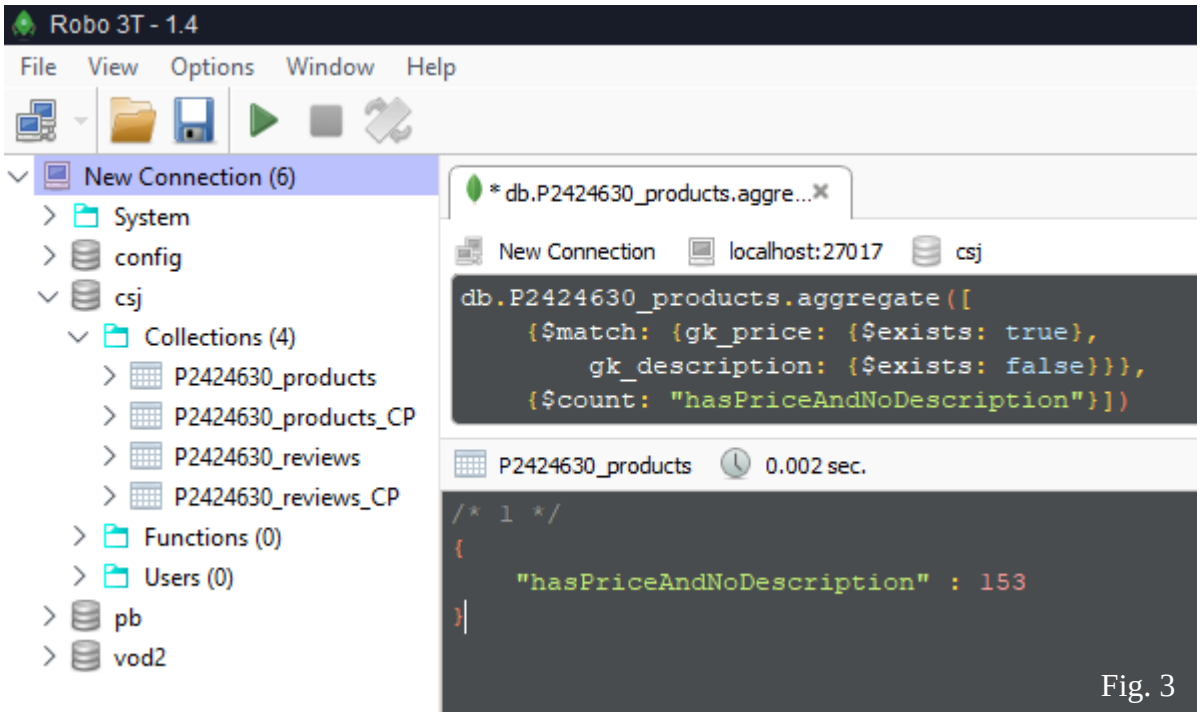

Fig. 3

Q4: List the product IDs of products that were reviewed before the 15th June 2000. All calculations should be performed by MongoDB. Only list the product IDs and do not show any duplicates.

```
var to_test = ISODate("2000-06-15T00:00:00.000Z").getTime() / 1000
db.P2424630_reviews.aggregate([
    {$match: {gk_unixReviewTime: {$lt: to_test}}},
    {$group: {_id: "$gk_asin"}}])
```

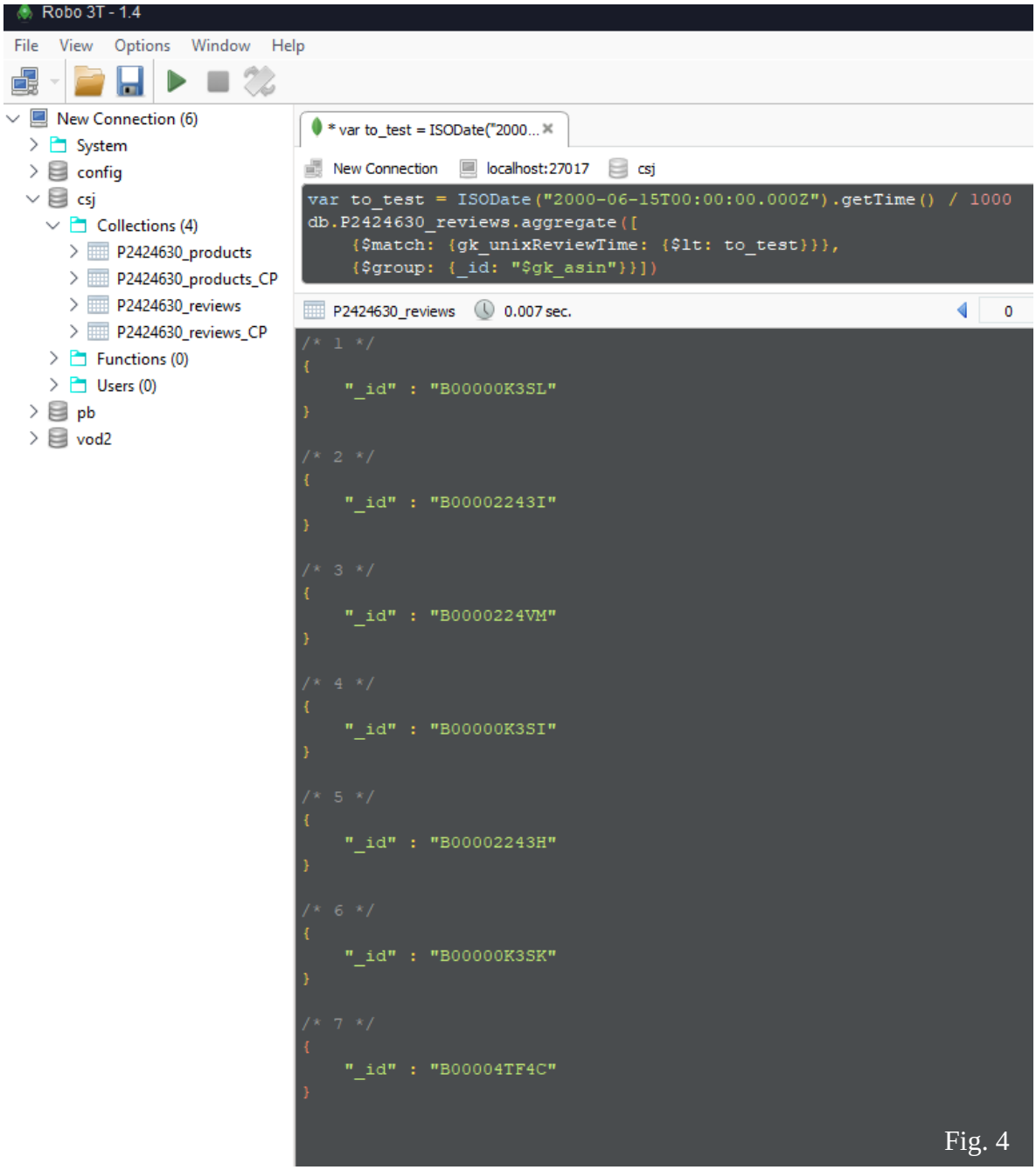

Fig. 4

Q5: List all of the reviewers who have written at least 3 reviews. Show the reviewer ID, reviewer name and the number of reviews for each one. List the names of reviewers in alphabetical order. [Hint: it is possible to solve using aggregate pipeline method].

```
db.P2424630_reviews.aggregate([
    {$group: {_id: {reviewerID: "$gk_reviewerID",
        reviewerName: "$gk_reviewerName"}, count: {$sum: 1}}},
    {$match: {count: {$gt: 2}}},
    {$sort: {"_id.reviewerName": 1}},
    {$project: {_id: 0, reviewerID: "$_id.reviewerID",
        reviewerName: "$_id.reviewerName",
        reviewCount: "$count"}}])
```

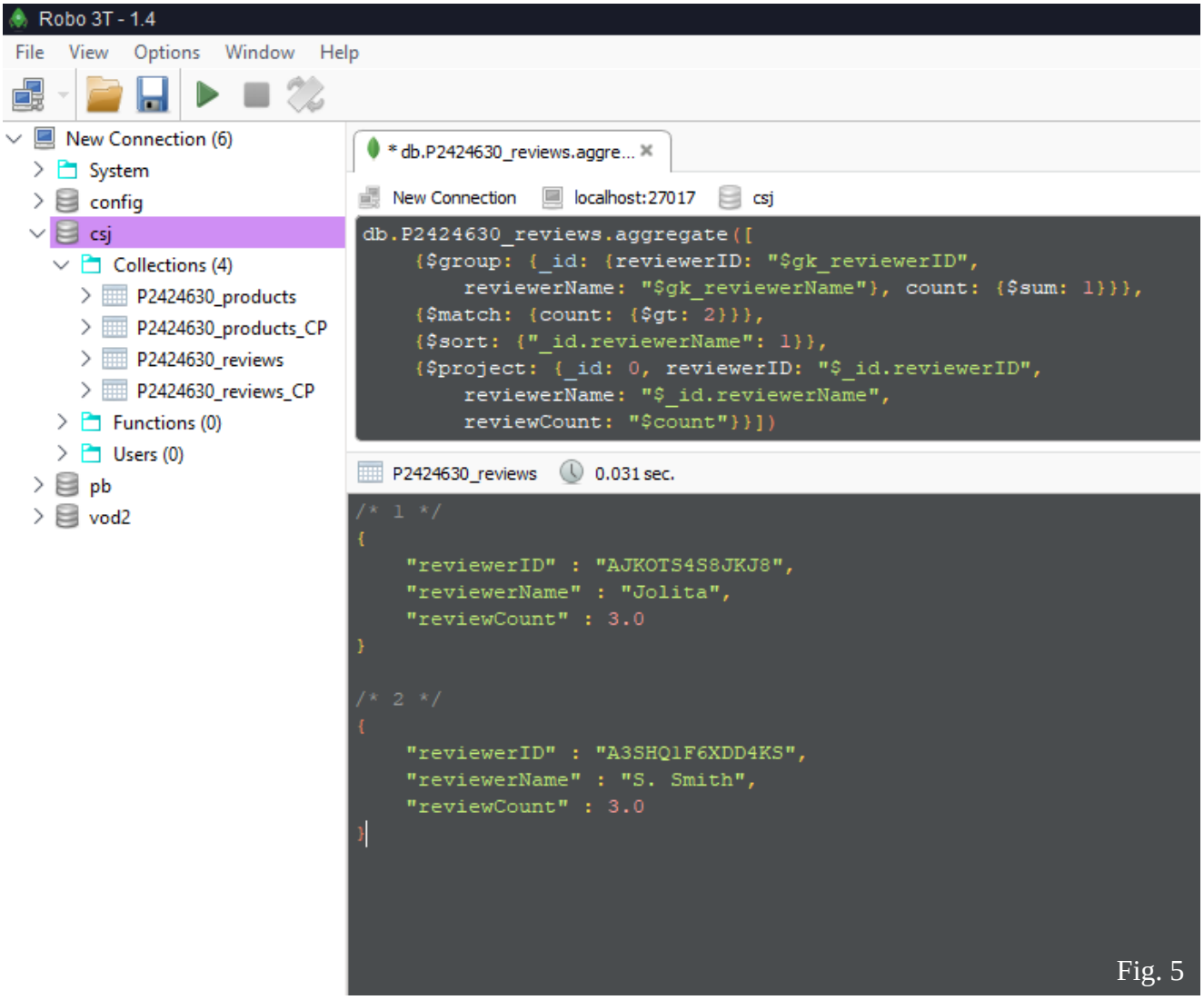

Fig. 5

Q6: Count the number of related products that were also viewed with the product “Mens Timex”. Only give the number of related products as your answer. [Hint: it is possible to solve using aggregate pipeline method].

```
db.P2424630_products.aggregate([
    {$match: {gk_title: "Mens Timex"}},
    {$project: {_id: 0, numRelated:
        {$size: "$gk_related.also_viewed"}}}])
```

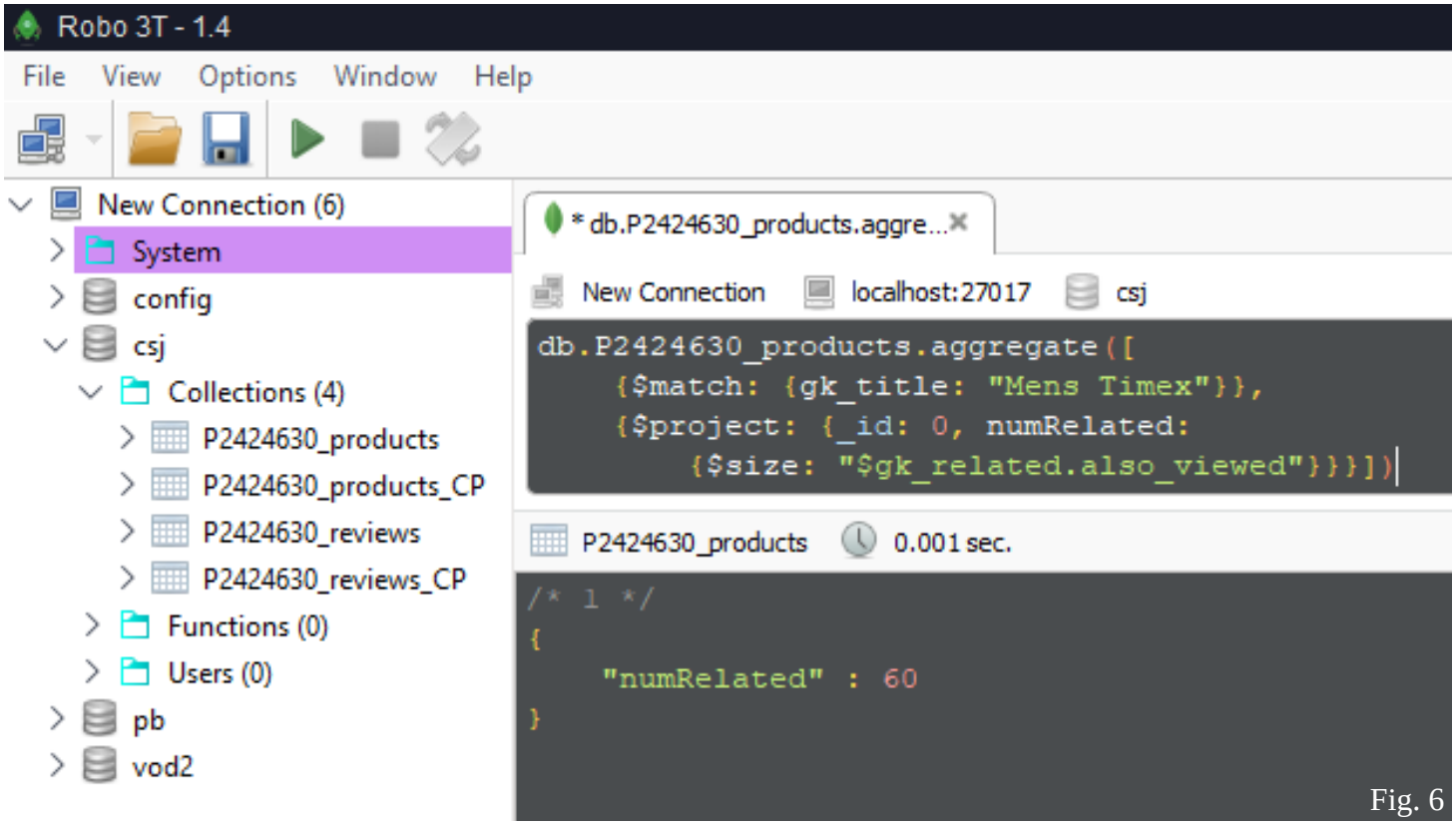

Fig. 6

Q7: List the product ID and title of all products in the “T-Shirts” category. No other product details are required. [Hint: it is possible to solve using aggregate pipeline method].

```
db.P2424630_products.aggregate([
    {$unwind: "$gk_categories"},
    {$match: {$expr: {$in: ["T-Shirts", "$gk_categories"]}}},
    {$project: {_id: 0, asin: "$gk_asin", title: "$gk_title"}}])
```

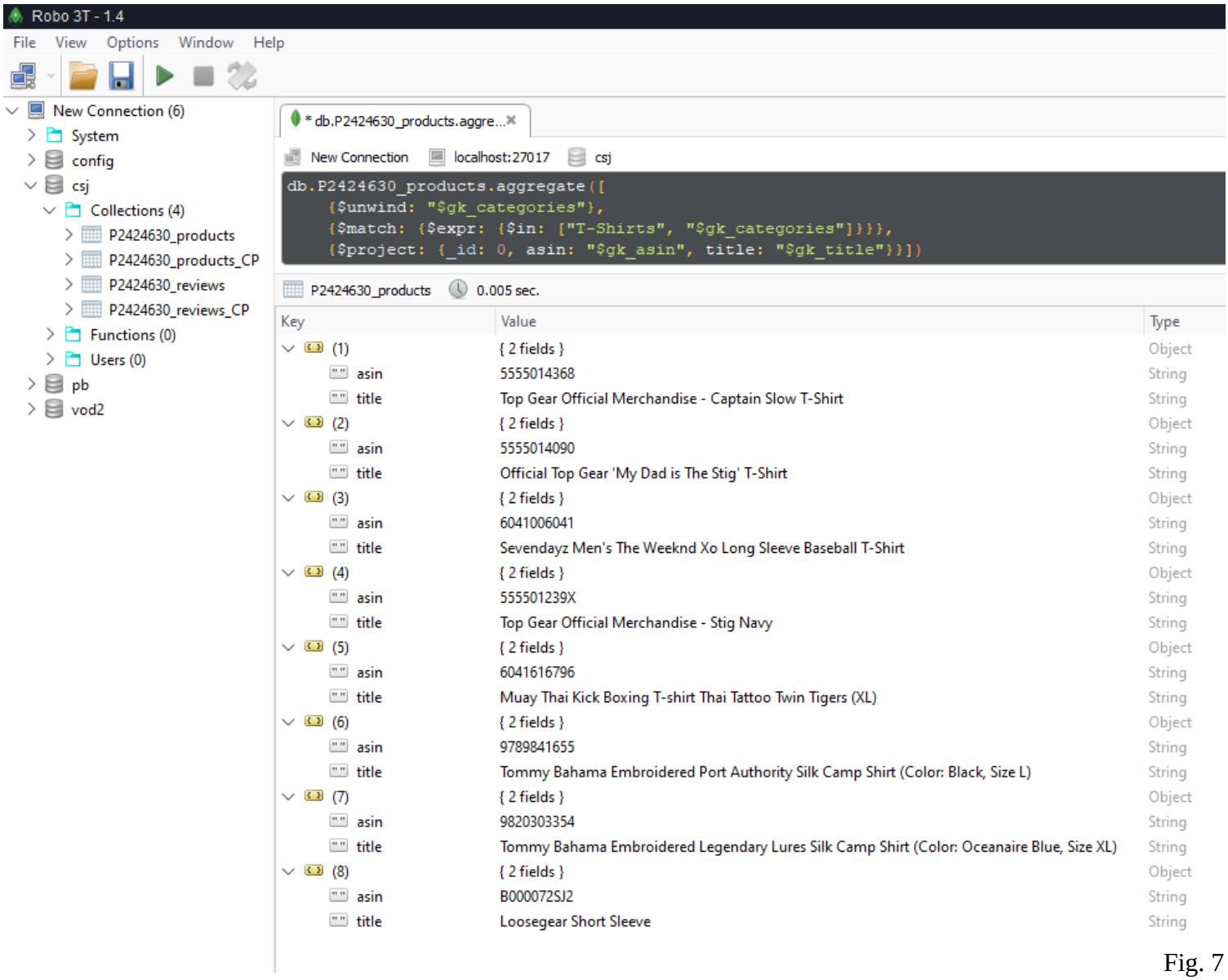

Fig. 7

Q8: Find the highest rating by the reviewer identified as “A2B3TVWTZ7609A” by using `map reduce` programming.

```
var map = function() {
        emit(this.gk_reviewerID, this.gk_overall)}
var reduce = function(key, values) {
    var maxRating = values[0]
    values.forEach(function(value){
        if (maxRating<value){
            maxRating = value
        }})
    return maxRating}
db.P2424630_reviews.mapReduce(map, reduce, {
        query: {gk_reviewerID: "A2B3TVWTZ7609A"},
        out: {inline: 1}})
```

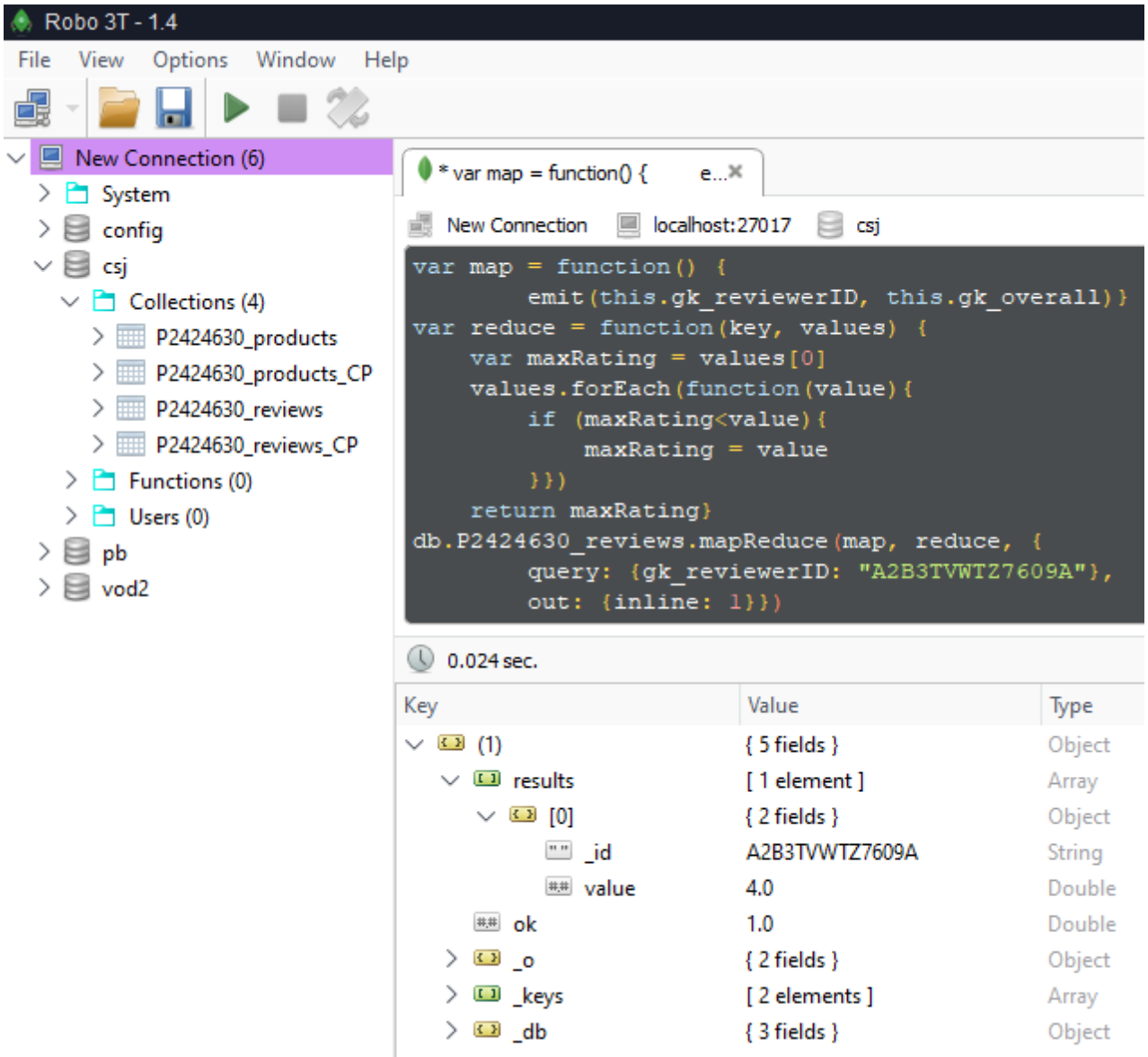

Fig. 8

Q9: Using both collections, find the title and description of products reviewed by the reviewer identified by “A2C65Q9IDNCSR”. No other product details are required.

```
var ASINcursor = db.P2424630_reviews.aggregate([
    {$match: {gk_reviewerID: "A2C65Q9IDNCSR"}},
    {$project: {gk_asin: "$gk_asin"}}])
while (ASINcursor.hasNext()){
    var nextDoc = ASINcursor.next()
    var nextAsin = nextDoc.gk_asin
    var prodCursor = (db.P2424630_products.aggregate([
        {$match: {gk_asin: nextAsin}},
        {$project: {_id: 0, title: "$gk_title",
                description: "$gk_description"}}
        ]))
    while (prodCursor.hasNext()){
        print(prodCursor.next())}}
```

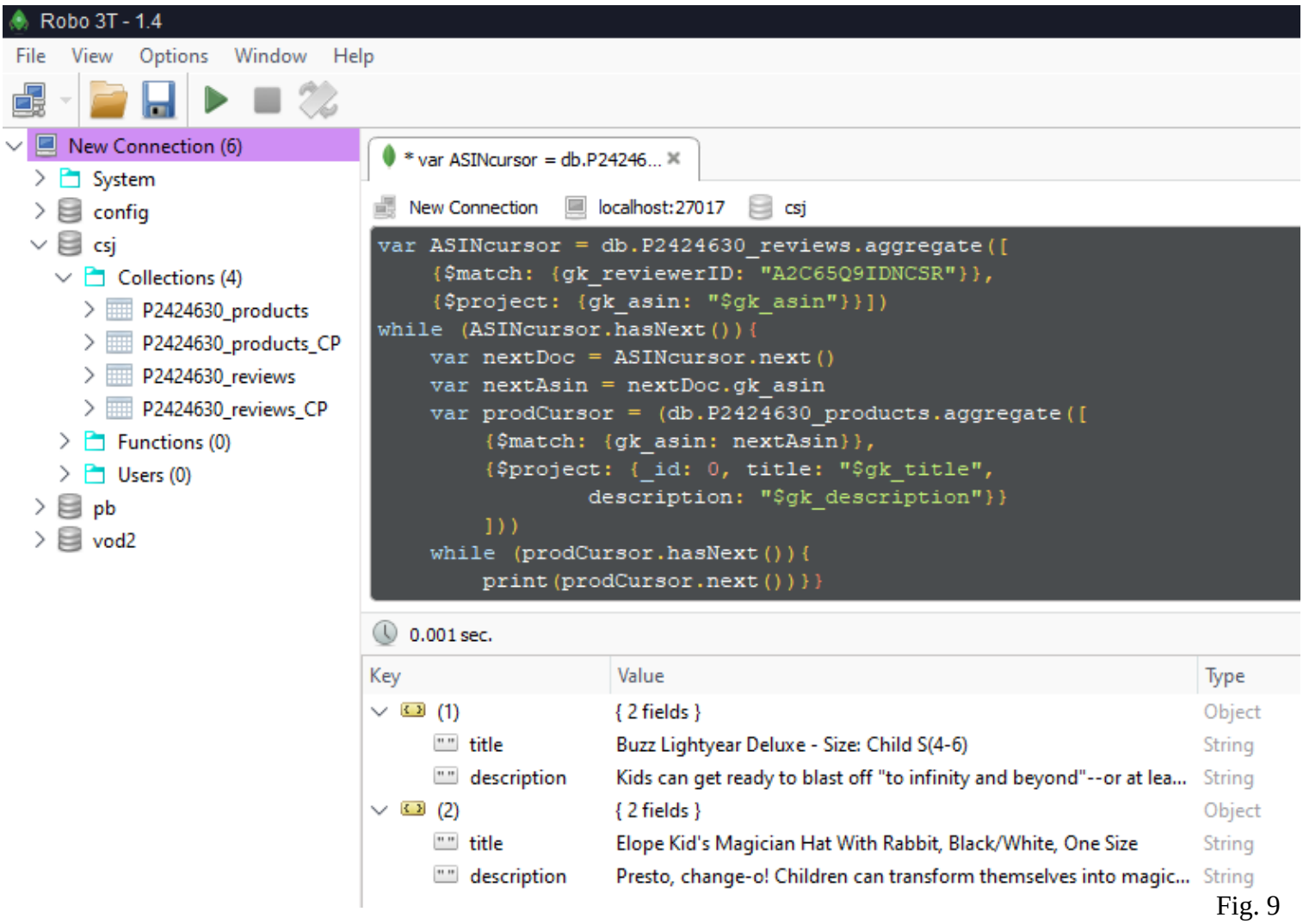

Fig. 9

E1: Implement one index that would improve the querying of the database based on one or more of the queries (Q1-Q9). Identify the chosen query or queries and explain and justify your choice of index. Present, explain and compare execution plans to support your choice and summarise your findings.

```
db.P2424630_reviews.createIndex({gk_unixReviewTime: 1})
```

I chose to use unixReviewTime as an index. Using this index I am expecting a significant reduction in time required for my question 4 query. As shown in the figure 4, the best time that I managed to achieve was 0.007 sec. While after creating the index, as shown in the figure below (fig. 10) the best time was reduced to 0.001 sec. Using the .explain("allPlansExecution") before the aggregate, it is clearly shown why the speed up is happening, as shown below in figures 11, before, and 12, after, the index was applied, totalDocsExamined were reduced from 7331 to 21, which are the returned documents.

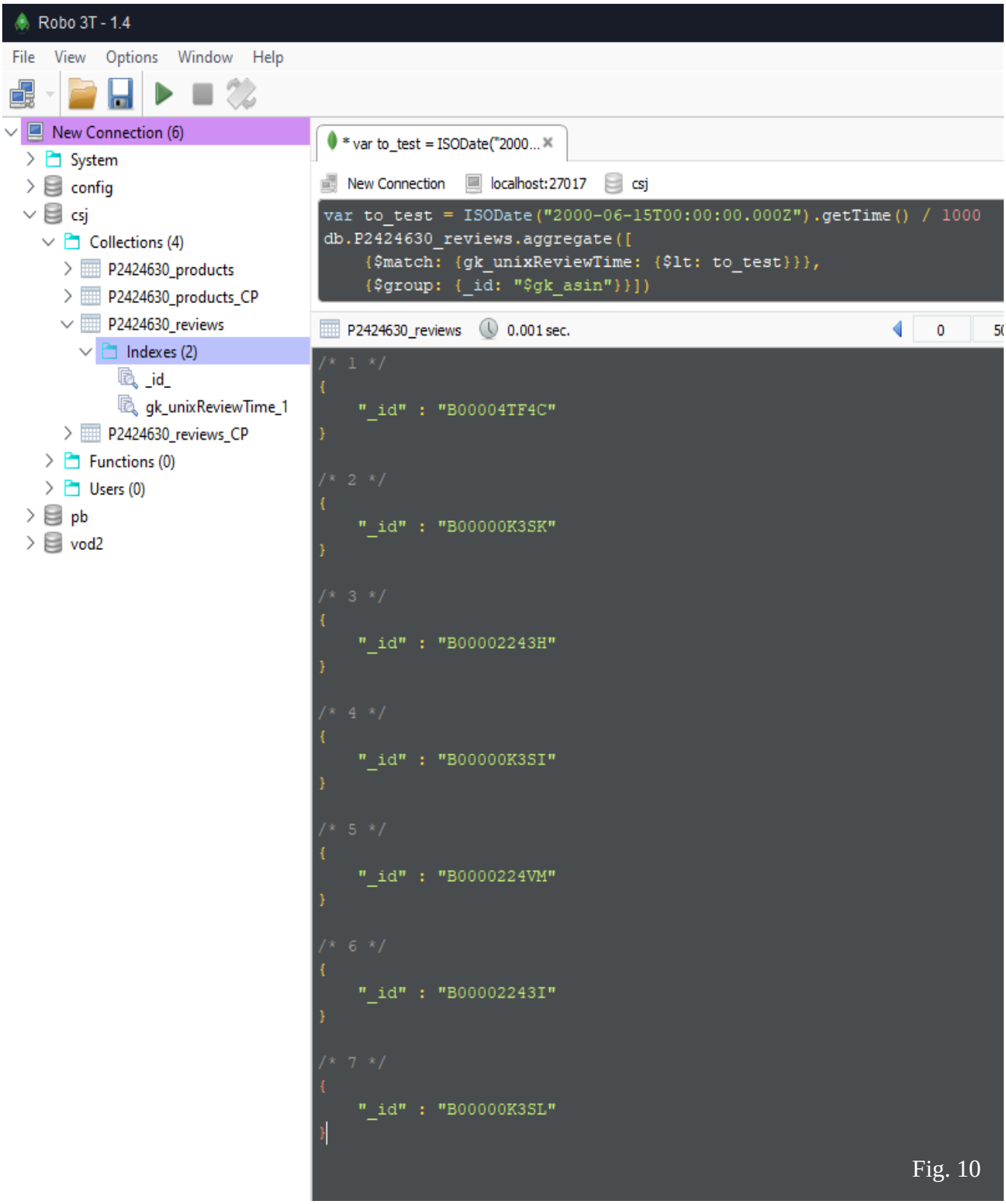
Robo 3T - 1.4
File View Options Window Help
New Connection (6)
System
config
csj
Collections (4)
P2424630_products
P2424630_products_CP
P2424630_reviews
Indexes (2)
_id_
gk_unixReviewTime_1
P2424630_reviews_CP
Functions (0)
Users (0)
pb
vod2
* var to_test = ISODate("2000...
New Connection
localhost:27017
csj
var to_test = ISODate("2000-06-15T00:00:00.000Z").getTime() / 1000
db.P2424630_reviews.aggregate([
{$match: {gk_unixReviewTime: {$lt: to_test}}},
{$group: {_id: "$gk_asin"}}])
P2424630_reviews
0.001 sec.
0
/* 1 */
{
"_id" : "B00004TF4C"
}
/* 2 */
{
"_id" : "B00000K3SK"
}
/* 3 */
{
"_id" : "B00002243H"
}
/* 4 */
{
"_id" : "B00000K3SI"
}
/* 5 */
{
"_id" : "B0000224VM"
}
/* 6 */
{
"_id" : "B00002243I"
}
/* 7 */
{
"_id" : "B00000K3SL"
}

Fig. 10

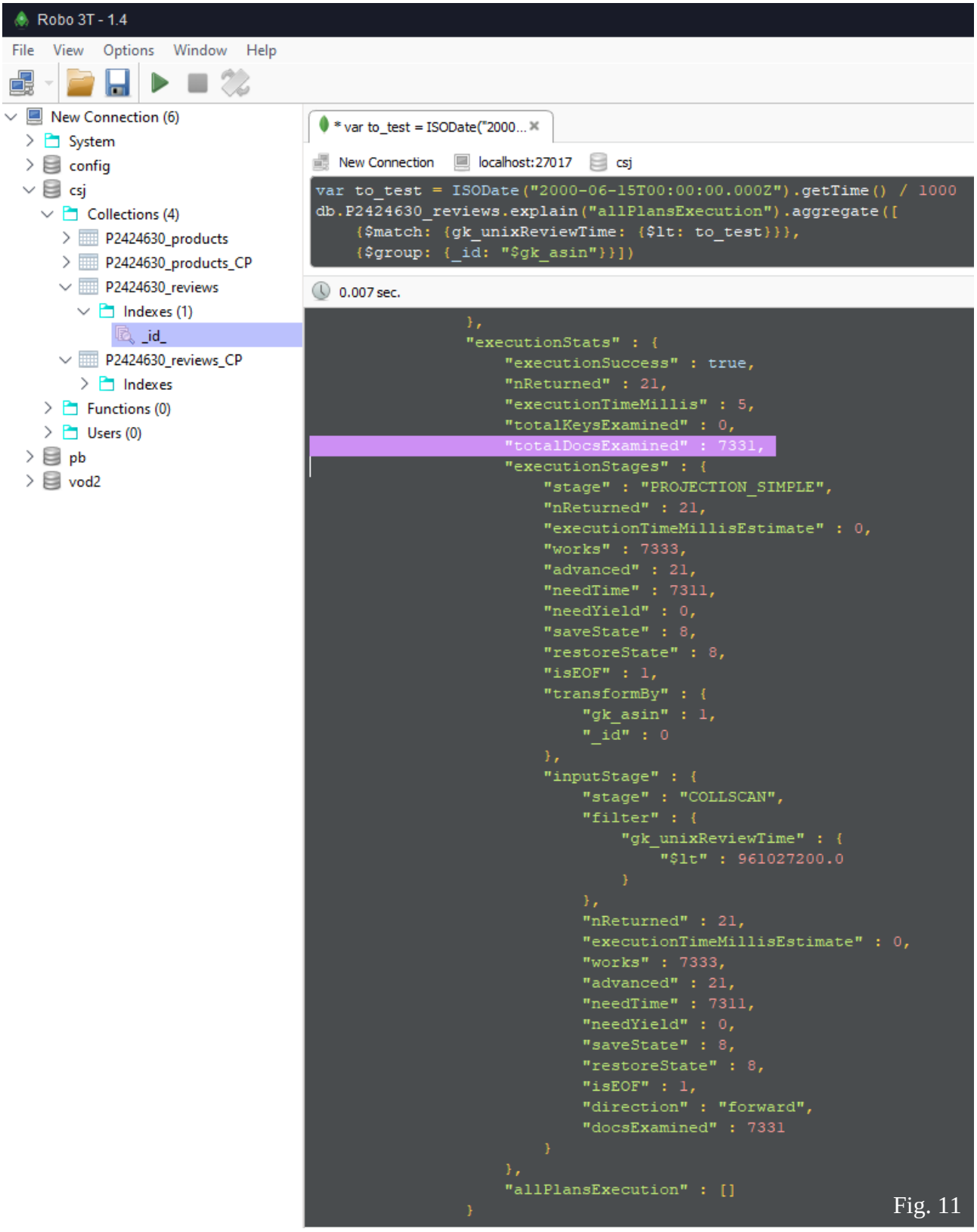
Robo 3T - 1.4
File View Options Window Help
New Connection (6)
System
config
csj
Collections (4)
P2424630_products
P2424630_products_CP
P2424630_reviews
Indexes (1)
_id_
P2424630_reviews_CP
Indexes
Functions (0)
Users (0)
pb
vod2
* var to_test = ISODate("2000...
New Connection
localhost:27017
csj
var to_test = ISODate("2000-06-15T00:00:00.000Z").getTime() / 1000
db.P2424630_reviews.explain("allPlansExecution").aggregate([
{$match: {gk_unixReviewTime: {$lt: to_test}}},
{$group: {_id: "$gk_asin"}}])
0.007 sec.
},
"executionStats" : {
"executionSuccess" : true,
"nReturned" : 21,
"executionTimeMillis" : 5,
"totalKeysExamined" : 0,
"totalDocsExamined" : 7331,
"executionStages" : {
"stage" : "PROJECTION_SIMPLE",
"nReturned" : 21,
"executionTimeMillisEstimate" : 0,
"works" : 7333,
"advanced" : 21,
"needTime" : 7311,
"needYield" : 0,
"saveState" : 8,
"restoreState" : 8,
"isEOF" : 1,
"transformBy" : {
"gk_asin" : 1,
"_id" : 0
},
"inputStage" : {
"stage" : "COLLSCAN",
"filter" : {
"gk_unixReviewTime" : {
"$lt" : 961027200.0
}
},
"nReturned" : 21,
"executionTimeMillisEstimate" : 0,
"works" : 7333,
"advanced" : 21,
"needTime" : 7311,
"needYield" : 0,
"saveState" : 8,
"restoreState" : 8,
"isEOF" : 1,
"direction" : "forward",
"docsExamined" : 7331
}
},
"allPlansExecution" : []
}

Fig. 11

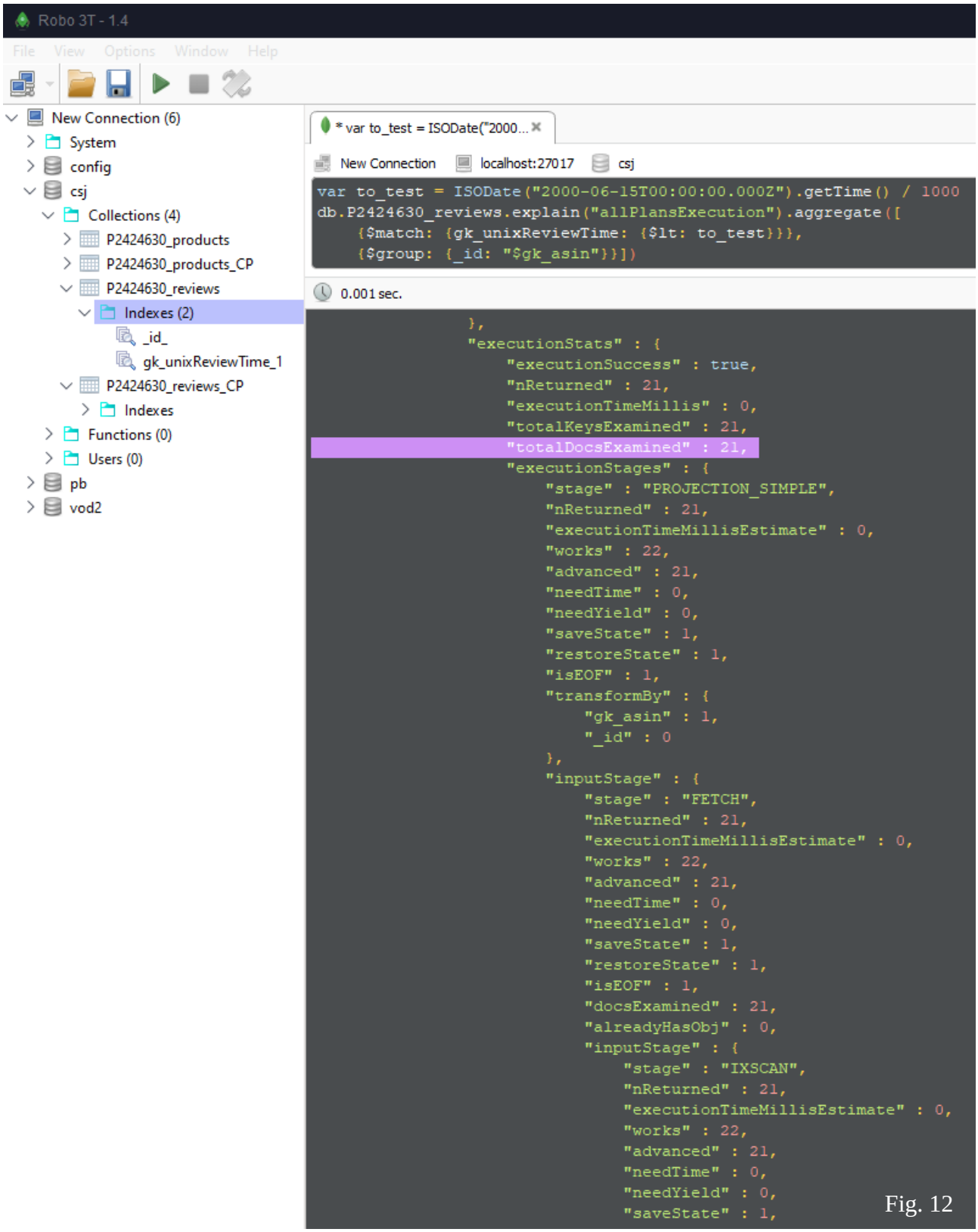
Robo 3T - 1.4
File View Options Window Help
New Connection (6)
System
config
csj
Collections (4)
P2424630_products
P2424630_products_CP
P2424630_reviews
Indexes (2)
_id_
gk_unixReviewTime_1
P2424630_reviews_CP
Indexes
Functions (0)
Users (0)
pb
vod2
* var to_test = ISODate("2000...
New Connection
localhost:27017
csj
var to_test = ISODate("2000-06-15T00:00:00.000Z").getTime() / 1000
db.P2424630_reviews.explain("allPlansExecution").aggregate([
{$match: {gk_unixReviewTime: {$lt: to_test}}},
{$group: {_id: "$gk_asin"}}])
0.001 sec.
},
"executionStats" : {
"executionSuccess" : true,
"nReturned" : 21,
"executionTimeMillis" : 0,
"totalKeysExamined" : 21,
"totalDocsExamined" : 21,
"executionStages" : {
"stage" : "PROJECTION_SIMPLE",
"nReturned" : 21,
"executionTimeMillisEstimate" : 0,
"works" : 22,
"advanced" : 21,
"needTime" : 0,
"needYield" : 0,
"saveState" : 1,
"restoreState" : 1,
"isEOF" : 1,
"transformBy" : {
"gk_asin" : 1,
"_id" : 0
},
"inputStage" : {
"stage" : "FETCH",
"nReturned" : 21,
"executionTimeMillisEstimate" : 0,
"works" : 22,
"advanced" : 21,
"needTime" : 0,
"needYield" : 0,
"saveState" : 1,
"restoreState" : 1,
"isEOF" : 1,
"docsExamined" : 21,
"alreadyHasObj" : 0,
"inputStage" : {
"stage" : "IXSCAN",
"nReturned" : 21,
"executionTimeMillisEstimate" : 0,
"works" : 22,
"advanced" : 21,
"needTime" : 0,
"needYield" : 0,
"saveState" : 1,

Fig. 12

D1: Write code in MongoDB to automatically embed the details of reviews from the reviews collection with their corresponding product in the product collection.Therefore, the products collection will contain both the product data and review data in a single collection of products.

```
var RVcursor = db.P2424630_reviews_CP.find({},{_id:0})
while (RVcursor.hasNext()){
    var nextRVDoc = RVcursor.next()
    var rv_gk_asin = nextRVDoc.gk_asin
    var rv_gk_reviewerID = nextRVDoc.gk_reviewerID
    var rv_gk_reviewerName = nextRVDoc.gk_reviewerName
    var rv_gk_helpful = nextRVDoc.gk_helpful
    var rv_gk_reviewText = nextRVDoc.gk_reviewText
    var rv_gk_overall = nextRVDoc.gk_overall
    var rv_gk_summary = nextRVDoc.gk_summary
    var rv_gk_unixReviewTime = nextRVDoc.gk_unixReviewTime
    var rv_gk_reviewTime = nextRVDoc.gk_reviewTime
    db.P2424630_products_CP.update({gk_asin: rv_gk_asin}, {$push:
        {gk_reviewerID: rv_gk_reviewerID,
         gk_reviewerName: rv_gk_reviewerName,
         gk_helpful: rv_gk_helpful,
         gk_reviewText: rv_gk_reviewText,
         gk_overall: rv_gk_overall,
         gk_summary: rv_gk_summary,
         gk_unixReviewTime: rv_gk_unixReviewTime,
         gk_reviewTime: rv_gk_reviewTime}})}
db.P2424630_products_CP.find({gk_asin: "B00005JHK9"},
    {_id:0, gk_title: 1, gk_reviewerID: 1, gk_reviewerName: 1,
     gk_helpful: 1, gk_reviewText: 1, gk_overall: 1,
     gk_summary: 1, gk_unixReviewTime: 1, gk_reviewTime: 1})
```

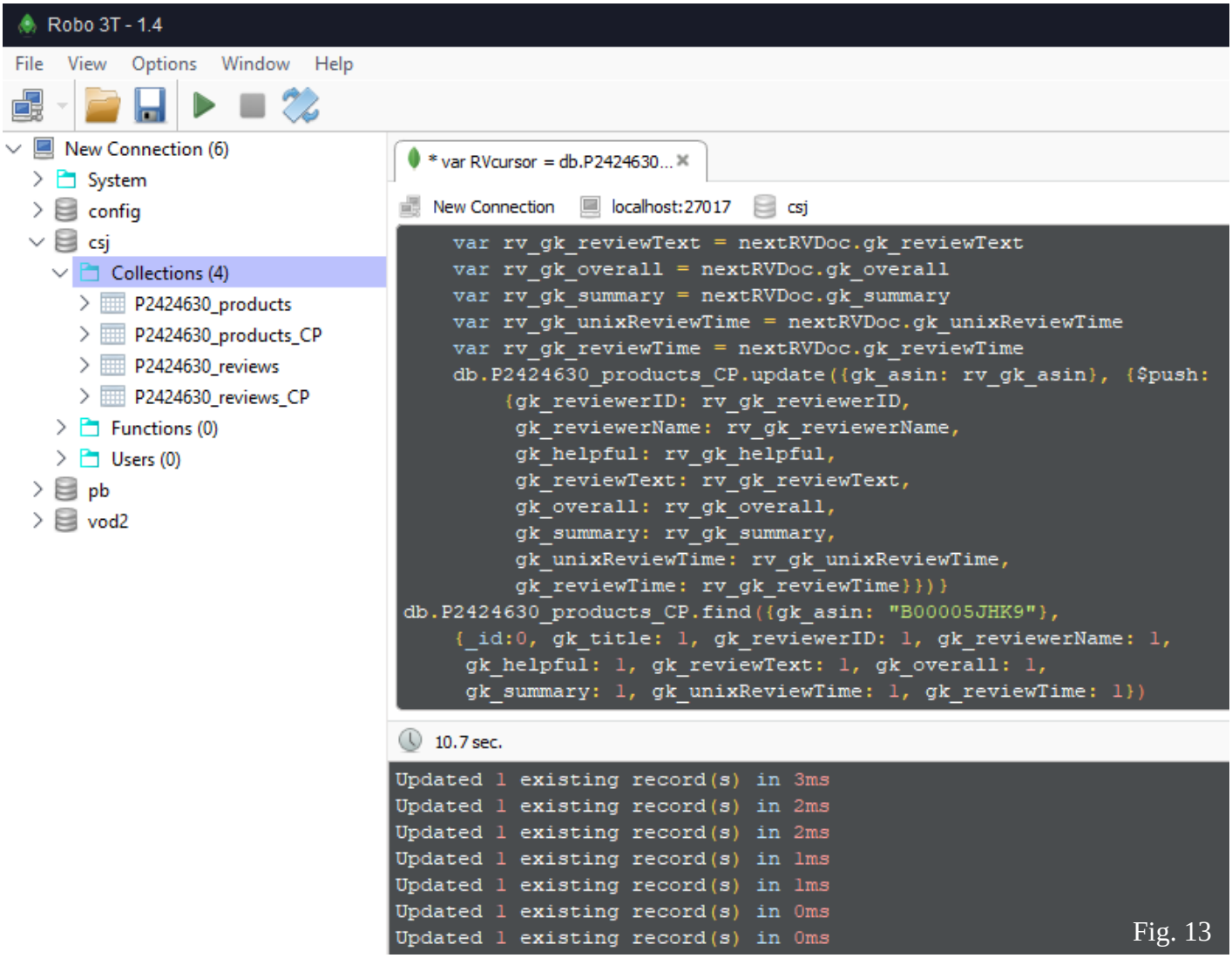
Robo 3T - 1.4
File View Options Window Help
New Connection (6)
System
config
csj
Collections (4)
P2424630_products
P2424630_products_CP
P2424630_reviews
P2424630_reviews_CP
Functions (0)
Users (0)
pb
vod2
* var RVcursor = db.P2424630...
New Connection
localhost:27017
csj
var rv_gk_reviewText = nextRVDoc.gk_reviewText
var rv_gk_overall = nextRVDoc.gk_overall
var rv_gk_summary = nextRVDoc.gk_summary
var rv_gk_unixReviewTime = nextRVDoc.gk_unixReviewTime
var rv_gk_reviewTime = nextRVDoc.gk_reviewTime
db.P2424630_products_CP.update({gk_asin: rv_gk_asin}, {$push:
{gk_reviewerID: rv_gk_reviewerID,
gk_reviewerName: rv_gk_reviewerName,
gk_helpful: rv_gk_helpful,
gk_reviewText: rv_gk_reviewText,
gk_overall: rv_gk_overall,
gk_summary: rv_gk_summary,
gk_unixReviewTime: rv_gk_unixReviewTime,
gk_reviewTime: rv_gk_reviewTime}})}
db.P2424630_products_CP.find({gk_asin: "B00005JHK9"},
{_id:0, gk_title: 1, gk_reviewerID: 1, gk_reviewerName: 1,
gk_helpful: 1, gk_reviewText: 1, gk_overall: 1,
gk_summary: 1, gk_unixReviewTime: 1, gk_reviewTime: 1})
10.7 sec.
Updated 1 existing record(s) in 3ms
Updated 1 existing record(s) in 2ms
Updated 1 existing record(s) in 2ms
Updated 1 existing record(s) in 1ms
Updated 1 existing record(s) in 1ms
Updated 1 existing record(s) in 0ms
Updated 1 existing record(s) in 0ms

Fig. 13

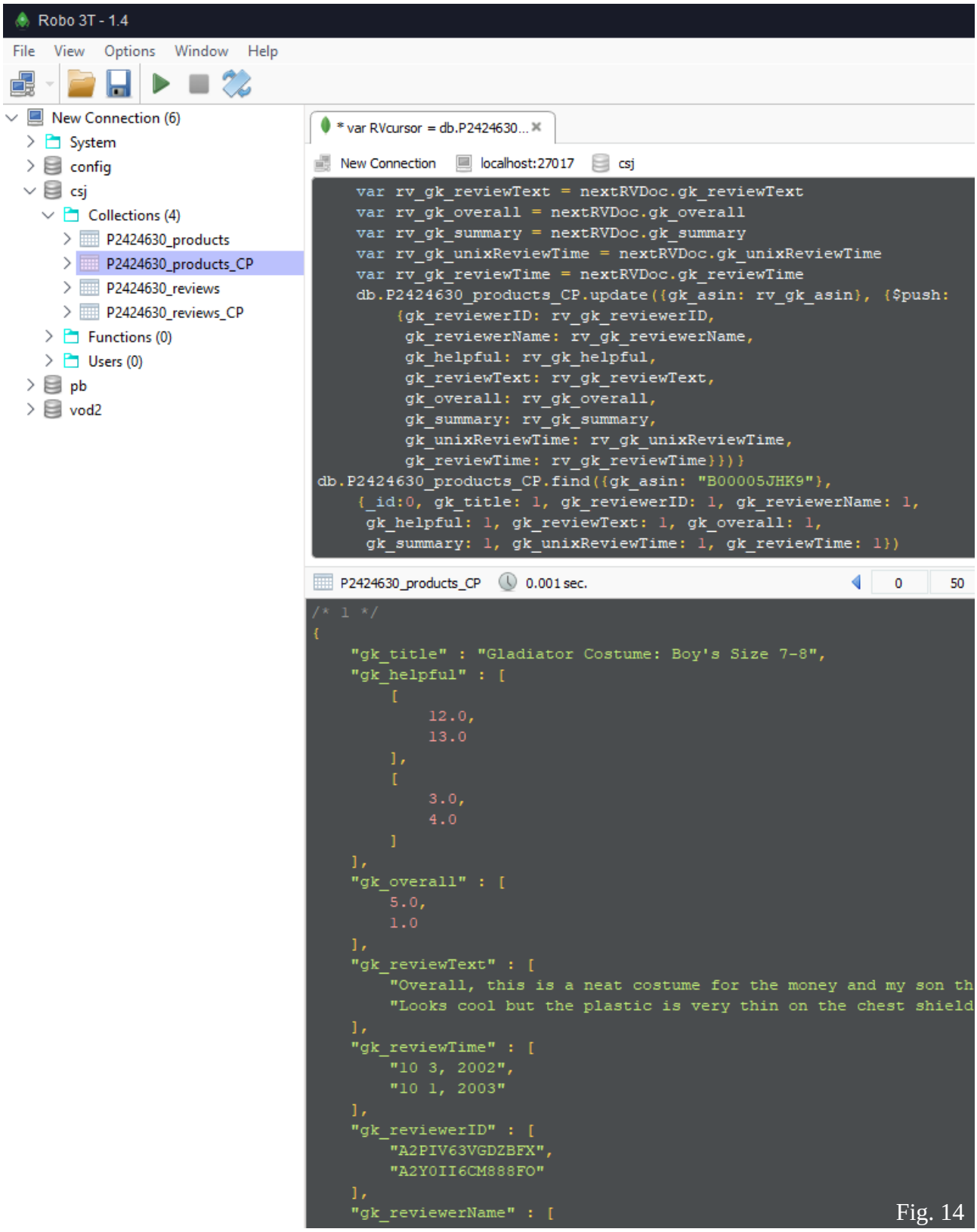
Robo 3T - 1.4
File View Options Window Help
New Connection (6)
System
config
csj
Collections (4)
P2424630_products
P2424630_products_CP
P2424630_reviews
P2424630_reviews_CP
Functions (0)
Users (0)
pb
vod2
* var RVcursor = db.P2424630...
New Connection
localhost:27017
csj
var rv_gk_reviewText = nextRVDoc.gk_reviewText
var rv_gk_overall = nextRVDoc.gk_overall
var rv_gk_summary = nextRVDoc.gk_summary
var rv_gk_unixReviewTime = nextRVDoc.gk_unixReviewTime
var rv_gk_reviewTime = nextRVDoc.gk_reviewTime
db.P2424630_products_CP.update({gk_asin: rv_gk_asin}, {$push:
{gk_reviewerID: rv_gk_reviewerID,
gk_reviewerName: rv_gk_reviewerName,
gk_helpful: rv_gk_helpful,
gk_reviewText: rv_gk_reviewText,
gk_overall: rv_gk_overall,
gk_summary: rv_gk_summary,
gk_unixReviewTime: rv_gk_unixReviewTime,
gk_reviewTime: rv_gk_reviewTime}})}
db.P2424630_products_CP.find({gk_asin: "B00005JHK9"},
{_id:0, gk_title: 1, gk_reviewerID: 1, gk_reviewerName: 1,
gk_helpful: 1, gk_reviewText: 1, gk_overall: 1,
gk_summary: 1, gk_unixReviewTime: 1, gk_reviewTime: 1})
P2424630_products_CP
0.001 sec.
0
50
/* 1 */
{
"gk_title" : "Gladiator Costume: Boy's Size 7-8",
"gk_helpful" : [
[
12.0,
13.0
],
[
3.0,
4.0
]
],
"gk_overall" : [
5.0,
1.0
],
"gk_reviewText" : [
"Overall, this is a neat costume for the money and my son th
"Looks cool but the plastic is very thin on the chest shield
],
"gk_reviewTime" : [
"10 3, 2002",
"10 1, 2003"
],
"gk_reviewerID" : [
"A2PIV63VGDZBFX",
"A2Y0II6CM888FO"
],
"gk_reviewerName" : [

Fig. 14