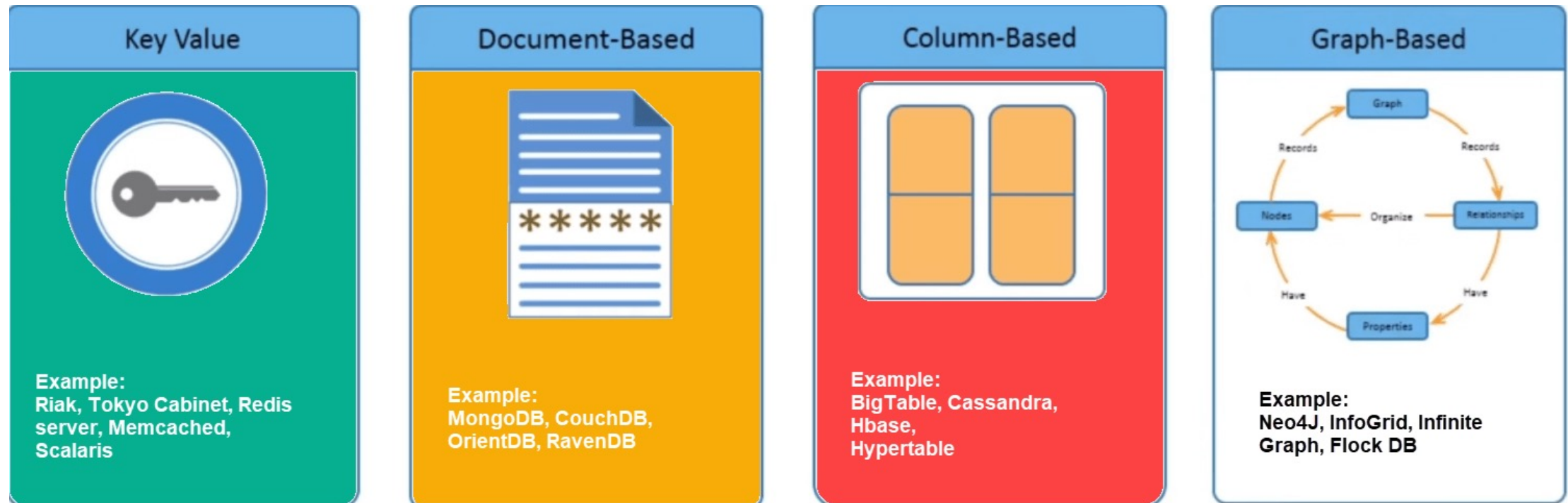
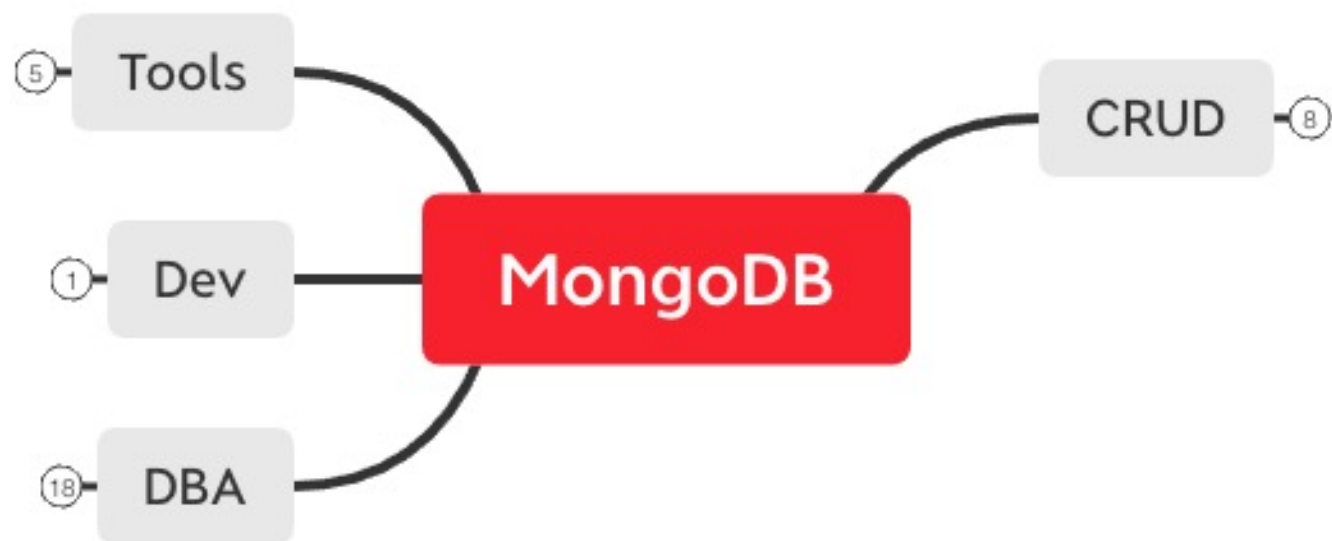


# MongoDB

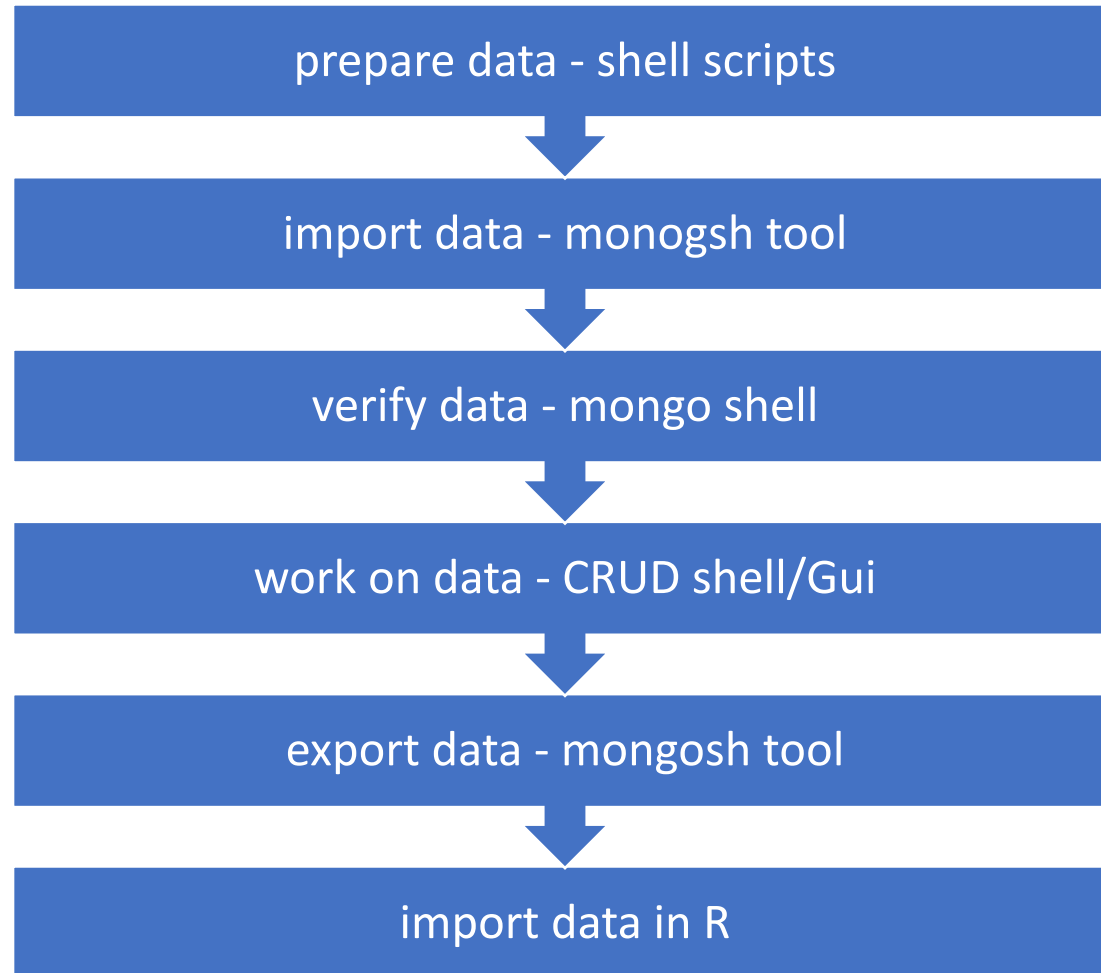
NoSQL Datastore

# NoSQL Landskab





# Flow



## prepare data - shell scripts

```
#!/usr/bin/bash -x
```

```
find ./matches/* -type f | while read x
```

```
do
```

```
    grep -l female $x
```

```
done | while read x
```

```
do
```

```
    grep match_id $x
```

```
done | cut -d\: -f2 | cut -d\, -f1 | sort | uniq > listoffemgames
```

```
cat listoffemgames | while read x
```

```
do
```

```
    cp events/$x.json tmp
```

```
done
```

```
for i in tmp/*.json; do sed -i 's/^\[//g' $i; done
```

```
for i in tmp/*.json; do sed -i 's/^\} \]/}/g' $i; done
```

```
for i in tmp/*.json; do sed -i 's/^\  \"id/  \"_id/g' $i;done
```

```
for i in tmp/*.json; do sed -i 's/^\}, {/}\n{/g' $i;done
```

```
#for i in *.json; do mongoimport -d statsbomb -c matches $i;done
```

```
43440 shot : {
43441     statsbomb_xg : 0.05243506,
43442     end_location : [ 120.0, 43.5, 1.3 ],
43443     key_pass_id : b1986be3-960d-4a11-bb9
43444     outcome : {
         id : 97,
         name : Goal
     },
     type : {
         id : 87,
         name : Open Play
     },
     body_part : {
         id : 40,
         name : Right Foot
     },
     technique : {
         id : 95,
         name : Volley
     },
     -
```

```
19796.json      2275106.json    3775579.json
(base) -----
| ~/Git/ML/statsbomb-open-data/data/tmp @ 1015
| => ls *.json | wc
      326      326      4021
```

## import data - monogsh tool



```
mongo
(base) -----
| ~/Downloads @ 10153-62 (thor)
| => mongoimport -d sales2 -c sales2 --jsonArray dumssp.json
2022-03-05T09:35:18.690+0100    connected to: mongod://localhost/
2022-03-05T09:35:19.122+0100    5000 document(s) imported successfully. 0 document(s) failed
```

Or GUI



Studio 3T Pro for MongoDB

Map-Reduce Compare Schema Reschema Tasks Export Import Data Masking SQL Migration

Schedule your first import, export or data comparison by clicking on "Tasks" in the toolbar.

Quickstart × matches × Modify Index: textidx × JSON Import\* × homes × matches × restaur

Save task Save task as... Schedule Run

**Target connection**

mydb (localhost:27017) Change target

☐ Validate JSON before import

**Select JSON sources to import:**

+ Add source — Remove source Paste from clipboard

JSON Source	Target Database	Target Collection
restaurants2.json	Downloads	restaurants2

# verify data - mongo shell



```
| => mongosh
Current Mongosh Log ID: 62232ffda2c0a03e1fa4584f
Connecting to:      mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTim
1.1.9
Using MongoDB:      5.0.3
Using Mongosh:      1.1.9
```

For mongosh info see: <https://docs.mongodb.com/mongodb-shell/>

-----

The server generated these startup warnings when booting:

2022-03-05T09:31:48.016+01:00: Access control is not enabled for the database. Read and v  
figuration is unrestricted

-----

Warning: Found ~/.mongorc.js, but not ~/.mongoshrc.js. ~/.mongorc.js will not be loaded.  
You may want to copy or rename ~/.mongorc.js to ~/.mongoshrc.js.

```
test> show dbs
Downloads 684 kB
admin     41 kB
cityflow  35.7 MB
config    36.9 kB
dining    668 kB
dining2   664 kB
dining3   664 kB
edchomes  2.29 MB
homes     360 kB
local     73.7 kB
sales     991 kB
sales2    991 kB
statsbomb 476 MB
test> use sales2
switched to db sales2
sales2> show collections
sales2
sales2> db.sales2.findOne()
{
  _id: ObjectId("5bd761dcae323e45a93ccfed"),
  saleDate: ISODate("2015-09-02T16:11:59.565Z"),
  items: [
    {
      name: 'binder',
      tags: [ 'school', 'general', 'organization' ],
      price: Decimal128("13.44"),
      quantity: 8
    },
    {
      name: 'binder',
      tags: [ 'school', 'general', 'organization' ],
      price: Decimal128("16.66"),
      quantity: 10
    }
  ],
  storeLocation: 'London',
  customer: { gender: 'M', age: 44, email: 'owtar@pu.cd', satisfaction: 2 },
  couponUsed: false,
  purchaseMethod: 'In store'
}
sales2> |
```

Or GUI



NEW: Task Scheduler is now available with your Pro license. Schedule your first import, export or data comparison.

Search Open Connections (%F) ...

Quickstart × matches × Modify Inc

mydb localhost:27017 [direct]

- Downloads
- admin
- cityflow
- config
- dining
- dining2
- dining3
- edchomes
- homes
- local
- sales
- sales2
  - Collections (1)
    - sales: Collection: sales2.sales2
      - Count: 5,000
      - Size: 4,1 MiB (4,328,222)
      - Storage Size: 900,0 KiB (921,600)
      - Indexes: 1
      - Total Index Size: 68,0 KiB (69,632)
      - Stats Fetched: 2022-03-05 10:42:34 CET
- statsbomb

mydb (localhost:27017) > sales2

Query { }

Projection { }

Skip

Result Query Code Explain

50

sales2 > saleDate

_id	saleDate
5bd761dcae3...	2015-09-02T...
5bd761dcae3...	2015-03-23T...
5bd761dcae3...	2017-03-21T...
5bd761dcae3...	2015-02-23T...
5bd761dcae3...	2017-06-22T...
5bd761dcae3...	2015-08-25T...
5bd761dcae3...	2017-12-08T...
5bd761dcae3...	2014-03-31T...
5bd761dcae3...	2014-11-11T...
5bd761dcae3...	2014-08-18T...
5bd761dcae3...	2015-07-22T...
5bd761dcae3...	2015-07-25T...

Refresh

## work on data - CRUD shell/Gui



```
sales2> db.sales2.find({"storeLocation":"/Lond/},{ "storeLocation":1,"items.quantity":1,saleDate:1,_id:0}).limit(2)
```

```
[
  {
    saleDate: ISODate("2014-11-11T02:13:51.893Z"),
    items: [
      { quantity: 4 },
      { quantity: 7 },
      { quantity: 5 },
      { quantity: 5 },
      { quantity: 3 },
      { quantity: 5 },
      { quantity: 1 },
      { quantity: 2 },
      { quantity: 3 },
      { quantity: 1 }
    ],
    storeLocation: 'London'
  },
]
```

```
sales2> db.sales2.find({"storeLocation":"/Lond/},{ "storeLocation":1,max:{$max:"$items.quantity"},saleDate:1, _id:0}).limit(2)
```

```
[
  {
    saleDate: ISODate("2014-11-11T02:13:51.893Z"),
    storeLocation: 'London',
    max: 7
  },
]
```

```
sales2> db.sales2.updateOne({_id:ObjectId("5bd761dcae323e45a93ccff3")},{ $set:{storeLocation:"Nyborg"}})
```



## export data - mongosh tool



```
| => mongoexport -v --collection=sales --db=sales -q='{"storeLocation":"London"}' --pretty --jsonArray --out dumplondon.json
2022-03-05T12:44:49.136+0100    will listen for SIGTERM, SIGINT, and SIGKILL
2022-03-05T12:44:49.142+0100    connected to: mongod://localhost/
2022-03-05T12:44:49.242+0100    exported 794 records
```

```
| => head dumplondon.json
[{
  "_id": {
    "$oid": "5bd761dcae323e45a93ccfed"
  },
  "saleDate": {
    "$date": "2015-09-02T16:11:59.565Z"
  },
  "items": [
    {
      "name": "binder",
```

# import data in R

RStudio

Project: (None)

```
1 library(jsonlite)
2 library(rjson)
3
4 london <- do.call(rbind,
5                   lapply(paste(readLines
6                               (file("/Users/thor/Downloads/dumplondon.json"), warn=FALSE),
7                               collapse=""),
8                           jsonlite::fromJSON))
```

Environment History Connections Tutorial

Import Dataset 307 MiB

Global Environment

info	Large list (1000 elements, 1.9 MB)
json_events	12841 obs. of 20 variables
london	794 obs. of 7 variables
\$ _id	: 'data.frame': 794 obs. of 1 variable:
.. \$ \$oid	: chr "5bd761dcae323e45a93ccfed" "5bd761dcae323e45a93ccfec" "5bd761dcae323e45a93ccfed"
\$ saleDate	: POSIXct, format: "2015-09-02 16:11:59" "2017-12-03 18:39:48" "2017-12-03 18:39:48"
\$ items	: List of 794
.. \$ : 'data.frame':	2 obs. of 4 variables:
.. .. \$ name	: chr "binder" "binder"
.. .. \$ tags	: List of 2
.. .. .. \$	: chr "school" "general" "organization"
.. .. .. \$	: chr "school" "general" "organization"
.. .. \$ price	: 'data.frame': 2 obs. of 1 variable:
.. .. .. \$ numberDecimal	: chr "13.44" "16.66"