



MongoDB Aggregation

Nuri Halperin | @nurih
KCDC 2016

Agenda

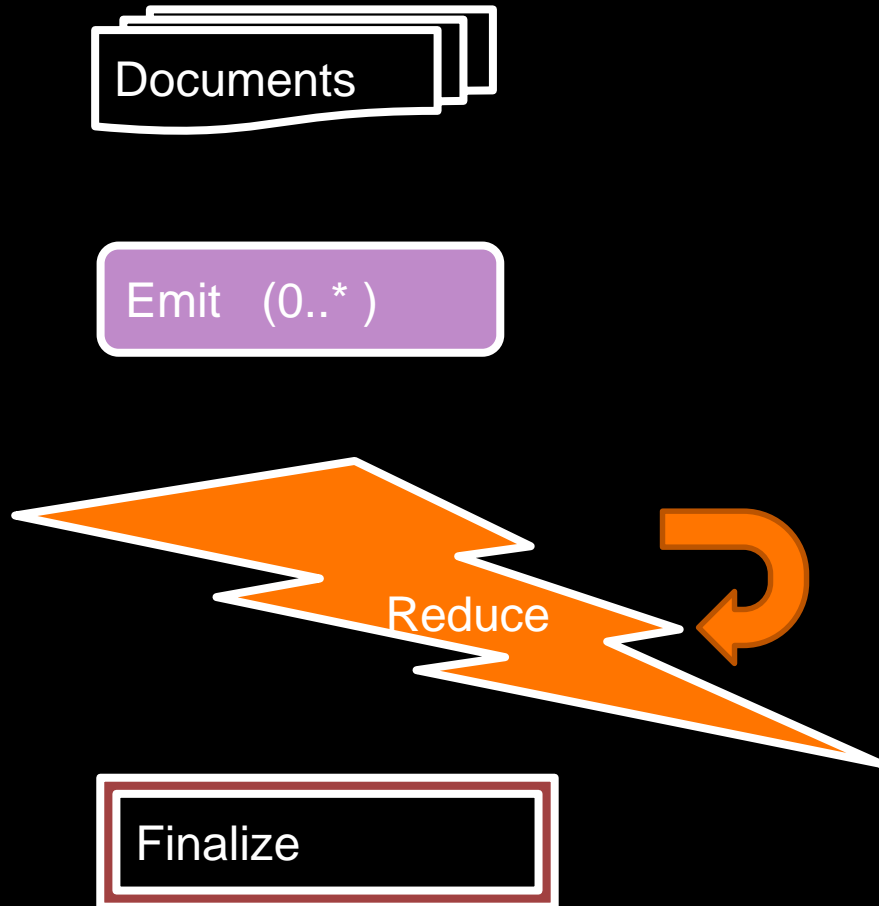
Map/Reduce pain

Aggregation framework

- Concepts
- Code demo

...Whatever comes our way

MapReduce



- Javascript: full programmability
- In/Out of Javascript costly
- Imparts "non-interactiveness"
- All I wanted was a **group-by!**

Aggregate

```
{ "_id" : "silver", "score" : 17, "house" : "A" }
```

```
> db.scores.aggregate(  
  [{ $group: {  
    _id: "$house",  
    avg: { $avg: "$score" }}}]);
```

```
{ "_id" : "B", "avg" : 20.333333333333332 },  
{ "_id" : "A", "avg" : 19.333333333333332 }
```



"Pipeline"

"Stages" (operators)



Characteristics



Transformer

Gate

Collector

Characteristics

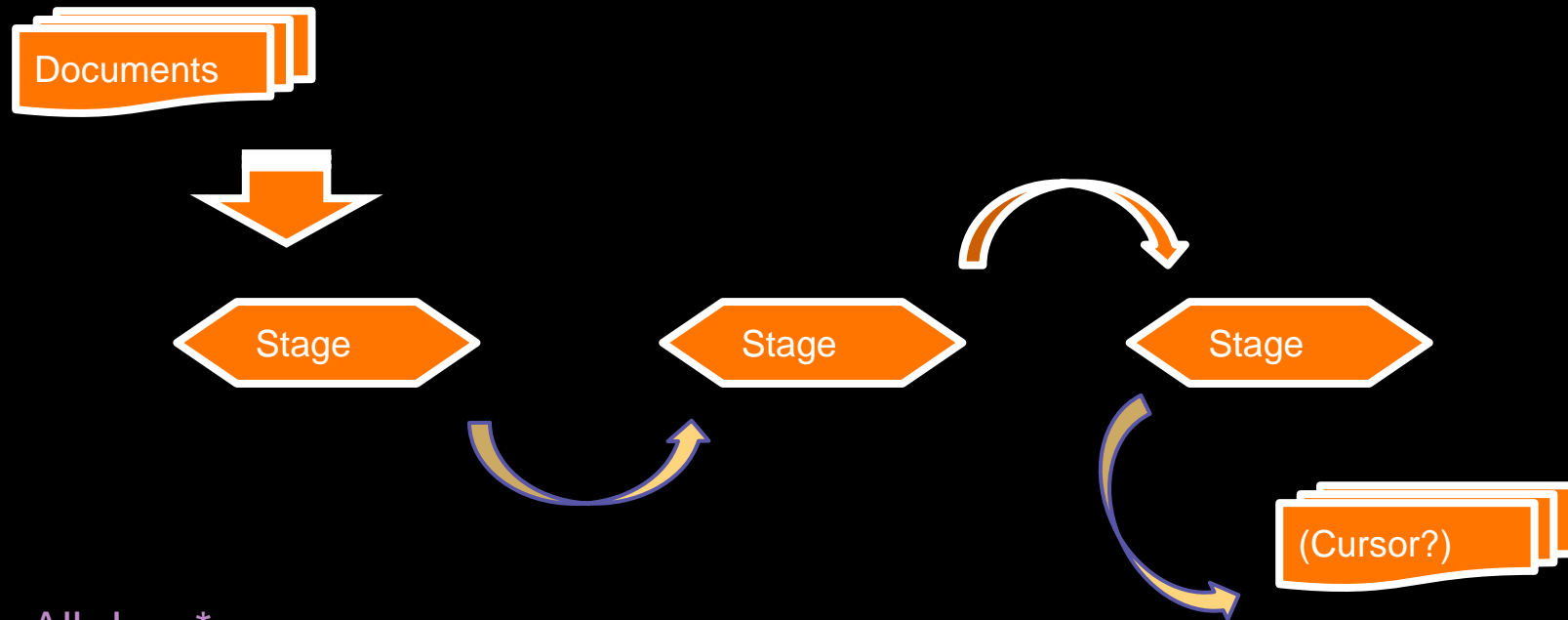


<< *Expression* >>

The *aggregate* Command

```
db.collection.aggregate(  
    [s1,s3, .. Sn]  
)
```

Aggregation pipeline



- All docs*
- One at a time*
- "The" => "A"

* (not really, not always, it depends, kinda.)

Pipeline Cast & Crew



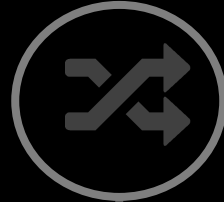
\$group



\$project
\$redact
\$unwind



\$match
\$limit
\$skip



\$sort



\$geoNear
\$out

Pipeline Cast & Crew



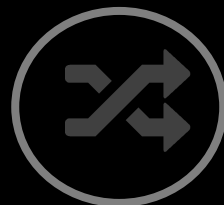
\$group



\$project
\$redact
\$unwind



\$match
\$limit
\$skip



\$sort



\$geoNear
\$out

Pipeline Cast & Crew



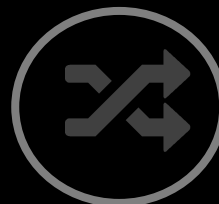
\$group



\$project
\$redact
\$unwind



\$match
\$limit
\$skip



\$sort



\$geoNear
\$out

Pipeline Cast & Crew



\$group



\$project
\$redact
\$unwind



\$match
\$limit
\$skip



\$sort



\$geoNear
\$out

Pipeline Cast & Crew



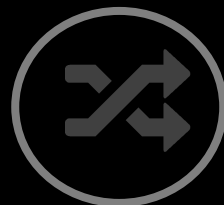
\$group



\$project
\$redact
\$unwind



\$match
\$limit
\$skip



\$sort



\$geoNear
\$out

Pipeline C



\$lookup

\$group

\$project

\$redact

\$unwind



Near

Demo



Filter early

Whittle
Down

Transformers
to the right

Arrays
over
Hashes



Rules of Thumb

Limitations

- Memory – killed if exceeds 10% of memory
- 100 MB mem limit => allowDiskUsage
- Predefined operators only (no scripting)
- Output options limited

Thank You!

Download: mongodb.org v 3.2



scan me!

