

# PIPELINE POWER

## Analytics with MongoDB Aggregation Framework



@asya999

by Asya Kamsky,

Lead MongoDB Maven

**MONGODB  
WORLD'17**

FOR GIANT IDEAS

#MDBW17

# APPLICATIONS & DATA

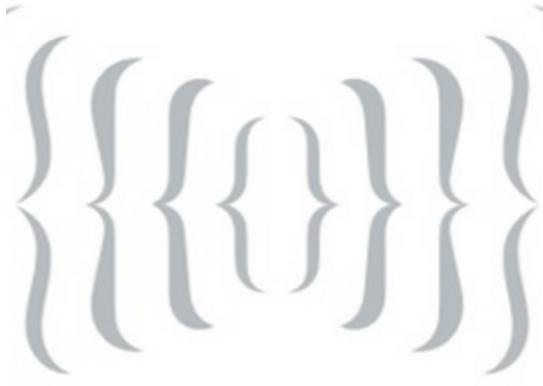




# OPTIONS FOR ANALYTICS



pre-aggregate

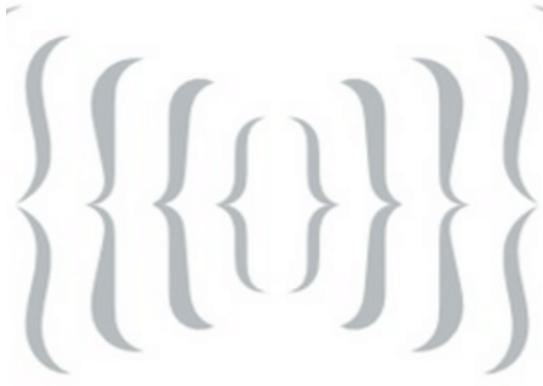


aggregate  
in MongoDB



aggregate  
elsewhere

# OPTIONS FOR ANALYTICS



aggregate  
in MongoDB



aggregate  
elsewhere

# OPTIONS FOR ANALYTICS



aggregate  
elsewhere

Secure <https://www.google.com/search?q=type+of+saw&oq=type+of+saw&aqs=chrome..69i57.1835j0j4&sourceid=ch...>

Show only

Available nearby  
 New items

Price

Up to \$60  
 \$60 – \$150  
 \$150 – \$450  
 \$450 – \$1,000  
 Over \$1,000  
\$  to  
\$  Go

Category

Saw Blades  
 Drill, Driver & Screwdriver Accessories

Brand

baileigh  
 bosch  
 dewalt  
 festool  
 milwaukee hand trucks

More

Seller

Acme Tools  
 eBay  
 HardwareAndTools.c  
 Home Depot  
 Neobits.com

More



**SKIL 3550-01 7 In. Wet Tile Saw with HydroLock System**

**\$999,999.99** from 25+ stores

★★★★★ 43 product reviews

SKIL

Skil has the unique tools you need to deliver exceptional results. Skilsaws not only build projects, they build confidence.



**17.25" SPECIAL 650 DI Semi Automatic Band Saw, Roller Table Support**

**\$38,995.00** from Circular Cold Saw Blades

17.25" SPECIAL 650 DI Semi Automatic Band Saw with Roller Table Support



**Baileigh CNC Table Saw - STS-14120-CNC**

**\$24,245.00** from 5+ stores

Baileigh · Cabinet · 14 inch · 7.5 horsepower

The STS-14120-CNC table saw from Baileigh Industrial is based from our other sliding table saw designs, but adds a CNC twist. The CNC function ...



**Safety Speed Cut Safety Speed ES5210M Panel Saw**

**\$19,949.00** from 3 stores

The ES5210M is the only panel saw in our lineup that features a carriage that traverses the full length of the machine frame. The traveling beam ...



**Hobart 6614-1 126" Vertical Meat Saw - 3 HP, 200-230/60/3V**

**\$8,481.55** from 4 stores

For precise cuts, simple operation and outstanding durability, look no further than this Hobart 6614-1 126" vertical meat saw! The 5/8" x 126 ...

Other size options: 142" (\$9,941)



**Baileigh 21 IN x 60 IN Combination Woodworking Machine MF-3005**

**\$8,295.00** from 3 stores

Baileigh · 12 inch

The MF-3005 from Baileigh Industrial takes care of that problem. This 5 in 1 combination woodworking machine combines all of the most popular ...



**Powermatic PM2000 Table Saw with Accu-Fence System, 10", 3 HP**

**\$7,179.98** from 20+ stores

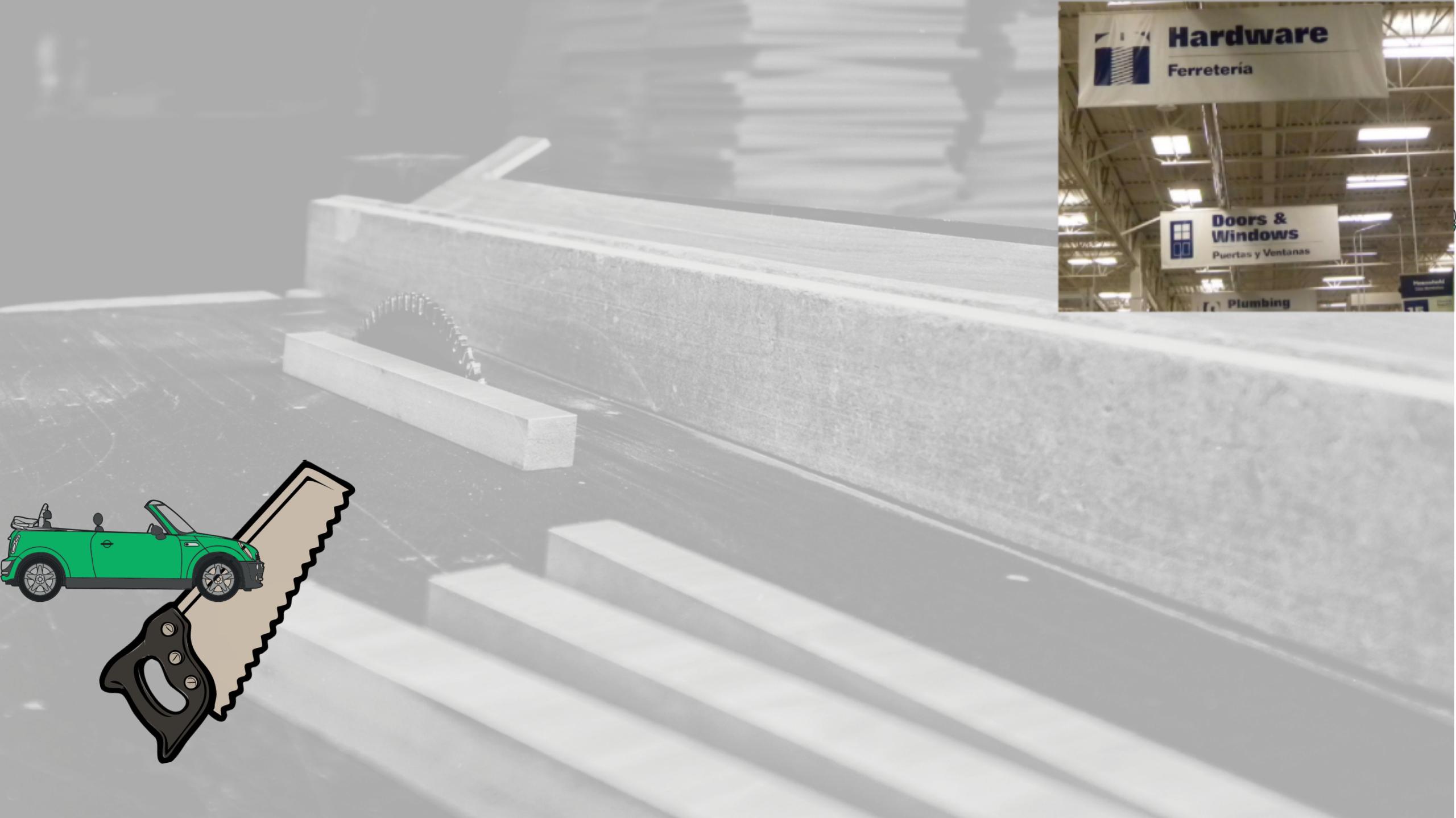
Powermatic · Cabinet · 10 inch · 3 horsepower

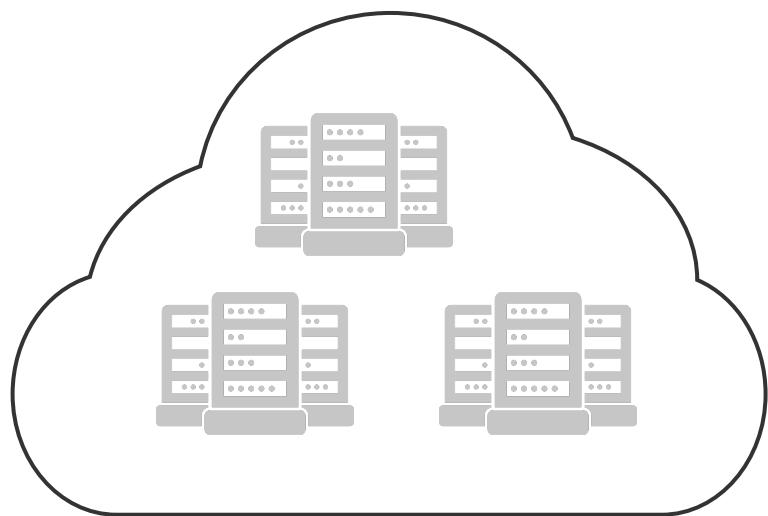
The PM2000 table saw has an array of features and massive design. It has the heaviest cast iron base that eliminates table vibration. It has a ...

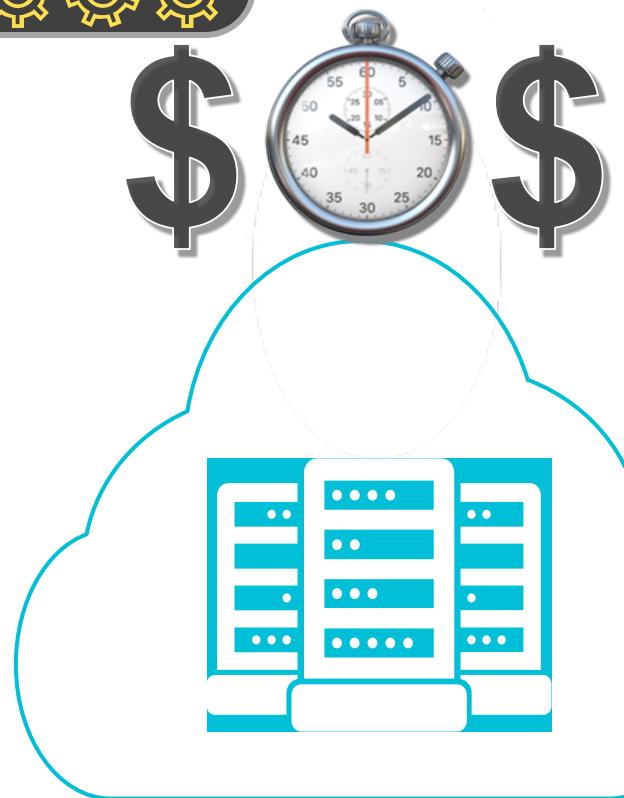
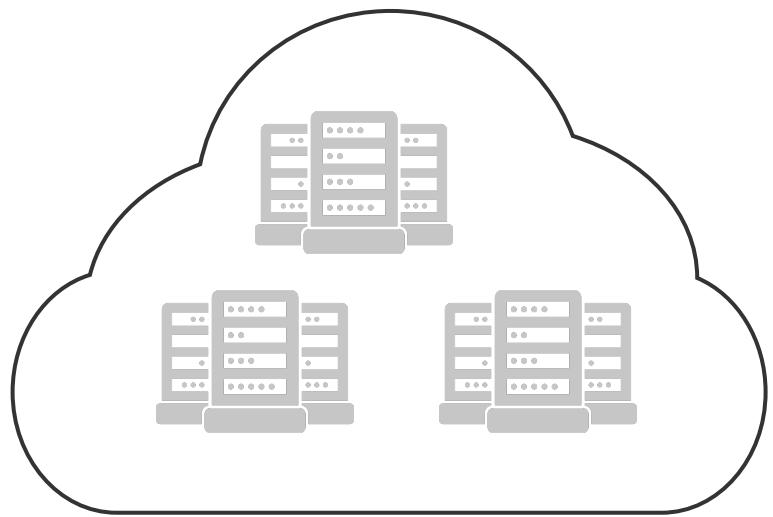


**Skid Steer Tree Saw Attachment Fixed Boom High Flow 20-30 gpm**

**\$6,995.00** from Spartan Equipment







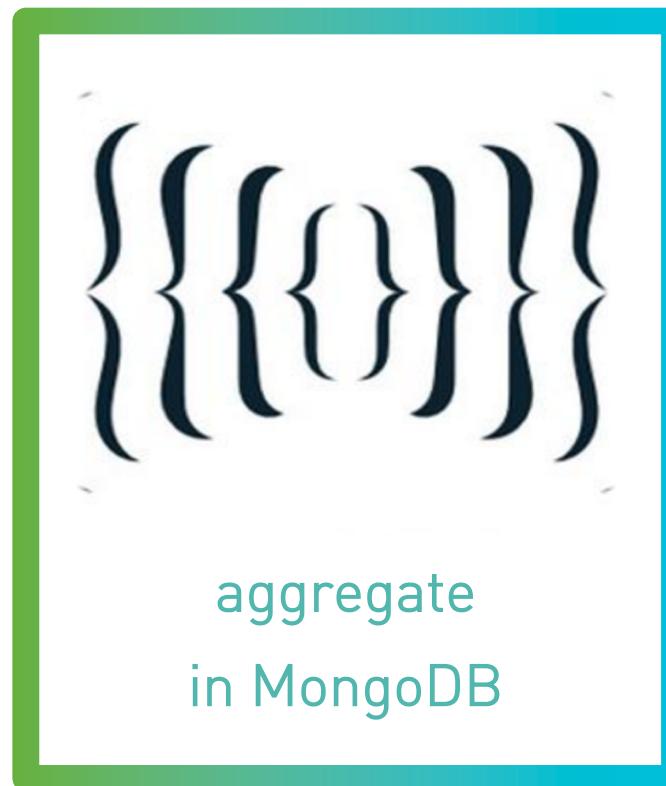
# OPTIONS FOR ANALYTICS



# OPTIONS FOR ANALYTICS



# ANALYTICS = AGGREGATION



# PIPELINE



*Ceci n'est pas une pipe.*



# PIPELINE

\*nix command line pipe

```
ps ax | grep mongod | head 1
```



# PIPELINE

MongoDB document pipeline

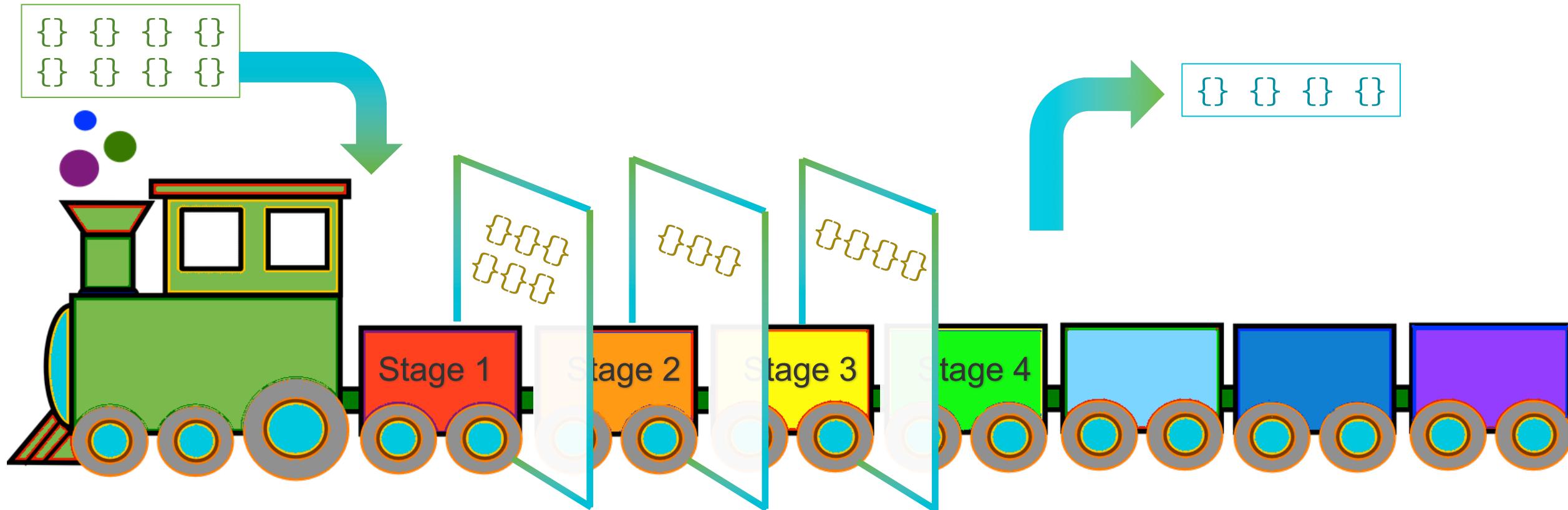
\$match | \$group | \$sort



Input stream {} {} {} {} → Result {} {} ...

START

Collection  
View  
Special stage



{"\$stage":{ ... }}

```
{ title: "The Great Gatsby",
language: "English",
subjects: [
  "Long Island",
  "New York",
  "1920s" ],
{ title: "War and Peace",
language: "Russian",
subjects: [
  "Russia",
  "War of 1812",
  "Napoleon" ],
{ title: "Open City",
language: "English",
subjects: [
  "New York",
  "Harlem" ] },
```

```
{title: "The Great Gatsby",
language: "English",
subjects: [
  "Long Island",
  "New York",
  "1920s" ],
{title: "Open City",
language: "English",
subjects: [
  "New York"
  "Harlem" ] }
```

```
{title: "The Great Gatsby",
language: "English",
subjects: "Long Island"}
{title: "The Great Gatsby",
language: "English",
subjects: "New York"}
{title: "The Great Gatsby",
language: "English",
subjects: "1920s"}
{title: "Open City",
language: "English",
subjects: "New York"}
{title: "Open City",
language: "English",
subjects: "Harlem"}
```

```
{ _id: "Long Island",
count: 1 },
{ _id: "New York",
count: 2 },
{ _id: "Long Island",
count: 1 },
{ _id: "1920s",
count: 1 },
{ _id: "Harlem",
count: 1 },
```

```
{"$unwind": "$subjects"} {"$sort": {"count": -1}} {"$limit": 3}
{"$match": {"language": "English"}} {"$group": {"_id": "$subjects", "count": {"$sum": 1}}} {"$project": ...}
```

\$match

\$unwind

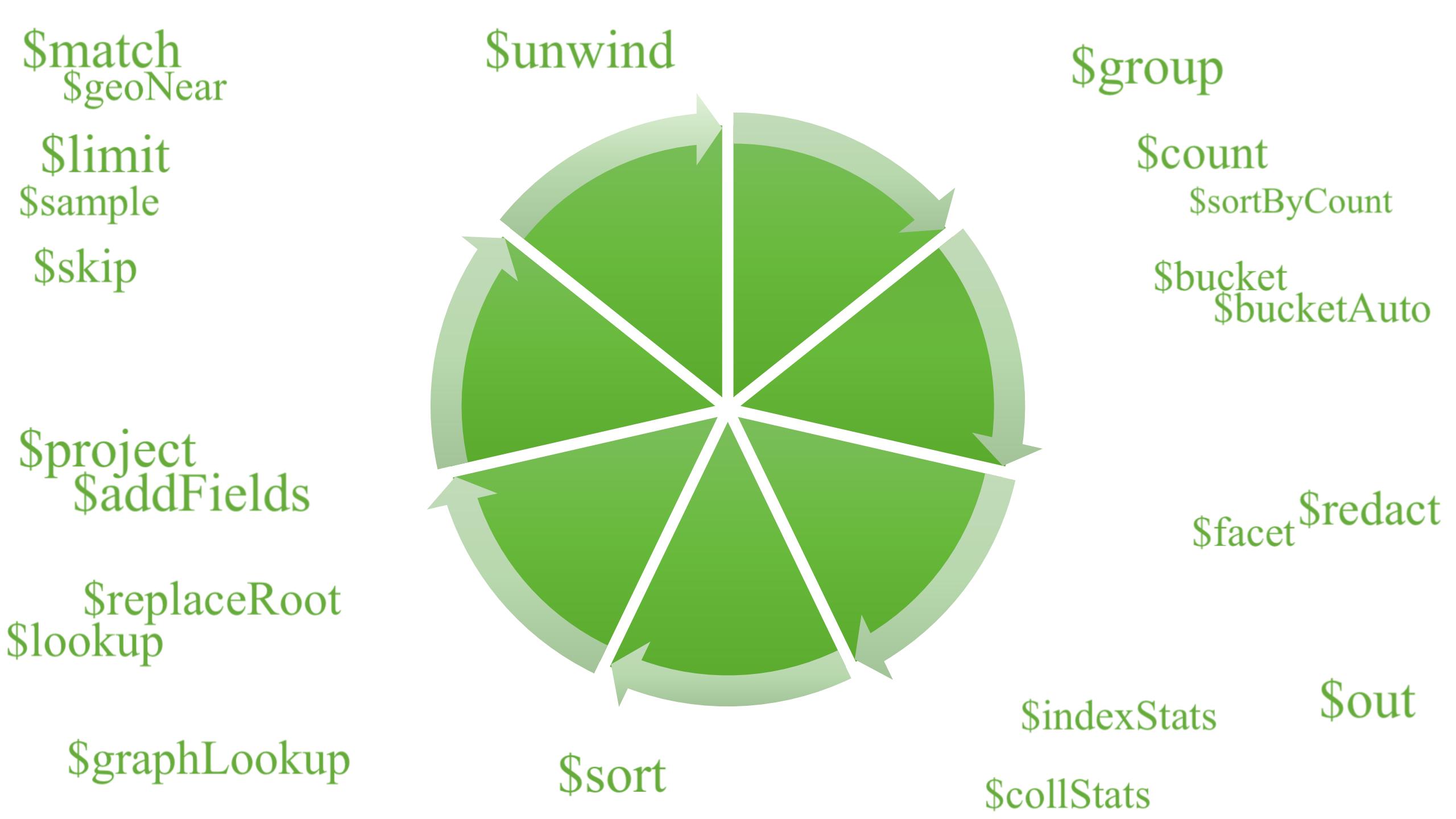
\$group

\$sort

\$limit

\$project

\$skip



\$match  
\$geoNear

\$limit  
\$sample

\$skip

\$project  
\$addFields

\$replaceRoot  
\$lookup

\$graphLookup

\$unwind

\$sort

\$indexStats  
\$collStats

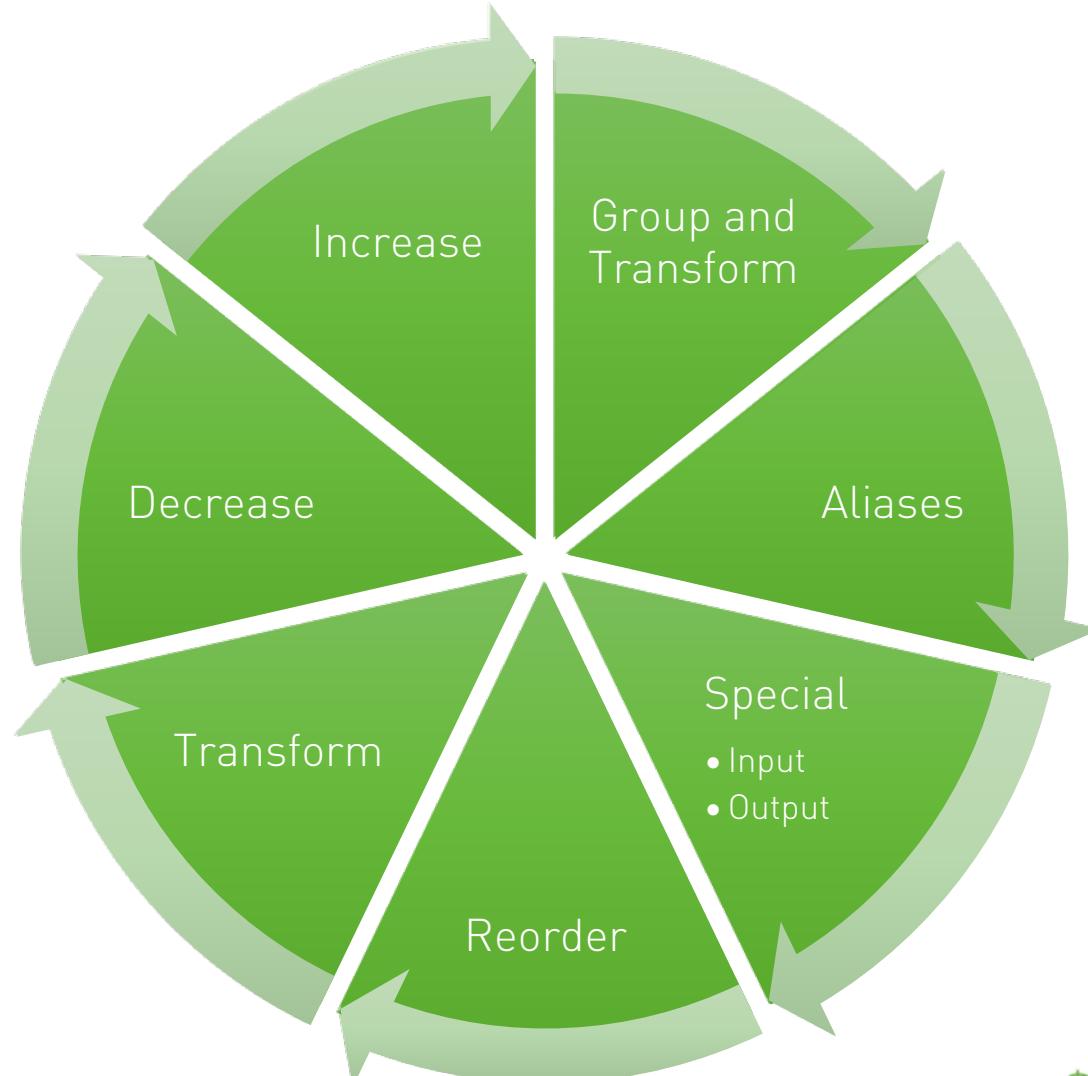
\$out

\$group

\$count  
\$sortByCount

\$bucket  
\$bucketAuto

\$facet \$redact



```
{ title: "The Great Gatsby",
  language: "English",
  subjects: [
    "Long Island",
    "New York",
    "1920s" ],
{ title: "War and Peace",
  language: "Russian",
  subjects: [
    "Russia",
    "War of 1812",
    "Napoleon" ],
{ title: "Open City",
  language: "English",
  subjects: [
    "New York"
    "Harlem" ] },
```

```
{title: "The Great Gatsby",
language: "English",
subjects: [
  "Long Island",
  "New York",
  "1920s" ],
{title: "Open City",
language: "English",
subjects: [
  "New York"
  "Harlem" ] }
```

```
{title: "The Great Gatsby",
language: "English".
  subjects: "Long Island"}
{title: "The Great Gatsby",
language: "English",
  subjects: "New York"}
{title: "The Great Gatsby",
language: "English",
  subjects: "1920s"}
{title: "Open City",
language: "English",
  subjects: "New York"}
{title: "Open City",
language: "English",
  subjects: "Harlem"}
```

```
{ _id: "Long Island",
  count: 1 },
{ _id: "New York",
  count: 1 },
{ _id: "1920s",
  count: 1 },
{ _id: "Harlem",
  count: 1 },
```

```
{ _id: "New York",
  count: 2 },
{ _id:"Long Island",
  count: 1 },
{ _id: "1920s",
  count: 1 },
{ _id: "Harlem",
  count: 1 },
```

```
  {"$unwind":"$subjects"} {"$sort":{"count": -1} {"$limit":3}}
{"$match":{"language": "English"} } {"$group":{"_id": "$subjects", "count": {"$sum": 1}} {"$project": ...}}
```

\$match

\$unwind

\$group

\$sort

\$limit

\$project

\$skip

\$match  
\$geoNear

\$limit  
\$sample

\$skip

\$project  
\$addFields

\$replaceRoot  
\$lookup

\$graphLookup

\$unwind

\$group

\$count  
\$sortByCount

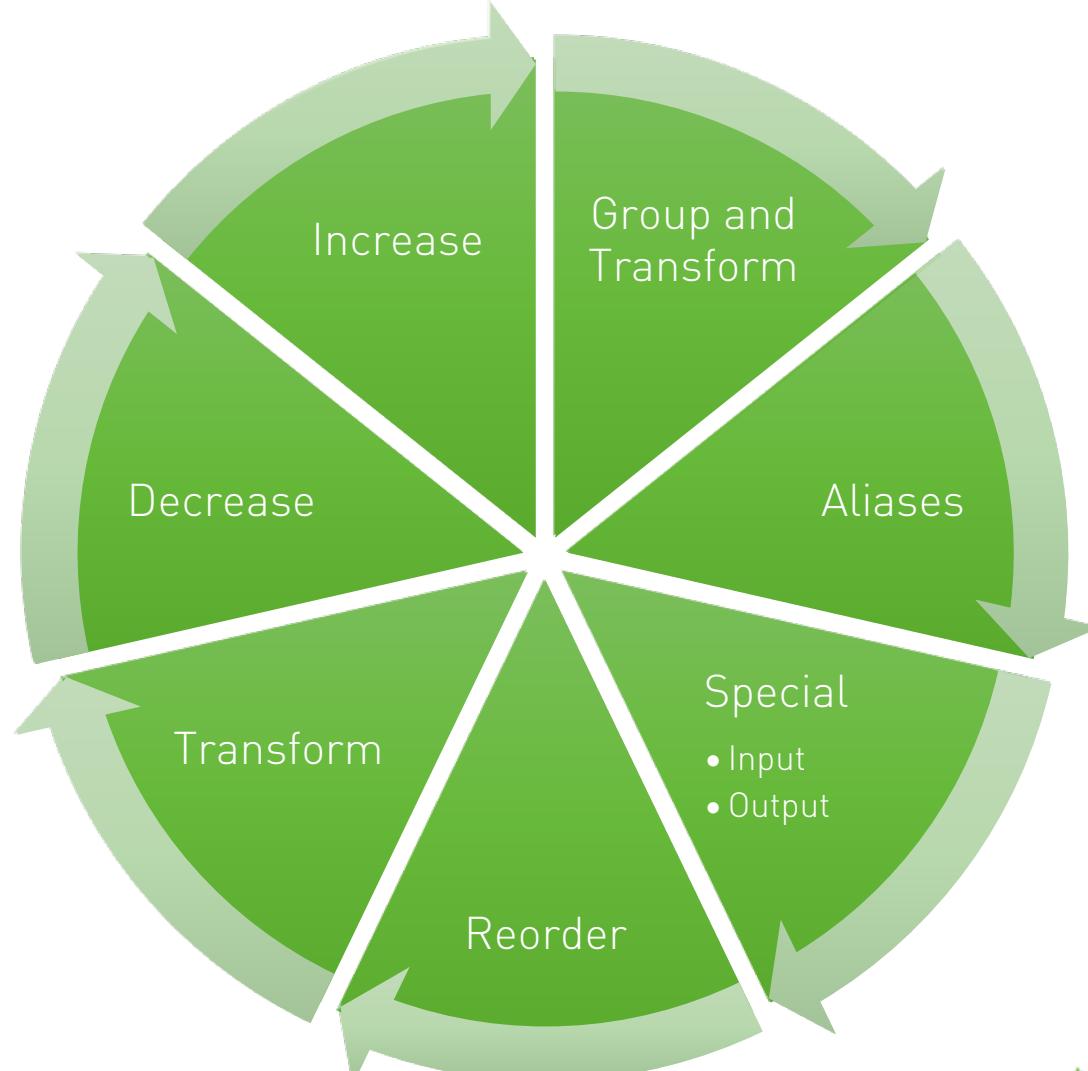
\$bucket  
\$bucketAuto

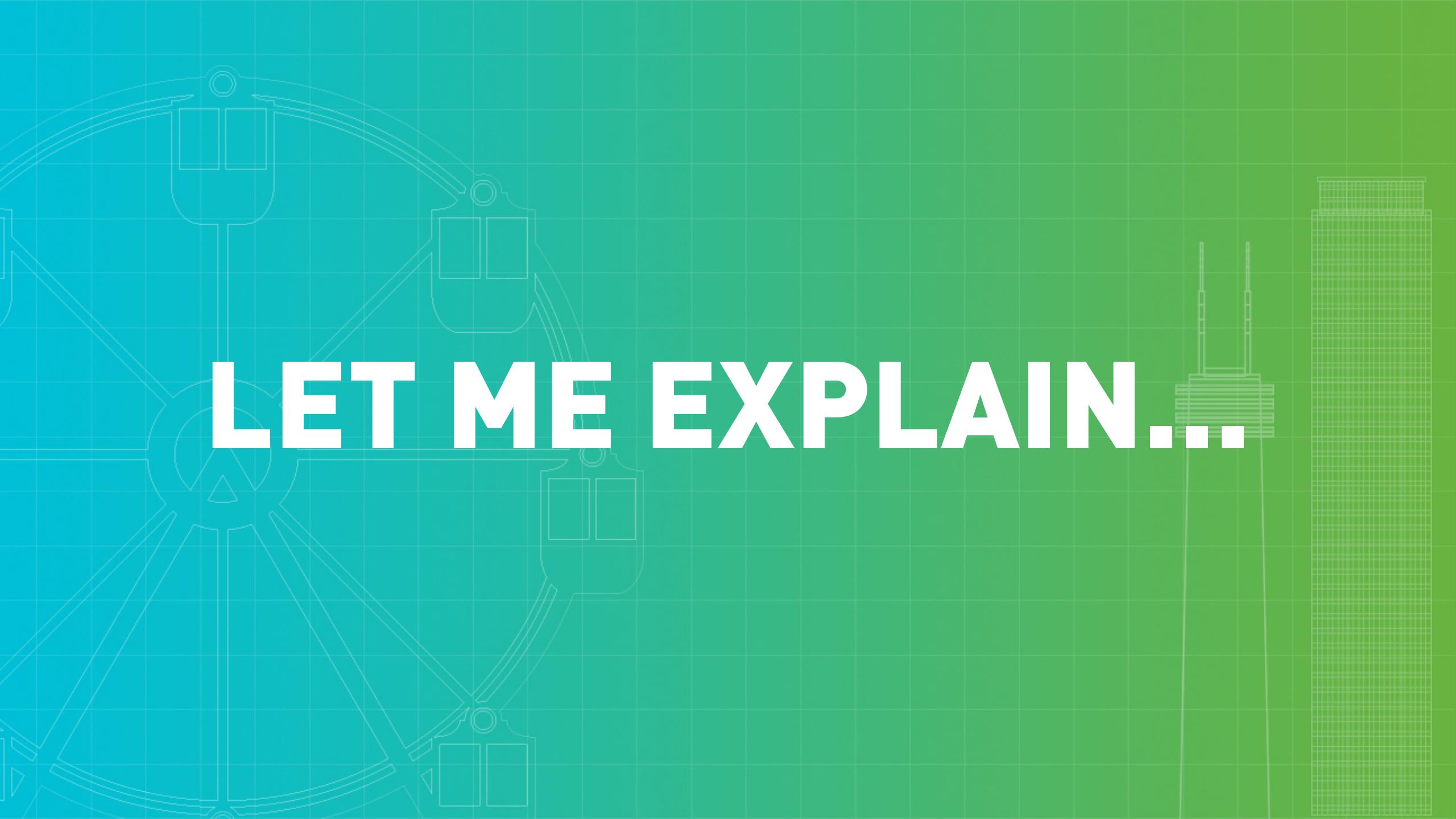
\$facet \$redact

\$indexStats  
\$collStats

\$out

\$sort





**LET ME EXPLAIN....**

```
db.books.aggregate([
  {$match: {"language": "English"}},
  {$unwind: "$subjects"},
  {$group: {"_id": "$subjects", "count": {"$sum": 1}}},
  {$sort: {"count": -1}},
  {$limit: 3}
], {explain: true})

{
  "stages": [
    {
      "$cursor": {
        "query": { "language": "English" },
        "fields": { "subjects": 1, "_id": 0 } ...
      }
    },
    {
      "$unwind": { "path": "$subjects" },
      "$group": { "_id": "$subjects", "count": { "$sum": { "$const": 1 } } },
      "$sort": {
        "sortKey": { "count": -1 },
        "limit": NumberLong(3)
      }
    }
  ]
}
```

```
db.books.aggregate([
  {$match:{"language":"English"}},
  {$unwind:"$subjects"},
  {$group:{_id:"$subjects",count:{$sum:1}}},
  {$sort:{count:-1}},
  {$limit:3}
],{explain:true})

{
  "stages": [
    {"$cursor": {"query": { "language": "English"}, "fields": { "subjects": 1,"_id": 0} ...}},
    {"$unwind": {"path": "$subjects"}},
    {"$group": {"_id": "$subjects","count": {"$sum": {"$const": 1}}}},
    {"$sort": {
      "sortKey": {"count": -1},
      "limit": NumberLong(3)
    }}
  ]
}
```

```
db.books.aggregate([
  {$match:{"language":"English"}},
  {$unwind:"$subjects"},
  {$group:{_id:"$subjects",count:{$sum:1}}},
  {$sort:{count:-1}},
  {$limit:3}
],{explain:true})
```

```
db.books.aggregate([
    {$unwind:"$subjects"},  

    {$group:{_id:"$subjects",count:{$sum:1}}},
    {$sort:{count:-1}},
    {$limit:3}
],{explain:true})  
  
{"stages": [  
    {"$cursor" : {"query" : { },  
                 "fields" : { "subjects" : 1,"_id" : 0} ...  
                }},  
    {"$unwind" : {"path" : "$subjects"}},  
    {"$group" : {"_id" : "$subjects","count" : {"$sum" : {"$const" : 1}}} },  
    {"$sort" : {  
        "sortKey" : {"count" : -1},  
        "limit" : NumberLong(3)  
    }}  
] }
```

```
db.books.aggregate([
  {$unwind:"$subjects"},

  {$group:{_id:"$subjects",count:{$sum:1}}},
  {$sort:{count:-1}},
  {$limit:3}
],{explain:true})
```

```
db.books.aggregate([
  {$unwind:"$subjects"},  

  {$match:{language:"English"}},  

  {$group:{_id:"$subjects",count:{$sum:1}}},  

  {$sort:{count:-1}},  

  {$limit:3}
],{explain:true})  
  
{"stages": [  
    {"$cursor": {"query": { "language": "English"},  
                 "fields": { "subjects": 1,"_id": 0} ...  
               },  
    {"$unwind": {"path": "$subjects"}},  
    {"$group": {"_id": "$subjects","count" : {"$sum" : {"$const" : 1}}} ,  
    {"$sort" : {  
        "sortKey" : {"count" : -1},  
        "limit" : NumberLong(3)  
      }  
  ] }
```

```
db.books.aggregate([
  {$unwind:"$subjects"},
  {$match:{language:"English",subjects:/^ABC/}},
  {$group:{_id:"$subjects",count:{$sum:1}}},
  {$sort:{count:-1}},
  {$limit:3}
],{explain:true})
```

```
{"stages": [
  {"$cursor": {"query": { "language": "English"}, "fields": { "subjects": 1," id": 0} ...}},
  {"$unwind": ["path": "$subjects"]},
  {"$match": {"subjects": {"$regex": "ABC"}}, {"$group": {"_id": "$subjects","count": {"$sum": {"$const": 1}}}}, {"$sort": {
    "sortKey": {"count": -1},
    "limit": NumberLong(3)
  }}}
] }
```

```
db.books.aggregate([
  {$unwind:"$subjects"},
  {$match:{language:"English",subjects:/^ABC/}},
  {$group:{_id:"$subjects",count:{$sum:1}}},
  {$sort:{count:-1}},
  {$limit:3}
],{explain:true})
```

```
{"stages": [
  {"$cursor": {"query": {language: "English"},  
              "fields": {subjects: 1,_id: 0} ...  
            },  
   {"$unwind": {"path": "$subjects"}},  
   {"$match": {"subjects": {"$regex": "^[ABC]"}},  
    {"$group": {"_id": "$subjects", "count": {"$sum": {"$const": 1}}}},  
    {"$sort": {
      "sortKey": {"count": -1},
      "limit": NumberLong(3)
    }}
]
```



```
db.books.aggregate([
  {$unwind:"$subjects"},
  {$match:{language:"English",subjects:/^ABC/}},
  {$group:{_id:"$subjects",count:{$sum:1}}},
  {$sort:{count:-1}},
  {$limit:3}
],{explain:true})
```

```
{"stages": [
  {"$cursor": {"query": {language: "English"},  

    "fields": {subjects: 1,_id: 0} ...  

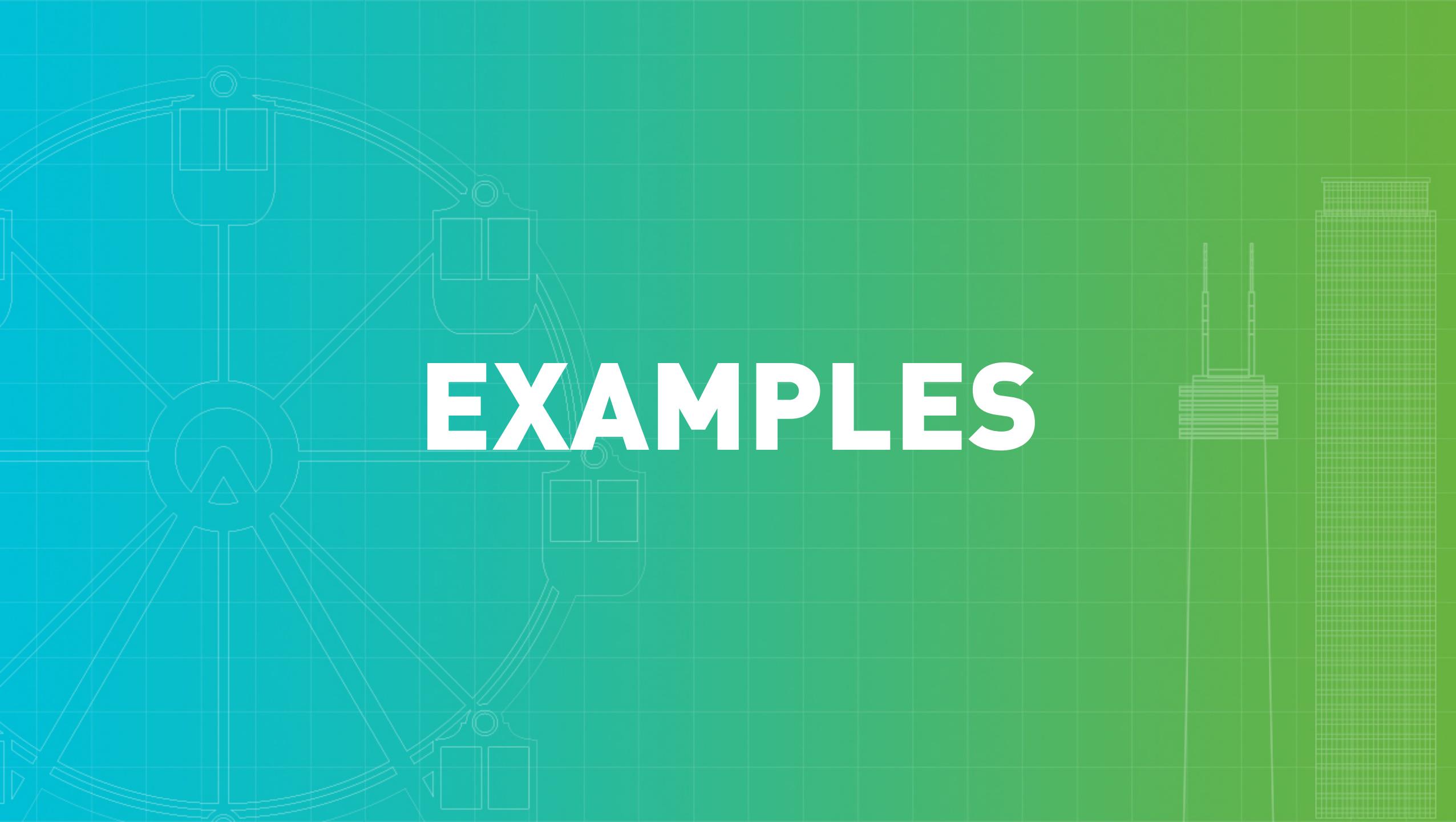
  }},
  {"$unwind": {"path": "$subjects"}},  

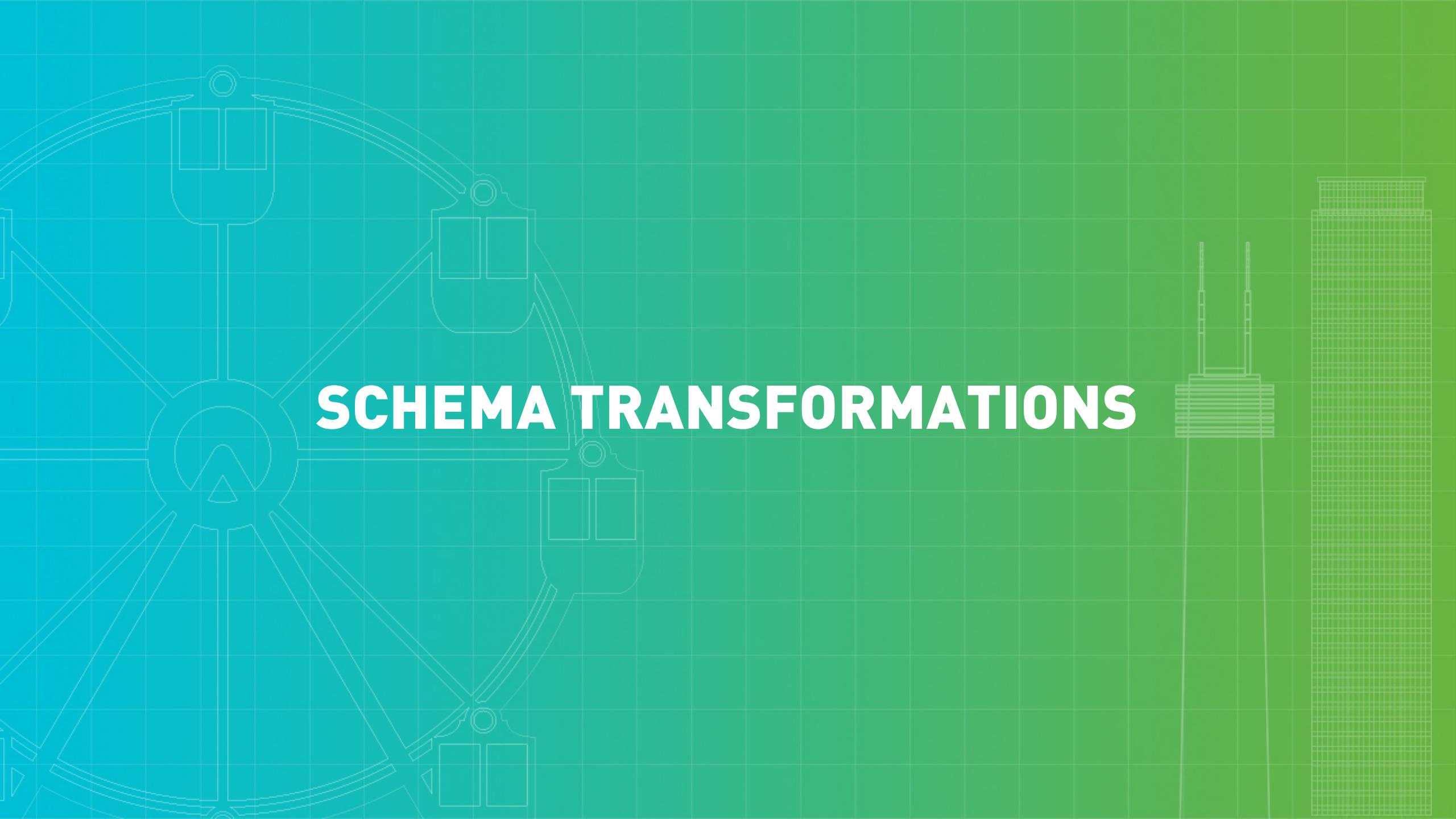
  {"$match": {"subjects": {"$regex": "^[ABC]"}},  

  {"$group": {"_id": "$subjects", "count": {"$sum": {"$const": 1}}}},  

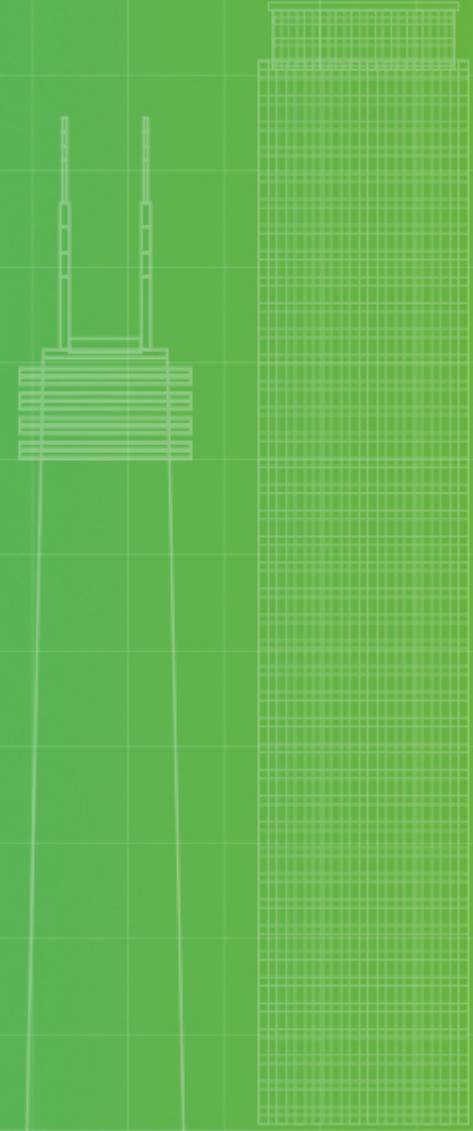
  {"$sort": {
    "sortKey": {"count": -1},
    "limit": NumberLong(3)
  }}
]
```

# EXAMPLES



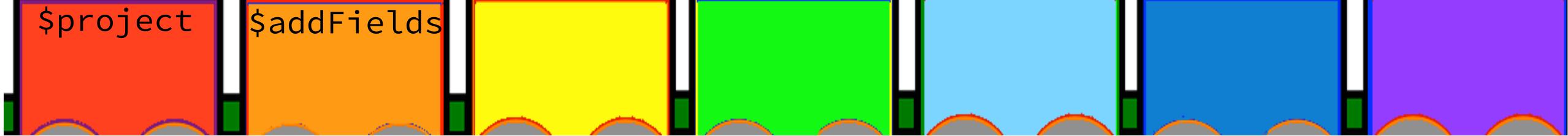


# SCHEMA TRANSFORMATIONS



```
{"body": {  
  "VMESSAGE": {  
    "072ade7d42d8": {  
      "msgId": "bcd1d9",  
      "date": 1486546629585,  
      "status": "accept",  
      "comment": "hi"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "595d0a56cff2": {  
      "msgId": "595d0a",  
      "date": 1486566646197,  
      "status": "reject",  
      "comment": "no good"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "52ffd09bf5b5": {  
      "msgId": "1dadce",  
      "date": 1486568943752,  
      "status": "accept"  
    } } } }
```

```
{"message": {  
  "vid": "072ade7d42d8",  
  "msgId": "bcd1d9",  
  "date": 1486546629585,  
  "status": "accept",  
  "comment": "hi"  
} },  
{"message": {  
  "vid": "595d0a56cff2",  
  "msgId": "595d0a",  
  "date": 1486566646197,  
  "status": "reject",  
  "comment": "no good"  
} },  
{"message": {  
  "vid": "52ffd09bf5b5",  
  "msgId": "1dadce",  
  "date": 1486568943752,  
  "status": "accept"  
} }
```



```
{"body": {  
  "VMESSAGE": {  
    "072ade7d42d8": {  
      "msgId": "bcd1d9",  
      "date": 1486546629585,  
      "status": "accept",  
      "comment": "hi"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "595d0a56cff2": {  
      "msgId": "595d0a",  
      "date": 1486566646197,  
      "status": "reject",  
      "comment": "no good"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "52ffd09bf5b5": {  
      "msgId": "1dadce",  
      "date": 1486568943752,  
      "status": "accept"  
    } } } }
```

New Expressions (3.)

```
{"$objectToArray":  
  => array of k/v pairs  
  
{$arrayToObject":  
  => object}
```

Object:

```
{  
  "072ade7d42d8": {  
    "messageId": "bcd1d9",  
    "date": 1486546629585,  
    "status": "accept",  
    "comment": "hi"  
  } }  
  
$objectToArray:  
[ {  
  "k": "072ade7d42d8",  
  "v": {  
    "messageId": "bcd1d991",  
    "date": 1486546629585,  
    "status": "accept",  
    "comment": "hi"  
  }  
}  
]  
  
Need array: [  
  {"k": "vid", "v": "072ade7d42d8"},  
  {"k": "messageId", "v": "bcd1d9"},  
  {"k": "date", "v": 1486546629585},  
  {"k": "status", "v": "accept"},  
  {"k": "comment", "v": "hi"}]
```

```
{"message": {  
  "vid": "072ade7d42d8",  
  "msgId": "bcd1d9",  
  "date": 1486546629585,  
  "status": "accept",  
  "comment": "hi"  
} },  
{"message": {  
  "vid": "595d0a56cff2",  
  "msgId": "595d0a",  
  "date": 1486566646197,  
  "status": "reject",  
  "comment": "no good"  
} },  
{"message": {  
  "vid": "52ffd09bf5b5",  
  "msgId": "1dadce",  
  "date": 1486568943752,  
  "status": "accept"  
} }
```

\$addFields

```
{"body": {  
  "VMESSAGE": {  
    "072ade7d42d8": {  
      "msgId": "bcd1d9",  
      "date": 1486546629585,  
      "status": "accept",  
      "comment": "hi"  
    } } },  
  {"body": {  
    "VMESSAGE": {  
      "595d0a56cff2": {  
        "msgId": "595d0a",  
        "date": 1486566646197,  
        "status": "reject",  
        "comment": "no good"  
      } } },  
  {"body": {  
    "VMESSAGE": {  
      "52ffd09bf5b5": {  
        "msgId": "1dadce",  
        "date": 1486568943752,  
        "status": "accept"  
      } } } }
```

```
  {"$addFields": {  
    "message": {  
      "$arrayToObject": ...  
      {"$objectToArray": "$body.VMESSAGE"}  
    } } }
```

```
  {"message": {  
    "vid": "072ade7d42d8",  
    "msgId": "bcd1d9",  
    "date": 1486546629585,  
    "status": "accept",  
    "comment": "hi" } },  
  {"message": {  
    "vid": "595d0a56cff2",  
    "msgId": "595d0a",  
    "date": 1486566646197,  
    "status": "reject",  
    "comment": "no good" } },  
  {"message": {  
    "vid": "52ffd09bf5b5",  
    "msgId": "1dadce",  
    "date": 1486568943752,  
    "status": "accept" } }
```

\$addFields

```
{"body": {  
  "VMESSAGE": {  
    "072ade7d42d8": {  
      "msgId": "bcd1d9",  
      "date": 1486546629585,  
      "status": "accept",  
      "comment": "hi"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "595d0a56cff2": {  
      "msgId": "595d0a",  
      "date": 1486566646197,  
      "status": "reject",  
      "comment": "no good"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "52ffd09bf5b5": {  
      "msgId": "1dadce",  
      "date": 1486568943752,  
      "status": "accept"  
    } } } }
```

```
{"$addFields": {  
  "bvm": {"$objectToArray": "$body.VMESSAGE"}  
} },  
{$addFields": {  
  "bvm2": {"$arrayElemAt": ["$bvm", 0]}  
} },  
{$addFields": {  
  "bvm": {"$objectToArray": "$bvm2.v"}  
} },  
{$addFields": {  
  "msgarr": {  
    }  
} },  
{$addFields": {  
  "message": {"$arrayToObject": "$msgarr"}  
} }
```

```
{"message": {  
  "vid": "072ade7d42d8",  
  "msgId": "bcd1d9",  
  "date": 1486546629585,  
  "status": "accept",  
  "comment": "hi"  
} },  
{"message": {  
  "vid": "595d0a56cff2",  
  "msgId": "595d0a",  
  "date": 1486566646197,  
  "status": "reject",  
  "comment": "no good"  
} },  
{"message": {  
  "vid": "52ffd09bf5b5",  
  "msgId": "1dadce",  
  "date": 1486568943752,  
  "status": "accept"  
} }
```

```
{"body": {  
  "VMESSAGE": {  
    "072ade7d42d8": {  
      "msgId": "bcd1d9",  
      "date": 1486546629585,  
      "status": "accept",  
      "comment": "hi"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "595d0a56cff2": {  
      "msgId": "595d0a",  
      "date": 1486566646197,  
      "status": "reject",  
      "comment": "no good"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "52ffd09bf5b5": {  
      "msgId": "1dadce",  
      "date": 1486568943752,  
      "status": "accept"  
    } } } }
```

```
{"$addFields": {  
  "bvm": {"$objectToArray": "$body.VMESSAGE"}  
} },  
{"$addFields": {  
  "bvm2": {"$arrayElemAt": ["$bvm", 0]}  
} },  
{"$addFields": {  
  "bvm": {"$objectToArray": "$bvm2.v"}  
} },  
{"$addFields": {  
  "msgarr": {  
    "$concatArrays": [  
      { "k": "vid", "v": "$bvm2.k" },  
      "$bvm"  
    ]  
  }  
} },  
{"$addFields": {  
  "message": {"$arrayToObject": "$msgarr"}  
} }
```

```
{"message": {  
  "vid": "072ade7d42d8",  
  "msgId": "bcd1d9",  
  "date": 1486546629585,  
  "status": "accept",  
  "comment": "hi"  
} },  
{"message": {  
  "vid": "595d0a56cff2",  
  "msgId": "595d0a",  
  "date": 1486566646197,  
  "status": "reject",  
  "comment": "no good"  
} },  
{"message": {  
  "vid": "52ffd09bf5b5",  
  "msgId": "1dadce",  
  "date": 1486568943752,  
  "status": "accept"  
} }
```

```
{"body": {  
  "VMESSAGE": {  
    "072ade7d42d8": {  
      "msgId": "bcd1d9",  
      "date": 1486546629585,  
      "status": "accept",  
      "comment": "hi"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "595d0a56cff2": {  
      "msgId": "595d0a",  
      "date": 1486566646197,  
      "status": "reject",  
      "comment": "no good"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "52ffd09bf5b5": {  
      "msgId": "1dadce",  
      "date": 1486568943752,  
      "status": "accept"  
    } } } }
```

```
{"$addFields":{  
  "message":{  
    "$arrayToObject":{  
      "$let":{  
        "vars":{  
          "elem": {"$arrayElemAt:[  
            {"$objectToArray":"$body.VMESSAGE"},  
            0  
          ]}  
        },  
        "in": {"$concatArrays": [  
          [ { k: "vid", v: "$$elem.k" } ],  
          {$objectToArray:"$$elem.v"}  
        ]}  
      }  
    }  
  }  
}
```

```
{"message": {  
  "vid": "072ade7d42d8",  
  "msgId": "bcd1d9",  
  "date": 1486546629585,  
  "status": "accept",  
  "comment": "hi" } },  
{"message": {  
  "vid": "595d0a56cff2",  
  "msgId": "595d0a",  
  "date": 1486566646197,  
  "status": "reject",  
  "comment": "no good" } },  
{"message": {  
  "vid": "52ffd09bf5b5",  
  "msgId": "1dadce",  
  "date": 1486568943752,  
  "status": "accept" } }
```

\$addFields

```
{"body": {  
  "VMESSAGE": {  
    "072ade7d42d8": {  
      "msgId": "bcd1d9",  
      "date": 1486546629585,  
      "status": "accept",  
      "comment": "hi"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "595d0a56cff2": {  
      "msgId": "595d0a",  
      "date": 1486566646197,  
      "status": "reject",  
      "comment": "no good"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "52ffd09bf5b5": {  
      "msgId": "1dadce",  
      "date": 1486568943752,  
      "status": "accept"  
    } } } }
```

```
{"$addFields":{  
  "message":{  
    "$arrayToObject":{  
      "$let":{  
        "vars":{  
          "elem": {"$arrayElemAt:[  
            {"$objectToArray":"$body.VMESSAGE"},  
            0  
          ]}  
        },  
        "in": {"$concatArrays": [  
          [ { k: "vid", v: "$$elem.k" } ],  
          {$objectToArray:"$$elem.v"}  
        ]}  
      }  
    }  
  }  
}
```

```
{"message": {  
  "vid": "072ade7d42d8",  
  "msgId": "bcd1d9",  
  "date": 1486546629585,  
  "status": "accept",  
  "comment": "hi" } },  
{"message": {  
  "vid": "595d0a56cff2",  
  "msgId": "595d0a",  
  "date": 1486566646197,  
  "status": "reject",  
  "comment": "no good" } },  
{"message": {  
  "vid": "52ffd09bf5b5",  
  "msgId": "1dadce",  
  "date": 1486568943752,  
  "status": "accept" } }
```

\$addFields

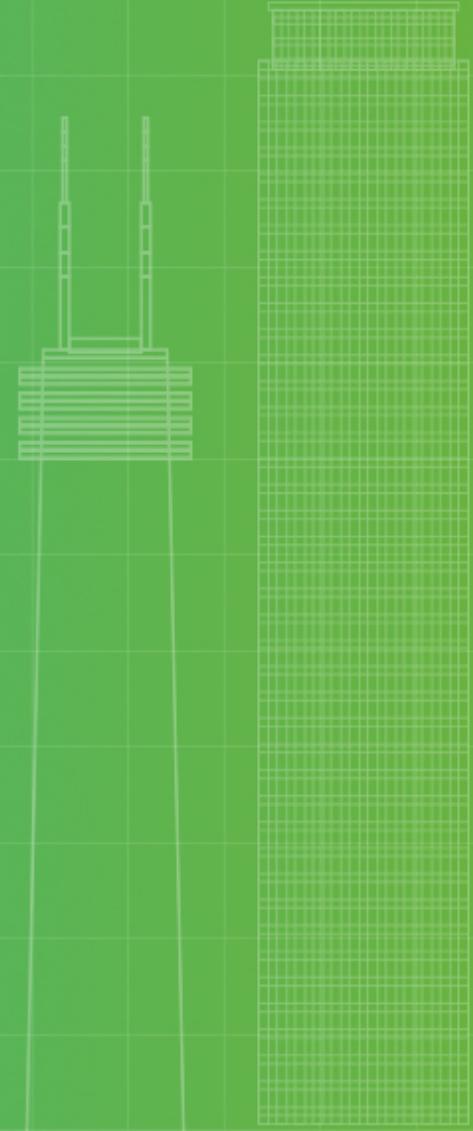
```
{"body": {  
  "VMESSAGE": {  
    "072ade7d42d8": {  
      "msgId": "bcd1d9",  
      "date": 1486546629585,  
      "status": "accept",  
      "comment": "hi"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "595d0a56cff2": {  
      "msgId": "595d0a",  
      "date": 1486566646197,  
      "status": "reject",  
      "comment": "no good"  
    } } } },  
{"body": {  
  "VMESSAGE": {  
    "52ffd09bf5b5": {  
      "msgId": "1dadce",  
      "date": 1486568943752,  
      "status": "accept"  
    } } } }
```

```
{"$addFields":{  
  "message":{  
    "$arrayToObject":{  
      "$let":{  
        "vars":{  
          "elem": {"$arrayElemAt:[  
            {"$objectToArray":"$body.VMESSAGE"},  
            0  
          ]}  
        },  
        "in": {"$concatArrays": [  
          { k: "vid", v: "$$elem.k" } ],  
          {$objectToArray:"$$elem.v"}  
        ]}  
      }  
    }  
  }  
}
```

```
{"message": {  
  "vid": "072ade7d42d8",  
  "msgId": "bcd1d9",  
  "date": 1486546629585,  
  "status": "accept",  
  "comment": "hi" } },  
{"message": {  
  "vid": "595d0a56cff2",  
  "msgId": "595d0a",  
  "date": 1486566646197,  
  "status": "reject",  
  "comment": "no good" } },  
{"message": {  
  "vid": "52ffd09bf5b5",  
  "msgId": "1dadce",  
  "date": 1486568943752,  
  "status": "accept" } }
```

\$addFields

# ARRAY MANIPULATIONS



# ARRAY MANIPULATIONS

`$arrayElemAt`

`$concatArrays`

`$indexOfArray`

`$isArray`

`$size`

`$range`

`$reverseArray`

`$map`

`$reduce`

`$filter`

`$slice`

`$zip`

`$in`

... plus all the set expressions

# ARRAY MANIPULATIONS

**\$arrayElemAt**

**\$concatArrays**

**\$indexOfArray**

**\$isArray**

**\$size**

**\$range**

**\$reverseArray**

**\$map**

**\$reduce**

**\$filter**

**\$slice**

**\$zip**

**\$in**

... plus all the set expressions

# ARRAY MANIPULATIONS

`$arrayElemAt`

`$concatArrays`

`$indexOfArray`

`$isArray`

`$size`

`$range`

`$reverseArray`

`$map`

`$reduce`

`$filter`

`$slice`

`$zip`

`$in`

... plus all the set expressions

# ARRAY MANIPULATIONS

\$map

input: array

output: array

\$filter

input: array

output: subset of array

\$reduce

input: array

output: anything

you

want

# ARRAY MANIPULATIONS: \$map

```
"arr": [{"a":1}, {"a":99}, {"a":5}, {"a":3}]
```

```
{"$map": {"$map": {"input": "$arr", "as": "var", "in": "$$this"}}, {"$map": {"input": "$arr", "as": "eachElement", "in": {"b": "$eachElement.a"}}, {"$map": {"input": {"$range": [0, {"$size": "$arr"}]}, "as": "index", "in": {"c": {"$arrayElemAt": ["$arr.a", "$$index"]}}}}
```

"a"	1
"a"	99
"a"	5
"a"	3

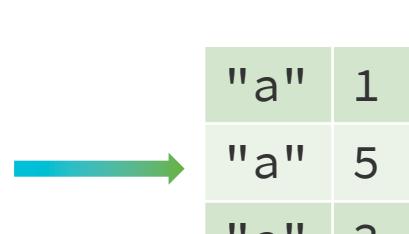
"a"	1
"a"	99
"a"	5
"a"	3

0	
1	
2	
3	

# ARRAY MANIPULATIONS: \$filter

```
"arr": [{"a":1}, {"a":99}, {"a":5}, {"a":3}]
```

```
{"$filter":  
  "input": "$arr",  
  "cond": {"$lt: ["$$this.a", 10]}  
}  
  
{"$filter":{  
  "input": "$arr",  
  "as": "elem",  
  "cond": {"$lt": ["$$elem.a", 10]}  
} }
```



"a"	1
"a"	99
"a"	5
"a"	3

"a"	1
"a"	5
"a"	3

# ARRAY MANIPULATIONS: \$reduce

```
"arr": [{"a":1}, {"a":99}, {"a":5}, {"a":3}]
```

```
{"$reduce": {  
    "input": "$arr",  
    "initialValue": 0,  
    "in": {$add: ["$$value", "$$this.a"]}  
} }
```

"a"	1
"a"	99
"a"	5
"a"	3



# ARRAY MANIPULATIONS: \$reduce

```
"arr": [{"a":1}, {"a":99}, {"a":5}, {"a":3}]
```

```
{"$reduce": {  
    "input": "$arr",  
    "initialValue": 0,  
    "in": {$add: ["$$value", "$$this.a"]}  
} }
```

```
{"$reduce":{  
    "input": "$arr",  
    "intialValue": [],  
    "in": {"$concatArrays": [  
        [ $$this ],  
        "$$value"  
    ]}  
} }
```

"a"	1
"a"	99
"a"	5
"a"	3



108

"a"	1
"a"	99
"a"	5
"a"	3

# ARRAY MANIPULATIONS: \$reduce

```
"arr": [{"a":1}, {"a":99}, {"a":5}, {"a":3}]
```

```
{"$reduce": {  
    "input": "$arr",  
    "initialValue": 0,  
    "in": {$add: ["$$value", "$$this.a"]}  
} }
```

"a"	1
"a"	99
"a"	5
"a"	3



108

```
{"$reduce": {  
    "input": "$arr",  
    "intialValue": [],  
    "in": {"$concatArrays": [  
        [ $$this ],  
        "$$value"  
    ]}  
} }
```

"a"	1		"a"	3
"a"	99		"a"	5
"a"	5		"a"	99
"a"	3		"a"	1

# READABILITY TIP



# ENCAPSULATE COMPLEXITY

Functions for expressions

```
reverseArray = function(input) {  
    return {"$reduce":{  
        "input": input,  
        "intialValue":[],  
        "in": {"$concatArrays": [  
            "$$this",  
            "$$value"  
        ]}  
    }};  
};  
  
db.c.aggregate([ {"$addFields":{  
    "revArray":reverseArray("$origArray")  
} } ])
```

# ENCAPSULATE COMPLEXITY

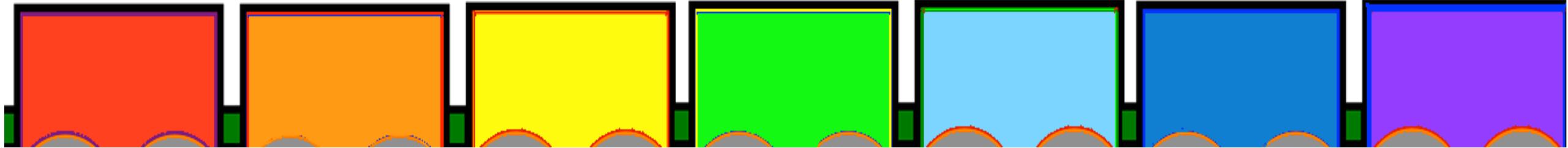
Functions for expressions

```
db.c.aggregate([ {"$addFields":{  
    "sortedArray":sortArray("$origArray")  
} } ])  
sortArray = function( inputArray, sortField="", asc=false)  
{  
    var suffix = "";  
    var maxF = MaxKey;  
    var minF = MinKey;  
    if (sortField != "") {  
        if (asc) {  
            maxF = MinKey;  
            minF = MaxKey;  
        }  
        for (var i = 0; i < inputArray.length; i++) {  
            var curF = inputArray[i][sortField];  
            if (curF > maxF) {  
                maxF = curF;  
            }  
            if (curF < minF) {  
                minF = curF;  
            }  
        }  
        for (var i = 0; i < inputArray.length; i++) {  
            var curF = inputArray[i][sortField];  
            if (curF == maxF) {  
                inputArray[i] = inputArray[minF];  
                inputArray[minF] = inputArray[i];  
            }  
        }  
    }  
    return inputArray;  
}
```

# NETWORK SUSPECT ACTIVITY DETECTION

```
start=ISODate("...")  
end=ISODate("...")  
{  
    user: "303900",  
    ipaddr: "71.56.112.56",  
    ts:ISODate("2017-05-08T05:28:13Z")  
}
```

---



```
start=ISODate("...")  
end=ISODate("...")  
{  
    user: "303900",  
    ipaddr: "71.56.112.56",  
    ts:ISODate("2017-05-08T...")  
}  
  
{$match:{ts:{$gte:start,$lt:end}}},  
{$sort:{ts:1}},  
{$group:{_id:"$user",ips:{$push:{ip:"$ipaddr", ts:"$ts"}}}},
```

```
{ _id: "303900",  
  ips: [  
    {ip:"71.56.112.56",  
     ts:ISODate("2017-05-08T08:54:04Z")  
    },  
    {ip:"71.56.112.56",  
     ts:ISODate("2017-05-09T09:01:11Z")  
    },  
    {ip:"12.130.117.87",  
     ts:ISODate("2017-05-09T09:04:59Z")  
    }  
]
```

\$match

\$sort

\$group

```
start=ISODate("...")  
end=ISODate("...")  
{  
    user: "303900",  
    ipaddr: "71.56.112.56",  
    ts:ISODate("2017-05-08T...")  
}  
  
[ { _id: "303900",  
    ips: [  
        {ip:"71.56.112.56",  
         ts:ISODate("2017-05-08T...")  
        },  
        {ip:"71.56.112.56",  
         ts:ISODate("2017-05-09T...")  
        },  
        {ip:"12.130.117.87",  
         ts:ISODate("2017-05-09T...")  
        }  
    ]}  
  
    {$match:{ts:{$gte:start,$lt:end}}},  
    {$sort:{ts:1}},  
    {$group:{_id:"$user",ips:{$push:{ip:"$ipaddr", ts:"$ts"}}}},  
    {$addFields:{diffs: {$filter:{  
        input:{$map:{  
            input: {$range:[0,{$subtract:[{$size:"$ips"},1]}]}, as:"i",  
            in:{$let:{vars:{ip1:$arrayElemAt:["$ips","$$i"]},  
                    ip2:$arrayElemAt:["$ips",{$add:[ "$$i",1]}]}},  
            in:{  
                diff:{$divide:[{$abs:{$subtract:[ $$ip1.ts,"$$ip2.ts"]}},60000]},  
                ip1:"$$ip1.ip", t1:"$$ip1.ts",  
                ip2:"$$ip2.ip", t2:"$$ip2.ts"  
            }}}},  
            cond:{$and:[{$lt:[ $$this.diff,10]},{$ne:[ $$this.ip1,"$$this.ip2"]}]}}},  
        }]}],
```

\$match

\$sort

\$group

\$addFields

\$match

\$project

```
start=ISODate("...")  
end=ISODate("...")  
{  
    user: "303900",  
    ipaddr: "71.56.112.56",  
    ts:ISODate("2017-05-08T...")  
}  
  
[{"_id": "303900",  
    "ips": [  
        {"ip": "71.56.112.56",  
            "ts": ISODate("2017-05-08T...")},  
        {"ip": "71.56.112.56",  
            "ts": ISODate("2017-05-09T...")},  
        {"ip": "12.130.117.87",  
            "ts": ISODate("2017-05-09T...")}  
    ]}  
  
{$match: {ts: {$gte: start, $lt: end}}},  
{$sort: {ts: 1}},  
{$group: {_id: "$user", ips: {$push: {ip: "$ipaddr", ts: "$ts"}}}},  
{$addFields: {diffs: {$filter: {  
        input: {$map: {  
            input: {$range: [0, {$subtract: [{size: "$ips"}, 1]}]}, as: "i"},  
            in: {$let: {vars: {ip1: {$arrayElemAt: ["$ips", "$$i"]},  
                ip2: {$arrayElemAt: ["$ips", {$add: ["$$i", 1]}]}},  
                in: {  
                    diff: {$divide: [{abs: {$subtract: ["$$ip1.ts", "$$ip2.ts"]}}, 60000]},  
                    ip1: "$$ip1.ip", t1: "$$ip1.ts",  
                    ip2: "$$ip2.ip", t2: "$$ip2.ts"}  
                }}}}},  
            cond: {$and: [{lt: ["$$this.diff", 10]}, {$ne: ["$$this.ip1", "$$this.ip2"]}]}  
        }]}},  
{$match: {"diffs": {$ne: []}}},  
{$project: {_id: 0, user: "$_id", suspectLogins: "$diffs"}}
```

\$match

\$sort

\$group

\$addFields

\$match

\$project

```
start=ISODate("...")  
end=ISODate("...")  
{  
    user: "303900",  
    ipaddr: "71.56.112.56",  
    ts:ISODate("2017-05-08T...")  
}  
  
{ "user" : "35237073",  
  "suspectLogins" : [  
    {"diff": 4.8333333333,  
     "ip1": "106.220.151.16",  
     "t1":"2017-05-08T06:58",  
     "ip2": "223.182.113.15"  
     "t2":"2017-05-08T07:03"},  
    {"diff": 8.3,  
     "ip1": "223.182.113.15",  
     "t1":"2017-05-08T07:03",  
     "ip2": "49.206.217.26",  
     "t1":"2017-05-08T07:11"}  
]  
}
```

```
    {$match:{ts:{$gte:start,$lt:end}}},  
    {$sort:{ts:1}},  
    {$group:{_id:"$user",ips:{$push:{ip:"$ipaddr", ts:"$ts"}}}},  
    {$addFields:{diffs: {$filter:{  
        input:{$map:{  
            input: {$range:[0,{$subtract:[{$size:"$ips"},1]}]}, as:"i",  
            in:{$let:{vars:{ip1:$arrayElemAt:["$ips","$$i"]},  
                    ip2:$arrayElemAt:["$ips",{$add:["$$i",1]}]}},  
            in:{  
                diff:{$divide:[{$abs:{$subtract:[ $$ip1.ts,"$$ip2.ts"]}},60000]},  
                ip1:"$$ip1.ip", t1:"$$ip1.ts",  
                ip2:"$$ip2.ip", t2:"$$ip2.ts"}  
            }}}},  
        cond:{$and:[{$lt:[$$this.diff,10]},{$ne:[$$this.ip1,"$$this.ip2"]}]}  
    }},  
    {$match:{diffs:$ne:[]}},  
    {$project:{_id:0, user:"$_id", suspectLogins:"$diffs"}}
```



```
start=ISODate("...")          {$match:{ts:{$gte:start,$lt:end}}},  
end=ISODate("...")           {$sort:{ts:1}},  
{                                {$group:{_id:"$user",ips:{$push:{ip:"$ipaddr", ts:"$ts"}}}},  
  user: "303900",  
  ipaddr: "71.56.112.56",  
  ts:ISODate("2017-05-08T...") {$addFields:{diffs: {$filter:{  
    input:{$map:{  
      input: {$range:[0,{$subtract:[{$size:"$ips"},1]}]}, as:"i",  
      in:{$let:{vars:{ip1:$arrayElemAt:["$ips","$$i"]},  
              ip2:$arrayElemAt:["$ips",{$add:["$$i",1]}]}},  
      in:{  
        diff:{$divide:[{$abs:{$subtract:["$$ip1.ts","$$ip2.ts"]}},60000]},  
        ip1:"$$ip1.ip", t1:"$$ip1.ts",  
        ip2:"$$ip2.ip", t2:"$$ip2.ts"  
      }}}},  
    cond:{$and:[{$lt:[$$this.diff,10]},{$ne:[$$this.ip1,"$$this.ip2"]}]}  
  }}},  
  {$match:{diffs:{$ne:[]}}},  
  {$project:{_id:0, user:"$_id", suspectLogins:"$diffs"}},  
}  
}  
]
```

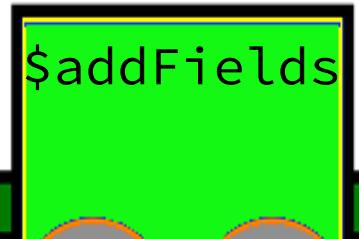
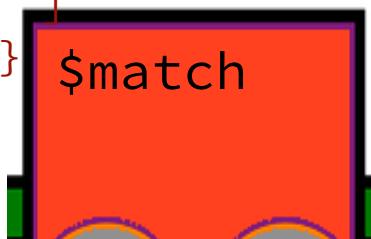


```
start=ISODate("...")  
end=ISODate("...")  
{  
    user: "303900",  
    ipaddr: "71.56.112.56",  
    ts:ISODate("2017-05-08T...")  
}
```

```
{ "user" : "35237073",  
  "suspectLogins" : [  
      {"diff": 4.8333333333,  
       "ip1": "106.220.151.16",  
       "t1": "2017-05-08T06:58",  
       "ip2": "223.182.113.15"  
       "t2": "2017-05-08T07:03"},  
      {"diff": 8.3,  
       "ip1": "223.182.113.15",  
       "t1": "2017-05-08T07:03",  
       "ip2": "49.206.217.26",  
       "t1": "2017-05-08T07:11"}]
```

```
}
```

```
    {$match:{ts:{$gte:start,$lt:end}}},  
    {$sort:{ts:1}},  
    {$group:{_id:"$user",ips:{$push:{ip:"$ipaddr", ts:"$ts"}}}},  
    {$addFields:{diffIpNum:{$size:{$setUnion:"$ips.ip"}}}},  
    {$match:{diffIpNum:{$gt:1}}},  
    {$addFields:{diffs: {$filter:{  
        input:{$map:{  
            input: {$range:[0,{$subtract:[{$size:"$ips"},1]}]}, as:"i",  
            in:{$let:{vars:{ip1:$arrayElemAt:["$ips","$$i"]},  
                    ip2:$arrayElemAt:["$ips",{$add:["$$i",1]}]}},  
            in:{  
                diff:{$divide:[{$abs:{$subtract:[$$ip1.ts,"$$ip2.ts"]}},60000]},  
                ip1:"$$ip1.ip", t1:"$$ip1.ts",  
                ip2:"$$ip2.ip", t2:"$$ip2.ts"}  
        }}}},  
        cond:{$and:[{$lt:[$$this.diff,10]},{$ne:[$$this.ip1,"$$this.ip2"]}]}  
    }}},  
    {$match:{diffs:{$ne:[]}}},  
    {$project:{_id:0, user:"$_id", suspectLogins:"$diffs"}}
```



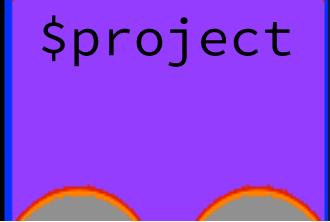
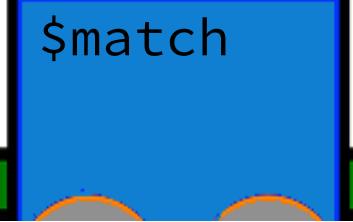
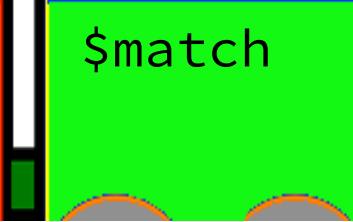
```
start=ISODate("...")  
end=ISODate("...")  
{  
    user: "303900",  
    ipaddr: "71.56.112.56",  
    ts:ISODate("2017-05-08T...")  
}
```

```
{ "user" : "35237073",  
  "suspectLogins" : [  
    {"diff": 4.8333333333,  
     "ip1": "106.220.151.16",  
     "t1": "2017-05-08T06:58",  
     "ip2": "223.182.113.15"  
     "t2": "2017-05-08T07:03"},  
    {"diff": 8.3,  
     "ip1": "223.182.113.15",  
     "t1": "2017-05-08T07:03",  
     "ip2": "49.206.217.26",  
     "t1": "2017-05-08T07:11"}]
```

```
}
```

```
}
```

```
    {$match:{ts:{$gte:start,$lt:end}}},  
    {$sort:{ts:1}},  
    {$group:{_id:"$user",ips:{$push:{ip:"$ipaddr", ts:"$ts"}},  
             diffIps:{$addToSet:"$ipaddr"}}},  
    {$match:{"diffIps.1":{$exists:true}}},  
    {$addFields:{diffs: {$filter:{  
        input:{$map:{  
            input: {$range:[0,{$subtract:[{$size:"$ips"},1]}]}, as:"i",  
            in:{$let:{vars:{ip1:$arrayElemAt:["$ips","$$i"]},  
                     ip2:$arrayElemAt:["$ips",{$add:["$$i",1]}]}},  
            in:{  
                diff:{$divide:[{$abs:{$subtract:[ $$ip1.ts,"$$ip2.ts"]}},60000]},  
                ip1:"$$ip1.ip", t1:"$$ip1.ts",  
                ip2:"$$ip2.ip", t2:"$$ip2.ts"}  
            }}}},  
        cond:{$and:[{$lt:[$$this.diff,10]},{$ne:[$$this.ip1,"$$this.ip2"]}]}  
    }}},  
    {$match:{diffs:{$ne:[]}}},  
    {$project:{_id:0, user:"$_id", suspectLogins:"$diffs"}}
```



```
start=ISODate("...")  
end=ISODate("...")  
{  
    user: "303900",  
    ipaddr: "71.56.112.56",  
    ts:ISODate("2017-05-08T...")  
}
```

```
{ "user" : "35237073",  
  "suspectLogins" : [  
    {"diff": 4.8333333333,  
     "ip1": "106.220.151.16",  
     "t1": "2017-05-08T06:58",  
     "ip2": "223.182.113.15"  
     "t2": "2017-05-08T07:03"},  
    {"diff": 8.3,  
     "ip1": "223.182.113.15",  
     "t1": "2017-05-08T07:03",  
     "ip2": "49.206.217.26",  
     "t1": "2017-05-08T07:11"}]
```

```
}
```

```
    {$match:{ts:{$gte:start,$lt:end}}},  
    {$sort:{ts:1}},  
    {$group:{_id:"$user",ips:{$push:{ip:"$ipaddr", ts:"$ts"}},  
             diffIps:{$addToSet:"$ipaddr"}}},  
    {$match:{"diffIps.1":{$exists:true}}},  
    {$addFields:{diffs: {$filter:{  
        input:{$map:{  
            input: {$range:[0,{$subtract:[{$size:"$ips"},1]}]}, as:"i",  
            in:{$let:{vars:{ip1:$arrayElemAt:["$ips","$$i"]},  
                     ip2:$arrayElemAt:["$ips",{$add:["$$i",1]}]}},  
            in:{  
                diff:{$divide:[{$abs:{$subtract:[$$ip1.ts,"$$ip2.ts"]}},60000]},  
                ip1:"$$ip1.ip", t1:"$$ip1.ts",  
                ip2:"$$ip2.ip", t2:"$$ip2.ts"}  
            }}}},  
        cond:{$and:[{$lt:[$$this.diff,10]},{$ne:[$$this.ip1,"$$this.ip2"]}]}  
    }}},  
    {$match:{diffs:{$ne:[]}}},  
    {$project:{_id:0, user:"$_id", suspectLogins:"$diffs"}}
```

\$match

\$sort

\$group

\$match

\$addFields

\$match

\$project

# POWERFUL AGGREGATIONS

## Understand Stages

- Best order for performance
- Avoid unnecessary "blocking"
  - keep "streaming"
- Maximize use of indexes
  - early stages get the index!
- Liberally check `explain()` output

## Understand Expressions

- Schema manipulation
- Array transformation

## Use functions

- Readable, debug-able, reusable

# THE FUTURE OF AGGREGATION

- Better performance & optimizations
- More stages & expressions
- More options for output
- Compass helper for aggregate

<https://github.com/asya999/mbw17>

\$sum \$minute  
\$pow \$strLenBytes \$substrBytes  
\$isoWeek \$reverseArray \$and \$addToSet  
\$max \$meta \$setDifference \$arrayElemAt  
\$allElementsTrue \$setDifference \$arrayElemAt  
\$add \$sortByCount \$setEquals  
\$literal \$gte \$reduce \$sample \$  
\$setUnion \$replaceRoot \$  
\$millisecond \$concat \$size  
\$second \$filter \$skip  
\$substringsCP \$multiply \$out  
\$log10 \$abs \$first  
\$dayOfWeek \$not \$gt  
\$dayOfYear \$slice \$last  
\$min \$divide \$in  
\$type \$slice \$or  
\$split \$concat \$indexStats  
\$min \$ne \$in  
\$isoWeekYear \$group \$redact  
\$ceil \$sort \$setIntersection  
\$ceil \$sort \$setIndexStats  
\$ceil \$sort \$setIsSubset  
\$ceil \$sort \$isArra  
\$ceil \$sort \$year  
\$ceil \$sort \$isoDayOfW  
\$ceil \$sort \$toLowerCase  
\$ceil \$sort \$anyElementTrue  
\$ceil \$sort \$stdDevSamp  
\$ceil \$sort \$dayOfMonth  
\$ceil \$sort \$stdDevP  
\$ceil \$sort \$push  
\$ceil \$sort \$avg  
\$ceil \$sort \$range  
\$ceil \$sort \$bucket  
\$ceil \$sort \$geoNear  
\$ceil \$sort \$bucket  
\$ceil \$sort \$indexCP  
\$ceil \$sort \$subtract  
\$ceil \$sort \$map  
\$ceil \$sort \$toUpperCase  
\$ceil \$sort \$switch  
\$ceil \$sort \$month  
\$ceil \$sort \$indexOfArray  
\$ceil \$sort \$switch  
\$ceil \$sort \$indexOfBytes

<https://github.com/asya999/mdbw17>

<https://github.com/asya999/mbw17>

# MONGODB WORLD'17

FOR GIANT IDEAS

<https://github.com/asya999/mbw17>

<https://github.com/asya999/mdbw17>



**MONGODB  
WORLD'17**  
FOR GIANT IDEAS



# MONGODB WORLD'17

FOR GIANT IDEAS

# AGGREGATE

is  
the  
new FIND