



# MongoDB

Шевцов Василий Викторович,  
директор ДИТ РУДН, [shevtsov-vv@rudn.ru](mailto:shevtsov-vv@rudn.ru), [vvshevtsov@fa.ru](mailto:vvshevtsov@fa.ru)

# Агрегация

## Функция count()

С помощью функции count() можно получить число элементов в коллекции:

```
> db.Students.count()  
69329
```

Число элементов в коллекции с учетом ограничений

```
db.Students.find({"Факультет":"Экономический"}).count()  
7587
```

```
db.Students.find({"Факультет":"Экономический","Статус":"Завершение"}).count()  
4092
```

```
db.Students.find().limit(2).count()  
69329
```

```
db.Students.find().limit(2).count(true)  
2
```

```
db.Students.find().skip(2).count()  
69329
```

```
db.Students.find().skip(2).count(true)  
69327
```

## Функция distinct

В коллекции могут быть документы, которые содержат одинаковые значения для одного или нескольких полей.

```
db.Students.distinct("Факультет")
```

```
[
```

```
  "Медицинский",
```

```
  "Экономический",
```

```
  "Институт гостиничного бизнеса и туризма",
```

```
  "Аграрный",
```

```
  "Физико-математических и естественных наук",
```

```
  "Гуманитарных и социальных наук",
```

```
  "Инженерный",
```

```
  "Юридический",
```

```
  "Экологический",
```

```
  "Филологический",
```

```
  "Институт иностранных языков",
```

```
  "Кафедра сравнительной образовательной политики",
```

```
  "Институт мировой экономики и бизнеса",
```

```
  "Институт международных программ",
```

```
  "Институт прикладных технико-экономических исследований и экспертиз",
```

```
  "Учебно-научный институт гравитации и космологии",
```

```
  "Научно-образовательный центр Нанотехнологии",
```

```
  "Курчатовский Институт"
```

```
]
```

## Функция distinct

```
db.Students.find({"Форма обучения":"Заочная"}).distinct("Форма обучения")
2019-11-21T11:53:14.854+0300 E QUERY [js] TypeError:
db.Students.find(...).distinct is not a function :
@(shell):1:1
```

```
> db.Students.distinct("Форма обучения",{"Форма обучения":"Заочная"})
[ "Заочная" ]
> db.Students.distinct("Факультет",{"Форма обучения":"Заочная"})
[
  "Медицинский",
  "Юридический",
  "Аграрный",
  "Экономический",
  "Институт международных программ",
  "Гуманитарных и социальных наук",
  "Инженерный",
  "Экологический",
  "Филологический",
  "Кафедра сравнительной образовательной политики",
  "Институт иностранных языков"
]
```

## Функция distinct

`db.collection.distinct(field, query, options)`

Parameter	Type	Description
field	string	The field for which to return distinct values.
query	document	A query that specifies the documents from which to retrieve the distinct values.
options	document	Optional. A document that specifies the options.

# Функция distinct

## Options

```
{ collation: <document> }
```

```
collation: {  
  locale: <string>,  
  caseLevel: <boolean>,  
  caseFirst: <string>,  
  strength: <int>,  
  numericOrdering: <boolean>,  
  alternate: <string>,  
  maxVariable: <string>,  
  backwards: <boolean>  
}
```

## Функция distinct

```
{ "_id": 1, "dept": "A", "item": { "sku": "111", "color": "red" }, "sizes": [ "S", "M" ] }  
{ "_id": 2, "dept": "A", "item": { "sku": "111", "color": "blue" }, "sizes": [ "M", "L" ] }  
{ "_id": 3, "dept": "B", "item": { "sku": "222", "color": "blue" }, "sizes": "S" }  
{ "_id": 4, "dept": "A", "item": { "sku": "333", "color": "black" }, "sizes": [ "S" ] }
```

```
db.inventory.distinct( "dept" )  
[ "A", "B" ]
```

```
db.inventory.distinct( "item.sku" )  
[ "111", "222", "333" ]
```

```
db.inventory.distinct( "sizes" )  
[ "M", "S", "L" ]
```

```
db.inventory.distinct( "item.sku", { dept: "A" } )  
[ "111", "333" ]
```



## Метод aggregate()

В MongoDB для агрегации используется метод aggregate(). Данный метод имеет следующий синтаксис:

```
db.ИМЯ_КОЛЛЕКЦИИ.aggregate(ОПЕРАЦИЯ_АГРЕГАЦИИ)
```

пример:

```
db.developers.aggregate  
( [ { $group : { _id : "Developers", total_salary: { $sum : "$salary" } } } ] )
```

результат:

```
{ "_id" : "Developers", "total_salary" : 10500 }
```

<https://docs.mongodb.com/manual/aggregation/>

<https://docs.mongodb.com/manual/reference/method/db.collection.aggregate/>

# Концепция фреймворка

Aggregation Framework работает по тому же принципу, что и оператор “GROUP BY” в SQL. Его две основные концепции — конвейеры и выражения. Конвейеры — это операторы, которые обрабатывают поток документов. А выражения возвращают обработанные документы после проведения вычислений над исходными документами. Вот некоторые конвейеры:

**\$match** — использует уточняющий предикат, так же как и `collection.find({})`;

**\$project** — позволяет менять форму результата, включая рассчитанные значения, суб-объекты и проч.;

**\$unwind** — разделяет элементы массива и добавляет его (их?) в выходной документ;

**\$sort** — сортирует документы;

**\$limit** — указывает максимальное число документов, которые будут возвращены;

**\$skip** — пропускает указанное число документов.

**\$group** — группирует данные по определенным ключам (полям)

# Конвейеры

```
db.Students.find({"Факультет":"Экономический"})
```

```
db.Students.aggregate({$match:{"Факультет":"Экономический"}})
```

```
db.Students.aggregate({$match:{"Факультет":"Экономический"}},{$sort:{"Курс":1}})
```

```
db.Students.aggregate({$match:{"Факультет":"Экономический"}},{$sort:{"Курс":-1}})
```

```
db.Students.aggregate(  
    {$match:{"Факультет":"Экономический"}},  
    {$sort:{"Курс":-1}},  
    {$limit:1}  
)
```

# Конвейеры

```
db.Students.aggregate(  
    {$match:{"Факультет":"Экономический"}},  
    {$project:{"Факультет":1,"Курс":1}},  
    {$sort:{"Курс":-1}},  
    {$limit:10}  
)
```

```
{ "_id": ObjectId("5bebd3f09135a16927cd7e8"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd752"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd88a"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd5d7"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd5c1"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd80b"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd8de"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd89a"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd750"), "Факультет": "Экономический", "Курс": 5 }  
{ "_id": ObjectId("5bebd3f09135a16927cd6ef"), "Факультет": "Экон  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }  
    { "Факультет": "Экономический", "Курс": 5 }
```

`{ $project: { "_id": 0, "Факультет": 1, "Курс": 1 } }`

# Конвейеры

```
[
{"_id":1,"Languages":["русский"]},
{"_id":2,"Languages":["русский","английский"]},
{"_id":3,"Languages":["русский","английский","немецкий"]},
{"_id":4,"Languages":["русский","английский","немецкий","французский"]},
{"_id":5,"Languages":["русский","английский","немецкий","французский","испанский"]}
]
```

```
db.Languages2.aggregate({$unwind:"$Languages"})
```

```
{ "_id" : 2, "Languages" : "русский" }
{ "_id" : 2, "Languages" : "английский" }
{ "_id" : 1, "Languages" : "русский" }
{ "_id" : 5, "Languages" : "русский" }
{ "_id" : 5, "Languages" : "английский" }
{ "_id" : 5, "Languages" : "немецкий" }
{ "_id" : 5, "Languages" : "французский" }
{ "_id" : 5, "Languages" : "испанский" }
{ "_id" : 3, "Languages" : "русский" }
{ "_id" : 3, "Languages" : "английский" }
{ "_id" : 3, "Languages" : "немецкий" }
{ "_id" : 4, "Languages" : "русский" }
{ "_id" : 4, "Languages" : "английский" }
{ "_id" : 4, "Languages" : "немецкий" }
{ "_id" : 4, "Languages" : "французский" }
```

```
db.Languages2.aggregate(
  {$unwind:"$Languages"},
  {$sort:{"_id":1}}
)
```

```
db.Languages2.aggregate(
  {$unwind:"$Languages"},
  {$match:{"Languages":"немецкий"}})
{ "_id" : 5, "Languages" : "немецкий" }
{ "_id" : 3, "Languages" : "немецкий" }
{ "_id" : 4, "Languages" : "немецкий" }
```

# \$unwind

```
db.test.insert({name: "Ivan", likes: ["Maria", "Anna"]})
```

```
db.test.insert({name: "Serge", likes: ["Anna"]})
```

```
db.test.aggregate({$unwind: "$likes"})
```

```
{ "result" : [  
  { "_id" : ObjectId("4f598de76a8f8bc74573e9fd"), "name" : "Ivan", "likes" : "Maria" },  
  { "_id" : ObjectId("4f598de76a8f8bc74573e9fd"), "name" : "Ivan", "likes" : "Anna" },  
  { "_id" : ObjectId("4f598e086a8f8bc74573e9fe"), "name" : "Serge", "likes" : "Anna" }  
], "ok" : 1 }
```

```
db.test.aggregate({$unwind: "$likes"}, {$project: {name:1, likes:1, count: {$add: [1]}}});
```

```
{ "_id" : ObjectId("4f598de76a8f8bc74573e9fd"), "name" : "Ivan", "likes" : "Maria", "count" : 1 },
```

## \$unwind

```
db.test.aggregate(  
  {$unwind: "$likes"},  
  {$project: {name:1, likes:1, count: {$add: [1]}}},  
  {$group: {_id: "$likes", number: {$sum: "$count"}}});
```

```
{ "result" : [  
  { "_id" : "Anna", "number" : 2 },  
  { "_id" : "Maria", "number" : 1 } ],  
  "ok" : 1 }
```

```
db.test.aggregate(  
  {$unwind: "$likes"},  
  {$project: {name:1, likes:1, count: {$add: [1]}}},  
  {$group: {_id: "$likes", number: {$sum: "$count"}}},  
  {$sort: {number: -1}},  
  {$limit: 1});
```

```
{ "result" : [ { "_id" : "Anna", "number" : 2 } ], "ok" : 1 }
```

# Метод aggregate()

Collection

```
db.orders.aggregate( [  
  $match stage → { $match: { status: "A" } },  
  $group stage → { $group: { _id: "$cust_id", total: { $sum: "$amount" } } }  
] )
```

{ cust_id: "A123", amount: 500, status: "A" }
{ cust_id: "A123", amount: 250, status: "A" }
{ cust_id: "B212", amount: 200, status: "A" }
{ cust_id: "A123", amount: 300, status: "D" }

orders

\$match →

{ cust_id: "A123", amount: 500, status: "A" }
{ cust_id: "A123", amount: 250, status: "A" }
{ cust_id: "B212", amount: 200, status: "A" }

\$group →

Results	
{	<code>_id: "A123",</code>
	total: 750
}	
{	<code>_id: "B212",</code>
	total: 200
}	



# Список операций для агрегации

Выражение	Описание	Пример
\$sum	Суммирует указанные значения всех документов в коллекции.	db.ИМЯ_КОЛЛЕКЦИИ.aggregate([{\$group : {_id : "\$title", total_salary : {\$sum : "\$salary"}}}])
\$avg	Рассчитывает среднее значение указанного поля для всех документов коллекции.	db.ИМЯ_КОЛЛЕКЦИИ.aggregate([{\$group : {_id : "\$title", avg_salary : {\$avg : "\$salary"}}}])
\$min	Получает минимальное значение указанного поля документа в коллекции	db.ИМЯ_КОЛЛЕКЦИИ.aggregate([{\$group : {_id : "\$title", min_salary : {\$min : "\$salary"}}}])
\$max	Получает максимальное значение указанного поля документа в коллекции	db.ИМЯ_КОЛЛЕКЦИИ.aggregate([{\$group : {_id : "\$title", max_salary : {\$max : "\$salary"}}}])

Выражение	Описание	Пример
\$push	Вставляет значение в массив в результирующем документе.	db.ИМЯ_КОЛЛЕКЦИИ.aggregate([{\$group : {_id : "\$title", skills : {\$push: "\$skills"}}}])
\$addToSet	Вставляет значение в массив в результирующем документе, но не создаёт дубликаты.	db.ИМЯ_КОЛЛЕКЦИИ.aggregate([{\$group : {_id : "\$title", skills : {\$addToSet : "\$skills"}}}])
\$first	Получает первый документ из сгруппированных. Обычно используется вместе с сортировкой.	db.ИМЯ_КОЛЛЕКЦИИ.aggregate([{\$group : {_id : "\$title", first_skill : {\$first : "\$skills"}}}])
\$last	Получает крайний документ из сгруппированных. Обычно используется вместе с сортировкой.	db.ИМЯ_КОЛЛЕКЦИИ.aggregate([{\$group : {_id : "\$title", last_skill : {\$last : "\$skills"}}}])

## Связанные функции

\$project	Используется для выбора некоторых специальных полей из коллекции.
\$match	Операция фильтрации, которая может уменьшить количество документов, которые передаются для ввода в следующий этап.
\$group	Агрегация
\$sort	Сортирует документы
\$skip	С помощью данной команды можно проигнорировать список документов в имеющемся множестве.
\$limit	Ограничивает количество документов для вывода на количество, переданное методу, начиная, с текущей позиции.
\$unwind	Используется для “разматывания” документов, который используют массивы.

## Пример aggregate

Подсчет количества по факультетам

```
db.Students.aggregate([{"$group":{"_id":"$Факультет",count:{$sum:1}}}]])
```

```
> db.Students.aggregate([{"$group":{"_id":"$Факультет",count:{$sum:1}}}]])
{ "_id" : "Курчатовский Институт", "count" : 12 }
{ "_id" : "Учебно-научный институт гравитации и космологии", "count" : 53 }
{ "_id" : "Институт иностранных языков", "count" : 2469 }
{ "_id" : "Институт прикладных технико-экономических исследований и экспертиз", "count" : 289 }
{ "_id" : "Кафедра сравнительной образовательной политики", "count" : 135 }
{ "_id" : "Филологический", "count" : 9450 }
{ "_id" : "Институт международных программ", "count" : 286 }
{ "_id" : "Экологический", "count" : 1900 }
{ "_id" : "Инженерный", "count" : 8727 }
{ "_id" : "Институт мировой экономики и бизнеса", "count" : 2253 }
{ "_id" : "Гуманитарных и социальных наук", "count" : 8581 }
{ "_id" : "Институт гостиничного бизнеса и туризма", "count" : 1812 }
{ "_id" : "Экономический", "count" : 7587 }
{ "_id" : "Физико-математических и естественных наук", "count" : 4263 }
{ "_id" : "Медицинский", "count" : 11681 }
{ "_id" : "Юридический", "count" : 6062 }
{ "_id" : "Научно-образовательный центр Нанотехнологии", "count" : 20 }
{ "_id" : "Аграрный", "count" : 3749 }
```

## Команда group

```
db.collection.group({ key, reduce, initial [, keyf] [, cond] [, finalize] })
```

<https://docs.mongodb.com/manual/reference/method/db.collection.group/>

## Команда group. Аргументы

- `key` - документ, описывающий, по каким полям группировать. Например, чтобы группировать по полю `category_id`, нужно в качестве ключа задать `{category id: true}`. Ключ может быть составным. Так, чтобы сгруппировать множество отзывов по `user_id` и `rating`, нужно задать ключ `{user_id: true, rating: true}`. Параметр `key` обязателен, если только не используется `keyf`.
- `keyf` - JavaScript-функция, которая, будучи применена к документу, генерирует ключ для этого документа. Это полезно, когда ключ группировки вычисляется динамически. Например, если требуется сгруппировать результирующий набор по дню недели, в который создавался документ, но это значение не хранится, то для генерации ключа можно воспользоваться функцией:

```
function ( doc) {return {day: doc.created_at.getDay()};
```

Эта функция генерирует ключи вида `{ day: 1}`. Отметим, что параметр `keyf` обязателен, если только не используется `key` для задания стандартного ключа.

## Команда group. Аргументы

- initial - документ, используемый в качестве основы для порождения результатов агрегирования. При первом обращении к функции reduce этот начальный документ используется в качестве первого значения агрегатора и обычно содержит все агрегируемые ключи. Например, если для каждой группы вычисляется сумма поданных голосов и общее число документов в группе, то начальный документ будет выглядеть так:

(vote\_surn: 0.0, doc count: 0).

Этот параметр обязателен.

## Команда group. Аргументы

- `reduce` - JavaScript-функция, выполняющая агрегирование. Она принимает два аргумента: текущий документ и документ-агрегатор, в котором хранятся результаты агрегирования. Начальным значением агрегатора будет документ `initial`. Вот пример функции `reduce` для агрегирования голосов и количества документов:

```
function ( doc, aggregator ) {  
    aggregator.doc_count += 1;  
    aggregator.vote_sum += doc.vote_count;  
}
```

Отметим, что функция `reduce` не обязана что-то возвращать; она просто должна модифицировать объект-агрегатор. Параметр `reduce` обязателен.



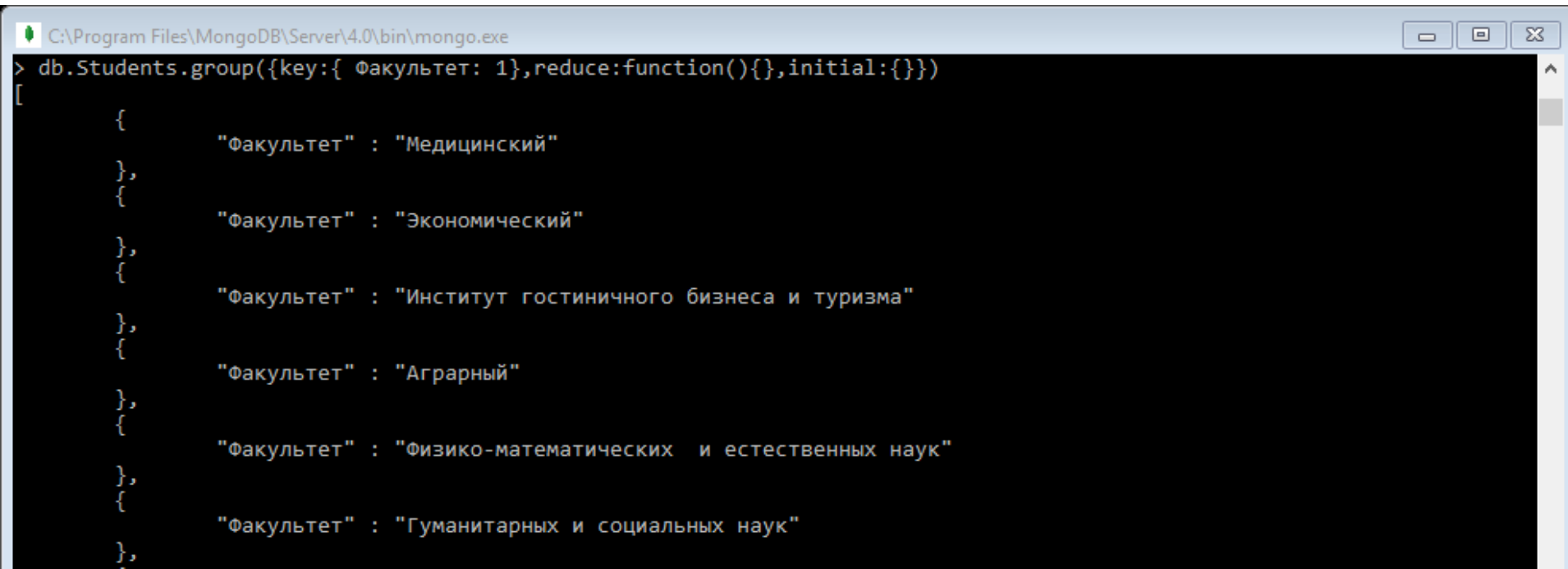
## Команда group. Аргументы

- `cond` - селектор запроса, который отбирает подлежащие агрегированию документы. Если вы не хотите агрегировать всю коллекцию, то должны передать селектор. Например, чтобы агрегировать только отзывы, за которые подано более пять голосов, нужно задать такой селектор запроса: `{ vote count:{$gt: 5}}`.
- `finalize` - JavaScript-функция, которая применяется ко всем результирующим документам перед возвратом результирующего набора. Эта функция позволяет выполнить постобработку результатов операции группировки. Типичный пример - усреднение. Можно использовать уже существующие в результате группировки значения для порождения нового содержащего среднее:

```
function(doc) {  
    doc.average = doc.vote_count / doc.doc_count;  
}
```

## Пример group

```
db.Students.group(  
  {  
    key: { Факультет: 1},  
    reduce: function () { },  
    initial: { }  
  }  
)
```



The screenshot shows a MongoDB shell window with the following content:

```
C:\Program Files\MongoDB\Server\4.0\bin\mongo.exe  
> db.Students.group({key:{ Факультет: 1},reduce:function(){},initial:{}})  
[  
  {  
    "Факультет" : "Медицинский"  
  },  
  {  
    "Факультет" : "Экономический"  
  },  
  {  
    "Факультет" : "Институт гостиничного бизнеса и туризма"  
  },  
  {  
    "Факультет" : "Аграрный"  
  },  
  {  
    "Факультет" : "Физико-математических и естественных наук"  
  },  
  {  
    "Факультет" : "Гуманитарных и социальных наук"  
  },  
]
```

## Пример group

```
db.Students.group(  
  {  
    key: { Факультет: 1 },  
    reduce: function( curr, result ) {  
      result.count++;  
    },  
    initial: { count: 0 }  
  }  
)
```

```
> db.Students.group({key: { Факультет: 1 },reduce: function( curr, result ) {result.count++;},initial: { count: 0 }})  
[  
  {  
    "Факультет" : "Медицинский",  
    "count" : 11681  
  },  
  {  
    "Факультет" : "Экономический",  
    "count" : 7587  
  },  
  {  
    "Факультет" : "Институт гостиничного бизнеса и туризма",  
    "count" : 1812  
  },  
  {  
    "Факультет" : "Аграрный",  
    "count" : 3749  
  },  
]
```

## Пример group с ограничением

```
db.orders.group(  
  {  
    key: { ord_dt: 1, 'item.sku': 1 },  
    cond: { ord_dt: { $gt: new Date( '01/01/2012' ) } },  
    reduce: function( curr, result ) {  
      result.total += curr.item.qty;  
    },  
    initial: { total : 0 }  
  }  
)
```

```
SELECT ord_dt, item_sku,  
SUM(item_qty) as total  
FROM orders  
WHERE ord_dt > '01/01/2012'  
GROUP BY ord_dt, item_sku
```

# Спасибо за внимание!



Шевцов Василий Викторович

[shevtsov-vv@rudn.ru](mailto:shevtsov-vv@rudn.ru), [vvshevtsov@fa.ru](mailto:vvshevtsov@fa.ru)

+7(903)144-53-57