



Data Analysis and Map-Reduce with MongoDB and pymongo

Alexander C. S. Hendorf, EuroPython 2015, Bilbao
[@opotoc](#)

Alexander C. S. Hendorf

- Mannheim, Germany
- IT is my 'second career'
- developer @my own company
opotoc IT GmbH
- mongoDB MUG organiser
- speaker, sometimes trainer
- EP2015 program WG co-chair



Today

1. mongoDB / document oriented database
2. What's the mongoDB aggregation framework?
3. Pipeline model
4. Pipeline stages
5. Map Reduce in mongoDB

some live demos

mongoDB aggregation framework

- introduced with mongoDB 2.2 in 2012
- framework for data aggregation
- documents enter a multi-stage pipeline that transforms the documents into an aggregated results
- it's designed 'straight-forward'
- all operations have an optimization phase which attempts to reshape the pipeline for improved performance

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	
{ _id: "A123", total: 750 }	
{ _id: "B212", total: 200 }	

Pipeline is like a relay race

\$match



get the baton

\$group



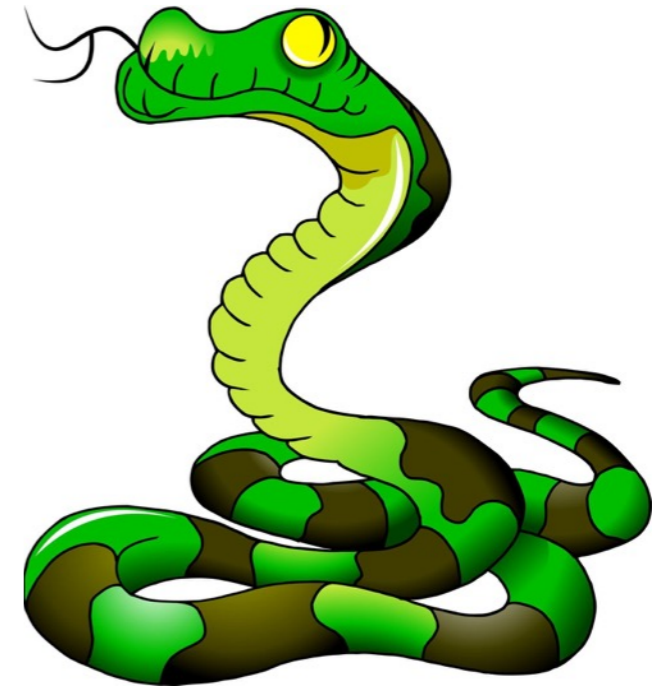
something smart

\$project



present nicely

- mongoDB 3.0
- WiredTiger storage engine
- driver: pymongo
- dataset 37GB, compressed with WT ~9GB
- collection of playlists from the iTunes Music Store
- playlists that appeared in some chart sometime in the past 3 years somewhere around the world



{'_id': ObjectId('5215d7f3ee6da1070d5cb88a'), 'adamId': 573885160, 'added': {'epoch_time': 'Thu 22.08.2013 09:20:51 UTC'}, 'headers': {'dict': {'apple-timing-app': '222 ms', 'cache-control': 'no-transform, max-age=60', 'connection': 'close', 'content-encoding': 'gzip', 'content-length': '17404', 'content-type': 'text/html; charset=UTF-8', 'date': 'Thu, 22 Aug 2013 09:20:51 GMT', 'last-modified': 'Thu, 22 Aug 2013 09:20:51 GMT', 'vary': 'Accept-Encoding', 'x-apple-aka-ttl': 'Generated Thu Aug 22 02:20:51 PDT 2013, Expires Thu Aug 22 02:21:51 PDT 2013, TTL 60s', 'x-apple-application-instance': '1009514', 'x-apple-application-site': 'NWK', 'x-apple-jingle-correlation-key': 'VASQDI34SUY5G', 'x-apple-lok-response-date': 'Thu Aug 22 02:20:51 PDT 2013', 'x-apple-orig-url': 'https://itunes.apple.com/co/album/id573885160', 'x-apple-partner': 'origin.0', 'x-apple-translated-wo-url': '/WebObjects/MZStore.woa/wa/viewAlbum?id=573885160&cc=co', 'x-webobjects-loadaverage': '0'}, 'encodingheader': None, 'fp': None, 'headers': {'no-transform, max-age=60', 'Connection': 'close', 'Content-Encoding': 'gzip', 'Content-Length': '17404', 'Content-Type': 'text/html; charset=UTF-8', 'Date': 'Thu, 22 Aug 2013 09:20:51 GMT', 'Last-Modified': 'Thu, 22 Aug 2013 09:20:51 GMT', 'Vary': 'Accept-Encoding', 'X-Apple-Partner': 'origin.0', 'apple-timing-app': '222 ms', 'x-apple-aka-ttl': 'Generated Thu Aug 22 02:20:51 PDT 2013, Expires Thu Aug 22 02:21:51 PDT 2013, TTL 60s', 'x-apple-application-instance': '1009514', 'x-apple-application-site': 'NWK', 'x-apple-jingle-correlation-key': 'VASQDI34SUY5G', 'x-apple-lok-response-date': 'Thu Aug 22 02:20:51 PDT 2013', 'x-apple-orig-url': 'https://itunes.apple.com/co/album/id573885160', 'x-apple-translated-wo-url': '/WebObjects/MZStore.woa/wa/viewAlbum?id=573885160&cc=co', 'x-webobjects-loadaverage': '0'}, 'maintype': 'text', 'plist': {'charset=UTF-8', 'plisttext': '; charset=UTF-8', 'seekable': 0, 'startofbody': None, 'startofheaders': None, 'status': '', 'subtype': 'html', 'type': 'text/html', 'typeheader': 'text/html; charset=UTF-8', 'unixfrom': ''}, 'info': {'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'artwork': [['200', 'http://a1.mzstatic.com/us/r30/Music/v4/9a/ce/66/9ace66e1-f14f-4981-ac6f-8acfd591960/cover200x200.jpeg'], [100, 'http://a5.mzstatic.com/us/r30/Music/v4/9a/ce/66/9ace66e1-f14f-4981-ac6f-8acfd591960/cover100x100.jpeg'], [250, 'http://a2.mzstatic.com/us/r30/Music/v4/9a/ce/66/9ace66e1-f14f-4981-ac6f-8acfd591960/cover250x250.jpeg'], [130, 'http://a4.mzstatic.com/us/r30/Music/v4/9a/ce/66/9ace66e1-f14f-4981-ac6f-8acfd591960/cover130x130.jpeg'], [400, 'http://a3.mzstatic.com/us/r30/Music/v4/9a/ce/66/9ace66e1-f14f-4981-ac6f-8acfd591960/cover400x400.jpeg'], [1400, 'http://a2.mzstatic.com/us/r30/Music/v4/9a/ce/66/9ace66e1-f14f-4981-ac6f-8acfd591960/cover1400x1400.jpeg'], [1200, 'http://a4.mzstatic.com/us/r30/Music/v4/9a/ce/66/9ace66e1-f14f-4981-ac6f-8acfd591960/cover1200x1200.jpeg']], 'children': [{'adamId': 573885322, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'bookletType': 'pdf', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'description': None, 'discNumber': None, 'genres': [20, 21, 1144], 'id': 573885322, 'kind': 'booklet', 'name': 'Digital Booklet - Night Visions', 'nameRaw': 'Digital Booklet - Night Visions', 'offers': [{'assets': [{'flavor': 'booklet', 'size': 2705648}], 'price': None, 'priceFormatted': '', 'type': 'buy', 'variant': 'PLUS'}], 'pieceId': None, 'popularity': 0, 'releaseDate': '2003-04-28', 'releaseDateEpoch': datetime.datetime(2003, 4, 28, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885322', 'trackNumber': None, 'url': 'https://itunes.apple.com/co/album/digital-booklet-night-visions/id573885160?i=573885322&l=en'}, {'adamId': 573885272, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon, Ben McKee, Alex Da Kid & Josh Mosser', 'url': 'https://itunes.apple.com/co/composer/id499982942?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885272, 'kind': 'song', 'name': 'Radioactive', 'nameRaw': 'Radioactive', 'offers': [{'assets': [{'duration': 186, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a840.phobos.apple.com/us/r2000/019/Music2/v4/4f/0d/30/4f0d30e9-ffa3-695c-44c8-d915f9e3fe98/mzaf_575316285755111697.aac.m4a'}, 'size': 6830469}], 'buyParams': {'productType=S&price=1290&salableAdamId=573885272&pricingParameters=PLUS', 'price': 1.29, 'priceFormatted': 'USD\xa01.29', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 1, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885272', 'trackNumber': 1, 'url': 'https://itunes.apple.com/co/album/radioactive/id573885160?i=573885272&l=en'}, {'adamId': 573885274, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon & Ben McKee', 'url': 'https://itunes.apple.com/co/composer/id499982939?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885274, 'kind': 'song', 'name': 'Tiptoe', 'nameRaw': 'Tiptoe', 'offers': [{'assets': [{'duration': 194, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a1623.phobos.apple.com/us/r2000/020/Music2/v4/5d/6c/3a/5d6c3a3c-7ea0-7f71-d100-cf90dc9e8433/mzaf_5720461395889014325.aac.m4a'}, 'size': 7244474}], 'buyParams': {'productType=S&price=990&salableAdamId=573885274&pricingParameters=PLUS', 'price': 0.99, 'priceFormatted': 'USD\xa00.99', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 1, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885274', 'trackNumber': 2, 'url': 'https://itunes.apple.com/co/album/tiptoe/id573885160?i=573885274&l=en'}, {'adamId': 573885275, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon & Ben McKee', 'url': 'https://itunes.apple.com/co/composer/id499982939?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885275, 'kind': 'song', 'name': 'It's Time', 'nameRaw': 'It's Time', 'offers': [{'assets': [{'duration': 240, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a1557.phobos.apple.com/us/r2000/006/Music2/v4/8b/c6/d9/8bc6d932-6ef4-166d-20fb-7cd5c5ba4c79a/mzaf_6099651544288202212.aac.m4a'}, 'size': 8452717}], 'buyParams': {'productType=S&price=1290&salableAdamId=573885275&pricingParameters=PLUS', 'price': 1.29, 'priceFormatted': 'USD\xa01.29', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.41357421875, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885275', 'trackNumber': 3, 'url': 'https://itunes.apple.com/co/album/its-time/id573885160?i=573885275&l=en'}, {'adamId': 573885278, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon, Ben McKee, Alex Da Kid & Josh Mosser', 'url': 'https://itunes.apple.com/co/composer/id499982942?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885278, 'kind': 'song', 'name': 'Demons', 'nameRaw': 'Demons', 'offers': [{'assets': [{'duration': 177, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a174.phobos.apple.com/us/r2000/016/Music/v4/e8/cb/a1/e8cb109-26ad-f7ea-f648-4a4bffd595f1/mzaf_6503879570199009699.aac.m4a'}, 'size': 6346043}], 'buyParams': {'productType=S&price=1290&salableAdamId=573885278&pricingParameters=PLUS', 'price': 1.29, 'priceFormatted': 'USD\xa01.29', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.1343994140625, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885278', 'trackNumber': 4, 'url': 'https://itunes.apple.com/co/album/demons/id573885160?i=573885278&l=en'}, {'adamId': 573885280, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon, Ben McKee & Alex Da Kid', 'url': 'https://itunes.apple.com/co/composer/id202856766?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885280, 'kind': 'song', 'name': 'On Top of the World', 'nameRaw': 'On Top of the World', 'offers': [{'assets': [{'duration': 192, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a1825.phobos.apple.com/us/r2000/015/Music2/v4/e1/36/20/e13620e1-31a2-5f9a-7766-769c97399b81/mzaf_7878115814185165018.aac.m4a'}, 'size': 6940151}], 'buyParams': {'productType=S&price=1290&salableAdamId=573885280&pricingParameters=PLUS', 'price': 1.29, 'priceFormatted': 'USD\xa01.29', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.1343994140625, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885280', 'trackNumber': 5, 'url': 'https://itunes.apple.com/co/album/on-top-of-the-world/id573885160?i=573885280&l=en'}, {'adamId': 573885281, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon & Ben McKee', 'url': 'https://itunes.apple.com/co/composer/id499982939?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885281, 'kind': 'song', 'name': 'Amsterdam', 'nameRaw': 'Amsterdam', 'offers': [{'assets': [{'duration': 241, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a80.phobos.apple.com/us/r2000/007/Music/v4/28/35/bc/2835bc7f-c8e2-8a8b-7cd3-aae132bd43f2/mzaf_8850126300550805333.aac.m4a'}, 'size': 8516981}], 'buyParams': {'productType=S&price=990&salableAdamId=573885281&pricingParameters=PLUS', 'price': 0.99, 'priceFormatted': 'USD\xa00.99', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.003662109375, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885281', 'trackNumber': 6, 'url': 'https://itunes.apple.com/co/album/amsterdam/id573885160?i=573885281&l=en'}, {'adamId': 573885283, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon & Ben McKee', 'url': 'https://itunes.apple.com/co/composer/id499982939?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885283, 'kind': 'song', 'name': 'Hear Me', 'nameRaw': 'Hear Me', 'offers': [{'assets': [{'duration': 235, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a157.phobos.apple.com/us/r2000/019/Music/v4/cf/6e/5d/cf6e5d87-fb86-55e2-2a9e-b62e94a2c4ea/mzaf_3208300556053684171.aac.m4a'}, 'size': 9043466}], 'buyParams': {'productType=S&price=990&salableAdamId=573885283&pricingParameters=PLUS', 'price': 0.99, 'priceFormatted': 'USD\xa00.99', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.00439453125, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885283', 'trackNumber': 7, 'url': 'https://itunes.apple.com/co/album/hear-me/id573885160?i=573885283&l=en'}, {'adamId': 573885284, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon & Ben McKee', 'url': 'https://itunes.apple.com/co/composer/id499982939?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885284, 'kind': 'song', 'name': 'Every Night', 'nameRaw': 'Every Night', 'offers': [{'assets': [{'duration': 217, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a962.phobos.apple.com/us/r2000/003/Music2/v4/ac/72/44/ac7244de-flee-4116-f494-acf5243a6e8f/mzaf_8514226465678986156.aac.m4a'}, 'size': 7730368}], 'buyParams': {'productType=S&price=990&salableAdamId=573885284&pricingParameters=PLUS', 'price': 0.99, 'priceFormatted': 'USD\xa00.99', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.000732421875, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/album/every-night/id573885160?i=573885284&l=en'}, {'adamId': 573885288, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon, Ben McKee, Alex Da Kid & Josh Mosser', 'url': 'https://itunes.apple.com/co/composer/id499982942?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885288, 'kind': 'song', 'name': 'Bleeding Out', 'nameRaw': 'Bleeding Out', 'offers': [{'assets': [{'duration': 90, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a1694.phobos.apple.com/us/r2000/007/Music/v4/6c/al/1d/6c11d2b-deb3-2afb-964b-7377f92ab57f/mzaf_6333841192228218638.aac.m4a'}, 'size': 7895431}], 'buyParams': {'productType=S&price=990&salableAdamId=573885288&pricingParameters=PLUS', 'price': 0.99, 'priceFormatted': 'USD\xa00.99', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.0108642578125, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885288', 'trackNumber': 9, 'url': 'https://itunes.apple.com/co/album/bleeding-out/id573885160?i=573885288&l=en'}, {'adamId': 573885309, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon & Ben McKee', 'url': 'https://itunes.apple.com/co/composer/id499982939?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885309, 'kind': 'song', 'name': 'Underdog', 'nameRaw': 'Underdog', 'offers': [{'assets': [{'duration': 209, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a1948.phobos.apple.com/us/r2000/008/Music2/v4/03/27/1f/03271f66-9dc2-4a43-894b-ec8dbb9cab84/mzaf_2556941019434400323.aac.m4a'}, 'size': 7569963}], 'buyParams': {'productType=S&price=990&salableAdamId=573885309&pricingParameters=PLUS', 'price': 0.99, 'priceFormatted': 'USD\xa00.99', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.00146484375, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885309', 'trackNumber': 10, 'url': 'https://itunes.apple.com/co/album/underdog/id573885160?i=573885309&l=en'}, {'adamId': 573885311, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon, Ben McKee & Clint Holgate', 'url': 'https://itunes.apple.com/co/composer/id573885315?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885311, 'kind': 'song', 'name': 'Nothing Left to Say / Rocks', 'nameRaw': 'Nothing Left to Say / Rocks', 'offers': [{'assets': [{'duration': 539, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a1480.phobos.apple.com/us/r2000/010/Music2/v4/f9/4e/65/f94e651a-1713-9288-6e73-0f9aba83cf76/mzaf_4695219378617698238.aac.m4a'}, 'size': 18730805}], 'buyParams': {'productType=S&price=990&salableAdamId=573885311&pricingParameters=PLUS', 'price': 0.99, 'priceFormatted': 'USD\xa00.99', 'type': 'buy', 'variant': 'PLUS'}, 'pieceId': None, 'popularity': 0.0025634765625, 'releaseDate': '2013-02-01', 'releaseDateEpoch': datetime.datetime(2013, 2, 1, 0, 0), 'shortUrl': 'https://itun.es/co/ORmni?i=573885311', 'trackNumber': 11, 'url': 'https://itunes.apple.com/co/album/nothing-left-to-say-rocks/id573885160?i=573885311&l=en'}, {'adamId': 573885312, 'artistId': 358714030, 'artistIdsIndex': 358714030, 'artistName': 'Imagine Dragons', 'artistUrl': 'https://itunes.apple.com/co/artist/imagine-dragons/id358714030?l=en', 'collectionId': 573885160, 'collectionName': 'Night Visions', 'composer': {'name': 'Dan Reynolds, Wayne Sermon & Ben McKee', 'url': 'https://itunes.apple.com/co/composer/id499982939?l=en'}, 'contentRating': {'system': 'RIAA'}, 'discNumber': 1, 'genres': [20, 34, 21, 1144], 'id': 573885312, 'kind': 'song', 'name': 'Working Man', 'nameRaw': 'Working Man', 'offers': [{'assets': [{'duration': 235, 'flavor': 'plusAudio', 'preview': {'duration': 90, 'url': 'http://a2.phobos.apple.com/us/r2000/000/Music2/v4/3b/c6/8e/3bc68e6c-0d26-6385-7155-37467fbafc22/

```
{'_id': 'ObjectId(5215d7f3ee6da1070d5cb88a)',
  'adamId': 573885160,

//release: album / single / playlist

  'info': {'artistId': 358714030,
           'artistIdsIndex': 358714030,
           'artistName': 'Imagine Dragons',
           'name': 'Night Visions',
           'offers': [{'price': 9.99, 'priceFormatted': 'USD\xa09.99'}],
           'releaseDate': '2013-02-01',
           'releaseDateEpoch': "ISODate('2013-02-01T00:00:00Z')",
           'userRating': {'ratingCount': 8, 'value': 5}}

// songs

  'children': [{'artistId': 358714030,
                'kind': 'song',
                'name': 'Amsterdam',
                'offers': [{'assets': [{'duration': 194}],
                          'price': 0.99,
                          'priceFormatted': 'USD\xa00.99'}],
                'releaseDate': '2013-02-01'}],

}
```



TAYLOR SWIFT

BORN IN 1989.



JOIN MAILING LIST

TAYLORSWIFT.COM

SHOP TAYLOR SWIFT

🕒 JUNE 21, 2015

❤️ 93,782 NOTES

TUMBLR RADAR!
ARCHIVE

To Apple, Love Taylor

I write this to explain why I'll be holding back my album, 1989, from the new streaming service, Apple Music. I feel this deserves an explanation because Apple has been and will continue to be one of my best partners in selling music and creating ways for me to connect with my fans. I respect the company and the truly ingenious minds that have created a legacy based on innovation and pushing the right boundaries.

I'm sure you are aware that Apple Music will be offering a free 3 month trial to anyone who signs up for the service. I'm not sure you know that Apple Music will not be paying writers, producers, or artists for those three months. I find it to be shocking, disappointing, and completely unlike this historically progressive and generous company.

pipeline = [

find in aggregation is \$match, sql: WHERE

{"\$match": {"info.artistName": **artist}},**

\$project, sql: SELECT

{"\$project": {"release": "\$info.name", "_id": 0}},

{"\$sort": {"release": ASCENDING}}

]

Aggregation stages



\$match

WHERE | HAVING

\$sort

ORDER BY

\$limit

LIMIT

\$project

SELECT

\$group

GROUP BY

\$unwind

(JOIN)

\$redact

\$out

```
{'_id': 'ObjectId(5215d7f3ee6da1070d5cb88a)',
  'adamId': 573885160,

  //release: album / single / playlist

  'info': {'artistId': 358714030,
    'artistIdsIndex': 358714030,
    'artistName': 'Imagine Dragons',
    'name': 'Night Visions',
    'offers': [{'price': 9.99, 'priceFormatted': 'USD\xa09.99'}],
    'releaseDate': '2013-02-01',
    'releaseDateEpoch': "ISODate('2013-02-01T00:00:00Z')",
    'userRating': {'ratingCount': 8, 'value': 5}}

  // songs

  'children': [{'artistId': 358714030,
    'kind': 'song',
    'name': 'Amsterdam',
    'offers': [{'assets': [{'duration': 194}],
      'price': 0.99,
      'priceFormatted': 'USD\xa00.99'}],
    'releaseDate': '2013-02-01'}],

}
```

pipeline = [

find in aggregation is \$match, sql: WHERE

{"\$match": {"info.artistName": **artist}},**

GROUP BY & COUNT()

{"\$group": {

"_id**": "\$info.name",**

"count**": {"\$sum": 1}}},**

\$project, sql: SELECT

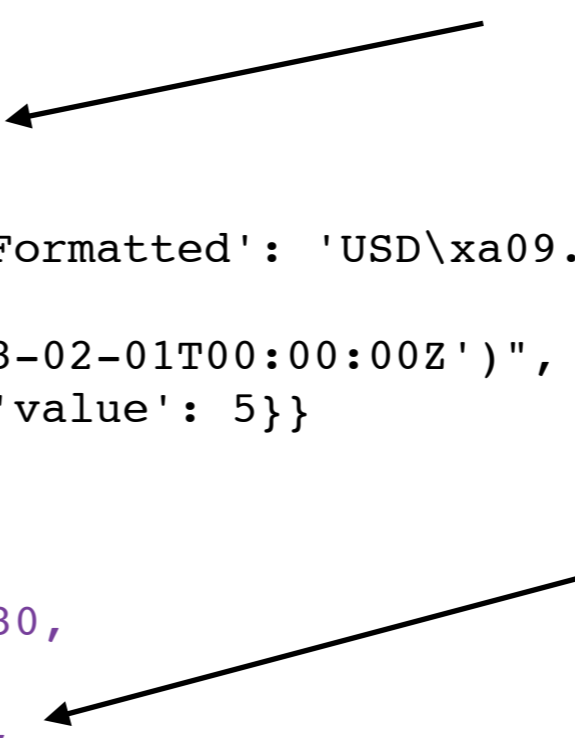
{"\$project": {"release**": "\$_id", "**_id**": 0}},**

{"\$sort": {"release": ASCENDING}}

]

working with lists of sub-documents

```
{ '_id': 'ObjectId(5215d7f3ee6da1070d5cb88a)',  
  'adamId': 573885160,  
  
  //release: album / single / playlist  
  
  'info': { 'artistId': 358714030,  
            'artistIdsIndex': 358714030,  
            'artistName': 'Imagine Dragons',  
            'name': 'Night Visions',  
            'offers': [{ 'price': 9.99, 'priceFormatted': 'USD\xa09.99' }],  
            'releaseDate': '2013-02-01',  
            'releaseDateEpoch': "ISODate('2013-02-01T00:00:00Z')",  
            'userRating': { 'ratingCount': 8, 'value': 5 } }  
  
  // songs  
  
  'children': [ { 'artistId': 358714030,  
                 'kind': 'song',  
                 'name': 'Amsterdam',  
                 'offers': [ { 'assets': [ { 'duration': 194 } ],  
                              'price': 0.99,  
                              'priceFormatted': 'USD\xa00.99' } ],  
                 'releaseDate': '2013-02-01',  
                 .....  
                 ],  
  
}
```




```
pipeline = [  
  {"$match": {"info.artistName": artist}},  
  # "explode" list  
  {"$unwind": "$info.children"},  
  {"$group": {  
    "_id": "$info.children.name"}},  
  {"$project": {"song": "$_id", "_id": 0}},  
  {"$sort": {"release": ASCENDING}}  
]
```

more stages...

\$skip: skip documents in found set

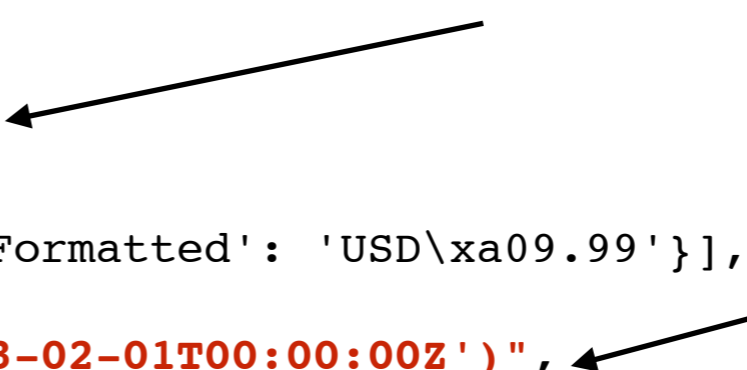
\$out: write the resulting documents of the aggregation pipeline to a collection, also incremental.

\$geoNear: returns an ordered stream of documents based on the proximity to a geospatial point

\$redact: reshapes each document in the stream by restricting the content for each document based on information stored in the documents themselves

\$min/\$max \$first/\$last

```
{'_id': 'ObjectId(5215d7f3ee6da1070d5cb88a)',  
  'adamId': 573885160,  
  
  //release: album / single / playlist  
  
  'info': {'artistId': 358714030,  
           'artistIdsIndex': 358714030,  
           'artistName': 'Imagine Dragons',  
           'name': 'Night Visions',  
           'offers': [{'price': 9.99, 'priceFormatted': 'USD\xa09.99'}],  
           'releaseDate': '2013-02-01',  
           'releaseDateEpoch': "ISODate('2013-02-01T00:00:00Z')",  
           'userRating': {'ratingCount': 8, 'value': 5}}  
  
  // songs  
  
  'children': [{'artistId': 358714030,  
               'kind': 'song',  
               'name': 'Amsterdam',  
               'offers': [{'assets': [{'duration': 194}],  
                           'price': 0.99,  
                           'priceFormatted': 'USD\xa00.99'}],  
               'releaseDate': '2013-02-01'},  
               .....  
               ],  
  
}
```



```
pipeline = [  
  {"$match": {"info.artistName": artist}},  
  {"$group": {  
    "_id": "",  
    "minDate": {"$min": "$info.releaseDateEpoch"},  
    "maxDate": {"$max": "$info.releaseDateEpoch"}},  
  {"$project": {"_id": 0, "minDate": 1, "maxDate": 1}},  
]
```

date operators

```
pipeline = [  
  {"$match": {"info.artistName": artist}},  
  {"$sort": SON([("info.releaseDate", ASCENDING)])},  
  {"$group": {  
    "_id": {"$year": "$info.releaseDateEpoch"},  
    "count": {"$sum": "1"}}},  
  {"$project": {"year": "$_id.year", "_id": 0, "count": 1}},  
]
```

date operators / multikey groups

```
pipeline = [  
  {"$match": {"info.artistName": artist}},  
  {"$sort": SON([("info.releaseDate", ASCENDING)])},  
  {"$group": {  
    "_id": { "year": {"$year": "$info.releaseDateEpoch",  
      "month": {"$month": "$info.releaseDateEpoch"}}},  
    "count": {"$sum": "1"},  
    {"$project": {"year": "$_id.year", "month": "$_id.month", "_id": 0,  
      "count": 1}}},  
  ]
```

The Nemesis. Google say



\$in

pipeline = [

 {"\$match": {"info.artistName": {"\$in": [artist,
nemesis]}}},

 ,

]

sub-sub-documents / \$avg

pipeline = [

```
  {"$match": {"info.artistName": {"$in": [artist, nemesis]}}},
```

```
  {"$unwind": "$info.children"},
```

```
  {"$unwind": "$info.children.offers"},
```

```
  {"$unwind": "$info.children.offers.assets"}
```

```
  {"$group": {"_id": "$info.children.name",
```

```
    "playtime": {"$avg": "$info.children.offers.assets.duration"},
```

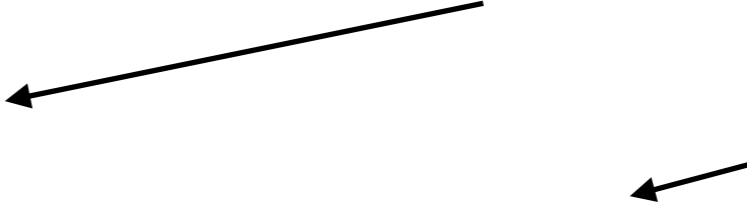
```
  }},
```

```
  {"$project":.....
```

```
]
```

only USD

```
{ '_id': 'ObjectId(5215d7f3ee6da1070d5cb88a)',  
  'adamId': 573885160,  
  
  //release: album / single / playlist  
  
  'info': { 'artistId': 358714030,  
            'artistIdsIndex': 358714030,  
            'artistName': 'Imagine Dragons',  
            'name': 'Night Visions',  
            'offers': [{ 'price': 9.99, 'priceFormatted': 'USD\xa09.99' }],  
            'releaseDate': '2013-02-01',  
            'releaseDateEpoch': "ISODate('2013-02-01T00:00:00Z')",  
            'userRating': { 'ratingCount': 8, 'value': 5 } }  
  
  // songs  
  
  'children': [ { 'artistId': 358714030,  
                  'kind': 'song',  
                  'name': 'Amsterdam',  
                  'offers': [ { 'assets': [ { 'duration': 194 } ],  
                               'price': 0.99,  
                               'priceFormatted': 'USD\xa00.99' } ],  
                  'releaseDate': '2013-02-01'},  
                .....  
                ],  
  
}
```



string operations / \$cmp

```
pipeline = [ {"$match": {"info.artistName": {"$in": [artist, nemesis]}}, {"$unwind": "$info.offers"},  
  
{"$project": {"info.offers.price": 1, "info.offers.priceFormatted": 1,  
  
"artist": "$info.artistName", "product": "$info.name",  
  
"isUSD": {"$cmp": [{"$toLower": {  
                "$substr": ["$info.offers.priceFormatted", 0, 3]}, "usd"]}}},  
  
{"$match": {"isUSD": 0}},  
  
{"$sort": {"info.offers.price": DESCENDING}},  
  
{"$group": {  
    "_id": {"artist": "$artist"},  
  
    "releases": {"$push": {"price": "$info.offers.price", "product": "$product"}}  
  
}}, {"$project": :.....]
```

a lot more operators...

- Stage Operators
 - Boolean Operators
 - Set Operators
 - Comparison Operators
 - Arithmetic Operators
 - String Operators
 - Array Operators
 - Text Search Operators
 - Variable Operators
 - Literal Operators
 - Date Operators
 - Conditional Expressions
 - Accumulators
- [see: MongoDB docs](#)

\$map

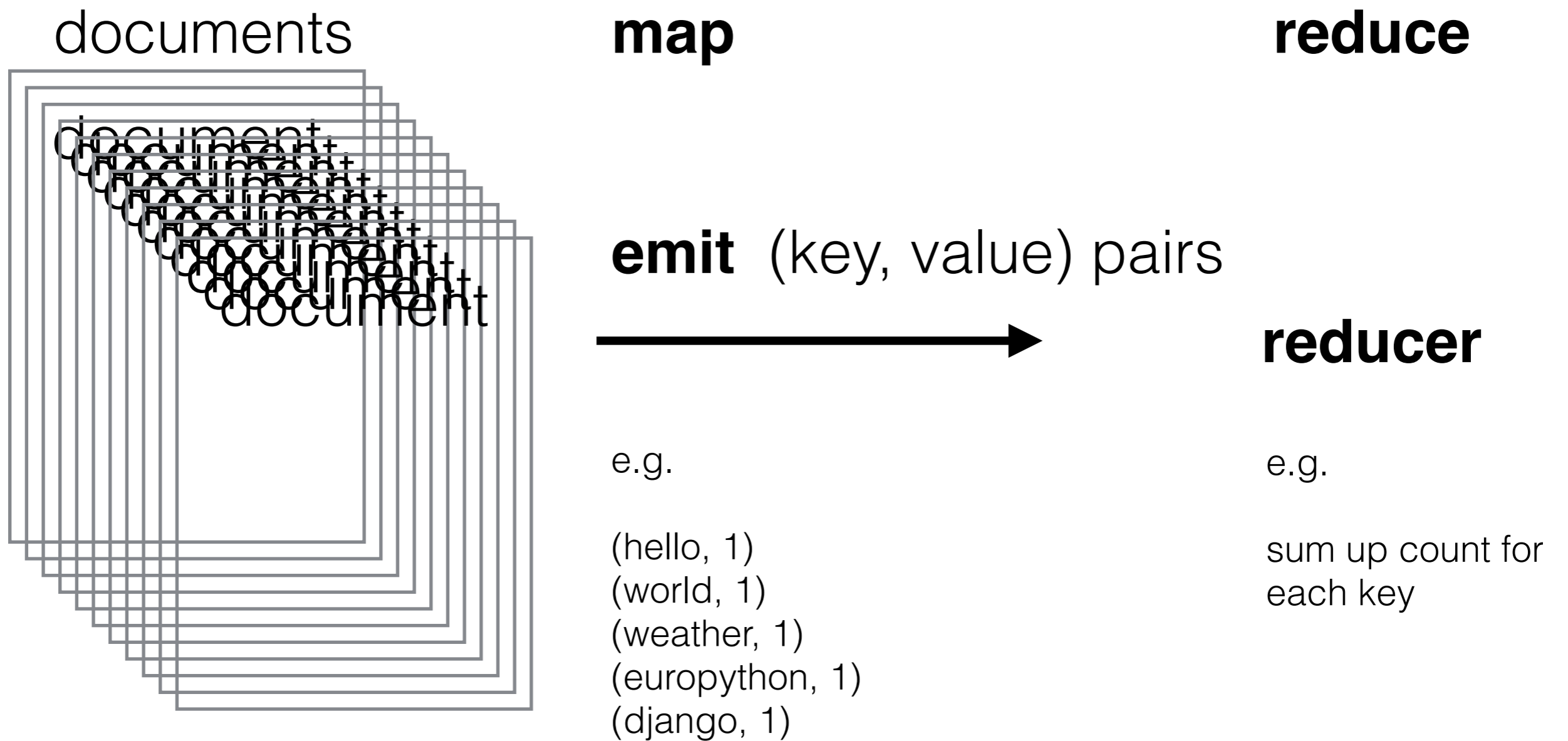
```
pipeline = [ {"$match": {"info.artistName": {"$in": [artist, nemesis]}},
{"$group": {
  "_id": "$info.artistName",
  "ratingCount": {"$push": "$info.userRating.ratingCount"}},
{"$project": {
  "adjustedRatingCount": {
    "$map": {"input": "$ratingCount",
  "as": "value",
  "in": {"$add": ["$$value", 10 ]}}}},
{"$unwind": "$adjustedRatingCount"},
{"$group": {
  "_id": "$_id", "totalRatingCount": {"$sum": "$adjustedRatingCount"}}}]
```

Map Reduce

- provides a map, reduce & finalize phase.
- for most operations the aggregation pipeline has better performance and is easier to handle
- output: inline or to a collection
- however, more flexibility via usage of JavaScript functions



Map Reduce in 15 sec



most popular words in release titles

```
{ '_id': 'ObjectId(5215d7f3ee6da1070d5cb88a)',  
  'adamId': 573885160,  
  
  //release: album / single / playlist  
  
  'info': { 'artistId': 358714030,  
            'artistIdsIndex': 358714030,  
            'artistName': 'Imagine Dragons',  
            'name': 'Night Visions',  
            'offers': [{ 'price': 9.99, 'priceFormatted': 'USD\xa09.99' }],  
            'releaseDate': '2013-02-01',  
            'releaseDateEpoch': "ISODate('2013-02-01T00:00:00Z')",  
            'userRating': { 'ratingCount': 8, 'value': 5 } }  
  
  // songs  
  
  'children': [ { 'artistId': 358714030,  
                 'kind': 'song',  
                 'name': 'Amsterdam',  
                 'offers': [ { 'assets': [ { 'duration': 194 } ],  
                              'price': 0.99,  
                              'priceFormatted': 'USD\xa00.99' } ],  
                 'releaseDate': '2013-02-01'},  
                 .....  
                 ],  
  
}
```



Some Best Practices & Tips

Database

- optimize your indexes, compound indexes
- mind the result limit of 16MB for final & intermediate results
- pipeline operator limit 100MB of RAM

Queries

- use `db.collection.aggregate().explain()` to get a better understanding of queries

Hardware

- mind RAM, more = better
- mind disk performance, faster = better, prefer SSD

Infrastructure

- work with dedicated server for aggregation
- can be e.g. a (hidden/delayed) member of replica set or standalone copy

Useful Sources

- **mongoDB docs**

<http://docs.mongodb.org/manual/core/aggregation-introduction/>

- **pymongo docs**

<http://api.mongodb.org/python/current/examples/aggregation.html>

- **Aysa Kamsky's blog**

<http://www.kamsky.org/stupid-tricks-with-mongodb>



Q&A

Alexander C. S. Hendorf
@opotoc