



Save The World with MongoDB Data Lake

The Goal



Save the World



Save Money

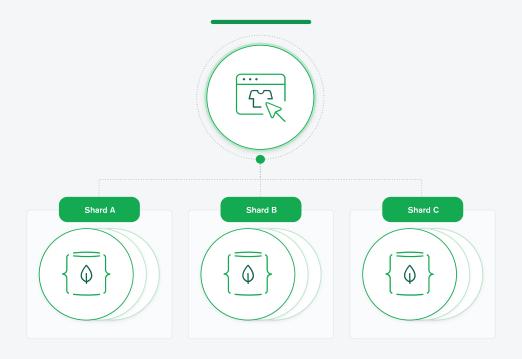




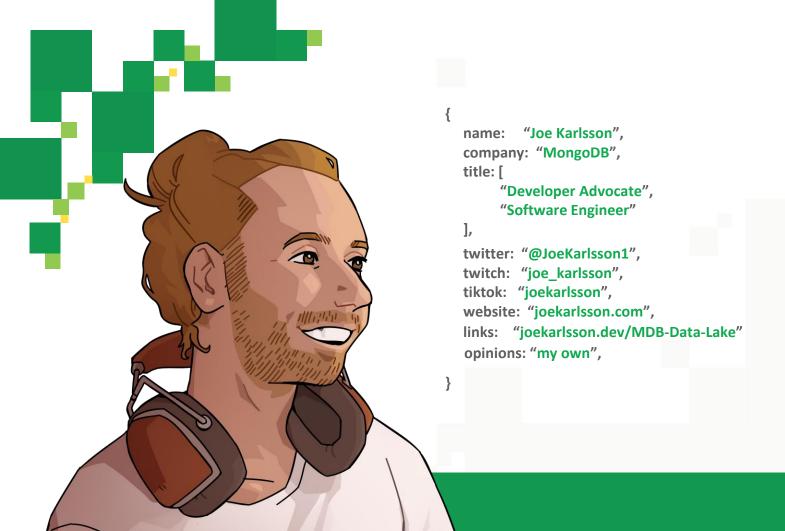












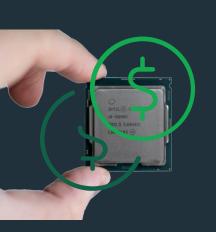


ioekarlsson.dev/MDB-Data-Lak



Hot Data

Cold Data





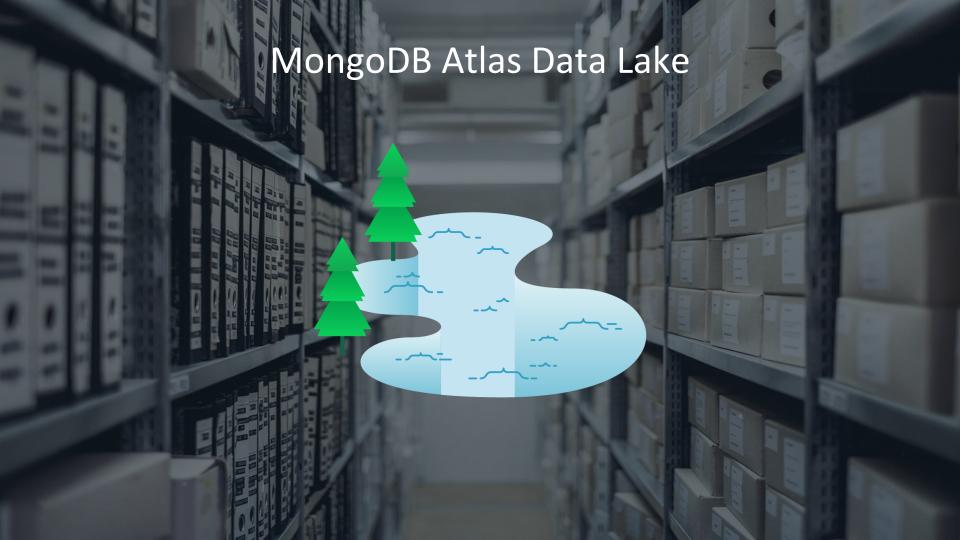














MongoDB Atlas Data Lake

Demo: Archive with a Realm function

Demo: Archive with \$out

Online Archive



Agenda



Atlas Data Lake - Features and Benefits

Work with rich data easily & intuitively



Leverage a serverless & scalable query service



Integrated with the MongoDB Cloud Platform



Eliminate cost & complexity of data movement

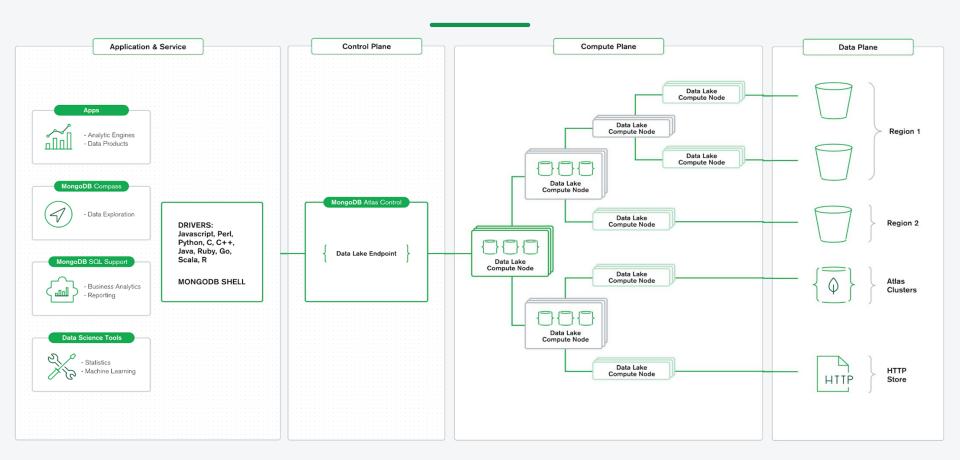


Easy to use with your favorite tools



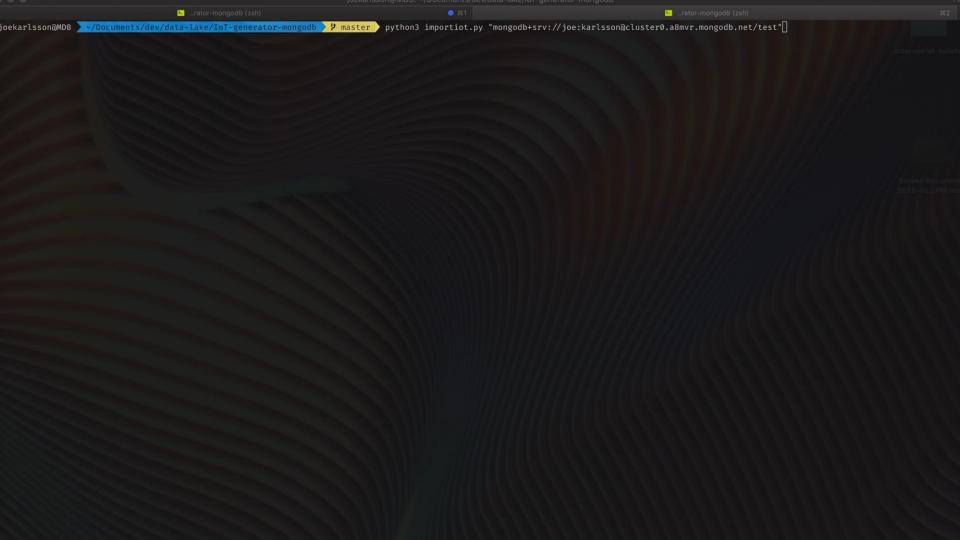


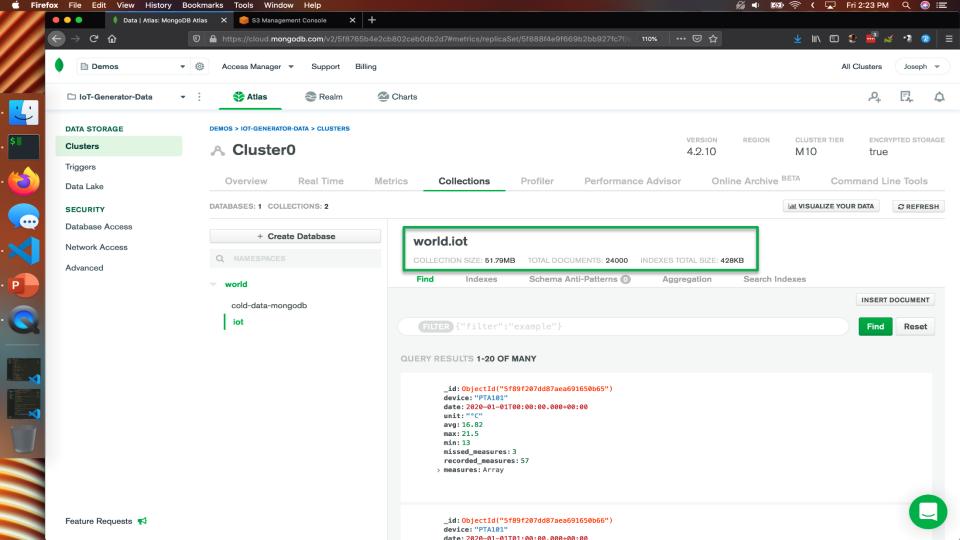
MongoDB Atlas Data Lake - Architecture



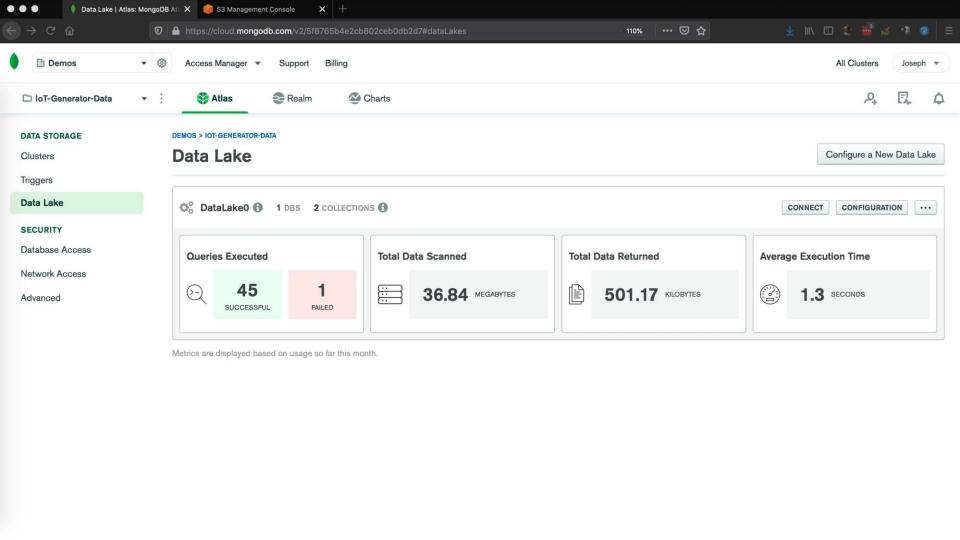


First, let's insert some data





Configuring Our Data Lake

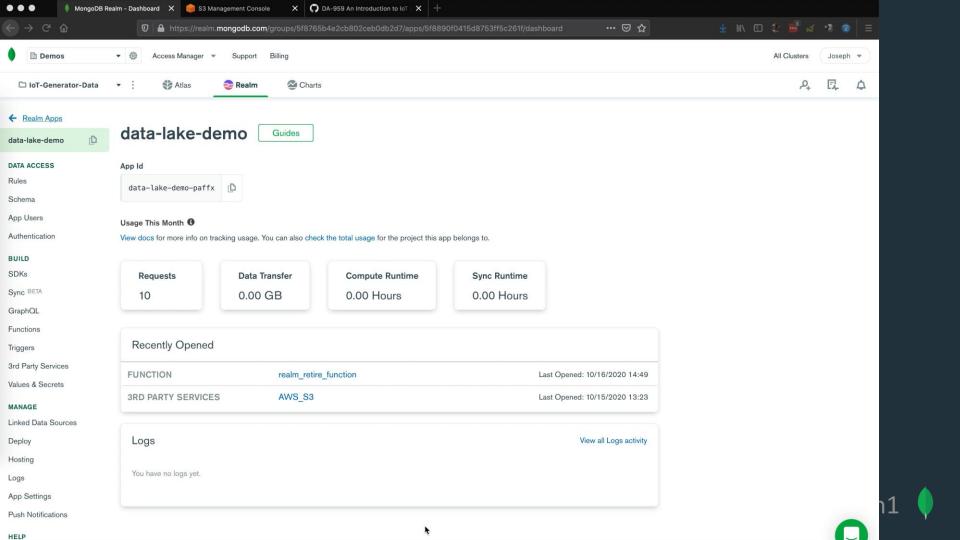


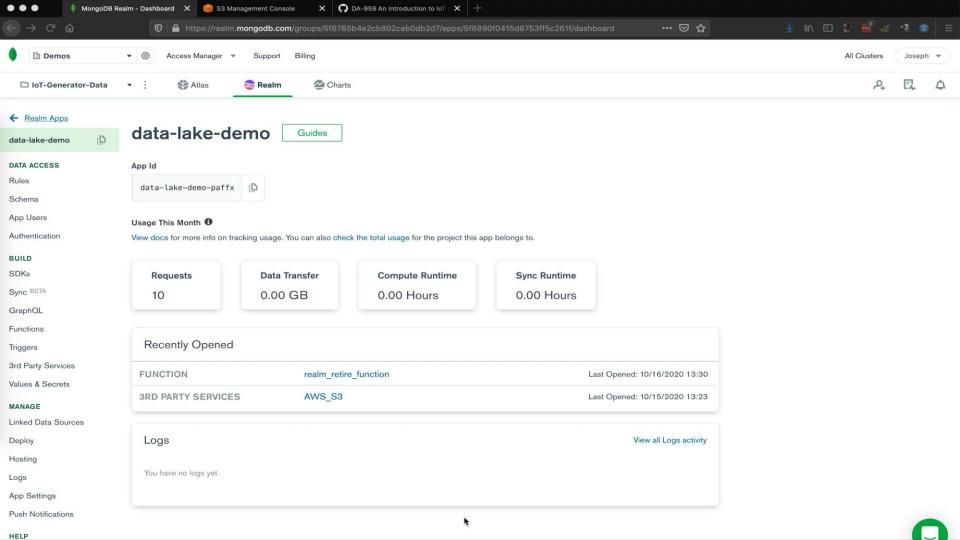
Archiving with a Realm Function



```
exports = async function(){
                 const s3 = context.services.get('AWS').s3("eu-west-1");
                  const iot = context.services.get("mongodb-atlas").db("world").collection("iot");
                  const first date = (await iot.find().sort({"date": 1}).limit(1).next()).date;
                  const next day = new Date(first date);
                  next day.setDate(next day.getDate() + 1);
                  console.log("Archiving from: " + first date);
                  console.log("Archiving to : " + next day);
                  const query = {"date": {"$gte": first date, "$lt": next day}};
                  const docs = await iot.find(query).sort({"date": 1}).toArray();
                  console.log("Sending" + docs.length + " docs to S3.");
                  const s3res = await s3.PutObject({
                                    "Bucket": "cold-data-mongodb",
                                     "Key": first_date.toISOString() + "-" + next_day.toISOString() + ".1.json",
                                    "Body": EJSON.stringify(docs)
                  console.log("S3 ETag: " + s3res.ETag);
                  return iot.deleteMany(query);
```



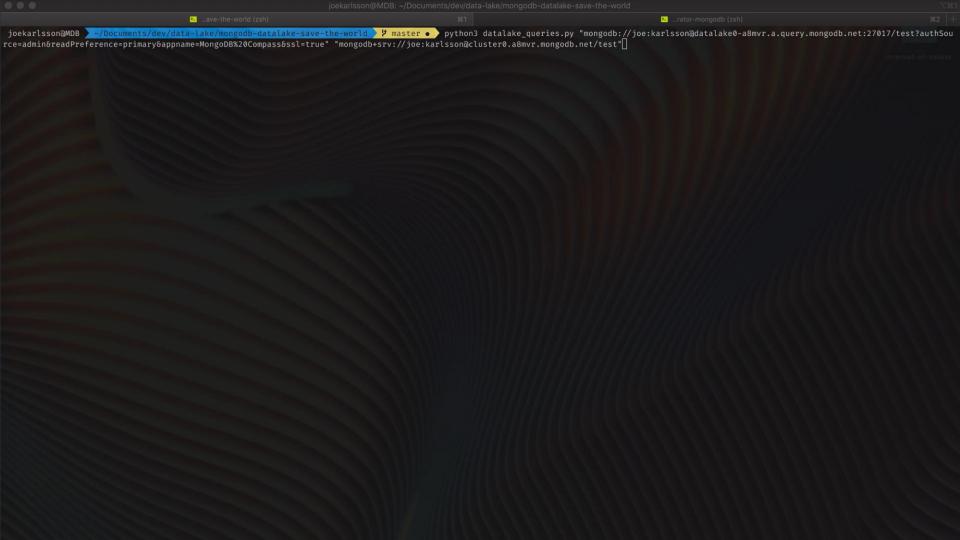


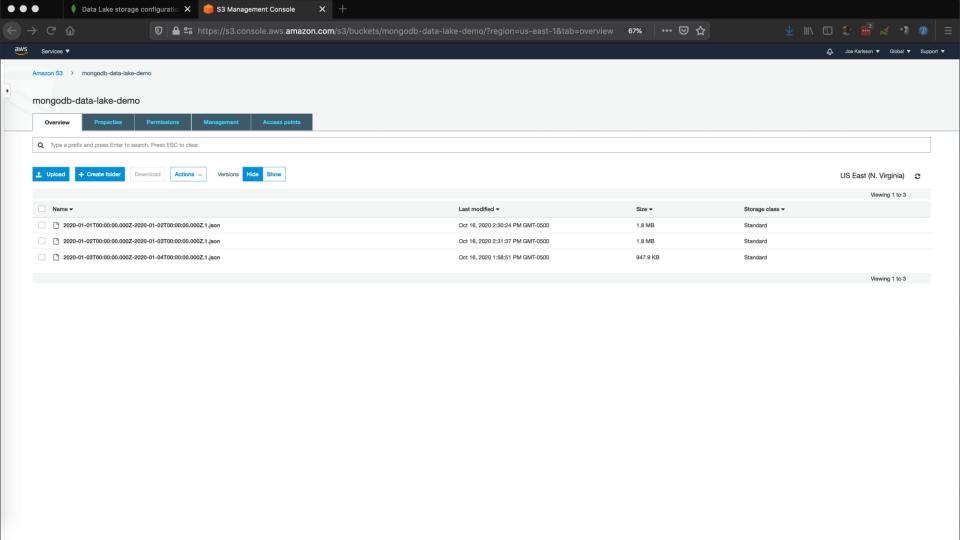


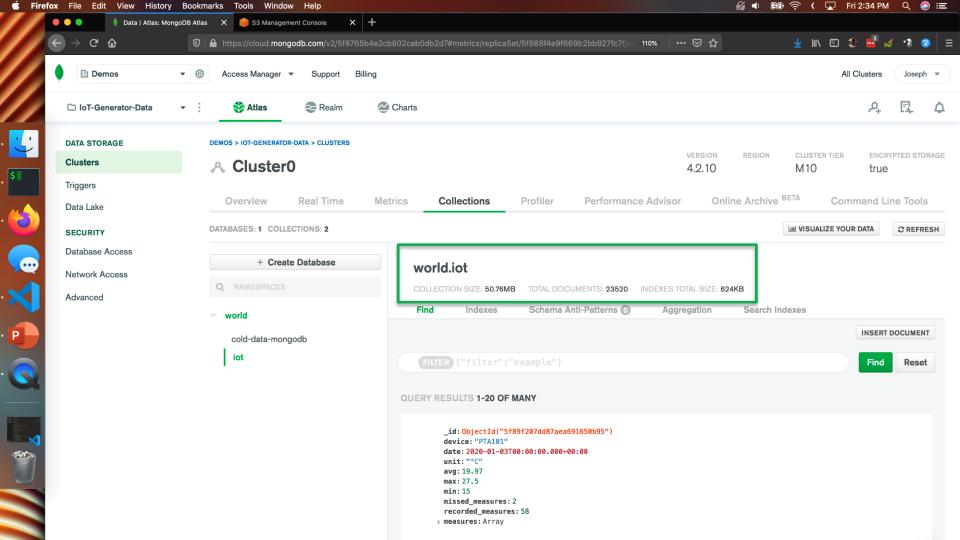
Archiving with pymongo & \$out

```
pipeline s3 = [
 {'$match': {'date': {'$gte': date start, '$lt': date stop}}},
      's3': {
         'filename': date_start.isoformat('T', 'milliseconds') + 'Z-' + date_stop.isoformat('T', 'milliseconds') + 'Z',
         'format': {'name': 'json', 'maxFileSize': '200MiB'}
iot_data_lake.aggregate(pipeline_s3)
```









Online Archive



OATest								4.2.6	N. Virginia (us-east-1)	M10 (General)	encrypted storage true
Overview	Real Time	Metrics	Collections	Profiler	Performance Advisor	Online Archive	Command Line Tools				
Automatically Archive Your Data											
Create rules to automatically move infrequently accessed immutable											
					data to queryable	e, read-only storage a	and free up space in your clusters.				
						Configure Onl	ne Archive				
Learn more ☑											

To Sum Up

We can archive in S3

We have access to ALL the data

Saving the World

Saving lot of money



Questions



What's next?







MongoDB Community

- joekarlsson.dev/MongoDBUniversity
- joekarlsson.dev/MongoDBDeveloper
- joekarlsson.dev/MongoDBCommunity
- Recommended Course:
 M320: Data Modeling







Want \$100 in FREE MongoDB Atlas credits?

Use code JoeK100





- [Docs] MongoDB Atlas Data Lake Documentation: https://docs.mongodb.com/datalake/
- [Docs] Archiving a MongoDB Cluster: https://docs.atlas.mongodb.com/online-archive/manage-online-archive/
- [DevHub Post] MongoDB Data Lake Setup Tutorial: https://developer.mongodb.com/how-to/atlas-data-lake-setup
- [GitHub] Save The World And Money With MongoDB Data Lake: https://github.com/JoeKarlsson/mongodb-datalake-save-the-world
- [GitHub] MongoDB IoT Sample Data Generator: https://github.com/joekarlsson/IoT-generator-mongodb

Additional Resources





name: "Joe Karlsson", company: "MongoDB", title: ["Developer Advocate", "Software Engineer" twitter: "@JoeKarlsson1", twitch: "joe_karlsson", tiktok: "joekarlsson", website: "joekarlsson.com",

links: "joekarlsson.dev/MDB-Data-Lake"





name: "Joe Karlsson", company: "MongoDB", title: ["Developer Advocate", "Software Engineer" twitter: "@JoeKarlsson1", twitch: "joe_karlsson", tiktok: "joekarlsson", website: "joekarlsson.com",

links: "joekarlsson.dev/MDB-Data-Lake"

