

Amazon DynamoDB: Scaling and Benchmarking

Celeste Castillo and Ben Kennel



DynamoDB



Idea and Objective

- NoSQL databases becoming popular
- Which database is most effective?
- Evaluate the performance of Amazon's DynamoDB by running various database read and write operations using a varying number of clients
- Determine in which situations DynamoDB offers a greater advantage than a traditional database

DynamoDB

- NoSQL database service on AWS
- Stores data in key-value pairs
- Automatically handles scaling
- Replicates data across multiple Availability Zones
- Used by: EA, Shazam, IMDb, and more

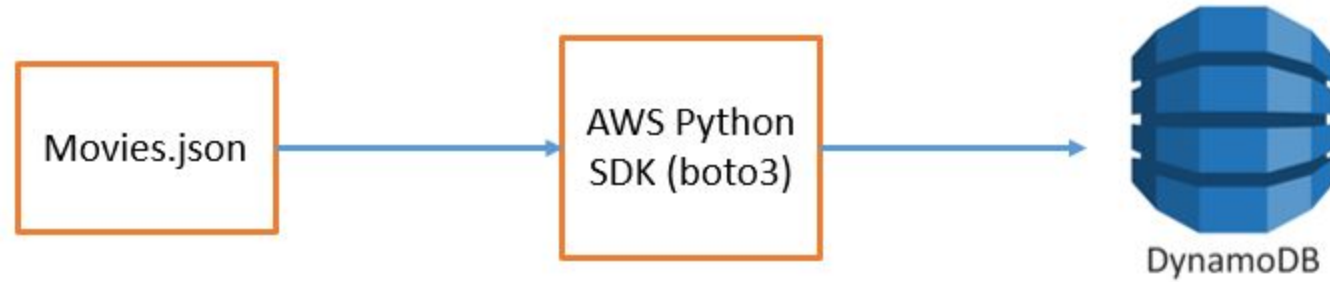
DynamoDB and Boto

- Data is indexed by the primary key
- Attempted to populate the database by importing .csv and JSON files
- Boto - AWS Python SDK, integrates Python and AWS services

Operations

- CreateTable
- UpdateTable
- DeleteTable
- DescribeTable
- ListTables
- BatchGetItem
- BatchWriteItem
- GetItem
- PutItem
- UpdateItem
- DeleteItem
- Query
- Scan

Uploading Data



JSON

```
1 {
2   "year": 1933,
3   "title": "King Kong",
4   "info": {
5     "directors": [
6       "Merian C. Cooper",
7       "Ernest B. Schoedsack"
8     ],
9     "release_date": "1933-03-07T00:00:00Z",
10    "rating": 8,
11    "genres": [
12      "Adventure",
13      "Fantasy",
14      "Horror"
15    ],
16    "image_url": "http://ia.media-
imdb.com/images/M/MV5BMTkxOTIxMDU2OV5BML5BanBnXkFtZTcwNjM5NjQyMg@@._V1_S
X400_.jpg",
17    "plot": "A film crew goes to a tropical island for an exotic
location shoot and discovers a colossal giant gorilla who takes a shine
to their female blonde star.",
18    "rank": 3551,
19    "running_time_secs": 6000,
20    "actors": [
21      "Fay Wray",
22      "Robert Armstrong",
23      "Bruce Cabot"
24    ]
25  }
26 }
```

```

1  {
2  "info": {
3    "M": {
4      "actors": {
5        "L": [
6          {
7            "S": "Fay Wray"
8          },
9          {
10           "S": "Robert Armstrong"
11         },
12         {
13           "S": "Bruce Cabot"
14         }
15       ]
16     },
17     "directors": {
18       "L": [
19         {
20           "S": "Merian C. Cooper"
21         },
22         {
23           "S": "Ernest B. Schoedsack"
24         }
25       ]
26     },
27     "genres": {
28       "L": [
29         {
30           "S": "Adventure"
31         },
32         {
33           "S": "Fantasy"
34         },
35         {
36           "S": "Horror"
37         }
38       ]
39     },
40     "image_url": {
41       "S": "http://ia.media-
42       imdb.com/images/M/MV5BMTkxOTIxMDU2OV5BML5BanBnXkFtZTcwNjM5NjQyMg@@._V1_S
43       X400_.jpg"
44     },
45     "plot": {
46       "S": "A film crew goes to a tropical island for an exotic
47       location shoot and discovers a colossal giant gorilla who takes a shine

```

```

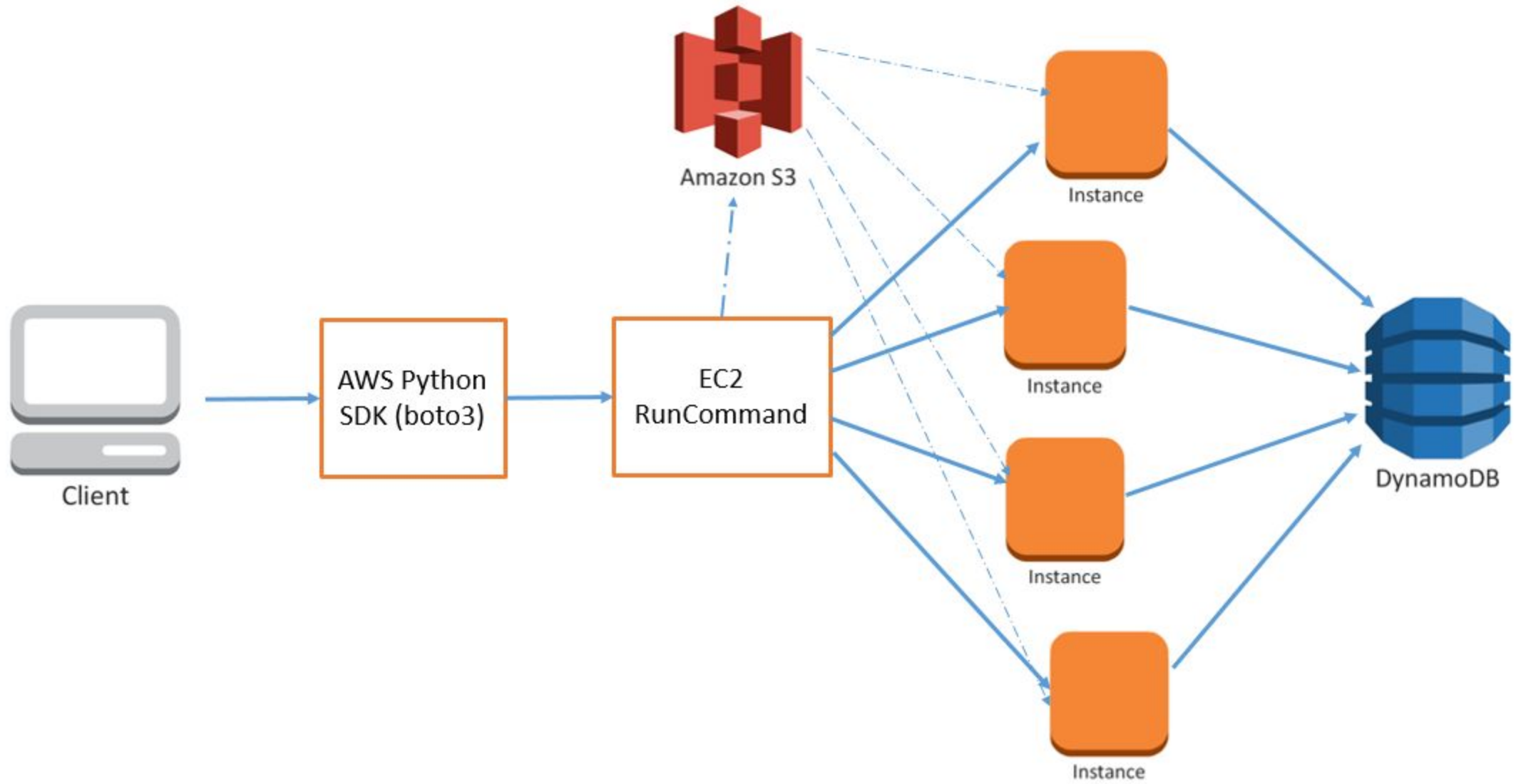
45     },
46     "rank": {
47       "N": "3551"
48     },
49     "rating": {
50       "N": "8"
51     },
52     "release_date": {
53       "S": "1933-03-07T00:00:00Z"
54     },
55     "running_time_secs": {
56       "N": "6000"
57     }
58   },
59 },
60 "title": {
61   "S": "King Kong"
62 },
63 "year": {
64   "N": "1933"
65 }
66 }

```

DynamoDB JSON

The Code

- Utilized Boto functions to integrate Python and DynamoDB operations
- Scripts incorporated a timing feature that outputs execution time
- Some operations were placed in a for loop so average execution time could be calculated and to collect many measurements
- Incorporated multiprocessing to simulate more clients





Amazon Linux AMI 2016.03.0 (HVM), SSD Volume Type - ami-c229c0a2

Select

Amazon Linux
Free tier eligible

The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages.

64-bit

Root device type: ebs Virtualization type: hvm

	Family	Type	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
<input type="checkbox"/>	General purpose	t2.micro Free tier eligible	1	1	EBS only	-	Low to Moderate
<input type="checkbox"/>	General purpose	m4.4xlarge	16	64	EBS only	Yes	High

Commands > Run a command

Run a command

A command document includes the information about the command you want to run. Select a command document from the following list and then specify parameters for the command.

Command document* AWS-RunShellScript

Description Run a shell script or specify the commands to run.

Target instances* No instances selected

Select instances

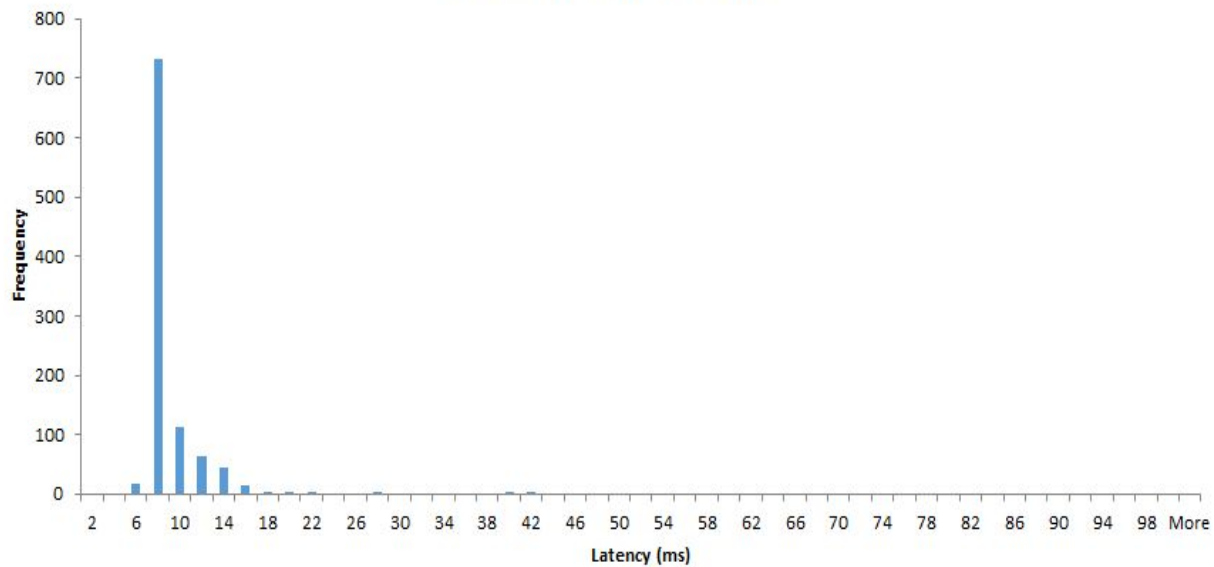
Commands*

```
aws s3 cp s3://cloudcomp-project/putitem.py .  
python putitem.py
```

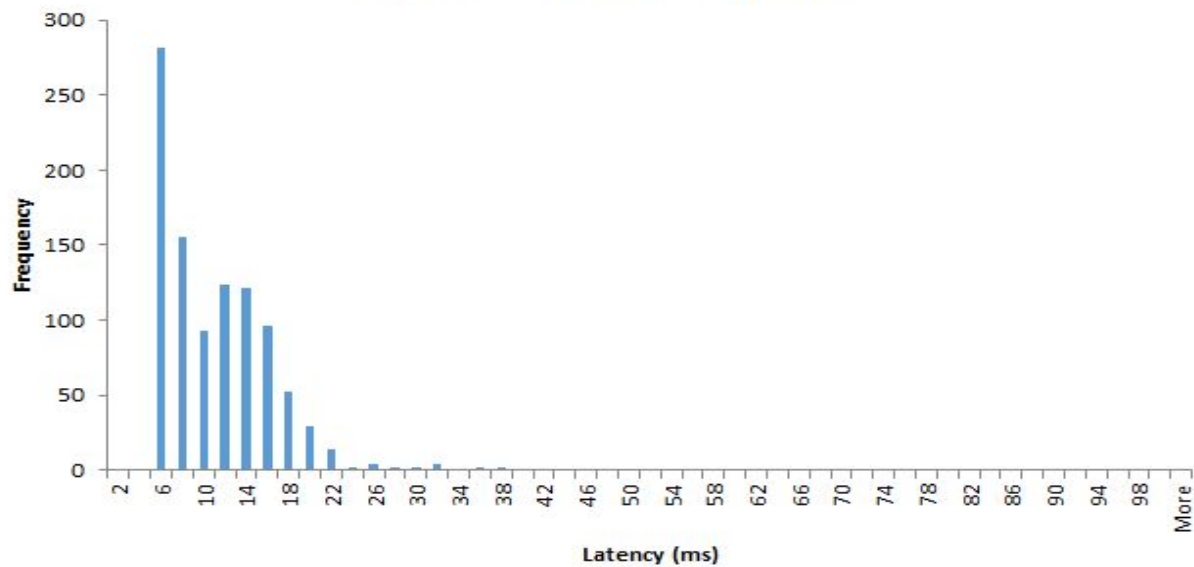
Challenges

- Scaling
 - EC2 instance constraints
 - Throughput capacity limits
- Configuring non-AWS technology to work with DynamoDB
 - SDKs

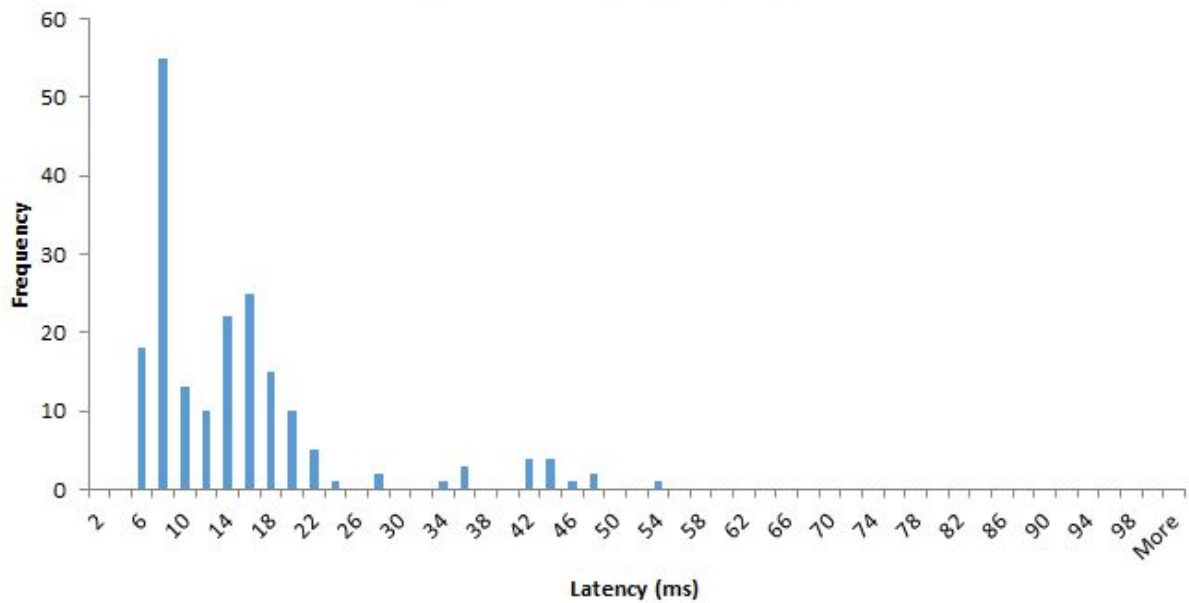
Get Item with 1 Client



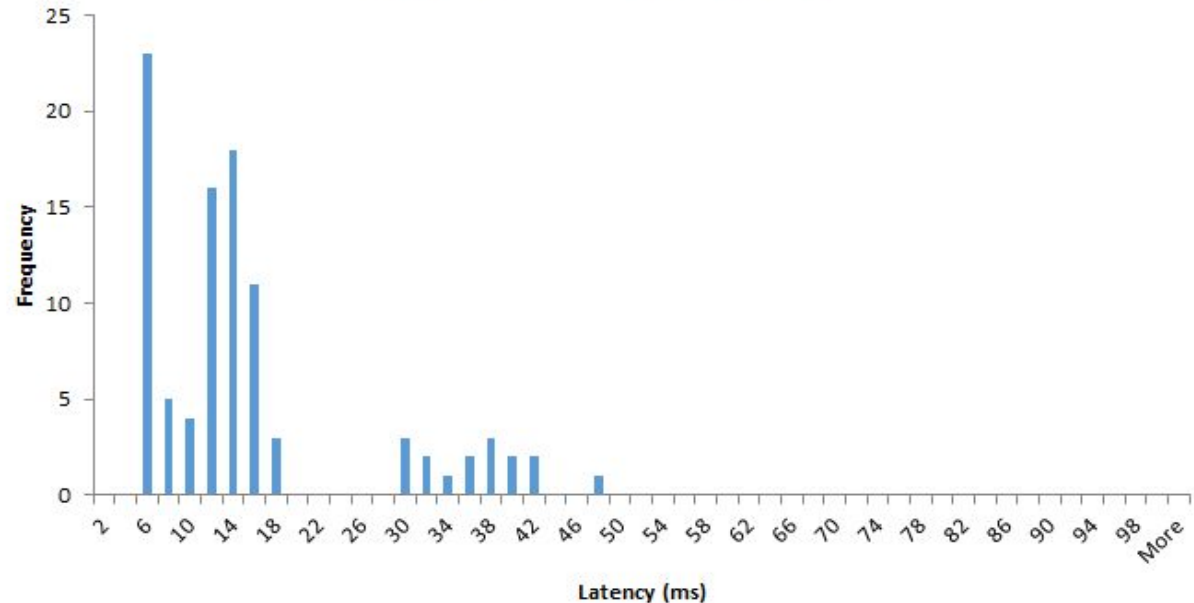
Get Item with 20 Clients



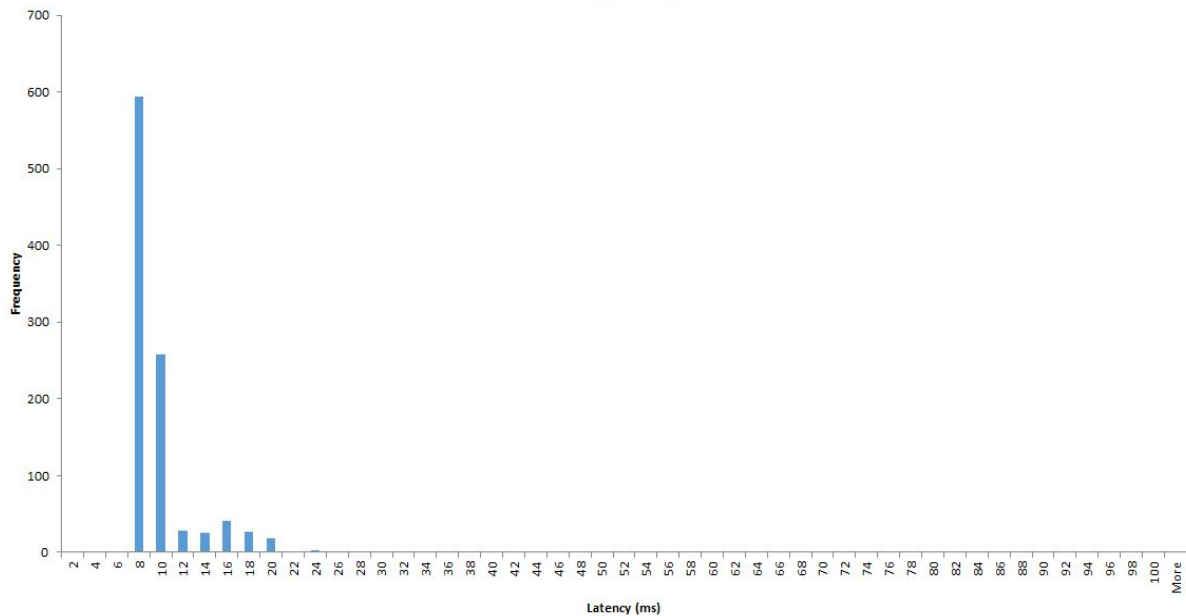
Get Item with 80 Clients



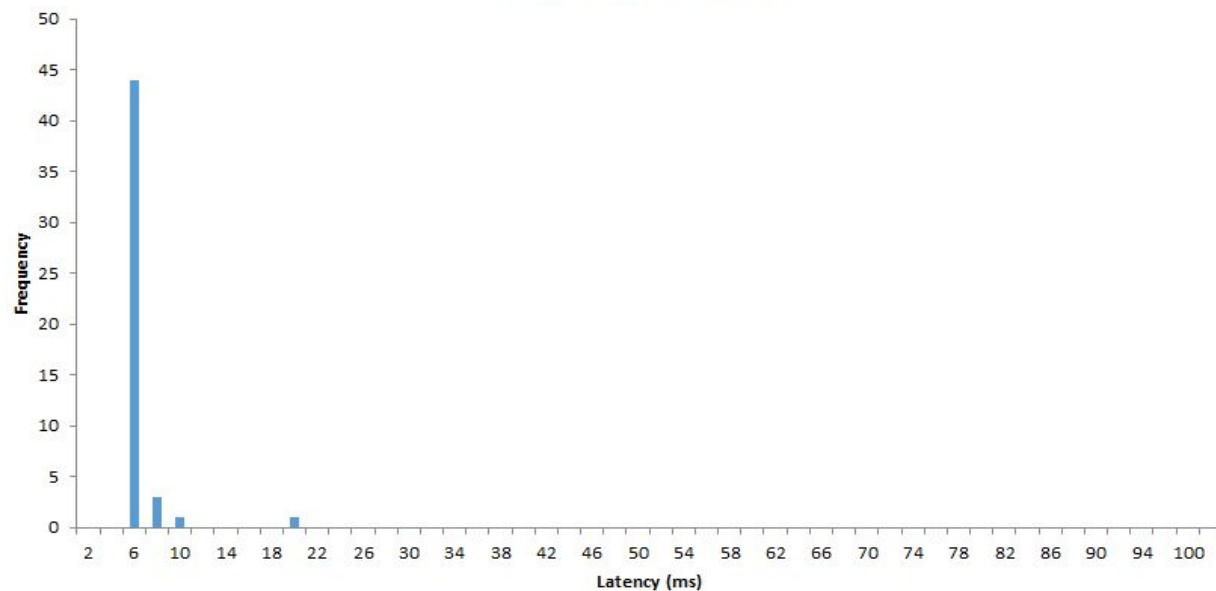
Get Item with 160 Clients



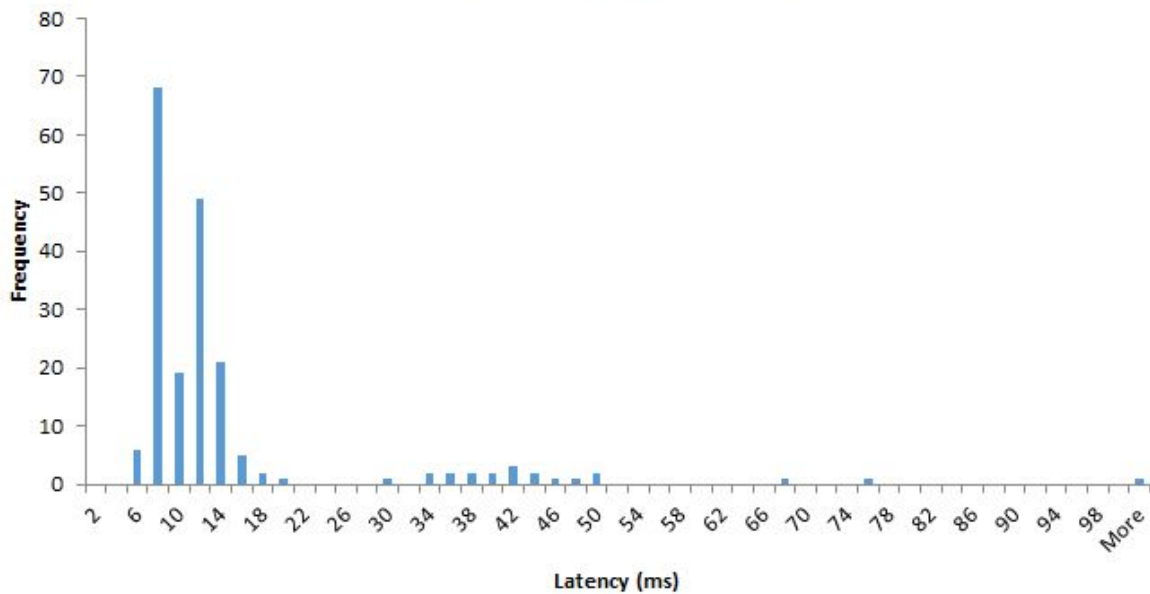
Put Item with 1 Client



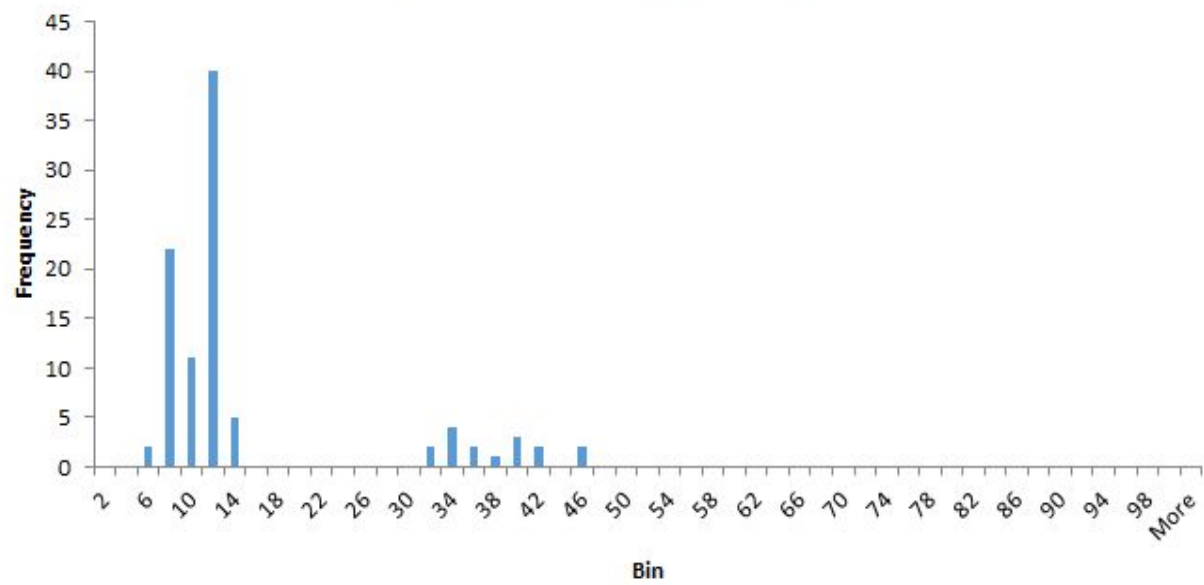
Put Item with 20 Clients



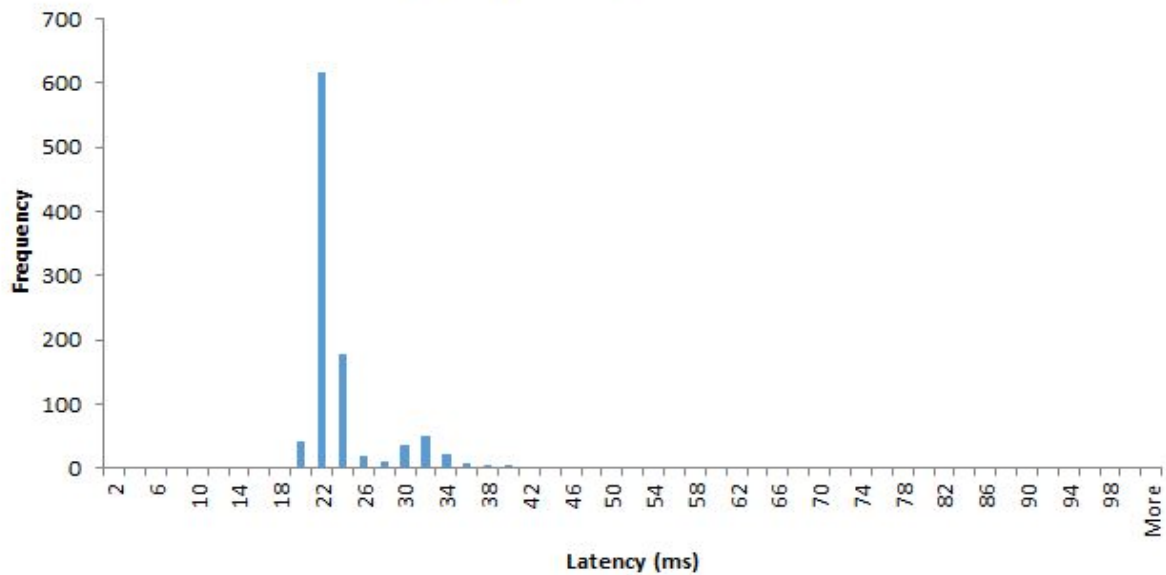
Put Item with 80 Clients



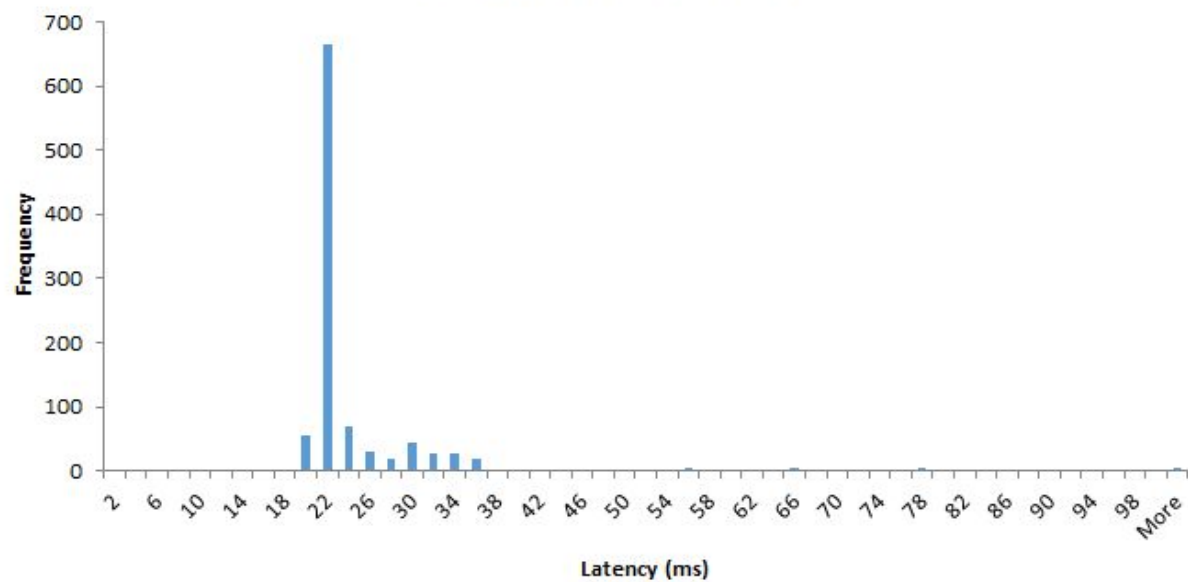
Put Item with 160 Clients



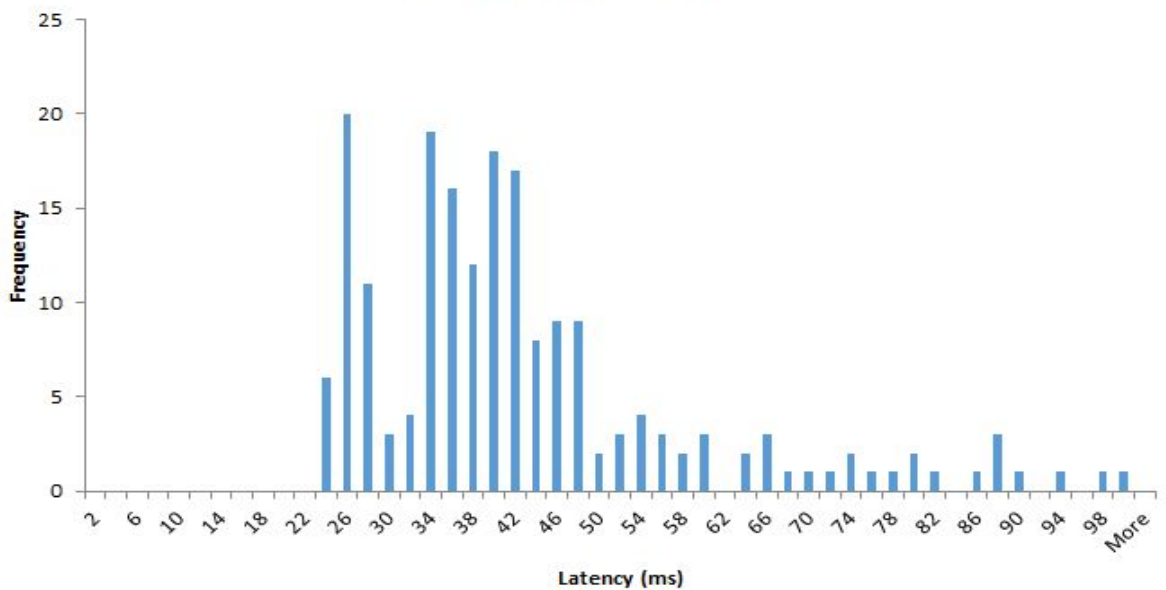
Query with 1 Client



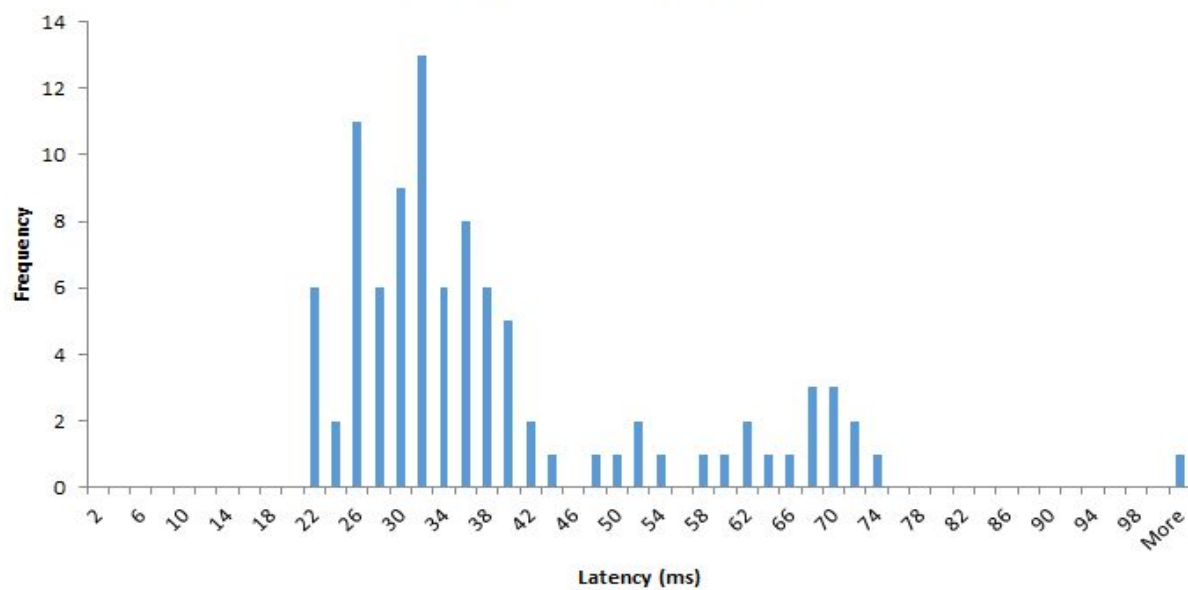
Query with 20 Clients



Query with 80 Clients



Query with 160 Clients



Results

- Latency did not change as clients increased
- More clients → more varied results
- Ideally verified with higher scalability

Conclusion

- DynamoDB is highly and easily scalable
- Difficult to interact with outside of AWS
- Use EC2/AWS resources to manage interaction
- Future research: DynamoDB vs. other NoSQL DBs