

Question 1

Amazon S3 > Buckets > bigdata-dzf-2 > bigdata_hw/

bigdata_hw/

Copy S3 URI

Objects Properties

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Refresh Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Find objects by prefix

< 1 > ⚙

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	Crimes_2001_to_Present.csv	csv	February 6, 2023, 18:47:55 (UTC-06:00)	1.7 GB	Standard

Question 5

Query 1 : X Query 2 : X + -

1 select * from (select primary_type, count(*) as cnt from chicago_crimes group by primary_type) t order by cnt desc limit 5;

SQL Ln 1, Col 124

Run again Explain Cancel Clear Create

Reuse query results
*Athena engine version 3 only

Query results Query stats

Completed Time in queue: 195 ms Run time: 1.391 sec Data scanned: 1.70 GB

Results (5)

Copy Download results

Search rows

< 1 > ⚙

#	primary_type	cnt
1	THEFT	1630499
2	BATTERY	1413736
3	CRIMINAL DAMAGE	881147
4	NARCOTICS	745909
5	ASSAULT	502622

Query 1 : X Query 2 : X + -

1 select * from (select primary_type, count(*) as cnt from chicago_crimes group by primary_type) t order by cnt limit 5;

SQL Ln 1, Col 119

Run again Explain Cancel Clear Create

Reuse query results
*Athena engine version 3 only

Query results Query stats

Completed Time in queue: 107 ms Run time: 1.1 sec Data scanned: 1.70 GB

Results (5)

Copy Download results

Search rows

< 1 > ⚙

#	primary_type	cnt
1	DOMESTIC VIOLENCE	1
2	NON-CRIMINAL (SUBJECT SPECIFIED)	9
3	RITUALISM	24
4	NON - CRIMINAL	38
5	HUMAN TRAFFICKING	99

Question 6

Query 1 : X

Query 2 : X

+ | ▾

1

```
select * from (select location_description, count(*) as cnt from chicago_crimes where primary_type = 'HOMICIDE' group by location_description) t order by cnt desc limit 5;
```

SQL Ln 1, Col 172

Run again

Explain

Cancel

Clear

Create ▾

Reuse query results

*Athena engine version 3 only

Query results

Query stats

Completed

Time in queue: 200 ms

Run time: 1.636 sec

Data scanned: 1.70 GB

Results (5)

Copy

Download results

Search rows

< 1 >

# ▾	location_description ▾	cnt ▾
1	STREET	6229
2	AUTO	1353
3	APARTMENT	1050
4	ALLEY	775
5	HOUSE	637

Question 9

Query 1 : X

Query 2 : X

Query 3 : X

Query 5 : X

+ | ▾

1

```
create table chicago_crimes_parquet
```

2

```
with (format = 'parquet')
```

3

```
as select * from chicago_crimes;
```

SQL Ln 3, Col 33

Run again

Explain

Cancel

Clear

Create ▾

Reuse query results

*Athena engine version 3 only

Query results

Query stats

Completed

Time in queue: 94 ms

Run time: 8.63 sec

Data scanned: 1.70 GB

Query successful.

Question 10

Query 1 : X

Query 5 : X

+ ▼

1 EXPLAIN ANALYZE

2 select max(occurred_at) from chicago_crimes;

SQL Ln 2, Col 45

Run again

Explain

Cancel

Clear

Create ▼

Reuse query results

Athena engine version 3 only

Query results

Query stats

Completed

Time in queue: 96 ms

Run time: 1.328 sec

Data scanned: 1.70 GB

Query Plan

Fragment 1

CPU: 12.91ms, Input: 55 rows (1.45kB), Data Scanned: 0B; per task: std.dev.: 0.00, Output: 1 row (27B)

Output layout: [max]

- Aggregate(FINAL) => [[max]]

CPU: 2.00ms (0.01%), Output: 1 row (27B)

Input avg.: 55.00 rows, Input std.dev.: 0.00%

max := "max"("max_4")

- LocalExchange[SINGLE] () => [[max_4]]

CPU: 2.00ms (0.01%), Output: 55 rows (1.45kB)

Input avg.: 13.75 rows, Input std.dev.: 51.78%

- RemoteSource[2] => [[max_4]]

CPU: 3.00ms (0.01%), Output: 55 rows (1.45kB)

Input avg.: 13.75 rows, Input std.dev.: 51.78%

Fragment 2

CPU: 23.82s, Input: 7728392 rows (199.00MB), Data Scanned: 1.70GB; per task: std.dev.: 36698.29, Output: 55 rows (1.45kB)

Output layout: [max_4]

- Aggregate(PARTIAL) => [[max_4]]

CPU: 327.00ms (1.37%), Output: 55 rows (1.45kB)

Input avg.: 140516.22 rows, Input std.dev.: 6.26%

max_4 := "max"("occurred_at")

- TableScan[awsdatacatalog:HiveTableHandle(schemaName=crime, tableName=chicago_crimes, analyzePartitionValues=Optional.empty), grouped = false] => [[occurred_at]]

CPU: 23.49s (98.60%), Output: 7728392 rows (199.00MB)

Input avg.: 140516.22 rows, Input std.dev.: 6.26%

LAYOUT: crime.chicago_crimes

occurred_at := occurred_at:string:2:REGULAR

Query 1 : X

Query 5 : X

+ ▼

1 EXPLAIN ANALYZE

2 select max(occurred_at) from chicago_crimes_parquet;

SQL Ln 2, Col 52

Run again

Explain

Cancel

Clear

Create ▼

Reuse query results

Athena engine version 3 only

Query results

Query stats

Completed

Time in queue: 185 ms

Run time: 1.121 sec

Data scanned: 26.59 MB

Query Plan

Fragment 1

CPU: 3.57ms, Input: 26 rows (702B), Data Scanned: 0B; per task: std.dev.: 0.00, Output: 1 row (27B)

Output layout: [max]

- Aggregate(FINAL) => [[max]]

CPU: 1.00ms (0.03%), Output: 1 row (27B)

Input avg.: 26.00 rows, Input std.dev.: 0.00%

max := "max"("max_4")

- LocalExchange[SINGLE] () => [[max_4]]

CPU: 0.00ms (0.00%), Output: 26 rows (702B)

Input avg.: 6.50 rows, Input std.dev.: 95.77%

- RemoteSource[2] => [[max_4]]

CPU: 0.00ms (0.00%), Output: 26 rows (702B)

Input avg.: 6.50 rows, Input std.dev.: 95.77%

Fragment 2

CPU: 3.40s, Input: 7728392 rows (199.00MB), Data Scanned: 26.59MB; per task: std.dev.: 11554.00, Output: 26 rows (702B)

Output layout: [max_4]

- Aggregate(PARTIAL) => [[max_4]]

CPU: 612.00ms (17.98%), Output: 26 rows (702B)

Input avg.: 297245.85 rows, Input std.dev.: 9.98%

max_4 := "max"("occurred_at")

- TableScan[awsdatacatalog:HiveTableHandle(schemaName=crime, tableName=chicago_crimes_parquet, analyzePartitionValues=Optional.empty), grouped = false] => [[occurred_at]]

CPU: 2.79s (81.99%), Output: 7728392 rows (199.00MB)

Input avg.: 297245.85 rows, Input std.dev.: 9.98%

LAYOUT: crime.chicago_crimes_parquet

occurred_at := occurred_at:string:2:REGULAR

Parquet format is more efficient than normal format when extracting the data from the tables.

Question 11

Query 1 : X

Query 5 : X

+ ▼

```
1 create table crimes_for_download as
2 SELECT
3     primary_type,
4     community_area,
5     count(*) as count
6 FROM chicago_crimes
7 GROUP BY primary_type, community_area;
```

SQL Ln 7, Col 39

Run again

Explain

Cancel

Clear

Create ▼

☐ Reuse query results
*Athena engine version 3 only

Query results

Query stats

Completed

Time in queue: 98 ms Run time: 2.603 sec Data scanned: 1.70 GB

Query successful.

I preview the table and download the result. The resulting file is csv.