DynamoDB

What is a Query?

A CLOUD GURU

- A Query operation finds items in a table based on the Primary Key attribute and a distinct value to search for.

- e.g. select an item where the user ID is equal to 212, will select all the attributes for that item, e.g. first name, surname, email etc.

What is a Query?

A CLOUD GURU

- Use an optional Sort Key name and value to refine the results.
- e.g. if your Sort Key is a timestamp, you can refine the query to only select items with a timestamp of the last 7 days.
- By default, a Query returns all the attributes for the items but you can use the ProjectionExpression parameter if you want the query to only return the specific attributes you want.
- e.g. if you only want to see the email address rather than all the attributes

DynamoDB

What is a Query?

A CLOUD GURU

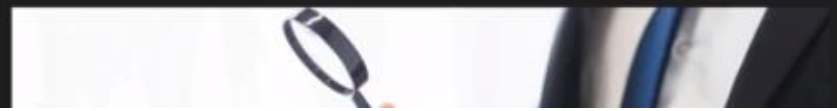
- Results are always sorted by the Sort Key
- Numeric order - by default in ascending order (1, 2, 3, 4)
- ASCII character code values
- You can reverse the order by setting the ScanIndexForward parameter to false
- By default, Queries are Eventually Consistent
- You need to explicitly set the query to be Strongly Consistent

DynamoDB

What is a Scan?

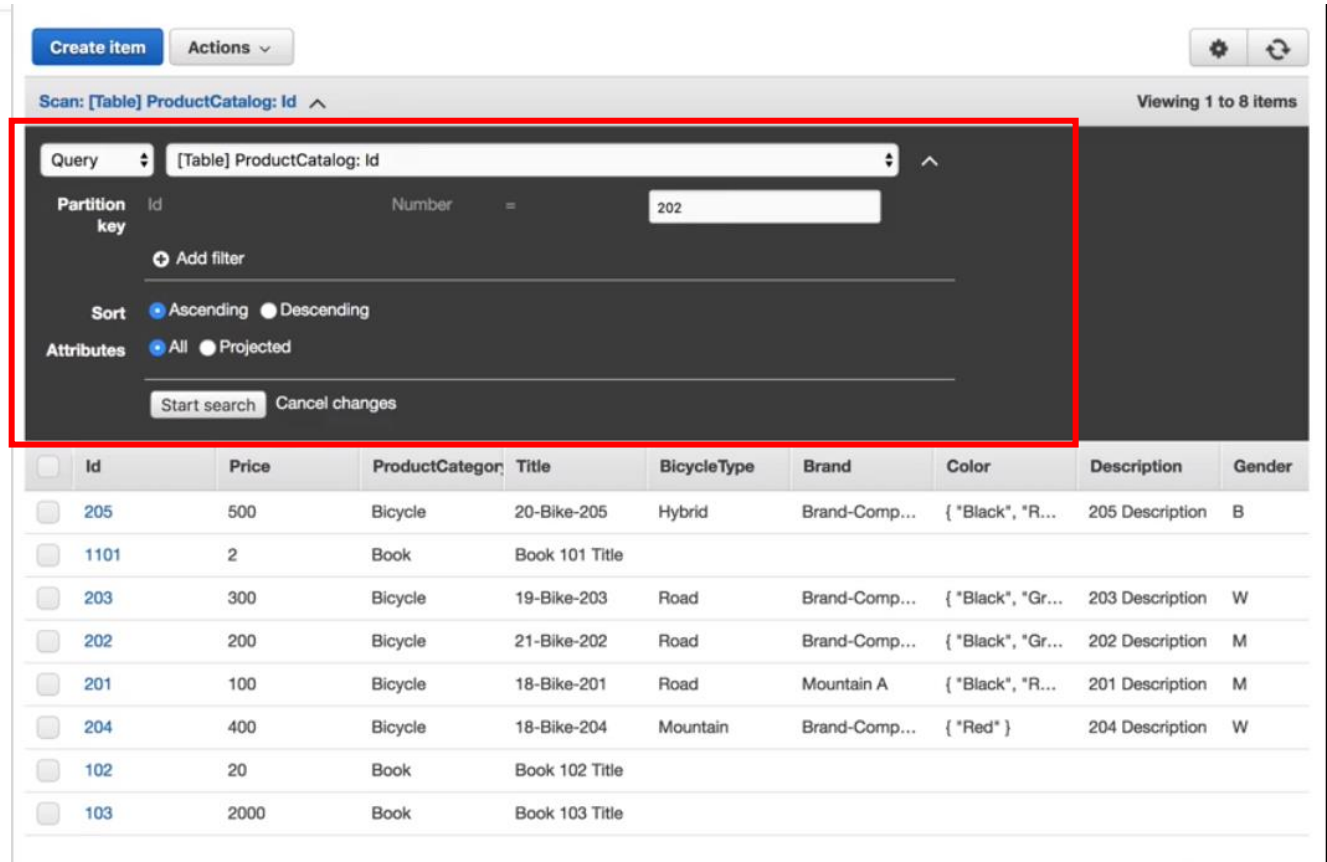
A CLOUD GURU

- A Scan operation examines every item in the table.
- By default returns all data attributes.
- Use the ProjectionExpression parameter to refine the scan to only return the attributes you want.



Id	Price	ProductCategor	Title	BicycleType	Brand	Color	Description	Gender
205	500	Bicycle	20-Bike-205	Hybrid	Brand-Comp...	{ "Black", "R...	205 Description	B
203	300	Bicycle	19-Bike-203	Road	Brand-Comp...	{ "Black", "Gr...	203 Description	W
202	200	Bicycle	21-Bike-202	Road	Brand-Comp...	{ "Black", "Gr...	202 Description	M
204	400	Bicycle	18-Bike-204	Mountain	Brand-Comp...	{ "Red" }	204 Description	W
103	2000	Book	Book 103 Title					

Scan vs Query



Id	Price	ProductCategor	Title	BicycleType	Brand	Color	Description	Gender
205	500	Bicycle	20-Bike-205	Hybrid	Brand-Comp...	{ "Black", "R...	205 Description	B
1101	2	Book	Book 101 Title					
203	300	Bicycle	19-Bike-203	Road	Brand-Comp...	{ "Black", "Gr...	203 Description	W
202	200	Bicycle	21-Bike-202	Road	Brand-Comp...	{ "Black", "Gr...	202 Description	M
201	100	Bicycle	18-Bike-201	Road	Mountain A	{ "Black", "R...	201 Description	M
204	400	Bicycle	18-Bike-204	Mountain	Brand-Comp...	{ "Red" }	204 Description	W
102	20	Book	Book 102 Title					
103	2000	Book	Book 103 Title					

Query or Scan?

- Query is more efficient than a Scan.
- Scan dumps the entire table, then filters out the values to provide the desired result - removing the unwanted data
- This adds an extra step of removing the data you don't want.
- As the table grows, the scan operation takes longer.
- Scan operation on a large table can use up the provisioned throughput for a large table in just a single operation.

How To Improve Scan Performance

A CLOUD GURU

- By default, a scan operation processes data sequentially in returning 1MB increments before moving on to retrieve the next 1 MB of data. It can only scan one partition at a time.

- You can configure DynamoDB to use Parallel scans instead by logically dividing a table or index into segments and scanning each segment in parallel.

- Best to avoid parallel scans if your table or index is already incurring heavy read / write activity from other applications.

Scan Vs Query Exam Tips

- A Query operation finds items in a table using only the Primary Key attribute.
- You provide the Primary Key name and a distinct value to search for.
- A Scan operation examines every item in the table.
- By default returns all data attributes.
- Use the ProjectionExpression parameter to refine the results.

Scan Vs Query Exam Tips

A CLOUD GURU

- Query results are always sorted by the Sort Key is there is one.
- Sorted in ascending order.
- Set ScanIndexForward parameter to false to reverse the order - queries only.
- Query operation is generally more efficient than a Scan.

Scan Vs Query Exam Tips

A CLOUD GURU

- Reduce the impact of a query or scan by setting a smaller page size which uses fewer read operations.
- Isolate scan operations to specific tables and segregate them from your mission-critical traffic.
- Try Parallel scans, rather than the default sequential scan.
- Avoid using scan operations if you can: design tables in a way that you can use the Query, Get, or BatchGetItem APIs.