CS 683 Emerging Technologies
Fall Semester, 2008
Doc 13 SimpleDb
Oct 16 2008

References

Amazon SimpleDB Developer Guide API Version 2007-11-07

Programming Amazon Web Services, James Murty O'Reilly Media, Inc, Chapter 13

SimpleDB

Simple database

Accessed via REST or SOAP

No SQL

No database schema

Simple queries

Data Model

Domains
Items
Attributes

Domain

Just a collection of items

- Sort of like a table
- ArrayList of item objects is better analogy

Items in a domain do not have to be related

Queries are on a single domain

No foreign keys to other domains

Item & Attribute

Item

- Collection of attributes
- Have names

Items in the same domain
can have different attributes

Attribute

- Has a name
- Multiple text values

Only text values
No numbers, etc.

Domain Example

	Category	Subcat.	Name	Color	Size	Make	Model
Item_01	Clothes	Sweater	Cathair Sweater	Siamese	Small, Medium,Large		
Item_02	Clothes	Pants	Designer Jeans	Paisley, Acid, Wash	30x32, 32x32, 32x34		
Item_03	Car Parts	Engine	Turbos			Audi	S4
Item_04	Car Parts	Emissions	02 Sensor			Audi	S4
Item_05	Motorcycle Parts	Bodywork	Fender Eliminator	Blue		Yamaha	R1
Item_06	Motorcycle Parts, Clothing	Clothing	Leather Pants	Black	Small, Medium		

Row = item

Column header = attribute name

Cell = attribute values

Eventual Consistency

SimpleDB domains are replicated on multiple machines

There can be a delay in propagating any changes

Some Limits

Domain size	10 GB per domain
Domain size	250,000,000 attribute name-value pairs
Domains per account	100
Attribute name-value pairs per item	256
Attribute name length	1024 bytes
Attribute value length	1024 bytes
Item name length	1024 bytes
Maximum query execution time	5 seconds

SimpleDB Operations

Create Domain

Delete Domain

List Domains

Put

Add, modify, or remove data

Delete

Remove items, attributes, or attribute values from your domain

Get

Retrieve the attributes and values of any item ID that you specify

Query

Query a domain against attribute values

QueryWithAttributes

Query a domain and retrieve the results for matching items

GetAttributes

Returns all of the attributes associated with the item.

Can restrict response to a particular attribute

Request Parameters

Item name	Required
Domain name	Required
Attribute Name	Optional

Queries

Basic Syntax

[attributeName comparison value]

['title' starts-with 'Erlang']
['size' > 10]
['name' != 'John']

Operators

=
!=
>
>=
<
<=
starts-with
does-not-start-with

and & or

Can and/or queries on the same attribute

['Year' > '1975' and 'Year' < '2008']

['Rating' = '***' or 'Rating' ='*****']

['Year' > '1950' and 'Year' <'1960' or 'Year' starts-with '193' or 'Year' = '2007']

Set Operators

intersection

union

not

['first name' = 'John'] intersection ['last name' = 'Smith']

['tag' starts-with 'Amazon'] union ['description' = 'SimpleDB']

not ['first name' = 'John']

not Verses !=

['Keyword' != 'Book']

Find all items that contain attribute 'Keyword'

Return those that don't have value 'Book'

All items returned have attribute 'Keyword'

not['Keyword' = 'Book']

Find all items that contain attribute 'Keyword'

And have a value equal to 'Book'

Return all other items

Will return items without attribute 'Keyword'

Numbers

"This provides application designers with the flexibility of enforcing data restrictions at the application layer without the data store enforcing constraints."

Translation

All that useful stuff with numbers you normally use does not work

Example - Comparing numbers

As Number

2 < 10

As String

2 > 10

All data in SimpleDb is a string

Number Problems

Any query on "numerical" data will give wrong results

['age' > '10']

will return item with age = '2'

Amazon's Recommendation

Negative Numbers Offsets

14.58, -12536.791, 20071109

+ 100,000

Make all number positive
Add a large integer

100014.58, 87463.209, 20171109

Zero Padding

00100014.58, 00087463.209, 20171109

Convert all queries ['attribute' > '500'] ⟶ ['attribute' > '00100500']

Convert all data when using it 00100014.58 ⟶ 14.58

- 100,000

Dates and other Types

Everything is a string

It is your job to

- Convert between types
- Structure strings so queries work

Query & QueryWithAttributes

Requests sent to SimpleDB

SOAP or Rest

Can request first N resuts

Query

Returns item names the satisfy query

QueryWithAttributes

Returns attributes the satisfy query

BoxUsage

Queries that take longer than 5 seconds to execute time out

All results return execution time taken - BoxUsage

Optimizing Queries

"The service also automatically indexes every piece of information stored within it to ensure queries will run as quickly as possible without the need for performance tuning."

The developers guide talks about query tuning

ands are faster than intersection

["attribute1" > "value1" and "attribute1" < "value2"]
["attribute1" > "value1"] intersection ["attribute1" < "value2"]

Quote from Programming Amazon Web Services

Scaling

Amazon claims SimpleDB scales

Not clear what makes SimpleDB scale more than other database

Amazon's replication of domains should help it scale

XML, REST, SOAP are slow