



DEV
DAY

DynamoDB and Schema Design

Yuriy Zubarev, AWS, Berlin | 2018





DEV DAY

Build your cloud skills with AWS

Use the discount code when booking online and get 20 % off:
MKBERSUM18-1-19P6OPOXN6D3Y

Date	Training	Location	Duration	Language
Aug 28-30	Developing on AWS	Berlin	3 days	English
Sep 5	Running Container-Enabled Microservices on AWS	Berlin	1 day	English
Sep 6	Deep Learning on AWS	Berlin	1 day	English
Sep 18	Building a Serverless Data Lake on AWS	Berlin	1 day	English
Oct 09-11	Systems Operations on AWS	Berlin	3 days	English
Oct 23-25	DevOps Engineering on AWS	Berlin	3 days	English

Terms & Conditions: Please note that this discount only applies to Public Schedule Classes sold by AWS in Germany. The promotion expires on 30.09.2018

Why are we talking about schemas?

 **Cory House**  @housecor · 10 Mar 2015 ▼

Schemaless DB's:

Promise: Don't worry about a **schema**.

Reality: It's now your responsibility to ensure data consistency.

#nosql

 9  115  55  

Why are we talking about schemas?



Cory House  @housecor · 10 Mar 2015

Schemaless DB's:

Promise: Don't worry about a **schema**.

Reality: It's now your responsibility to ensure data consistency.

“... for maximizing performance and minimizing throughput costs”

9

115

55

Mindset

SQL: normalize until it
hurts, de-normalize until
it works

- old adage

NoSQL: aggregate until
it hurts, de-aggregate
until it works

- new adage?

**NoSQL - you shouldn't start designing your schema
until you know the questions it will need to answer**

What and How; SQL

- Data Redundancy (eliminate)
- Data Integrity and Accuracy (ensure)
 - Relationships
 - Normalization
 - 1NF
 - ...
 - Integrity Rules

What and How; NoSQL

- Data size
- Data shape
- Data velocity
- Keep related data together
- Use sort order
- Use GSIs
- Distribute queries

DynamoDB Core Components

- DynamoDB
 - Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability.
- Tables, Items, Attributes
- Primary Key
 - simple – Partition Key only
 - composite – Partition Key + Sort Key
- Secondary Indexes
 - Global Secondary Index (GSI)
 - Local Secondary Index (LSI)
- RCU
- WCU

1:1 relationships

Users table

<i>Partition Key</i>	<i>Attributes</i>
SSN = 1234567	name = Anna; email = anna@gmx.net; occupation = ...
SSN = 2345678	name = Jan; email = jan@t-online.de; occupation = ...

1:N relationships

Device-measurements table

<i>Partition Key</i>	<i>Sort Key</i>	<i>Attributes</i>
DeviceId = 1	epoch = 5513A97C	temp = 30; pressure = 90
DeviceId = 1	epoch = 5513A9DB	temp = 29; pressure = 85

1:N relationships (non-homogenous)

Users table

<i>Partition Key</i>	<i>Sort Key</i>	<i>Attributes</i>
SSN = 1234567	info_type = pers	eye_color = blue; height: 170
	info_type = proff	occupation = Developer; skills: { “fixing code”, ...
SSN = 2345678	info_type = pers	eye_color = green; height: 175
	info_type = proff	occupation = Doctor; skills: { “fixing people”, ...

Indexes; GSI Overloading

Users table

<i>Partition Key</i>	<i>Sort Key</i>	<i>Attributes</i>
SSN = 1234567	info_type = pers	attr1 = blue; attr2: 170
	info_type = proff	attr1 = Developer; attr2: 100000
SSN = 2345678	info_type = pers	attr1 = green; attr2: 175
	info_type = proff	attr1 = Doctor; attr2: 120000

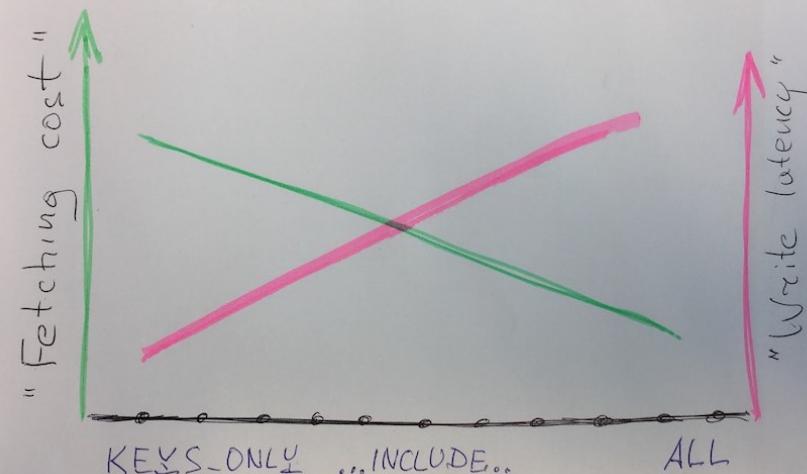
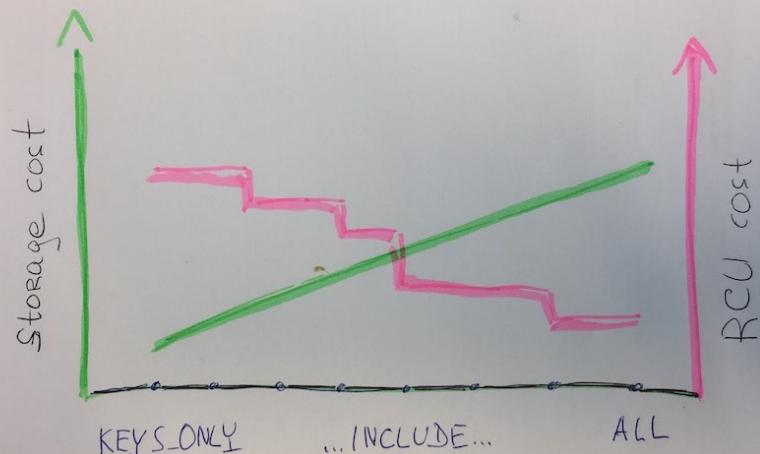
N:M relationships

users-accounts table			accounts-users-gsi	
<i>Partition Key</i>	<i>Sort Key</i>		<i>Partition Key</i>	<i>Sort Key</i>
SSN = 1234567	account_id = 1		account_id = 1	SSN = 1234567
SSN = 1234567	account_id = 3		account_id = 1	SSN = 2345678
SSN = 2345678	account_id = 1		account_id = 2	SSN = 3456789
SSN = 3456789	account_id = 2		account_id = 3	SSN = 1234567

Indexes; Attribute Projections

Tradeoff: cost for provisioned throughput vs cost of storing extra attributes

“... for maximizing performance and minimizing throughput costs”



Indexes; Sparse Indexes

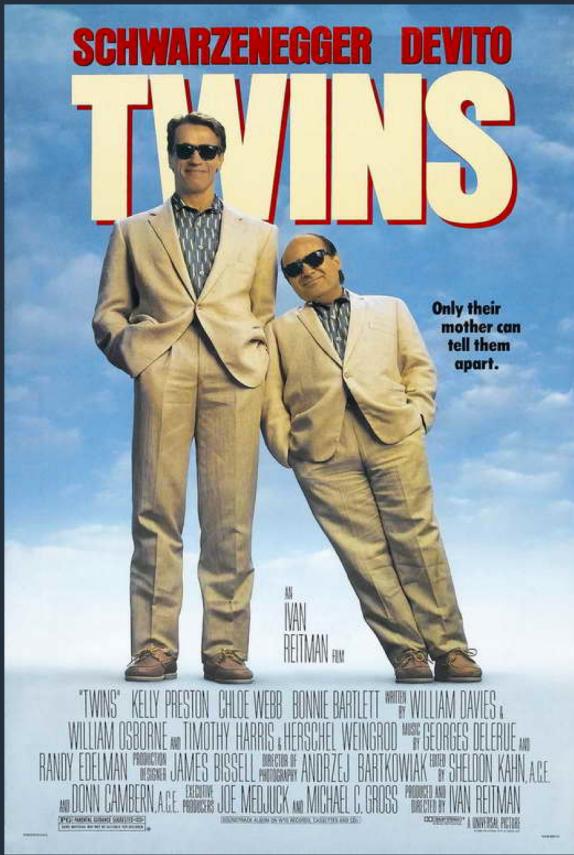
Table

Primary Key		Data Attr		GSI	Projected Attributes...				
Partition Key	Sort Key				Primary Key				
Player_ID	Game_ID	Attribute 1	Attribute 2		Partition Key				
Rick	Game_1	Score: 36,750 <i>(game score)</i>	Date: 2018-01-19 <i>(date of game)</i>		Award	Player_ID	Game_ID	Score	
		Score: 69,450 <i>(game score)</i>	Date: 2018-02-02 <i>(date of game)</i>		Champ	Rick	Game_3	135,900	
	Game_2	Score: 135,900 <i>(game score)</i>	Date: 2018-03-10 <i>(date of game)</i>			Padma	Game_6	147,300	
Padma	Game_3	Score: 25,350 <i>(game score)</i>	Date: 2018-01-27 <i>(date of game)</i>		Award: Champ <i>(type of award)</i>	Padma	Game_7	169,100	
		Score: 69,450 <i>(game score)</i>	Date: 2028-01-19 <i>(date of game)</i>						
	Game_4	Score: 147,300 <i>(game score)</i>	Date: 2018-02-02 <i>(date of game)</i>		Award: Champ <i>(type of award)</i>				
	Game_5	Score: 169,100 <i>(game score)</i>	Date: 2018-03-10 <i>(date of game)</i>						

Indexes; GSI Sharding

Table	Primary Key	Data Attributes...								
	Partition Key									
	Event_ID	Attribute 1		Attribute 2		Attribute 3		Attribute 4		...
EID_12345	Time: event timestamp)	State:	INFO	GSI PK:	(random: 0-N) (random GSI-PK value)	GSI SK:	INFO#2018-02-07T08:42:40 (composite state-time)	...etc.		
		(event state)								
EID_12346	Time: event timestamp)	State:	CRITICAL	GSI PK:	(random: 0-N) (random GSI-PK value)	GSI SK:	CRITICAL#2018-02-07T08:32:40 (composite state-time)	...etc.		
EID_12347	Time: event timestamp)	State:	WARN	GSI PK:	(random: 0-N) (random GSI-PK value)					
		(event state)				GSI SK: WARN#2018-02-07T08:22:40 (composite state-time)		...etc.		
EID_12348	Time: event timestamp)	State:	INFO	GSI PK:	(random: 0-N) (random GSI-PK value)	GSI SK:	INFO#2018-02-07T08:12:40 (composite state-time)	...etc.		
		(event state)								

Indexes; Replica



Time-Series Data

Current table

Provisioned at: WCU=750 and RCU=300

Primary Key		Attributes		
Partition Key	Sort Key	Radiant Intensity	Wavelength	...
2018-03-15	00:00:00.000	17.373 W/Sr	713	
2018-03-15	00:00:00.001	17.373 W/Sr	713	
2018-03-15	00:00:00.002	17.373 W/Sr	713	
2018-03-15	00:00:00.003	17.373 W/Sr	713	
...				

Previous table

Provisioned at: WCU=1 and RCU=100

Primary Key		Attributes		
Partition Key	Sort Key	Radiant Intensity	Wavelength	...
2018-03-14	00:00:00.001	13.669 W/Sr	456	
2018-03-14	00:00:00.002	13.522 W/Sr	459	
2018-03-14	00:00:00.003	13.596 W/Sr	457	
2018-03-14	00:00:00.004	15.721 W/Sr	425	
...				

Older table

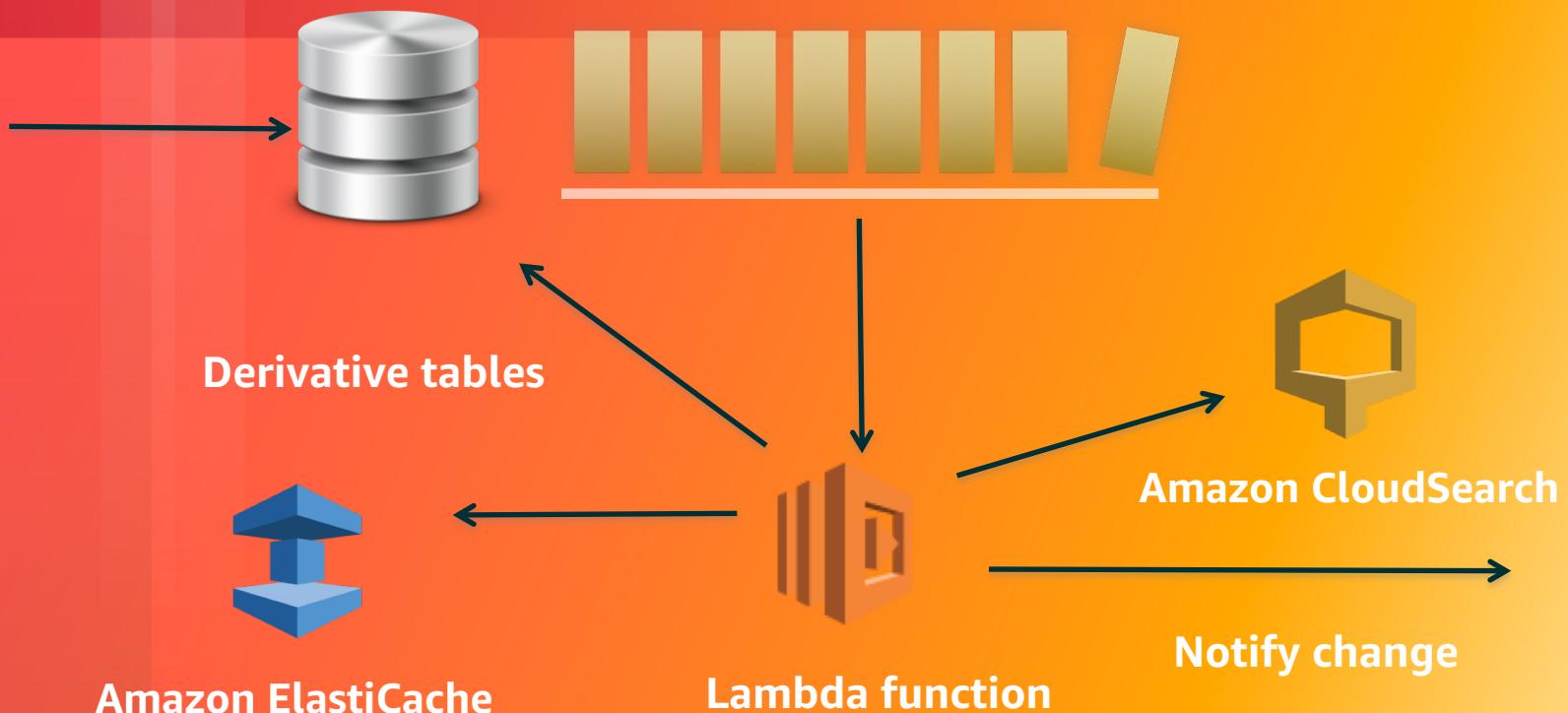
Provisioned at: WCU=1 and RCU=1

Primary Key		Attributes		
Partition Key	Sort Key	Radiant Intensity	Wavelength	...
2018-03-10	00:00:00.001	13.669 W/Sr	456	
2018-03-10	00:00:00.002	13.522 W/Sr	459	
2018-03-10	00:00:00.003	13.596 W/Sr	457	
2018-03-10	00:00:00.004	15.721 W/Sr	425	
...				

What and How

- Data size
- Data shape
- Data velocity
- Keep related data together
- Use sort order
- Use GSIs
- Distribute queries

Mindset





DEV DAY



Yuriy Zubarev, AWS

03-July-2018

Danke. Thank you.

Use the discount code when booking online and get 20 % off:
MKBERSUM18-1-19P6OP0XN6D3Y

Date	Training	Location	Duration	Language
Aug 28-30	Developing on AWS	Berlin	3 days	English
Sep 5	Running Container-Enabled Microservices on AWS	Berlin	1 day	English
Sep 6	Deep Learning on AWS	Berlin	1 day	English
Sep 18	Building a Serverless Data Lake on AWS	Berlin	1 day	English
Oct 09-11	Systems Operations on AWS	Berlin	3 days	English
Oct 23-25	DevOps Engineering on AWS	Berlin	3 days	English

Terms & Conditions: Please note that this discount only applies to Public Schedule Classes sold by AWS in Germany.
The promotion expires on 30.09.2018