



C* Keys: Partitioning, Clustering, & CrossFit

Adam Hutson - Data Architect, DataScale Inc.

The DataScale logo, featuring the word "datascale" in a bold, white, sans-serif font. A white swoosh underline starts under the "d", goes under the "a", "t", "a", and "s", and then loops around the "c" and "a" at the end.

Who am I & What do we do?



Adam Hutson

Data Architect @ DataScale -> www.datascale.io

DataStax MVP for Apache Cassandra

DataScale provides hosted data platforms as a service

Offering Cassandra & Spark, with more to come

Currently hosted in Amazon & Azure





CASSANDRA
SUMMIT 2016

Overview



- | | |
|---|-------------------------|
| 1 | Why |
| 2 | Partition |
| 3 | Partition Key |
| 4 | Composite Partition Key |
| 5 | Clustering Columns |

Why give this presentation?

Partitioning & Clustering should be the foundation.

Too often glossed over.

Has the biggest impact to performance of the cluster



CASSANDRA
SUMMIT **2016**

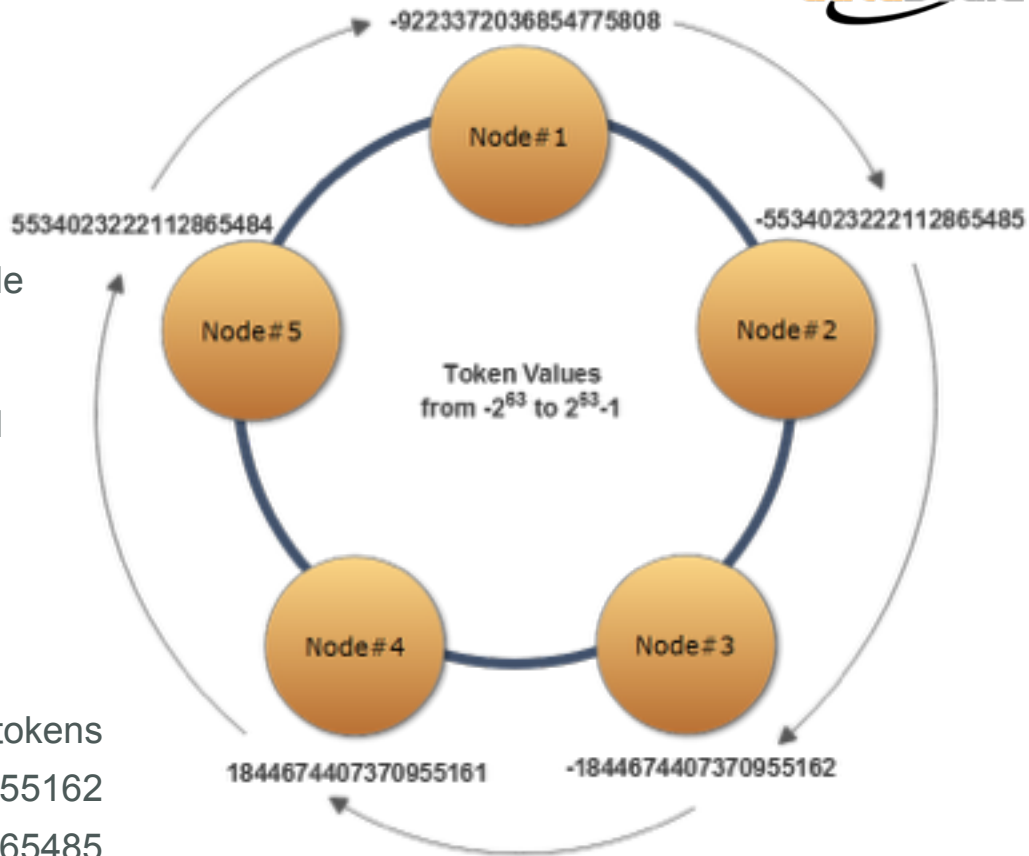
Partition



Partition Explained

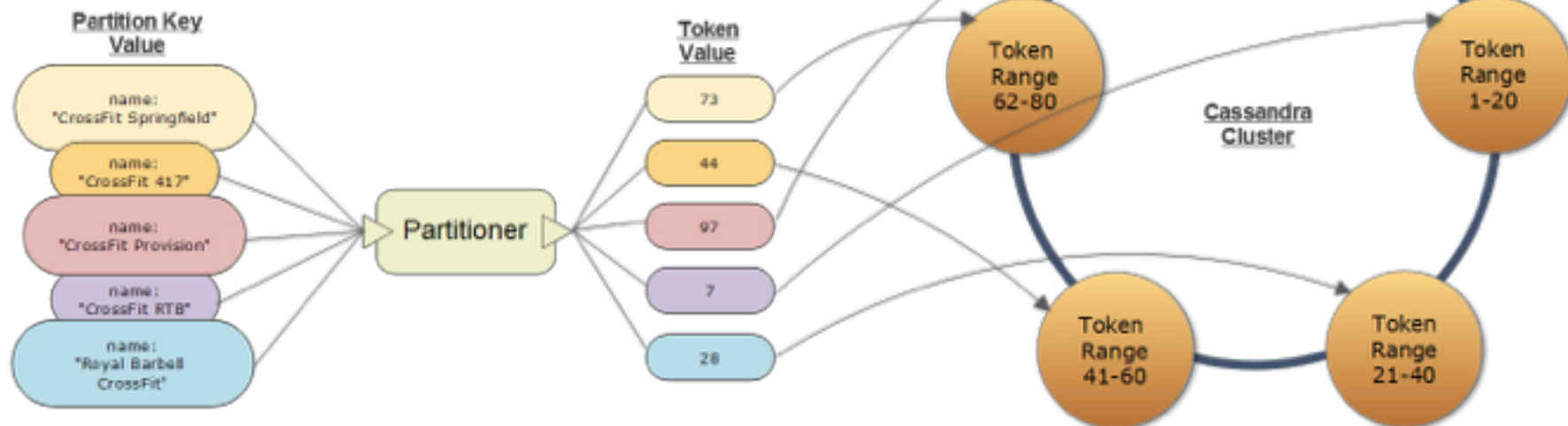
- Token values can range from -2^{63} to $2^{63}-1$.
- Nodes in the cluster/ring are assigned a single token.
- A node is responsible for the token value and expands to the previous node's token.
- A Partitioner decides where a partition key maps onto the cluster/ring.

Node #3 is responsible for tokens
from -1844674407370955162
to -5534023222112865485



Partition Explained

```
CREATE TABLE crossfit_gyms
(
  name text,
  city text,
  state text,
  PRIMARY KEY (name)
)
```





Partition Key



Partition Key Explained

The Partition Key is:

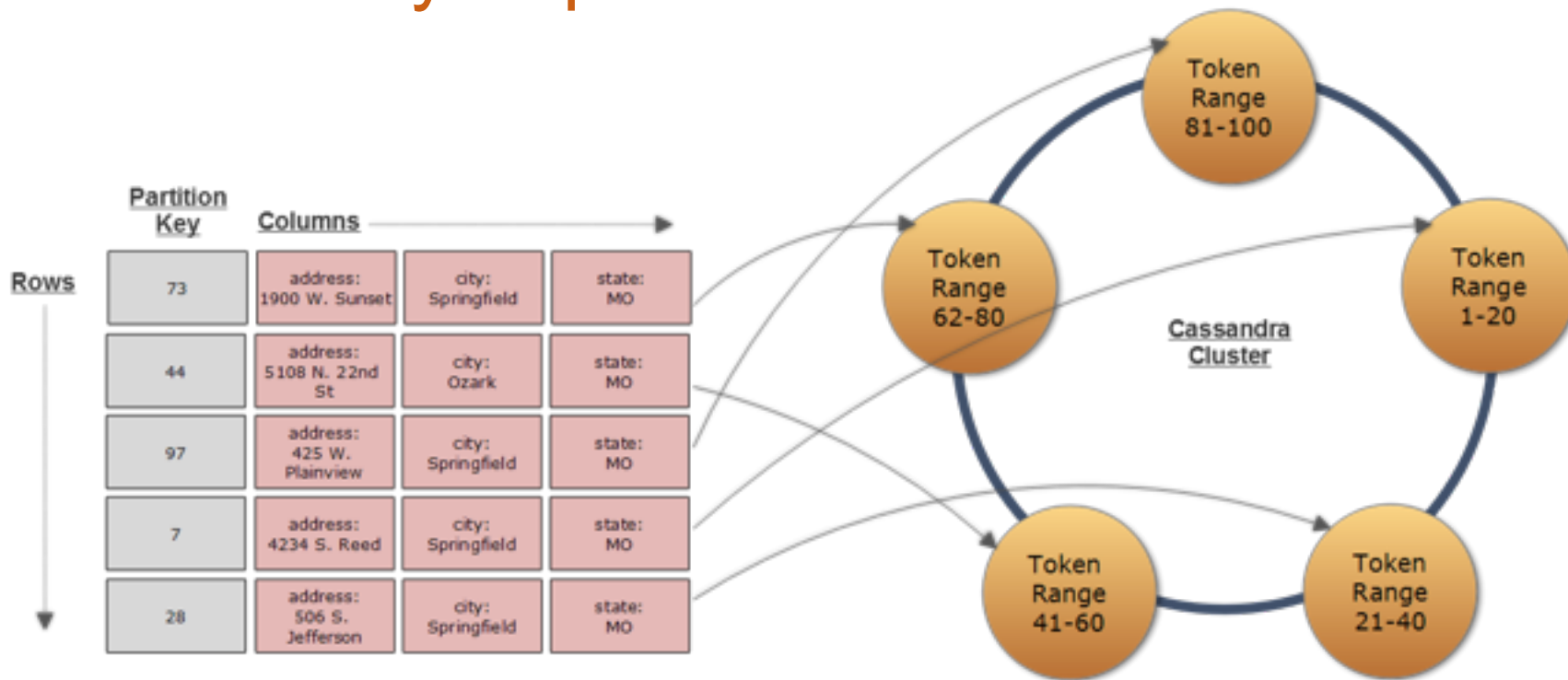
- responsible for distribution of data amongst the nodes
- the first column defined in the PRIMARY KEY

crossfit_gyms

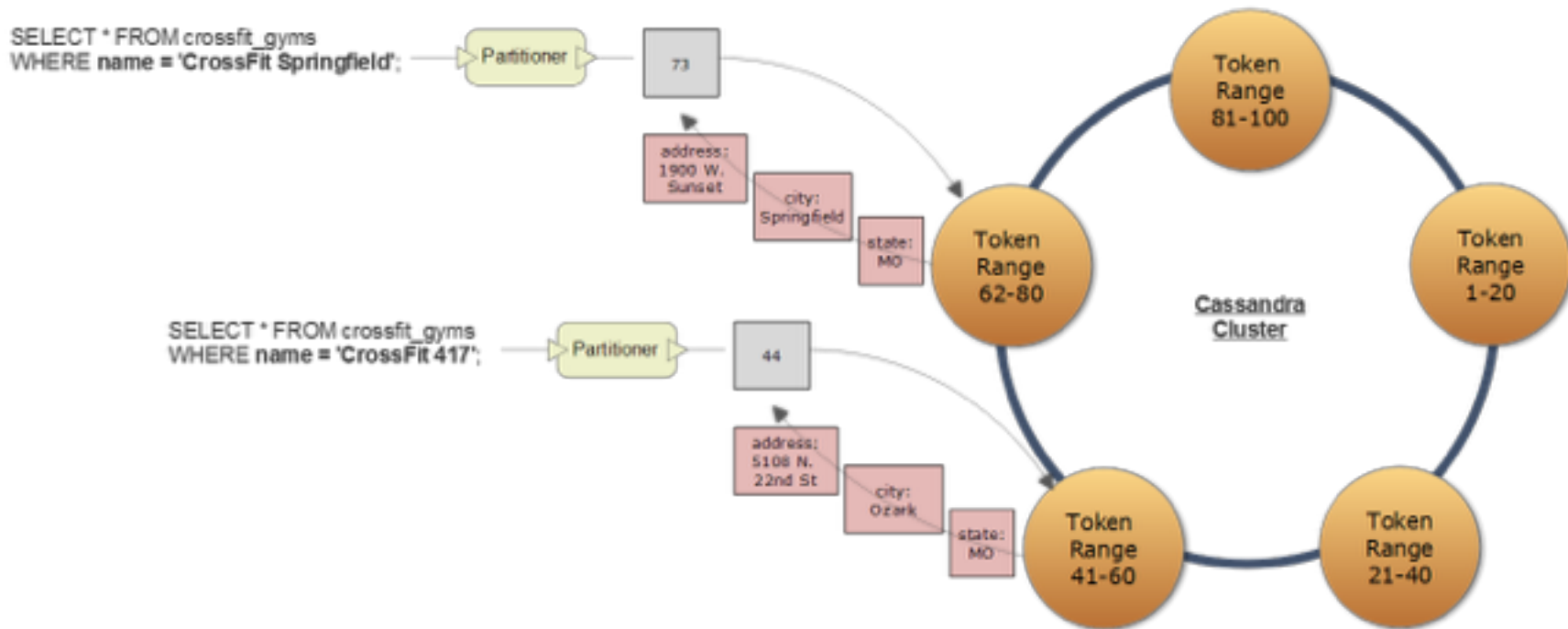
name	address	city	state
CrossFit Springfield	1900 W. Sunset	Springfield	MO
CrossFit 417	5108 N. 22nd St	Ozark	MO
CrossFit Provision	425 W. Plainview	Springfield	MO
CrossFit RTB	4234 S. Reed	Springfield	MO
Royal Barbell CrossFit	506 S. Jefferson	Springfield	MO

name: CrossFit Springfield	address: 1900 W. Sunset	city: Springfield	state: MO
name: CrossFit 417	address: 5108 N. 22nd St	city: Ozark	state: MO
name: CrossFit Provision	address: 425 W. Plainview	city: Springfield	state: MO
name: CrossFit RTB	address: 4234 S. Reed	city: Springfield	state: MO
name: Royal Barbell CrossFit	address: 506 S. Jefferson	city: Springfield	state: MO

Partition Key Explained



Partition Key Explained





Composite Partition Key



Composite Partition Key Explained

Using multiple columns for the token hash value.

```
CREATE TABLE crossfit_gyms
(
  name text,
  address text,
  city text,
  state text,
  PRIMARY KEY ((name, state))
);
```

Composite Partition Key

~~SELECT * FROM crossfit_gyms
WHERE name = 'CrossFit Springfield';~~

SELECT * FROM crossfit_gyms
WHERE name = 'CrossFit Springfield'
AND state = 'MO';

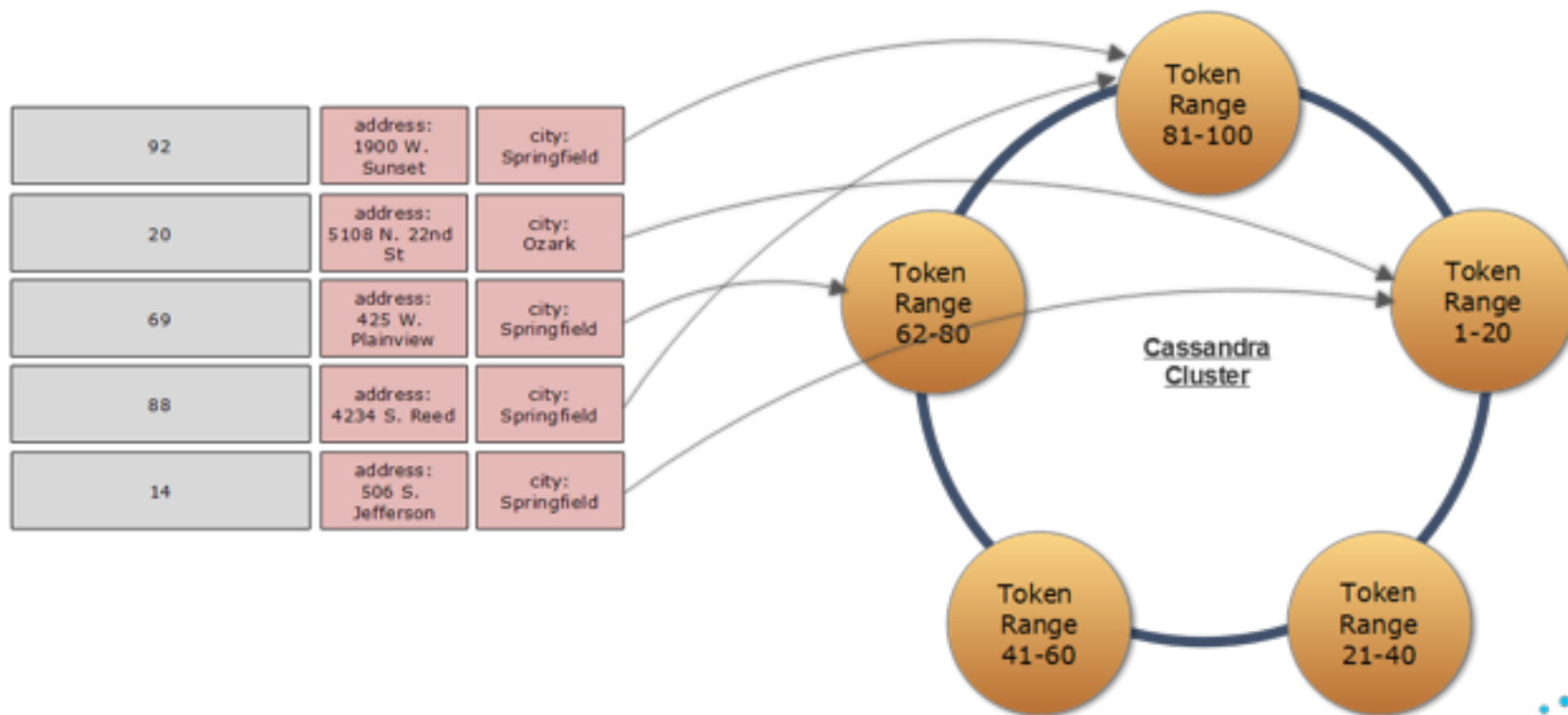
Rows

Partition Key

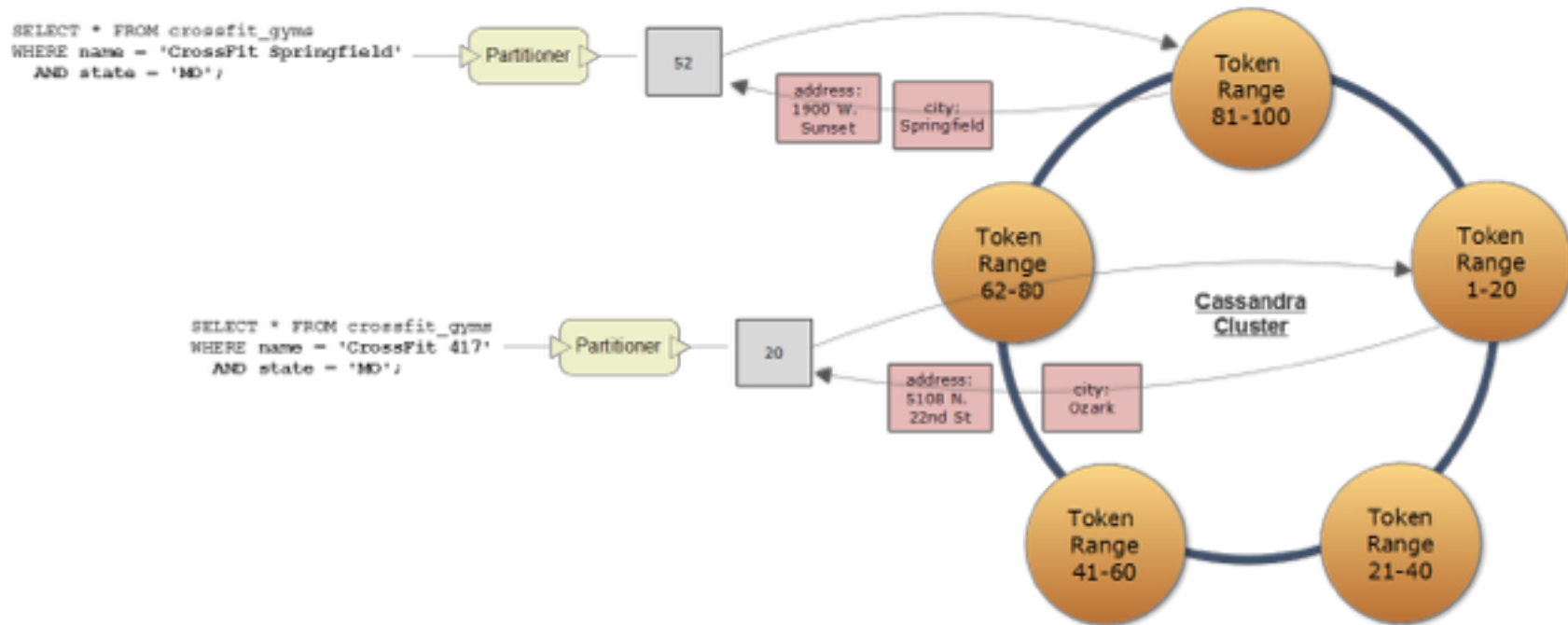
Columns

name: CrossFit Springfield	state: MO	address: 1900 W. Sunset	city: Springfield
name: CrossFit 417	state: MO	address: 5108 N. 22nd St	city: Ozark
name: CrossFit Provision	state: MO	address: 425 W. Plainview	city: Springfield
name: CrossFit RTB	state: MO	address: 4234 S. Reed	city: Springfield
name: Royal Barbell CrossFit	state: MO	address: 506 S. Jefferson	city: Springfield

Composite Partition Key Explained



Composite Partition Key Explained





Clustering Columns



Clustering Columns Explained

Clustering Columns are:

- responsible for sorting within the partition
- any column added to the Primary Key, past the first column

crossfit_gyms_by_location			
state	city	name	address
MO	Ozark	CrossFit 417	5108 N. 22nd St
MO	Springfield	CrossFit Provision	425 W. Plainview
MO	Springfield	CrossFit RTB	4234 S. Reed
MO	Springfield	CrossFit Springfield	1900 W. Sunset
MO	Springfield	Royal Barbell CrossFit	506 S. Jefferson

state: MO	city: Ozark name: Springfield 417 address: 5108 N. 22nd St	city: Springfield name: CrossFit Provision address: 425 W. Plainview	city: Springfield name: CrossFit RTB address: 4234 S. Reed	city: Springfield name: CrossFit Springfield address: 1900 W. Sunset	city: Springfield name: Royal Barbell CrossFit address: 506 S. Jefferson
--------------	--	--	--	--	--

Clustering Columns Explained

Can be used for Hierarchical structured data.

PRIMARY KEY (country, state, city)

WHERE country = ? AND state = ? AND city = ? **YES**

WHERE country = ? AND state = ? **YES**

WHERE country = ? **YES**

WHERE country = ? AND city = ? **NO**

WHERE state = ? AND city = ? **NO**

Clustering Columns Explained

Can be used for Time Series structured data.

```
CREATE TABLE member_log
(
  member text,
  workout_date timestamp,
  workout_duration text,
  PRIMARY KEY (member, workout_date)
) WITH CLUSTERING ORDER BY (workout_date DESC);
```

Member1	workout_date: 2016-09-02 07:00:00 workout_duration: 2 hr	workout_date: 2016-09-01 07:00:00 workout_duration: 2 hr	workout_date: 2016-08-31 07:00:00 workout_duration: 2 hr
Member2	workout_date: 2016-06-13 16:00:00 workout_duration: 2 hr		
Member3	workout_date: 2016-08-23 05:00:00 workout_duration: 1.75 hr	workout_date: 2016-08-21 07:00:00 workout_duration: 1.5 hr	

Clustering Columns Explained

```
CREATE TABLE crossfit_gyms_by_locations
(
  state text,
  city text,
  name text,
  address text,
  phone text,
  PRIMARY KEY (state, city, name)
);
```

state: MO	city: Ozark name: Springfield 417 address: 5108 N. 22nd St	city: Springfield name: CrossFit Provision address: 425 W. Plainview	city: Springfield name: CrossFit RTB address: 4234 S. Reed	city: Springfield name: CrossFit Springfield address: 1900 W. Sunset	city: Springfield name: Royal Barbell CrossFit address: 506 S. Jefferson
--------------	--	--	--	--	--

```
UPDATE crossfit_gyms_by_locations SET phone = '417-848-7791'
WHERE state = 'MO' AND city = 'Springfield' AND name = 'CrossFit Springfield'
```

state: MO	city: Ozark name: Springfield 417 address: 5108 N. 22nd St	city: Springfield name: CrossFit Provision address: 425 W. Plainview	city: Springfield name: CrossFit RTB address: 4234 S. Reed	city: Springfield name: CrossFit Springfield address: 1900 W. Sunset	city: Springfield name: CrossFit Springfield phone: 417-848-7791	city: Springfield name: Royal Barbell CrossFit address: 506 S. Jefferson
--------------	--	--	--	--	--	--



Thank You!
Questions?

Adam Hutson

adam@datascale.io

[@AdamHutson](https://twitter.com/AdamHutson)

[@DataScaleInc](https://twitter.com/DataScaleInc)

