



Architecture and Administration Basics

Workshop Day 2 - FTS (Full Text Search)

Full Text Search



- Very Flexible, but more concepts to learn initially
- Inverted Index
- Similar to Elasticsearch

Couchbase users need to search their documents



elastic





Full Text Search

SEARCH



Full Text Search

best hotel location

SEARCH

Search results

Scoring	Document ID	Description Matches
1.25	hotel_1234	best <u>location</u>
1.37	hotel_2345	loved <u>hotel location</u>
1.82	hotel_3456	<u>location</u> is awesome
1.88	hotel_4557	hard to <u>locate</u>



Index, Analyze, Query

- 1. Index fields of a document
"...located in the heart of the new City Quay development. The hotel has views of ..."
- 2. Analyze terms for index
located ... heart .. new City Quay development. .. hotel has views
- 3. Query the index
description: *location*

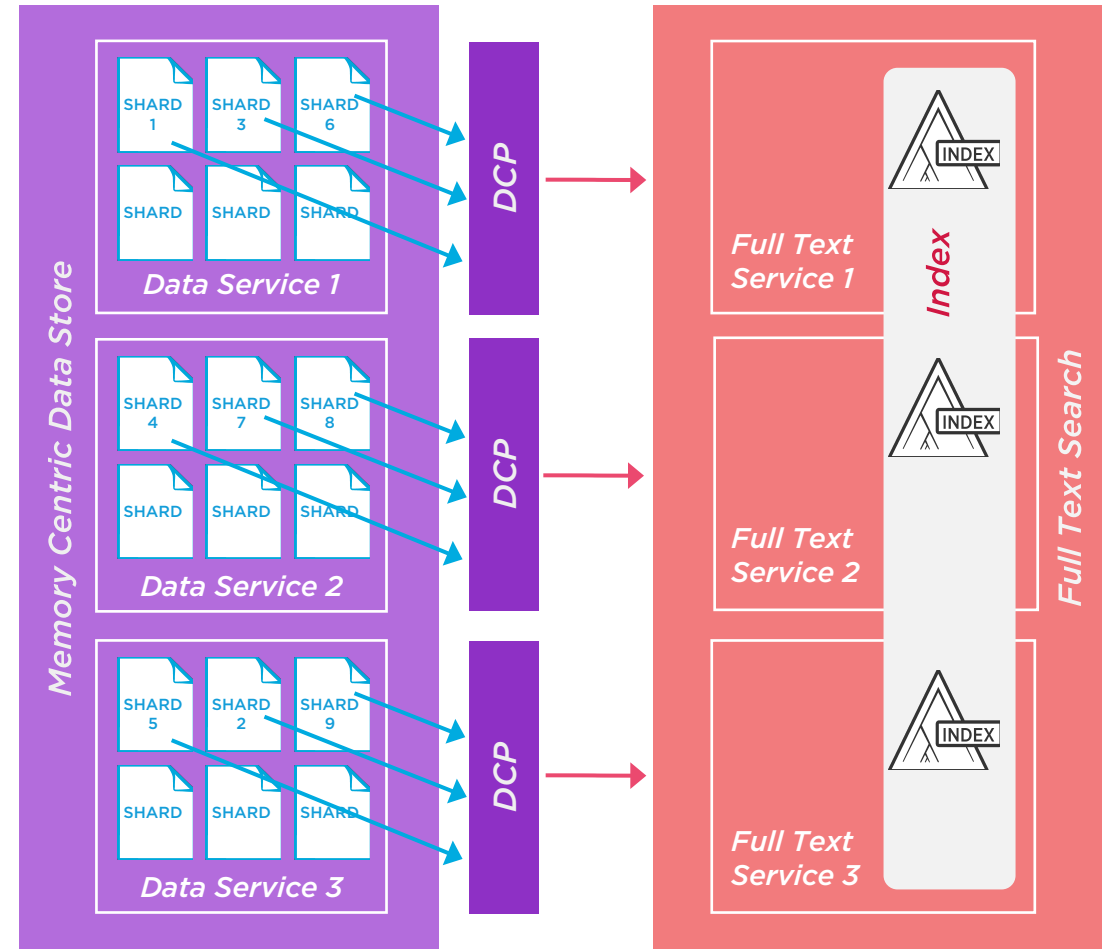
Return scored documents list

Scoring	Document ID	Description Matches
1.25	hotel_1234	best <u>location</u>
1.37	hotel_2345	loved <u>hotel location</u>
1.82	hotel_3456	<u>location</u> is awesome
1.88	hotel_4557	hard to <u>locate</u>



Full-Text Data Projection

- Each FTS node is a DCP stream subscriber
- Indexing work is distributed across FTS nodes
- Any FTS service can receive FTS query, “scatters” to other nodes and “gathers” response
- Application sees a single logical Index



Full-Text Search - setup



Couchbase Enterprise Edition

[Documentation](#) • [Support](#) • [About](#) • [Sign Out](#)

[Overview](#)

[Server Nodes](#)

[Data Buckets](#)

[Query](#)

[Indexes](#)

[XDCR](#)

[Security](#)

[Log](#)

[Settings](#)

[Global Indexes](#)

[Views](#)

[Full Text](#)

ix-travel

Index Name ix-travel

Bucket

travel-sample

Type Mappings

Add Type Mapping

hotel | only index specified fields

description | text | index | store | include in *_all* field | include term vectors

name | text | index | store | include in *_all* field | include term vectors

{ } **reviews** | only index specified fields

content | text | index | store | include in *_all* field | include term vectors

default | disabled

Analyzers

Custom Filters

Advanced

☐ Show advanced settings

Cancel

Update Index

Choose the fields to be included in the search

edit

insert child mapping

insert child field

Full-Text Search - search results



Couchbase Console (4.5.0) | localhost:8091/ui/index.html#/fts_search/hotels/_search?query=best%20western&q=best%20western&p=1

Couchbase | Classic UI • Documentation • Support • About • Sign Out

Overview | Server Nodes | Data Buckets | Query | **Indexes** | XDCR | Security | Log | Settings

Global Indexes | Views | Full Text

Full Text Search: hotels

best western ☐ Advanced
[full text query syntax help](#)

473 results (17ms server-side)

hotel_26019

Description
Chain motel with a few more amenities than the typical **Best Western**, outdoor swimming pool, internet access, cafe on-site, pet friendly.
Liked publicly by...
Tristin Hickie, Miss Luigi Kessler, Marcia Shields, Webster Hills, Gussie Kunde, Aileen Wolf, Tamara Terry

hotel_27821

Description
...would be ideal. The more affordable mid-range option compared to it's neighbouring **Best Western**. They offer a pool and hot tub, 24-front desk services, and a complimentary continental breakfast. Know ...
Liked publicly by...
Matilde Heller, Mrs. Reanna Gutmann, Thelma Rippin, Santa Gutkowski, Marion Olson, Catherine Lindgren, Sheld...nger, Elizabeth Gutmann

hotel_27819

Description
... **Best Western**. A few blocks from Sonoma Plaza, this hotel is pet friendly and has a pool, jacuzzi rooms and free breakfast. Free wifi and...
Liked publicly by...
Wilbert Wolf DDS, Gilles Skiles II, Maria Botsford, Jovanny Adams, Salma Klehn, Lera Mayer

hotel_21859

Description
... exit Hotel **Best Western** Opéra d'Antin and leave the large department stores behind you (Galeries Lafayette & Le Printemps), just cross boulevard...
Liked publicly by...
Eldred Graham, Declan Fisher, Jaily Witting, Erick Rogahn, Jackson Dickinson, Margarete Aufderhar, Elroy Koch

Result Text Snippets

Highlighted Search Terms

How does it work?



- Inverted indexes
- Language awareness
- Scoring

Inverted index

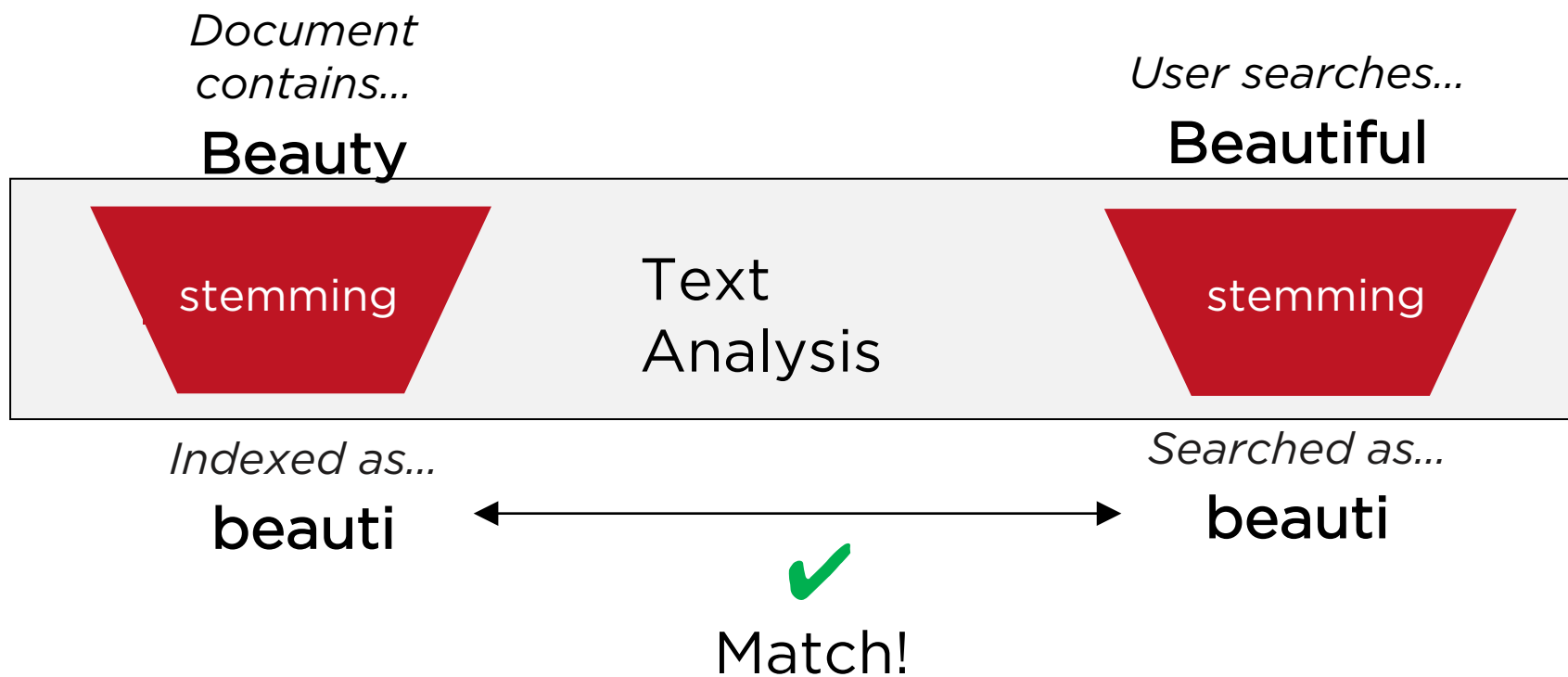


Terms Where found



my: Doc 1, Doc 2, Doc 3
dog: Doc 1, Doc 2, Doc 81
has: Doc 1, Doc 2, Doc 3
fleas: Doc 1, Doc 81
...

language aware



Scoring



The screenshot shows the Couchbase Console interface. The browser address bar displays `localhost:8091/ui/index.html#/fts_search/hotels/_search?query=best%20western&q=best%20western&p=1`. The Couchbase logo is in the top left, and navigation links (Classic UI, Documentation, Support, About, Sign Out) are in the top right. A horizontal menu contains links for Overview, Server Nodes, Data Buckets, Query, Indexes (selected), XDCR, Security, Log, and Settings. Below this, sub-links for Global Indexes, Views, and Full Text are visible. The main section is titled 'Full Text Search: hotels'. A search input field contains 'best western', followed by a 'Search' button and an unchecked 'Advanced' checkbox. A link for 'full text query syntax help' is below the input. A callout box on the left points to the results count, displaying '473 results'. The results are listed as '473 results (17ms server-side)'. A 'Show Scoring' checkbox is on the right. The first result, 'hotel_26019', shows a description mentioning 'Best Western', a list of amenities, and a list of users who liked it. The second result, 'hotel_27821', also mentions 'Best Western' and lists amenities and users. The third result, 'hotel_27819', mentions 'Best Western' and lists amenities and users. The fourth result, 'hotel_21859', mentions 'Best Western' and lists amenities and users.

473 results

Full Text Search: hotels

best western ☐ Advanced

[full text query syntax help](#)

473 results (17ms server-side) ☐ Show Scoring

hotel_26019

Description
Chain motel with a few more amenities than the typical **Best Western**; outdoor swimming pool, internet access, cafe on-site, pet friendly.

Liked publicly by...
Tristin Hickie, Miss Luigi Kessler, Marcia Shields, Webster Hills, Gussie Kunde, Aileen Wolf, Tamara Terry

hotel_27821

Description
...would be ideal. The more affordable mid-range option compared to it's neighbouring **Best Western**. They offer a pool and hot tub, 24-front desk services, and a complimentary continental breakfast. Know ...

Liked publicly by...
Matilde Heller, Mrs. Reanna Gutmann, Thelma Rippin, Santa Gutkowski, Marion Olson, Catherine Lindgren, Sheldon Renner, Elizabeth Gutmann

hotel_27819

Description
... **Best Western**. A few blocks from Sonoma Plaza, this hotel is pet friendly and has a pool, jacuzzi rooms and free breakfast. Free wifi and laundry facilities you can use. Expensive for a **Best Western** ...

Liked publicly by...
Wilbert Wolf DDS, Gilles Skiles II, Maria Botsford, Jovanny Adams, Salma Klehn, Lera Mayer

hotel_21859

Description
... exit Hotel **Best Western** Opéra d'Antin and leave the large department stores behind you (Galeries Lafayette & Le Printemps), just cross boulevard Haussmann and you will find yourself before the Opera ...

Liked publicly by...
Eldred Graham, Declan Fisher, Jaily Witting, Erick Rogahn, Jackson Dickinson, Margarete Aufderhar, Elroy Koch



TF/IDF scoring

- **TF = Term Frequency**
 - How often does a term occur in a document?
 - More often yields a higher score
- **IDF = Inverse Document Frequency**
 - How many documents have this term?
 - More documents yields lower score
 - (because it means the term is more common)

Index mapping



▼ Type Mappings

default

repositories | text

status | text | index | store | include in_*all* field | include term vectors

createdAt | text | index | store | include in_*all* field | include term vectors

title | text | index | store | include in_*all* field | include term vectors

assignedId | text | index | store | include in_*all* field | include term vectors

id | text | index | store | include in_*all* field | include term vectors

type | text | index | store | include in_*all* field | include term vectors

{} **developerInfo** | only index specified fields

email | text | index | store | include in_*all* field | include term vectors

lastName | text | index | store | include in_*all* field | include term vectors

firstName | text | index | store | include in_*all* field | include term vectors



Index mapping

- Exclude fields/sub-sections
- Configure indexing behavior by type of document (beer vs brewery)
- Configure indexing behavior per-field
 - Index Fields
 - Nested structures
 - Arrays



Index mapping

Before you search – you have to tell Couchbase how to index your docs...

```
{
```

```
  "json": "You know it!",
```



```
  "some_junk": "ignore me"
```



```
}
```

Type mapping



```
{  
  "name": "21st Amendment Brewery Cafe",  
  "type": "brewery",  
  ...  
}
```

Default type = no type or no type mapping for type

Simple types



Index as “fields”

```
{  
  “type” : “brewery”,  
  “name” : “21st Ammendment”,  
  “random_number” : 4,  
  “edible” : false  
}
```

Nested values



Use a **child mapping**,
then index fields

```
{  
  "type" : "brewery",  
  "geo": {  
    "accuracy": "ROOFTOP",  
    "lat": 37.7825,  
    "lon": -122.393  
  }  
}
```

Arrays



- Array of simple types

Add a field for
“pachyderms”

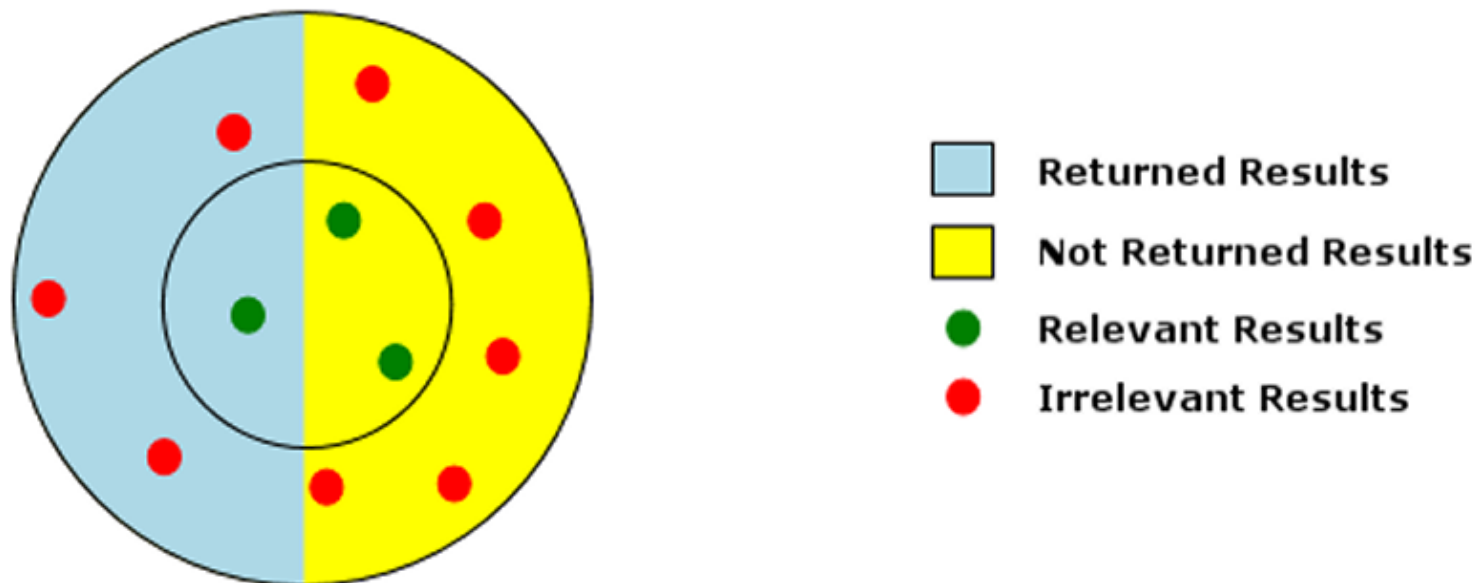
```
{  
  “pachyderms”: [  
    “elephant”, “rhinoceros”, “hippopotamus”  
  ]  
}
```

- Array of Objects

Use a child mapping
for “alphabet”

```
{  
  “alphabet”: [  
    {“letter”:“b”},{“letter”:“c”},{“letter”:“d”}  
  ]  
}
```

precision vs. recall



- Precision – ratio of document matches that are actually relevant
- Recall – ratio of relevant documents that are actually matched
- High quality results depend on performing the right analysis for your text
- Beware: increasing precision may reduce recall (and vice versa)



How does it scale?

- ✓ auto index partitioning (hash partitioning)
- ✓ to multiple FTS nodes (auto-placement)
- ✓ rebalance (add/swap/remove)
- ✓ scatter/gather queries (partial results ok)
- ✓ replicas (only primaries queried)
- ✓ failover (replicas promoted)

FTS design / index partitioning

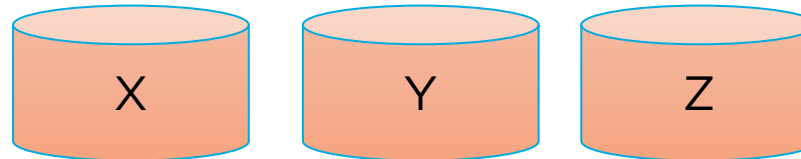


bucket partitions:
(1024 vbuckets)

0, 1, 2, 3, 4, ...

... ,1021, 1022, 1023

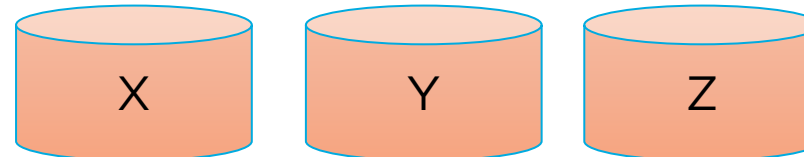
FTS nodes:





... ,1021, 1022, 1023

FTS nodes:



FTS design / index partitioning

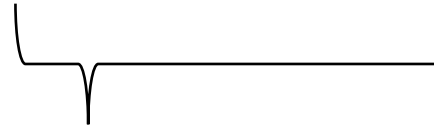


bucket partitions:
(1024 vbuckets)

0, 1, 2, 3, 4, ...

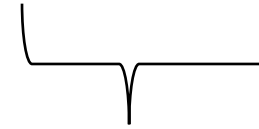
... ,1021, 1022, 1023

index partitions:
(groups of vbuckets)



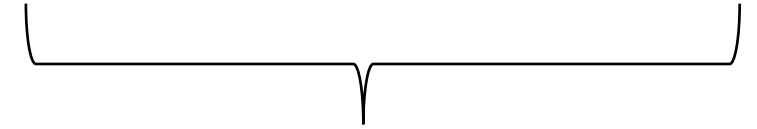
A

0-399



B

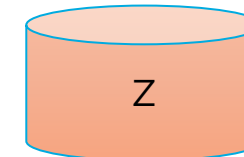
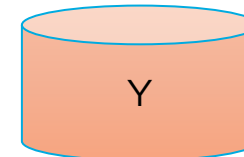
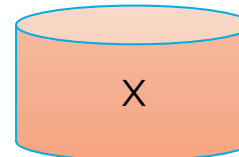
400-799



C

800-1023

FTS nodes:



FTS design / index partitioning

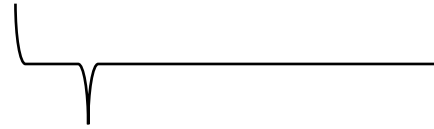


bucket partitions:
(1024 vbuckets)

0, 1, 2, 3, 4, ...

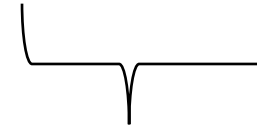
... ,1021, 1022, 1023

index partitions:
(groups of vbuckets)



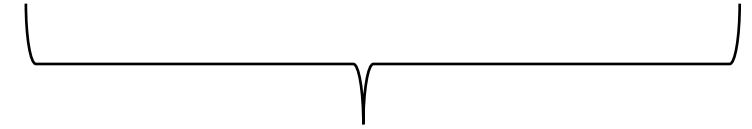
A

0-399



B

400-799

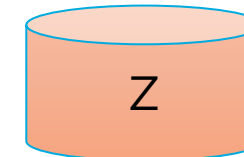
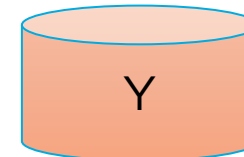
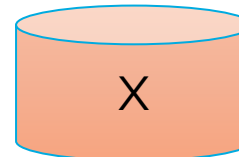


C

800-1023

assign to FTS nodes:

FTS nodes:



FTS design / index partitioning



bucket partitions:
(1024 vbuckets)

0, 1, 2, 3, 4, ...

... ,1021, 1022, 1023

index partitions:
(groups of vbuckets)

A

B

C

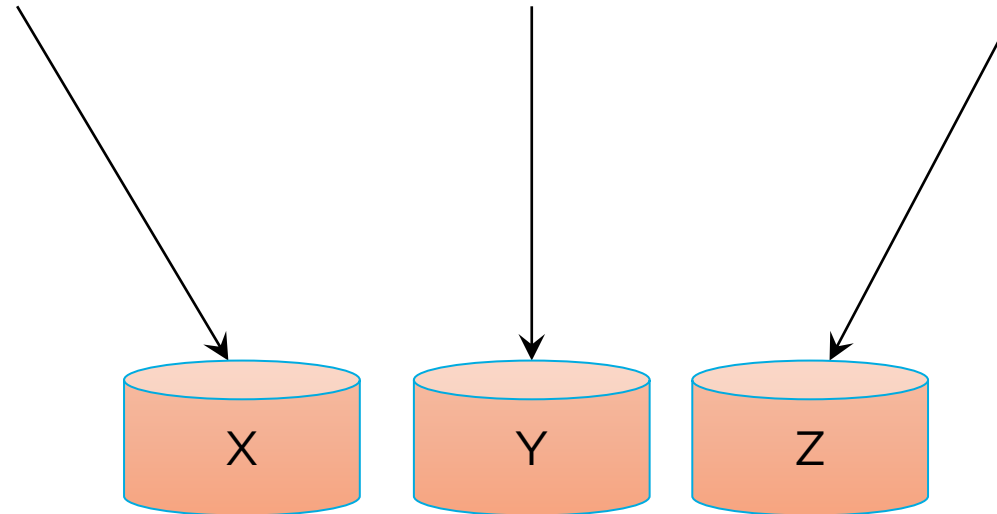
0-399

400-799

800-1023

assign to FTS nodes:

FTS nodes:



FTS design / index partitioning



bucket partitions:
(1024 vbuckets)

0, 1, 2, 3, 4, ...

... ,1021, 1022, 1023

index partitions:
(groups of vbuckets)

A

B

C

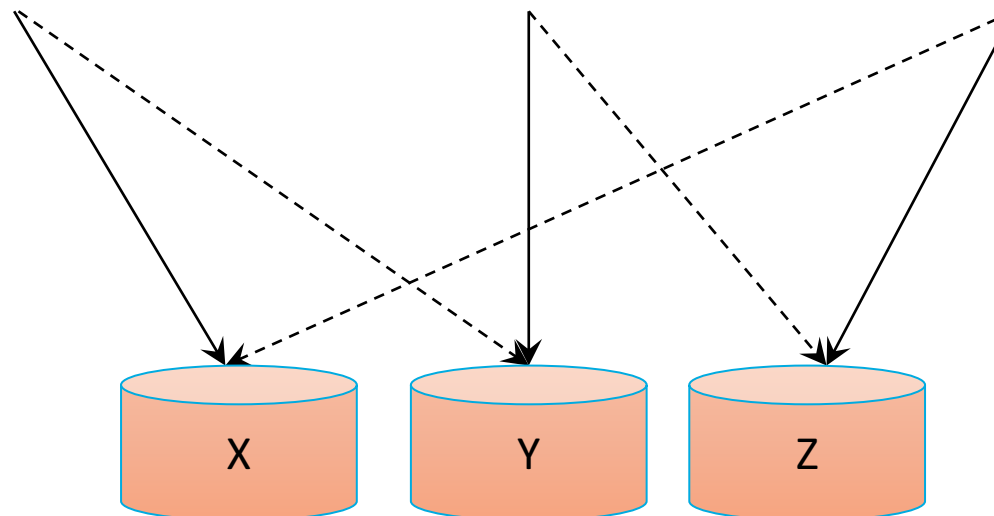
0-399

400-799

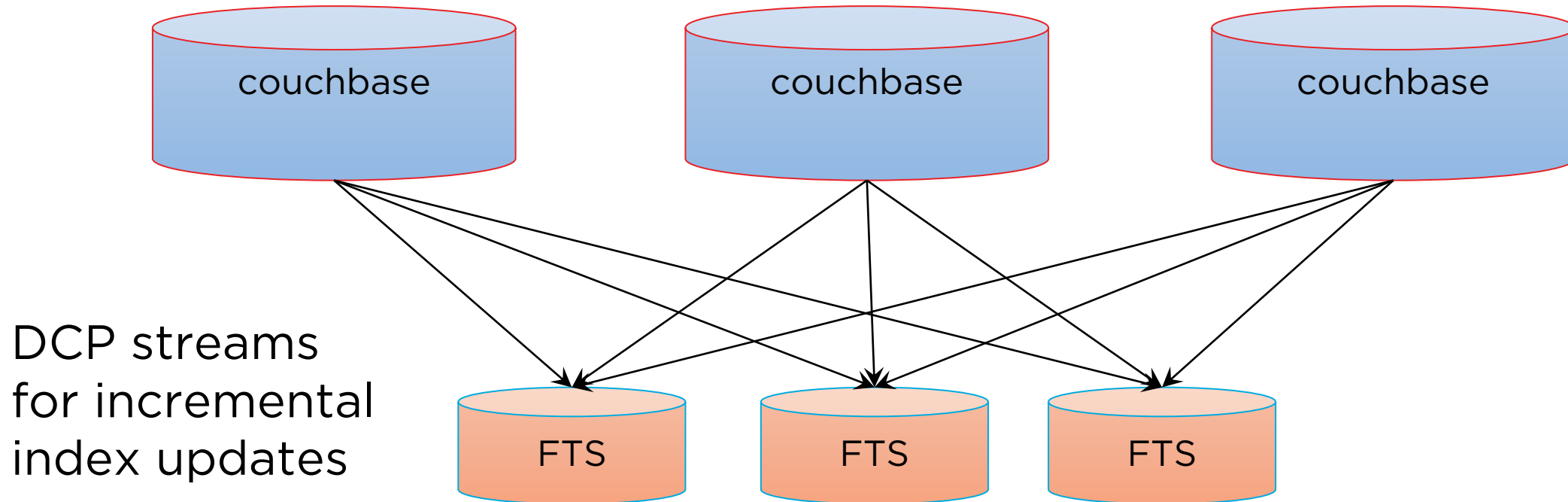
800-1023

assign to FTS nodes:
replicas, too:

FTS nodes:



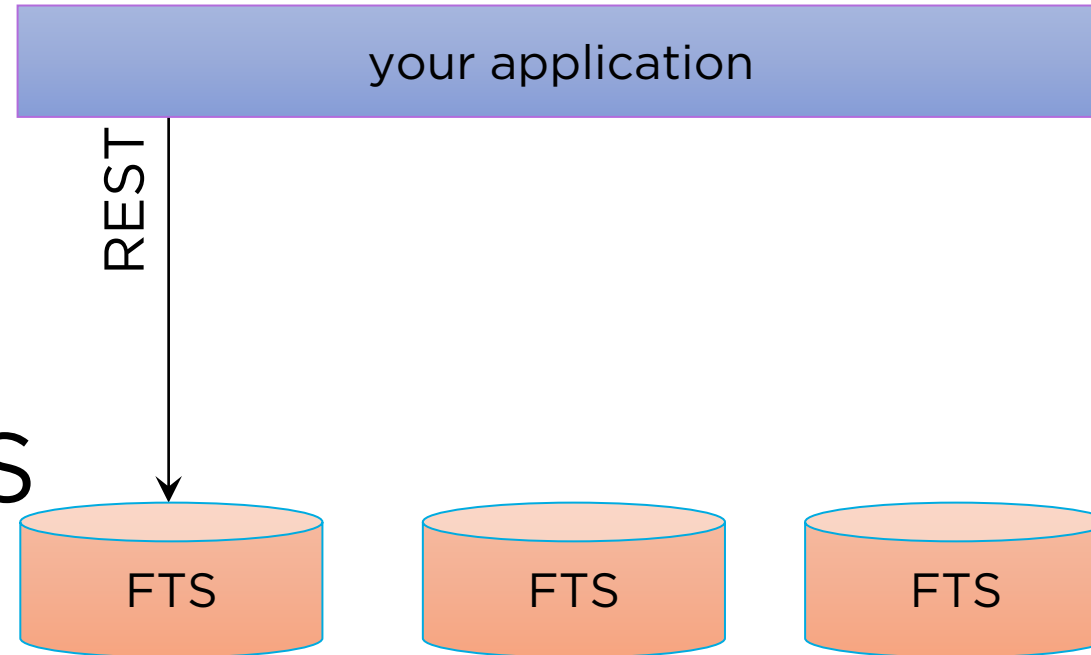
FTS design / indexing



FTS design / queries



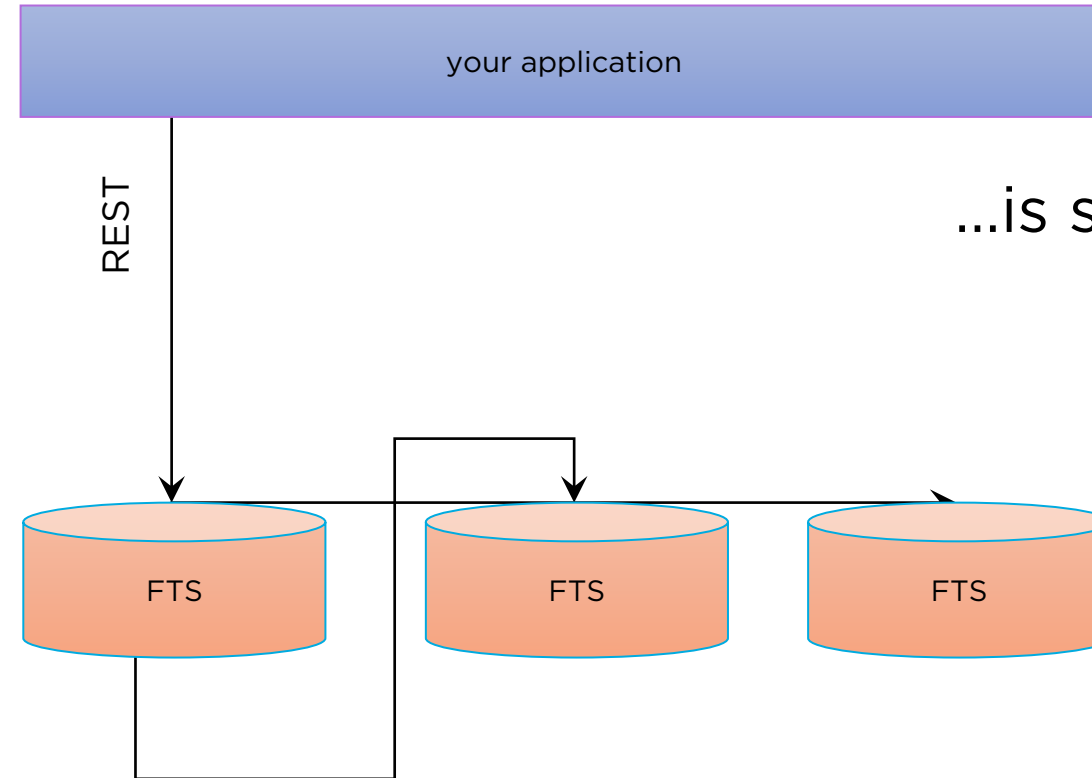
a query
sent
to any FTS
node...



FTS design / queries



a query sent
to any FTS
node...



...is scatter / gathered
to the other
FTS nodes

Thank you

