



**Couchbase**



# Full Text Search Integration

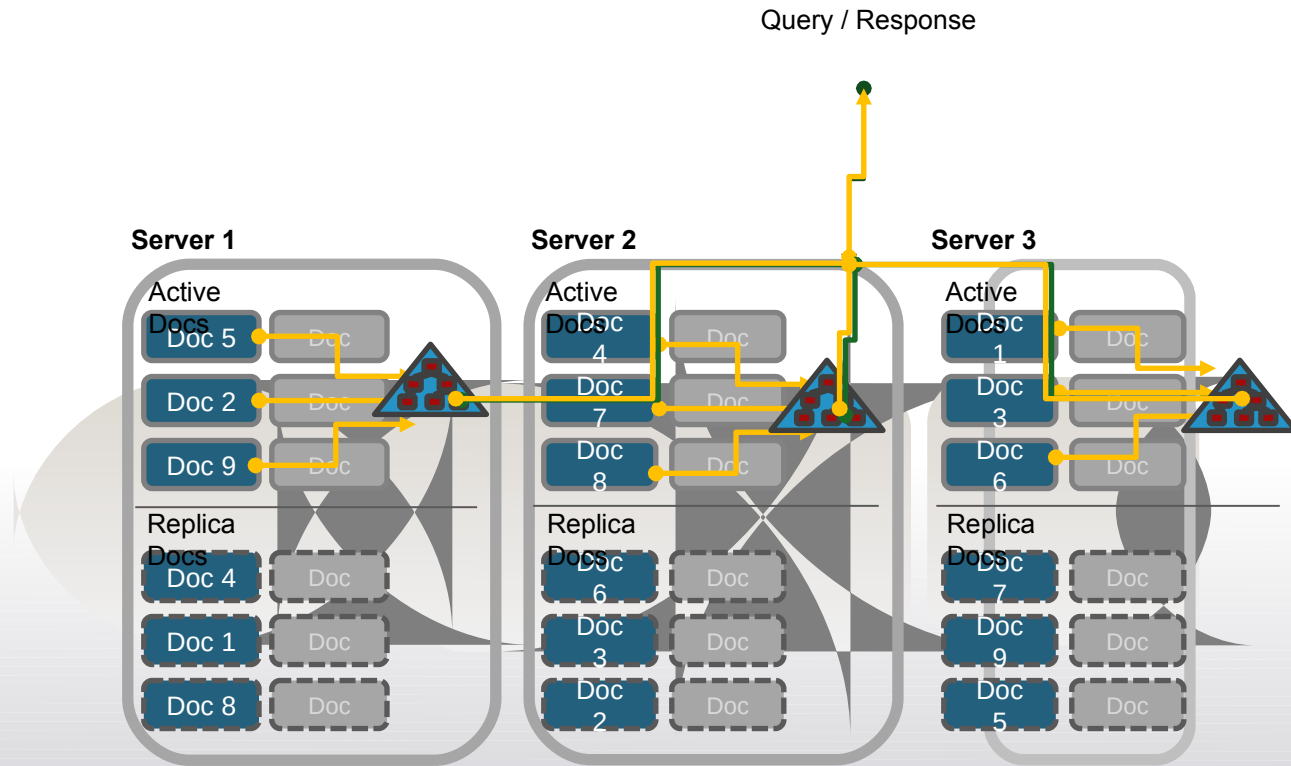
**Tugdual Grall**  
**Technical Evangelist**



**Couchbase**



# Distributed Indexing and Querying Using Incremental Map Reduce



# Search Across Full JSON Body

```
{
  "name": "Abbey Belgian Style Ale",
  "description": "Winner of four World Beer
Cup
                medals and eight medals at
the
                Great American Beer Fest,
Abbey
                Belgian Ale is the Mark
Spitz
                of New Belgium's lineup –
but
                it didn't start out that
way."
}
```

Search term: abbey

# Search Across Full JSON Body

```
{  
  "name": "Abbey Belgian Style Ale",  
  "description": "Winner of four World Beer  
Cup  
the  
Abbey  
Spitz  
but  
way."  
}
```

Search term: abbey

# Integrate with Elasticsearch for Full Text Search

- Based on proven Apache Lucene technology
- Apache 2 Licensed with commercial support available
- Distributed
- Schema Free JSON Documents
- RESTful API



# ElasticSearch Terminology

- **Document**

- Schema-less JSON...
- Contains a set of fields

- **Type**

- Contains a set of mappings describing how fields are indexed

- **Index**

- Logical namespace for scoping indexing/searching
- May contain documents of different types
- Uniqueness by ID/Type



# How does it work?

- Unidirectional Cross Data Center Replication



**Couchbase**



**ElasticSearch**  
ch



# Getting Started



# Install the Couchbase Plug-In

- **Pre-requisite**

- Existing Couchbase and ElasticSearch Clusters

- **Install the ElasticSearch Couchbase Transport Plug-in**

- bin/plugin -install

`couchbaselabs/elasticsearch-transport-couchbase/1.0.0-beta`

- **Configure the Plug-in**

- Set a password
- Install the Couchbase Index Template

- **Restart ElasticSearch**

# Configure XDCR (part 1)

The screenshot shows the Couchbase XDCR configuration interface. A modal dialog box titled "Create Cluster Reference" is open, allowing the user to configure a new cluster reference. The dialog contains the following fields and options:

- Cluster Name:** ElasticSearch
- IP/hostname:** 127.0.0.1:9091 [What's this?](#)
- Security:** [What's this?](#)
- Username:** Administrator
- Password:** (masked with dots)

At the bottom of the dialog, there are "Cancel" and "Save" buttons. The background interface shows sections for "REMOTE CLUSTERS", "ONGOING REPLICATIONS", and "PAST REPLICATIONS".

# Configure XDCR (part 2)

Replications

▼ REMOTE CLUSTERS Create Cluster Reference

Name	IP/hostname	
ElasticSearch	127.0.0.1:9091	Delete Edit

ONGOING REPLICATION

Bucket	From	To	Status	When
beer-sample	this cluster	ES	Cancelled	on change 16:40:38 - Wed Sep 12, 2012

▼ PAST REPLICATION

Bucket From

beer-sample this cluster ES Cancelled on change 16:40:38 - Wed Sep 12, 2012

### Create Replication

Replicate changes from: To:

Cluster: **this cluster** Cluster: ElasticSearch

Bucket: beer-sample Bucket: indexName

Cancel Replicate

# Documents are now being indexed

Overview | Browser | Structured Query [+] | API Request [+]

**Cluster Overview** [New Index](#)

**Indexer-sample**  
Size: 3.8mb (3.8mb)  
Docs: 7479 (7479)  
[Info](#) [Actions](#)

**Azazel** PcZHKV\_2TyysdjkNUxencw  
inet[/192.168.1.8:9200]  
[Info](#) [Actions](#)

**Unassigned**

0	1	2	3	4
0	1	2	3	4

Document  
Count  
Increasing



---

**What Now?**



# Document from Beer Sample Dataset

```
{  
  "name": "Pabst Blue Ribbon",  
  "abv": 4.74,  
  "ibu": 0,  
  "srm": 0,  
  "upc": 0,  
  "type": "beer",  
  "brewery_id": "110f1d5dc2",  
  "updated": "2010-07-22 20:00:20",  
  "description": "PBR is not just any  
beer...",  
  "style": "American-Style Light Lager",  
  "category": "North American Lager"  
}
```



# Sample ES Query with HTTP

- Search for any beer matching the term “lager”
  - **GET** `http://127.0.0.1:9200/beer-sample/_search?q=lager`

```
{  
  "took": 7,  
  "timed_out": false,  
  "_shards": { ... },  
  "hits": {  
    "total": 1271,  
    "max_score": 1.1145955,  
    "hits": [...]  
  }  
}
```



# Sample ES Query with HTTP

- Search for any beer matching the term “lager”
  - **GET** `http://127.0.0.1:9200/beer-sample/_search?q=lager`

```
{  
  "took": 7,  
  "timed_out": false,  
  "_shards": { ... },  
  "hits": {  
    "total": 1271,  
    "max_score": 1.1145955,  
    "hits": [...]  
  }  
}
```



Total Search  
Execution Time

# Sample ES Query with HTTP

- Search for any beer matching the term “lager”
  - **GET** `http://127.0.0.1:9200/beer-sample/_search?q=lager`

```
{  
  "took": 7,  
  "timed_out": false,  
  "_shards": { ... },  
  "hits": {  
    "total": 1271,  
    "max_score": 1.1145955,  
    "hits": [...]  
  }  
}
```



Total Number of Documents Matching Query

# Sample ES Query with HTTP

- Search for any beer matching the term “lager”
  - **GET** `http://127.0.0.1:9200/beer-sample/_search?q=lager`

```
{
  "took": 7,
  "timed_out": false,
  "_shards": { ... },
  "hits": {
    "total": 1271,
    "max_score": 1.1111111
  }
}
```



Maximum Score of  
All Matching  
Documents

# Sample ES Query with HTTP

- Search for any beer matching the term “lager”
  - **GET** `http://127.0.0.1:9200/beer-sample/_search?q=lager`

```
{
  "took": 7,
  "timed_out": false,
  "_shards": { ... },
  "hits": {
    "total": 1271,
    "max_score": 1.1145955
    "hits": [...]
  }
}
```



Array of Matching Documents



# Single Search Result

```
"hits": [ {  
  "_index": "beer-sample",  
  "_type": "couchbaseDocument",  
  "_id": "110fc4b16b",  
  "_score": 1.1145955,  
  "_source": {  
    "meta": {  
      "id": "110fc4b16b",  
      "rev": "1-001ba0044ce30dd5000000000000000000",  
      "flags": 0,  
      "expiration": 0  
    }  
  }  
}, ... ]
```



ID of Matching Document

# Single Search Result

```
"hits": [ {  
  "_index": "beer-sample",  
  "_type": "couchbaseDocument",  
  "_id": "110fc4b16b",  
  "_score": 1.1145955,  
  "_source": {  
    "meta": {  
      "id": "110fc4b16b",  
      "rev": "1-001ba0044ce30dd50000000000000000000",  
      "flags": 0,  
      "expiration": 0  
    }  
  }  
}, ... ]
```



Where's the document body?

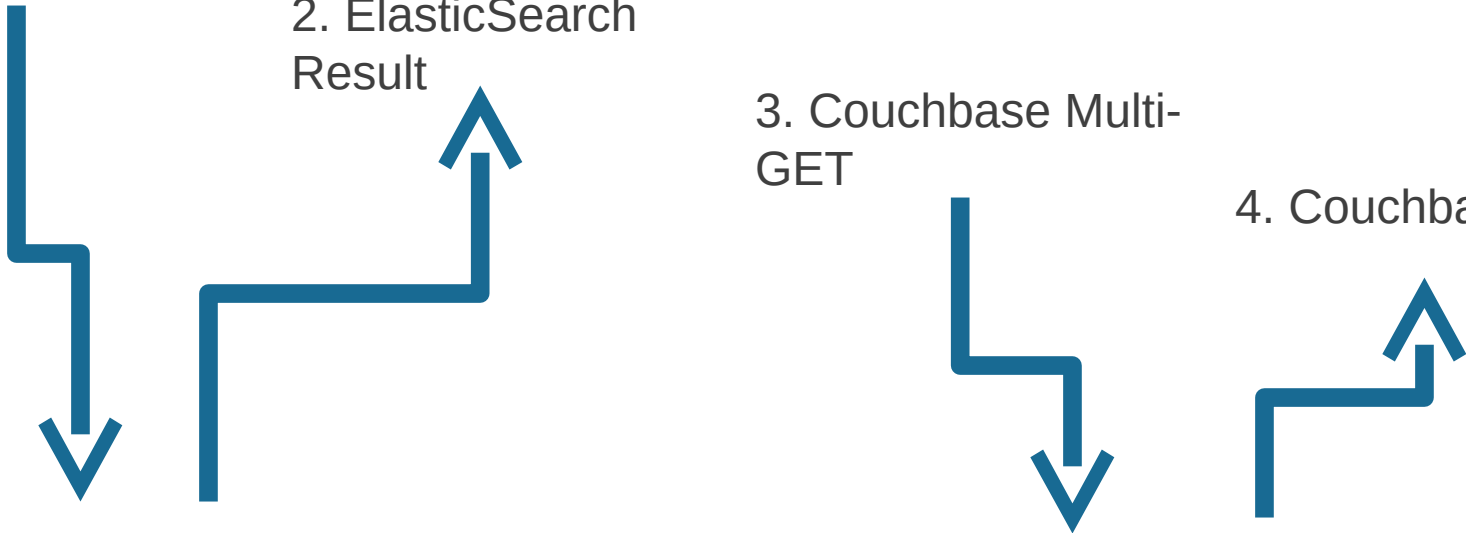
# Recommended Usage Pattern

1. Elasticsearch  
Query

2. Elasticsearch  
Result

3. Couchbase Multi-  
GET

4. Couchbase Result



ElasticSearch

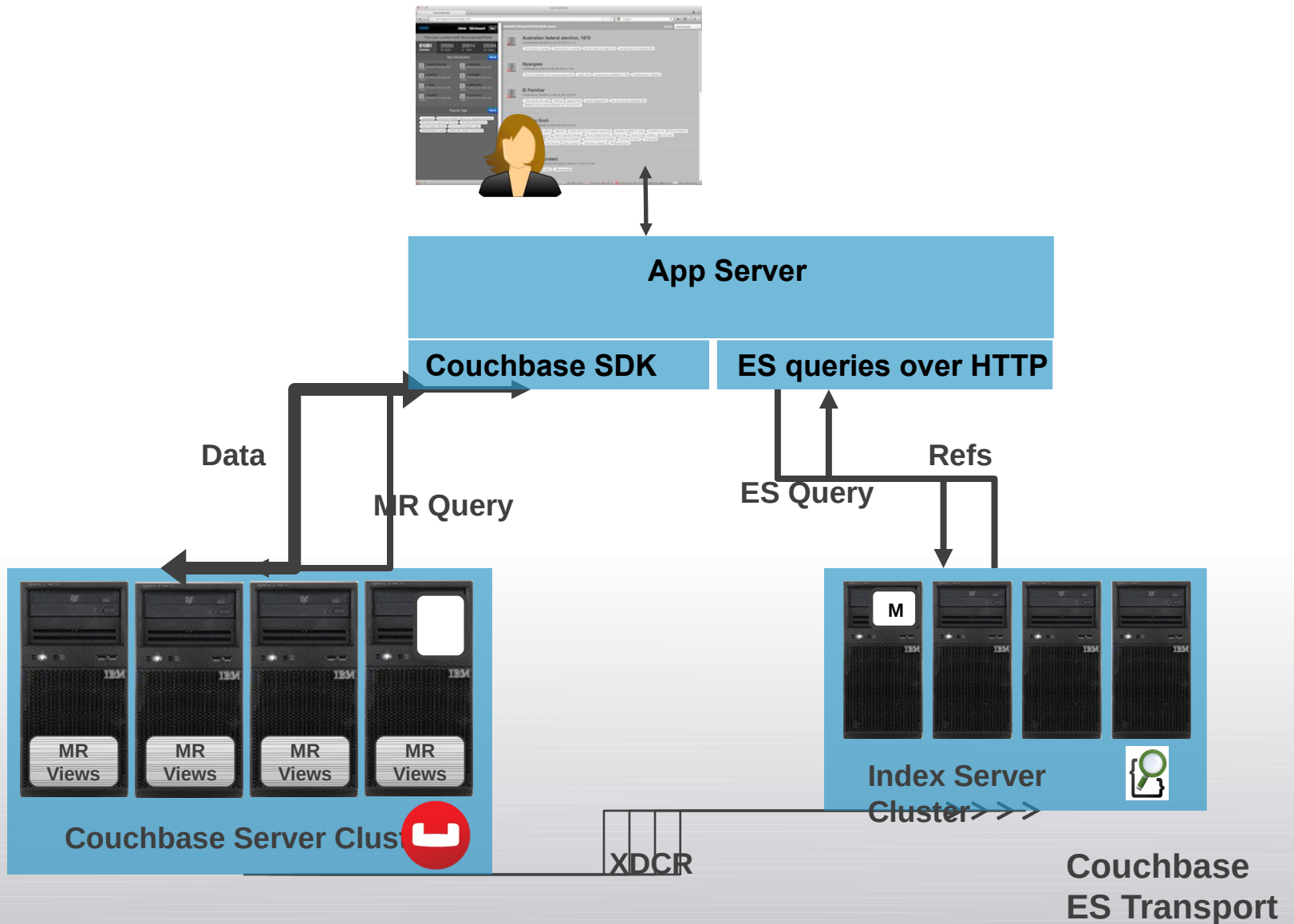


Couchbase



Couchbase

# Architecture Overview





# More Advanced Capabilities



# Another Query with HTTP

- **POST** `http://127.0.0.1:9200/default/_search`

```
{
  "query": {
    "query_string": {
      "query": "style: lambic AND description:
blueberry"
    }
  }
}
{
  "name": "Wild Blue Blueberry Lager",
  "abv": 8,
  "type": "beer",
  "brewery_id": "110f01abce",
  "updated": "2010-07-22 20:00:20",
  "description": "...ripe blueberry aroma...",
  "style": "Belgian-Style Fruit Lambic",
  "category": "Belgian and French Ale"
}
```



# Faceted Search

Categories

## Product Results

### Brand

- ASUS (1)
- Antec (1)
- Cooler Master (1)
- Corsair (2)
- Crucial Technology (1)
- Iomega (2)
- JVC (1)
- Marshall (1)
- Ocz Technology (1)
- PNY (1)

### Price

- \$20 - \$50 (3)
- \$50 - \$100 (5)
- \$100 - \$150 (8)
- \$200 - \$250 (1)
- \$250 - \$500 (1)
- \$500 - \$750 (4)
- \$750 - \$1,000 (1)
- \$3,000+ (2)

### Customer Rating

- 4 - 5 Stars (5)
- 3 - 3.9 Stars (1)

Range Facets

Items with Counts

# Faceted Search Query – Beer Style

```
{
  "query": {
    "query_string": {
      "query": "bud"
    }
  },
  "facets" : {
    "styles" : {
      "terms" : {
        "field" : "style",
        "size" : 3
      }
    }
  }
}
```

# Faceted Search Results - Incorrect

```
"terms": [  
  {  
    "term": "style"  
    "count": 8  
  }  
  {  
    "term": "lager"  
    "count": 6  
  }  
  {  
    "term": "american"  
    "count": 4  
  }  
]
```

Style was **“American-Style Lager”**

# Update the Mapping

- **PUT** /beer-sample/couchbaseDocument/\_mapping

```
{
  "couchbaseDocument": {
    "properties": {
      "doc": {
        "properties": {
          "style": {
            "type": "string",
            "index": "not_analyzed"
          }
        }
      }
    }
  }
}
```

**NOTE:** When you change the mapping you **MUST** re-index.

# Faceted Search Results – Correct

```
"terms": [  
  {  
    "term": "American-Style Light Lager",  
    "count": 5  
  },  
  {  
    "term": "American-Style Lager",  
    "count": 2  
  },  
  {  
    "term": "Belgian-Style White",  
    "count": 1  
  }  
]
```



# Faceted Search Query – % Alcohol Range

```
{
  "query": {
    "query_string": {
      "query": "bud"
    }
  },
  "facets" : {
    "abv" : {
      "range" : {
        "abv" : [
          { "to" : 3 },
          { "from" : 3, "to" : 5 },
          { "from" : 5 }
        ]
      }
    }
  }
}
```



# Faceted Search Results – % Alcohol Range

```
"ranges": [  
  {  
    "to": 3,  
    "count": 1  
  },  
  {  
    "from": 3,  
    "to": 5,  
    "count": 5  
  },  
  {  
    "from": 5,  
    "count": 3  
  }  
]
```



# Search Result Scoring

- Each matching document is assigned a score based on how well it matches the query

```
hits: [  
  {  
    "_index": "default",  
    "_type": "couchbaseDocument",  
    "_id": "35addbc374",  
    "_score": 1.1306798,  
    ...  
  }  
]
```

# Custom Scoring – Document Properties

- Each document has a numerical field “abv”
- Let’s use this field to boost the beers natural score

```
{
  "query": {
    "custom_score" : {
      "query": {
        "query_string": {
          "query": "bud"
        }
      },
      "script" : "_score * doc['abv'].value"
    }
  }
}
```

# Custom Scoring – User Preferences

- Let users could rank beer styles from 1-10
- User with no preferences set searches for “bud”

Name	Style	Score
Bud Extra		1.5409653
Bud Light Lime	American-Style Light Lager	1.513119
Bud Light Golden Wheat	Belgian-Style White	1.3208274
Bud Ice	American-Style Lager	1.2839241
Bud Ice Light	American-Style Lager	1.2839241
Bud Light	American-Style Light Lager	1.245288
Bud Dry	American-Style Light Lager	1.1968427
Budweiser Select	American-Style Light Lager	0.8559494
Miller Lite	American-Style Light Lager	0.7201389

# Custom Scoring – User Preferences

- User ranks “Belgian-Style White” with value 10

```
{
  "query": {
    "custom_filters_score" : {
      "query" : {
        "text" : { "_all": "bud" }
      },
      "filters" : [
        {
          "filter" : { "term" : { "style" : "Belgian-Style White" } },
          "boost" : "10"
        }
      ],
      "score_mode" : "first"
    }
  }
}
```

# Custom Scoring – User Preferences

Name	Style	Score
Bud Light Golden Wheat	Belgian-Style White	13.208274
Bud Extra	American-Style Light Lager	1.5409653
Bud Light Lime	American-Style Light Lager	1.513119
<del>Bud Light Golden Wheat</del>	<del>Belgian-Style White</del>	<del>1.3208274</del>
Bud Ice	American-Style Lager	1.2839241
Bud Ice Light	American-Style Lager	1.2839241
Bud Light	American-Style Light Lager	1.245288
Bud Dry	American-Style Light Lager	1.1968427
Budweiser Select	American-Style Light Lager	0.8559494
Miller Lite	American-Style Light Lager	0.7201389



# Learning Portal – Proof of Concept



# Next Steps



# Explore Elasticsearch Capabilities

- **Customize Document Mappings**
  - Default behavior isn't always what you want
  - Index one field multiple ways
- **Advanced Cluster Topologies**
  - Dedicate nodes for routing/querying
- **Rich Query DSL**

ElasticSearch Guide:  
<http://www.elasticsearch.org/guide/>

# Couchbase Elasticsearch Future

- **Release 1.0.0**
- **Possible features for future**
  - More fine-grained cluster configuration
  - More index-level configuration
  - Pre-index script execution
  - Indexing non-JSON data
- **Give us your feedback!**

# Resources



- Marty Schoch's blog:  
<http://blog.couchbase.com/couchbase-and-full->



- <https://github.com/couchbaselabs/elasticsearch>

- [tug@couchbase.com](mailto:tug@couchbase.com)



- [@tgrall](#)





**Couchbase**

