



# APP MODERNISATION DAY

09/04/2019 — MICROSOFT HOUSE, MILANO

## Chi cerca trova: come cercare meglio

Alessandro Melchiori

alessandro@codiceplastico.com

<https://github.com/melkio>

Twitter: @amelchiori

#APPMODERNISATION

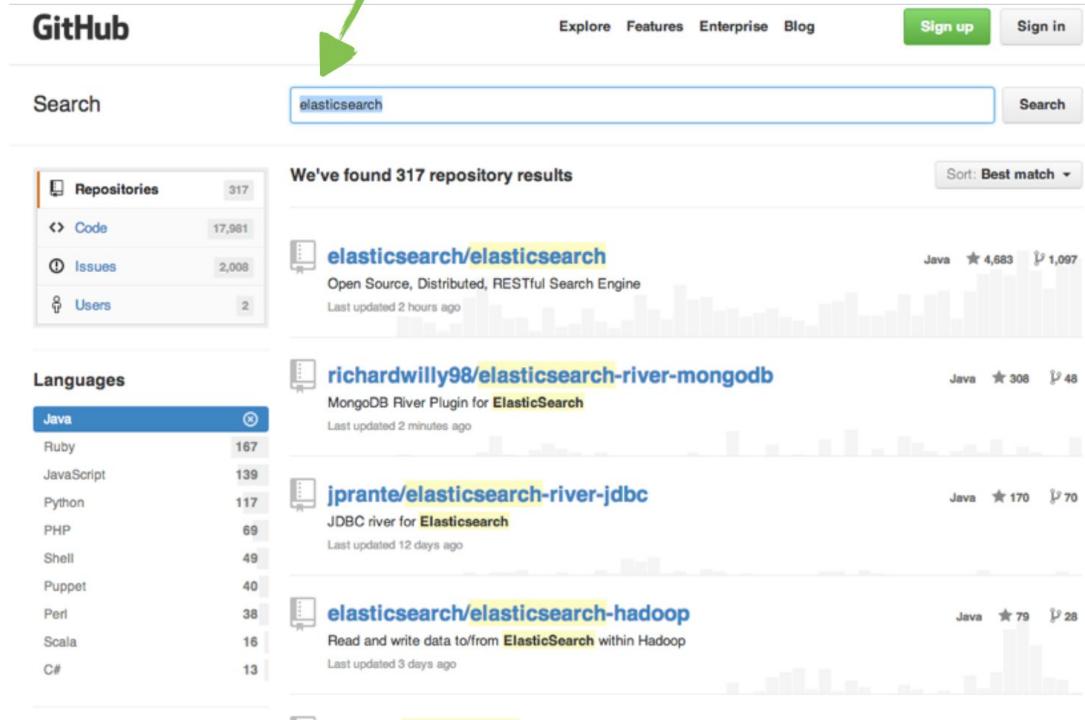
#UGIDOTNET

# Kudos

# COMPUTER GROSS

## managed/designs

# Unstructured search

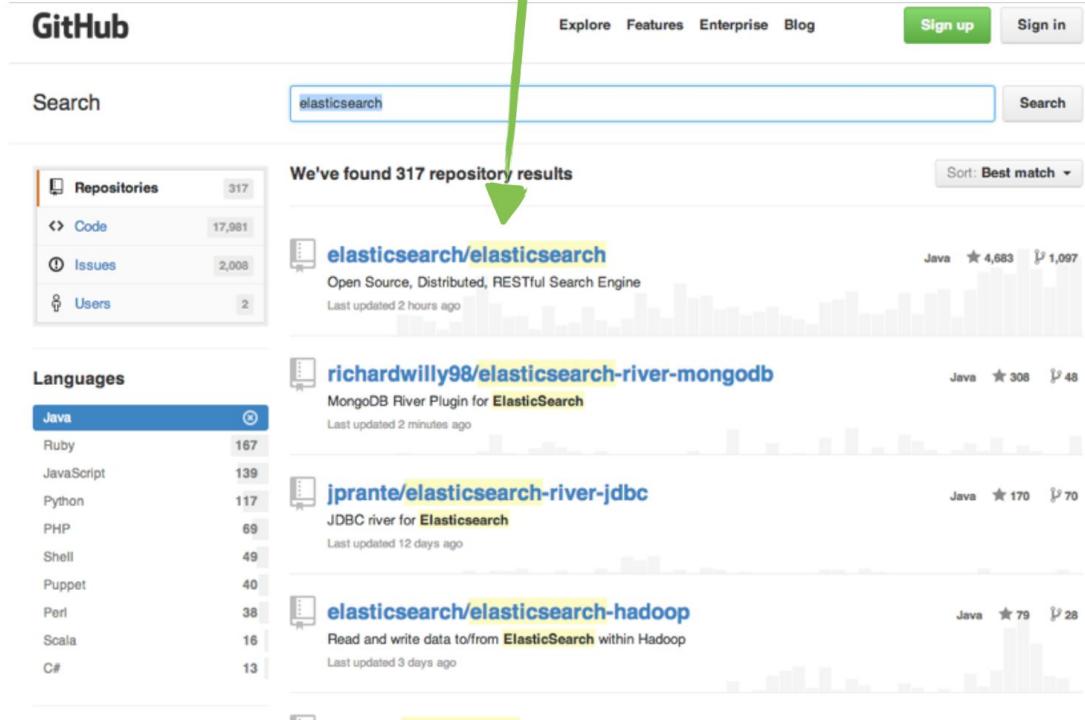


A screenshot of the GitHub search interface. A green arrow points from the title 'Unstructured search' down to the search bar. The search bar contains the text 'elasticsearch'. Below the search bar, the text 'We've found 317 repository results' is displayed. The results are listed in a grid format:

Repository	Description	Language	Stars	Forks
<a href="#">elasticsearch/elasticsearch</a>	Open Source, Distributed, RESTful Search Engine	Java	4,683	1,097
<a href="#">richardwilly98/elasticsearch-river-mongodb</a>	MongoDB River Plugin for ElasticSearch	Java	308	48
<a href="#">jprante/elasticsearch-river-jdbc</a>	JDBC river for Elasticsearch	Java	170	70
<a href="#">elasticsearch/elasticsearch-hadoop</a>	Read and write data to/from Elasticsearch within Hadoop	Java	79	28

On the left side of the page, there are two sidebar sections: 'Repositories' and 'Languages'. The 'Languages' section has 'Java' selected, with other languages listed below it.

# Enrichment



A screenshot of the GitHub search interface. The search bar at the top contains the query "elasticsearch". Below the search bar, a message says "We've found 317 repository results". A green arrow points from the word "elasticsearch" in the search bar down to the first repository result. The results are sorted by "Best match".

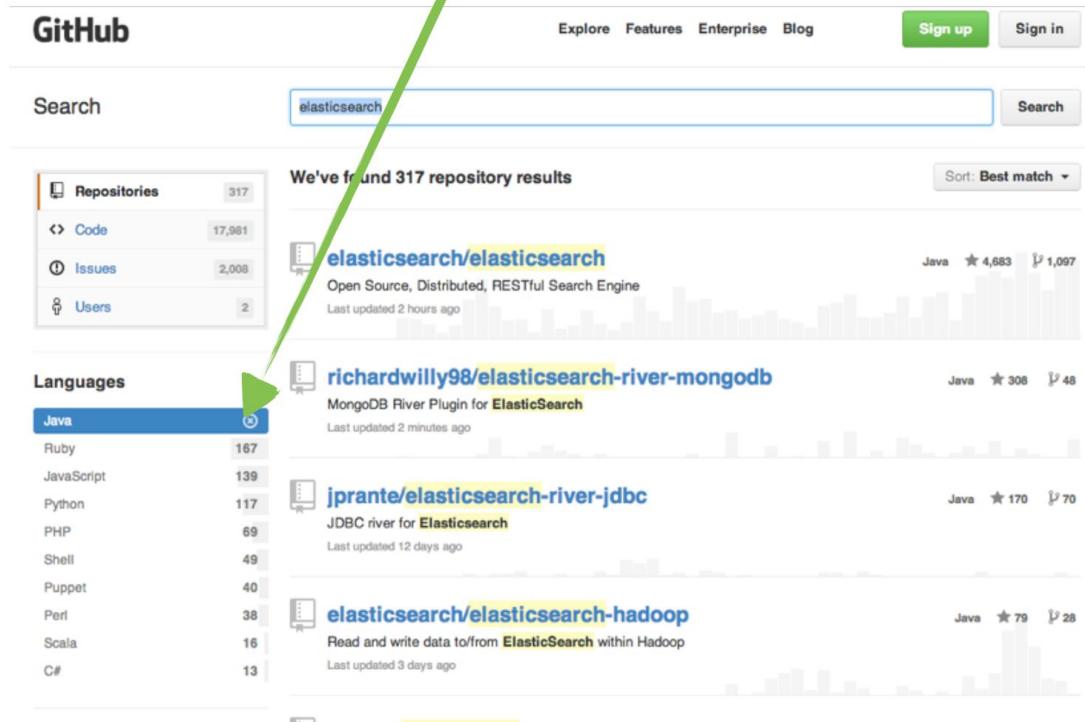
**Repositories**

- elasticsearch/elasticsearch
- richardwilly98/elasticsearch-river-mongodb
- jprante/elasticsearch-river-jdbc
- elasticsearch/elasticsearch-hadoop

**Languages**

Language	Count
Java	167
Ruby	139
JavaScript	117
Python	69
PHP	49
Shell	40
Puppet	38
Perl	16
Scala	13
C#	

# Aggregation



A screenshot of the GitHub search interface. The search bar at the top contains the query "elasticsearch". Below the search bar, there are sections for "Repositories", "Code", "Issues", and "Users". A green arrow points from the "Languages" section on the left towards the search results. The "Languages" section shows Java selected, with a count of 317 repositories. The main search results show 317 repository results found. The first result is "elasticsearch/elasticsearch" by "elasticsearch", described as an "Open Source, Distributed, RESTful Search Engine" in Java, with 4,683 stars and 1,097 forks. The second result is "richardwilly98/elasticsearch-river-mongodb" by "richardwilly98", described as a "MongoDB River Plugin for ElasticSearch" in Java, with 308 stars and 48 forks. The third result is "jprante/elasticsearch-river-jdbc" by "jprante", described as a "JDBC river for Elasticsearch" in Java, with 170 stars and 70 forks. The fourth result is "elasticsearch/elasticsearch-hadoop" by "elasticsearch", described as a "Read and write data to/from Elasticsearch within Hadoop" in Java, with 79 stars and 28 forks.

Language	Count
Java	317
Ruby	167
JavaScript	139
Python	117
PHP	69
Shell	49
Puppet	40
Perl	38
Scala	16
C#	13

# Suggestions



GitHub This repository ▾ debian

Sign up Sign in

Star 4,683 Fork 1,097

New Issue

elasticsearch

Browse Issues

Everyone's Issues

Labels

- Lucene 4.5 Upgr...
- breaking
- bug
- enhancement
- feature
- non-issue

This repository ▾ debian

① elasticsearch/elasticsearch#1726 debian package violates naming convention

① elasticsearch/elasticsearch#3571 debian package init-script: start-stop-daemon ne...

① elasticsearch/elasticsearch#1681 Debian pkg

① elasticsearch/elasticsearch#3286 There is no official debian/ubuntu repository

① elasticsearch/elasticsearch#3500 Elasticsearch should include debian's standard j...

① elasticsearch/elasticsearch#1526 Moving debian package to maven

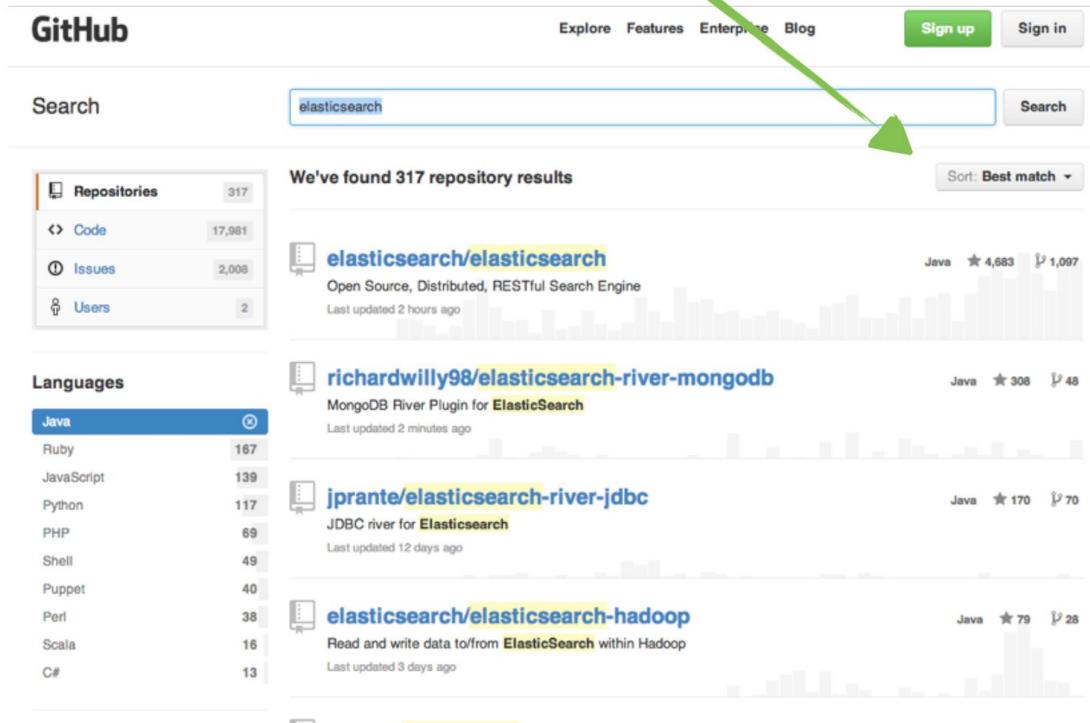
Search elasticsearch/elasticsearch for 'debian'

Search GitHub for 'debian'

① NoShardAvailableActionException in ES 0.90.3 on startup #3700  
Opened by richardwilly98 a day ago

① Feature Request: Don't reindex the document when updating non-indexed fields #3696  
Opened by ddorian 2 days ago 4 comments

# Sorting



A screenshot of the GitHub search results page for the query "elasticsearch". A green arrow points from the text "Sorting" at the top left towards the "Sort: Best match" dropdown menu in the header.

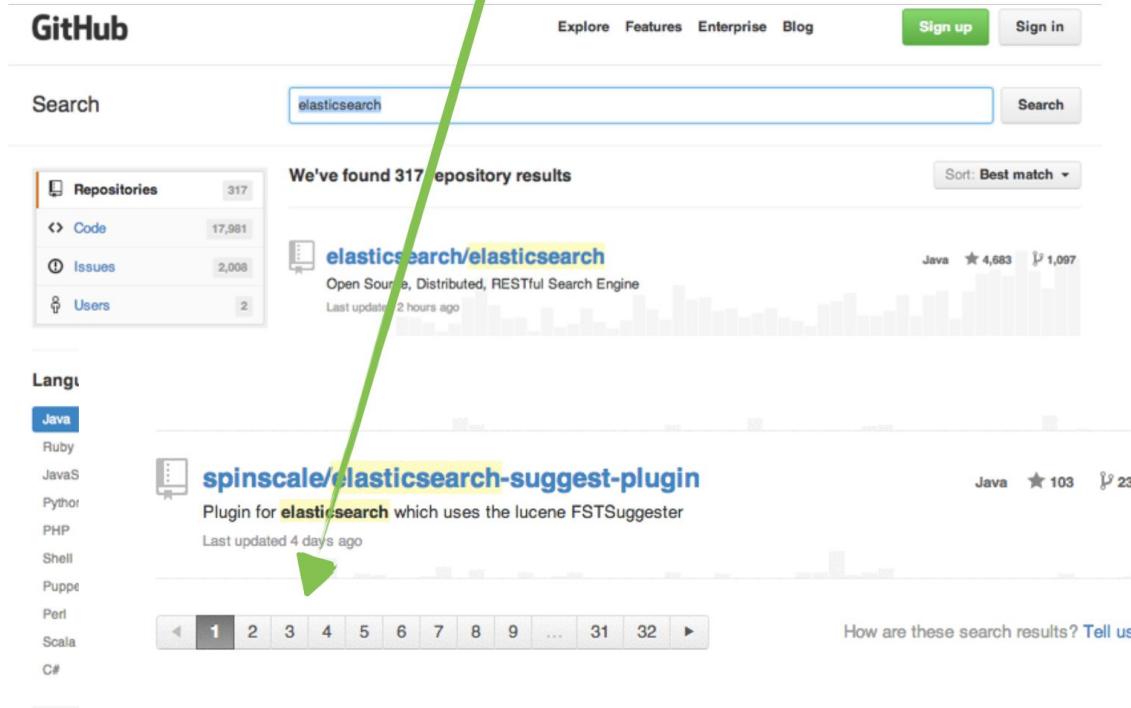
The GitHub interface includes a navigation bar with links for Explore, Features, Enterprise, and Blog, along with Sign up and Sign in buttons. The search bar contains the query "elasticsearch".

The search results section displays 317 repository results. The first result is the official Elasticsearch repository, followed by three Java-based plugins: "elasticsearch-river-mongodb", "elasticsearch-river-jdbc", and "elasticsearch-hadoop". Each result card shows the repository name, description, programming language (Java), star count, and fork count.

On the left side, there are two sidebar sections: "Repositories" (317 total) and "Languages" (Java is selected). The "Languages" section lists various programming languages and their counts: Java (167), Ruby (167), JavaScript (139), Python (117), PHP (69), Shell (49), Puppet (40), Perl (38), Scala (16), and C# (13).

Language	Count
Java	167
Ruby	167
JavaScript	139
Python	117
PHP	69
Shell	49
Puppet	40
Perl	38
Scala	16
C#	13

# Pagination



The screenshot shows the GitHub search interface with the query "elasticsearch" entered in the search bar. The results page displays 317 repository results, sorted by "Best match". A large green arrow points from the top right towards the bottom right, highlighting the pagination area.

**GitHub**

Explore Features Enterprise Blog Sign up Sign in

Search elasticsearch Search

We've found 317 repository results Sort: Best match ▾

**Repositories** 317

- Code 17,981
- Issues 2,008
- Users 2

**elasticsearch/elasticsearch**  
Open Source, Distributed, RESTful Search Engine Java ★ 4,683 ⚡ 1,097  
Last updated 2 hours ago

**Lang**

**Java** (selected)

- Ruby
- JavaScript
- Python
- PHP
- Shell
- Puppet
- Perl
- Scala
- C#

**spinscale/elasticsearch-suggest-plugin**  
Plugin for **elasticsearch** which uses the lucene FSTsuggester Java ★ 103 ⚡ 23  
Last updated 4 days ago

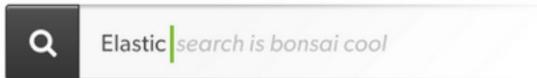
1 2 3 4 5 6 7 8 9 ... 31 32 ►

How are these search results? Tell us!

# Welcome Elasticsearch!

## full text search

Elasticsearch uses Lucene under the covers to provide the most powerful full text search capabilities available in any open source product. Search comes with multi-language support, a powerful query language, support for geolocation, context aware did-you-mean suggestions, autocomplete and search snippets.



## conflict management

Optimistic version control can be used where needed to ensure that data is never lost due to conflicting changes from multiple processes



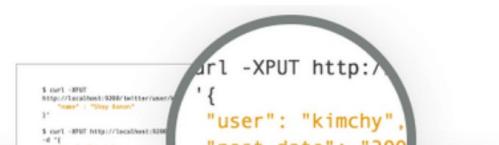
## document oriented

Store complex real world entities in Elasticsearch as structured JSON documents. All fields are indexed by default, and all the indices can be used in a single query, to return results at breath taking speed.

```
Document oriented  
  
$ curl -XPUT http://localhost:9200/twitter/user/kimchy -d  
'  
  {  
    "name" : "Shay Banon"  
  }'
```

## schema free

Elasticsearch allows you to get started easily. Toss it a JSON document and it will try to detect the data structure, index the data and make it searchable. Later, apply your domain specific knowledge of your data to customise how your data is indexed.



```
curl -XPUT http://localhost:9200/twitter/user/kimchy -d  
'  
  {  
    "user": "kimchy",  
    "postDate": "2012-09-10T13:25:00",  
    "titles": [ "Stranger in a Strange Land" ],  
    "text": "I am a Strange Loop"  
  }'
```

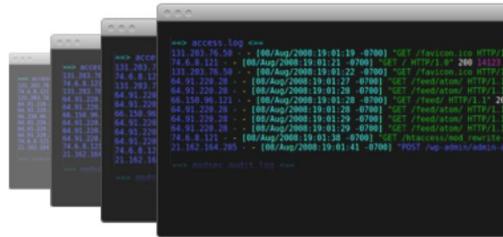
## restful api

Elasticsearch is API driven. Almost any action can be performed using a simple RESTful API using JSON over HTTP. An API already exists in the language of your choice.



## per-operation persistence

Elasticsearch puts your data safety first. Document changes are recorded in transaction logs on multiple nodes in the cluster to minimise the chance of any data loss.



## apache 2 open source license

Elasticsearch can be downloaded, used and modified free of charge. It is available under the Apache 2 license, one of the most flexible open source licenses available.



## build on top of apache lucene™

Apache Lucene is a high performance, full-featured Information Retrieval library, written in Java. Elasticsearch uses Lucene internally to build its state of the art distributed search and analytics capabilities.



```
> curl XGET http://localhost:9200?pretty
```



**verb**

```
> curl XGET http://localhost:9200?pretty
```



**node**

```
> curl XGET http://localhost:9200?pretty
```



**port**

```
> curl XGET http://localhost:9200?pretty
```



**query string**

# DEMO

```
> GET /twitter/_search?q=2014
```

```
11 results
```

```
> GET /twitter/_search?q=2014
```

11 results

```
> GET /twitter/_search?q=2014-09-13
```

11 results

```
> GET /twitter/_search?q=2014
```

11 results

```
> GET /twitter/_search?q=2014-09-13
```

11 results

```
> GET /twitter/_search?q=date:2014-09-13
```

1 result

```
> GET /twitter/_search?q=2014
11 results

> GET /twitter/_search?q=2014-09-13
11 results

> GET /twitter/_search?q=date:2014-09-13
1 result

> GET /twitter/_search?q=date:2014
no result
```

# DEMO

[mapping]

**date** = type:**date**

**\_all** = type:**string**

**date** = exact value

**\_all** = full text

# Inverted index

“The quick brown fox jumped over the lazy dog”

“Quick brown foxes leap over lazy dogs in summer”

# Inverted index: **separate words / terms**

[The,quick,brown,fox,jumped,over,the,lazy,dog]

[Quick,brown,foxes,leap,over,lazy,dogs,in,summer]

# Inverted index: **sort unique terms**

[The,brown,dog,fox,jumped,lazy,over,quick,the]

[Quick,brown,dogs,foxes,in,lazy,leap,over,summer]

# Inverted index: list docs containing terms

term	doc-01	doc-02
Quick		X
The	X	
brown	X	X
dog	X	
dogs		X
fox	X	
...	...	...

# analysis

=

tokenization + normalization

# analyzer

=

tokenizer + token filters

# Standard analyzer

“The Quick Brown Fox jumped over the Lazy Dog!”

# Standard analyzer **standard tokenizer**

“The,Quick,Brown,Fox,jumped,over,the,Lazy,Dog”

# Standard analyzer **lowercase filter**

“the,quick,brown,fox,jumped,over,the,lazy,dog”

# Standard analyzer **stopword filter**

“quick,brown,fox,jumped,over,lazy,dog”

# English analyzer

“The Quick Brown Fox jumped over the Lazy Dog!”

# English analyzer **standard tokenizer**

“The,Quick,Brown,Fox,jumped,over,the,Lazy,Dog”

# English analyzer **lowercase filter**

“the,quick,brown,fox,jumped,over,the,lazy,dog”

# English analyzer **english stemmer**

“the,quick,brown,fox,jumped,over,the,lazy,dog”

# English analyzer **english stemmer**

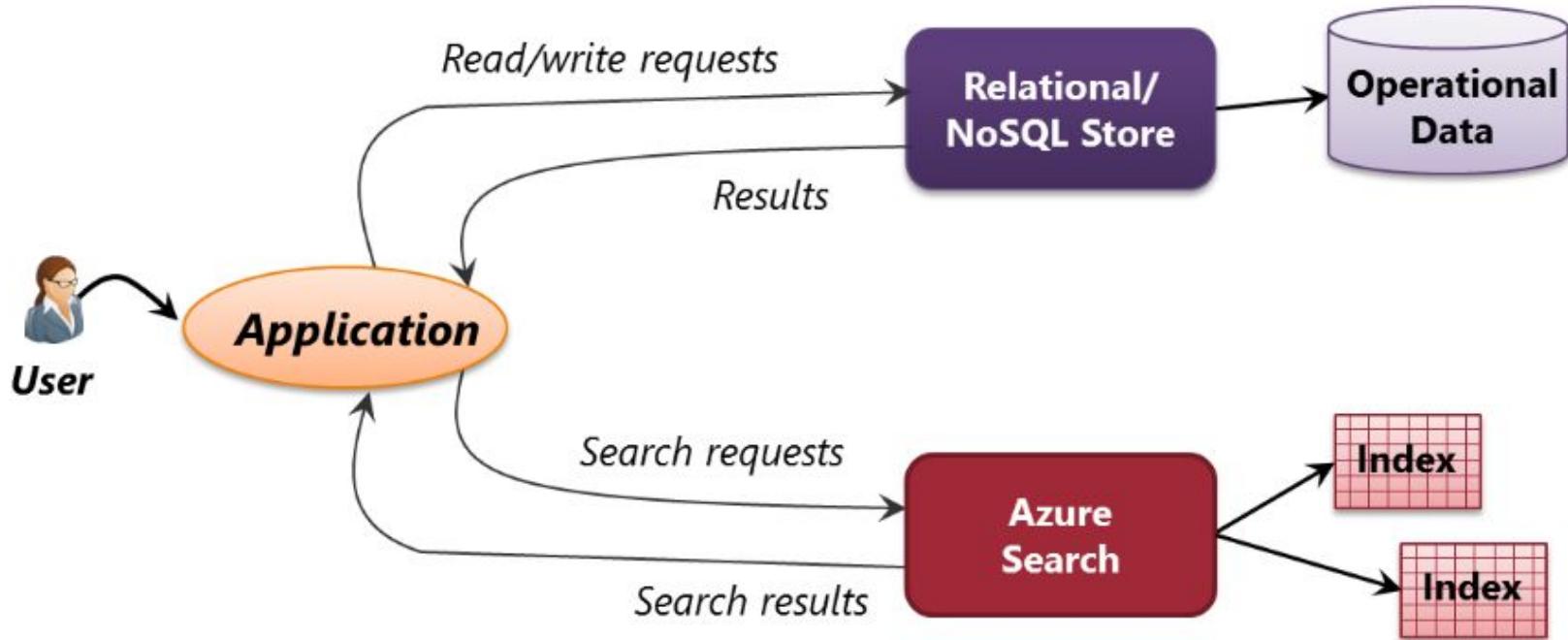
“the,quick,brown,fox,jump,over,the,lazy,dog”

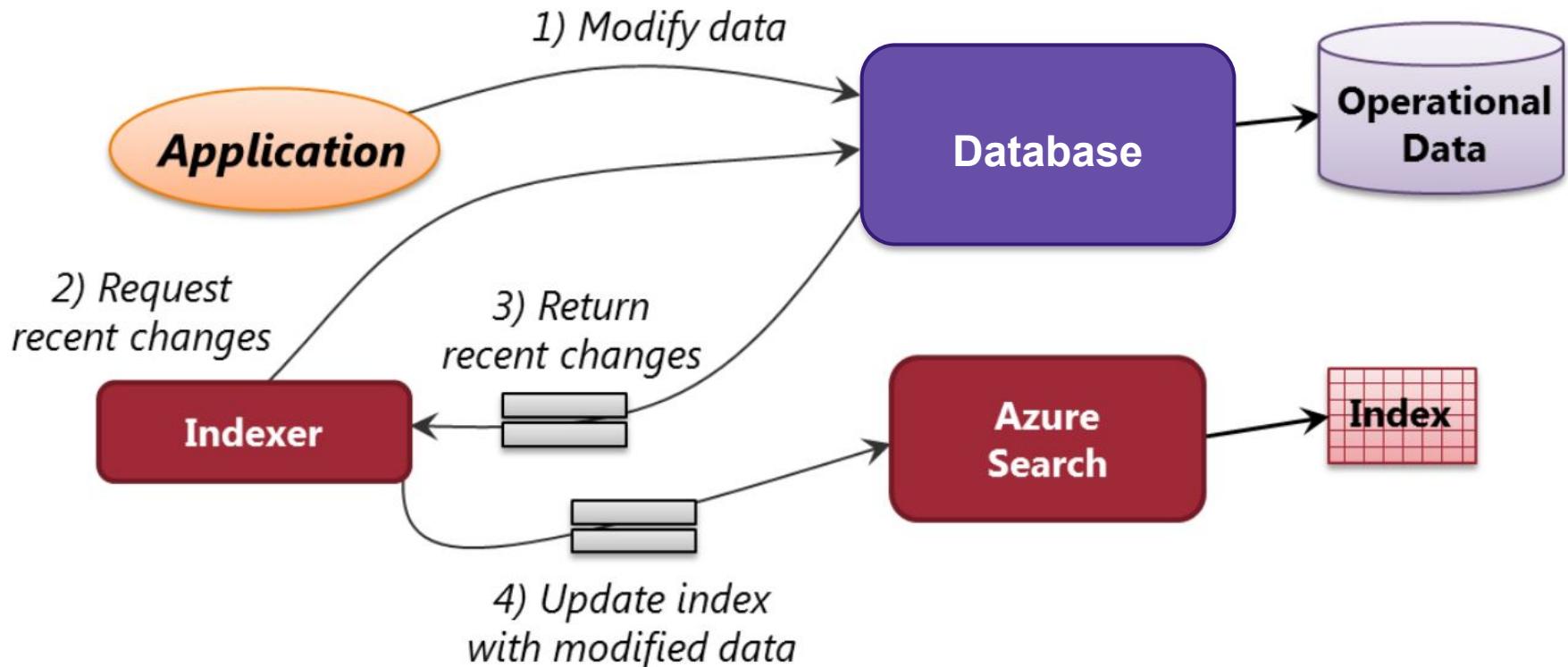
# English analyzer **english stopwords**

“quick,brown,fox,jump,over,lazy,dog”

# DEMO

[analyzer]





# DEMO

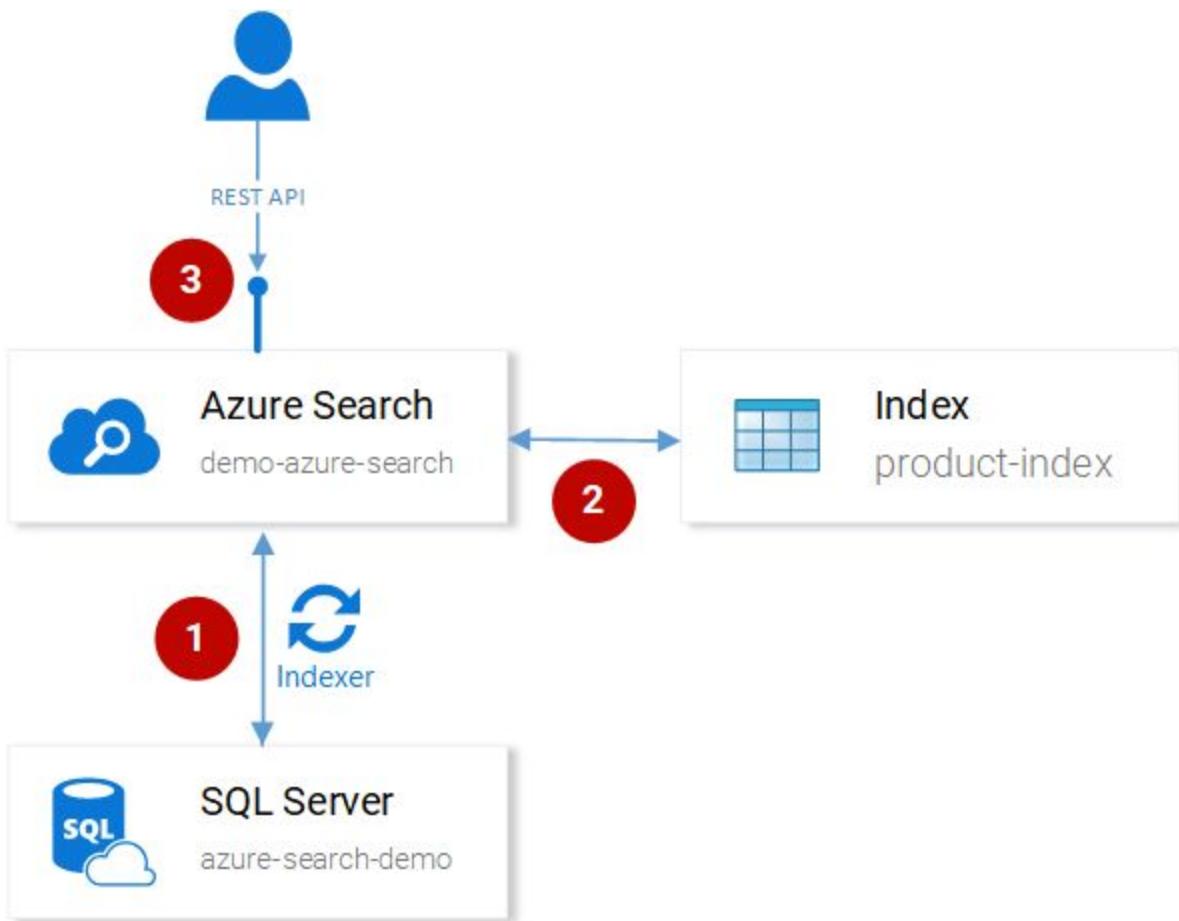
[manual indexer]

# Welcome Azure Search!

Azure Search is a search-as-a-service cloud solution  
that gives developers APIs and tools for adding a rich search experience  
over private, heterogenous content in web, mobile, and enterprise applications.

# DEMO

[searching]



# DEMO

[indexer]



# APP MODERNISATION DAY

09/04/2019 — MICROSOFT HOUSE, MILANO

L'EVENTO È STATO POSSIBILE GRAZIE A:

COMPUTER  
GROSS

managed/designs