Harnessing Flexible Data in the Cloud

Matthias Brantner (brantner@28msec.com)

28 msec

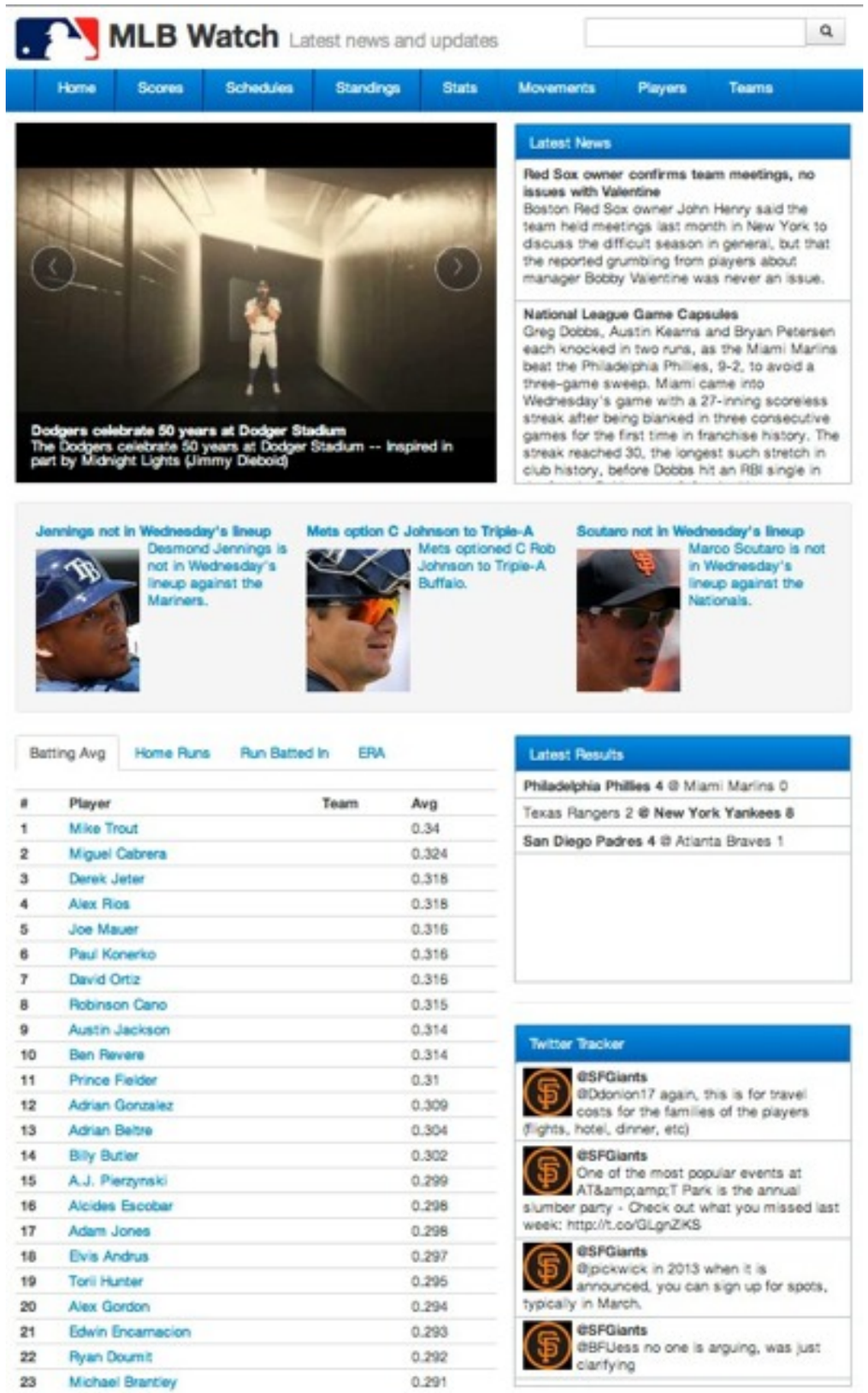
Demo

Flexible Data

JSON

SportsML

SportsML



Text

NITF
(XML)

SportsML

JSON

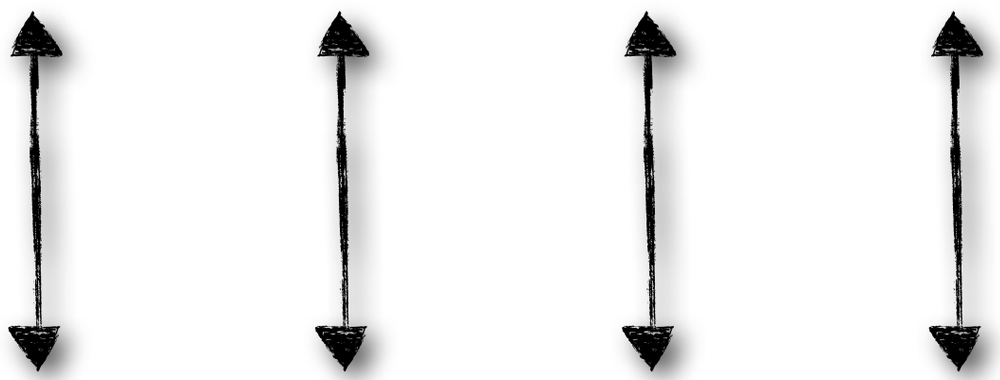
Harness Data

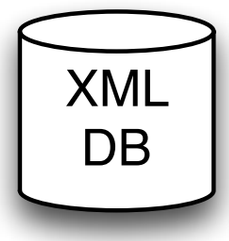
28 msec

- store
- query & update
- full-text search
- complex queries
- transform

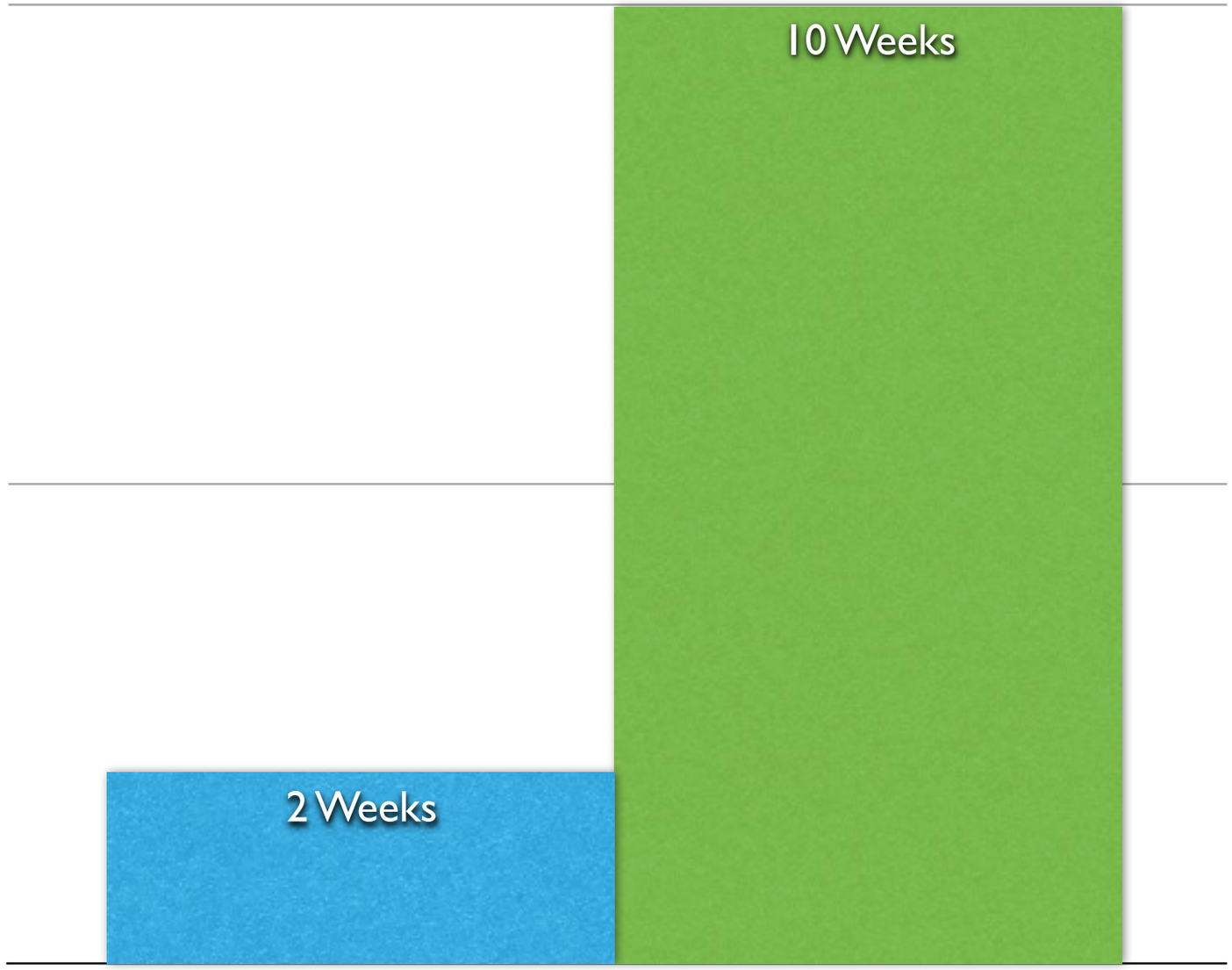
reliability, availability, performance, and scalability

Possible Solution Stack

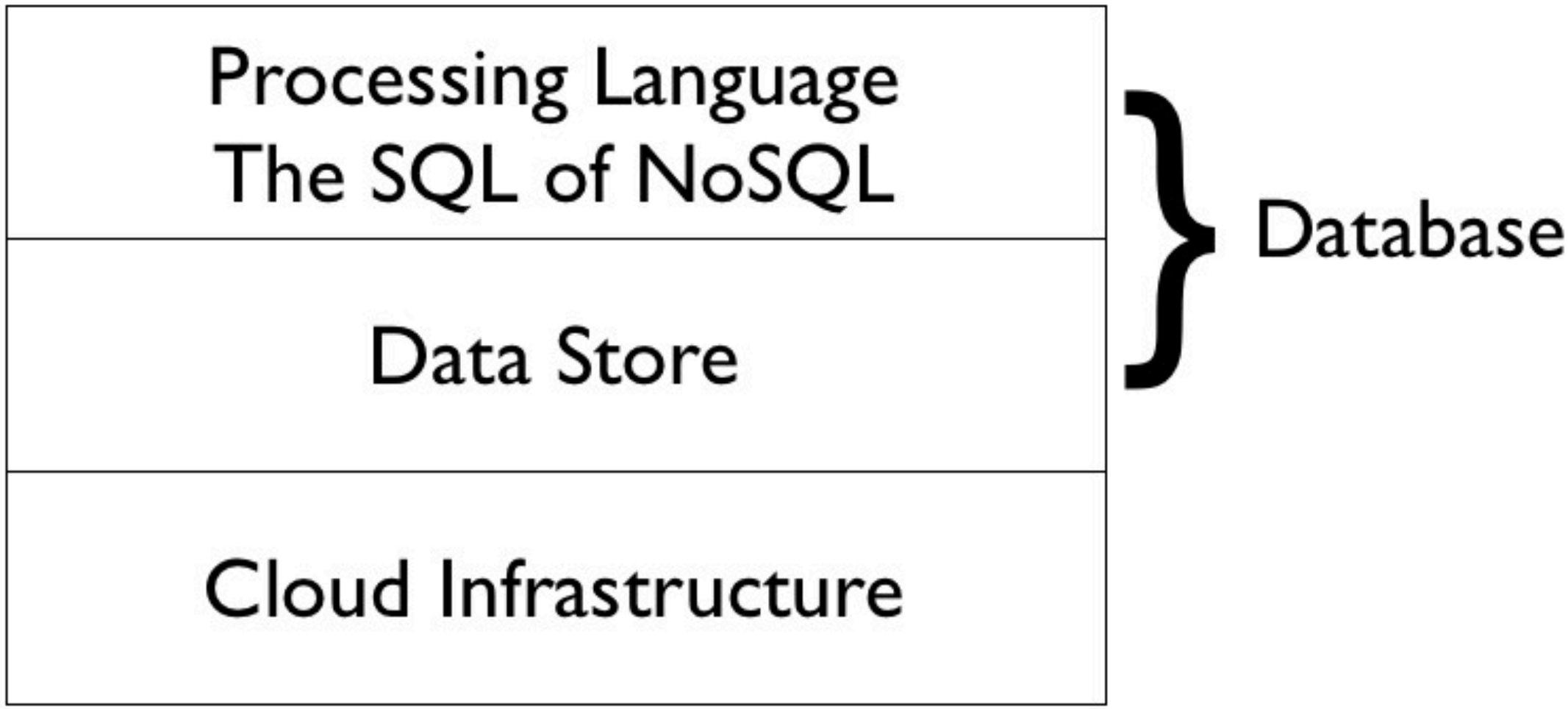




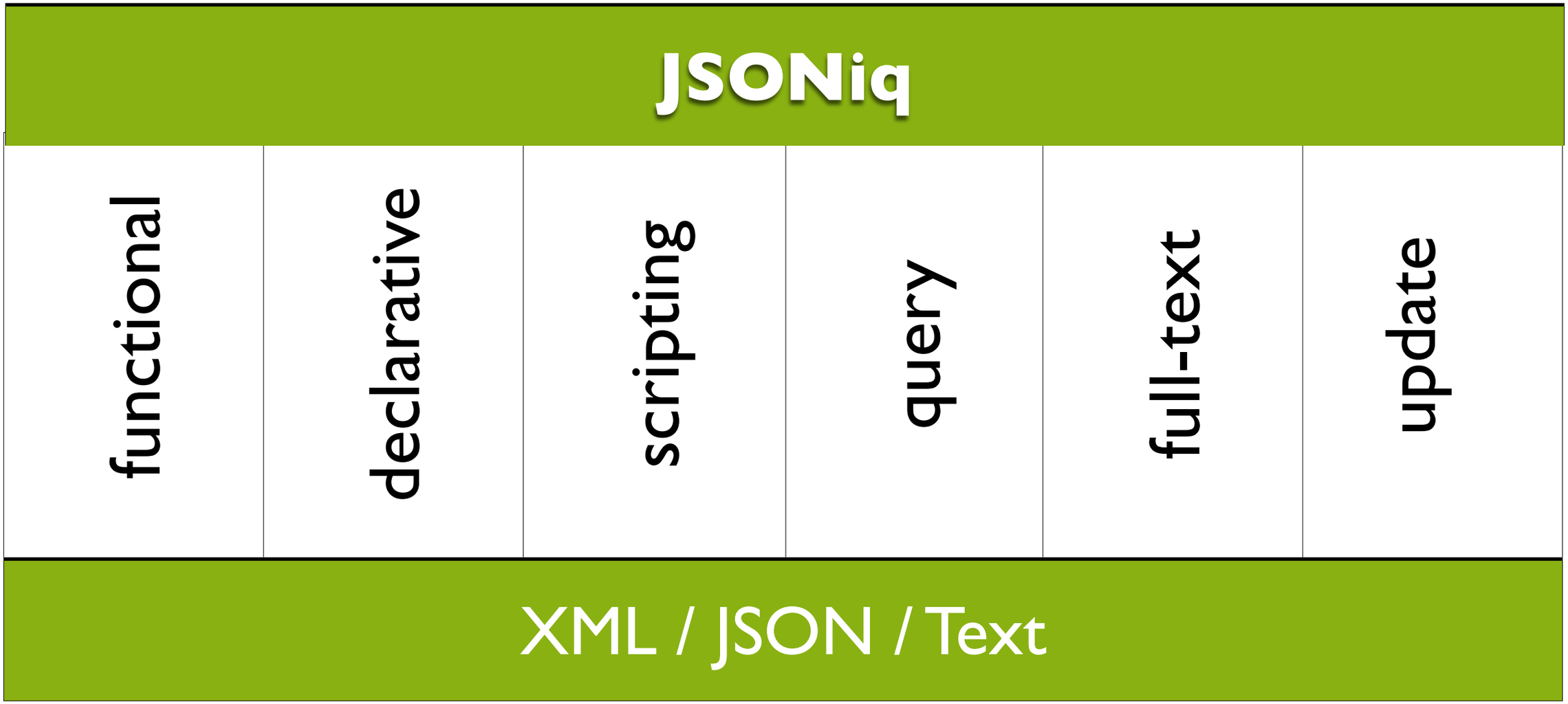
Development Time



How do we do this?



Processing Language



Example (1)

28 msec

```
1  for $zip in db:collection("zips")
2  group by $zip("state")
3  let $count := count($zip)
4  order by $count descending
5  return { “state” : $state, “count” : $count }
```

```
{ "state" : "TX",  "count" : 1676 }
{ "state" : "NY",  "count" : 1596 }
{ "state" : "CA",  "count" : 1500 }
...
```

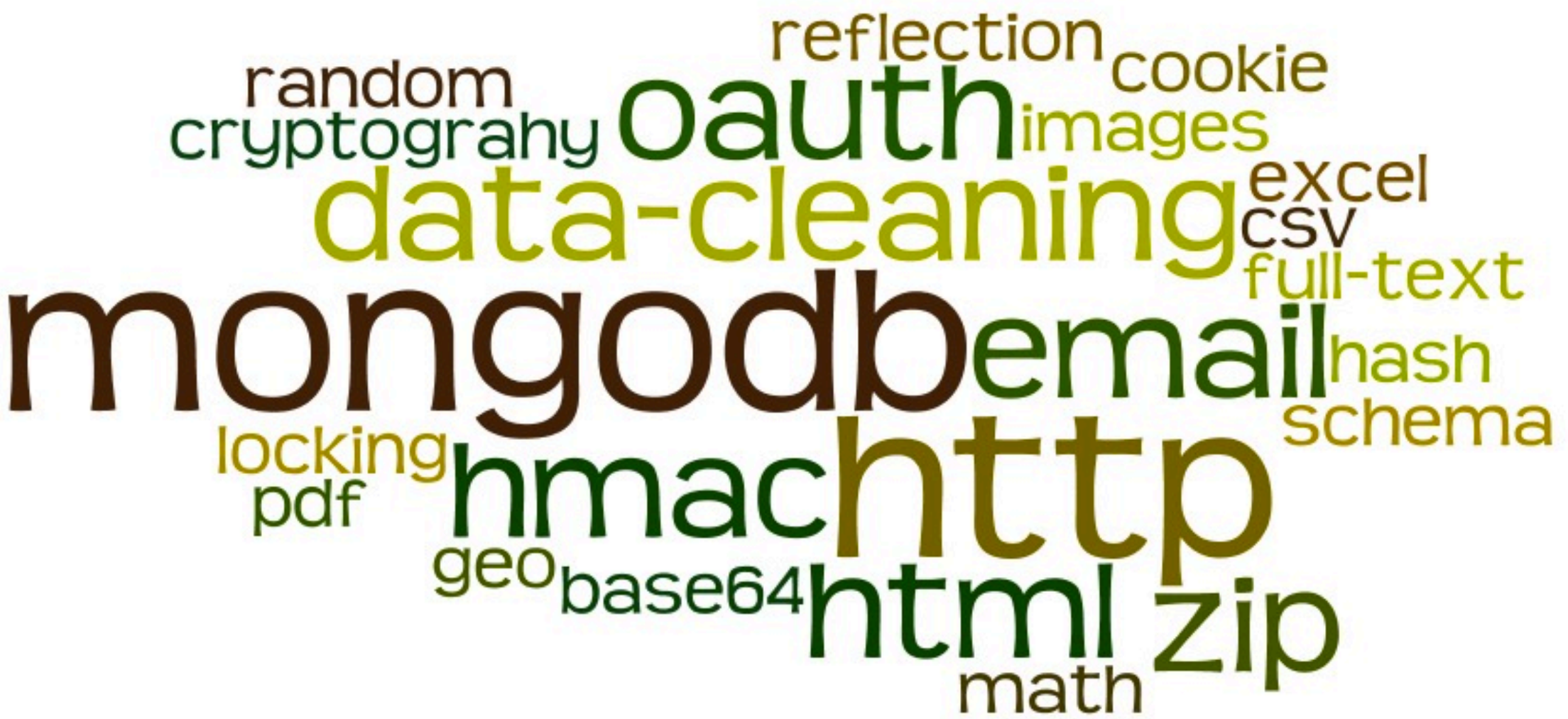
Example (2)

28 msec

```
1  let $search-result := jn:parse-json(
2    http:get-text("twitter.com/search.json?q=.")[2])
3  for $result in jn:members($search-result("results"))
4  let $text := $result("text")
5  for $token in ft:tokenize-string($text)
6  where not(ft:is-stop-word($token))
7  let $lc := ft:strip-diacritics(lower-case($token))
8  group by $lc
9  let $count := count($token)
10 order by $count descending
11 return { “token” : $lc, “count” : $count }
```

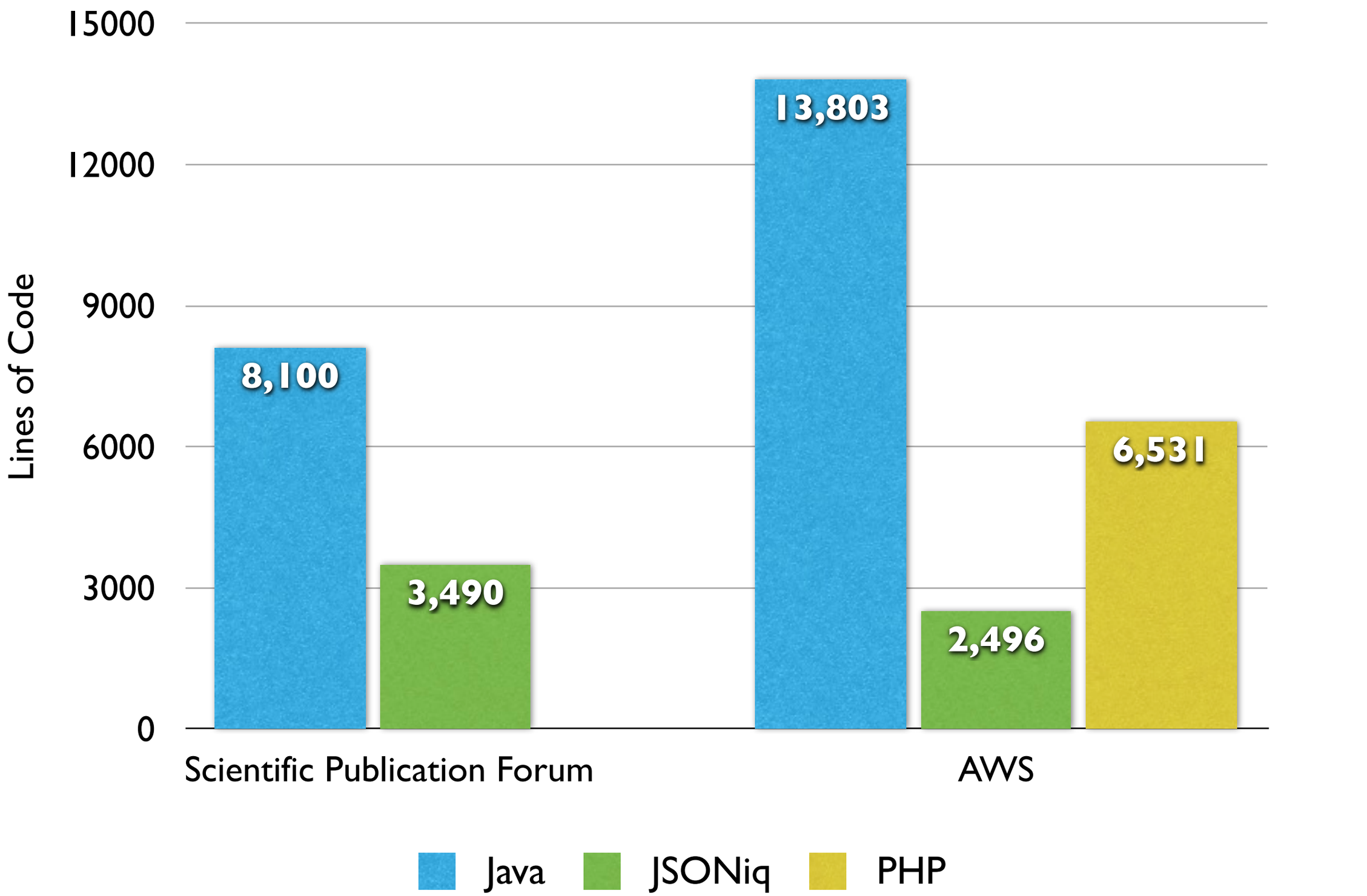
```
{ "token" : "hard", "count" : 3 }
{ "token" : "truths", "count" : 3 }
{ "token" : "sql", "count" : 3 }
{ "token" : "revolution", "count" : 2 }
```

Module Library



Productivity

28 msec



document store

performant

high availability

scalable

atomic updates

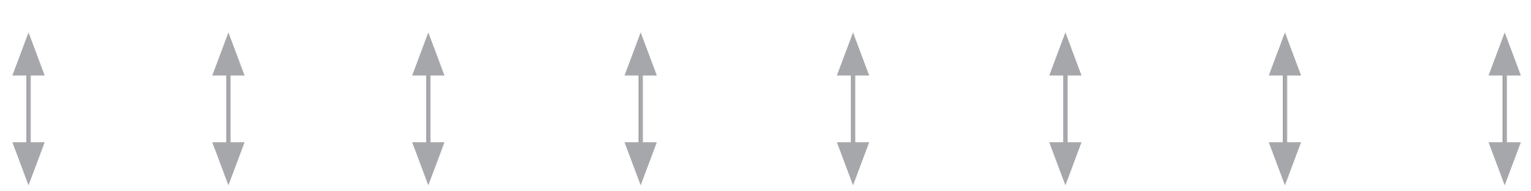
JSON / XML

Storage Layout (1)

JSONiq Collection

```
{
  "name" : "Matthias"
  "dob" : 1978-09-12
}
```

```
<person>
  <name>Matthias</name>
  <dob>1978-09-12</dob>
</person>
```



MongoDB Collection

```
{
  "name" : "Matthias"
  "dob" : 1978-09-12
}
```

...

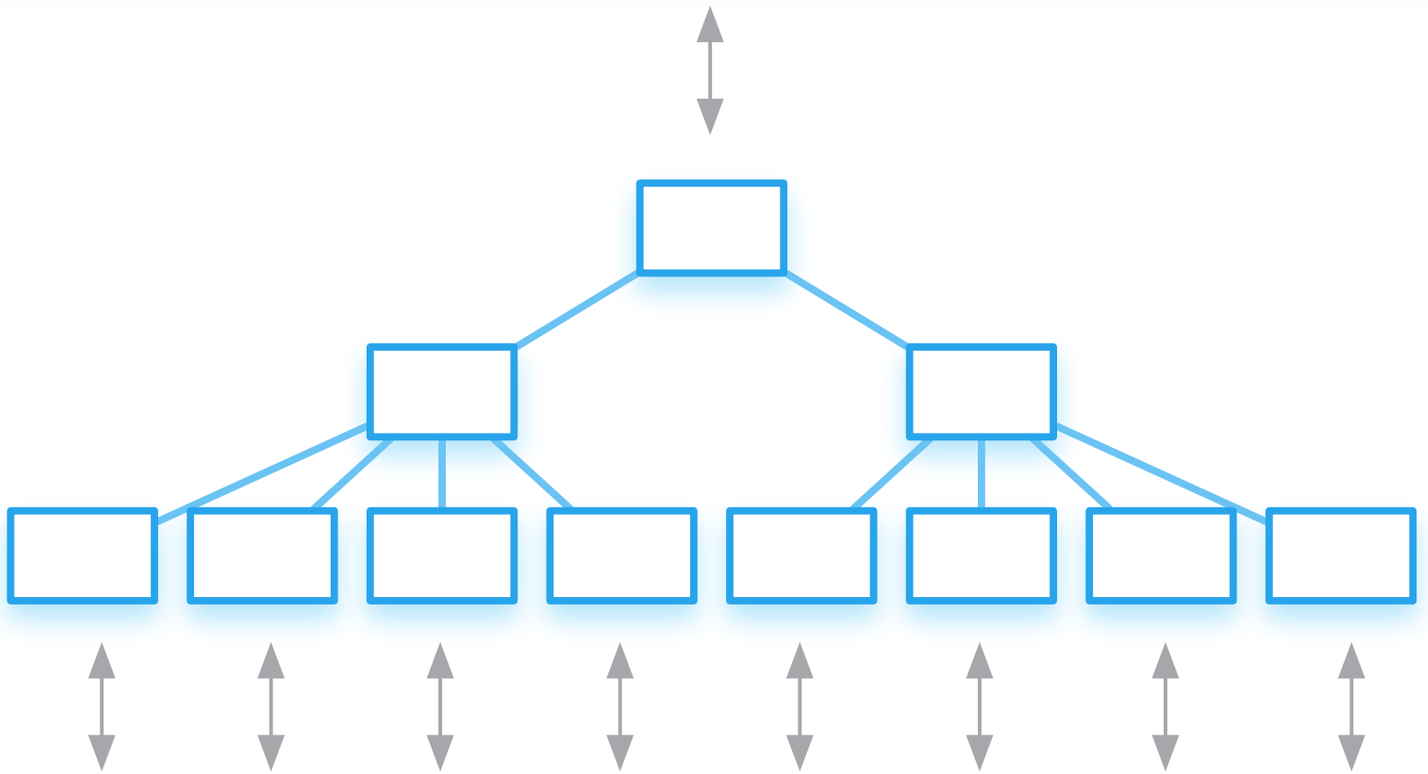
```
<person>
  <name>Matthias</name>
  <dob>1978-09-12</dob>
</person>
```

Storage Layout (2)

JSONiq Collection

```
{
  "name" : "Matthias"
  "dob" : 1978-09-12
}
```

```
<person>
  <name>Matthias</name>
  <dob>1978-09-12</dob>
</person>
```



Index (BTree)

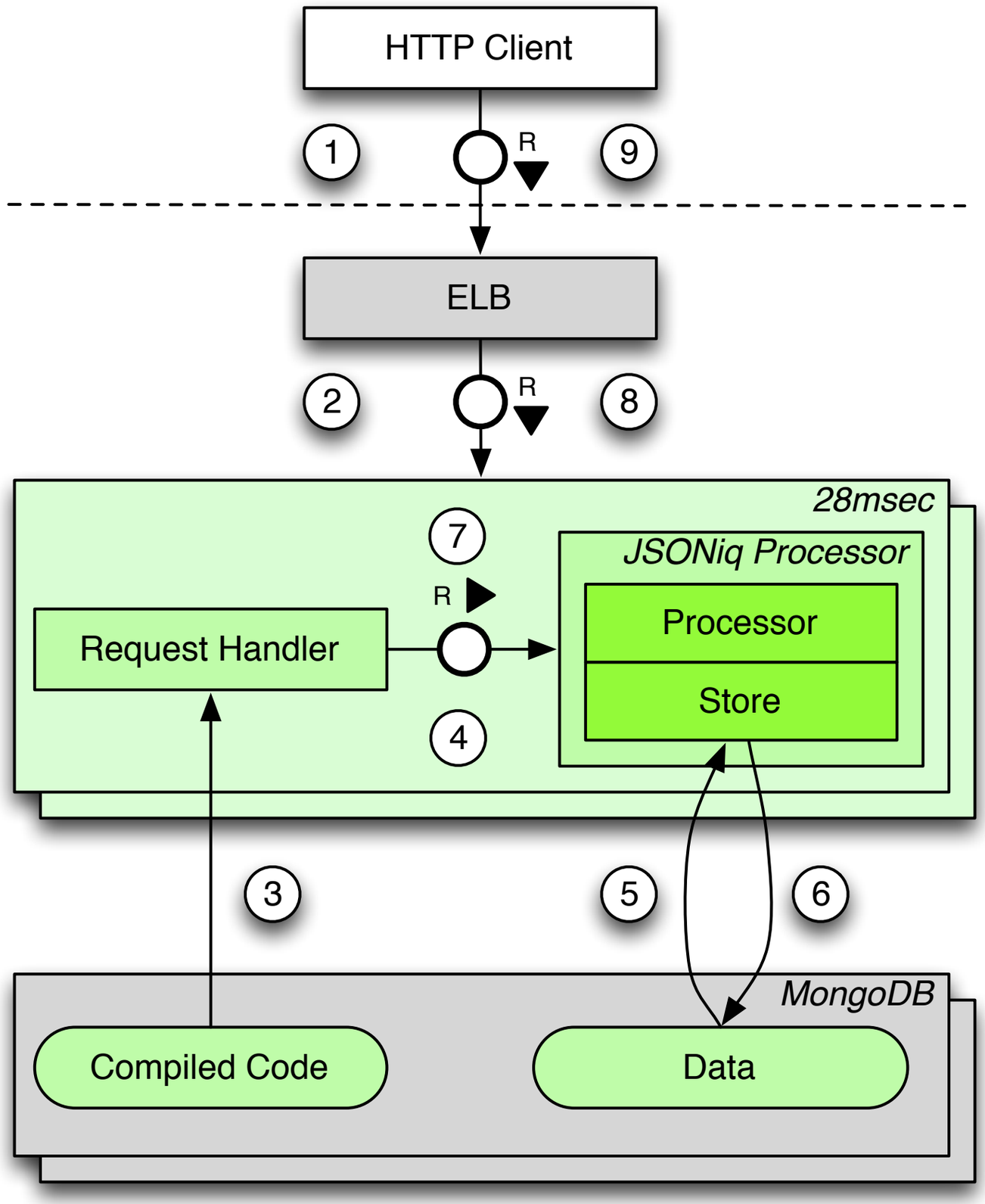
MongoDB Collection

```
{
  "name" : "Matthias"
  "dob" : 1978-09-12
}
```

...

```
<person>
  <name>Matthias</name>
  <dob>1978-09-12</dob>
</person>
```

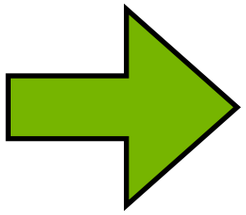
Cloud Infrastructure - Request Processing



Conclusion

28 msec

- JSONiq to process flexible data
- using MongoDB as data store and index



Database

- up to 5x more (cost) effective
- scalable infrastructure on AWS

JSONiq talks

- 11:00 am
  - Jonathan Robie (EMC)
- 11:45am
  - Chris Hillery (FLWOR Foundation)

28 msec

Thank You!

Questions?

Matthias Brantner (brantner@28msec.com)

Visit us at our booth!