

Realtime Availability for eShop


Mikhail Khlodnev



Grid Dynamics

Scalable eCommerce Platform Solutions







Agenda

- Availability Updates Problem
 - Approaches
 - Solr Nuts&Bolts
 - Architecture
- 

▼ Style

- Polos (1)
- T-shirts (1)
- Button-front (1)

▼ Color

-  Red(2)
-  Beige (1)
-  Black (2)
-  Blue (2)
-  Brown (1)
-  Gray (1)

See More ▶

▼ Current Offers

- Clearance (2)
- All Items On Sale (2)
- Free Shipping (2)

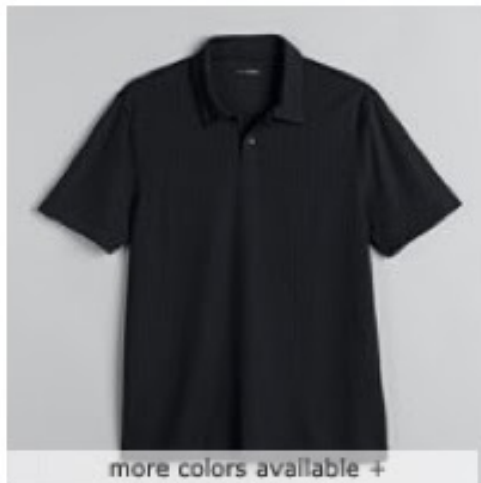
▶ Discount

▼ Size

- XL (2)
- S (2)
- M (2)
- L (2)
- XXL (2)

Shirts

All Products (2) (2) Sort by Date Added to Site ▾



John Henry Men's Sport Shirt - Modern Fit



\$7.99 - \$42.00

Details



John Henry Men's Short Sleeve Crewneck Tee



\$7.99 - \$32.00

Details

add to compare

John Henry Men's Short Sleeve Crewneck Tee

0 Reviews | | [Write a review](#)

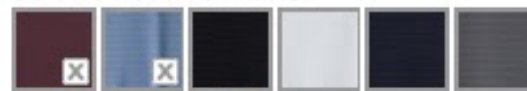
From RUB 323.76 To RUB 1,296.68

This item is not available in this selection


Extended Sizes: Men's

Size: Sleeve Length: Quantity:

Color: Select from below



[Size Guide](#) | [Reset Selection](#)

 Add to Cart

Please select Color

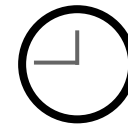
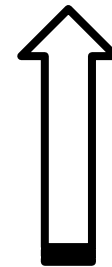
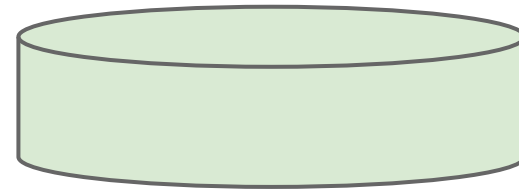
Ship

JOHN HENRY.

John Henry Shirts
John Henry Men's
All John Henry

Availability Updates

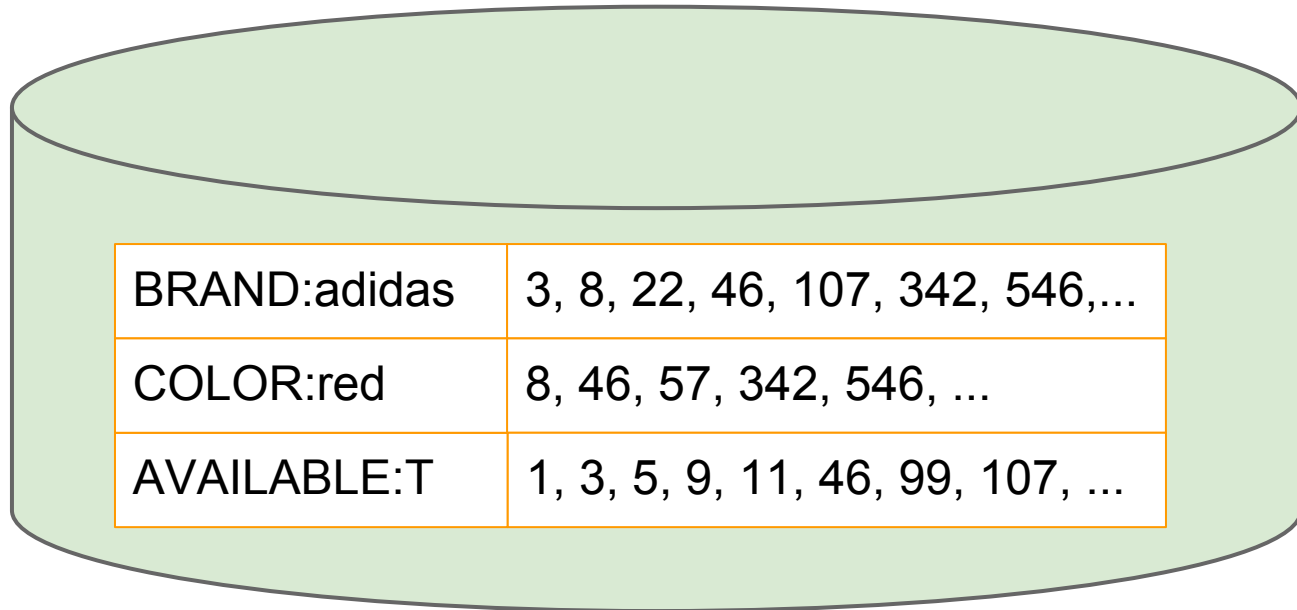
$q=polo\&$
→
 $fq=AVAILABLE:T$



**Inventory
Tracking**

CMS

What the fq?



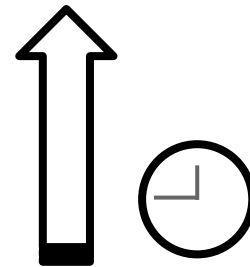
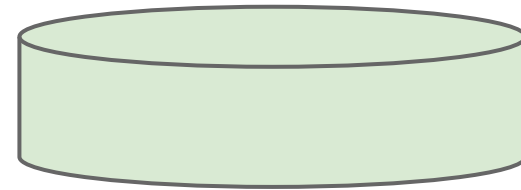
BRAND:adidas	3, 8, 22, 46, 107, 342, 546,...
COLOR:red	8, 46, 57, 342, 546, ...
AVAILABLE:T	1, 3, 5, 9, 11, 46, 99, 107, ...

AVAILABLE:T

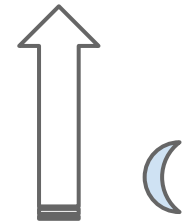
0	1	0	1	0	1	0	0	0	1	0	0	...
---	---	---	---	---	---	---	---	---	---	---	---	-----

Availability Updates

$q = \text{polo} \&$
—————→
 $fq = \text{AVAILABLE:T}$



Inventory
Tracking

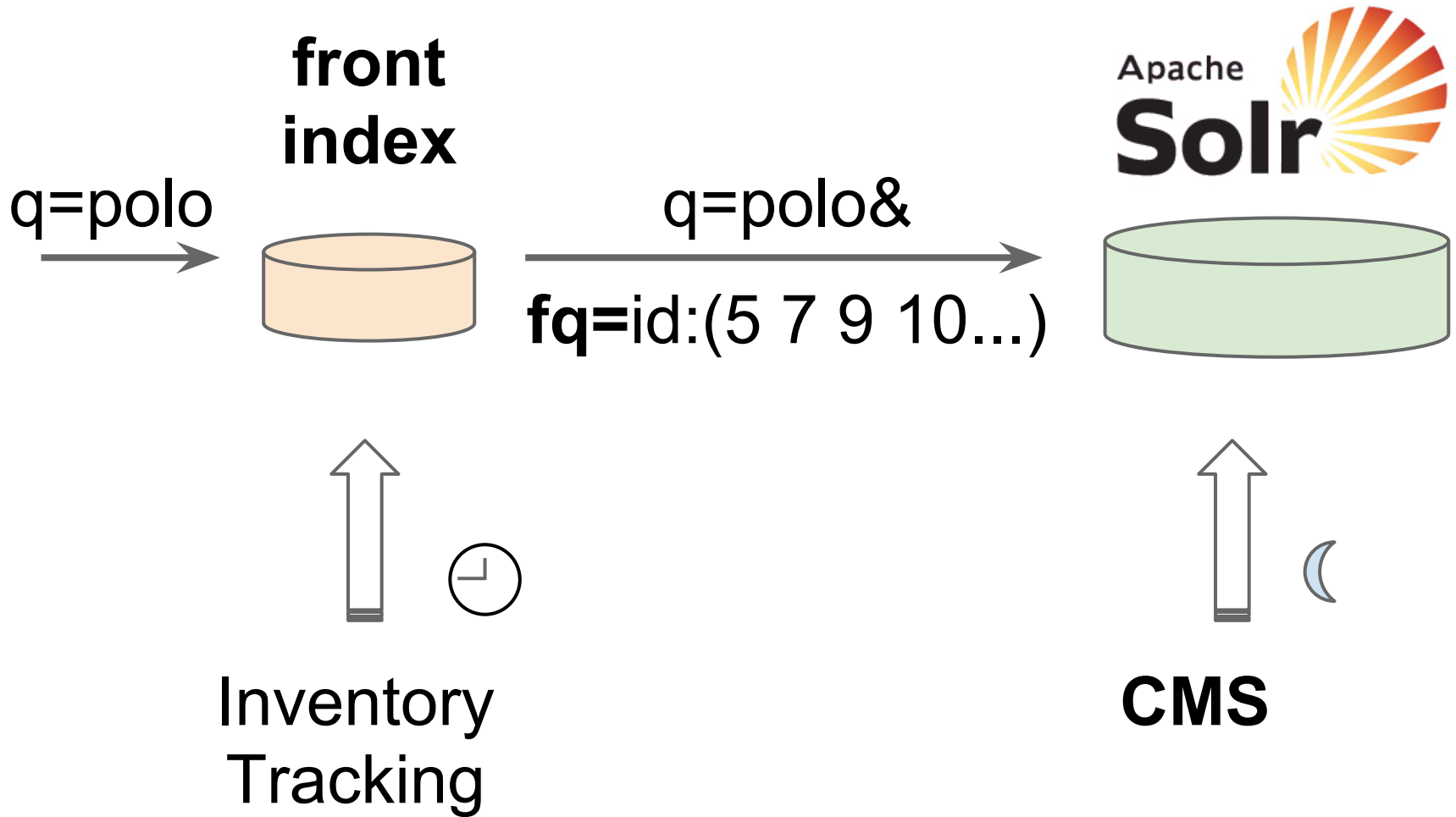


CMS

Frequent Updates are HARD

- reindex whole docs
- replicate often
- tune for merges
- regenerate caches
- reserve heap for all above

Two Indexes



Core Problem

**..&fq=id:4740 OR id:456 OR id:5667
OR id:28224 OR id:1038
OR id:45283 OR id:3435..&..**

It's Over 9000!

- hard to parse
- seeking terms many times

Isn't there a cache?

- low hit ratio
- cache hit costs $O(n_clauses)$

What's the difference?

fq=id:(5 7 9 10...)

VS

fq=AVAILABLE:T

They both are filters!

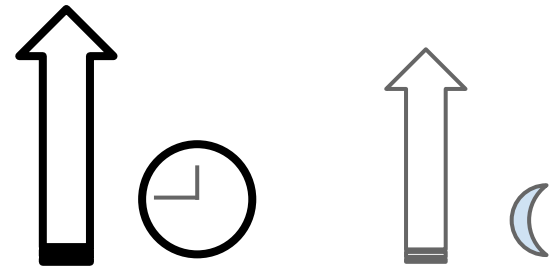
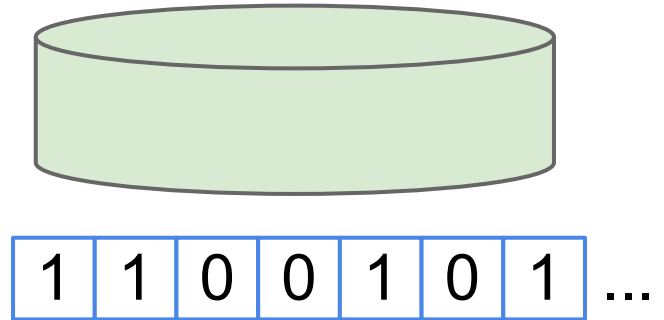
fq=id:(5 7 9 10...)

1	1	0	0	1	0	1	0	0	0	1	1	...
---	---	---	---	---	---	---	---	---	---	---	---	-----

fq=AVAILABLE:T

Availability Updates 2.0

$q = \text{polo} \&$
→
 $f_q = \text{AVAILABLE} : T$



Inventory
Tracking

CMS

Solr Nuts&Bolts



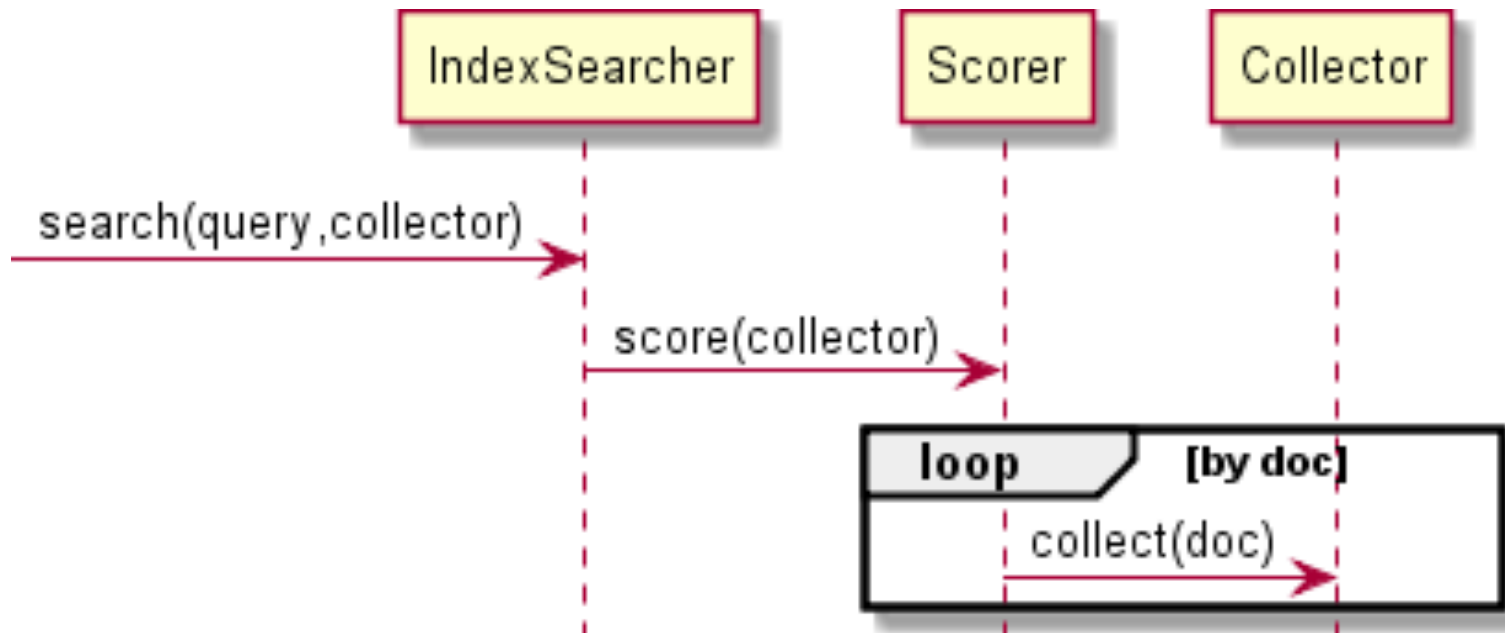
QParserPlugin

q={!myparser}short sleeve polo

Nested Queries

q=**text:polo _query_:"{!myparser}blueberry"**

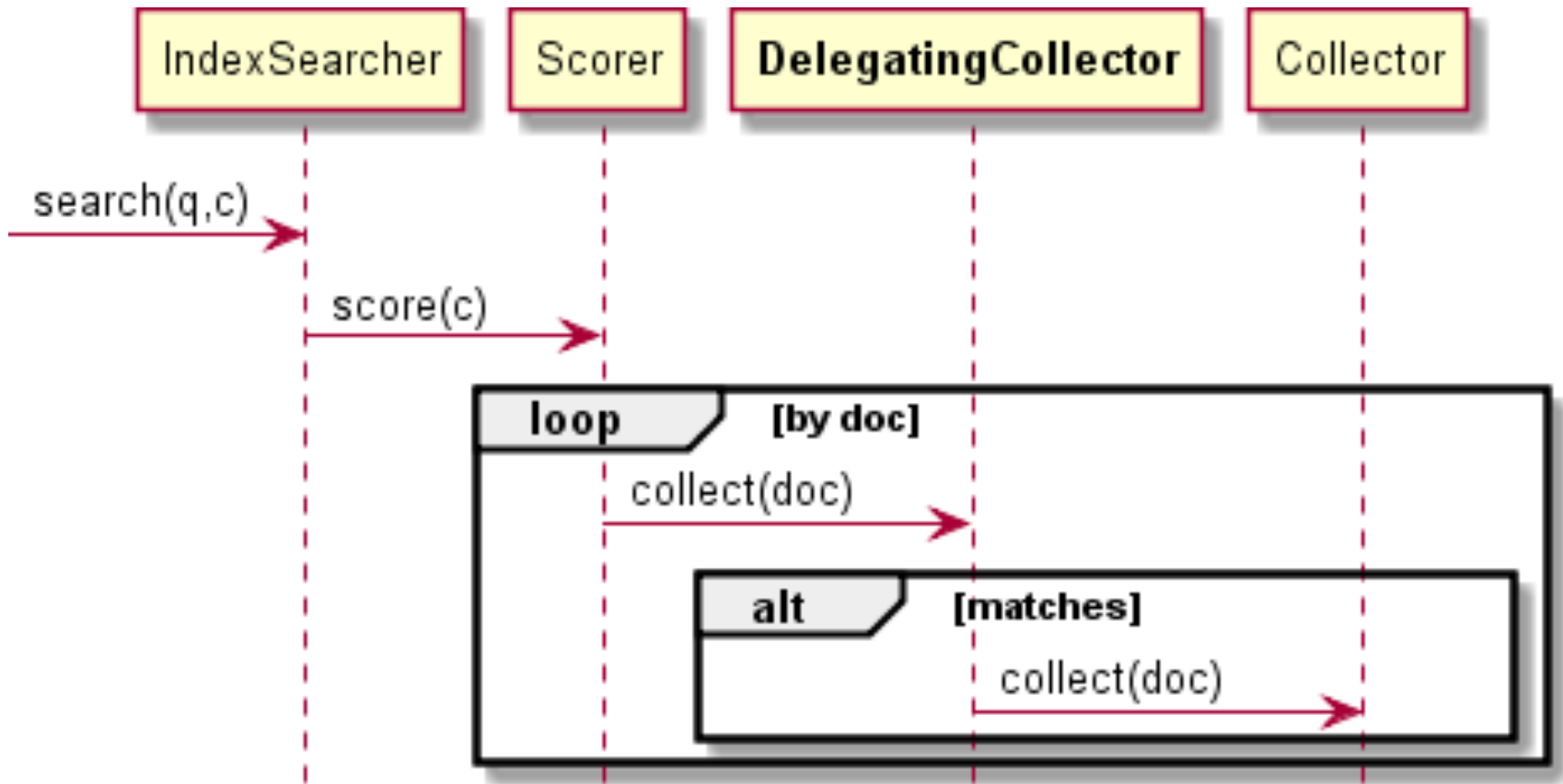
Lucene Search



<http://www.plantuml.com:>

[80/plantuml/img/YzQrK_3Clqcj24vDB4hEI2soKYW6ipGAIrEBAAdMlytDoKfDByeiqURbGr2a0DGOdvnS1MK0j88O6ijgS_Fm2XQHAXPJyP5ukg3gW3cUO4g0cg78De1AWXjly55wkZCoKl8Mq_FolfFmA0000](http://www.plantuml.com/img/YzQrK_3Clqcj24vDB4hEI2soKYW6ipGAIrEBAAdMlytDoKfDByeiqURbGr2a0DGOdvnS1MK0j88O6ijgS_Fm2XQHAXPJyP5ukg3gW3cUO4g0cg78De1AWXjly55wkZCoKl8Mq_FolfFmA0000)

PostFilter



<http://www.plantuml.com>:

[80/plantuml/img/T0un3i8m34NtlBc3rQb887PKTO65kIDJufWsKeXBwg7S7XUw89N5UjB_zlFjqC8r1vexSWK7AYUOFbG_zxYpvYS3BNV8HKbh2zHB9p6Fu5yG6AqPNP68SNHPe6fyUw54IPEO-pEdH2XScgDIAt0JR6JMh09LRSJgnl_NQbEtIm8F9_hTPCtMWjxvlqqvh6E8WU3EB9xdDm00](http://www.plantuml.com/img/T0un3i8m34NtlBc3rQb887PKTO65kIDJufWsKeXBwg7S7XUw89N5UjB_zlFjqC8r1vexSWK7AYUOFbG_zxYpvYS3BNV8HKbh2zHB9p6Fu5yG6AqPNP68SNHPe6fyUw54IPEO-pEdH2XScgDIAt0JR6JMh09LRSJgnl_NQbEtIm8F9_hTPCtMWjxvlqqvh6E8WU3EB9xdDm00)

ExternalFileField + FuncRangeQuery

..&fq={!frange l=1}AVAILABILITY&...

external_AVAILABILITY.20121108.txt

424=0

125=1

672=0

18=0

FileFloatSource\$ReloadCacheRequestHandler

Implementation

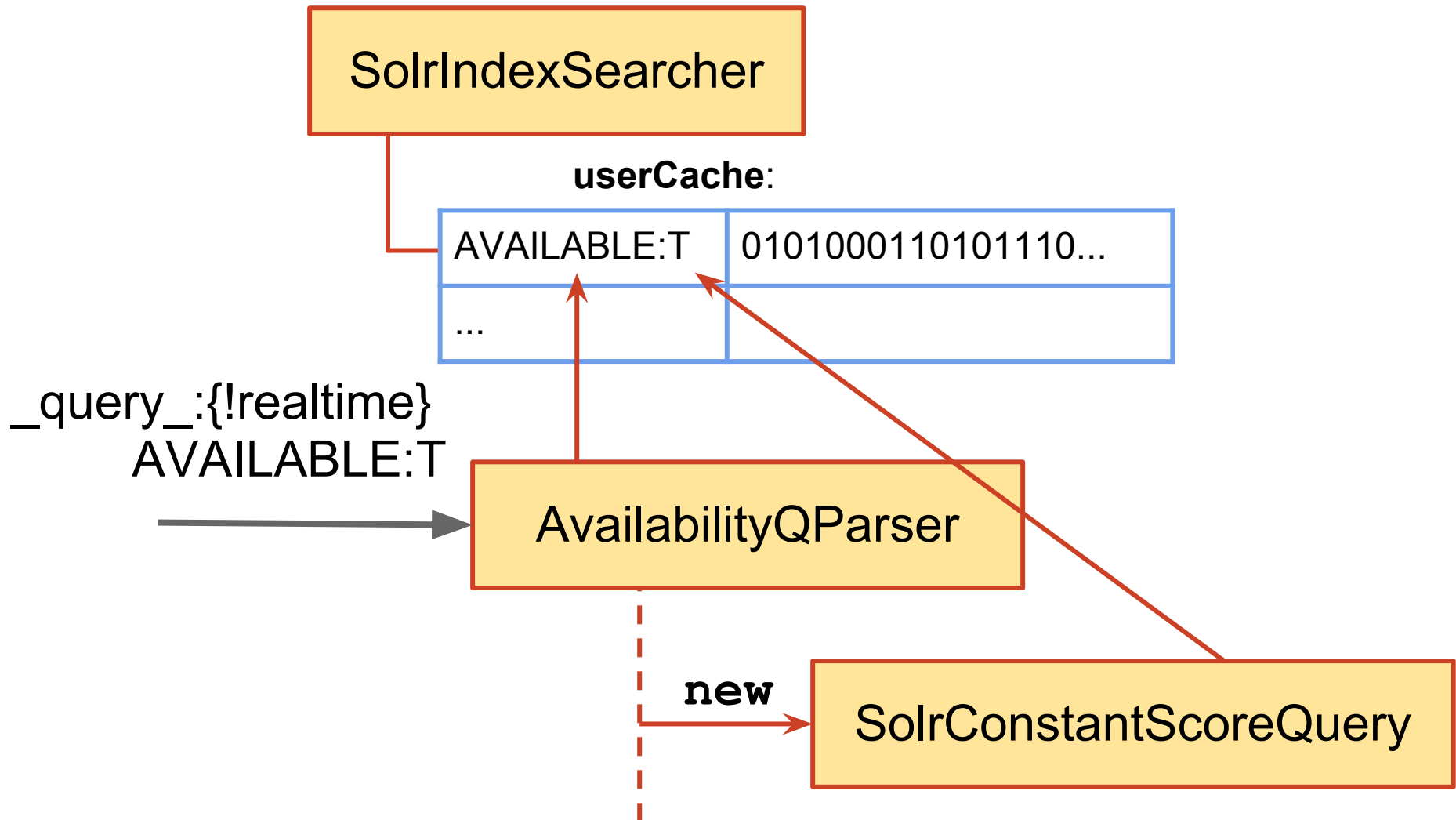


QParser + Nested Query

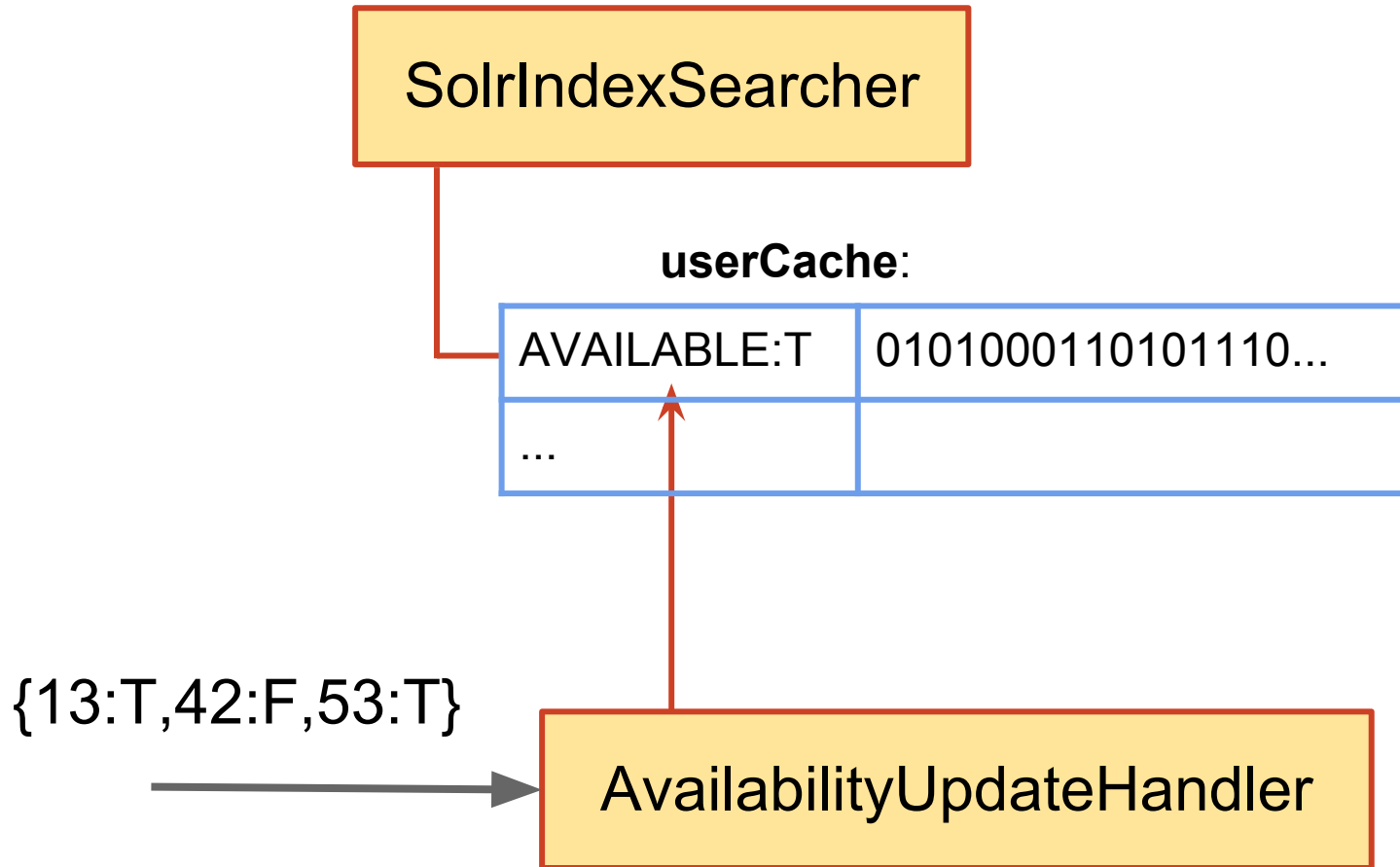
q=short sleeve polo AND

query:"{!realtime}AVAILABILITY:T"

Query



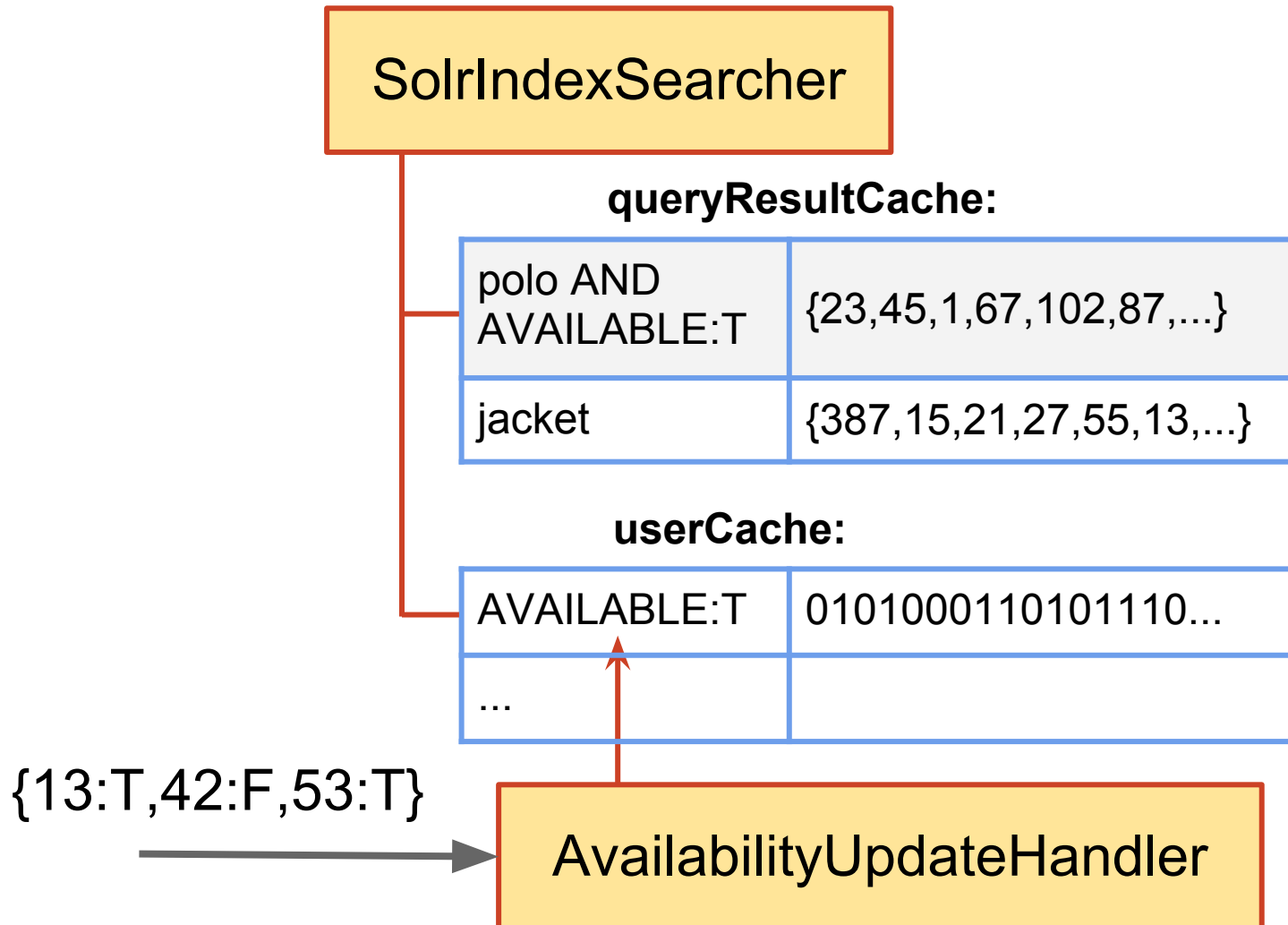
Update



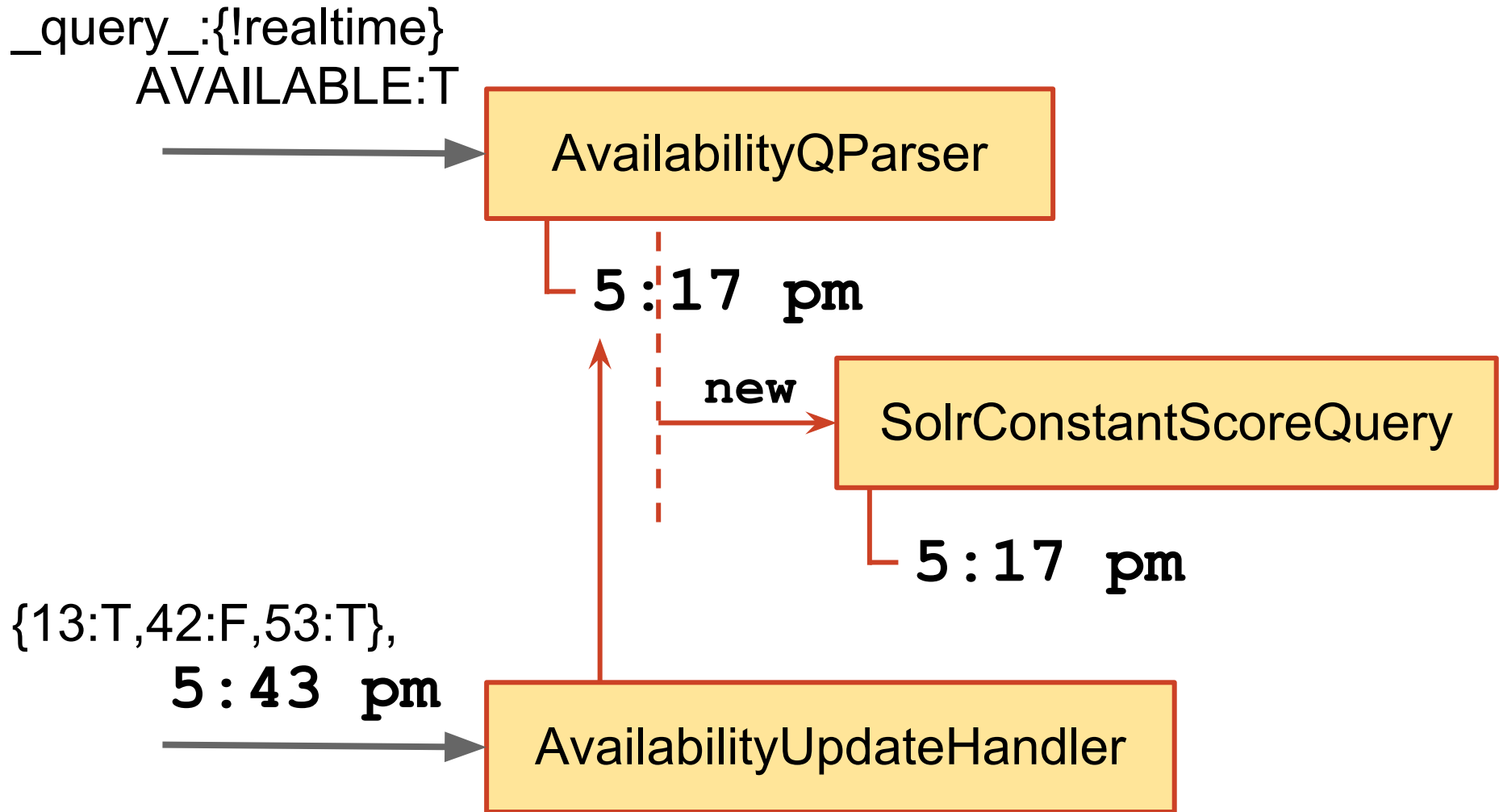
Update Message

```
{ query : "AVAILABILITY:T",  
  lookup-field: "id",  
  updates: {  
    3 : T,  
    25 : F, ...  
  }  
  timestamp: 1351978373455L  
}
```

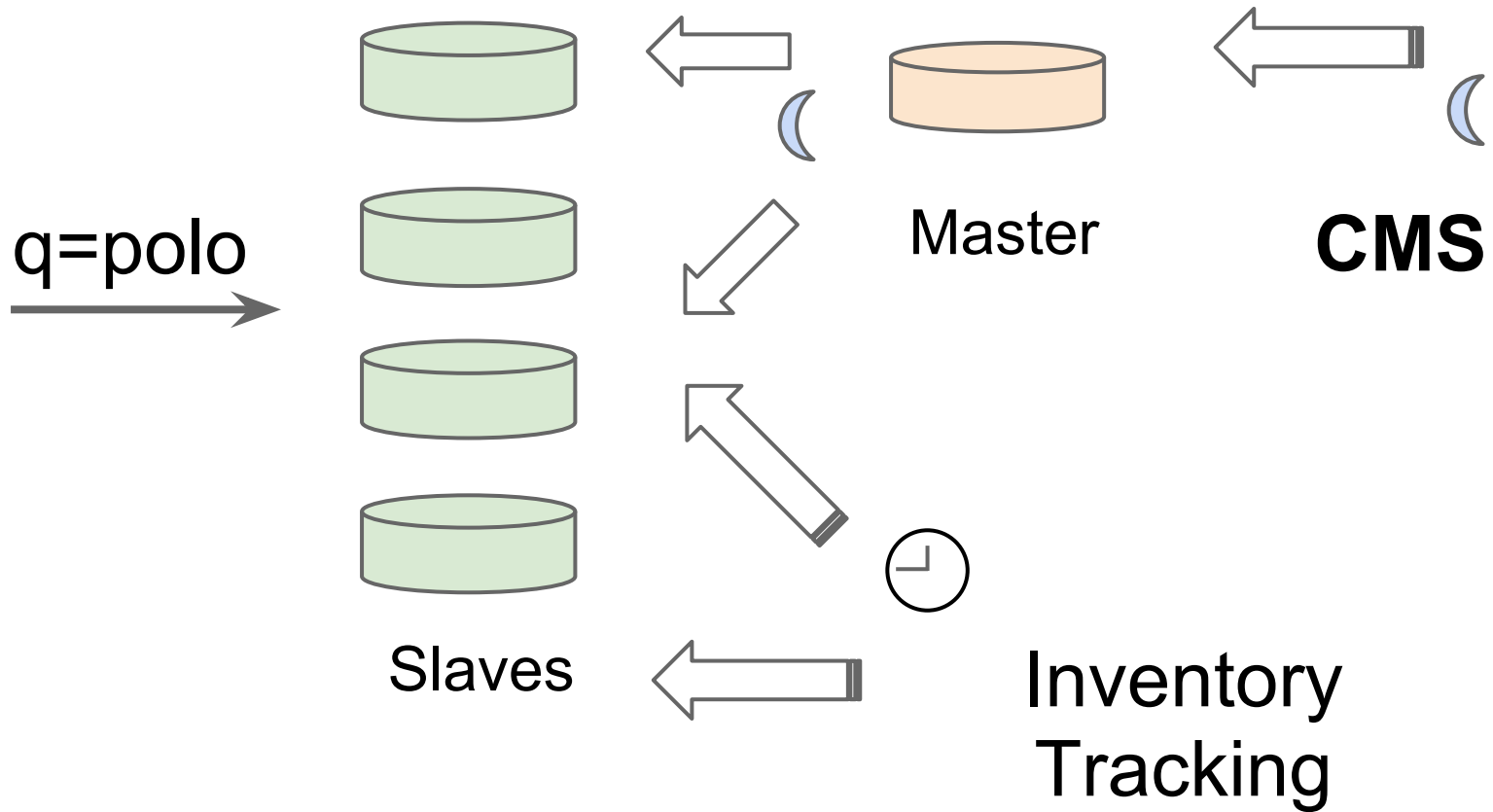
Stale queryQueryResult Cache



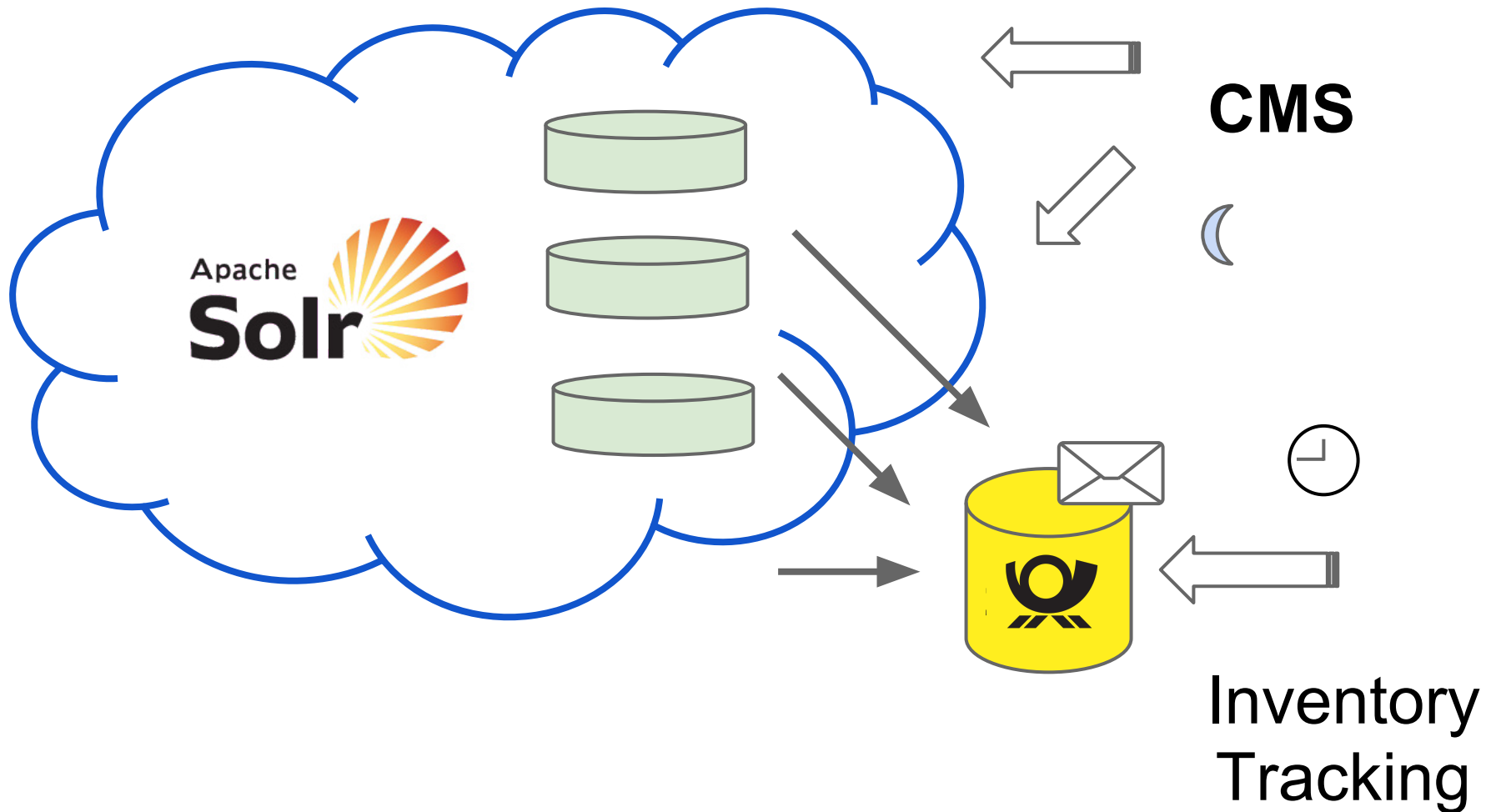
Stamp Time on It!



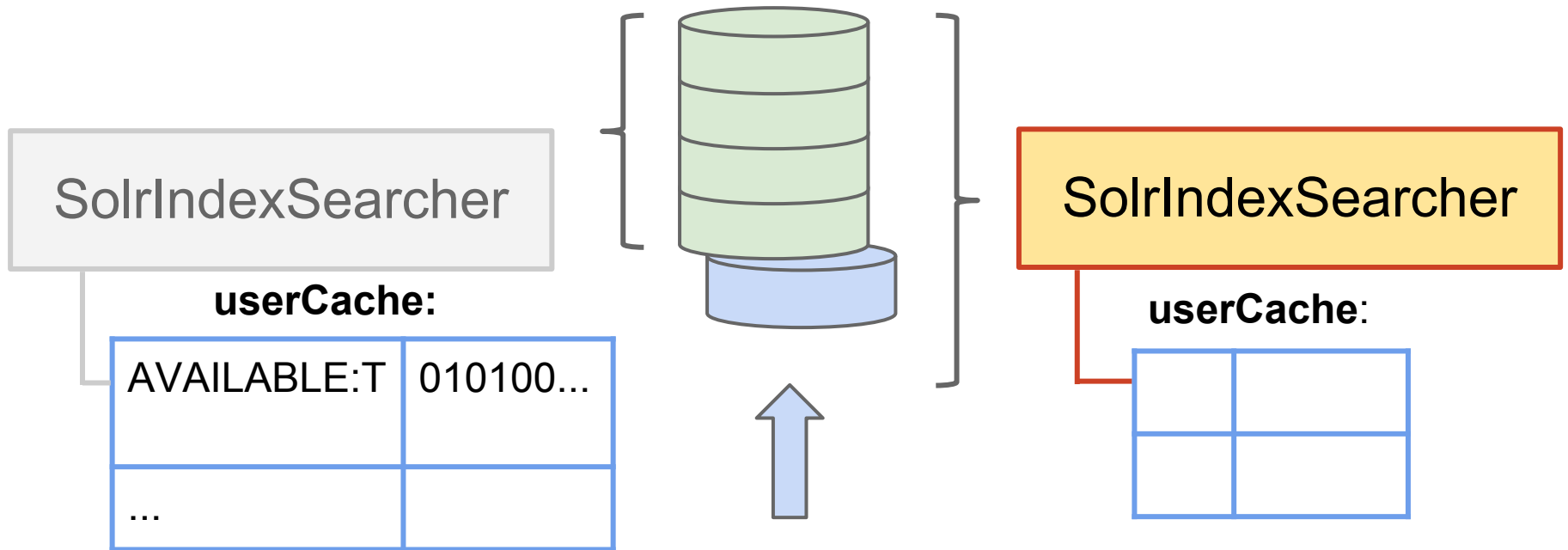
Updates Broadcasting



JMS Topic



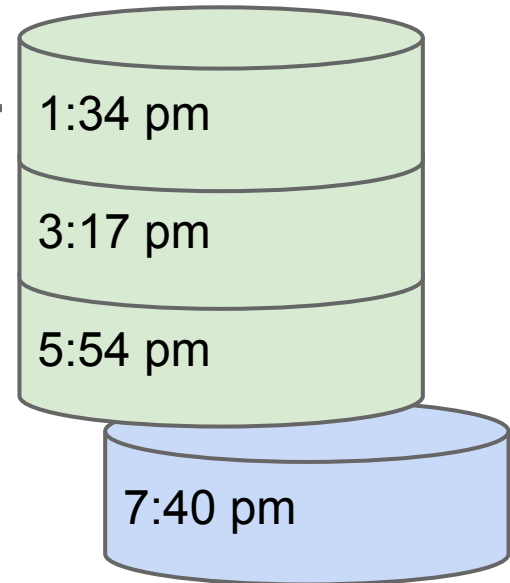
Oops! New Segment Arrives



Replay Update Messages

`availability-updates.log`

```
1:15 pm 3:T, 56:F, 73:T..  
1:25 pm 13:F, 47:T..  
1:42 pm 2:F, 6:F, 23:T..  
1:57 pm 5:T, 19:T, 79:F..  
2:05 pm 7:T, 12:T, 47:F..  
2:12 pm 78:T, 241:T..  
2:23 pm 34:T..  
...
```



If you need to bounce

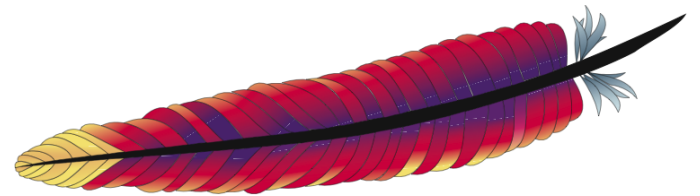
Updates Replay

+

Durable Subscription

OR

Apache Kaffka



Wrap-Up

- frequent index updates
- filter for available products
- transient availability
- Solr Nuts&Bolts
- query timestamp
- updates broadcasting
- updates replaying
- durable subscription



<http://goo.gl/0Ppem>