



# How to tackle performance issues when implementing high traffic multi-language search engine with Solr/Lucene

André Bois-Crettez  
Software Architect

Anca Kopetz  
Software Engineer

**Berlin Buzzwords 2014**

# Outline



Context and setup



Benchmarks

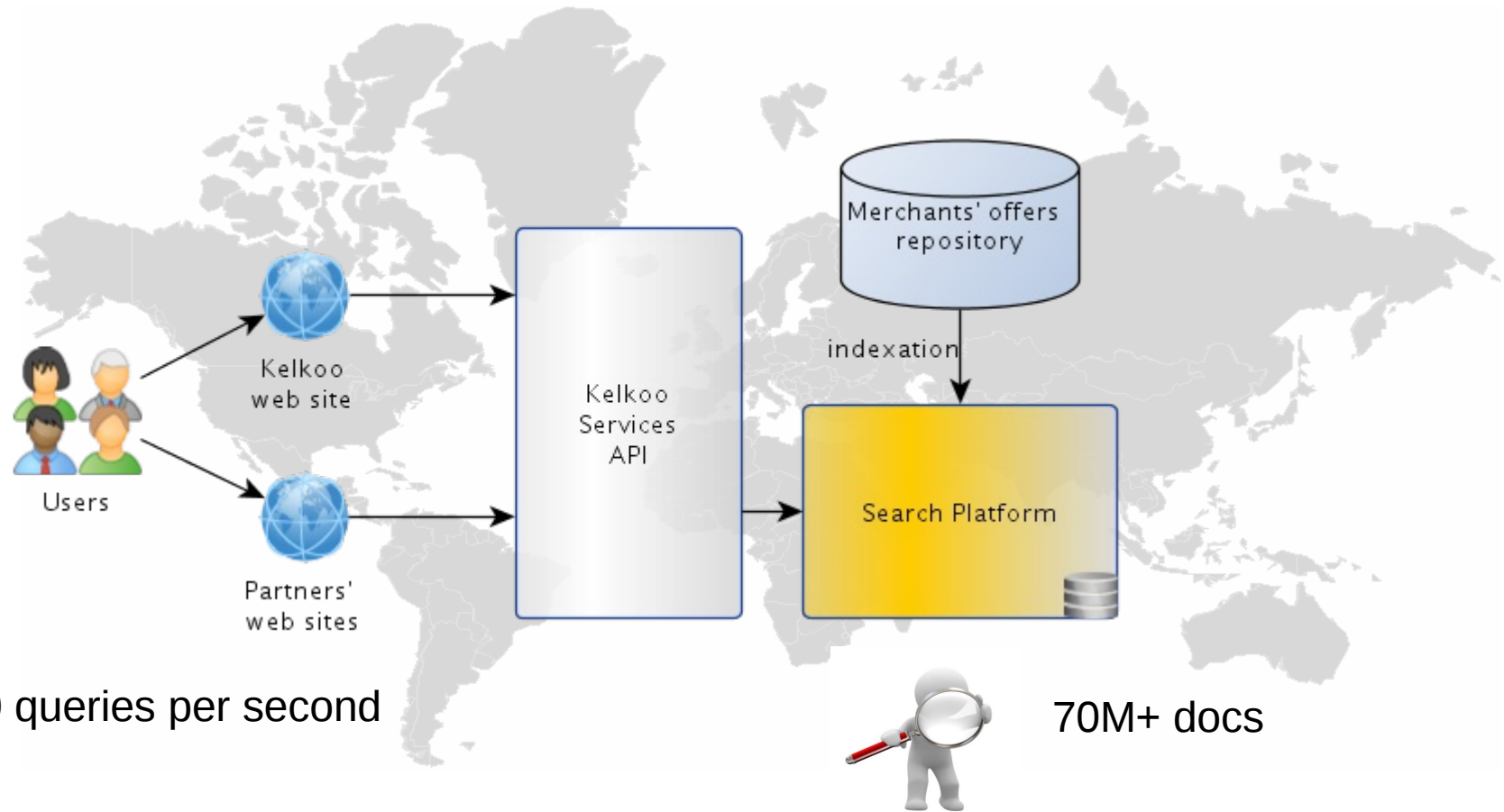


Performance solutions



Production

# Kelkoo shopping platform



3000 queries per second

70M+ docs

12 countries



# “Appareils à fondue en céramique”

Nous sommes une plate-forme permettant aux internautes de comparer les offres de nos partenaires e-marchands. Les résultats affichés proviennent des informations fournies par nos partenaires et ne reflètent donc pas l'intégralité des offres disponibles sur le marché.

58368 Résultats sponsorisés chez 395 marchands partenaires

Trier par **Pertinence**

Catégorie

















Prix

Marchand

- Priceminister Neuf (31579)
- Amazon.fr (9418)
- GSM55.com (8108)
- PriceMinister Occasions (528)
- Delamaison.fr\_\_ (525)
- Arthierro (406)
- Plus...

Recherches associées

**Facets**

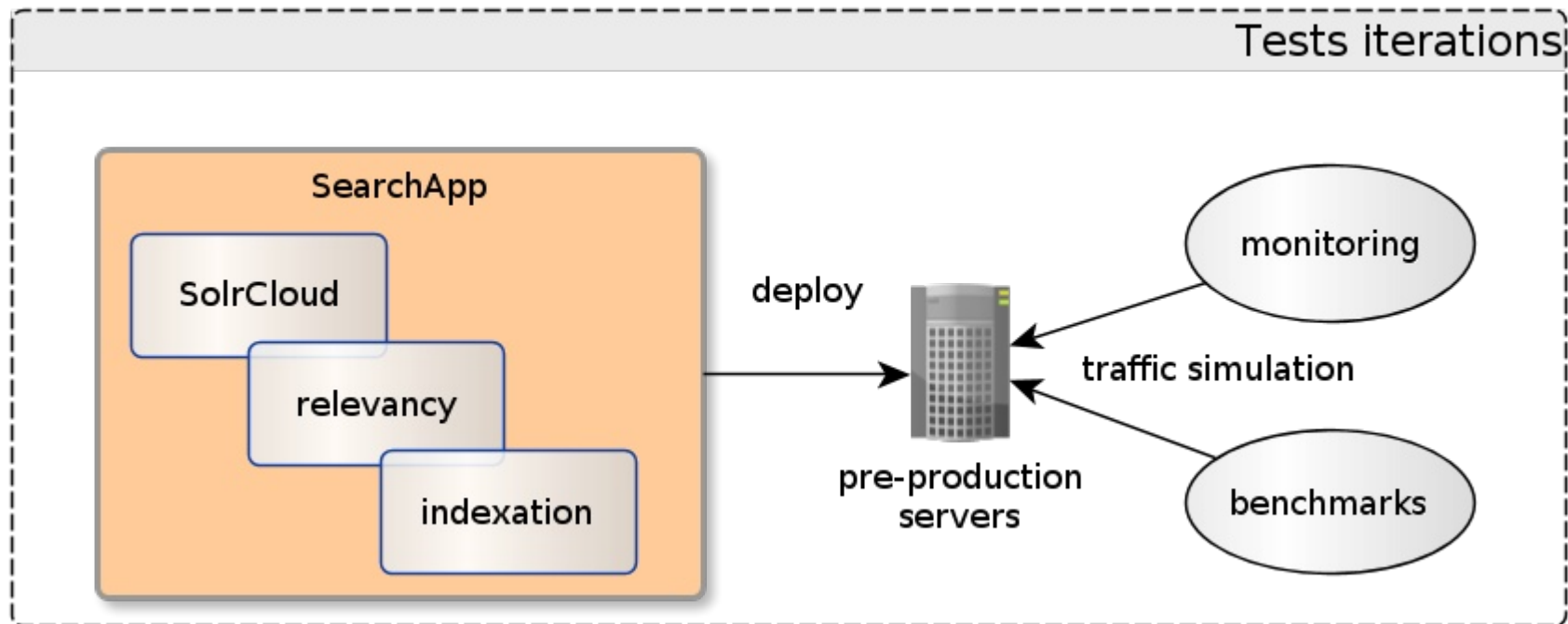
 <p>Küchenprofi 2125102200 St.Gallen Appareil à Fondue <b>Amazon.fr</b> <b>30,88 €</b> Port: Gratuit • 30,88 € <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>	 <p>Tefal EF351412 SIMPLY INVENTS Fondue <b>Darty.com</b> <b>66,39 €</b> Port: Gratuit • 66,39 € <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>	 <p>Küchenprofi 2125101000 Appenzell Appareil à Fondue <b>Amazon.fr</b> <b>50,91 €</b> Port: Gratuit • 50,91 € <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>	 <p>Kela Manège à fondue pour 6 personnes en inox et bois <b>Delamaison.fr__</b> <b>119,50 €</b> Port: Gratuit • 119,50 € <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>	 <p>Emile HENRY Service à fondue en céramique - D. 24 <b>Delamaison.fr__</b> <b>90,50 €</b> Port: Gratuit • 90,50 € <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>	 <p>Kela Manège à fondue pour 6 personnes en inox avec <b>Delamaison.fr__</b> <b>87,50 €</b> Port: Gratuit • 87,50 € <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>
 <p>Kela Manège à fondue pour 6 personnes en inox et coupelle <b>Delamaison.fr__</b> <b>79,50 €</b> Port: Gratuit • 79,50 € <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>	 <p>Kela Mini service à fondue chocolat en céramique spécial <b>Delamaison.fr__</b> <b>14,90 €</b> Port: 3,90 € • 18,80 € TFC <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>	 <p>Kela Mini service à fondue chocolat pour 4 personnes en <b>Delamaison.fr__</b> <b>19,90 €</b> Port: 4,30 € • 24,20 € TFC <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>	 <p>Emile HENRY Service à fondue en céramique - D. 24 <b>Delamaison.fr__</b> <b>90,50 €</b> Port: Gratuit • 90,50 € <a href="#">Voir l'offre</a> Offre valable aujourd'hui</p>		
					

# Search engines at Kelkoo

- Past : Yahoo vertical search engine
- **2013: New implementation using Solr 4.x**
  - Targetting Christmas high traffic
  - Without NRT indexation

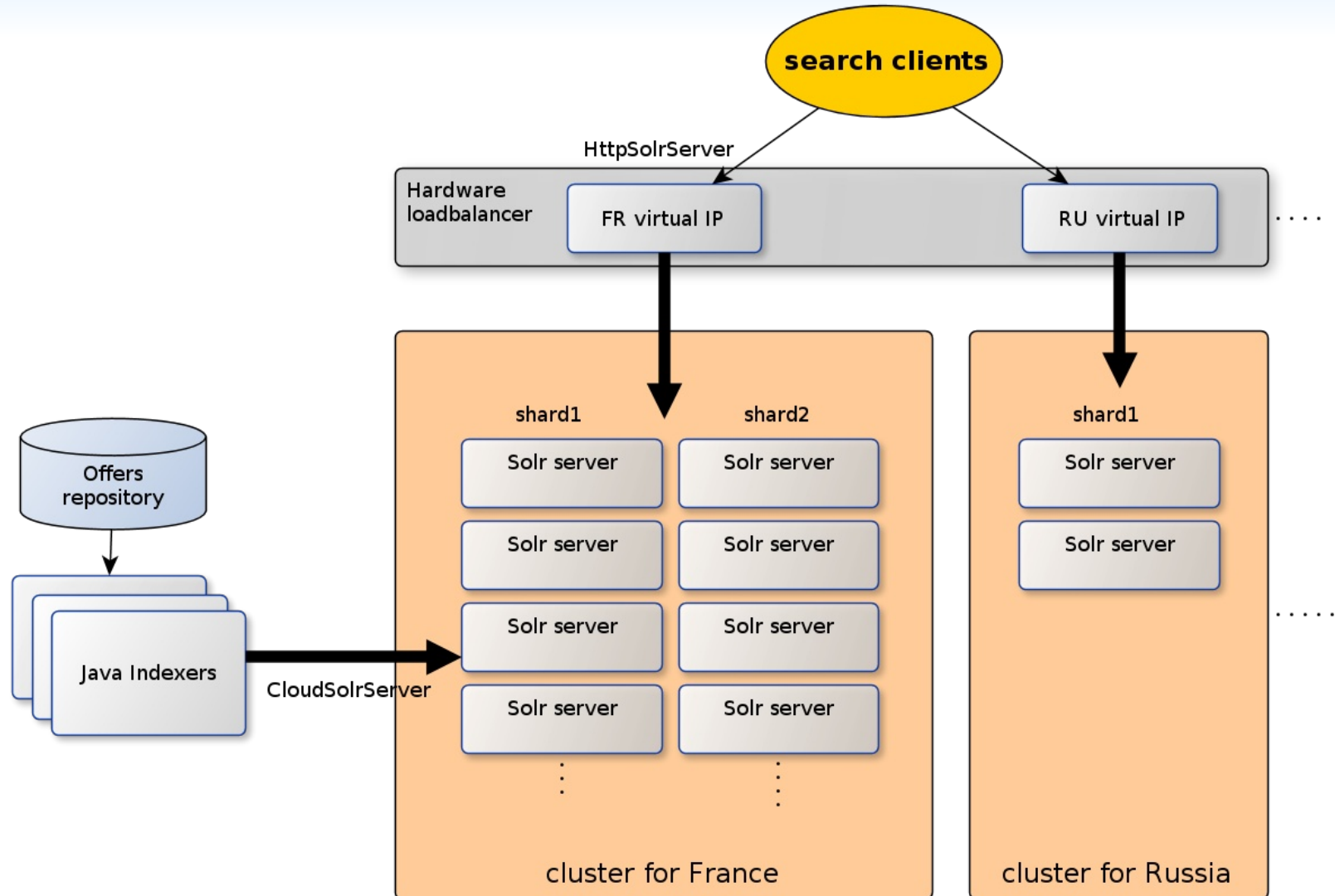
# Before performance evaluation

- Target for biggest country cluster :
  - 600+ qps, avg qtime ~200ms
  - 12 servers, 10M+ docs



Frequent interaction with SOLR community

# Cluster architecture



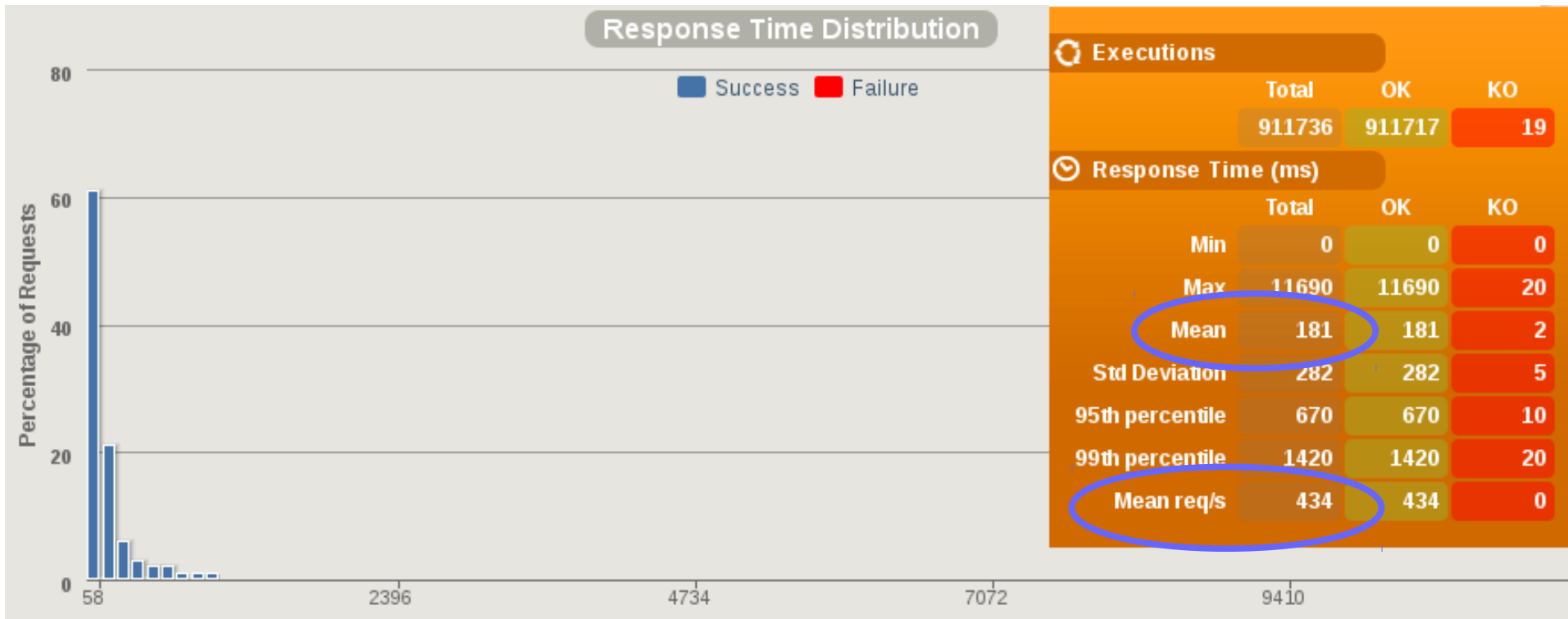
# Benchmarks



- Number of users in parallel : impact on CPU load
- Duration of test : important in order to have relevant results

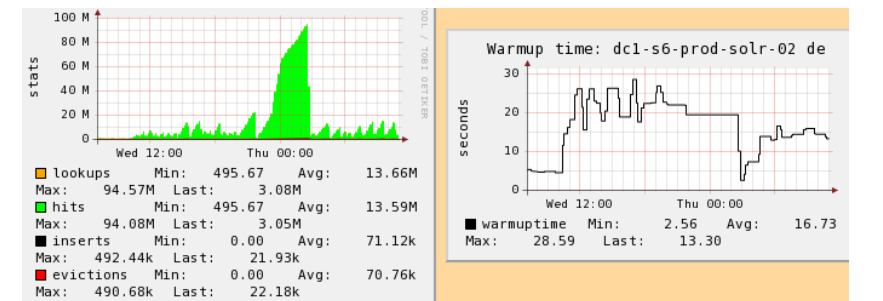
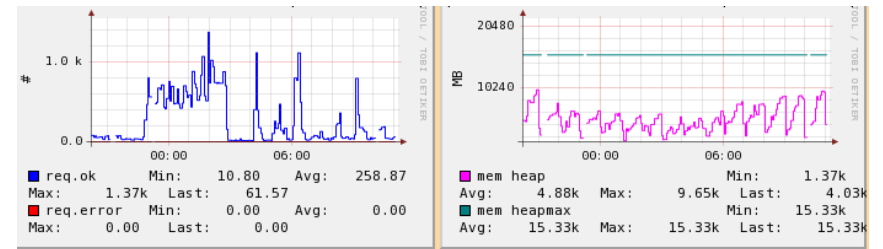
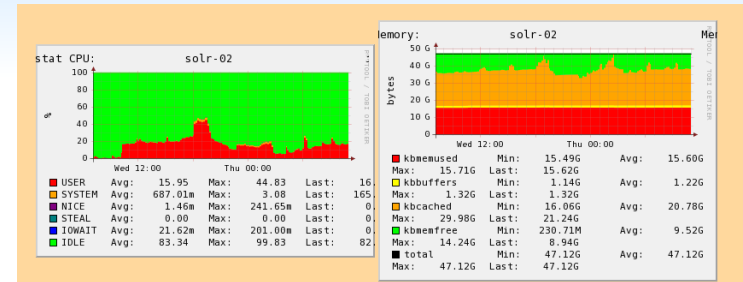


# Gatling report



# Monitored resources

- Target
  - Queries per second, response time
- System
  - Memory and JVM heap
  - CPU load
  - I/O disk
- Solr
  - Caches (hitratio, evictions)
  - Warmup time
  - Number of index segments



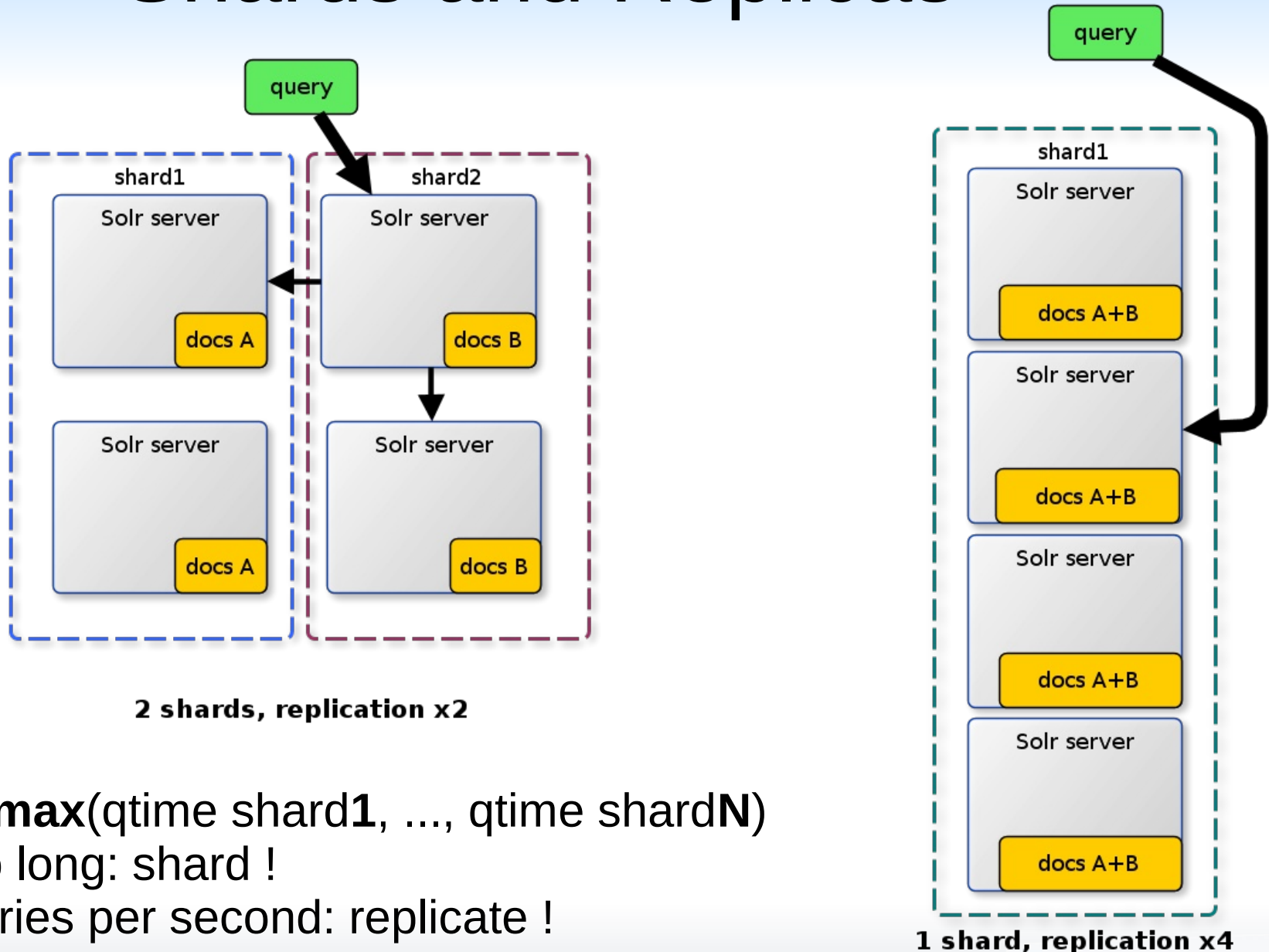
# Launch benchmarks



## Benchmark results:

- Query performance evaluation
- Explore ways to improve it

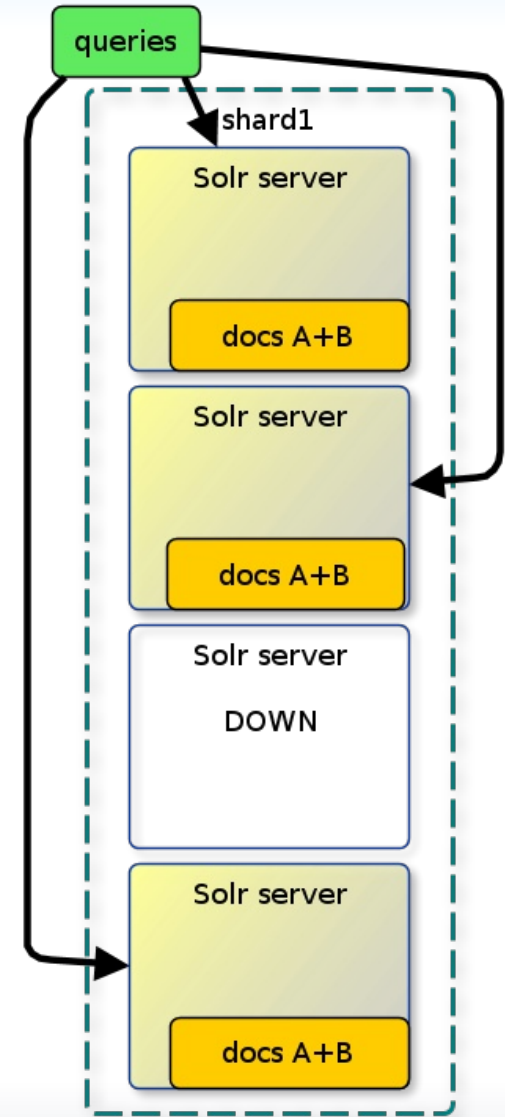
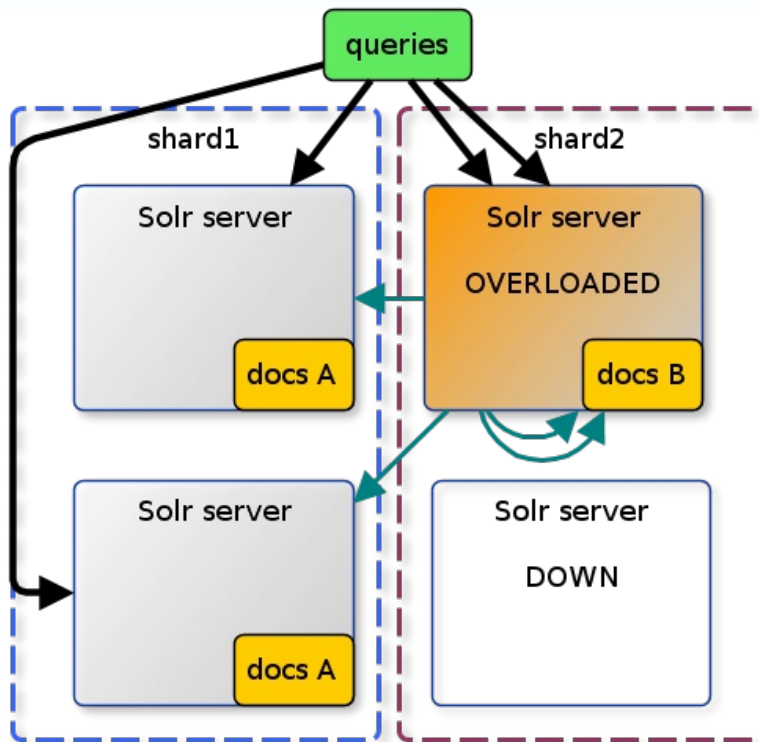
# Shards and Replicas



- QTime  $\approx \max(\text{qtime shard1}, \dots, \text{qtime shardN})$
- Qtime too long: shard !
- More queries per second: replicate !

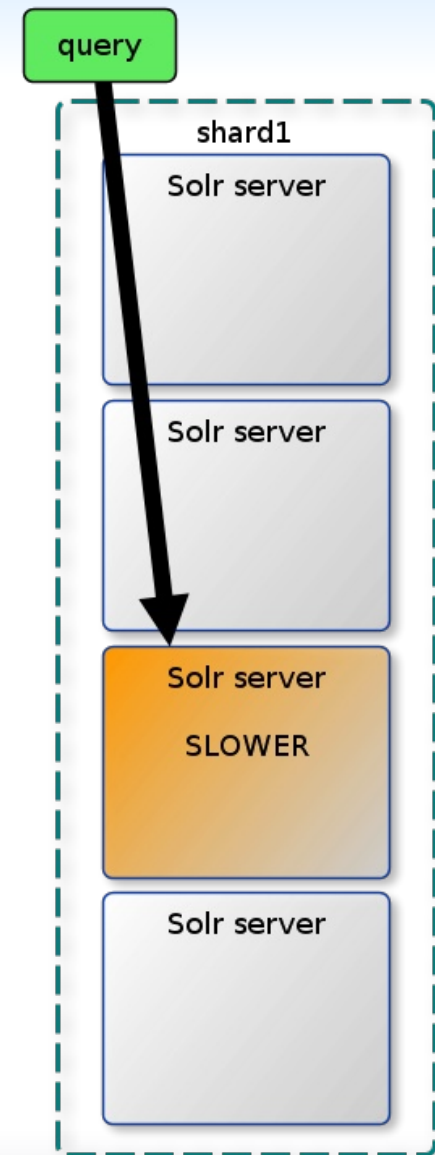
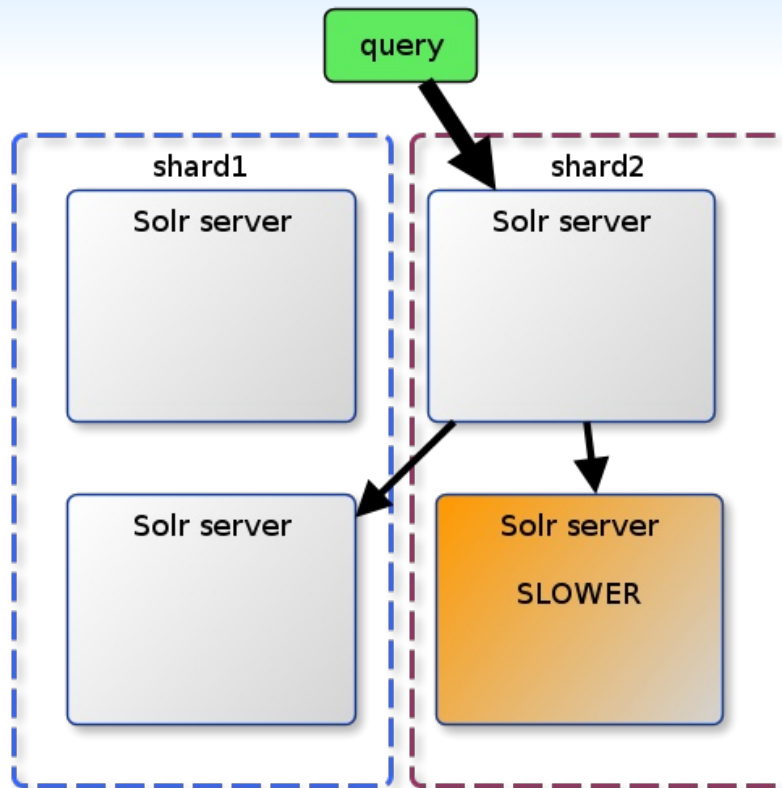
Performance solutions

# Fault tolerance



- Need resilience: replicate !
- Performance drops faster when sharding

# Heterogeneous hardware

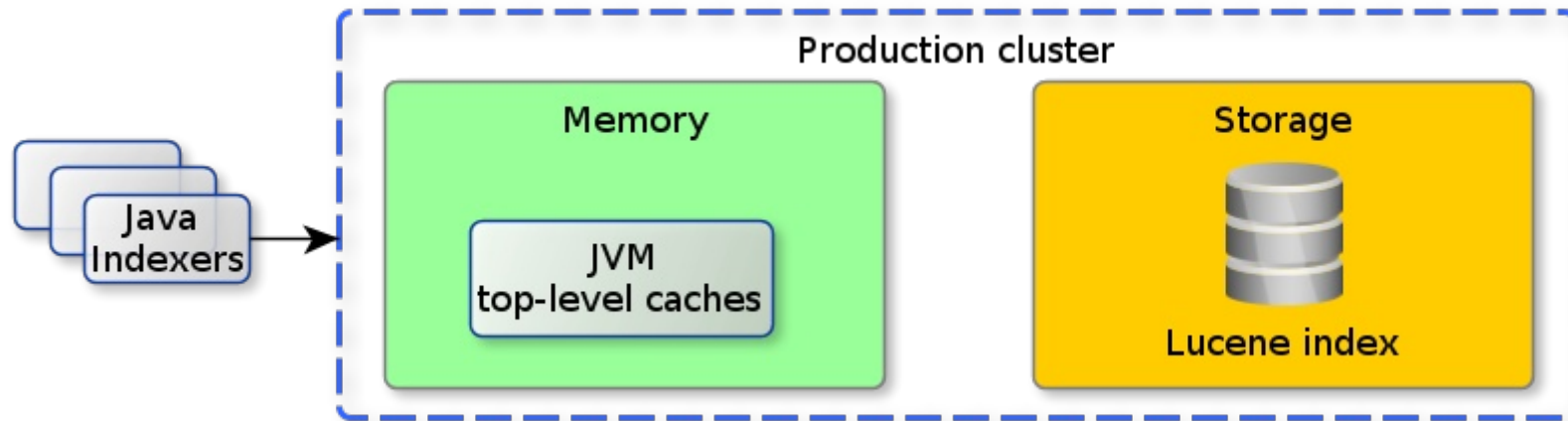


Slowest server is **the** bottleneck !

- **ALL** other servers will be underused
- Fun fact: some servers had 15% higher CPU usage due to “energy saving” mode setting in BIOS !

Performance solutions

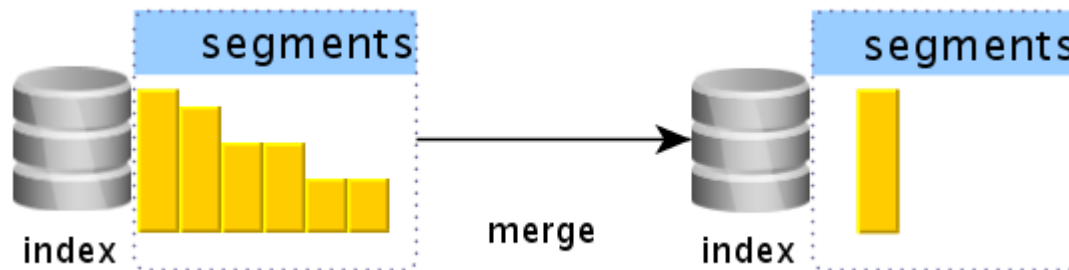
# Indexation



Visibility: soft-commit	Durability: hard-commit
openSearcher = true	openSearcher = false
commitWithin = 30min	autoCommit = 15min
tlogs not truncated	tlogs truncated
⚠ caches flushed & auto-warmed	segment merges initiated

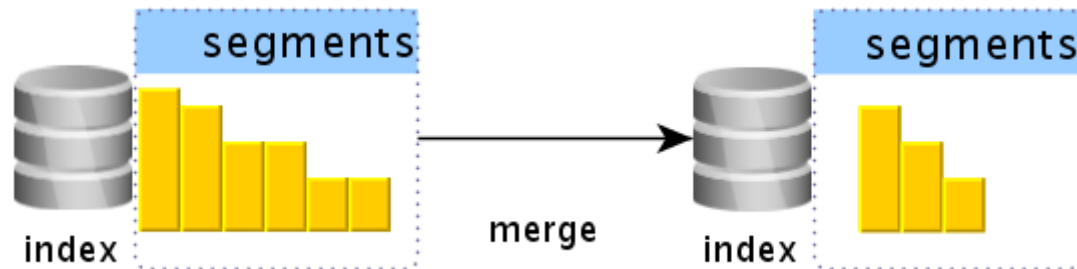
# Merge policy

- Reducing the number of index segments impacts the query performance
- ~~Optimize index~~ : short-term benefits





# Aggressive merge policy



```
<mergePolicy class="org.apache.lucene.index.TieredMergePolicy">  
  <int name="maxMergeAtOnce">6</int> <!-- default : 10 -->  
  <int name="segmentsPerTier">3</int> <!-- default : 10 -->  
  <double name="reclaimDeletesWeight">5</double>  
</mergePolicy>
```

# Search Caches

- *MMapDirectory* Lucene files: OS cache in RAM
  - Less I/O wait, without increasing JVM memory used
  - With enough RAM, SSD not useful for us
- *DocumentCache*: disabled
  - MMapDirectory efficient enough to load stored fields
- *QueryResultCache*: very small
  - Already a cache above Kelkoo Search API

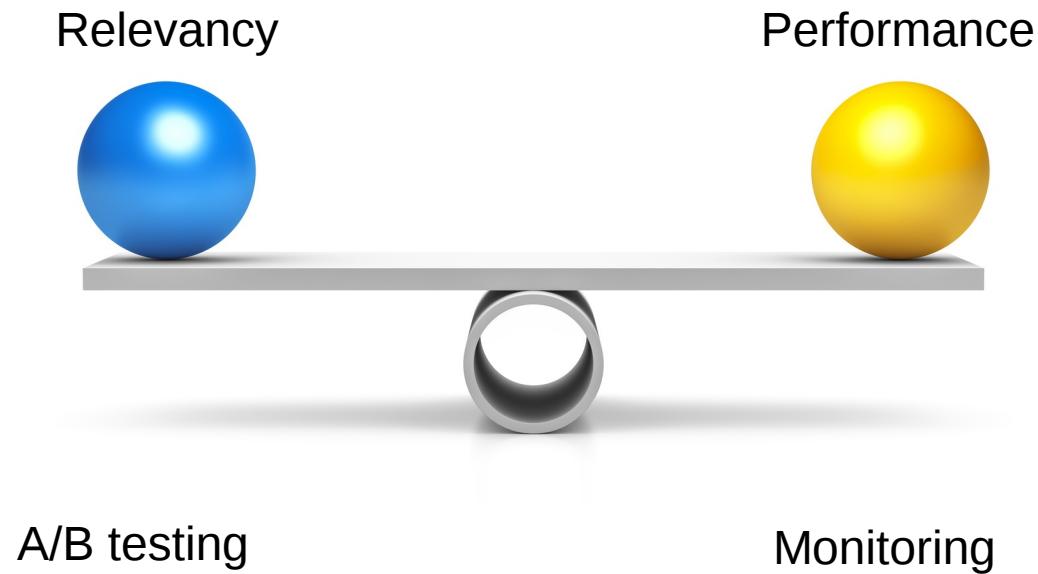
# Search Caches

&q=iphone  
&fq=merchantid:112233  
&facet.field=categoryid  
&facet.field=feature\_color

- Filters : **filterCache**
- Facet on single-valued field (*categoryid*)
  - *facet.method=fc* : **fieldCache**
- Facet on dynamic multi-valued fields (*feature\_\**)
  - ✗ *facet.method=fc* : fieldValueCache
    - Higher memory usage per field
    - Not good for low cardinality fields
  - ✓ *facet.method=enum* : **filterCache**
    - Optimal memory usage for lots of dynamic fields
    - Adequate processing time

# Query features

*Get a good balance between ...*



“Leather accessories for iphone5”



leather accessory iphone 5

8965 Results from 8 retailers

**Category**

- Mobile Phone Accessories (8943)
- MP3 Player Accessories (6)
- Digital Camera Accessories (3)
- Toys (1)
- Home Telephone Accessories (1)

**Price**

**Merchant**

**Related Searches**

<p>Vandot Poposh Phone Mobile Accessory 3in1 For Apple Amazon Marketplace <b>£9.99</b> inc P&amp;P: £12.98 <a href="#">Visit Store</a></p>	<p>Yosave Accessories PU Leather Wallet Case for iPhone Amazon UK <b>£7.01</b> inc P&amp;P: £11 <a href="#">Visit Store</a></p>	<p>Aspinal of London Leather iPhone 5 Case John Lewis <b>£45</b> inc P&amp;P: £54.95 <a href="#">Visit Store</a></p>	<p>Hugo Boss Booklet Gracious Leather iPhone 5/5S Case Shop Direct Group Limited <b>£45</b> inc P&amp;P: £48.95 <a href="#">Visit Store</a></p>	<p>Yosave Accessories PU Leather Flip Case for iPhone 5 Amazon Marketplace <b>£4.99</b> inc P&amp;P: £5.98 <a href="#">Visit Store</a></p>	<p>Slick Accessories New! Faux Leather Wallet Case for iPhone Amazon Marketplace <b>£6.99</b> Free shipping <a href="#">Visit Store</a></p>
<p>Yosave Accessories PU Leather Wallet Case for iPhone Amazon Marketplace <b>£4.90</b> inc P&amp;P: £5.89 <a href="#">Visit Store</a></p>	<p>Yosave Accessories PU Leather Wallet Case for iPhone Amazon Marketplace <b>£5.90</b> inc P&amp;P: £6.89 <a href="#">Visit Store</a></p>	<p>Yosave Accessories PU Leather Wallet Case for iPhone Amazon UK <b>£8.40</b> inc P&amp;P: £12.39 <a href="#">Visit Store</a></p>	<p>Yosave Accessories PU Leather Wallet Case for iPhone Amazon Marketplace <b>£5.90</b> inc P&amp;P: £6.89 <a href="#">Visit Store</a></p>		
<p>Yosave Accessories PU Leather Wallet Case for iPhone Amazon Marketplace <b>£4.90</b> inc P&amp;P: £5.89 <a href="#">Visit Store</a></p>	<p>Slick Accessories New! Faux Leather Wallet Case for iPhone Amazon Marketplace <b>£6.99</b> Free shipping <a href="#">Visit Store</a></p>	<p>Yosave Accessories PU Leather Flip Case for iPhone 5 Amazon Marketplace <b>£4.99</b> inc P&amp;P: £5.98 <a href="#">Visit Store</a></p>	<p>Accessories4electronics PU Leather Case for iPhone 5 Amazon Marketplace <b>£5.99</b> inc P&amp;P: £6.89 <a href="#">Visit Store</a></p>	<p>Yosave Accessories PU Leather Wallet Case for iPhone Amazon Marketplace <b>£4.99</b> inc P&amp;P: £5.98 <a href="#">Visit Store</a></p>	<p>Yosave Accessories PU Leather Wallet Case for iPhone Amazon Marketplace <b>£4.99</b> inc P&amp;P: £5.98 <a href="#">Visit Store</a></p>

- AND/OR
- facets
- group by merchant

# Improvements during benchmarks

- SolrCloud architecture
- Indexation policy
- Merge policy
- Search caches
- Solr features

# Production

- OutOfMemory due to filterCache
  - Production behaviour different from benchmarks
  - Use of -Xms30G -Xmx30G, reduce number of filterCache entries
- GC params :  
[http://wiki.apache.org/solr/ShawnHeisey#GC\\_Tuning](http://wiki.apache.org/solr/ShawnHeisey#GC_Tuning)
  - Concurrent Mark and Sweep is important for Solr
- Servers in recovery: impact on performance
- Updated Solr to latest versions

# Conclusions

- Success story with Lucene/Solr
  - 1 year, 3000 qps peaks, 70M+ docs
- Benchmarks vs. Production
- Monitor resources
- Identify the bad guy and deal with him!
- New ideas for performance optimization



# Questions ?

- Search ? Big Data ? Join us !  
<http://jobs.kelkoo.fr/>



André Bois-Crettez <andre.bois@kelkoo.com>

Anca Kopetz <anca.kopetz@kelkoo.com>

**kelkoo**