Hibernate Search

Emmanuel Bernard
JBoss, a Division of Red Hat

Doer

Copyright Red Hat Middleware 2008

Search: left over of today's apps

Add a search dimension to your persistent domain model

"Frankly, search sucks on this project"

-- Anonymous' boss

Why? How to fix that? For cheap?

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Integrate full-text search and persistent domain

- SQL Search vs Full-text Search
- Object model / Full-text Search mismatches
- *Demo*
- Hibernate Search architecture
- Configuration and mapping
- Full-text based object queries
- Some more features

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

SQL search limits

- Wildcard / word search
  - ‘%hibernate%’
- Approximation (or synonym)
  - ‘hybernat’
- Proximity
  - ‘Java’ close to ‘Persistence’
- Relevance or (result scoring)
- multi-”column” search

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Full Text Search

- Search information
  - by word
  - inverted indices (word frequency, position)

- In RDBMS engines
  - portability (proprietary add-on on top of SQL)
  - flexibility
- Standalone engine
  - Apache Lucene(tm) http://lucene.apache.org

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Mismatches with a domain model

- Structural mismatch
  - full text index are text only
  - no reference/association between document
- Synchronization mismatch
  - keeping index and database up to date
- Retrieval mismatch
  - the index does not store objects
  - certainly not managed objects

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

DEMO

Let’s google our application!

Copyright Red Hat Middleware 2008

Hibernate Search

- Under the Hibernate platform
  - LGPL
- Built on top of Hibernate Core
- Use Apache Lucene(tm) under the hood
  - In top 10 downloaded at Apache
  - Very powerful
  - Somewhat low level
  - easy to use it the “wrong” way
- Solve the mismatches

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Architecture

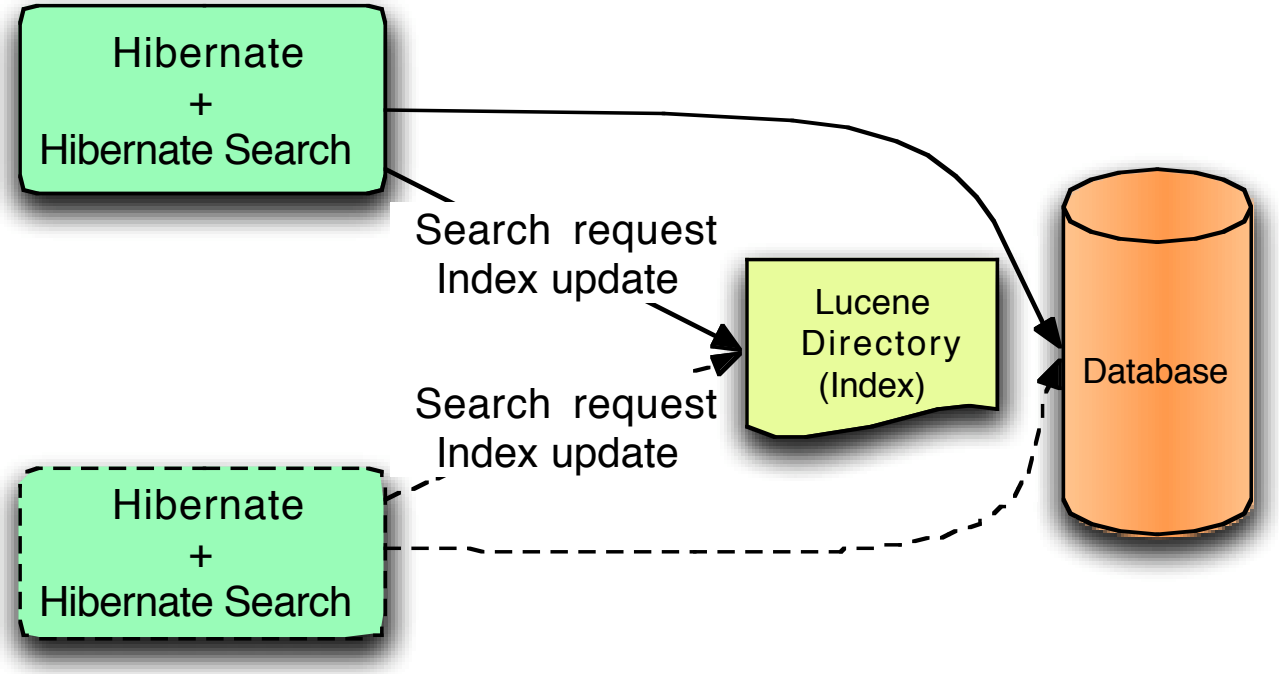
- Transparent indexing through event system (JPA)
  - PERSIST / UPDATE / DELETE
- Convert the object structure into Index structure
- Operation batching per transaction
  - better Lucene performance
  - “ACID”-ity
  - (pluggable scope)
- Backend
  - synchronous / asynchronous mode
  - Lucene, JMS

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Architecture (Backend)

- Lucene Directory
  - standalone or symmetric cluster
  - immediate visibility
  - Can affect front end runtime
  - no good cluster scalability

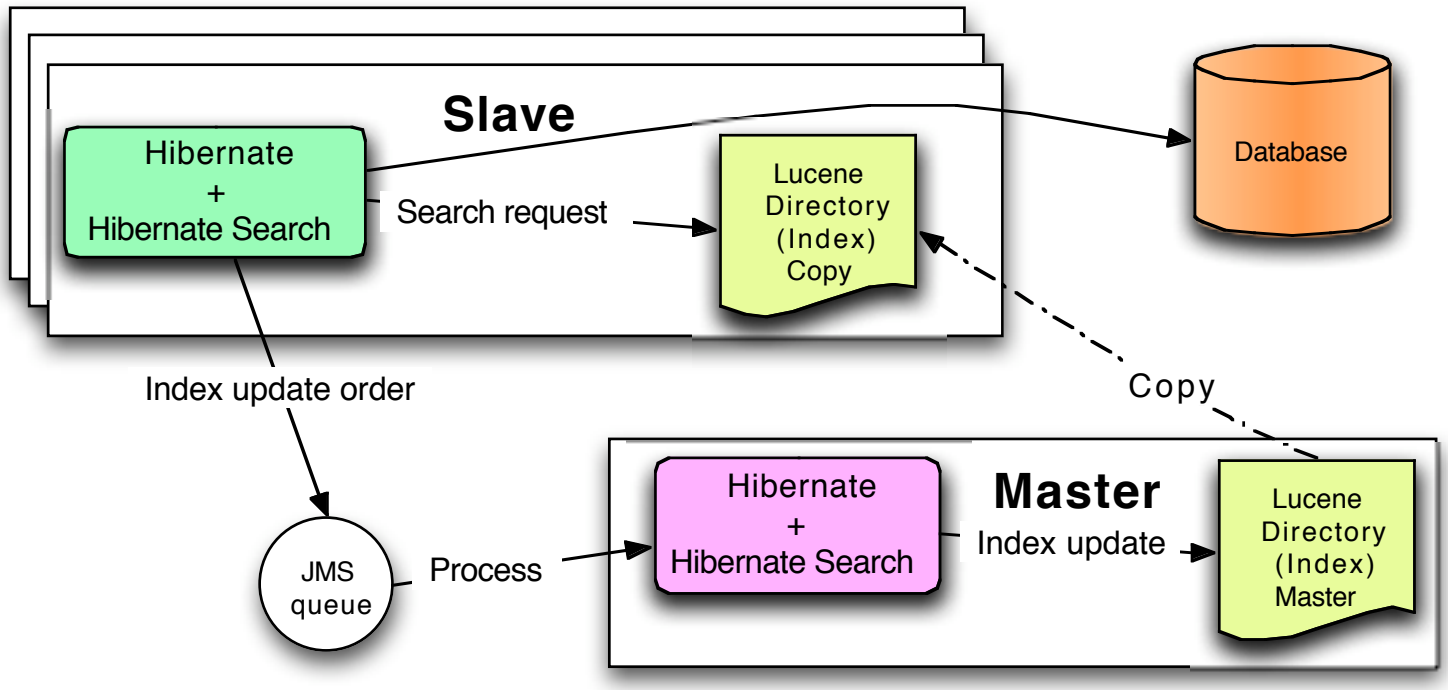


JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Architecture (Backend)

- JMS (Cluster)
  - Search processed locally
  - Changes sent to a master node (JMS)
  - asynchronous indexing (delay)
  - No front end extra cost / good cluster scalability



Configuration and Mapping

- Configuration
  - event listener wiring
    - transparent in Hibernate Annotations
  - Backend configuration

- Mapping: annotation based
  - @Indexed
  - @Field(store, index)
  - @IndexedEmbedded
  - @FieldBridge
  - @Boost / @Analyzer

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Query

- Retrieve objects, not documents
  - no boilerplate conversion code!
- Objects from the Persistence Context
  - same semantic as a JPA-QL or Criteria query
- Use org.hibernate.Query/javax.persistence.Query
  - common API for all your queries
  - pagination support
  - list / scroll / iterate support
- Query on correlated objects
  - “JOIN”-like query

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Query example

```
org.apache.lucene.search.Query luceneQuery;
String queryString = "summary:Festina Or brand:Seiko"
luceneQuery = parser.parse( queryString );

org.hibernate.Query fullTextQuery =
fullTextSession.createFullTextQuery( luceneQuery );

fullTextQuery.setMaxResult(200);

List result = fullTextQuery.list();
//return a list of managed objects

queryString = "title:hybernat~" //Approximate search
queryString = "\"Hibernate JBoss\"~10" //Proximity search
queryString = "author.address.city:Atlanta"
//correlated search
```


JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

More query: Lucene under the hood

- BooleanQuery (optional, must, must not)
- TermQuery
- RangeQuery
- PrefixQuery
- PhraseQuery
- WildcardQuery
- FuzzyQuery
- BoostingQuery

- and many more

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

More query features

- Fine grained fetching strategy
- Sort by property rather than relevance
- Projection (both field and metadata)
  - metadata: SCORE, ID, BOOST etc
  - no DB / Persistence Context access
- Filters
  - security / temporal data / category
- Total number of results

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Some more features

- Automatic index optimization
- Index sharding
- Manual indexing and purging
  - non event-based system
- Shared Lucene resources
- Native Lucene access

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Full-Text search without the hassle

- Transparent index synchronization
- Automatic Structural conversion through Mapping
- No paradigm shift when retrieving data

- Clustering capability out of the box

- Easier / transparent optimized Lucene use

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

Q&A

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008

For More Information

- Hibernate Search
  - http://search.hibernate.org
  - book at Manning (shhhh)
- JBoss Seam
  - http://seamframework.org
- Apache Lucene
  - http://lucene.apache.org
  - Lucene In Action
- http://in.relation.to
- http://parleys.com

JBoss WORLD ORLANDO 2008 PRESENTED BY RED HAT

Copyright Red Hat Middleware 2008