

JUDCon

JBoss Users & Developers Conference

Boston:2011

Who am I?

- Sanne Grinovero
- Software Engineer at Red Hat
 - Hibernate, especially Search
 - Infinispan, focus on Search and Query integrations
 - Hibernate OGM
 - Apache Lucene
 - JGroups

Infinispan



Index

A

- AbstractJMSHibernateSearch-Controller 323
- Accents handling 130
- access strategy 41
- ACID 141
- acronyms 125, 128
- active-passive 317
- ad hoc queries 386
- generation 212
- adapter class 369, 430
- Adobe 418
- Aelfred2 430
- AJAX 143
- AliastoBeanConstructorResult-Transformer 192
- AliastoBeanResultTransformer
- Altavista 12
- Amazon 5, 183, 249
- @Analyzer 218
- @AnalyzerDef 127
- analyzers 15, 31, 45, 53, 83, 116, 125, 216–223, 415
- field 83
- filter 127
- non-English language 130
- performance 277
- query-synonyms 217
- specify a definition 127
- tokenizer 127
- annotations 67
- ANT 400
- Apache Commons
- Codec 131
- Collection 270
- Apache Lucene. See Lucene
- Apache Software Foundation 29
- Apache Solr 20, 126
- Apache Solr 20, 126
- apache-solr-analyzer.jar 216
- apostrophe 125, 128
- appliance solutions 19
- application server 30
- approximation 44
- architecture 121, 145, 310
- associated objects 110
- association 284
- bidirectional 110
- performance 277
- async 146
- asynchronous 277
- clustering 314
- copy (index) 124
- Automatic optimization 291

Our Index

- Searching in Infinispan
 - Map/Reduce
 - Fulltext indexing
- Infinispan Query engine
- Clustering a Lucene index
 - Dynamic load balancing (demo)
 - Performance & future
- Cloud deployed applications
- Future

Infinispan

- An advanced clusterable cache
- A very fast, transactional scalable datagrid
- A “NoSQL”, a *key-value* store
 - How do you query a key value store?

```
SELECT * FROM GRID
```

To Query a “Grid”

- What's in C7 ?

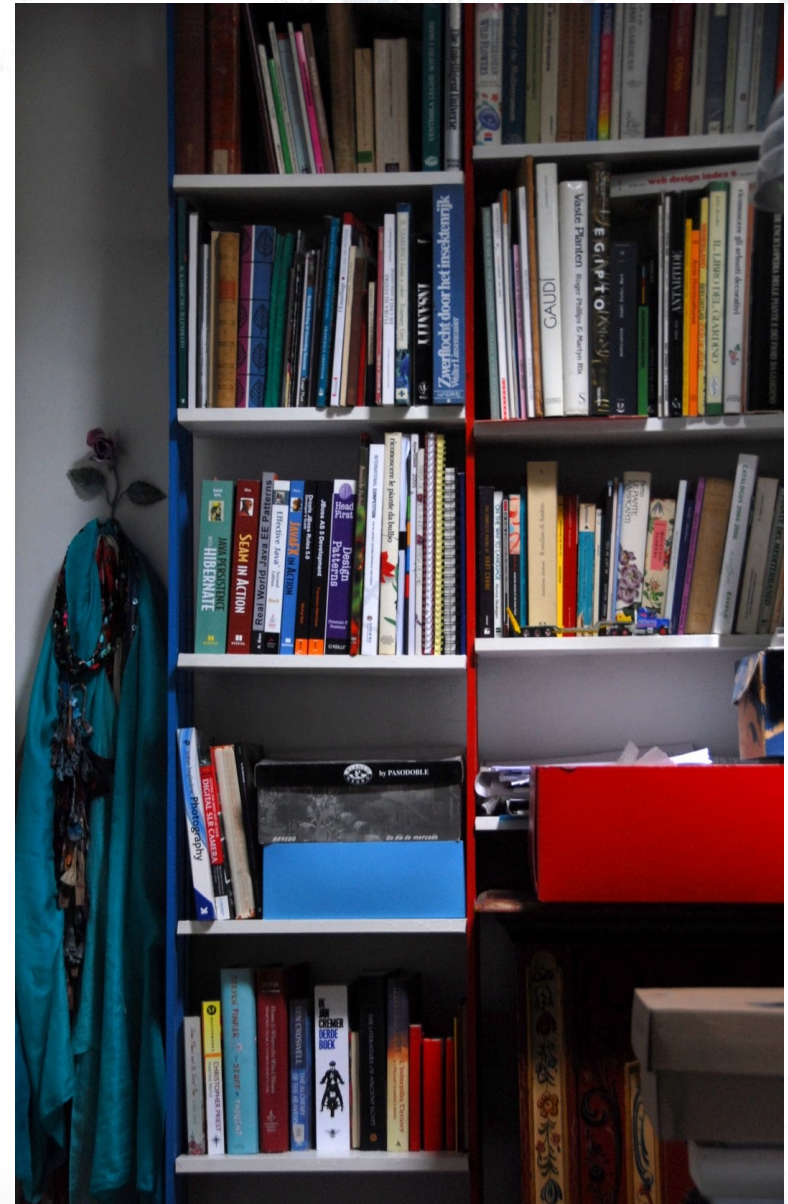
```
Object =  
  cache.get("c7");
```



If you don't know the key, no
way to find the value

Let's test my bookshelf

- Where's Hibernate Search in Action?
- Could you hand me ISBN 978-1-933988-17-7 ?
- How many books about Gaudí ?



Everybody knows that
bookshelves don't scale

How to implement the bookshelf features on a k/v?

- Where's "Hibernate Search in Action"?
- Can you hand me "ISBN 978-1-933988-17-7" ?
- How many books about Gaudí ?

Most document based NoSQLs support Map/Reduce

- Infinispan does not focus on documents
 - That won't stop you from using any format
JSON, XML, YAML, Java:

```
public class Book implements Serializable {  
  
    final String title;  
    final String author;  
    final String editor;  
  
    public Book(String title, String author, String editor) {  
        this.title = title;  
        this.author = author;  
        this.editor = editor;  
    }  
  
}
```

Iterate & collect

```
class TitleBookSearcher implements Mapper<String, Book, String, Book> {
    final String title;
    public TitleBookSearcher(String title) {
        this.title = title;
    }
    public void map(String key, Book value, Collector collector) {
        if ( title.equals( value.title ) ) {
            collector.emit( key, value );
        }
    }
}

class BookReducer implements Reducer<String, Book> {
    public Book reduce(String reducedKey, Iterator<Book> iter) {
        return iter.next();
    }
}
```

How to implement the bookshelf features on a k/v?

- ✓ Where's "Hibernate Search in Action"?
- ✓ Can you hand me "ISBN 978-1-933988-17-7" ?
- ✗ How many books *about* "Gaudí" ?
 - To properly score fulltext results we need to consider relative term frequencies on the whole corpus
 - Pre-tagging is a poor choice

Would you want a web
Search engine to return hits
in alphabetical order?

[A - Wikipedia](#) ☆

A o a (nome italiano "a" /a/) è la 1ª lettera dell'alfabeto latino e italiano e anche nella maggior parte degli alfabeti derivanti da quello fenicio. ...

[Origine](#) - [Usò nelle lingue](#) - [Codici informatici](#) - [Voci correlate](#)

it.wikipedia.org/wiki/A - [Copia cache](#) - [Simili](#)

[Home ATM, Azienda Trasporti Milanese ATM, Azienda Trasporti Milanese](#) ☆

Immagine campagna e link a nuova pagina sito BikeMi ... Mercoledì di Champions a San Siro. Il servizio Atm per Milan-Real Madrid - 02/11 ...

www.atm-mi.it/ - [Copia cache](#) - [Simili](#)

[A come AMORE - Frasi D'Amore, Poesie D'Amore, Sms D'Amore](#) ☆

A Come Amore, Tutto Sull'Amore: Frasi D'Amore, Poesie D'Amore, SMS Sull'Amore, Forum Amore, Consigli D'Amore, Segreti D'Amore, Affirma Di Coppo.

www.icconeanche.com/ - [Copia cache](#) - [Simili](#)

[A sbafo](#) ☆

Prova a contattare l'appresentante del casino online tramite le chat room, telefono o la posta elettronica. Cerca di porre le giuste domande e ...

www.sbafo.com/ - [Copia cache](#) - [Simili](#)

[adieta.it | dieta, alimentazione, fitness, salute e bellezza](#) ☆

Dimagrire, mettersi a dieta, mangiare sano, dieta on-line, calcolare il peso ideale, ricette light: tutti i modi e ancor di più quello che preferite!

www.adieta.it/ - [Copia cache](#) - [Simili](#)

[Home Page - Scherzi a parte](#) ☆

Direttore Format, Programmi e Sit-com R.T.I., Fatma Ruffini è nata a Reggio Emilia e si è laureata in Storia dell'Arte. E' approdata a Canale 5, nel 1981, ...

www.scherziaparte.mediaset.it/ - [Copia cache](#) - [Simili](#)

[Dove Andiamo A Cena Stasera? - Guida Ufficiale Ai Ristoranti](#) ☆

ristorante guida, ristoranti guida, guida ristoranti, guida ufficiale ai ristoranti in internet, official restaurants guide of internet, first born,

www.acena.it/ - [Copia cache](#) - [Simili](#)

Apache Lucene

- Open source Apache™ top level project
- Vibrant community
- Countless products and sites use it
- Integrates in Hibernate via Hibernate Search
- Clusterable via Infinispan



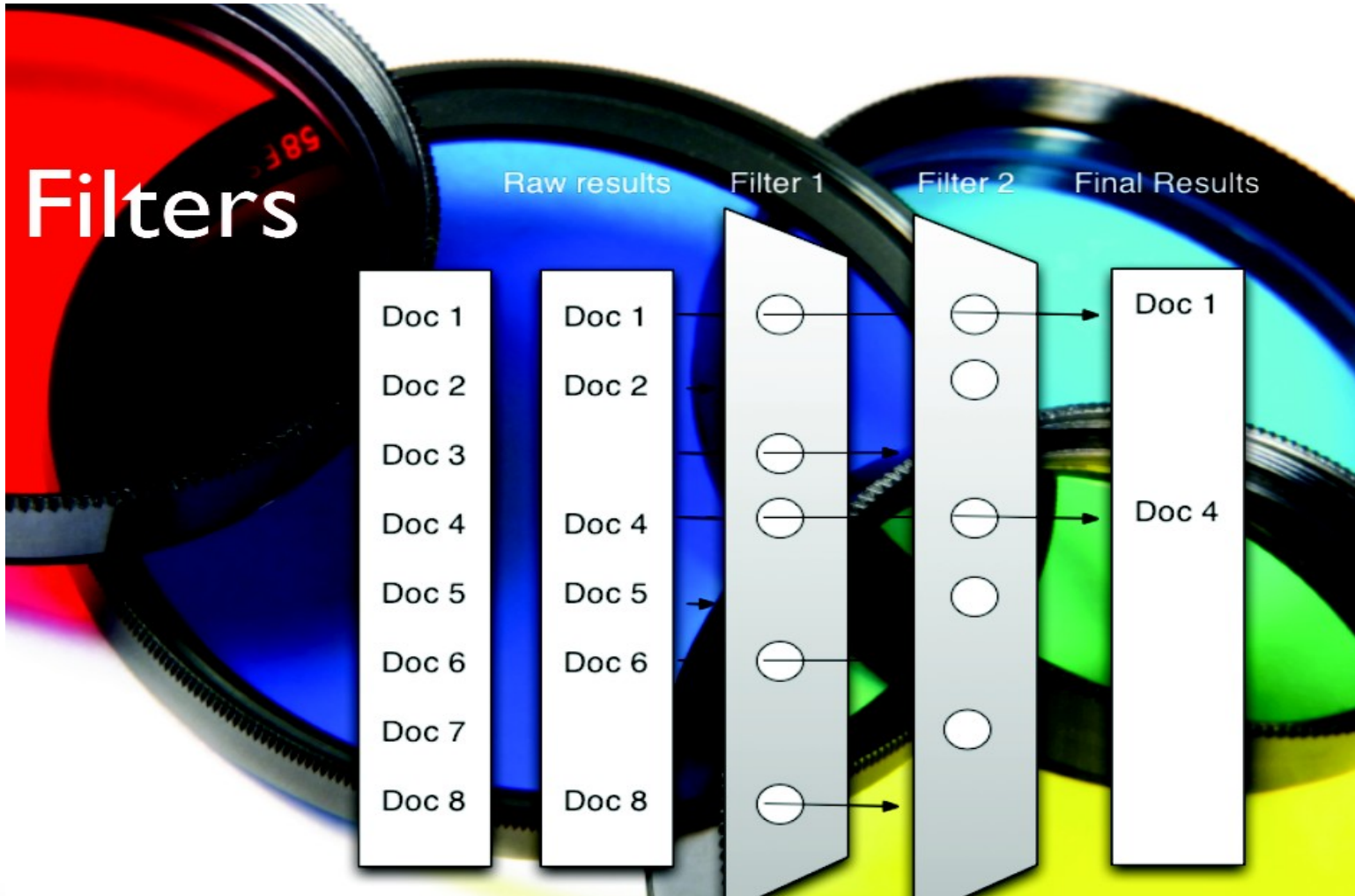
What does Lucene get us?

- Similarity scoring searches
- Advanced text analysis
 - Synonyms, Stopwords, Stemming, ...
- Reusable declarative Filters
- TermVectors
- MoreLikeThis
- Faceted Search
- Speed!

Lucene: Stopwords

a, able, about, across, after, all, almost, also, am, among, an, and, any, are, as, at, be, because, been, but, by, can, cannot, could, dear, did, do, does, either, else, ever, every, for, from, get, got, had, has, have, he, her, hers, him, his, how, however, i, if, in, into, is, it, its, just, least, let, like, likely, may, me, might, most, must, my, neither, no, nor, not, of, off, often, on, only, or, other, our, own, rather, said, say, says, she, should, since, so, some, than, that, the, their, them, then, there, these, they, this, tis, to, too, twas, us, wants, was, we, were, what, when, where, which, while, who, whom, why, will, with, would, yet, you, your

Filters



Faceted Search

Shop All Departments ▼ Search Hibernate Search

Books Advanced Search Browse Subjects New Releases Bestsellers TI

Department
◀ Any Department
◀ Books
Computers & Internet
Programming (14)
Computer Science (4)
Databases (2)
Software (2)
Web Development (2)
Networking (1)
Home Computing (1)

Format
 Paperback (15)

Author
Any Author
Joe Vitale (1)

Shipping Option [\(What's this?\)](#)
Any Shipping Option
Free Super Saver Shipping

Books > Computers & Internet > "Hibernate Search"

Showing 1 - 12 of 15 Results

- 

LOOK INSIDE!

Hibernate Search in Action
★★★★★ (3 customer reviews)
Formats
Paperback
Order in the next **2 hours** to get it by **Monday, Apr 18.** ~~\$49.99~~
Only 1 left in stock - order soon.

Eligible for **FREE** Super Saver Shipping.
Excerpt - Page 1: "... breaking the sus...
Surprise me! See a random page in thi
- 

LOOK INSIDE!

Spring Persistence with Hibernate
★★★★☆ (5 customer reviews)
Formats
Paperback

The downsides

- Requires an Index
 - on filesystem
 - in memory
 - in Infinispan
- Made of immutable segments
 - Optimized for search speed, not for updates
- A world of strings and frequencies

Infinispan Query quickstart

- Enable it in configuration
- Have infinispan-query.jar in your classpath
- Annotate your POJO values to specify what to index

```
<dependency>  
  <groupId>org.infinispan</groupId>  
  <artifactId>infinispan-query</artifactId>  
  <version>5.0.0.CR1</version>  
</dependency>
```

Enable Infinispan Query, programmatically

```
Configuration c = new Configuration()  
    .fluent()  
    .indexing()  
    .addProperty(  
        "hibernate.search.option", "valueForOption" )  
    .build();  
CacheManager manager = new DefaultCacheManager( c );
```

Enable Query in Infinispan XML configurations

```
<?xml version="1.0" encoding="UTF-8"?>
<infinispan
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="urn:infinispan:config:5.0
http://www.infinispan.org/schemas/infinispan-config-5.0.xsd"
  xmlns="urn:infinispan:config:5.0">
  <default>
    <indexing enabled="true">
      <properties>
        <property name="hibernate.search.option" value="value" />
      </properties>
    </indexing>
  </default>
```

Annotate your model

```
@ProvidedId @Indexed
public class Book implements Serializable {

    @Field String title;
    @Field String author;
    @Field String editor;

    public Book(String title, String author, String editor) {
        this.title = title;
        this.author = author;
        this.editor = editor;
    }
}
```


Run a Query

```
SearchManager qf = Search.getSearchManager(cache);

Query query = qf.buildQueryBuilderForClass(Book.class)
    .get()
    .phrase()
        .onField("title")
        .sentence("in action")
    .createQuery();

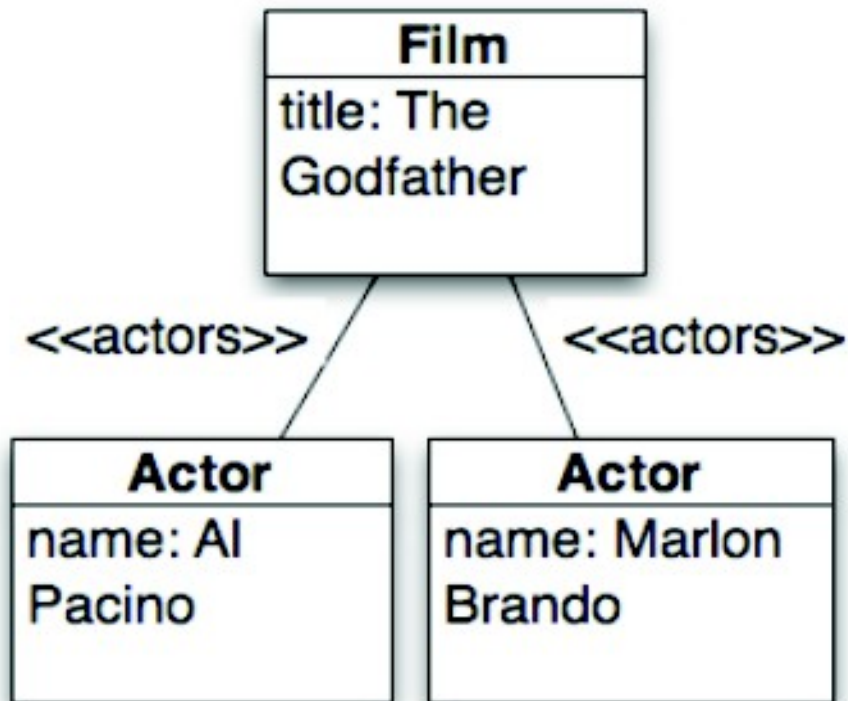
List<Object> list = qf.getQuery(query).list();
```

Architecture

- Integrates Hibernate Search
 - Listen to Hibernate events & transactions
 - Infinispan events & transactions
 - Maps Java types and model graphs to Lucene Documents
 - Thin-layer design

Index mapping

Object world



Index world

Film Document	
title	The Godfather
actor	Al Pacino
actor	Marlon Brando

declarative analyzers

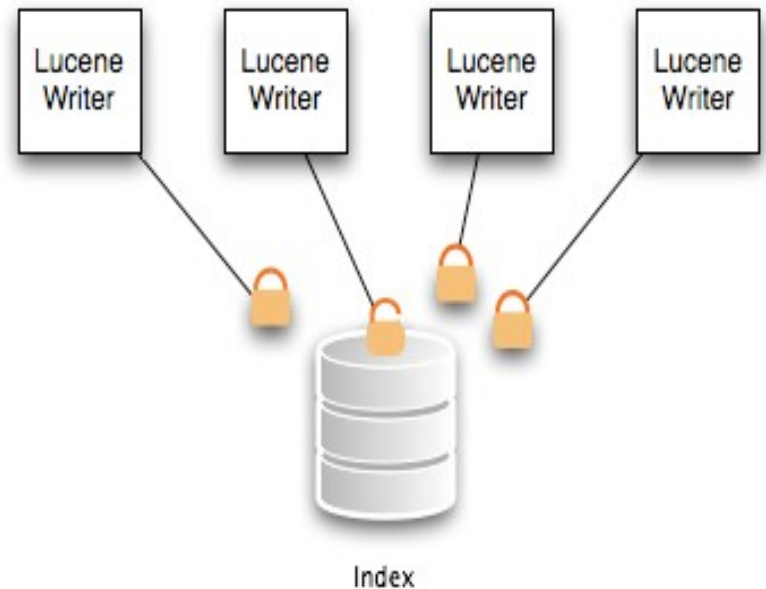
```
@Entity @Indexed
@AnalyzerDef(name = "frenchAnalyzer", tokenizer =
    @TokenizerDef(factory=StandardTokenizerFactory.class), filters = {
        @TokenFilterDef(factory = LowerCaseFilterFactory.class),
        @TokenFilterDef(factory = SnowballPorterFilterFactory.class,
            params = {@Parameter(name = "language", value = "French")})
    })
public class Book {

    @Field(index=Index.TOKENIZED, store=Store.NO)
    @Analyzer(definition = "frenchAnalyzer")
```

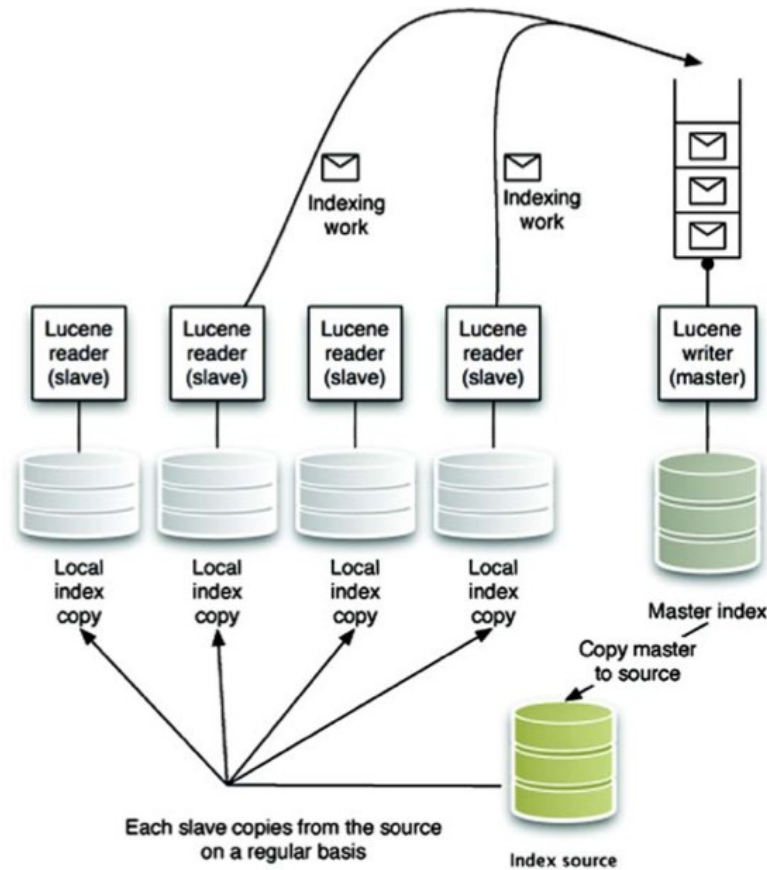
Query demo

Scalability issues

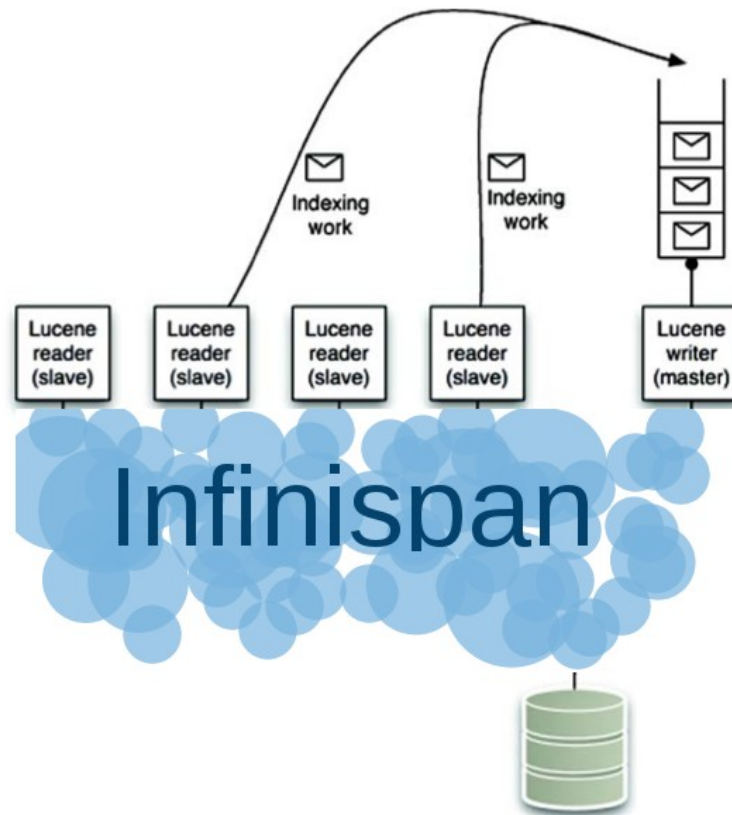
- Global writer locks
- NFS based index sharing *very* tricky



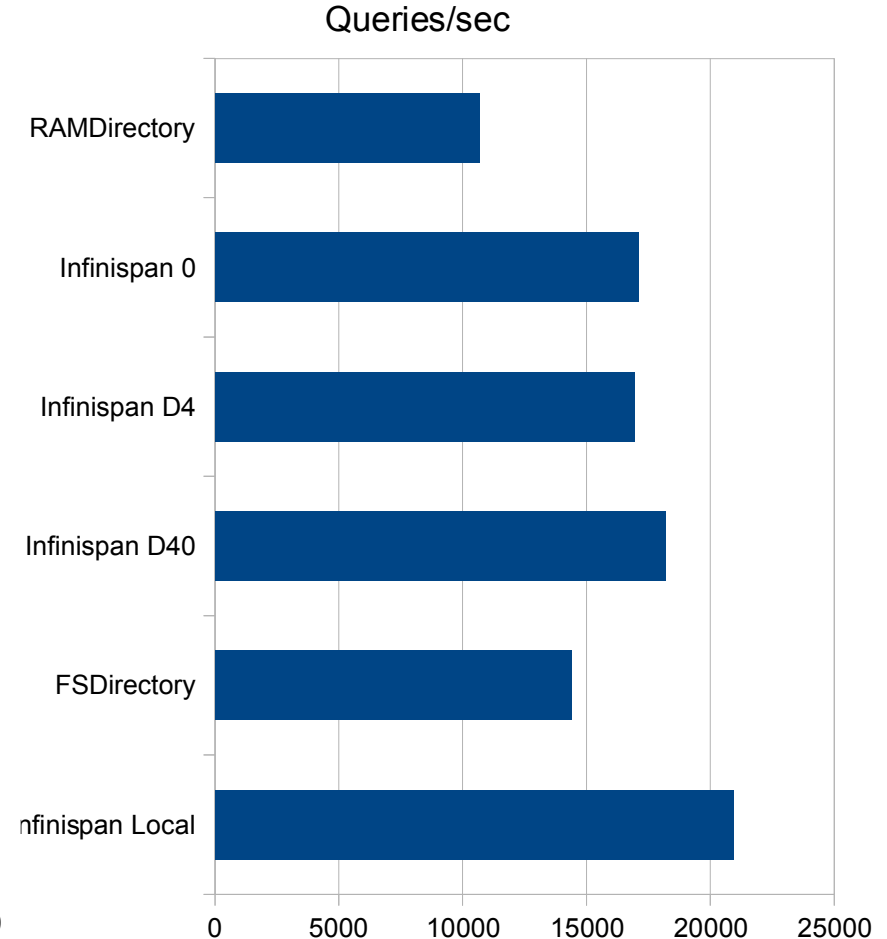
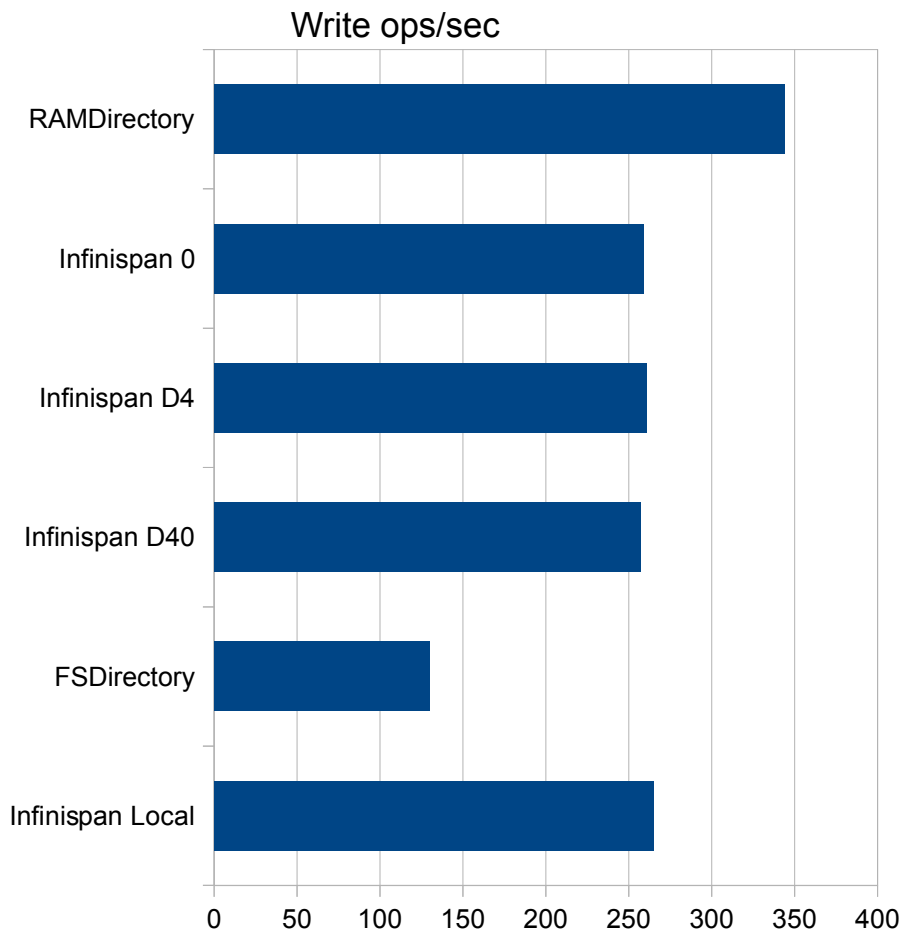
Queue-based clustering (via filesystem)



Index stored in Infinispan

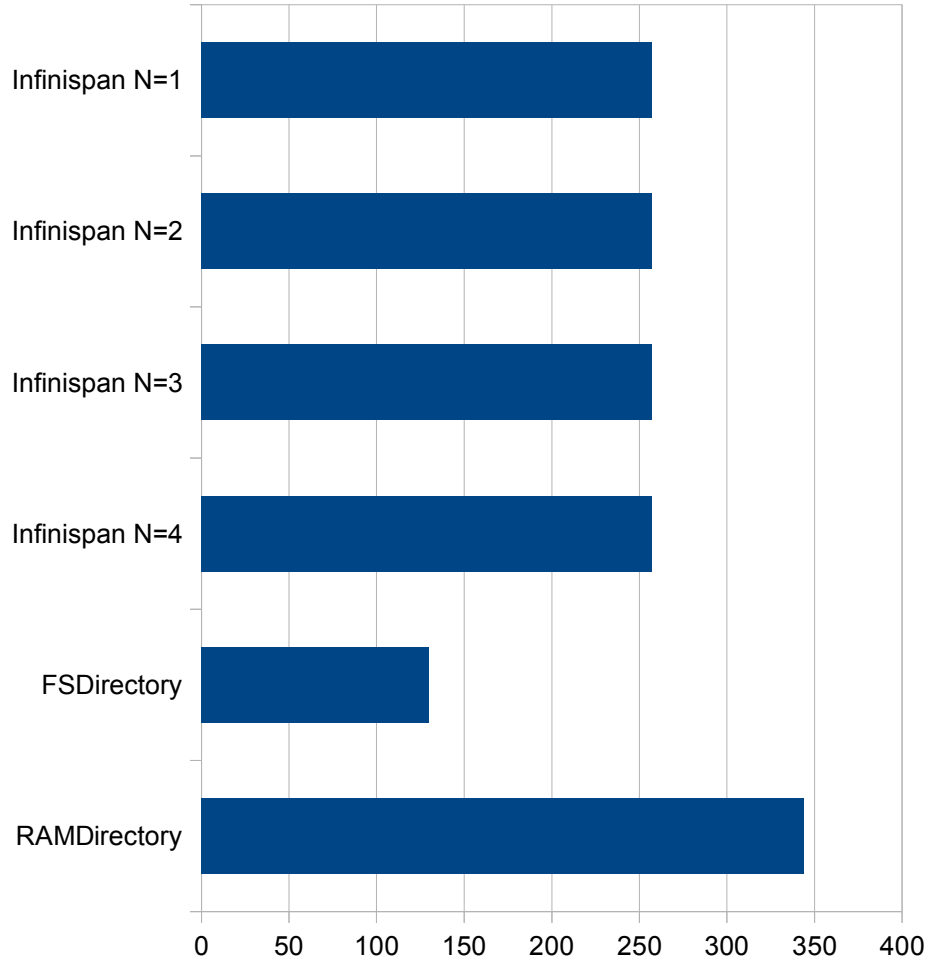


Single node *performance idea*

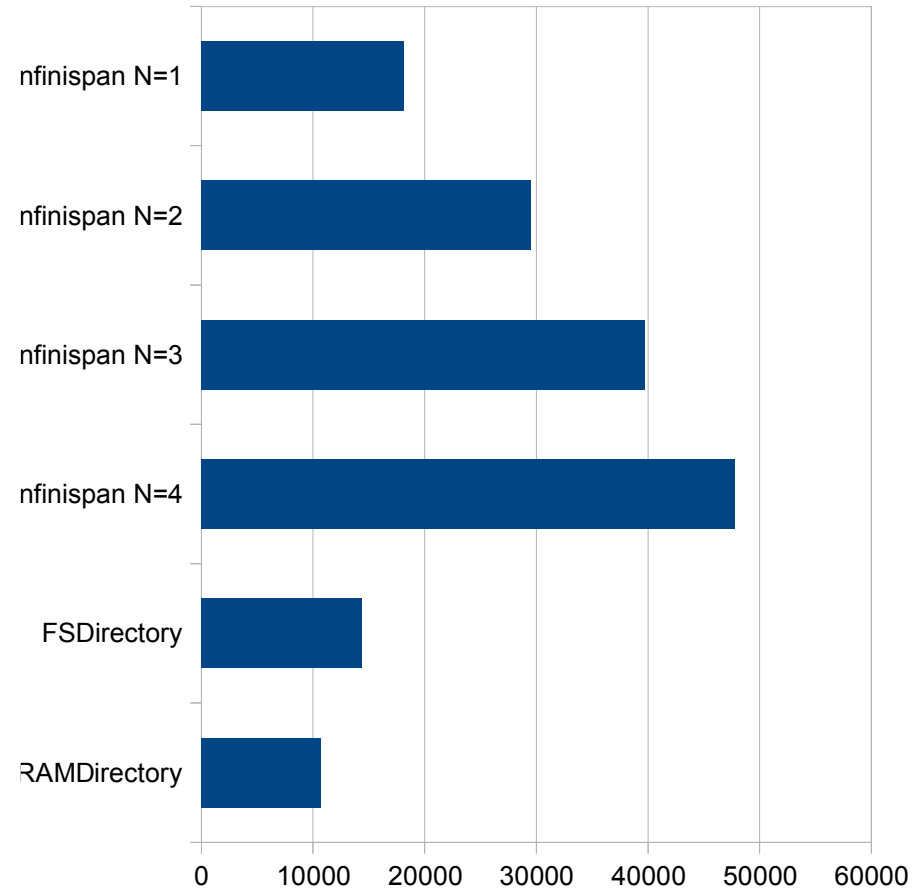


multi-node setup

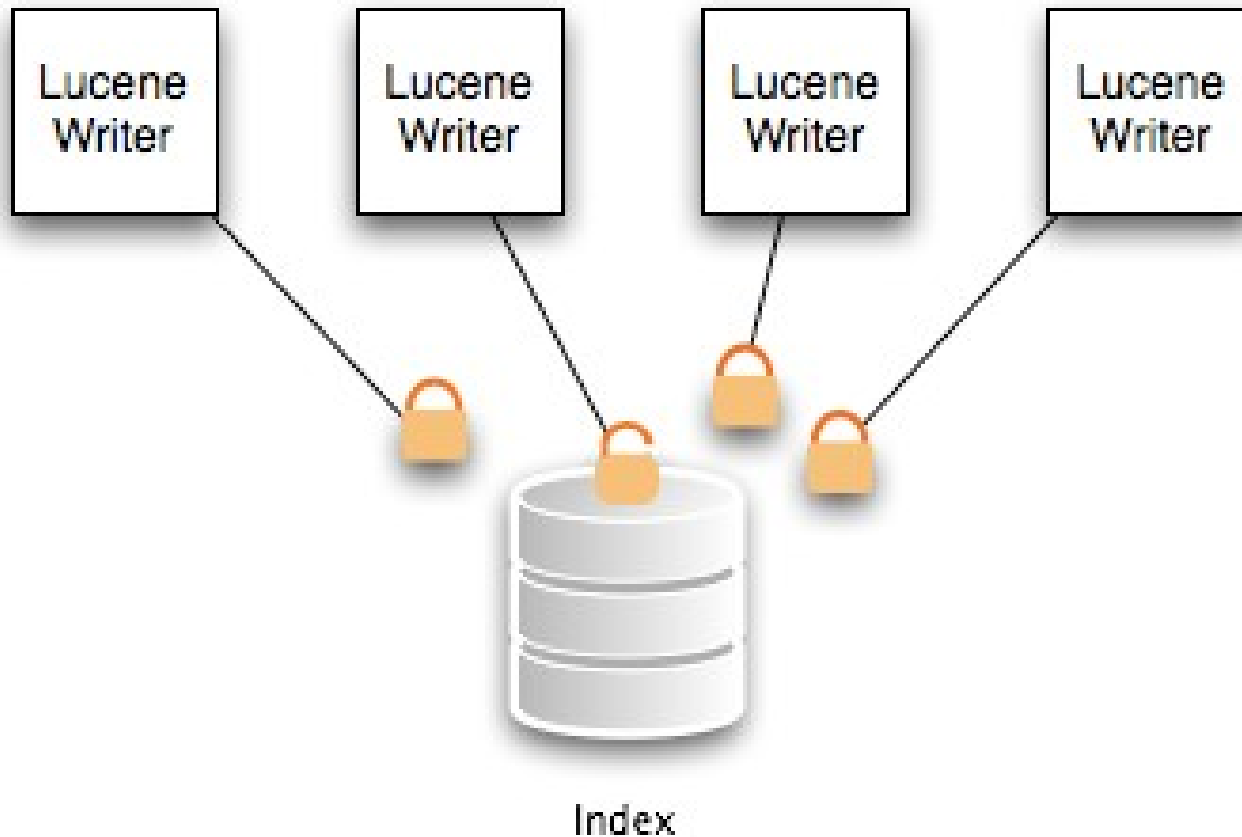
Write ops/sec



Queries/sec



Why does writing not scale?



Performance warnings

- Set Lucene's maximum segment size to fit in LuceneDirectory *chunk_size* to avoid readlocks
- Verify blob sizes fit in JGroups network packets
- Check for CacheStores “sweet spot” size

Memory requirements

- RAMDirectory: all must fit in a single VM's memory
- FSDirectory: OS does a great caching job – but if it doesn't fit in memory
 - Garbage collection
- Infinispan: comparable to FSDirectory
 - Flexible
 - Faster
 - Network vs. disk

Ingredients for a cloud

- setup once and reuse Infinispan
 - To store indexes
 - As Hibernate second level cache
 - As application cache
 - As datagrid
 - As a JPA “store” via Hibernate OGM

Ingredients for a cloud

- JGroups discovery protocol
 - TCP_PING
 - JDBC_PING
 - S3_PING
- Choose a CacheLoader
 - Database based
 - JClouds
 - Cassandra

Plan for improvements

- Have writing scale too
- Ease configuration aspects for clustering
- Parallel searching
- Continuous querying
- A component of
 - <http://www.cloudtm.eu>



Related events at JBossWorld

- Keynote demo
- Highly Scalable Data Grids and Distributed Caching with Infinispan
- Infinispan – Optimizing Performance & Consistency at the Chicago Board Options Exchange
- JPA on Infinispan: When PaaS Persistence Meets Java EE

Q&A?

Infinispan