

# APACHE LUCENE EUROCON



Solr at



Robin Bramley, Ixxus  
October 2011

Presented by



# What I Will Cover

- How Solr helps charities raise money
- Who I am
- The challenges in launching a new not-for-profit business to an aggressive timeframe
- Search with an Open Source SOA
  - **Loosely coupled, event-driven indexing**
  - **Solr topology**
  - **High Availability / Disaster Recovery**
- Lessons learnt and alternative solutions

# My Background

- Worked at consultancies for 13 years
  - Worked with Mark Harwood (Lucene committer) at Rubus
  - Lead Architect on Virgin Money Giving project
- Around the edges I also write stuff
  - GroovyMag author - articles inc. *Using Lucene in Grails*
  - DZone Most Valuable Blogger
  - Open Source contributions
    - *OpenID support in Acegi / Spring Security*
    - *Codenarc support in Hudson / Jenkins CI Violations*



# – from vinyl to space...

- One of the world's most recognised brands
- Sir Richard Branson is an iconic entrepreneur



- Founded in 1970 to sell mail order records
- Virgin group launched over 300 branded companies
  - **Virgin Money provides financial services**
  - **Virgin Galactic is the world's first spaceline**

# The London Marathon

- Established in 1981
- One of 5 World Marathon Majors (with Berlin, Boston, Chicago & New York)
- Largest annual fund raising event in the world (Guinness world record)
  - **2010 – 36,549 people raised over £50m**
- Virgin signed a 5 year sponsorship deal



# Virgin Money Giving

- A brand new not-for-profit business
- Launched in direct competition with an established profit-making firm
- Principle of “Everyone’s Better Off”
  - **Fundraisers & donors: more cash going to good causes**
  - **Charities: pay lowest fees**
  - **Virgin Money: brand awareness**





# The Challenge

- The business needed to launch for the start of the marathon campaign
- Search is a fundamental requirement of the public site
  - **Fundraisers need to find charities and/or events**
  - **Donors need to find charities**
  - **Sponsors need to find fundraisers**


# Front Page Search

**Find a charity or friend**

Find a friend  

Find a charity  


[Browse charities](#)




## Make a donation

We've found 13 matches for 'headway'

Areas to search  Fundraisers  Fundraising events  Charities

Search  

Showing results 1 - 10 of 13

1 2 [Next](#) 



### **Headway Thames Valley Limited**

**Reg No. 900591**

Headway Thames Valley was founded twenty two years ago by the mother of a head injured young man whose life had been dramatically changed by a road traffic accident. Medical technology had saved his life and rehabilitation had taught him to walk again, but on discharge from hospital, there was nothing to help him continue the progress already made, just a void. Headway Thames Valley started to fill that gap, initially as a small support group. Affiliated to Headway UK, its base is now well established in Henley with satellites in Reading and Wokingham. There are 9 Trustees, 11 staff including 2 outreach workers, 5 clinical

# Technical Constraints

- Offload full text search from the database
- The search solution needed to fit into an SOA
- Open Source preferred for cost effectiveness

# Technical Goals

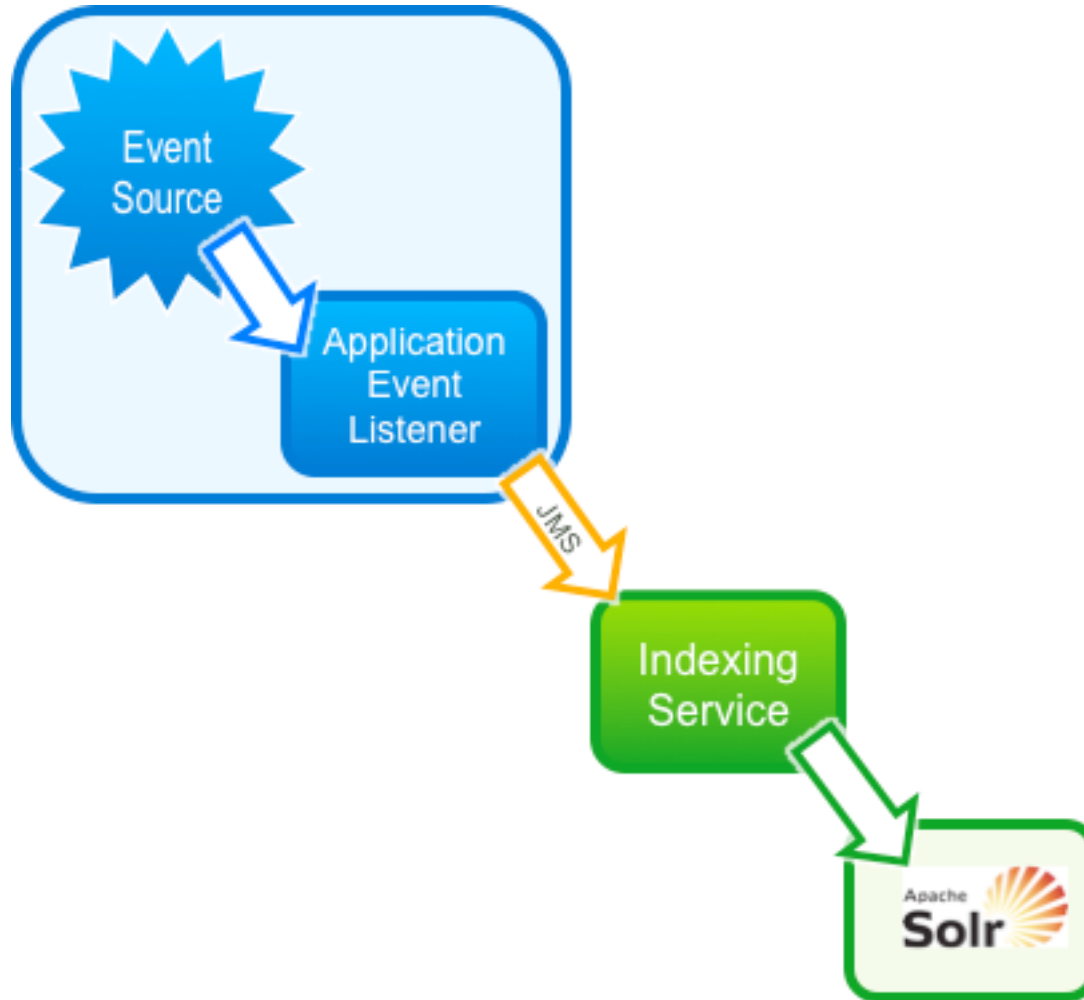
The key aims:

1. Not polluting business logic with search APIs
2. Direct presentation of results
3. Queries served by read-only replicas
4. Keep the secondary data centre indices fresh

# Technical Goal - 1

- **Goal:** *Not polluting business logic or tightly coupling it to a search engine API*
- **Why:** indexing is a technical concern
- **Approach:** Event-driven indexing
- **How:**
  - Spring Application Events
  - JMS
  - Message Driven POJO
  - SolrJ

# Event-driven indexing



# Event Publishing in Spring

```
public class Xyz implements ApplicationEventPublisherAware {  
  
    private ApplicationEventPublisher applicationEventPublisher;  
  
    public  
    setApplicationEventPublisher(ApplicationEventPublisher aep)  
    {  
        applicationEventPublisher = aep;  
    }  
  
    ...  
  
    public create(Map<String,String> foo) {  
        ...  
  
        XyzCreatedEvent evt = new XyzCreatedEvent(foo);  
  
        applicationEventPublisher.publishEvent(evt);  
    }  
}
```

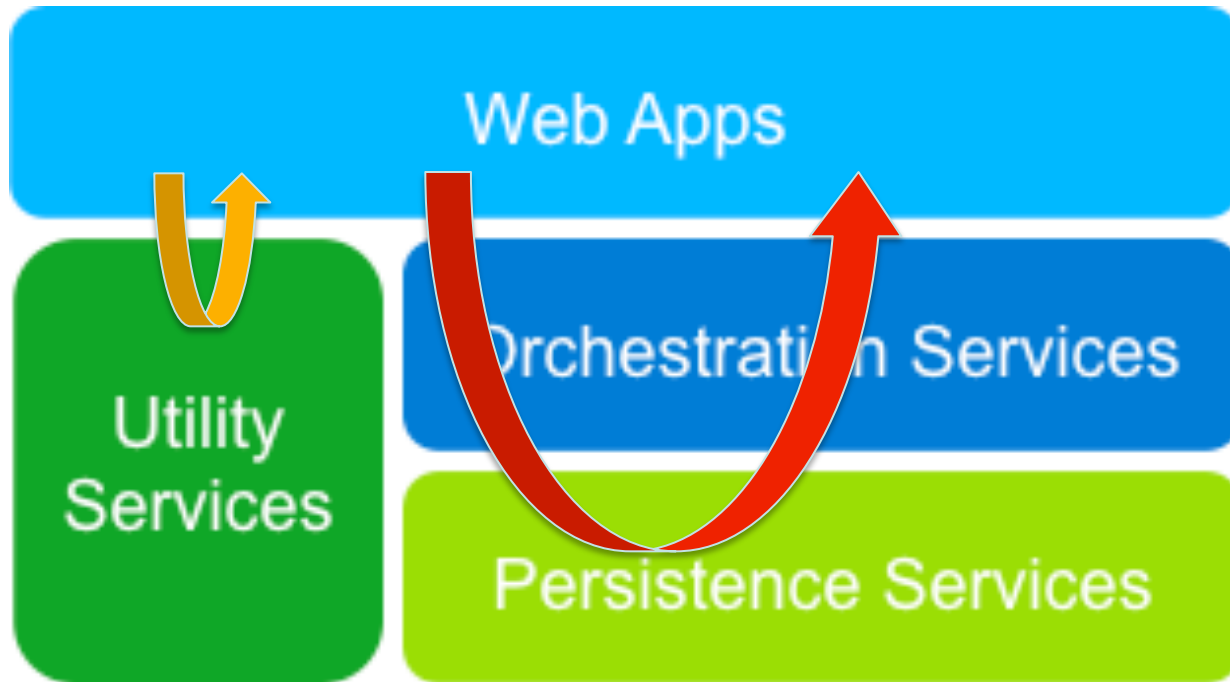
# Event Listener

```
public class IndexingListener implements ApplicationListener {  
    public onApplicationEvent(ApplicationEvent event) {  
        if (event instanceof XYZCreatedEvent) {  
            XYZCreatedEvent evt = (XYZCreatedEvent) evt;  
            ...  
            // handle created event  
        } else if (event instanceof XYZDeletedEvent) {  
            ...  
        }  
    }  
}
```

# Technical Goal - 2

- **Goal:** *Direct results presentation*
- **Why:** prevent additional database reads
- **How:**
  - Careful schema design

# Results delivery



Key

Direct results

Avoid this!

# Schema

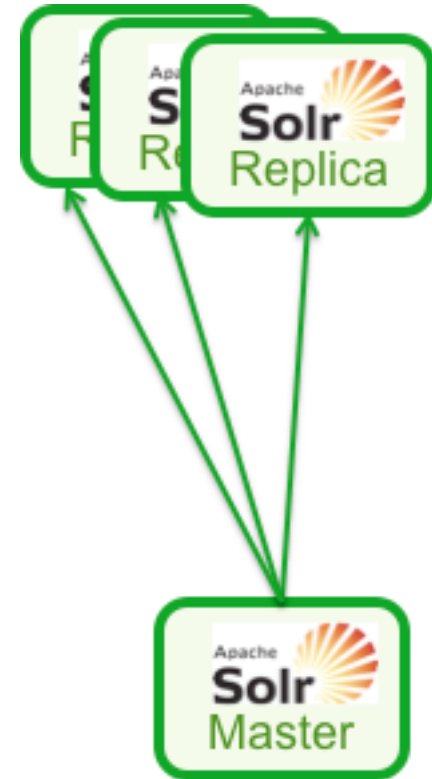
- Keep it simple
- Document types in one collection
- 4 field types
  - boolean, date, string, text
  - *Simple analysis on text*
- Selectively indexed stored fields
  - **Presentation tier can deliver results straight from Solr**

# Technical Goal - 3

- **Goal:** *Read-only replica search nodes for scale out*
- **Why:** separate read and write load
- **How:**
  - Snap Pull Replication

# Snap Pull Replication

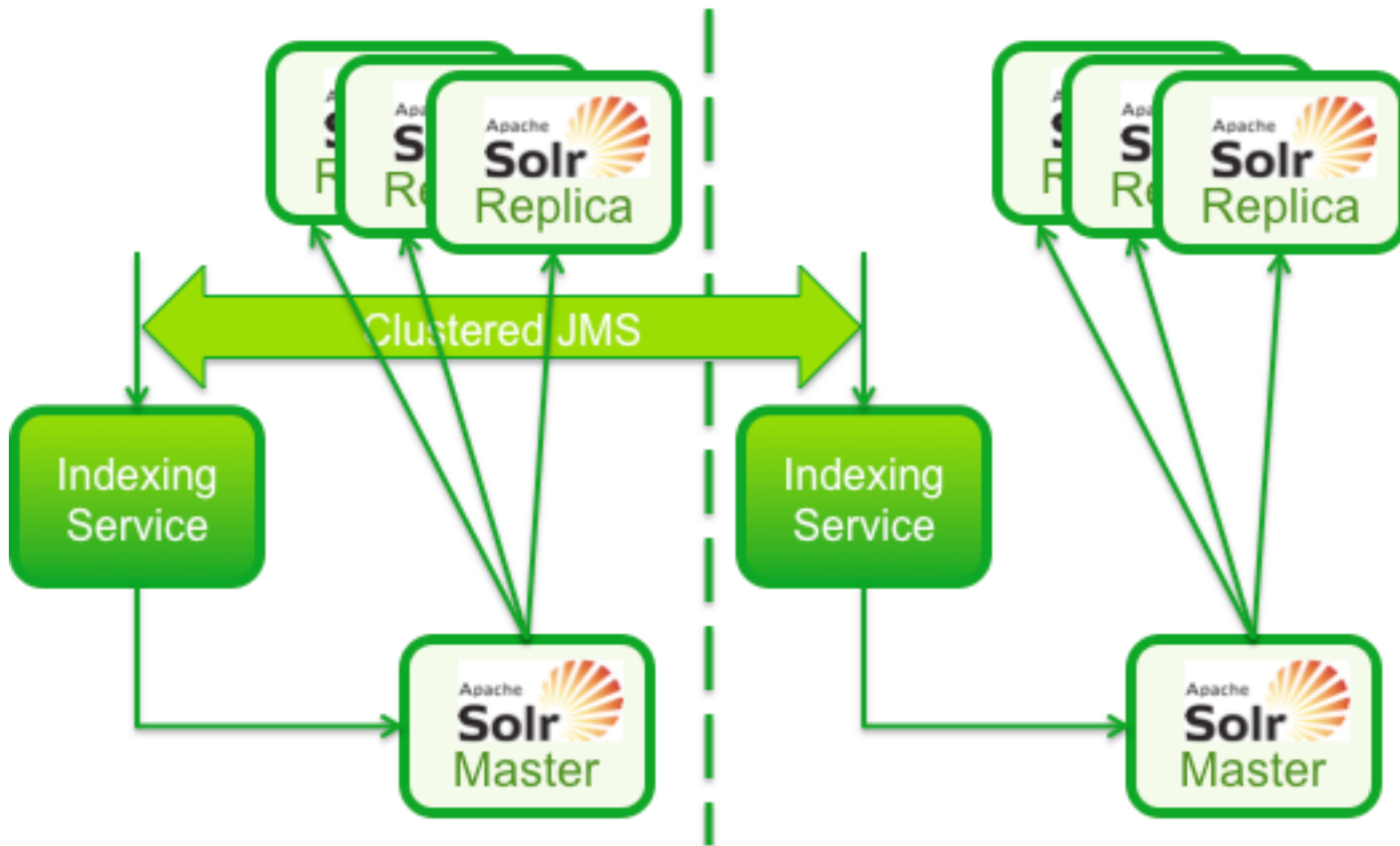
- Collection distribution
  - Snap shooter
  - Snap puller
  - Snap installer
  - Snap cleaner
- Rsync transport



# Technical Goal - 4

- **Goal:** *Propagate index updates to secondary site*
- **Why:** site-level failover / disaster recovery
- **How:**
  - **Clustered JMS**

# Solr Topology



# Business Results

- Business launched to aggressive timescales
- Raised over £10m for charity in the first 6 months
- Project awarded a medal in the British Computer Society 2010 Computing Awards for an outstanding community IT project

# Lessons Learnt

- **Currency**
  - **Solr replication has a number of challenges**
    - *Trade-offs with frequent index updates*
  
- **Consistency**
  - **Inconsistency across replicas**
  - **Clustered JMS presented some operational issues**

# Alternative options

- Alternative options now available:
  - Solr Cloud with ZooKeeper ensemble
  - Solandra
  - ElasticSearch

# References

1. <http://uk.virginmoneygiving.com>
2. <http://wiki.apache.org/solr/CollectionDistribution>
3. <http://www.ixxus.com/news/2011/02/bcs-medallist/>
4. <http://wiki.apache.org/solr/SolrCloud>
5. <https://github.com/tjake/Solandra>
6. <http://www.elasticsearch.org/>

# Contact

- Robin Bramley
  - <http://www.ixxus.com>
  - Twitter: [@rbramley](https://twitter.com/rbramley)
  - Blog: <http://leanjavaengineering.com>