In the Mix - Developing open source search technologies on the Microsoft platform

Matthew Brumpton
Senior Systems and Applications Developer
UK Data Archive

IASSIST 2013, Cologne
29 May 2013
Overview

• Why use open source?
• The Solr project
• Integrating Solr with Windows
• Application architecture
• Server architecture
• Debugging and monitoring
• Questions
Why use open source?

• Features
• Performance
• Community
• Roadmap
• Cost
• Tooling
The Solr project

- Open-source solution, largely supported by the Apache Software Foundation.
- Written in Java
- Created in 2004, current version 4.3
- SOLR works with Hadoop, Lucece, Tika, Nutch
- Capable of indexing billions of items in a clustered environment.
- Features include:
  - Full-text search
  - Faceted search
  - Highlighting
  - Rich document handling
  - Distributed search (Solr Cloud)
  - Highly scalable
  - NoSQL
Solr on Windows

- Requires Java and web application server (Tomcat, Jetty)
- Rest web service API
  
  http://dasolr/solr/Catalogue/select? q="Social Science"& sort=Date+desc

- Manual installation

- Plugins might only run on Linux or require an emulator (Cygwin)

- Configured through XML

- Reduce dependencies through Service Bus
Application Architecture

- **Application**
  - Umbraco / ASP.MVC
  - HTML
  - Jquery

- **Business / Data Access**
  - .NET Libraries
    - UKDA.Search.Library
    - SolrNet
    - Mule

- **Solr Cloud**
  - Solr 4.1
  - Lucene
  - JVM 7
  - Tomcat 7

- **Data**
  - MS SQL
  - .NET 4 Console
  - Entity framework / WebAPI
Debugging and monitoring

• Query Logs (Enterprise Service Bus)
• Solr Logs
• Elmah (.NET error logging)
• Fiddler (HTTP Monitoring)
• WireShark (Network Monitoring)
Questions