



# Search Engine Death Match

## Solr (Lucene) vs. Verity

Kevin Penny, Jobs2Web, Inc

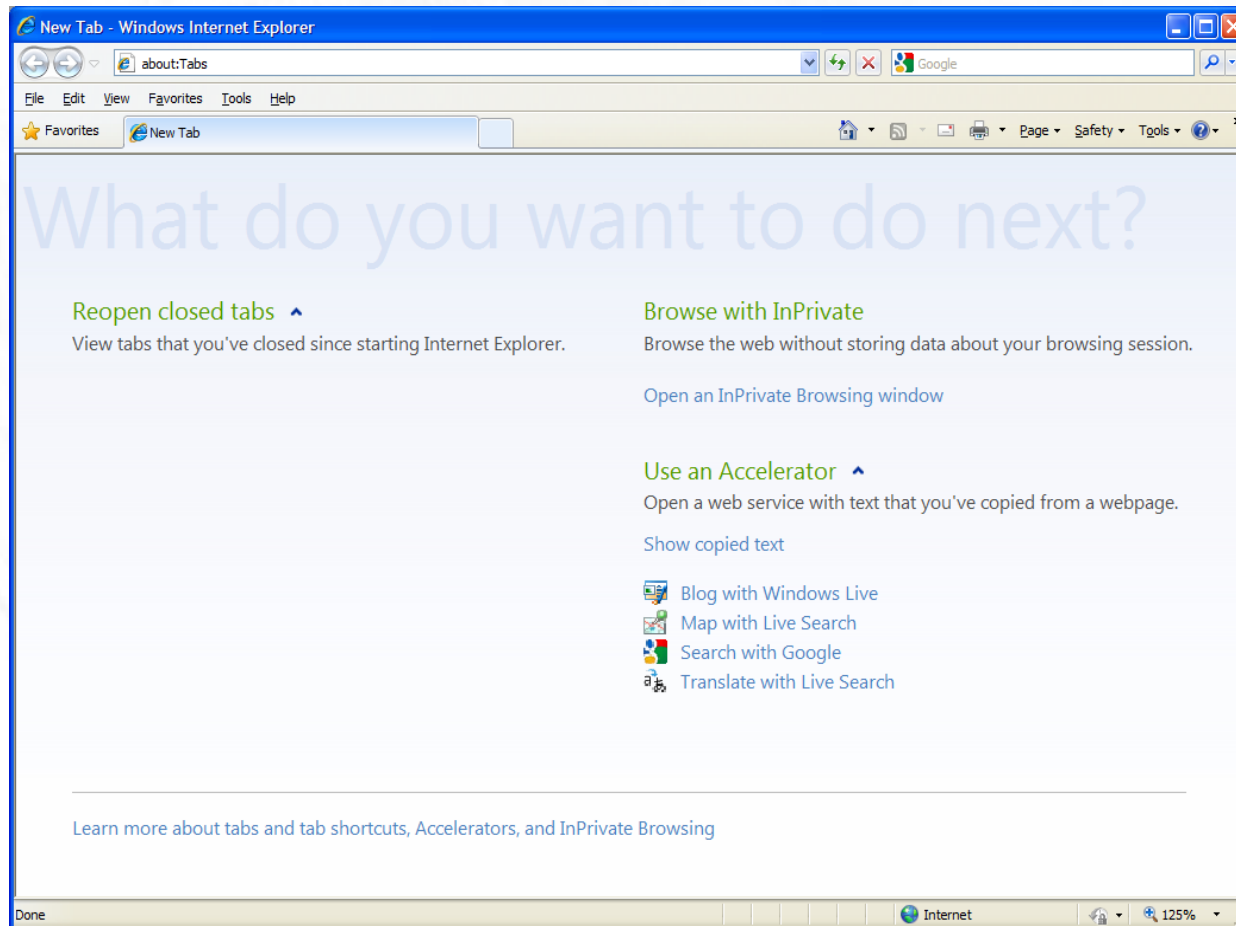
# Tale of the Tape

- Sr. Web Architect – Jobs2Web, Inc (5 years)
- Carlson Marketing Group  
Plymouth, MN
- ComMark, Inc – North Dakota
- I am Canadian
- CFML Developer since 1998
- Alpha Tester for Bolt (ColdFusion Builder)
- [CFProperty Inspector](#)
- Alpha Tester for Centaur (CF9)
- Blogger @ [codeFusionLLC.blogspot.com](http://codeFusionLLC.blogspot.com)

# Strength of Search?

- Search Rules All
- Google has made a living (quite a nice one) with Search
- Everyone needs it
- Hard to get right
- Internet would be difficult to use without it

# Please Type Your URL Here...



# And in this corner...

- **Solr** (Lucene's Child)
- Searching On Lucene w/Replication (SOLR)
- XML/HTTP Interface
- Open Source Project
- Java Based (Lucene)
- XML / HTTP / DB / Docs
- FAST
- **Verity** (OEM w/CF)
- Bundled
- Stand alone Installation (1 IP)
- Cfsearch -able
- Docs / DB
- EASY

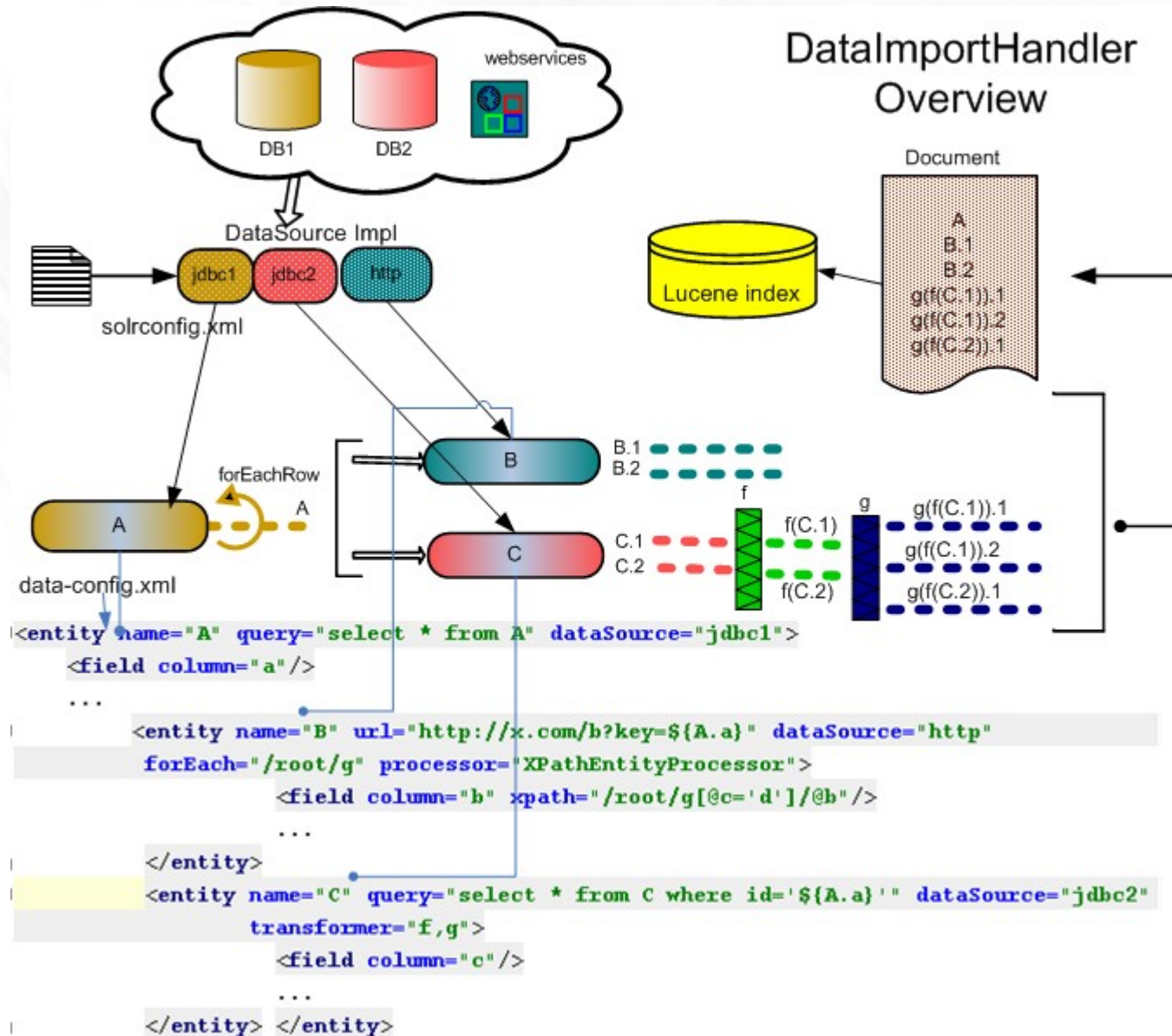
# Knocks

- **Solr** (Lucene's Child)
  - Many options mean more configuration
  - Multiple Collections require multiple Solr Cores
  - Not CF Integrated (\*)
  - More difficult to get perfect
  - Relevancy differs from Verity
  - \* Or is it?
- **Verity** (OEM w/CF)
  - Not Cross Platform (mac)
  - Limited to Threads/Hardware
  - Collection Limitations
  - Record Limitations
  - Column Limitations (custom1,2,3,4)
  - Expensive if you outgrow OEM

# Solr Features

- Stemming Support i.e. (Manager~Manag)
- Phonetic Search
- Schema Defined
- Sorting/Faceting
- Field searching
- XML Result
- Choice of Servlet Container (i.e. Jetty, Tomcat, Resin)
- Familiar / Simple query String/http calls (? command=full-import)
- Multiple Sources of Data
- Extensive Caching
- Index Replication

# Solr Architecture



# Solr -Who's Using it?

E-trade

Goldman Sachs

Disney

Apple, Inc

AT&T

NASA

Netflix

Citysearch

StubHub!

AOL

# Verity -Who's Using it?

- Arden Business Consultants
- BayLogin
- US Gov't
- (Not Jobs2Web ;)

# Solr Syntax

- Q – Main search parameter (what you will query for)
- Fq – filter query – used to further subfilter the original resultset
- DebugQuery - (true) if you want to see what Solr is doing under the hood (relevancy/phonetic matches/stemming etc)
- Start – point in the recordset in which you want to view the data
- Rows – how many rows do you want to return
- Indent – on/off – compact the xml result
- Fl – field list – which fields to you want to retrieve data for (\*,score)

# Ready to Rumble

- PART 1

- Indexing XML Data Raw (sample/iTunes Library)
- [Schema/Data-config](#)
- Indexing XML Data in Cfadmin Collection
- Output (search syntax)
- Debugging/XML Output

- PART 2

- [Indexing Database Data](#)
- Schema Design Setup
- Delta/DeletePK Query
- Output (syntax)
- Debugging/XML Output
- Faceting

# Facets (dynamic clustering)

**Digital cameras**

Refine your results

**Manufacturer**

- Canon USA (5)
- Sony (2)
- Nikon (2)
- Olympus (6)
- Pentax (2)

**Resolution**

- 6 megapixels (3)
- 8 megapixels and up (14)

**Zoom range**

- 3X to 4X (11)
- 8X to 12X (1)

**More**

- LCD size
- Image stabilizer
- Flash memory
- Still image format
- Maximum ISO

See all >

you selected: \$400 - \$500 SLR remove all

17 results

Regular search results list

Show 10 results per page Sort by: Review date

COMPARE SELECTED

**Canon EOS Rebel XS (silver, with 18-55mm lens)** \$459 to \$699 at 15 stores

**Annotations:**

- Manufacturer is a **facet**, a way of categorizing the results
- Canon, Sony, and Nikon are **constraints**, or facet values
- The **facet count** or constraint count shows how many results match each value
- The **breadcrumb** trail shows what constraints have already been applied and allows for their removal

# The Bloody Truth (Verity)

Collection Limit Hit (82-88)

Collections Offline (Randomly)

Collection Random X-Fer (verity bug)

Errors & Thread counts (3 per collection)

# Final Round

[Solr FAQ - Relevancy Cookbook](#)

[Lucene FAQ](#)

[CF 9 \(Solr Integrated\)](#)

[SolColdFusion.RIAForge.org](#)

[www.CodeFusionLLC.com Blog](#)

[Ferret - Ruby](#)

[Namazu](#)

[Faceted Searching with Solr](#)

[CodeFusionLLC.BlogSpot.com](#)

[jobs2web](#)

# And the Winner is...

- What Search Solutions do you Use?
- Would you use Solr?
- Would you continue to use Verity?