

APACHE
LUCENE
EUROCON



Just the Job – Employing Apache Solr for Recruitment Search

charlie@flax.co.uk

Charlie Hull, Flax

@FlaxSearch

19th October 2011

Presented by

lucid
IMAGINATION



What I Will Cover

- Who are Flax?

What I Will Cover

- Who are Flax?
- The Project & The Solution

What I Will Cover

- Who are Flax?
- The Project & The Solution
- How we did it
 - A flexible pipeline in two parts
 - Transforming the UI
 - Performance
 - Issues
 - Results & benefits

What I Will Cover

- Who are Flax?
- The Project & The Solution
- How we did it
 - A flexible pipeline in two parts
 - Transforming the UI
 - Performance
 - Issues
 - Results & benefits
- Conclusions & Lessons Learned
 - Learning to love open source search

Who are Flax?

- Search engine specialists with decades of experience
- Based in Cambridge, U.K.
- Customers include Financial Times, Durrants Ltd., Accenture, University of Cambridge
- UK Authorised Partner of Lucid Imagination

We also run a Search Meetup:



Start your own - add to www.searchmeetups.com !

The Project

- The client: *Reed Specialist Recruitment*

The Project

- The client: *Reed Specialist Recruitment*
- The data
 - Hundreds of millions of items to search
 - Hundreds of fields in the database schema (which will change in the future)
 - CVs (resumés) in Word, PDF formats
 - Multiple languages

The Project

- The client: *Reed Specialist Recruitment*
- The data
 - Hundreds of millions of items to search
 - Hundreds of fields in the database schema (which will change in the future)
 - CVs (resumés) in Word, PDF formats
 - Multiple languages
- The problem
 - Search takes several minutes
 - 3000+ users familiar with the old system
 - No foundation for innovation



The Solution – Apache Solr

- Flexible and extendable
 - This is only the first wave of development
 - A need for complex business rules to drive the search – Boosts & FunctionQueries

The Solution – Apache Solr

- Flexible and extendable
 - This is only the first wave of development
 - A need for complex business rules to drive the search – Boosts & FunctionQueries
- Economically scalable
 - Much more data to come
 - Too hard to predict future cost of commercial, closed source alternatives

The Solution – Apache Solr

- Flexible and extendable
 - This is only the first wave of development
 - A need for complex business rules to drive the search – Boosts & FunctionQueries
- Economically scalable
 - Much more data to come
 - Too hard to predict future cost of commercial, closed source alternatives
- Great support available - from



A flexible pipeline - in two parts

A flexible pipeline - in two parts

1. Indexer

- Reads an XML settings file
- Extracts data from Oracle
- Processes if necessary
- Adds to a Solr index

A flexible pipeline - in two parts

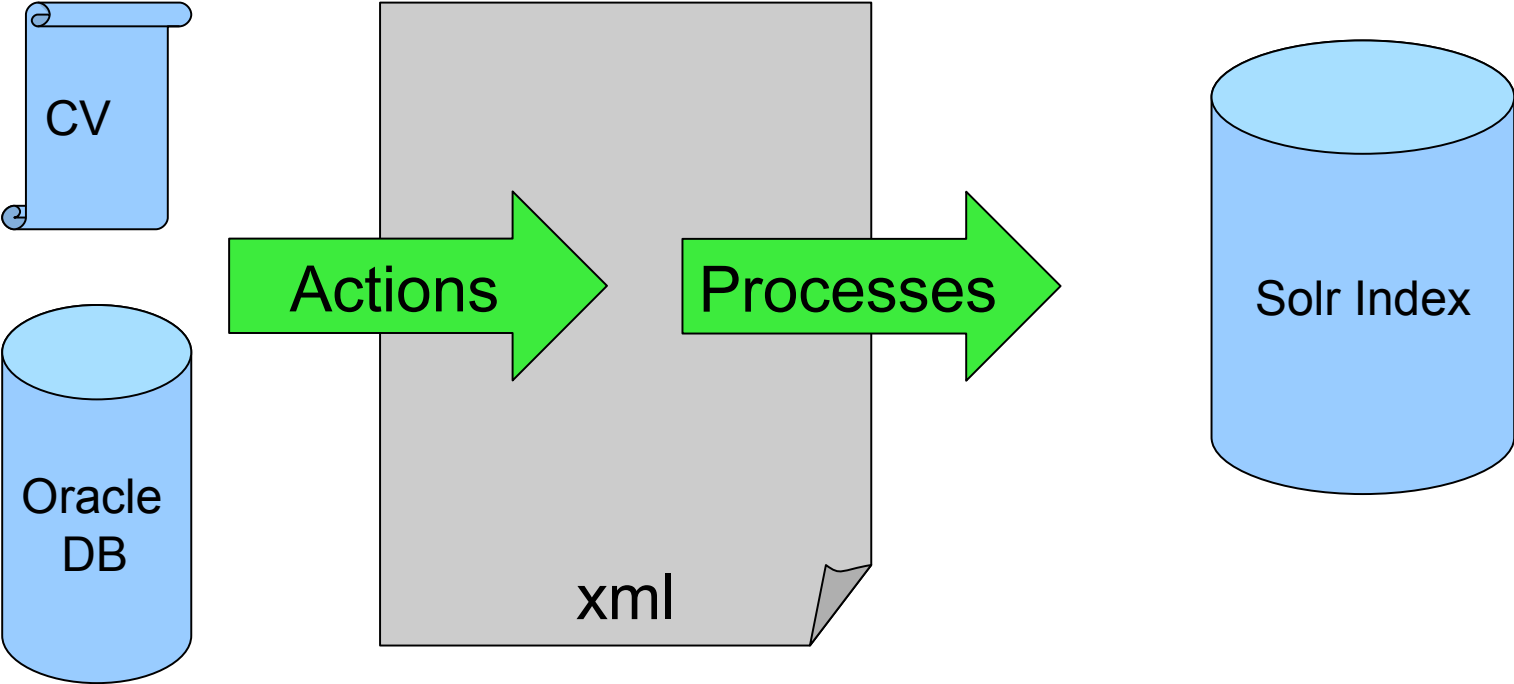
1. Indexer

- Reads an XML settings file
- Extracts data from Oracle
- Processes if necessary
- Adds to a Solr index

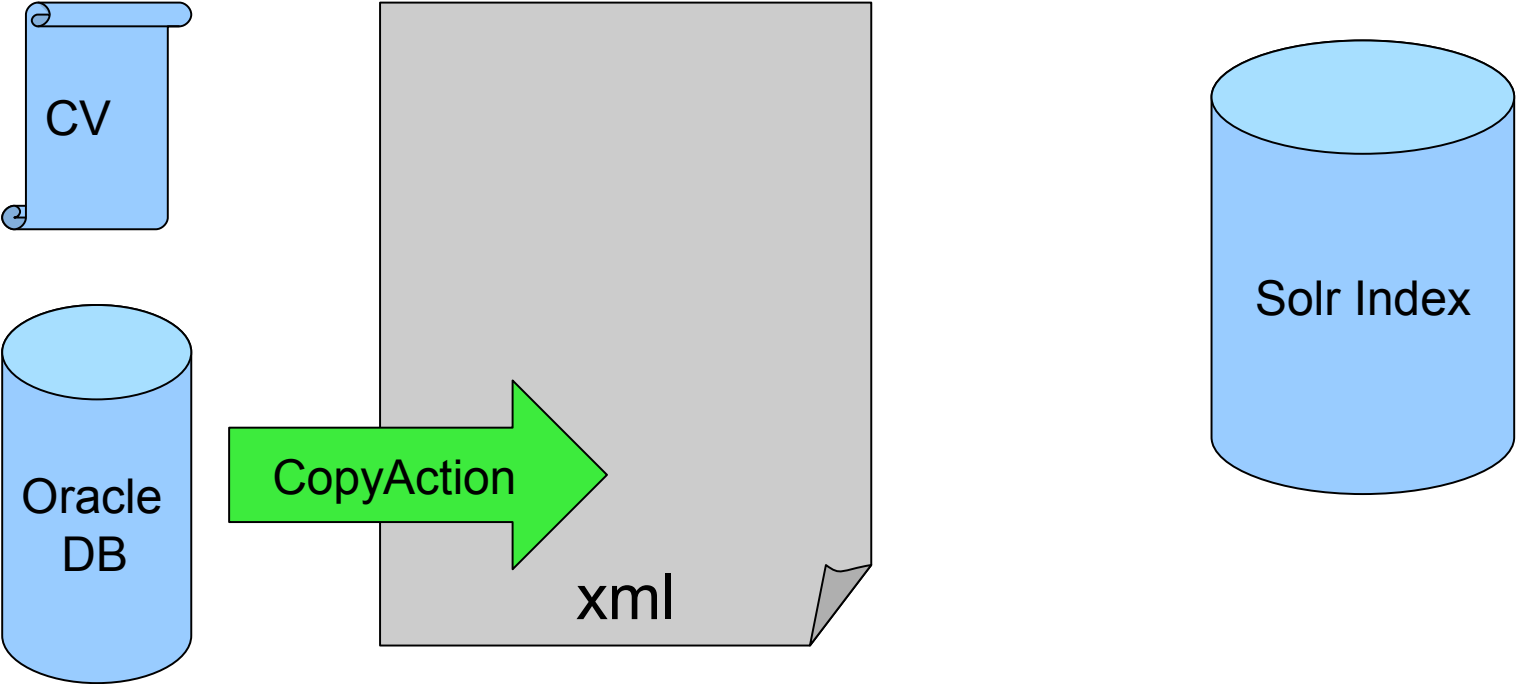
2. Config tool

- Creates a Solr schema from the Indexer settings
- Verifies types and checks for conflicts

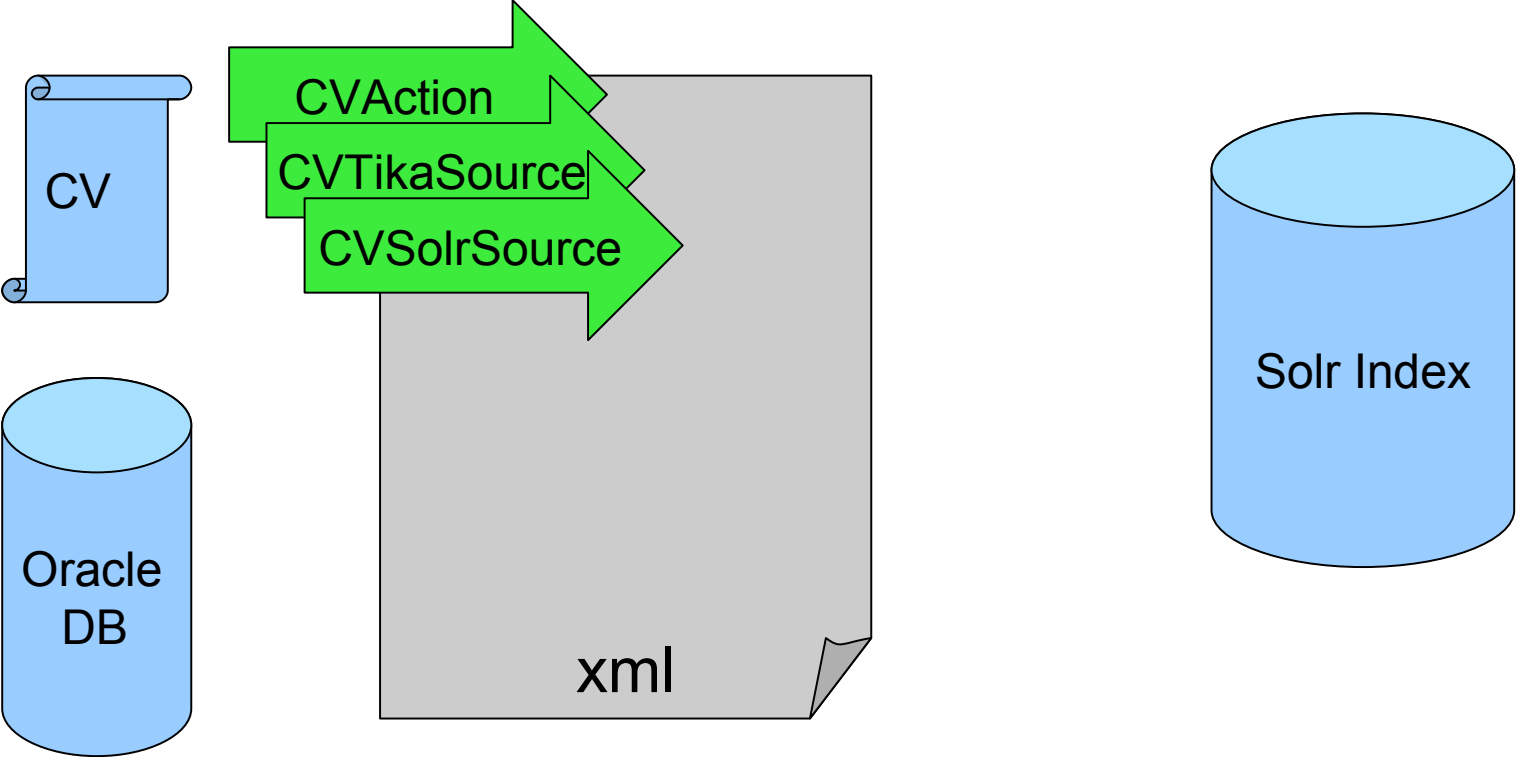
The Indexer



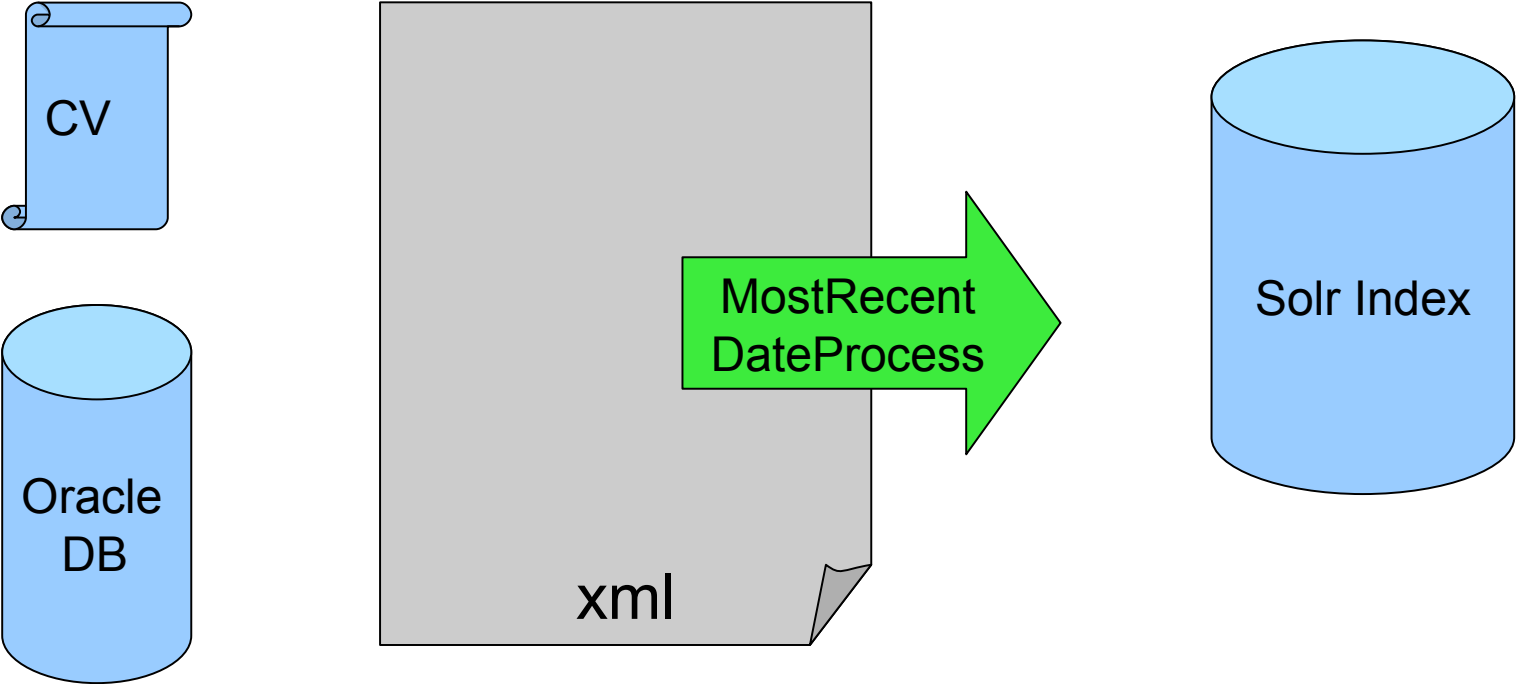
The Indexer



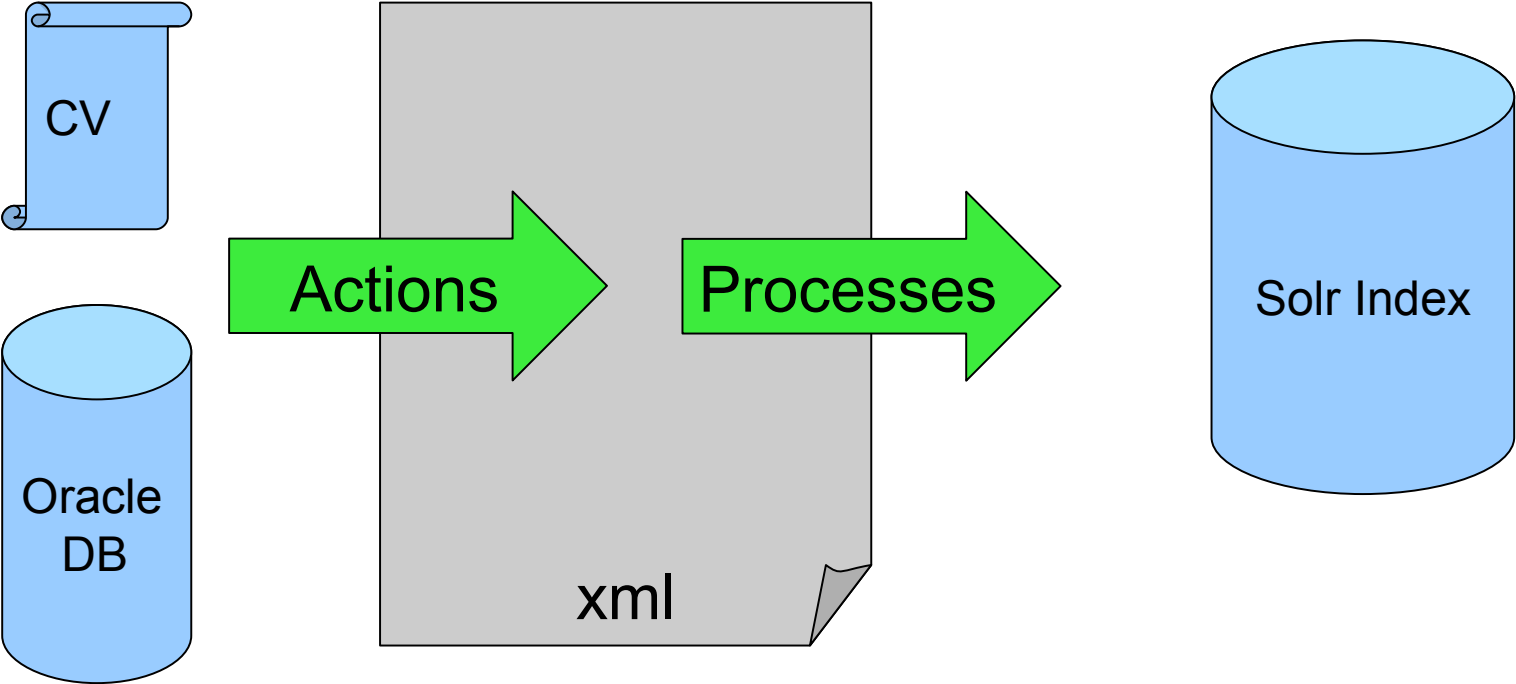
The Indexer



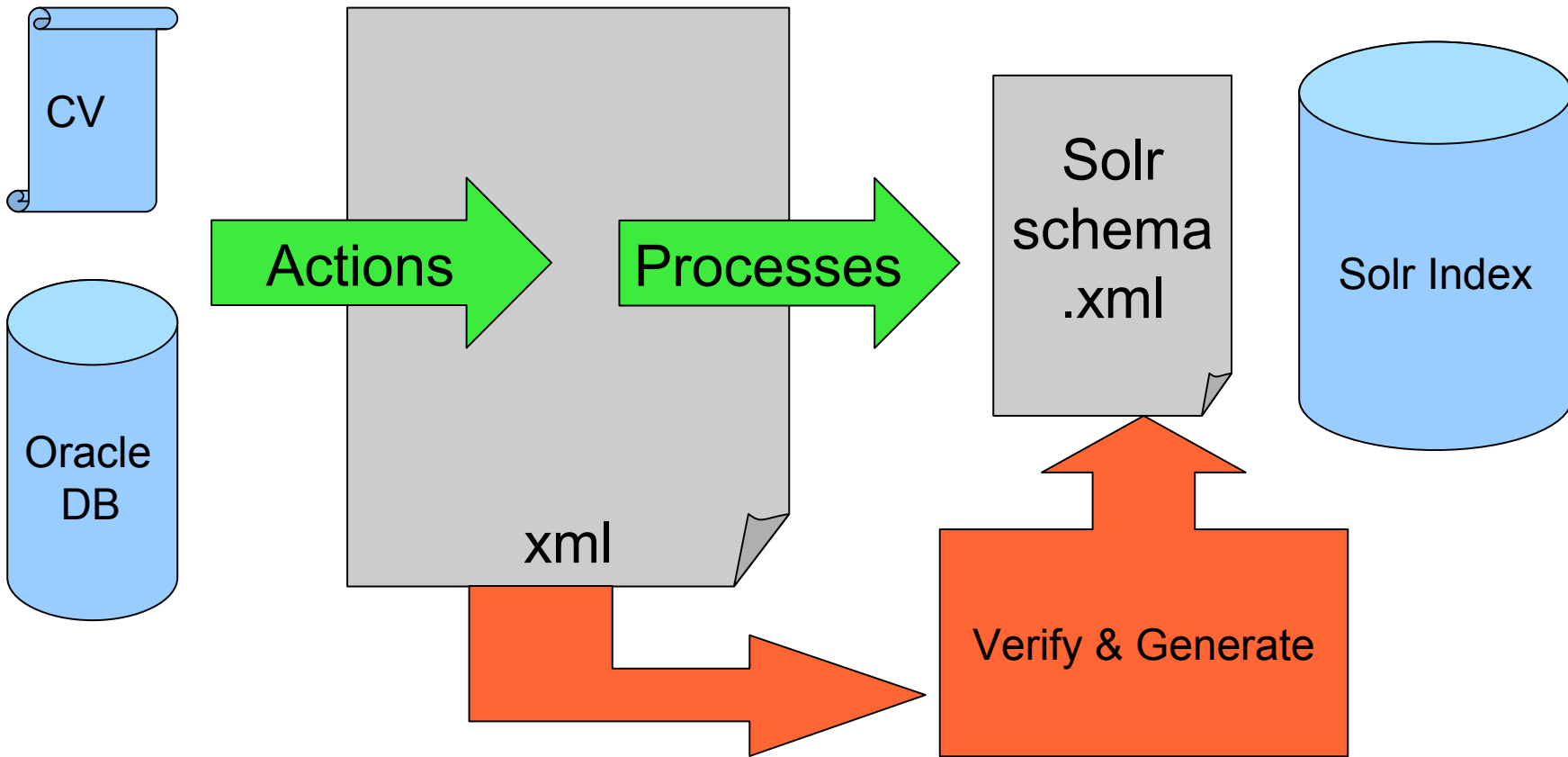
The Indexer



The Indexer



The Indexer & The Config Tool



The pipeline in code...

Actions

```
<action ref="copyAction" column="EMAIL" field="email" />
```

Processes

```
<process-map>  
  <process field="boost_date">  
    <beans:bean class="...MostRecentDateProcess">  
      ...  
      <beans:value>updateddate</beans:value>  
      <beans:value>createddate</beans:value>  
      ...  
    </process>  
</process-map>
```

The pipeline in code...

Actions

```
<action ref="copyAction" column="EMAIL" field="email"
type="string" indexed="true" stored="true"/>
```

Processes

```
<process-map>
  <process field="boost_date" type="tdate"
indexed="true" stored="false">
    <beans:bean class="...MostRecentDateProcess">
      ...
      <beans:value>updateddate</beans:value>
      <beans:value>createddate</beans:value>
    </beans:bean>
  </process>
</process-map>
```

...and a Solr schema

```
<?xml version="1.0" encoding="UTF-8" ?>
  <schema>
    <fields>
      <field name="email" type="string" indexed="true"
stored="true" />
      <field name="boost_date" type="tdate" indexed="true"
stored="false"/>
    </fields>
  </schema>
```

Transforming the UI



A world class desktop for recruitment professionals

reed.co.uk | reedglobal.com | Your Details | Contact Us | Help | About Xpress | iLearnIT | 

Switch: TEMP75701.. 0757 Charing Cross Nurse

Search internal directory

Wed, 12 October 2011 - Good Morning, Neil

My taskbar

Menu Lists

Candidates : Results ->

Home

- Jobs
- Candidates
- Managers
- Business
- Shortlist
- QuickLinks

The request to this fragment failed.

Results

Total No of Records: 500

<input type="checkbox"/>	<u>Name</u> <u>Req ID</u> <u>Reed Global ID</u> <u>Registered</u> <u>Client Xpress ID</u>	<u>Activity</u> <u>Activation</u> <u>ID</u> <u>Activated</u>	<u>Last Login</u> <u>Last</u> <u>Contacted</u> <u>Availability</u>	<u>Reed Title</u> <u>Current Job Title</u> <u>Sector</u> <u>Emp Type</u>	<u>Exp</u> <u>Salary</u> <u>Min Rate</u> <u>Available</u>	<u>Transport</u> <u>Home Town</u> <u>Branch</u> <u>Visa Expiry Date</u>
<input type="checkbox"/>	4781220 03/03/2008 	0 6193695 --/--/----	10/12/2008 PROSPECTIVE	Air Cabin Crew BANKING TEMPORARY	12,000 8.00	FULL LICENCE/VEHICLE WASHINGTON 0000 - ONLINE
<input type="checkbox"/>	4779947 03/03/2008 	0 6191730 --/--/----	07/07/2008 PROSPECTIVE	at the moment no wor ACCOUNTANCY BOTH	9,600 5.00	FULL LICENCE/VEHICLE HARROW 0000 - ONLINE
<input type="checkbox"/>	4780623 03/03/2008 	0 6192777 --/--/----	27/05/2008 PROSPECTIVE	Admin Assistant ADMIN,SECRETARIAL & PA PERMANENT	16,000 9.00	FULL LICENCE/VEHICLE BRIDLINGTON 0000 - ONLINE
<input type="checkbox"/>	4779772 03/03/2008	0 6191384 --/--/----	09/05/2008 PROSPECTIVE	retail HEALTH AND MEDICINE PERMANENT	13,000 700.00	FULL LICENCE/VEHICLE GILLINGHAM 0000 - ONLINE

Transforming the UI



REED x2 Switch: TEMP11101.. 0111 Baker Street Specialist Search Lock to My Office

Search Candidate Manager Jobs Business Admin Activity External

Candidates > keyword: *, office:9223, avail-status:PROSPECTIVE SELF SUITED ... Refine Search

43 results Show me: 10 results Sort by: Last Updated

1 2 3 4 5

View CV **View Details**

LONDON, UNITED KINGDOM Activation ID: 7319450 Req ID:
Branch: 9223 - It Projects Consultant:
Team
15.0 PERMANENT Visa Exp: Registered: 27/10/2008
Availability: PROSPECTIVE Transport:
Last Login: Last Contact:

[Simple, tax effective giving to the charity of your choice through your payroll Welcome to our Give As You Earn Scheme Give As You Earn is a simple, efficient and flexible scheme, which allows you to donate to any charity of

Tags: [ADMINISTRATOR]

Show Results

- Search
- Add
- Tasks

TOP-TIPS

- POWERFUL WILDCARD SEARCHING
- TRACKING AGAINST MULTIPLE CANDIDATES

Filter by Availability Status:

- PROSPECTIVE (41)
- AVAILABLE (10)
- SELF SUITED (2)
- STOP LIST (2)

Filter by Desired Salary:

- below 20K (5)
- 20K-25K (1)
- 25K-30K (1)
- 30K-40K (0)
- 40K+ (1)

Filter by Date Registered:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (1)

Filter by Last Login:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (0)

Filter by Office:

- ONLINE (537077)
- ST ALBANS (19891)
- ENGINEERING
- GLOBAL RESOURCING MELB (18477)
- Dublin Specialist (16665)
- MANCHESTER (14204)
- CONSTRUCTION
- Brighton Specialist (13831)
- Manchester Specialist (13711)
- Guildford Testhouse (12045)

CONSULTANT **View Details**

INTEGRATION ENGINEER **View Details**

CHEMICAL ANALYST **View Details**

BAR TENDER **View Details**

ANALYST PROGRAMMER **View Details**


PERSONNEL OFFICER **View Details**

Candidates > keyword: *, office:9223, avail-status:PROSPECTIVE SELF SUITED ...

 Refine Search

43 results

Show me: 10 results

Sort by: **Last Updated** 

Navigation: 1 2 3 4 5

CONSULTANT  [View CV](#) [View Details](#)

LONDON, UNITED KINGDOM Activation ID: 7319450 Req ID:
Branch: 9223 - It Projects Consultant:
Team

15.0 Visa Exp: Registered: 27/10/2008
PERMANENT Availability: PROSPECTIVE Transport:
Last Login: Last Contact:

[Simple, tax effective giving to the charity of your choice through your payroll Welcome to our Give As You Earn Scheme Give As You Earn is a simple, efficient and flexible scheme, which allows you to donate to any charity of

Tags: [ADMINISTRATOR]

- CONSULTANT**  [View Details](#)
- INTEGRATION ENGINEER**  [View Details](#)
- CHEMICAL ANALYST**  [View Details](#)
- BAR TENDER**  [View Details](#)
- ANALYST PROGRAMMER**  [View Details](#)
- PERSONNEL OFFICER**  [View Details](#)
- PROJECT MANAGER**  [View CV](#) [View Details](#)

Filter by Availability Status:

- PROSPECTIVE (41)
- AVAILABLE (10)
- SELF SUITED (2)
- STOP LIST (2)

Filter by Desired Salary:

- below 20K (5)
- 20K-25K (1)
- 25K-30K (1)
- 30K-40K (0)
- 40K+ (1)

Filter by Date Registered:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (1)

Filter by Last Login:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (0)

Filter by Office:

- ONLINE (537077)
- ST ALBANS (19891)
- ENGINEERING
- GLOBAL RESOURCING MELB (18477)
- Dublin Specialist (16665)
- MANCHESTER (14204)
- CONSTRUCTION
- Brighton Specialist (13831)
- Manchester Specialist (13711)
- Guildford Technology (13018)
- Manchester Technology (12729)
- LONDON & CITY (12194)
- INSURANCE
- Nottingham Specialist (12087)
- Chancery Lane Specialist (11917)
- Croydon Insurance (11860)
- Baker Street Specialist (5413)
- IT Projects Team (43)

Show Results 

Search

Add


Tasks

TOP-TIPS 

- POWERFUL WILDCARD SEARCHING
- TRACKING AGAINST MULTIPLE CANDIDATES
- AND MANY MORE...

READ NOW!

Candidates > keyword: *, office:9223, avail-status:PROSPECTIVE SELF SUITED ...

 Refine Search

43 results

Show me: 10 results Sort by: Last Updated

1 2 3 4 5

LONDON, UNITED KINGDOM

15.0 PERMANENT

Activation ID: 7319450

Branch: 9223 - It Projects Team

Visa Exp:

Availability: PROSPECTIVE

Last Login:

Req ID:

Consultant:

Registered: 27/10/2008

Transport:

Last Contact:

[Simple, tax effective giving to the charity of your choice through your payroll Welcome to our Give As You Earn Scheme Give As You Earn is a simple, efficient and flexible scheme, which allows you to donate to any charity of

Tags: [ADMINISTRATOR]

Filter by Availability Status:

- PROSPECTIVE (41)
- AVAILABLE (10)
- SELF SUITED (2)
- STOP LIST (2)

Filter by Desired Salary:

- below 20K (5)
- 20K-25K (1)
- 25K-30K (1)
- 30K-40K (0)
- 40K+ (1)

Filter by Date Registered:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (1)

Filter by Last Login:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (0)

Filter by Office:

- ONLINE (537077)
- ST ALBANS (19891)
- ENGINEERING
- GLOBAL RESOURCING MELB (18477)
- Dublin Specialist (16665)
- MANCHESTER (14204)
- CONSTRUCTION
- Brighton Specialist (13831)
- Manchester Specialist (13711)
- Guildford Technology (13018)
- Manchester Technology (12729)
- LONDON & CITY (12194)
- INSURANCE
- Nottingham Specialist (12087)
- Chancery Lane Specialist (11917)
- Croydon Insurance (11860)
- Baker Street Specialist (5413)
- IT Projects Team (43)

Show Results

Search

Add

Tasks

TOP-TIPS

- POWERFUL WILDCARD SEARCHING
- TRACKING AGAINST MULTIPLE CANDIDATES
- AND MANY MORE...

READ NOW!

- CONSULTANT** View Details (+)
- INTEGRATION ENGINEER** View Details (+)
- CHEMICAL ANALYST** View Details (+)
- BAR TENDER** View Details (+)
- ANALYST PROGRAMMER** View Details (+)
- PERSONNEL OFFICER** View Details (+)
- PROJECT MANAGER** View CV View Details (+)

Candidates > keyword: *, office:9223, avail-status:PROSPECTIVE SELF SUITED ...

 Refine Search

43 results

Show me: 10 results Sort by: Last Updated

1 2 3 4 5

 [View CV](#) [View Details](#)

LONDON, UNITED KINGDOM Activation ID: 7319450 Req ID:
 Branch: 9223 - It Projects Team Consultant:
 15.0 Visa Exp: Registered: 27/10/2008
 PERMANENT Availability: PROSPECTIVE Transport:
 Last Login: Last Contact:

[Simple, tax effective giving to the charity of your choice through your payroll Welcome to our Give As You Earn Scheme Give As You Earn is a simple, efficient and flexible scheme, which allows you to donate to any charity of

Tags: [ADMINISTRATOR]

-  [View Details](#) **CONSULTANT**
-  [View Details](#) **INTEGRATION ENGINEER**
-  [View Details](#) **CHEMICAL ANALYST**
-  [View Details](#) **BAR TENDER**
-  [View Details](#) **ANALYST PROGRAMMER**
-  [View Details](#) **PERSONNEL OFFICER**
-  [View CV](#) [View Details](#) **PROJECT MANAGER**

Filter by Availability Status:

- PROSPECTIVE (41)
- AVAILABLE (10)
- SELF SUITED (2)
- STOP LIST (2)

Filter by Desired Salary:

- below 20K (5)
- 20K-25K (1)
- 25K-30K (1)
- 30K-40K (0)
- 40K+ (1)

Filter by Date Registered:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (1)

Filter by Last Login:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (0)

Filter by Office:


- ONLINE (537077)
- ST ALBANS (19891)
- ENGINEERING
- GLOBAL RESOURCING MELB (18477)
- Dublin Specialist (16665)
- MANCHESTER (14204)
- CONSTRUCTION
- Brighton Specialist (13831)
- Manchester Specialist (13711)
- Guildford Technology (13018)
- Manchester Technology (12729)
- LONDON & CITY (12194)
- INSURANCE
- Nottingham Specialist (12087)
- Chancery Lane Specialist (11917)
- Croydon Insurance (11860)
- Baker Street Specialist (5413)
- IT Projects Team (43)

Show Results

Search

Add

Task

TOP-TIPS 

- POWERFUL WILDCARD SEARCHING
- TRACKING AGAINST MULTIPLE COMPANIES
- AND MANY MORE...

[READ NOW!](#)

Candidates > keyword: *, office:5223 avail-status:PROSPECTIVE SELF SUITED ...

 Refine Search

43 results

Show me: 10 results Sort by: Last Updated

1 2 3 4 5

 [View CV](#) [View Details](#)

LONDON, UNITED KINGDOM Activation ID: 7319450 Req ID:

15.0 Branch: 9223 - It Projects Team Consultant:

PERMANENT Visa Exp: Registered: 27/10/2008

Availability: PROSPECTIVE Transport:

Last Login: Last Contact:

[Simple, tax effective giving to the charity of your choice through your payroll Welcome to our Give As You Earn Scheme Give As You Earn is a simple, efficient and flexible scheme, which allows you to donate to any charity of

Tags: [ADMINISTRATOR]

-  [View Details](#) (+) **CONSULTANT**
-  [View Details](#) (+) **INTEGRATION ENGINEER**
-  [View Details](#) (+) **CHEMICAL ANALYST**
-  [View Details](#) (+) **BAR TENDER**
-  [View Details](#) (+) **ANALYST PROGRAMMER**
-  [View Details](#) (+) **PERSONNEL OFFICER**
-  [View CV](#) [View Details](#) (+) **PROJECT MANAGER**

Filter by Availability Status:

- PROSPECTIVE (41)
- AVAILABLE (10)
- SELF SUITED (2)
- STOP LIST (2)

Filter by Desired Salary:

- below 20K (5)
- 20K-25K (1)
- 25K-30K (1)
- 30K-40K (0)
- 40K+ (1)

Filter by Date Registered:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (1)

Filter by Last Login:

- Last Day (0)
- Last Week (0)
- Last Fortnight (0)
- Last Month (0)

Filter by Office:

- ONLINE (537077)
- ST ALBANS (19891)
- ENGINEERING
- GLOBAL RESOURCING MELB (18477)
- Dublin Specialist (16665)
- MANCHESTER (14204)
- CONSTRUCTION
- Brighton Specialist (13831)
- Manchester Specialist (13711)
- Guildford Technology (13018)
- Manchester Technology (12729)
- LONDON & CITY (12194)
- INSURANCE
- Nottingham Specialist (12087)
- Chancery Lane Specialist (11917)
- Croydon Insurance (11860)
- Baker Street Specialist (5413)
- IT Projects Team (43)

Show Results

Search

Add

Tasks

TOP-TIPS

- POWERFUL WILDCARD SEARCHING
- TRACKING AGAINST MULTIPLE CANDIDATES
- AND MANY MORE...
- READ NOW!

Performance

- Many factors can affect search performance...

Performance

- Many factors can affect search performance...
- ...so we built a test framework
 - Randomly generated queries based on terms in the index
 - Average query times & number of results recorded
 - Allows for direct comparison of boost functions, for example

Performance...much improved!



- Sub-second searches
- Only a single server required
- So fast that the thin client hardware had to be upgraded as it became a bottleneck!
- Still work to be done on improving indexing speed

Issues

- Users don't always understand their new freedoms
 - Training can be required on free text search, faceting...
 - Any issues reduce user confidence in new systems

Issues

- Users don't always understand their new freedoms
 - Training can be required on free text search, faceting...
 - Any issues reduce user confidence in new systems
- Solr features can conflict with each other
 - Make sure you understand how features interact – i.e. recency over relevance, synonyms, stopwords
 - Get the basics working first

Results & benefits

- Project delivered on time and under budget
- Now live across 350 offices UK & worldwide
- 24/7/365 support provided by Lucid Imagination

Results & benefits

- Project delivered on time and under budget
- Now live across 350 offices UK & worldwide
- 24/7/365 support provided by Lucid Imagination
- **A very happy client!**

Conclusions & Lessons Learned

- What we learned
 - A flexible pipeline is essential
 - Get the basics working first - watch out for feature conflict

Conclusions & Lessons Learned

- What we learned
 - A flexible pipeline is essential
 - Get the basics working first - watch out for feature conflict
- What Reed learned
 - User training is important - even if the new system is “simpler”
 - To love Open Source Search...

Conclusions & Lessons Learned

"The transition to Solr was the latest step in our strategy to develop a truly worldclass search application. We believe it provides a robust architecture that meets our future aims, it will scale economically and is a welcome addition to our existing suite of Open Source systems."

The End

- Thanks for listening!
- For more information please contact me:

Charlie Hull, Managing Director, Flax
charlie@flax.co.uk
<http://www.flax.co.uk/blog>
@FlaxSearch