



# Building a Real-time, Big Data Analytics Platform with Solr

Trey Grainger

Manager, Search Technology Development

@careerbuilder®

Lucene Revolution 2013 - San Diego, CA

# Overview

- Intro to Analytics with Solr
- Real-world examples & Faceting deep dive
- Solr enhancements we're contributing:
  - Distributed Pivot Faceting
  - Pivoted Percentile/Stats Faceting
- Data Analytics with Solr... the next frontier

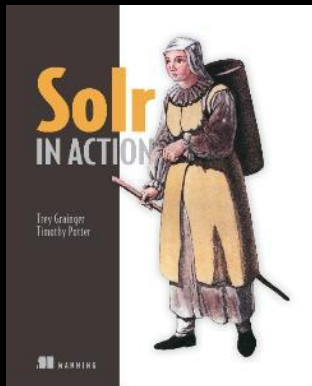
# My Background

Trey Grainger

Manager, Search Technology Development  
@ **CareerBuilder**.com

## Relevant Background

- Search & Recommendations
- High-volume, Distributed Systems
- NLP, Relevancy Tuning, User Group Testing, & Machine Learning



## Other Projects

- Co-author: *Solr in Action*
- Founder and Chief Engineer @ *Celaccess*.com

# About Search @careerbuilder®

- Over 2.5 million new jobs each month
- Over 50 million actively searchable resumes
- ~300 globally distributed search servers (in the U.S., Europe, & Asia)
- Thousands of unique, dynamically generated indexes
- Over ½ Billion actively searchable documents
- Over 1 million searches an hour

# Our Search Platform

- Generic Search API wrapping Solr + our domain stack
- Goal: Abstract away search into a simple API so that any engineer can build search-based product with no prior search background
- 3 Supported Methods (with rich syntax):
  - AddDocument
  - DeleteDocument
  - Search

*\*users pass along their own dynamically-defined schemas on each call*
- Building out as Restful API so search can be used “from anywhere”

# Paradigm shift?

- Originally, search engines were designed to return a ranked list of relevant documents

[Search Again](#) [Advanced Search](#)

Keywords:  Location:  Posted within:  [Find Jobs](#)

### Search Results

Page 1 of 160 | [Next Page](#)

	Job Title / Description ( <a href="#">show titles only</a> )	Company	Location (Distance)	Posted
	<b>Software Engineer – Search Technology Software Developer</b> - <a href="#">View similar jobs</a> Job type: Full-Time <b>Solr</b> to create an environment to cluster documents and retrieve high-level, meaningful information. We ... search experience with Fast, ... <a href="#">View full job description</a> <a href="#">Save to MyCareerBuilder</a> <a href="#">Email to a friend</a>	<b>CareerBuilder-US</b>	GA - Norcross	1 Week Ago
	<b>Java Developer with SOLR/Lucene</b> - <a href="#">View similar jobs</a> Job type: Contractor Required Qualifications:Skills Requirements:2 years of experience with JAVA, <b>SOLR</b> /Lucene... <a href="#">View full job description</a> <a href="#">Save to MyCareerBuilder</a> <a href="#">Email to a friend</a>	<b>V.L.S. Systems, Inc.</b>	NC - Raleigh	2 Weeks Ago
	<b>Director Product Management - Internal Site Search</b> - <a href="#">View similar jobs</a> Job type: Full-Time Linguistics Site Search Platforms – <b>Solr</b> Preferred• Ability to use analytics and competitive research to ... -line results• Ability... <a href="#">View full job description</a> <a href="#">Save to MyCareerBuilder</a> <a href="#">Email to a friend</a>	<b>Sears Holdings Management Corporation</b>	IL - Hoffman Estates	1 Week Ago

# Paradigm shift?

- Then, features like faceting were added to augment the search experience with aggregate information...

[Search Again](#) [Advanced Search](#)

Keywords:








Location:

Posted within:

[Find Jobs](#)

**Search Results**

Page 1 of 160 | [Next Page](#)

	Job Title / Description ( <a href="#">show titles only</a> )	Company	Location (Distance)	Posted
 	<b>Software Engineer – Search Technology Software Developer</b> - <a href="#">View similar jobs</a> Job type: Full-Time <b>Solr</b> to create an environment to cluster documents and retrieve high-level, meaningful information. We ... search experience with Fast, ... <a href="#">View full job description</a> <a href="#">Save to MyCareerBuilder</a> <a href="#">Email to a friend</a>	 <b>CareerBuilder-US</b>	GA - Norcross	1 Week Ago
 	<b>Java Developer with SOLR/Lucene</b> - <a href="#">View similar jobs</a> Job type: Contractor Required Qualifications:Skills Requirements:2 years of experience with JAVA, <b>SOLR</b> /Lucene... <a href="#">View full job description</a> <a href="#">Save to MyCareerBuilder</a> <a href="#">Email to a friend</a>	<b>V.L.S. Systems, Inc.</b>	NC - Raleigh	2 Weeks Ago
 	<b>Director Product Management - Internal Site Search</b> - <a href="#">View similar jobs</a> Job type: Full-Time Linguisticso Site Search Platforms – <b>Solr</b> Preferred• Ability to use analytics and competitive research to ... -line results• Ability... <a href="#">View full job description</a> <a href="#">Save to MyCareerBuilder</a> <a href="#">Email to a friend</a>	<b>Sears Holdings Management Corporation</b>	IL - Hoffman Estates	1 Week Ago

# Paradigm shift?

- But... what kinds of products could you build if you only cared about the aggregate calculations?

[Search Again](#) [Advanced Search](#)

Keywords:

Location:

Posted within:

[Find Jobs](#)

### Narrow Search

Category:	Company:	City:	State:
Management (81413)	Army National Guard (10389)	Houston (6445)	California (33967)
Sales (75209)	Pizza Hut (6242)	Chicago (5492)	Texas (33603)
Health Care (58682)	Aerotek (3429)	New York (4992)	Florida (22269)
Customer Service (52322)	UnitedHealth Group (3155)	Phoenix (3894)	Illinois (17988)
Information Technology (40010)	Confidential (2608)	Dallas (3881)	New York (15852)
Skilled Labor - Trades (36463)	CardPayment.Solutions (2100)	Atlanta (3781)	Pennsylvania (15417)

### Search Results

Page 1 of 160 | [Next Page](#)

Job Title / Description ( <a href="#">show titles only</a> )	Company	Location (Distance)	Posted
<div></div> <b>Software Engineer – Search Technology Software Developer</b> - <a href="#">View similar jobs</a> Job type: Full-Time Solr to create an environment to cluster documents and retrieve high-level, meaningful information. We ... search experience with Fast, ... <a href="#">View full job description</a> <a href="#">Save to MyCareerBuilder</a> <a href="#">Email to a friend</a>	CareerBuilder-US	GA - Norcross	1 Week Ago
<div></div> <b>Java Developer with SOLR/Lucene</b> - <a href="#">View similar jobs</a> Job type: Contractor Required Qualifications:Skills Requirements:2 years of experience with JAVA, SOLR/Lucene... <a href="#">View full job description</a> <a href="#">Save to MyCareerBuilder</a> <a href="#">Email to a friend</a>	V.L.S. Systems, Inc.	NC - Raleigh	2 Weeks Ago



# Workforce Supply & Demand

## Supply & Demand Report

[What's New](#)[Create a New Report](#)[Compensation Report](#)[Saved Searches](#)[Export](#)[Help](#)

### Applied Filters

☒ Construction Laborers (655)

[Clear all filters...](#)

### Filter By

#### Occupation (ONET)

- ☐ All Occupations
- ☐ Registered Nurses (28,332)
- ☐ Retail Salespersons (16,712)
- ☐ First-Line Supervisors/Managers of Retail Sales Workers (11,949)
- ☐ First-Line Supervisors/Managers of Food Preparation and Serving Workers (11,344)
- ☐ Customer Service Representatives (10,865)
- ☐ Telemarketers (10,304)
- ☐ Executive Secretaries and Administrative Assistants (9,177)
- ☐ Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products (7,264)
- ☐ First-Line Supervisors/Managers of Office and Administrative Support Workers (7,026)
- ☐ Waiters and Waitresses (5,910)
- ☒ Construction Laborers (655)

#### Company

- ☒ All Companies
- ☐ Harrah's Entertainment, Inc. (4)
- ☐ National Park Service (3)
- ☐ Lowe's Companies, Inc. (2)
- ☐ Republic Services, Inc. (2)
- ☐ Winder Farms (2)
- ☐ Wynn Resorts, Limited (2)
- ☐ American Electric Power Company, Inc. (1)
- ☐ American Greetings Corporation (1)
- ☐ Apptis Holdings, Inc. (1)
- ☐ Brinderson Engineering (1)

### Overview

Supply & Demand

### Supply

Workforce - Active and Total

### Demand

Job Postings

### Labor Pressure

Supply Over Demand

BETA

### Hiring Indicator

Compare Markets

## Construction Laborers

Las Vegas Nevada Area for April 2011 to March 2013

### Supply ?

437

Active Candidates

[View Resumes](#)

### Demand ?

655

Job Postings

5,562 Total Available Workforce ? ▼

### Hiring Indicator ?

69



## Job Postings and Recruitment

### Top Job Posters

Harrah's Entertainment, Inc.	4
National Park Service	3
Wynn Resorts, Limited	2
Winder Farms	2
Republic Services, Inc.	2
Lowe's Companies, Inc.	2
Brinderson Engineering	1
Apptis Holdings, Inc.	1
American Greetings Corporation	1
American Electric Power Company, Inc.	1

### Easiest Cities to Recruit

Henderson, NV

Las Vegas, NV

## Profile of Candidates Looking for a Job

### What company do they work for?

Perini Corporation	6
Laborers Union Local 304	6
Laborers Local	5
Labor Max Staffing	5

### What are their most recent job titles?

Laborer	56
General Laborer	9
Construction Worker	7
Construction Laborer	6

# Faceting Overview

/solr/select/?q=...&facet=true

//Field Faceting  
&facet.field=city

```
"facet_fields":{  
  "city":[  
    "new york, ny",2337,  
    "los angeles, ca",1693,  
    "chicago, il",1535,  
    ... ]}  
}
```

//Query Faceting:

```
&facet.query={!frange key="0 to 10 km" l=0 u=10 incll=false}geodist()  
&facet.query={!frange key="10 to 25 km" l=10 u=25 incll=false}geodist()  
&facet.query={!frange key="25 to 50 km" l=25 u=50 incll=false}geodist()  
&facet.query={!frange key="50+" l=50 incll=false}geodist()  
&sfield=location  
&pt=37.7770,-122.4200
```

```
"facet_queries":{  
  "0 to 10 km":1187,  
  "10 to 25 km":462,  
  "25 to 50 km":794,  
  "50+":105296  
},
```

//Range Faceting  
&facet.range=years\_experience  
&facet.range.start=0  
&facet.range.end=10  
&facet.range.gap=1  
&facet.range.other=after

```
"facet_ranges":{  
  "years_experience":{  
    "counts":[  
      "0",1010035,  
      "1",343831,  
      ...  
      "9",121090  
    ], ...  
    "after":59462}}  
}
```

# Multi-select Faceting with Tags / Excludes

## Faceting with No Filters Selected:

<b>► State</b> <input type="checkbox"/> Georgia (6) <input type="checkbox"/> California (4) <input type="checkbox"/> New York (4) <input type="checkbox"/> Texas (3) <input type="checkbox"/> Illinois (2) ...	<b>► Price Range</b> <input type="checkbox"/> < \$5 (6) <input type="checkbox"/> \$5 - \$10 (5) <input type="checkbox"/> \$10 - \$20 (3) <input type="checkbox"/> \$20 - \$50 (6) <input type="checkbox"/> \$50+ (0)	<b>► City</b> <input type="checkbox"/> Atlanta, GA (6) <input type="checkbox"/> New York, NY (4) <input type="checkbox"/> Austin, TX (3) <input type="checkbox"/> San Francisco (3) <input type="checkbox"/> Chicago (2) ...
--	---	--

&facet=true

&facet.field=state

&facet.field=price\_range

&facet.field=city

## Faceting with Filter of state:California selected:

<b>► State</b> <input checked="" type="checkbox"/> California (4)	<b>► Price Range</b> <input type="checkbox"/> < \$5 (0) <input type="checkbox"/> \$5 - \$10 (2) <input type="checkbox"/> \$10 - \$20 (0) <input type="checkbox"/> \$20 - \$50 (2) <input type="checkbox"/> \$50+ (0)	<b>► City</b> <input type="checkbox"/> San Francisco (3) <input type="checkbox"/> Los Angeles (1)
--	---	---

&facet=true

&facet.field=state

&facet.field=price\_range

&facet.field=city

&fq=state:California

## Multi-Select Faceting with state:California selected:

<b>► State</b> <input type="checkbox"/> Georgia (6) <input checked="" type="checkbox"/> California (4) <input type="checkbox"/> New York (4) <input type="checkbox"/> Texas (3) <input type="checkbox"/> Illinois (2) ...	<b>► Price Range</b> <input type="checkbox"/> < \$5 (0) <input type="checkbox"/> \$5 - \$10 (2) <input type="checkbox"/> \$10 - \$20 (0) <input type="checkbox"/> \$20 - \$50 (2) <input type="checkbox"/> \$50+ (0)	<b>► City</b> <input type="checkbox"/> San Francisco (3) <input type="checkbox"/> Los Angeles (1)
---	---	---

&facet=true

&facet.field={!ex="stateTag"}state

&facet.field=price\_range

&facet.field=city

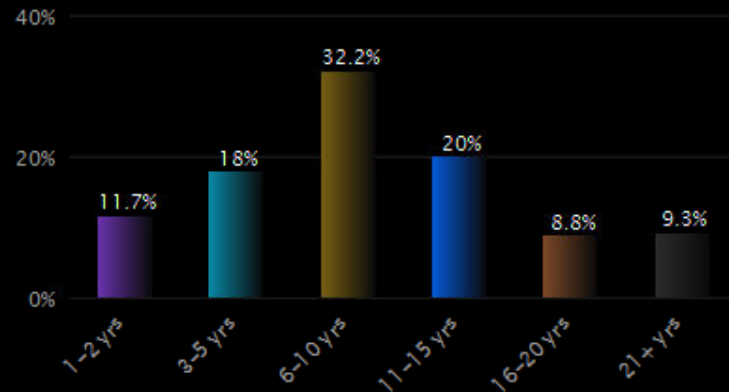
&fq={!tag="stateTag"}state:California

# Supply of Candidates

## Profile of Candidates Looking for a Job

[View Demographics Reports](#)[Company](#)[Experience](#)[Location](#)[Education](#)[Demographics](#)

### How many years of experience?



### What are their most recent job titles?

Laborer	56
General Laborer	9
Construction Worker	7
Construction Laborer	6
Union Laborer	5
Skilled Laborer	4
Crew Chief	2
Certified Union Laborer Local 872	2
Carpentry/Icf Foreman	2
Carpentry Helper	2

### Do they have management experience?



### What is their military experience?

Veteran	6
Retired Military	3
National Guard-Drilling	1
Active Duty	1

# Why Solr for Analytics?

- Allows “ad-hoc” querying of data by keywords
- Is good at on-the-fly aggregate calculations (facets + stats + functions + grouping)
- Solr is horizontally scalable, and thus able to handle billions of documents
- Insanely Fast queries, encouraging user exploration

# Supply of Candidates

## Supply & Demand Report

[What's New](#)[Create a New Report](#)[Compensation Report](#)[Saved Searches](#)[Export](#)[Help](#)

### Applied Filters

☒ Construction Laborers (655)  
[Clear all filters...](#)

### Filter By

#### Occupation (ONET)

- ☐ All Occupations
- ☐ Registered Nurses (28,332)
- ☐ Retail Salespersons (16,712)
- ☐ First-Line Supervisors/Managers of Retail Sales Workers (11,949)
- ☐ First-Line Supervisors/Managers of Food Preparation and Serving Workers (11,344)
- ☐ Customer Service Representatives (10,865)
- ☐ Telemarketers (10,304)
- ☐ Executive, Secretaries and Administrative Assistants (9,177)
- ☐ Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products (7,264)
- ☐ First-Line Supervisors/Managers of Office and Administrative Support Workers (7,026)
- ☐ Waiters and Waitresses (5,910)
- ☒ Construction Laborers (655)

#### Company

- ☒ All Companies
- ☐ Harrah's Entertainment, Inc. (4)
- ☐ National Park Service (3)
- ☐ Lowe's Companies, Inc. (2)
- ☐ Republic Services, Inc. (2)
- ☐ Winder Farms (2)
- ☐ Wynn Resorts, Limited (2)
- ☐ American Electric Power Company, Inc. (1)
- ☐ American Greetings Corporation (1)
- ☐ Apptis Holdings, Inc. (1)
- ☐ Brinderson Engineering (1)

### Overview

Supply & Demand

### Supply

Workforce - Active and Total

### Demand

Job Postings

### Labor Pressure

Supply Over Demand

### Hiring Indicator

Compare Markets

☒ For Active Candidates ☐ For Total Workforce

### Supply

Active Candidates Trend Over Last 24 Months

[View the Full Report](#)

Source: CareerBuilder.com

### Active Candidates Nationwide

☒ State ☐ MSA Compare by time period: **Annually**

Less More No Data

Rank	State	Apr. 2011 - Mar. 2012	Apr. 2012 - Mar. 2013	% Change
1	California	1,925	2,111	+9.66 %
2	Florida	1,697	1,957	+15.32 %
3	Illinois	1,481	1,746	+17.89 %
4	Pennsylvania	1,208	1,506	+24.67 %
5	Ohio	1,202	1,371	+14.06 %
6	Texas	1,177	1,367	+16.14 %
7	Indiana	1,110	1,247	+12.34 %
8	New York	932	1,086	+16.52 %
9	Michigan	747	969	+29.72 %
10	Georgia	669	854	+27.65 %

### Profile of Candidates Looking for a Job

[View Demographics Reports](#)

#### Company

What company do they work for?

#### Experience

#### Location

#### Education

#### Demographics

What industry do they work in?

# Demand for Jobs

Overview  
Supply & Demand

Supply  
Workforce - Active and Total

**Demand**  
Job Postings

Labor Pressure  
Supply Over Demand

BETA Hiring Indicator  
Compare Markets

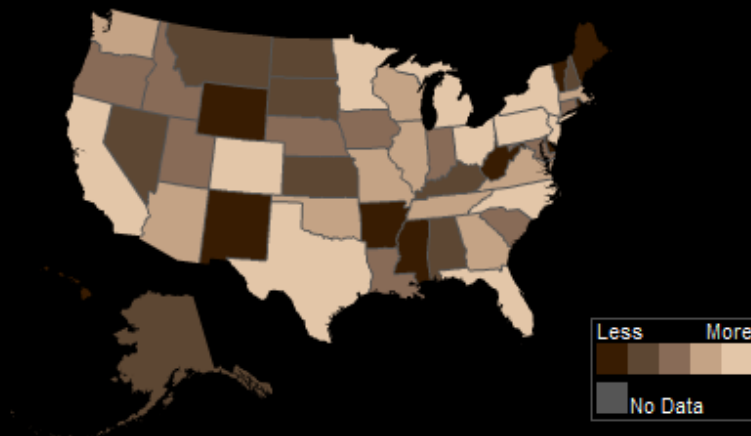
## Demand

### Job Posting Trend Over Last 24 Months [? What is this?](#)

[View the Full Report](#)

### Demand Geography Details

☒ State ☐ MSA Compare by time period: **Annually**



Rank	State	Apr, 2011 - Mar, 2012	Apr, 2012 - Mar, 2013	% Change
1	Texas	5,343	8,009	+49.90 %
2	California	4,743	6,466	+36.33 %
3	Florida	3,965	5,785	+45.90 %
4	Pennsylvania	4,816	4,616	-4.15 %
5	Ohio	4,022	4,517	+12.31 %
6	New York	3,880	4,434	+14.28 %
7	Michigan	2,984	3,637	+21.88 %
8	Colorado	1,961	3,407	+73.74 %
9	North Carolina	2,129	2,790	+31.05 %
10	New Jersey	2,344	2,764	+17.92 %
11	Minnesota	2,344	2,657	+13.35 %

Sources: CareerBuilder.com, Wanted Technologies

# Supply over Demand (Labor Pressure)

[Overview  
Supply & Demand](#)
[Supply  
Workforce - Active and Total](#)
[Demand  
Job Postings](#)
[Labor Pressure  
Supply Over Demand](#)

**BETA**
[Hiring Indicator  
Compare Markets](#)

## Labor Pressure

☒ For Active Candidates
 ☐ For Total Workforce

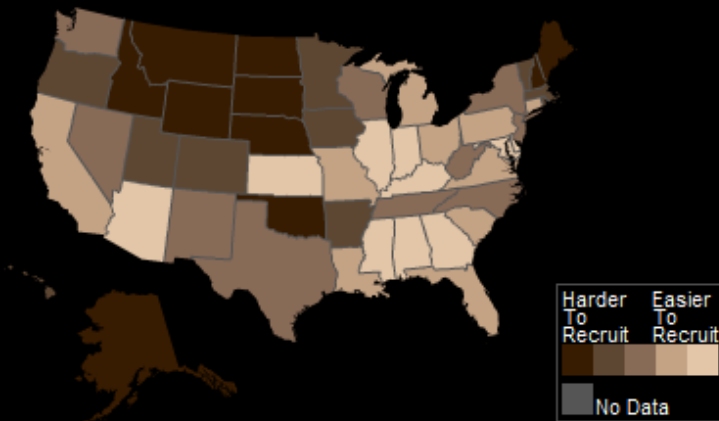
### Trend Over the Last 24 Months

[View the Full Report](#)



### Active Labor Pressure Geography Details

☒ State
 ☐ MSA
 Compare by time period: Annually



Rank	State	Apr, 2011 - Mar, 2012	Apr, 2012 - Mar, 2013	% Change
1	Illinois	1.05	0.92	-12.38 %
2	Delaware	0.94	0.88	-6.38 %
3	Indiana	0.94	0.84	-10.64 %
4	Kentucky	0.59	0.52	-11.86 %
5	Georgia	0.54	0.49	-9.26 %
6	Mississippi	0.56	0.46	-17.86 %
7	Kansas	0.54	0.45	-16.67 %
8	Maryland	0.41	0.44	+7.32 %
9	Arizona	0.51	0.40	-21.57 %
10	Alabama	0.45	0.39	-13.33 %
11	South Carolina	0.38	0.34	-10.53 %

Sources: CareerBuilder.com, Wanted Technologies

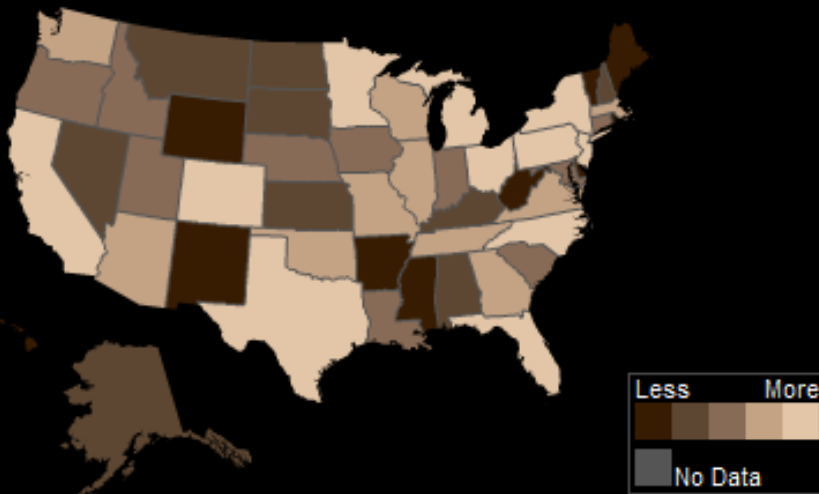


Wait, how'd you do that?

# Building Blocks...



/solr/select/?q=...&facet=true&facet.field=month\*



/solr/select/q=...&facet=true&facet.field=state

What is their military experience?

Veteran	6
Retired Military	3
National Guard-Drilling	1
Active Duty	1

/solr/select/?q=...&facet=true&  
facet.field=military\_experience

# Building Blocks...

## Profile of Candidates Looking for a Job

### What company do they work for?

Perini Corporation	6
Laborers Union Local 304	6
Laborers Local	5
Labor Max Staffing	5
Teamsters & Chauffeurs Local No 962	3
Modesto Junior College	3
Maycock Construction	3
Laborers International Local 89	3
Knoll Construction	3
Aaa Metal & Glass Inc	3

### What are their most recent job titles?

Laborer	56
General Laborer	9
Construction Worker	7
Construction Laborer	6
Union Laborer	5
Skilled Laborer	4
Crew Chief	2
Certified Union Laborer Local 872	2
Carpentry/Icf Foreman	2
Carpentry Helper	2

/solr/select/?

q="construction worker"&  
fq=city:"las vegas, nv"&  
facet=true&  
facet.field=company

/solr/select/?

q="construction worker"&  
fq=city:"las vegas, nv"&  
facet=true&  
facet.field=lastjobtitle

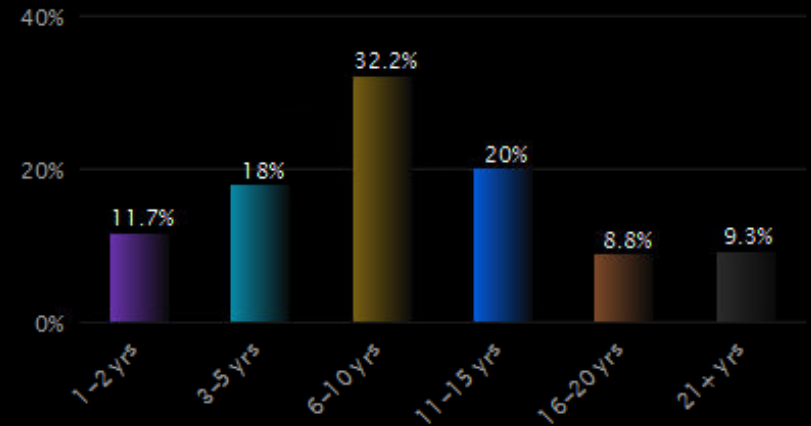
# Building Blocks...

Do they have management experience?



/solr/select/?q=...&facet=true&  
facet.field=management\_experience

How many years of experience?



/solr/select/? q=...&  
facet=true&facet.field=experience\_ranges

That's pretty basic... what else can you do?

# Radius Faceting

## Applied Filters

☒ Food Service Managers (2,253)

[Clear all filters...](#)

## Filter By

### Occupation (ONET)

- ☐ All Occupations
- ☐ Registered Nurses (29,372)
- ☐ Computer Systems Analysts (18,682)
- ☐ Web Developers (17,673)
- ☐ Truck Drivers, Heavy and Tractor-Trailer (17,147)
- ☐ Retail Salespersons (15,279)
- ☐ Accountants (12,949)
- ☐ Sales Managers (12,329)
- ☐ Computer Support Specialists (11,761)
- ☐ Customer Service Representatives (11,595)
- ☐ Executive Secretaries and Administrative Assistants (11,382)
- ☒ Food Service Managers (2,253)

### Company

- ☒ All Companies
- ☐ Aramark Corporation (73)
- ☐ Compass Group Llc (37)
- ☐ Compass Group International Llc (28)
- ☐ Compass Group Usa, Inc. (28)
- ☐ Compass Group Usa Investments Llp (17)
- ☐ Omni Hotels Corporation (15)
- ☐ Premier Solutions Inc (13)
- ☐ Panda Express, Inc. (12)
- ☐ Culinaire International, Inc. (11)
- ☐ Panda Restaurant Group, Inc. (11)

### Industry (NAICS)

### Staffing

Overview  
Supply & Demand

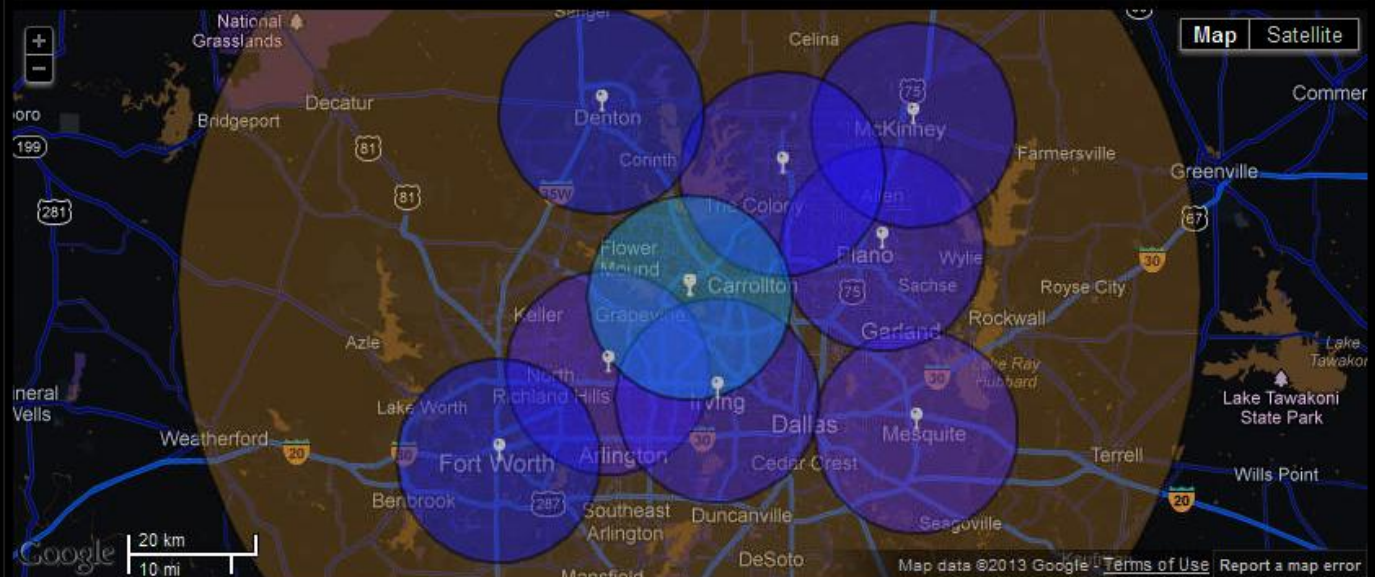
Supply  
Workforce - Active and Total

Demand  
Job Postings

Labor Pressure  
Supply Over Demand

**BETA** Hiring Indicator  
Compare Markets

## Nearby Cities with Hiring Indicator Data



Indicator	City	State	Distance (mi)	Active Supply	Demand	Compensation	Apps/Job
72	Mesquite	TX	25.72	168	235	38k - 60k	Average
72	Bedford	TX	10.86	112	189	38k - 60k	Below Average
70	Irving	TX	10.4	295	500	38k - 60k	Average
68	Plano	TX	19.36	372	660	38k - 60k	Above Average
64	Mckinney	TX	27.53	184	265	38k - 60k	Below Average
63	Frisco	TX	15.06	133	311	38k - 60k	Average
55	Fort Worth	TX	24.59	1,354	1,838	38k - 60k	Above Average
49	Denton	TX	20.01	182	371	38k - 60k	Below Average
18	Dallas	TX	--	1,612	12,445	38k - 60k	Above Average
0	Westlake	TX	11.79	0	122	38k - 60k	Insufficient Data

Harder to Recruit Easier to Recruit Insufficient Comparable Data



# Hiring Comparison per Market

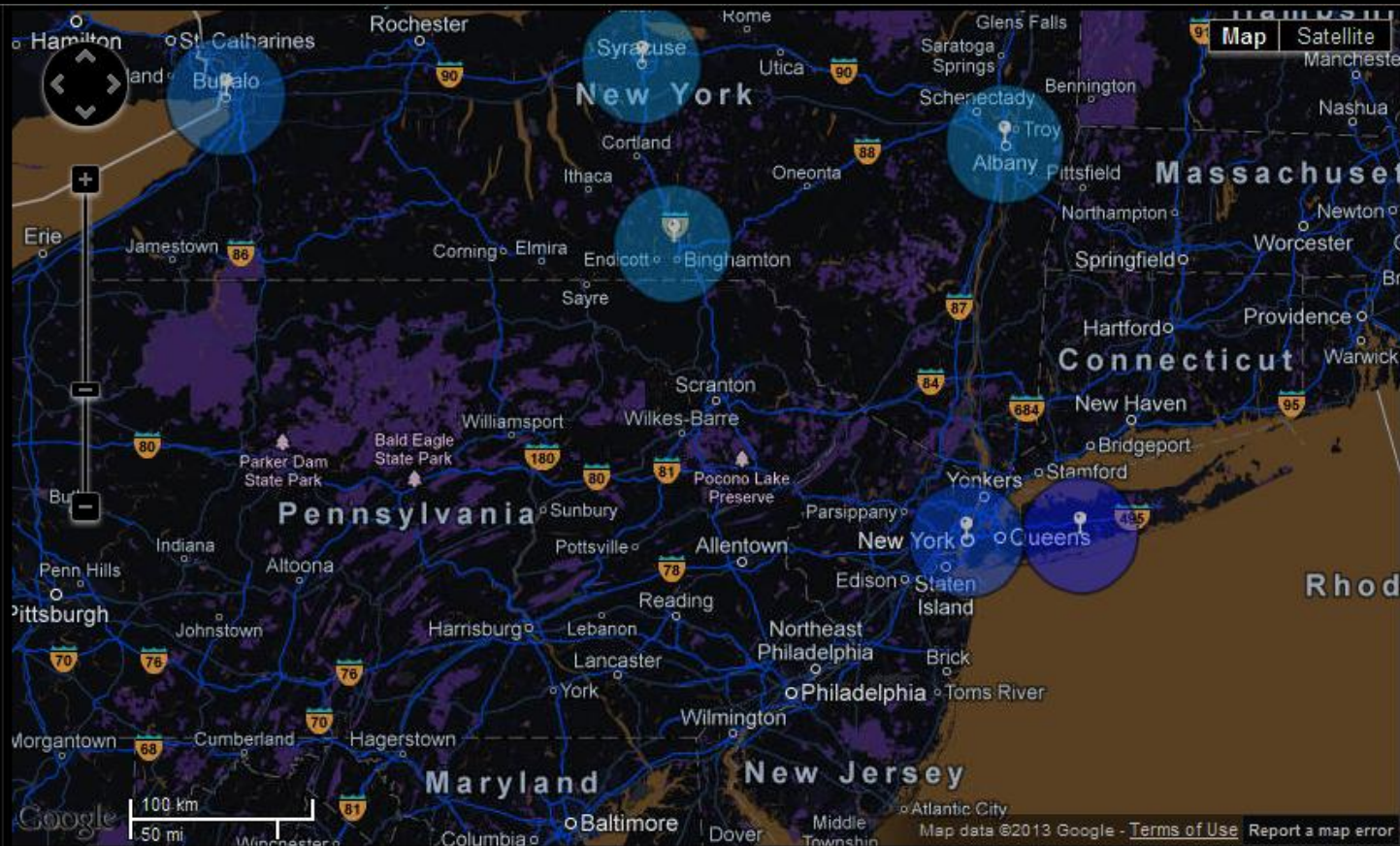
- ☒ Registered Nurses (53,112)
- ☐ Medical and Health Services Managers (4,475)
- ☐ Licensed Practical and Licensed Vocational Nurses (3,466)
- ☐ Nursing Aides, Orderlies, and Attendants (1,411)
- ☐ Home Health Aides (985)
- ☐ Physician Assistants (977)
- ☐ Medical Assistants (510)
- ☐ Dietitians and Nutritionists (417)
- ☐ Nursing Instructors and Teachers, Postsecondary (367)
- ☐ Nurse Practitioners (366)
- ☒ Advanced Practice Psychiatric Nurses (90)
- ☒ Acute Care Nurses (18)
- ☒ Critical Care Nurses (14)

## Company

- ☒ All Companies
- ☐ North Shore- Long Island Jewish Health System (5,194)
- ☐ Lenox Hill Hospital (1,277)
- ☐ Newyork- Presbyterian Hospital (1,060)
- ☐ Visiting Nurse Service Of New York (628)
- ☐ Onward Healthcare (578)
- ☐ Catholic Health System (356)
- ☐ Us Department Of Veterans Affairs (348)
- ☐ Bayada Nurses, Inc (311)
- ☐ Health Quest Management Corporation Vii (296)
- ☐ Lourdes Hospital (294)

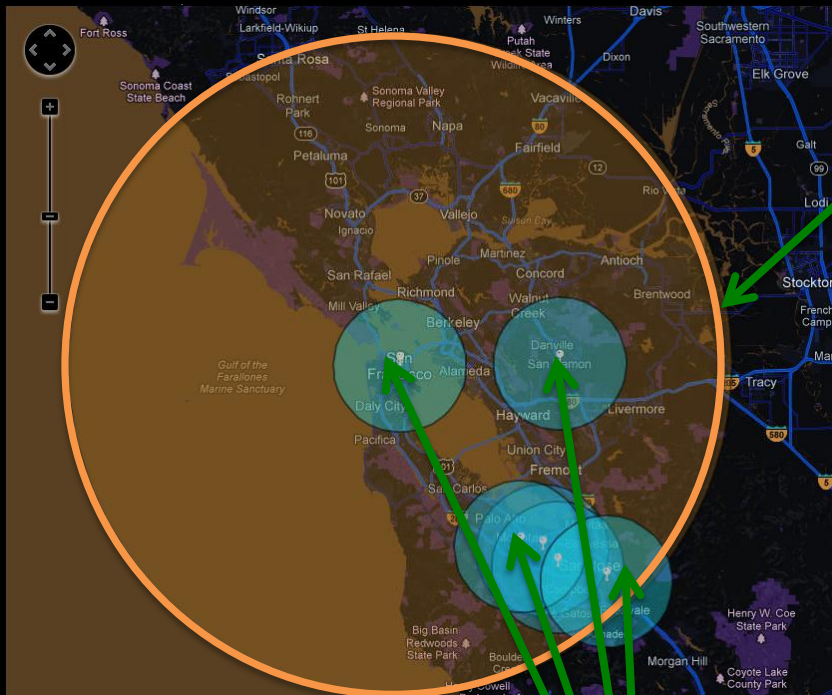
## Industry (NAICS)

## Staffing



Indicator	City	State	Active Supply	Demand	Compensation	Apps/Job
47	Bay Shore	NY	70	613	55k - 91k	Below Average
31	New York	NY	366	22,167	54k - 89k	Average
22	Buffalo	NY	43	4,537	47k - 67k	Weak
19	Binghamton	NY	37	1,431	42k - 65k	Below Average
17	Syracuse	NY	59	3,642	46k - 65k	Weak
16	Albany	NY	63	5,599	42k - 69k	Weak

# Geo-spatial Analytics



## Query 1:

/solr/select/?...

fq={!geofilt sfield=latlong pt=37.777,-122.420 d=80}  
&facet=true&facet.field=city&

"facet\_fields":{  
 "city":[  
 "san francisco, ca",11713,  
 "san jose, ca",3071,  
 "oakland, ca",1482,  
 "palo alto, ca",1318,  
 "santa clara, ca",1212,  
 "mountain view, ca",1045,  
 "sunnyvale, ca",1004,  
 "fremont, ca",726,  
 "redwood city, ca",633,  
 "berkeley, ca",599]}

## Query 2:

/solr/select/?...

&facet=true&facet.field=city&

fq=( \_query\_:"{!geofilt sfield=latlong pt=37.7770,-122.4200 d=20}" //san francisco  
OR \_query\_:"{!geofilt sfield=latlong pt=37.338,-121.886 d=20}" //san jose  
...  
OR \_query\_:"{!geofilt sfield=latlong pt=37.870,-122.271 d=20}" //berkeley  
)



# Customer-specific Analytics



Member Search

Dashboard

Recruitment Tools

## Member Search

- ☒ Talent Networks
- ☒ Employee Networks

Keywords, job titles, and/or categories

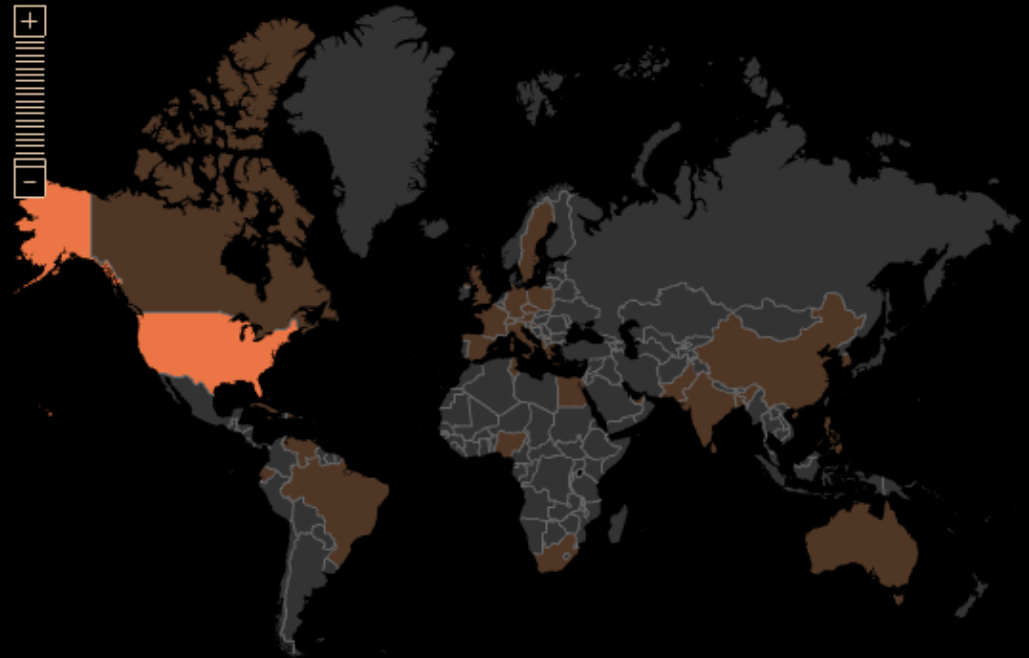
labor and delivery nurse or mechanical engineer

Location

ex: Chicago, IL or 60601

Search

You have 8,114 members in your Talent Network



# Customer-specific Analytics



Networks:



Member Search

Dashboard

Recruitment Tools

Summary

Traffic

Activity

Members

Jobs

Source

Lead Export

The "Summary Tab" provides an overview of the growth of your Talent Network, a snapshot of key member information and a comparison of the supply of talent versus your demand. For a quick video tutorial, please click [here](#).

Total Members

7,786

Members in Last 60 Days

496

See Who's Recently Joined

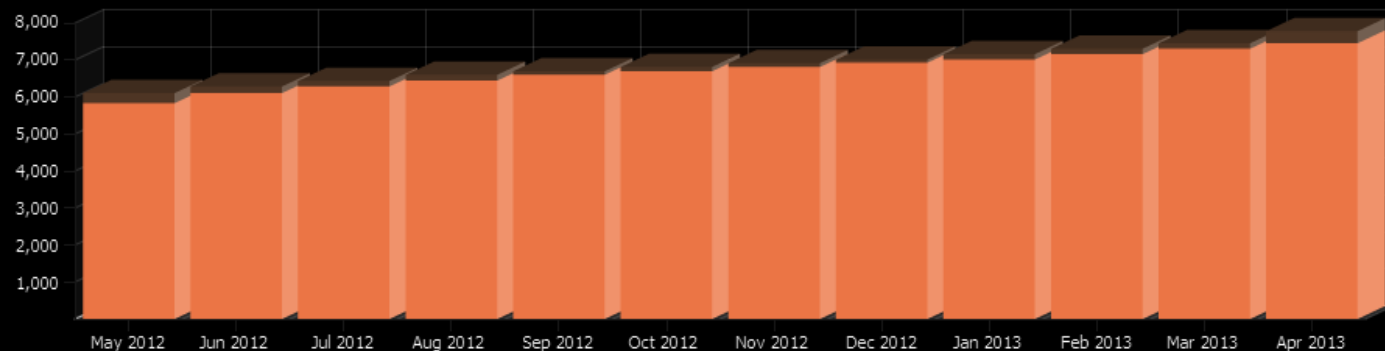
**John Doe**

Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products



## Member Growth

Previous Total New Members



Export

Source: CareerBuilder

# Customer-specific Analytics



Networks:

TN Company XYZ  
TALENT NETWORK

Member Search

Dashboard

Recruitment Tools

Summary

Traffic

Activity

Members

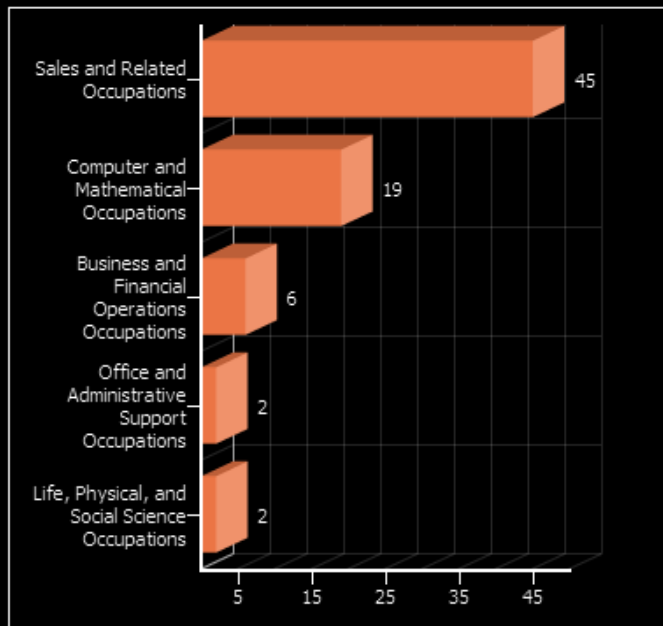
Jobs

Source

Lead Export

The "Jobs Tab" allows you to analyze the types and locations of jobs in your Talent Network. For a quick video tutorial, please click [here](#).

## Jobs By Category



Jobs by Category	Total Active Jobs
Sales and Related Occupations	45
Computer and Mathematical Occupations	19
Business and Financial Operations Occupations	6
Office and Administrative Support Occupations	2
Life, Physical, and Social Science Occupations	2
Arts, Design, Entertainment, Sports, and Media Occupations	1

Export

Source: CareerBuilder

# Customer's Performance vs. Competition

## PROFILE SNAPSHOT

### Education

	YOU	COMPETITORS
High School	16%	15%
Bachelor	59%	41%
Master	12%	32%

...

### Years Of Experience

	YOU	COMPETITORS
3-5	17%	19%
6-10	37%	30%
11-15	21%	17%

...

### Ethnicity

	YOU	COMPETITORS
African-American	44%	39%
Caucasian	22%	31%
Asian	19%	15%

...

## POSTING PERFORMANCE SNAPSHOT

You Are Averaging

17

Applications Per Posting for  
Petroleum Engineers

Underperforming  
Companies



Top Performing  
Companies

[View Job Titles for Current Job Group](#)

### Choose Your Filters: ?

Petroleum Engineers ?

Enter a location ?

Date Applied: 03/25/2013 To 04/25/2013

Submit

Reset



[View Detailed Report and Additional Filters](#)

## JOB GROUPS

Looking Good!

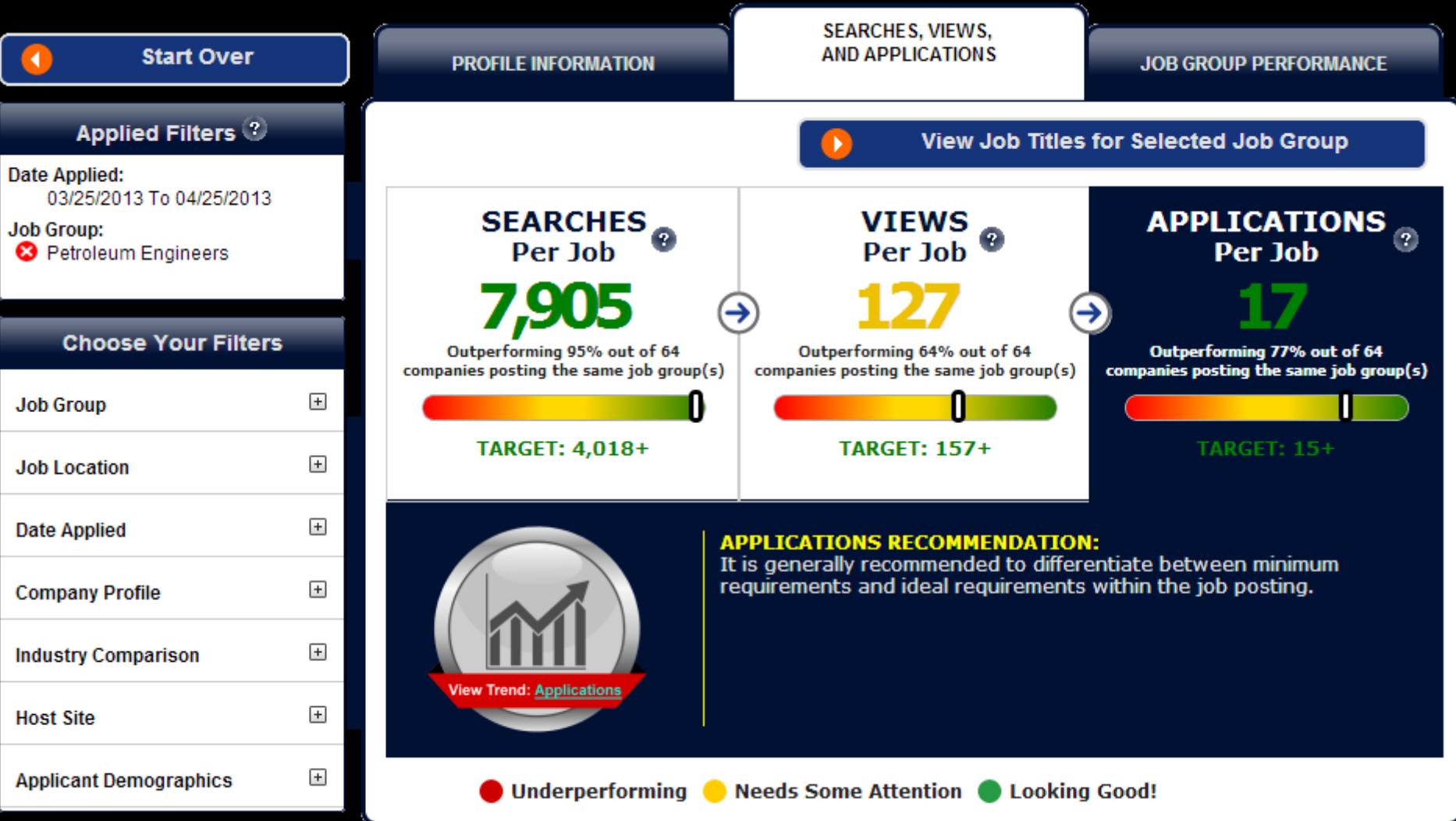
16



Job Group	Number Of Postings
Telecommunications Specialists	8
Petroleum Engineers	7
Law Clerks	3
Actuaries	2
Industrial Engineering Technicians	2
Producers	2

[▶ BACK TO ALL](#)

# Customer's Performance vs. Competition



# Customer's Performance vs. Competition

PROFILE INFORMATION

SEARCHES, VIEWS,  
AND APPLICATIONS

JOB GROUP PERFORMANCE

Below are how each of your job groups perform compared to other companies posting in the same job group. The number beside each job group is the number of jobs associated with it. Click on a job group to view a more detailed posting performance report. **Click on the number of job postings next to a job group to view the job titles included in the job group.**

UNDERPERFORMING

129



- Accountants(1067 )
- Bookkeeping, Accounting, and Auditing Clerks(863 )
- Financial Managers, Branch or Department(463 )
- Web Developers(194 )
- Network and Computer Systems Administrators(154 )

NEEDS SOME ATTENTION

87



- Medical Records and Health Information Technicians(8 )
- Art Directors(8 )
- Copy Writers(7 )
- Network Systems and Data Communications Analysts(5 )
- Advertising and Promotions Managers(4 )
- Operations Research

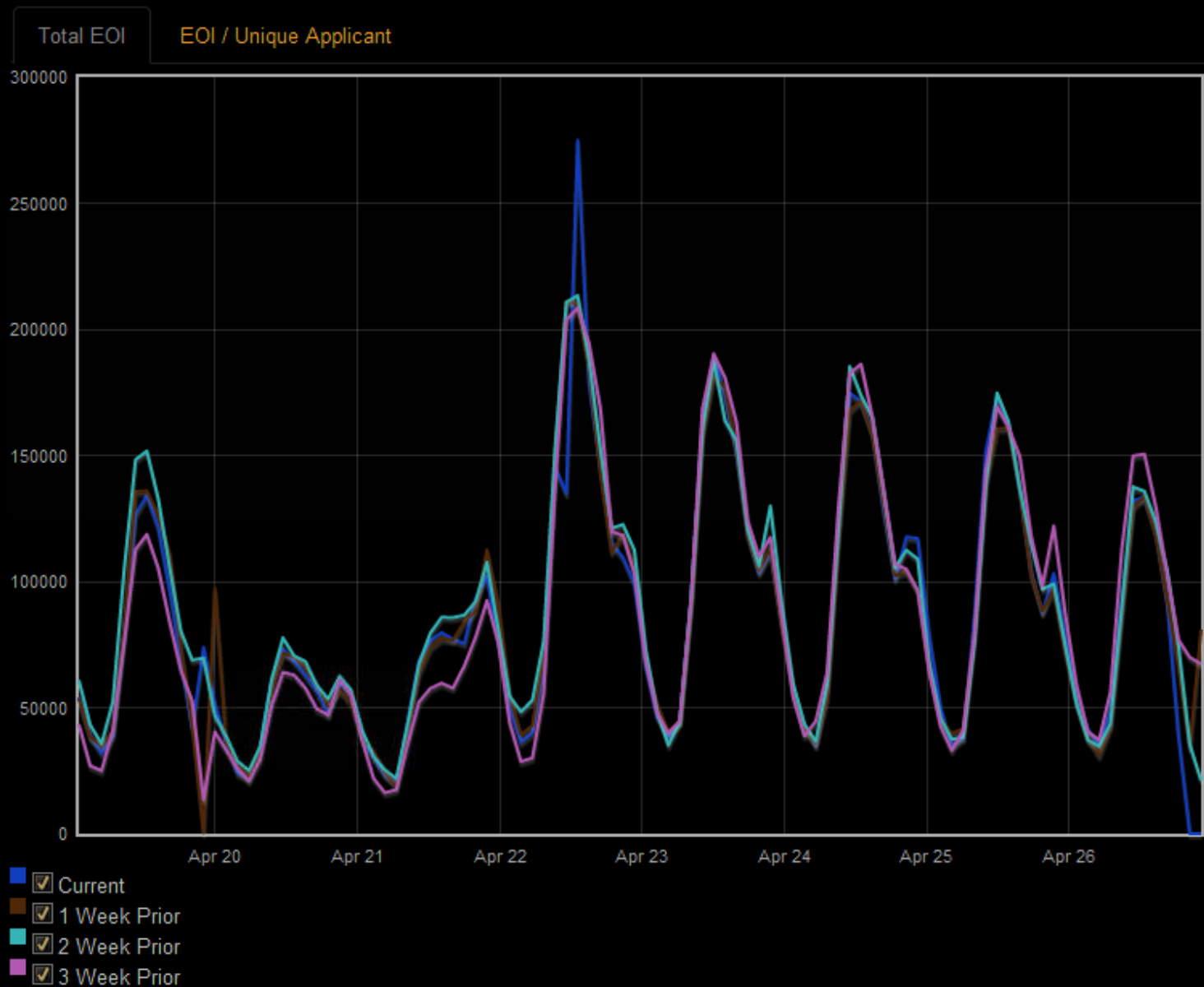
LOOKING GOOD!

16



- Electronics Engineering Technicians(1)
- Information and Record Clerks, All Other(1)
- Regulatory Affairs Specialists(1)
- Stock Clerks, Sales Floor(1)
- Document Management Specialists(1)
- Structural Metal Fabricators and Fitters(1)

# Event Stream Analysis



# A/B Testing

## Date Range:

START:

04/01/2013

END:

04/04/2013

## Filters:

HOSTSITES

US

IPATHS

SITEIDS

JOBS

## Data to Plot:

EXPERIMENT

Search

DocToDocExperimentType

US

2013-02-25 23:58 TO 2013-04-04 09:01

☒ EOI Per Experiment Group

OTHER

☐ EOI Per Week

☐ EOI Per Hostsite

☐ EOI Per IPath

☐ EOI Per SiteId

☐ EOI Per Job



# A/B Testing

Date Range:

START:

04/01/2013

END:

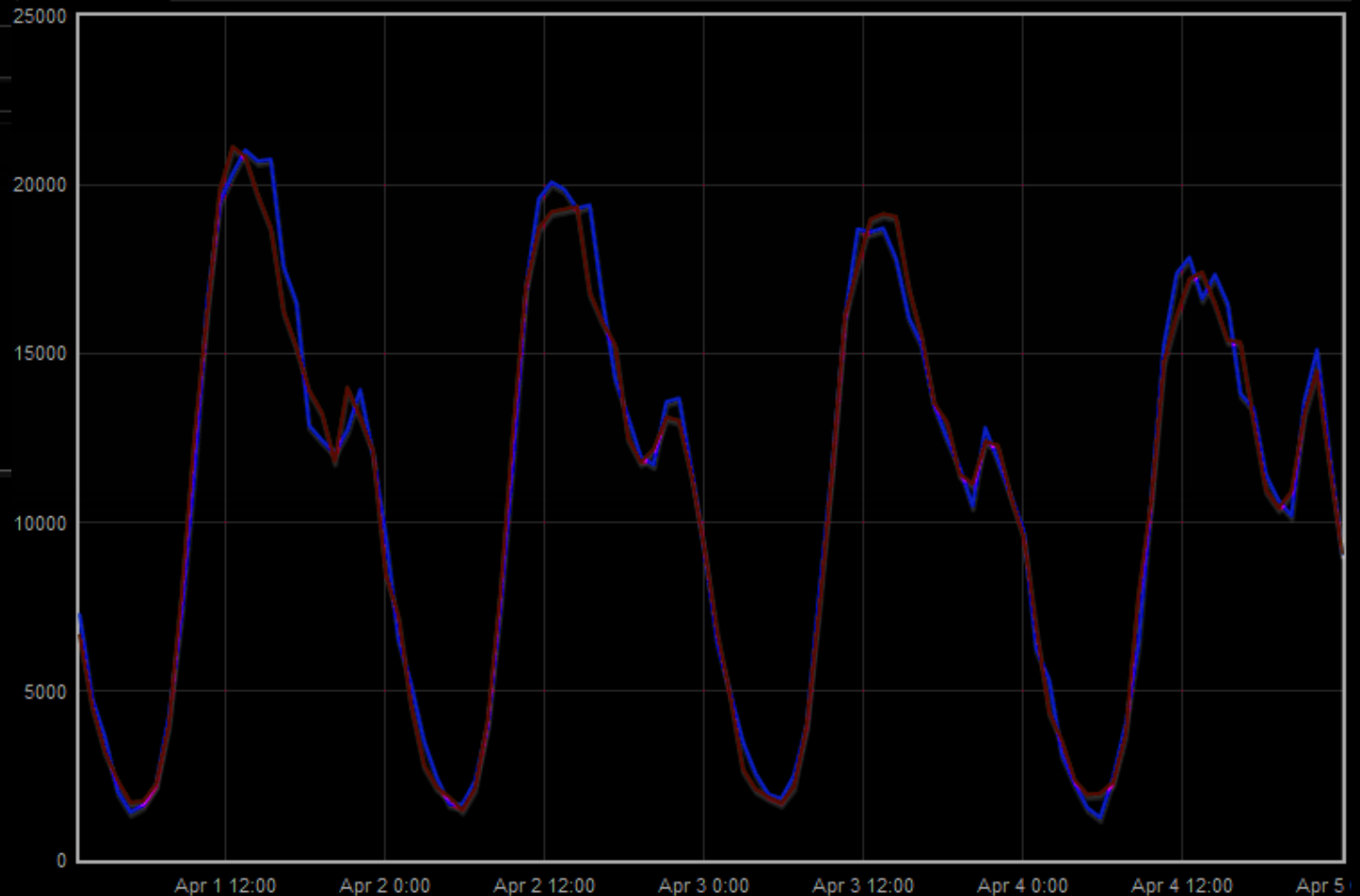
04/04/2013

Filters:

Total EOI

EOI / Unique Applicant

Data to Plot:



- ☒ Group 0: MultipleONetsNoTfIdf (25% traffic)
- ☒ Group 1: MultipleONetsWithTfIdf (25% traffic)
- ☐ Group 2: Off (25% traffic)
- ☐ Group 3: TfIdfOnly (25% traffic)

# A/B Testing

1. Hash all users based upon user identifier in application stack
2. Flow user event streams into Solr with user identifier
3. Implement custom hashing function in Solr
4. Facet based upon custom function
5. Profit!

# A/B Testing – Adding Function to Solr

## //Custom Solr Plugin

```
public class ExperimentGroupFunctionParser extends ValueSourceParser {  
    public ValueSource parse(FunctionQParser fqp) throws SyntaxError{  
        String experimentName = fqp.parseArg();  
        ValueSource uniqueID = fqp.parseValueSource();  
        int numberOfGroups = fqp.parseInt();  
        double ratioPerGroup = fqp.parseDouble();  
        return new ExperimentGroupFunction (experimentName, uniqueID, numberOfGroups, ratioPerGroup);  
    }  
}  
  
public class ExperimentGroupFunction extends ValueSource {  
    ...  
    @Override  
    public FunctionValues getValues(Map context, AtomicReaderContext readerContext) throws IOException {  
        final FunctionValues vals = uniqueIdFieldValueSource.getValues(context, readerContext);  
        return new IntDocValues(this) {  
            @Override  
            public int intVal(int doc) {  
                //returns an deterministically hashed integer indicating which test group a user is in  
                return GetExperimentGroup(experimentName, vals.strVal(doc), numGroups, ratioPerGroup);  
            }  
        }  
    }  
    ...  
}
```

## //SolrConfig.xml:

```
<valueSourceParser name="experimentgroup"  
    class="com.careerbuilder.solr.functions. ExperimentGroupFunctionParser" />
```

# A/B Testing – Faceting on Function

//Group 1 Time Series Facet Query:

/solr/select?

facet.range=eventDate&

facet.range.start=2013-04-01T01:00:00.000Z&

facet.range.end=2013-04-07T01:00:00.000Z&

facet.range.gap=+1Hour&

fq={!frange l=1 u=1 key="Group 1"}experimentgroup("EXPERIMENT\_NAME", userIdField, 4, 0.25)

//repeat for groups 2-4

...fq={!frange l=2 u=2 key="Group 2"}experimentgroup("EXPERIMENT\_NAME", userIdField, 4, 0.25)

...fq={!frange l=3 u=3 key="Group 3"}experimentgroup("EXPERIMENT\_NAME", userIdField, 4,

0.25)...fq={!frange l=4 u=4 key="Group 4"}experimentgroup("EXPERIMENT\_NAME", userIdField, 4, 0.25)

# A/B Testing – Faceting on Function

//Group 1 Time Series

/solr/select?

facet.range=eventD

facet.range.start=2

facet.range.end=20

facet.range.gap=+1

fq={!frange l=1 u=1 l=

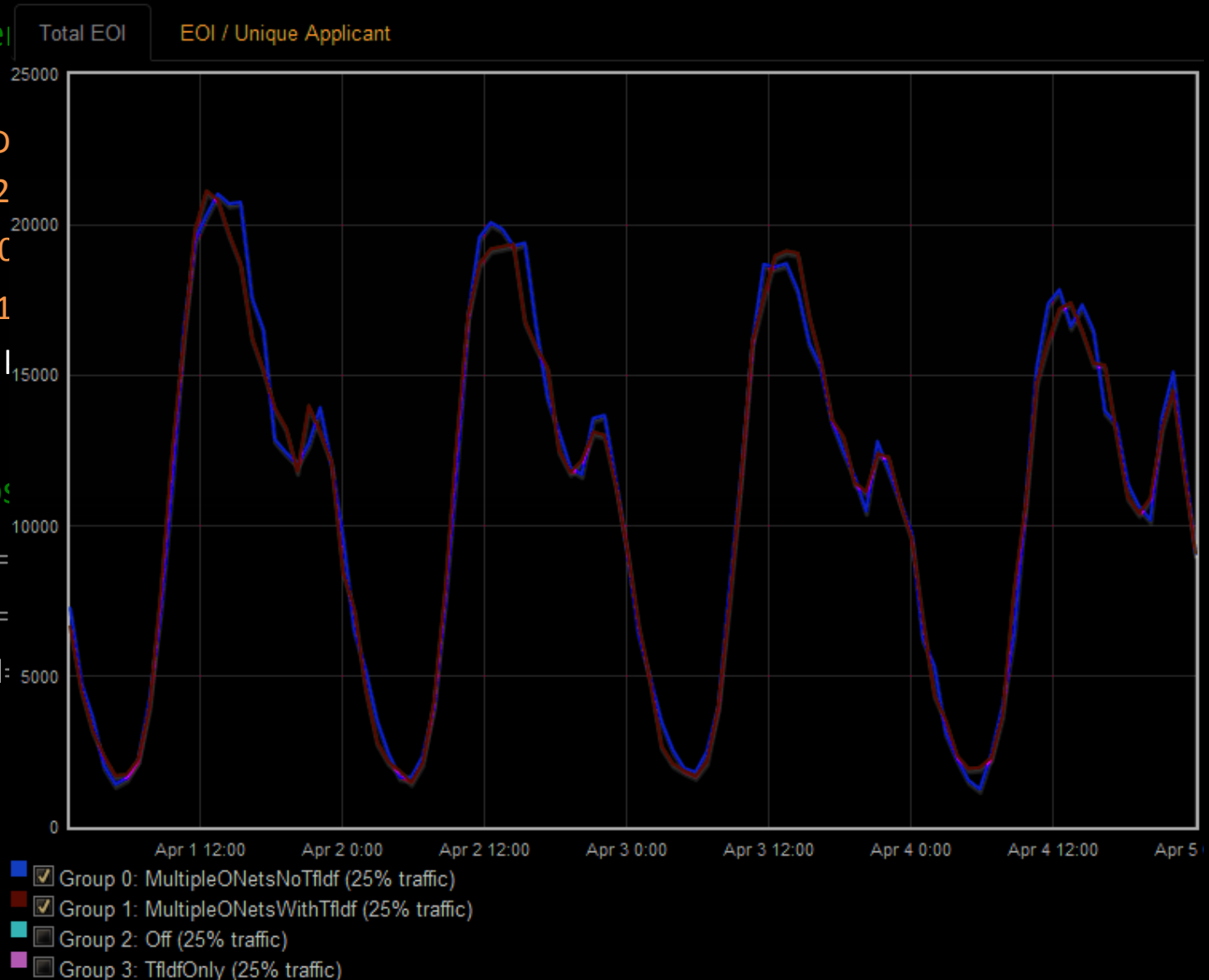
//repeat for group:

...fq={!frange l=2 u=

...fq={!frange l=3 u=

0.25)...fq={!frange l=

0.25)



Solr Patches in progress

# SOLR-2894: “Distributed Pivot Faceting”



Lead: Yonik Seeley | Category: Lucene | URL: <http://lucene.apache.org/solr/>

Summary

Issues

Road Map

Agile

Change Log

Popular Issues

Versions

Components

Labels

Source

## Popular Issues

Shows the unresolved issues sorted by number of votes

Votes	T	Key	Summary
37		SOLR-64	strict hierarchical facets
25		SOLR-2894	Implement distributed pivot faceting

#2 Most popular issue

**Status:** We have submitted a stable patch (including distributed refinement) which is working in production at [CareerBuilder](#).

Note: The community has reported Issues with non-text fields (i.e. dates) which need to be resolved before the patch is committed.

# SOLR-3583: “Stats within (pivot) facets”

Solr / SOLR-3583




## Percentiles for facets, pivot facets, and distributed pivot facets


Comment


Agile Board

More Actions ▾

### ▼ Details

Type:  Improvement

Status:  Open

Priority:  Minor

Resolution: Unresolved

Affects Version/s: None

Fix Version/s: 4.3

Component/s: None


Labels: newbie patch

### ▼ Description

Built on top of [SOLR-2894](#), this patch adds percentiles and averages to facets, pivot facets, and distributed pivot facets by making use of range facet internals.

### ▼ Attachments



 SOLR-3583.patch

81 kB

29/Mar/13 15:24

**Status:** We have submitted an early patch (built on top of SOLR-2894, distributed pivot facets), which is in production at [CareerBuilder](#).



# SOLR-3583: “Stats within (pivot) facets”

```
/solr/select?q=...&  
facet=true&  
facet.pivot=state,city&  
facet.stats.percentiles=true&  
facet.stats.percentiles.averages=true&  
facet.stats.percentiles.field=compensation&  
f.compensation.stats.percentiles.requested=10,25,50,75,90&  
f.compensation.stats.percentiles.lower.fence=1000&  
f.compensation.stats.percentiles.upper.fence=200000&  
f.compensation.stats.percentiles.gap=1000
```

```
"facet_pivot":{  
  "state,city":[{  
    "field":"state",  
    "value":"california",  
    "count":1872280,  
    "statistics":[  
      "compensation",[  
        "percentiles",[  
          "10.0","26000.0",  
          "25.0","31000.0",  
          "50.0","43000.0",  
          "75.0","66000.0",  
          "90.0","94000.0"],  
        "percentiles_average",52613.72,  
        "percentiles_count",1514592]],  
    "pivot":[{  
      "field":"city",  
      "value":"los angeles, ca",  
      "count":134851,  
      "statistics":{  
        "compensation":[  
          "percentiles",[  
            "10.0","26000.0",  
            "25.0","31000.0",  
            "50.0","45000.0",  
            "75.0","70000.0",  
            "90.0","95000.0"],  
          "percentiles_average",54122.45,  
          "percentiles_count",213481]]}  
    ]  
  ]  
}
```

# Real-world Use Case

Compensation by Years of Experience

\* Figures may not add up to total due to missing or unavailable values

Years of Experience	# Reporting	Avg.	Salary Distribution (percentile) ⓘ				
			10th	25th	50th	75th	90th
1 - 2 years	2,134	\$32,052	\$19,250	\$21,250	\$24,000	\$42,000	\$57,000
3 - 5 years	3,742	\$30,284	\$19,250	\$21,500	\$25,500	\$34,000	\$49,000
6 - 10 years	5,652	\$33,572	\$20,250	\$21,500	\$26,000	\$34,250	\$66,000
11 - 15 years	3,424	\$34,296	\$21,500	\$23,750	\$29,250	\$38,500	\$56,000
16 - 20 years	1,642	\$37,403	\$21,250	\$23,500	\$31,500	\$43,000	\$65,000
21+ years	2,040	\$41,869	\$21,250	\$25,750	\$33,500	\$52,000	\$74,500

Another Pivot...

Compensation

Field Facet

Stats Pivot Faceting (Average)

Stats Pivot Faceting (Percentiles)

Metro Area (MSA)	# Reporting	Avg. ▼	Salary Distribution (percentile) ⓘ				
			10th	25th	50th	75th	90th
Greater New York City Area	846	\$36,276	\$21,500	\$25,500	\$32,000	\$39,500	\$56,000
Indianapolis Indiana Area	444	\$36,068	\$20,250	\$21,250	\$26,500	\$40,000	\$80,000
Greater Philadelphia Area	801	\$33,836	\$21,250	\$23,500	\$28,000	\$40,000	\$58,000
Greater Chicago Area	1,140	\$33,635	\$20,750	\$22,000	\$27,000	\$38,000	\$55,500
Washington D.C. Metro Area	505	\$32,925	\$21,250	\$23,250	\$27,000	\$37,000	\$45,500
Greater Los Angeles Area	512	\$30,566	\$20,000	\$21,000	\$29,000	\$34,000	\$46,000
Houston Texas Area	448	\$30,042	\$19,500	\$21,750	\$26,250	\$35,000	\$45,000
Dallas/Fort Worth Area	435	\$29,805	\$19,250	\$21,250	\$25,500	\$32,000	\$48,000
Orlando Florida Area	471	\$28,838	\$21,250	\$23,500	\$23,750	\$30,000	\$43,000
Phoenix Arizona Area	696	\$27,144	\$20,000	\$21,000	\$25,000	\$30,000	\$40,000

1 - 2 y

3 - 5 y

6 - 10 y

11 - 15 y

16 - 20 y

21+ y

Data Analytics with Solr... the next frontier...

# Solr Analytics Wish List...

- Faceting Architecture Redesign:
  - “All” Facets should be pivot facets (default pivot depth = 1)
  - Each “Pivot” could be a field, query, or range
  - Meta information (like the Stats patch) could be nested in each pivot
  - Backwards compatibility with current facet response format for a while...
- Grouping
  - Grouping should also support multiple levels (Pivot Grouping)
  - If “Pivot Grouping” were supported... what about faceting within each of the pivot groups?
- Remaining caches should be converted to “per-segment” to enable NRT
- These kinds of changes would enable Solr to return rich Data Analytics calculations in a single search call where many calls are required today.

# Recap

Billions of documents

+

Ad-hoc querying by keyword

+

Horizontal Scalability

+

Sub-second query responses

+

Facet on “anything” – field, function, etc.

+

User friendly visualizations/UX

---

Building a real-time, Big Data Analytics Platform with Solr

# Contact Info

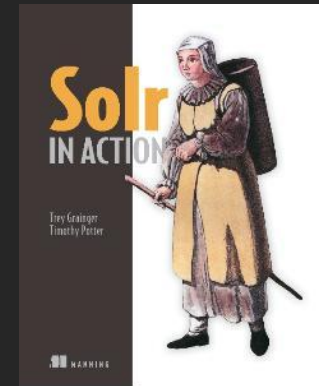
## ■ Trey Grainger



trey.grainger@careerbuilder.com  
@treygrainger

Other presentations:

<http://www.treygrainger.com>



<http://solrinaction.com>

Yes, **we are hiring** @CareerBuilder. Come talk with me if you are interested...