### **Source of the Data**

The engine behind LinkUp’s Job Search platform is driven by a very sophisticated and complex technology infrastructure that systematically “spiders” corporate web sites to gather information about open job postings.

While scraping an individual website once is not a technological feat today, doing it daily for greater than 32,300 companies is. LinkUp has spent years developing technology to deliver the most reliable, accurate and timely jobs data in the industry.

We add roughly 50,000 to 100,000 per day.We also capture all of the job description data for each job that provides it. Our Job Description database goes back to March 17, 2014, and has descriptions for over 21 million jobs. As of December 31, 2015, the LinkUp Raw trial data feed consists of approximately 70 million unique job records from roughly 32,300 companies spanning the time period from August 2,2007 through December 31, 2015.

### Schema of the Data

#### jobs

Contains the raw data. Each row contains the job id, company id and when the posting was added and taken down.

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **Description** |
| job\_id | STRING | id of the job |
| title | STRING | job title |
| company\_id | STRING | company id |
| company\_name | STRING | company name |
| city | STRING | city of the job |
| state | STRING | state of the job |
| zip | STRING | zip of the job |
| country | STRING | country of the job |
| category\_id | STRING | job category id |
| category\_name | STRING | job category name |
| created | TIMESTAMP | date+time job was first added |
| created\_date | STRING | date job was first added |
| created\_day | INTEGER | day job was added |
| created\_month | INTEGER | month job was added |
| created\_year | INTEGER | year job was added |
| last\_checked | TIMESTAMP | date+time job was last checked |
| last\_checked\_date | STRING | date job was last checked |
| last\_checked\_day | INTEGER | day job was last checked |
| last\_checked\_month | INTEGER | month job was last checked |
| last\_checked\_year | INTEGER | year job was last checked |

#### company\_daily

Daily aggregation of the raw jobs data on a company level. Each row contains the number of jobs added/removed for each company on the given day.

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **Description** |
| company\_id | INTEGER | id of the company |
| date | TIMESTAMP | date the analysis refers to |
| jobs\_active | INTEGER | number of active jobs |
| net\_jobs\_active\_added | INTEGER | the net number of active jobs |
| jobs\_created | INTEGER | the number of job postings created |
| jobs\_removed | INTEGER | the number of job postings created |

#### categories

A list of job categories found in the dataset.

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **Description** |
| id | STRING | job category id |
| description | STRING | job category name |

### 

### 

### companies

A list of companies comprising the dataset.

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **Description** |
| id | STRING | company id |
| name | STRING | company name |
| earliest\_scrape\_date | TIMESTAMP | earliest job scrape date |
| last\_scrape\_date | TIMESTAMP | last job scrape date |
| city | STRING | most common posted city in the job postings |
| state | STRING | most common posted state in the job postings |
| country | STRING | most common country in the job postings |
| official\_name | STRING | official name of company |
| ticker\_main | STRING | the associated base ticker/symbol |
| ticker\_ric\_exchg | STRING | RIC exchange code |
| ticker\_ric | STRING | base RIC exchange code |
| bloomberg\_id | STRING | unique global Bloomberg entity identifier |
| company\_scraping\_url | STRING | the unique base URL LinkUp uses to spider/scrape |
| trc\_legal\_entity\_identifier | STRING | Thomson Reuters legal identifier |
| trc\_business\_sector | STRING | Thomson Business Sector classification |
| trc\_economic\_sector | STRING | Thomson Economic Sector classification |
| trc\_industry\_sector | STRING | Thomson Industry Sector classification |
| trc\_industry\_group | STRING | Thomson Industry Group classification |

#### Descriptions

Description of the scraped jobs

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **Description** |
| id | STRING | id of the job |
| description | STRING | description of the job |