# Title: Fuzzy de-duplication of leads and contact - May 2019

## 

|  |
| --- |
| **Aspirations - Q2 2019 - mid quarter update** Freddy-Freshsales, Dedupe, 21/05/2019 |
|  |

### Feature

Build scalable framework which can identify duplicates in spite of spelling mistakes, abbreviations, fields mismatch, phonetic similarity, etc.

### 

|  |  |  |
| --- | --- | --- |
| **Squad/Team-name-here** | **Name** | **Git home (if applicable)** |
| Freddy | Bharathi | <https://github.com/freshdesk/fs_dedupe> |
| Freddy | Srivatsa |  |
| Freddy | Suvrat |  |
| Freshsales | Ajeet |  |
| Freshsales - QA | Logesh |  |
| Freshsales | Sivalingam |  |
| Freshsales - Devops | Madhav |  |
| Freshsales - Devops | Sivakumar |  |
| Freddy | Swaminathan |  |
| Freshsales | Sudharshan |  |
| Freshsales - Product | Aditi |  |
| Freshsales - Product | Srivatsan |  |

### **Timeline**

|  |  |
| --- | --- |
| 1st to 23rd April 2019 | Deployment on docker/staging and testing |
| 23rd to 30 April 2019 | Production setup, offline scripts |
| 2nd May, 2019 | System went live for US accounts and is running stably. |
| 11th May 2019 - Ongoing | Live on Freshdesk. Metric calculation/Monitoring stats. |
| 27th May - 31st May | Go live on 100+ accounts. |

## **Success metrics - KPIs (Key Performance Indicators)**

| # | Goal area (direction/theme) | Individual goals (list of features) | Metrics |
| --- | --- | --- | --- |
| 1 | Lead organisation | * (Q1) To de-duplicate leads using fuzzy matching techniques | * Merge number is oscillating week on week. * Standout is (number of merges / number duplicates checks)   + 7x increase in ease of finding duplicates |

### **New capabilities**

| Capability | ES based | Freddy based |
| --- | --- | --- |
| Phonetic match | No | Yes |
| Field mismatch | No | Yes |
| Abbreviation | No | Yes |
| Spelling mistake | No | Yes |
| Domain match | No | Yes |
| Smart ranking | No | Yes |
| Pure string match on name | Yes | No |

## **Previous Aspirations Link:**

[Fuzzy de-duplication of leads and contacts](https://confluence.freshworks.com/pages/viewpage.action?pageId=210018712)

## **Quarterly Goals & Progress:**

|  |  |  |
| --- | --- | --- |
| Goals | Status | Comments |
| Fuzzy de-duplication | Beta released | Feature is released for internal accounts. We are monitoring and calculating impact. Unit test coverage ~80% |
| Extend for foreign language | In Progress |  |
| Fuzzy de-duplication of accounts | Not started |  |

## **Learnings (Innovation/R&D/Architecture/Design/...)**

(Please list items that you/your team's accomplishments that they are proud of and share it Engineering Org wide)

|  |  |  |  |
| --- | --- | --- | --- |
| Item | Description | Impact | Comments |
| Patent | We are working with legal team to apply for patent on deduplication system |  |  |
| New team | Development team was new and was also working first time with spark, kafka, graph. Also, we were located in remote location. Among 2 developers and a data scientist we churned 70k+ lines of code and several model iteration in 1 quarter. | Given the team in Bangalore now  understands different components (baikal, central, devops), we will be able to increase our development speed.  If same project is to be done again, it would take 50%-75% of time. |  |
| Graph database to store duplicates | We used AWS Neptune to store duplicates. This helped in more efficient data manipulation. | We can extend it to build knowledge graph of leads in Freshsales |  |

## **Do you have any specific asks? (any help needed to address constraints, unblocking, ...)**

**(Please keep this list engineering related - this section is optional)**

|  |  |  |
| --- | --- | --- |
| Ask | Impact | Recurring |
| Communication - conference calls | Conference calls are difficult to follow on zoom. There should be better mic/system. | Yes |

## **Appreciation**

(Please list appreciations for your team/squad by you/team members/others)

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Appreciated by | Impact | In their own words |
| Bharathi | Suvrat | Faster project delivery | During deployment of project Bharathi spent 15+ hours each day for a week, ensuring project is deployed on time. Also, she is aggressive and independent in resolving issues. |
| Ajeet | Suvrat | Responsive to queries | Ajeet is very responsive to queries and easy to reach out on chat. |

## **Demo/Deep dives**

* Specific QA test cases can be found [here](https://docs.google.com/spreadsheets/d/1ebsaVc2Xf-AL5cYVbx4Q891ChlOmkXJmaGjxNAkM9HI/edit?usp=sharing).
* Online and offline metric can be found [here](https://docs.google.com/spreadsheets/d/1dxdYQk1EiMQy1gnYLQ5FbnZ6T7EqY_ksIvsmyzPzIko/edit?usp=sharing).