

# Job Market Analysis Platform

Baochen Hu, Ning Zhang

# Schedule

- Crawl Job Info from Website (Craiglist, Glassdoor)
- Ingest the data using any pipeline(kafka)
- distribute to backend
- Hbase for report
- Spring Boot Web app to demo
- Spark( Real Time Analysis)

# Technology Stack

- Hbase
- Kafka
- Spark-Streaming
- Spark-ML
- Scrapy Crawler

# Questions

- What is highest salary
- Which company has the best review
- Which company is the best match

# Expectation: Report

## CORE INSIGHT ENTERPRISE

## SECURITY TESTING EXECUTIVE REPORT

SUMMARY OF VERIFIED EXPOSURES FOUND FROM 1/1/2012 TO 2/22/2013

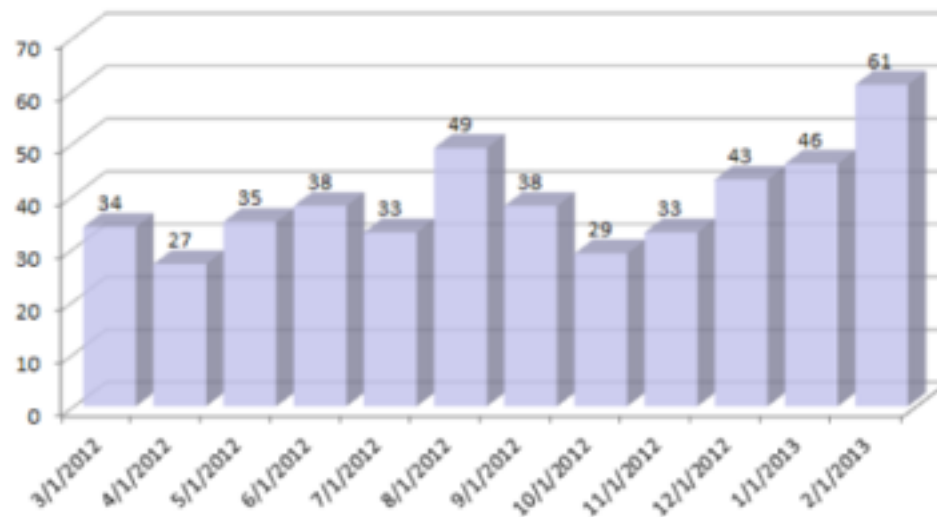
### TESTING SUMMARY

Test Objective	RISK	RISK TREND	RISK CHANGE
Customer Password Compromise	✓ Low		↑ Up
Credit Card Risk	✓ Low		↓ Down
Low value data	▲ Medium		↑ Up
Intellectual Property Risk	✓ Low		↑ Up
SSN Compromise	✓ Low		— No Change
Website Defacement	✗ High		↑ Up

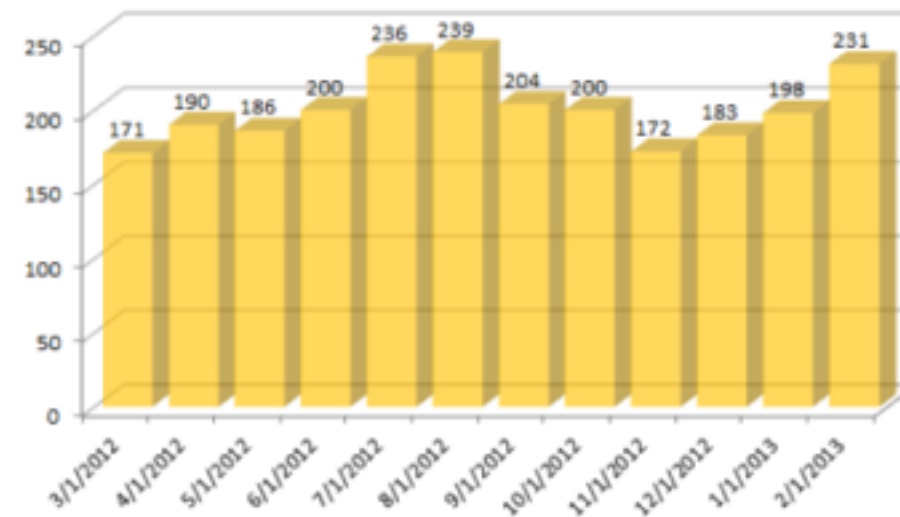
### EXPOSURE DISTRIBUTION



### COMPROMISED ASSETS



### EXPOSURE OVER TIME



# Recommendation

email@example.com

## 1 Heading on level 1 (section)

---

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies

### 1.1 Heading on level 2 (subsection)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem.

#### 1.1.1 Heading on level 3 (subsubsection)

Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate