Jose Chinchilla
MCITP/MCSE: Database Administrator, SQL Server
MCITP/MCSE: Business Intelligence SQL Server

Current Positions:
President, Agile Bay, Inc.
President, Tampa Bay Business Intelligence User Group
Regional Mentor, PASS LATAM

Blog: http://www.sqljoe.com
Twitter: @sqljoe
Linked-in: http://www.linkedin.com/in/josechinchilla
Email: jchinchilla@agilebay.com

Customers & Partners
Agenda

1. Top 5 Client Use Cases of Big Data
2. Overview of Big Data Ecosystem
3. Overview of Sentiment Analysis
4. Demo: Sentiment Analysis using Twitter
Top 5 Client Use Cases of Big Data
Top 5 Client Use Cases

1) Modern Data Warehouse
   • Enterprise Data Warehouse Hadoop integration
   • Long term data staging and archiving

2) Sentiment Analysis
   • Opinion Mining
   • Twitter, Facebook, Google+, Yelp, UrbanSpoon, TripAdvisor

3) Market Basket Analysis
   • Product Affinity Analysis
   • Recommendation Engine

4) Clickstream Analysis
   • Website visitor behavior
   • Click patterns

5) Risk Analysis
   • Consumer behavior
   • Fraud detection
Big Data Ecosystem
Big Data Ecosystem

Apache Hadoop
• Open-source
• http://hadoop.apache.org

Commercial Distributions
• Microsoft
• Horton Works
• Cloudera
• Amazon Web Services
• Greenplum
• Talend
• MapR
• Intel
• IBM
Big Data Ecosystem

The Zoo

- Oozie
- Flume
- Sqoop
- Hive
- Pig
- Falcon

- Mahout
- Impala
- Cheetah
- Giraph
- Stinger
- Phoenix
Big Data Ecosystem

The Zoo

- Oozie
- Flume
- Sqoop
- Hive
- Pig
- Falcon

- Mahout
- Impala
- Cheetah
- Giraph
- Stinger
- Phoenix
Sentiment Analysis Overview
“Everything we hear is an opinion, not a fact. Everything we see is a perspective, not the truth.”

—Marcus Aurelius
What is Sentiment Analysis

The #TMNT movie was great! Highly recommend.

Watching the #TMNT movie with @SQLJoe

That #TMNT movie was a waste of time. #fail
Sentiment Analysis 101

- Feelings
- Emotions
- Attitudes
- Opinions
- Judgments
- Orientation
- Polarity
Sentiment Analysis 101

• What are my customers saying about my products and services?
• Are customers talking positively or negatively about my products and services?
• What other brands are people talking about positively or negatively?
• Who is influencing the public opinion or perception about my products and services?
Sentiment Analysis 101

- Opinion mining
- Emotion analysis
- Opinion extraction
- Sentiment detection
- Sentiment categorization

- Sentiment classification
- Sentiment polarity
- Judgment analysis
- Subjectivity analysis
Sentiment Analysis 101: Process
Sentiment Analysis 101

What about?

• LOL
• OMG
• #FAIL
• #AWESOMESAUCE
• :(  
• :)
Sentiment Analysis:
Hadoop and Twitter data #TMNT
Twitter Demo: Steps

1) **You** will tweet using #tmnt
2) Extract tweets containing #tmnt hashtag via Flume job
3) Stage tweets in text files in the TMNT folder in HDFS (1 file each 90 secs or every 1000 tweets)
4) Load tweets into Hcatalog (cloudera JSON SerDe)
5) Break down tweets into sentences
6) Break down tweets into words
7) Lookup each word in lexical dictionary to get polarity value (1,0,-1)
8) Add polarity value for each word and get overall tweet polarity
   Positive = > 0, Neutral = 0, Negative < 0
The movie was great! Highly recommend #TMNT
# Twitter Demo: Example

<table>
<thead>
<tr>
<th>TweetID</th>
<th>LineNum</th>
<th>WordNum</th>
<th>Word</th>
<th>Word</th>
<th>Polarity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>The</td>
<td>good</td>
<td>positive</td>
<td>+1</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>2</td>
<td>movie</td>
<td>awesome</td>
<td>positive</td>
<td>+1</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>3</td>
<td>was</td>
<td>best</td>
<td>positive</td>
<td>+1</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>4</td>
<td>great</td>
<td>great</td>
<td>positive</td>
<td>+1</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>1</td>
<td>Highly</td>
<td>nice</td>
<td>positive</td>
<td>+1</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>2</td>
<td>recommend</td>
<td>highly</td>
<td>positive</td>
<td>+1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>excellent</td>
<td>positive</td>
<td>+1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>awful</td>
<td>negative</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>bad</td>
<td>negative</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>worse</td>
<td>negative</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>recommend</td>
<td>neutral</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>the</td>
<td>neutral</td>
<td>0</td>
</tr>
</tbody>
</table>

...
Twitter Demo: Example

The movie was great!

\[0 + 0 + 0 + 1 = +1\]

Highly recommend!

\[1 + 0 = +1\]

\[\text{= +2 (Positive)}\]
DEMO:
Sentiment Analysis using Flume and Hive
Resources
Resources

• Analyzing Twitter Data Using CDH
  https://github.com/cloudera/cdh-twitter-example

• Anatomy of a tweet

• Hortonworks Sandbox & Tutorials
  http://hortonworks.com/sandbox/
  http://hortonworks.com/tutorials/

• Flume Twitter API Configuration

• WordNet Lexical database (Princeton University)
  http://wordnet.princeton.edu/
Analyzing Twitter data

Twitter REST API v1.1

- API Resources
  https://dev.twitter.com/docs/api/1.1

- Field documentation:
  https://dev.twitter.com/docs/platform-objects/tweets

- Search (https://api.twitter.com/1.1/search/tweets.json)
  - GET search/tweets
    https://dev.twitter.com/docs/api/1.1/get/search/tweets

- Using the Twitter Search API
  https://dev.twitter.com/docs/using-search
Analyzing Twitter data

Twitter Streaming APIs

• Public stream
  https://dev.twitter.com/docs/streaming-apis/streams/public

• User stream
  https://dev.twitter.com/docs/streaming-apis/streams/user

• Site stream
  https://dev.twitter.com/docs/streaming-apis/streams/site
Analyzing Twitter data

Twitter Streaming APIs

- Public stream
  [https://dev.twitter.com/docs/streaming-apis/streams/public](https://dev.twitter.com/docs/streaming-apis/streams/public)

- Endpoints
  - POST statuses/filter
  - GET statuses/sample
  - GET statuses/firehose
Thank You!