MemoryKloud.

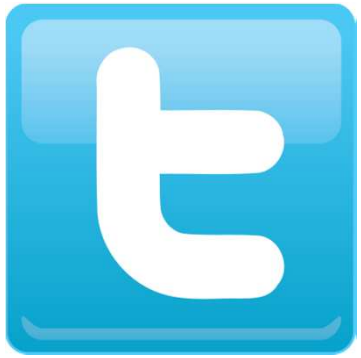
Presented By: Rami Taibah & Jie Wu
Advisor: Ray Larson

http://memorykloud.com/ischool | @memorykloud

When It all Began!

When It all Began!

What is the PROBLEM?



Precious memories are hard to track and ultimately lost.

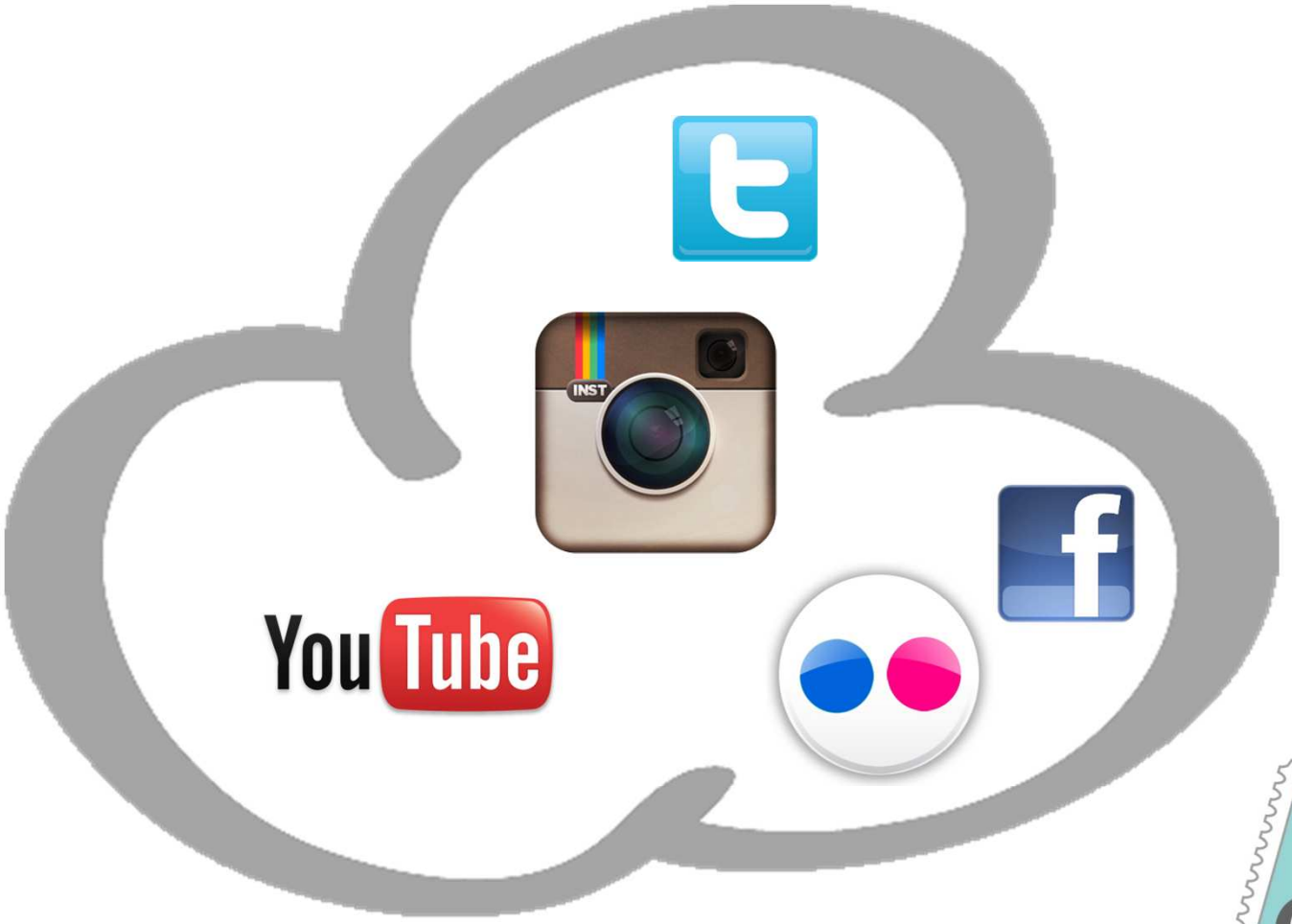
SO

What can We

DO

?

CREATE a Digital ScrapBook



Option #1: Mobile Solution

Option #2: Analytics

HOW Did we Go About IT ?

Solutions Design

Android Mobile Application

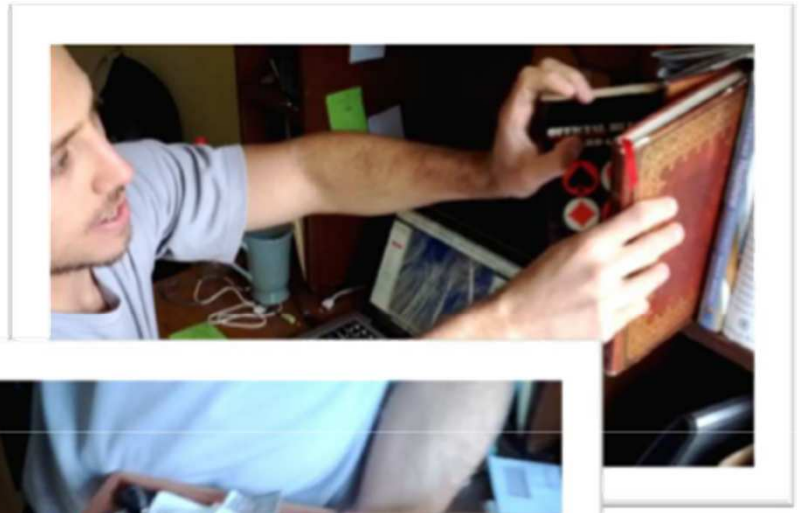
Web Interface

Data Analytics

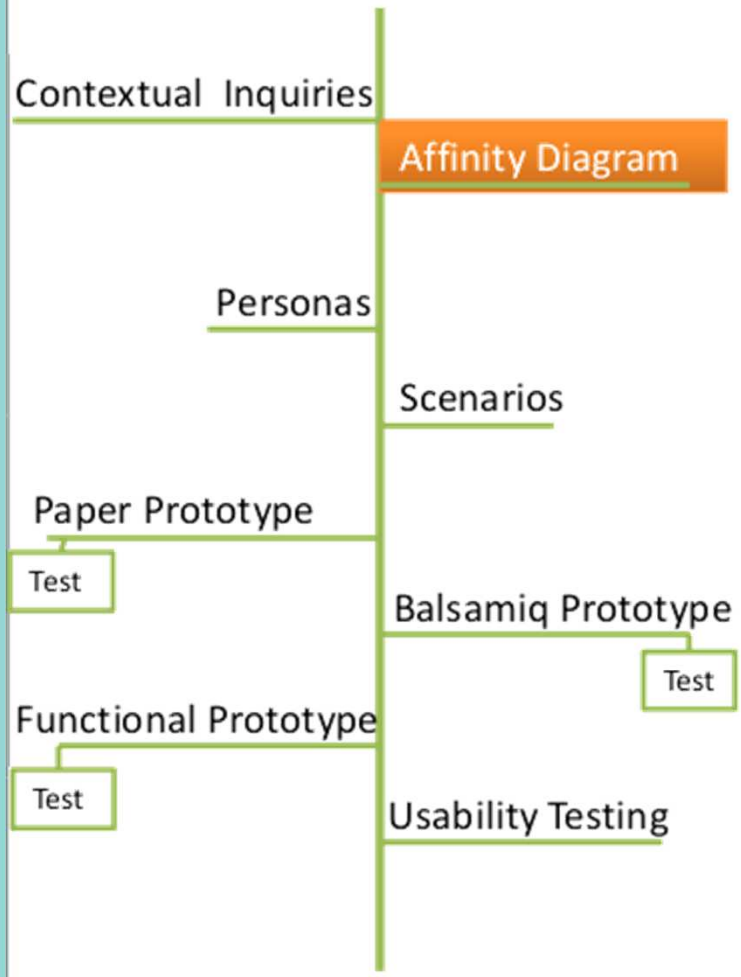
MOBILE
Application
DESIGN.

Contextual Interviews

- Contextual Inquiries
  - Affinity Diagram
  - Personas
  - Scenarios
  - Paper Prototype
    - Test
  - Balsamiq Prototype
    - Test
  - Functional Prototype
    - Test
  - Usability Testing



Affinity Diagram



Personas Design

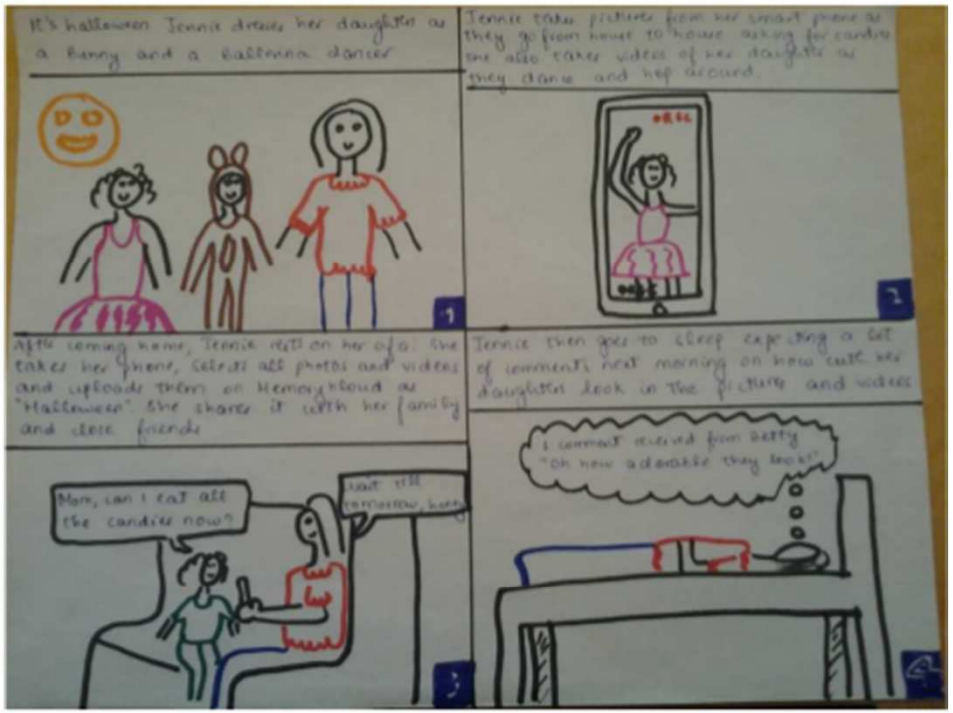
- Contextual Inquiries
- Affinity Diagram
- **Personas**
- Scenarios
- Paper Prototype
  - Test
- Balsamiq Prototype
  - Test
- Functional Prototype
  - Test
- Usability Testing

Brett

Jennie

Designing Scenarios

- Contextual Inquiries
- Affinity Diagram
- Personas
- **Scenarios**
- Paper Prototype
  - Test
- Balsamiq Prototype
  - Test
- Functional Prototype
  - Test
- Usability Testing



Paper Prototyping

- Contextual Inquiries
- Affinity Diagram
- Personas
- Scenarios
- Paper Prototype
  - Test
- Balsamiq Prototype
  - Test
- Functional Prototype
  - Test
- Usability Testing

Balsamiq Prototyping

Contextual Inquiries

Affinity Diagram

Personas

Scenarios

Paper Prototype

Test

Balsamiq Prototype

Test

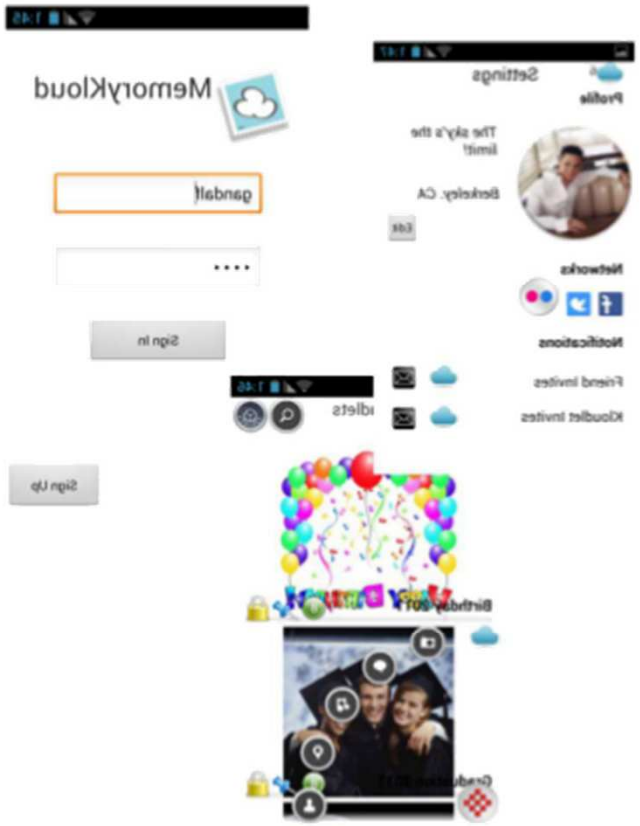
Functional Prototype

Test

Usability Testing

Functional Prototyping

- Contextual Inquiries
- Affinity Diagram
- Personas
- Scenarios
- Paper Prototype
  - Test
- Balsamiq Prototype
  - Test
- **Functional Prototype**
  - Test
- Usability Testing



Formative Evaluation

Contextual Inquiries

Affinity Diagram

Personas

Scenarios

Paper Prototype

Test

Balsamiq Prototype

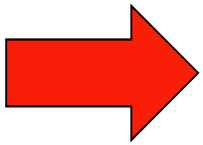
Test

Functional Prototype

Test

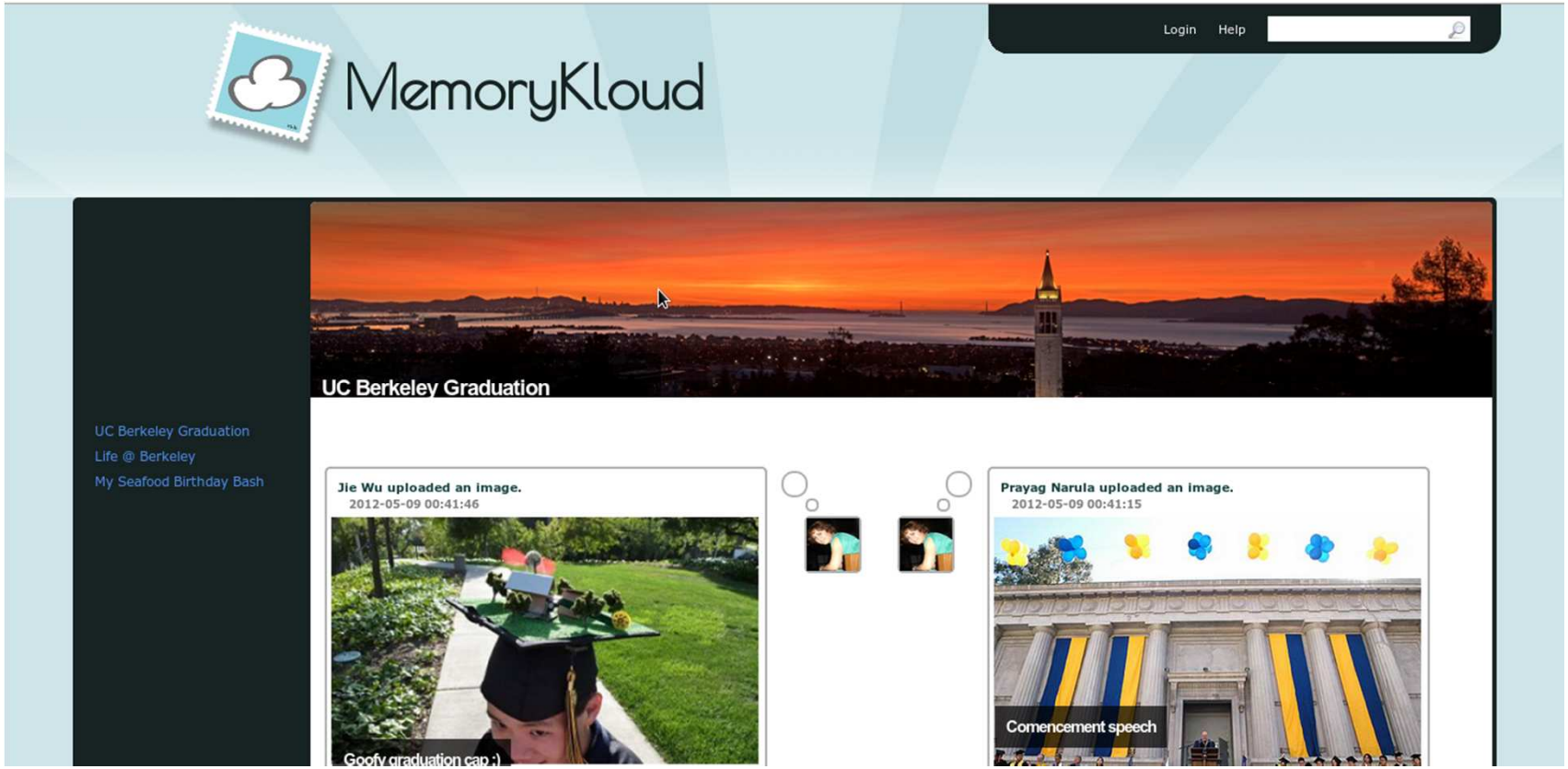
Usability Testing

Video Demo



The Web INTERFACE.

Web App


MemoryKloud
Login
Help
UC Berkeley Graduation
UC Berkeley Graduation
Life @ Berkeley
My Seafood Birthday Bash
Jie Wu uploaded an image.
2012-05-09 00:41:46
Goofy graduation cap :)
Prayag Narula uploaded an image.
2012-05-09 00:41:15
Comencement speech

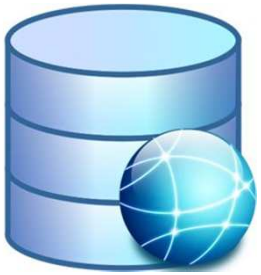
THE Technology Behind IT ?

Architecture

Mobile App

RESTFUL API
+
PYTHON

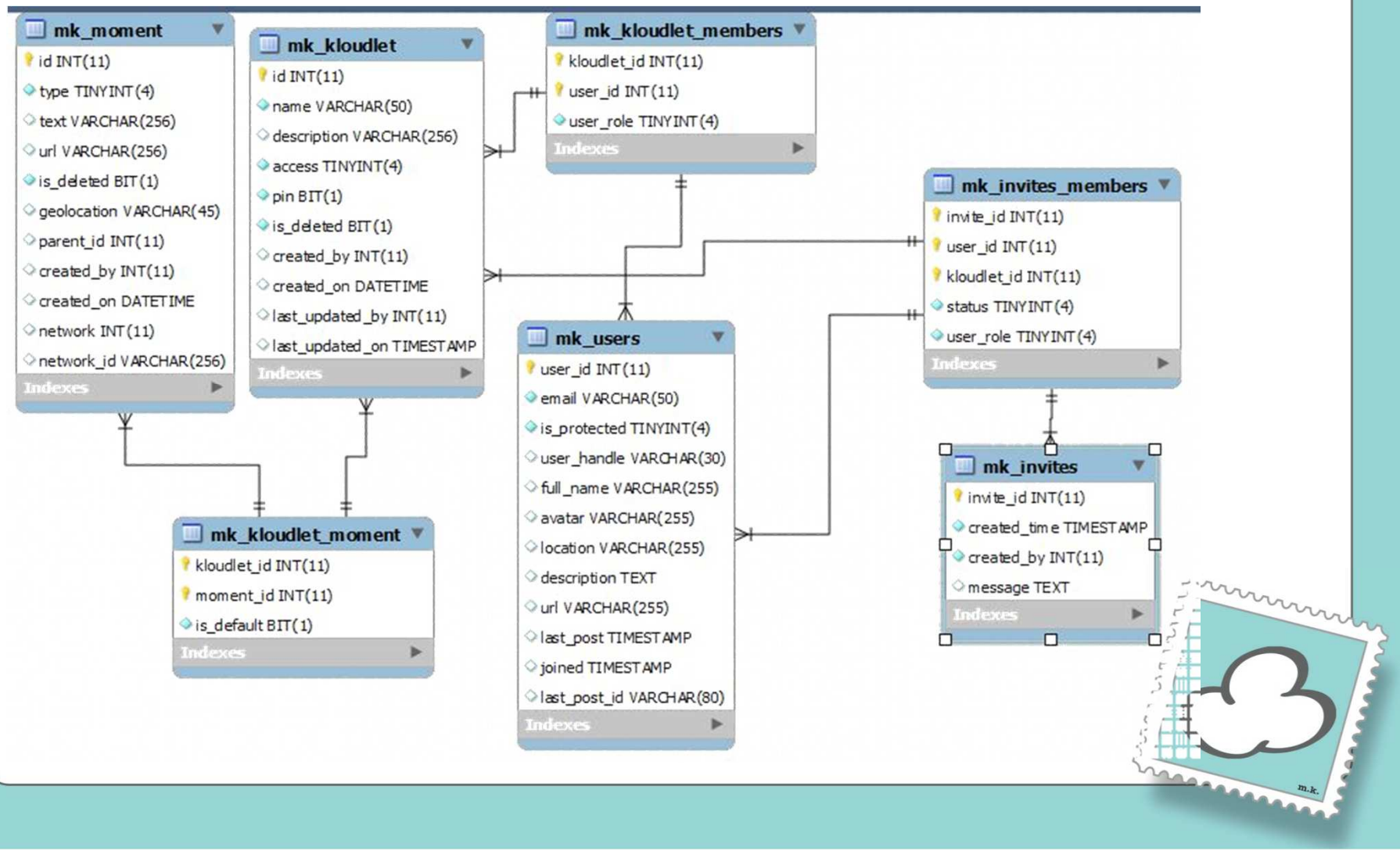
MySQL



Web App

Curation/Analytics

Database Schema

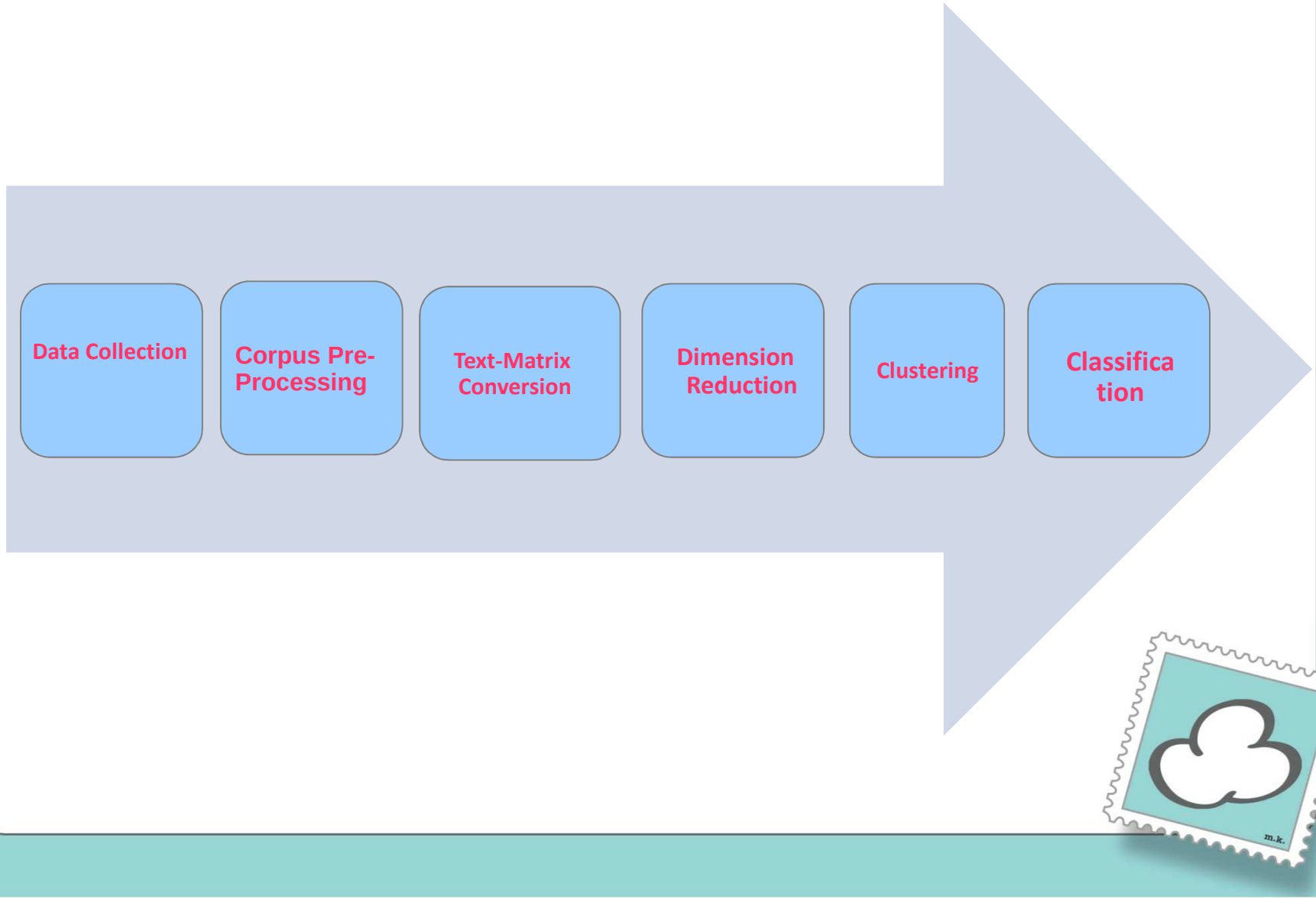


DATA
Analytics
SOLUTION.

Research Question

How can we group moments that are relevant together?

The Process


Data Collection
Corpus Pre-Processing
Text-Matrix Conversion
Dimension Reduction
Clustering
Classification

Data Collection

- Sample tweets data on a single day 01/23/2011
- Twitter indices published by National Institute of Standards and Technology (*NIST*)
- Run *Twitter-Corpus-Tools* on Amazon EMR to crawl Twitter.com for content
- 100 MB raw tweets data.

Corpus Preprocessing

- Remove *null* entries
- Remove *Non-English* Texts (any tweet that contains non-ASCII code
- Remove stop words
- Remove words with length less than 4
- If the remaining non-duplicated word count is great than 15, we use that for our analysis

Text-Matrix Conversion

Goal: Tweets -> a *document-term matrix*

Generate a *tf-idf*（term frequency–inverse document frequency）matrix X

$$\mathrm{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad \mathrm{idf}_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|}$$

$$\mathrm{tfidf}_{i,j} = \mathrm{tf}_{i,j} \times \mathrm{idf}_i$$

Expect 6432 * 2527 document-term matrix

Dimension Reduction

SVD **(Singular Value Decomposition):**

s <- svd(X, 10, 10)

Convert X (n * p) 2-D matrix into U*d*V'

1		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
2	1	0.001672	-0.01169	0.003933	-0.01196	0.003382	-0.00541	0.002831	-9.15E-05	9.47E-05	-0.00333
3	2	0.002904	-0.00696	0.002683	-0.00756	-8.11E-05	-0.00318	-0.00156	-0.00043	-0.00072	0.000186
4	3	0.006987	0.000373	0.00017	-0.00167	-0.01275	0.001236	0.00037	0.01149	-0.01099	-0.00034
5	4	0.002969	-0.01467	0.001472	-0.00816	0.001616	-0.00455	0.000618	-0.00039	-0.00033	-0.00091
6	5	0.000685	-0.00444	0.001308	-0.00425	0.000998	-0.00199	0.000903	6.23E-05	-1.26E-05	-0.00075
7	6	0.001903	-0.00103	0.000447	-0.00206	-0.00558	-0.00047	-0.00495	0.000598	0.000196	-0.00049
8	7	0.001525	-0.01021	0.003603	-0.01057	0.003459	0.007594	0.001069	-2.09E-07	-7.64E-05	-0.00175
9	8	0.001183	-0.00834	0.002711	-0.00867	0.002401	-0.00363	0.001937	-1.03E-05	-4.68E-05	-0.00158
10	9	0.019443	0.002066	1.13E-05	-0.00054	0.00098	-0.00261	-0.02868	0.005674	-0.0045	-0.00105
11	10	0.002484	-0.0009	0.000247	-0.00177	-0.00597	-0.00027	-0.00554	0.001043	0.0013	-0.00041
12	11	0.001491	-0.00273	0.000974	-0.00305	-0.00059	-0.00121	-0.00125	-0.00014	-0.00031	0.000411
13	12	0.011583	-0.00259	0.001365	-0.00522	-0.00863	-0.00202	-0.01462	0.007955	0.007939	0.001302
14	13	0.002429	-0.01901	0.00822	-0.00341	0.002338	-0.00488	0.002216	-3.58E-05	-7.68E-05	0.000581

Clustering

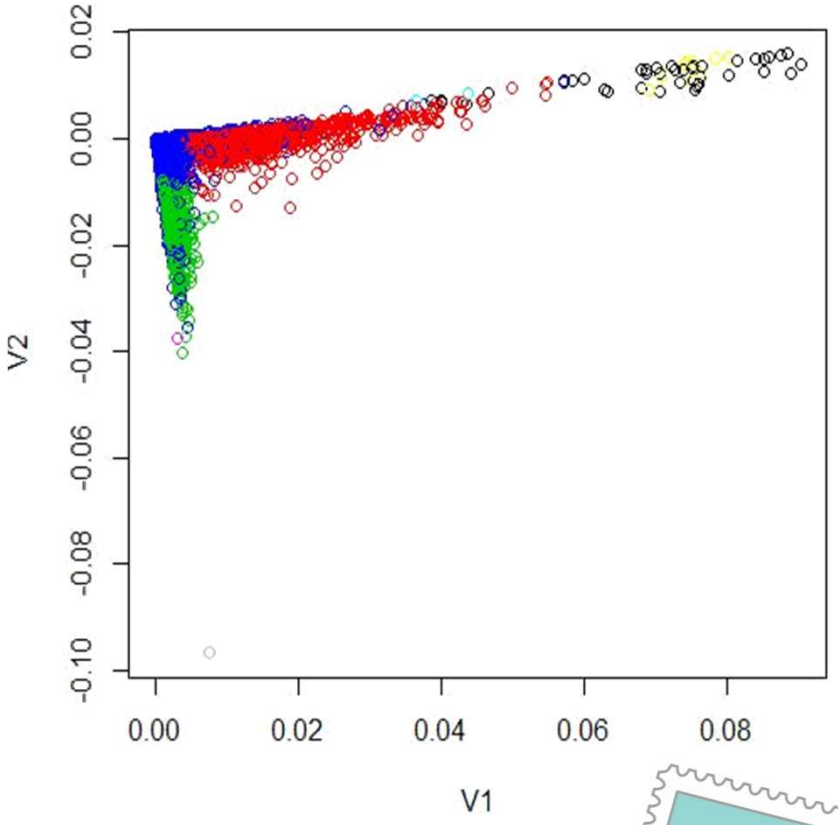
Partitioning Algorithms

K-means (k = 10)

Blue Cluster: **indonesia, ujian, ngiler, makanan, maen**

Green Cluster: **packers, match, season, Chicago, #packers**

Red Cluster: **bitch, fight, hospital, hurt, jail, army**



Classification

- Validation: check test data to see how good the clustering is.
- 10% testing, 90% training
- ***SVM*** (Support Vector Machine) algorithm is applied
- Confusion matrix after SVM

	true									
pred	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	0	0	0	1	0	0	0	0	0	0
3	0	2	255	12	0	0	0	0	0	0
4	0	0	10	209	0	0	0	0	0	0
5	0	0	2	12	0	0	0	0	0	0
6	0	0	0	8	1	1	0	0	0	3
7	0	1	0	7	0	0	5	0	0	4
8	0	0	0	2	0	0	0	4	2	17
9	0	0	0	0	0	0	0	0	4	20
10	0	0	0	1	0	0	0	0	0	43
11	0	0	0	0	0	0	0	0	0	13
12	0	0	0	0	0	0	0	0	0	3

MemoryKloud Conclusions

- Mobile
  - Enhancing the APIs
  - Authentication & Invitation
  - Android & Website UI
  - iOS app
- Data Analytics
  - Limitations
  - Alternative Algorithms

Thanks!

Question?