

Data Analytics as a Service for Data Scientists

Prof. Chen Li



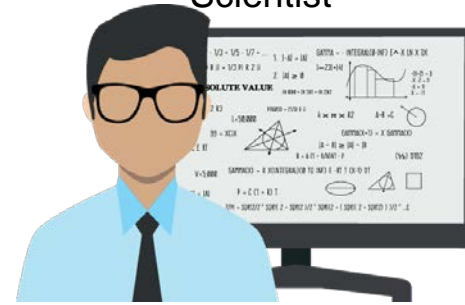
UCIRVINE

A real story

Sue: Public Health
researcher



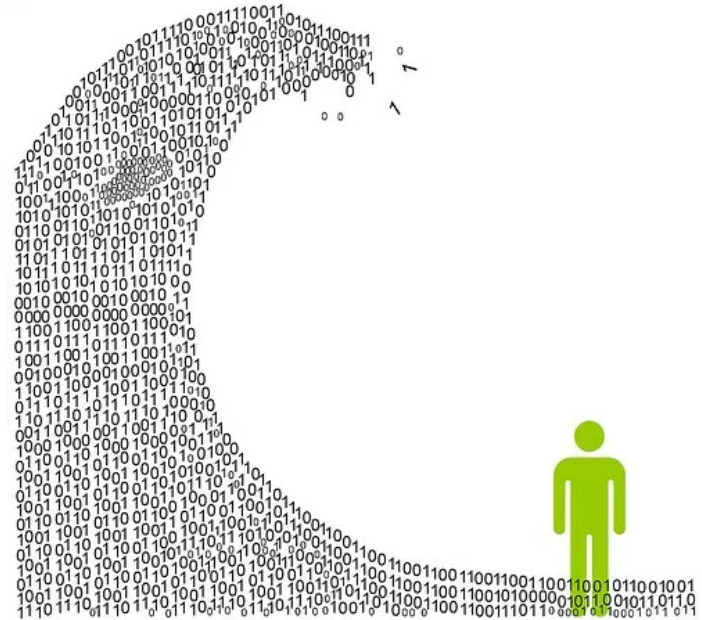
Adam: Data
Scientist



Challenges ...

- Infrastructure
- Data collection
- Large scale
- Machine learning

Not enough IT background!





Software solutions

Cloudberry:
Big data visualization



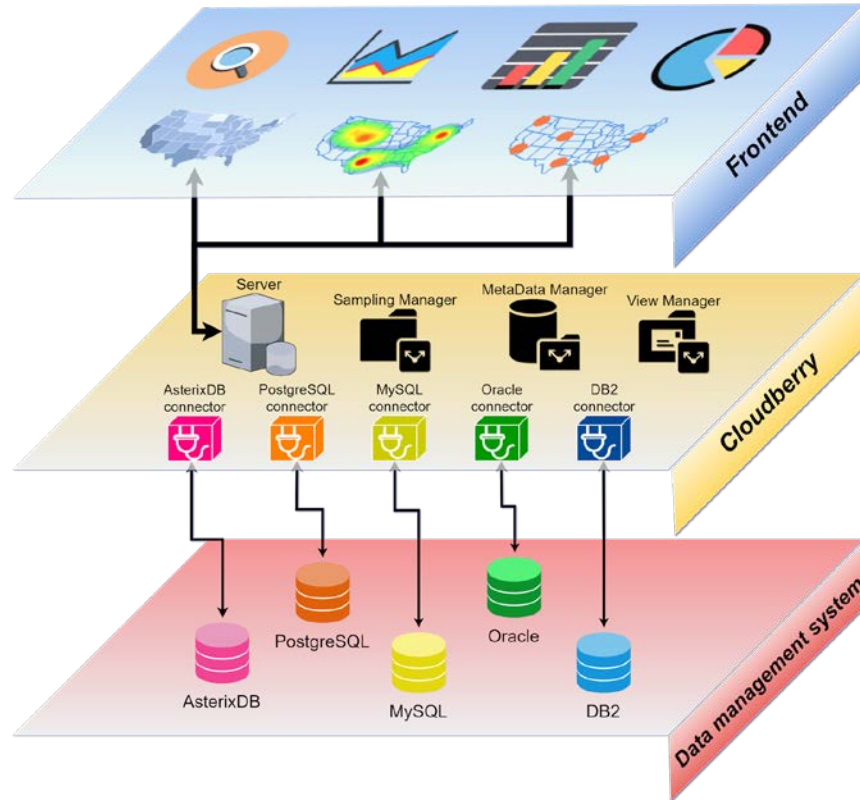
AsterixDB:
parallel database



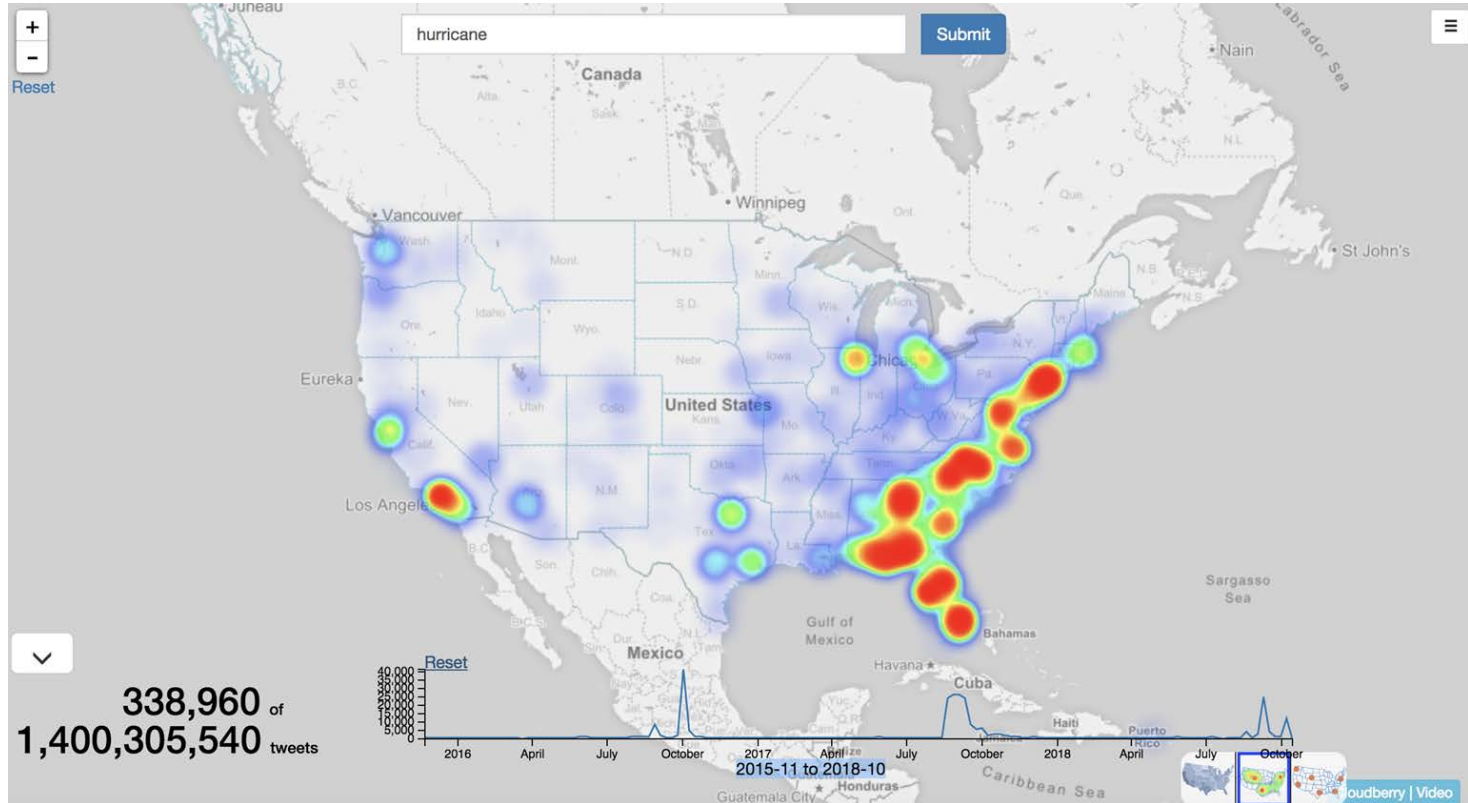
Texera:
Analytics using workflows

Users: researchers from UCI, UCLA

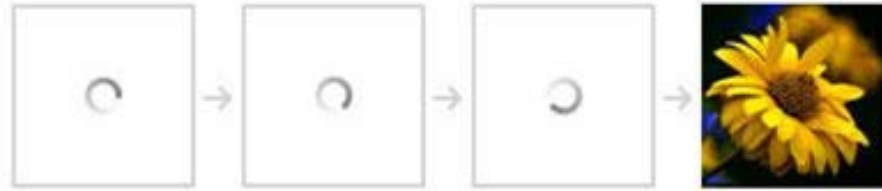
Cloudberry: Big Data Visualization



TwitterMap system

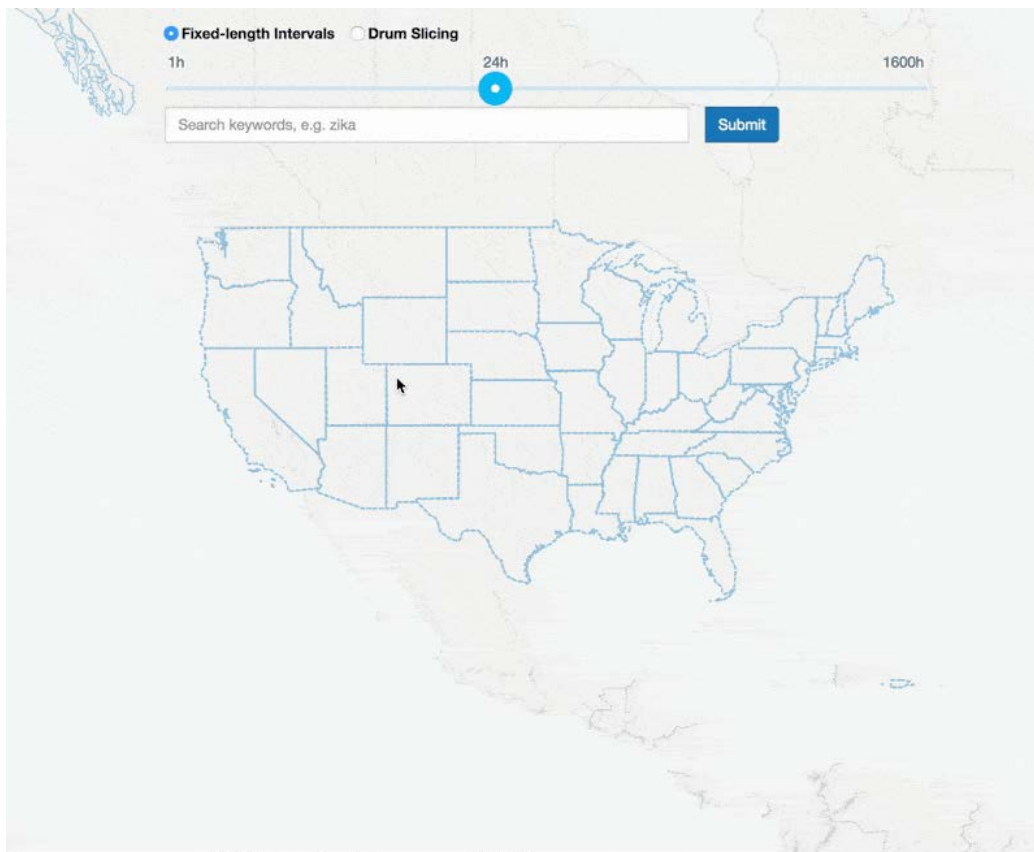


Takes too much time?

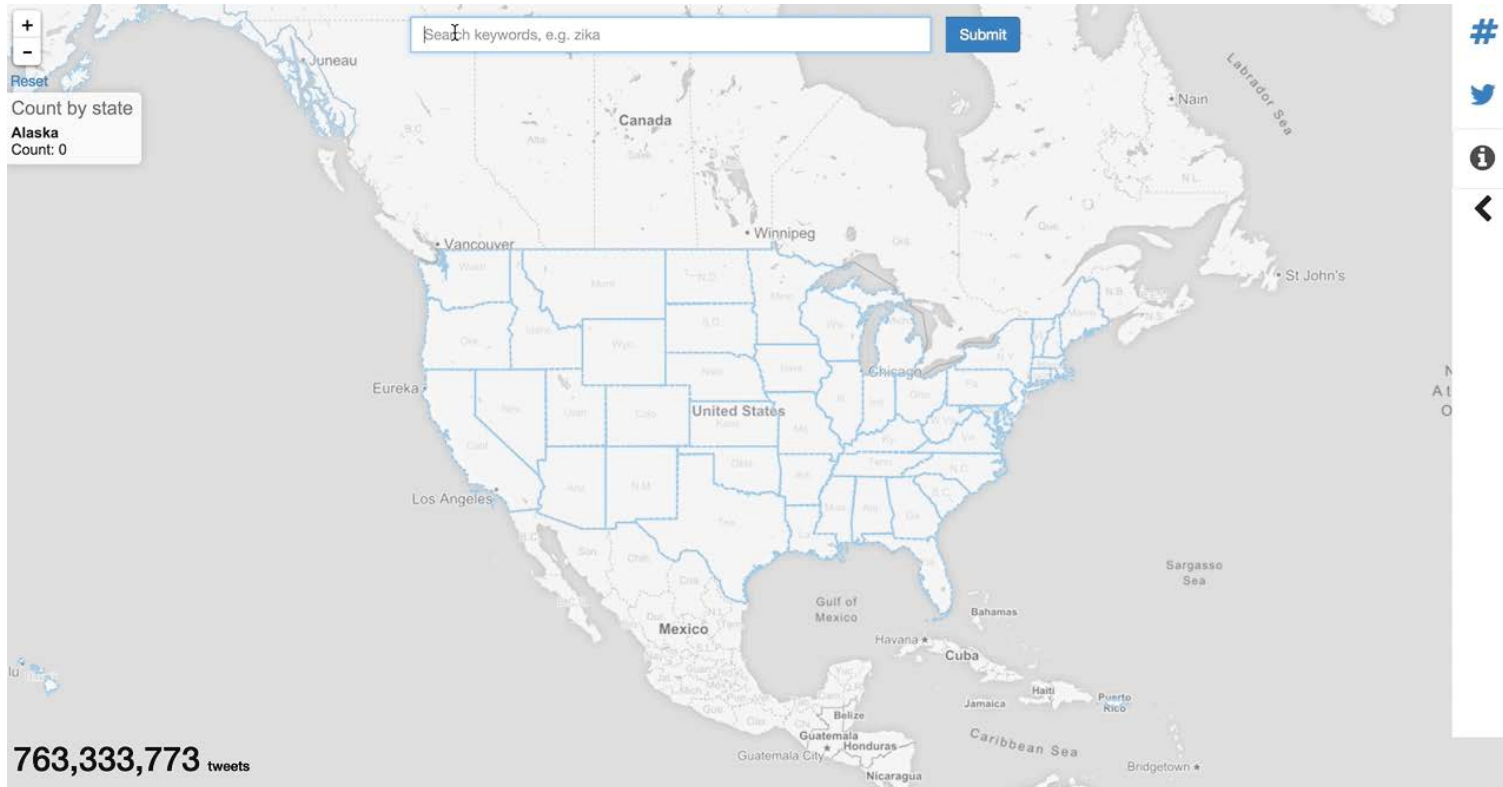


✓ Simple JPEG

Fixed-length slicing?

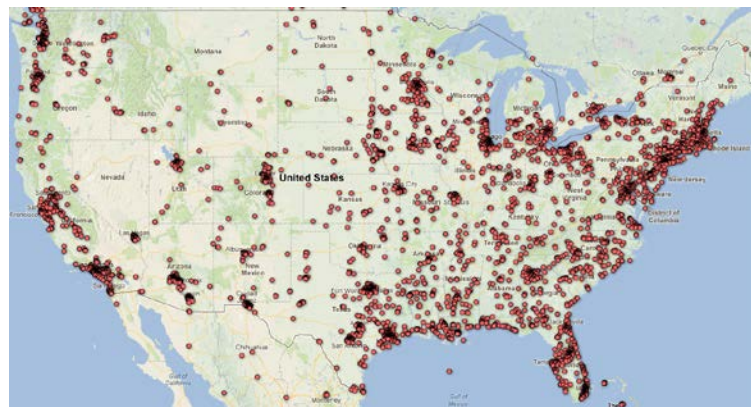


Query slicing with a rhythm



Open challenges

- Modeling DB for approximation viz
- Visualizing large number of records
- Integrating computing between middleware and frontend

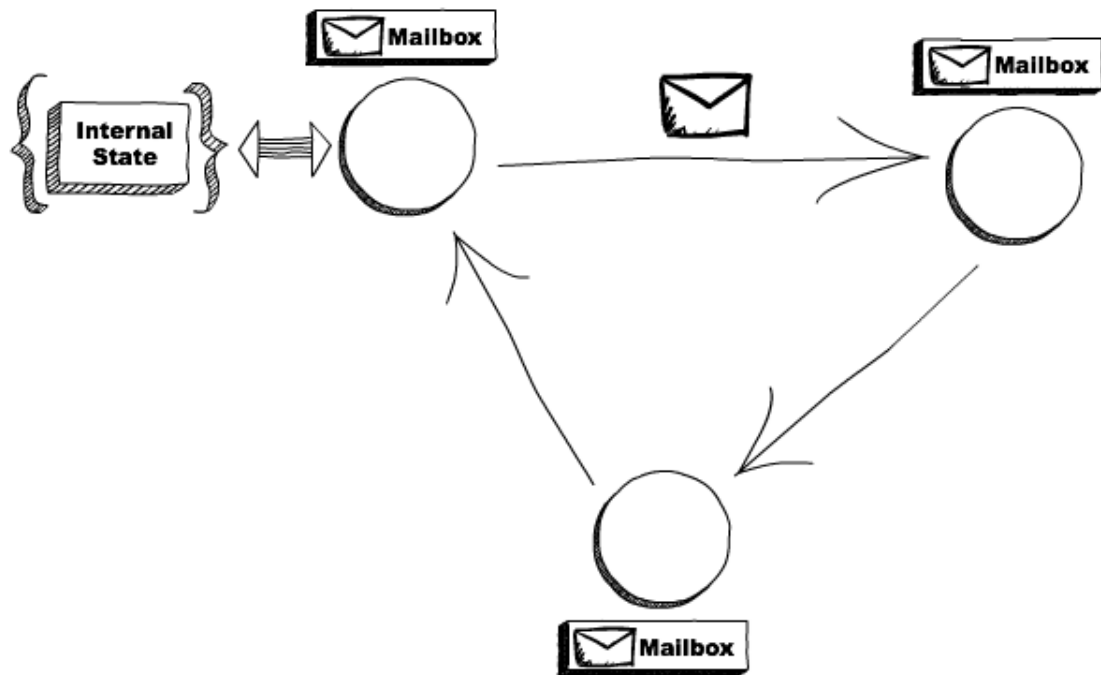


Texera: big data analytics using interactive workflows

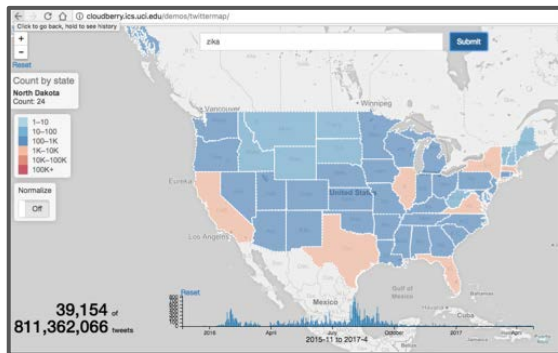
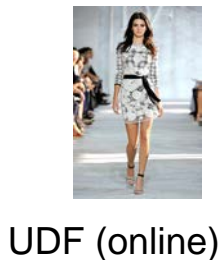
The screenshot displays the Texera workflow editor. At the top, there are buttons for 'Dictionary Upload' and 'Run'. Below this is a toolbar with actions: Sources, Search, Analyze, Split, Other, Join, Write Database, and View Results. The main workspace shows a workflow with four steps: 'Source: Scan', 'Keyword Se...', 'Sentiment A...', and 'View Results'. A dropdown menu is open under 'Sources', listing 'Source: Scan', 'Source: File', and 'Source: Keyword'. On the left, there are panels for 'Attributes' (with a dropdown and 'text' input), 'Query' (with 'ok' input), and 'SpanListName' (with 'result' input and a 'Delete' button). At the bottom right, there is a 'Download As Excel' button. The bottom section is a 'Result Window' showing a table of data.

text	tweet_link	user_link	user_screen_name	user_name	user_description	user_f
Ok.	https://twitter.com/statu	https://twitter.com/melis	melissatruth	Melissa	Scientist.	89

Actor Model



Integrate ML Models



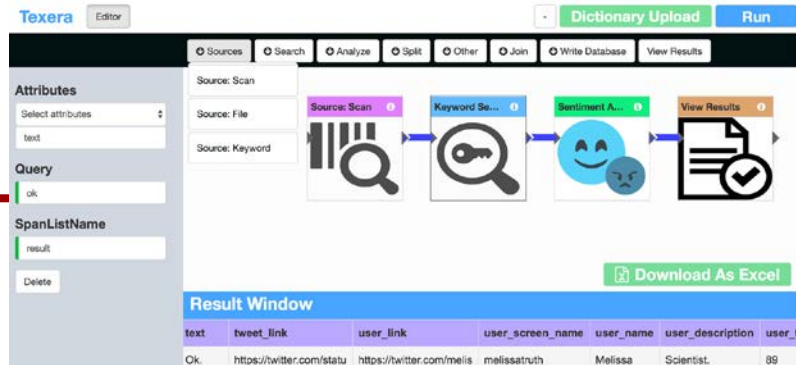
Data preparation for training

Training Instances

Included as operators



UDF (feed)



Tezra Editor

Dictionary Upload Run

Sources Search Analyze Split Other Join Write Database View Results

Attributes
Select attributes: text

Query
ok

SpanListName
result

Delete

Source: Scan
Source: File
Source: Keyword

Source: Scan Keyword Se... Sentiment A... View Results

Download As Excel

Result Window

text	tweet_link	user_link	user_screen_name	user_name	user_description	user_f
Ok	https://twitter.com/statu	https://twitter.com/melis	melissatruth	Melissa	Scientist.	89

Conclusion: data analytics as a service

