Titan

CRDT-based computations for Participatory Sensing applications

David Navalho (FCT-UNL) intern @ LIP6/INRIA

Introduction

- Increasing amount of available information
  - Content-generators/aggregators
    - e.g.: Google Now - maps, traffic, flights, hotels, restaurants, movies, concerts, etc...
  - User-generated data (Microblogging)
    - e.g.: Foursquare, Twitter
  - More sensor data every day
    - User's mobile devices

Participatory Sensing

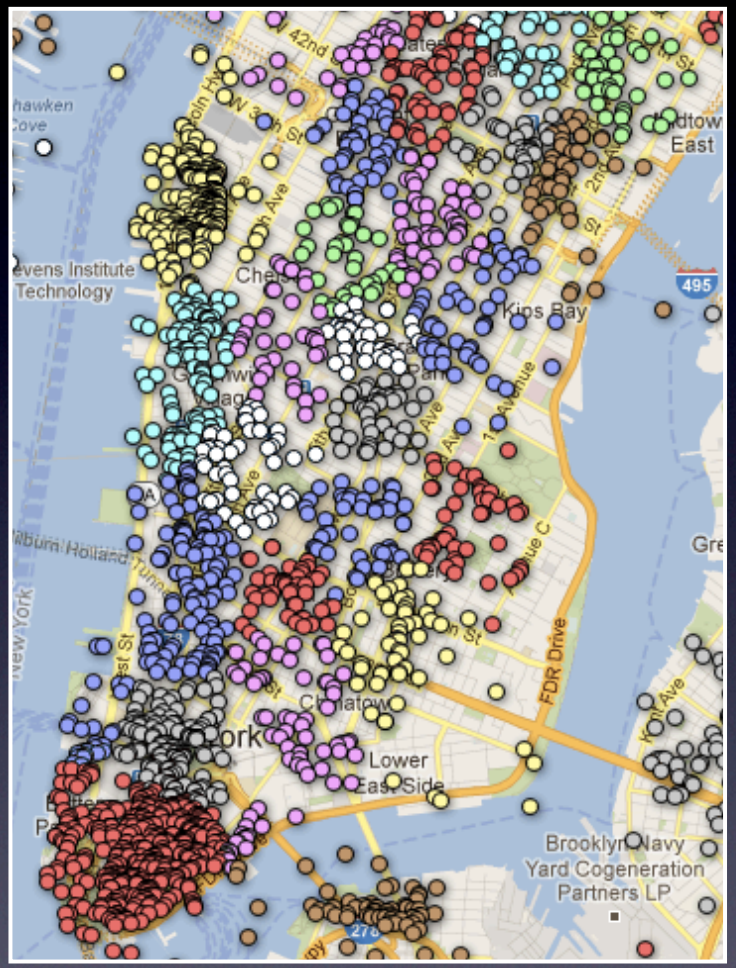
- Extracting information from sensors:
  - Mapping data (Wi-Fi, Cellular), Historical traffic
  - Traffic, road bumps (accelerometer+GPS)
  - Noise levels
  - ...

Participatory Sensing

- Leveraging sensors in mobile devices for performing wide-area sensing tasks
  - Process and cross reference sensor data from multiple sources to detect patterns and infer new data

Example

"Livehoods reveal how the people and places of a city come together to form the dynamic character of local urban areas."



Titan: Objectives

- Efficient, (near) real-time computations
- Building data-flows easily
- Quickly Bootstrap new applications by feeding from previously available information and computations

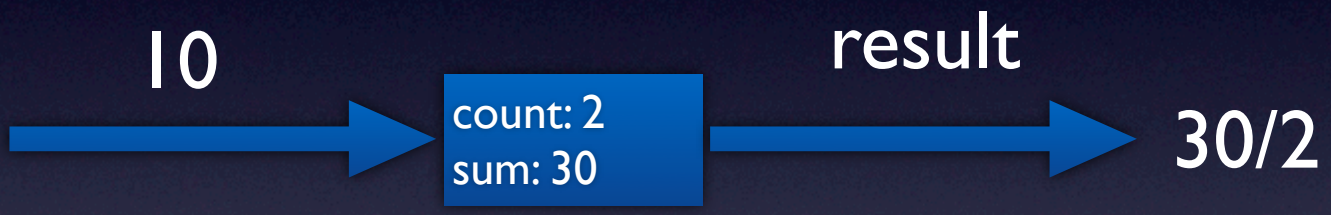
Computations

- Efficient Computation
  - Needs to be processed in (near) real-time
  - Most computations are highly commutative
    - e.g., sums, averages, aggregations...
- CRDTs are used for both storage and computation

Example: Average

count: 1
sum: 20

Example: Average

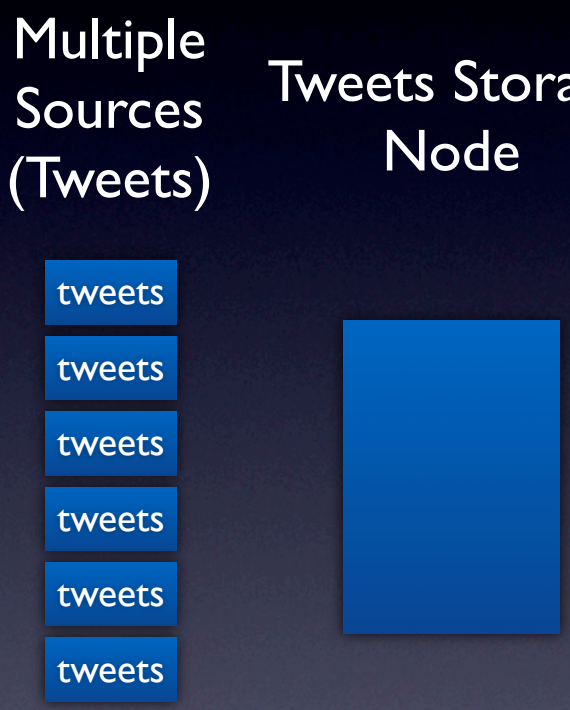


Computations

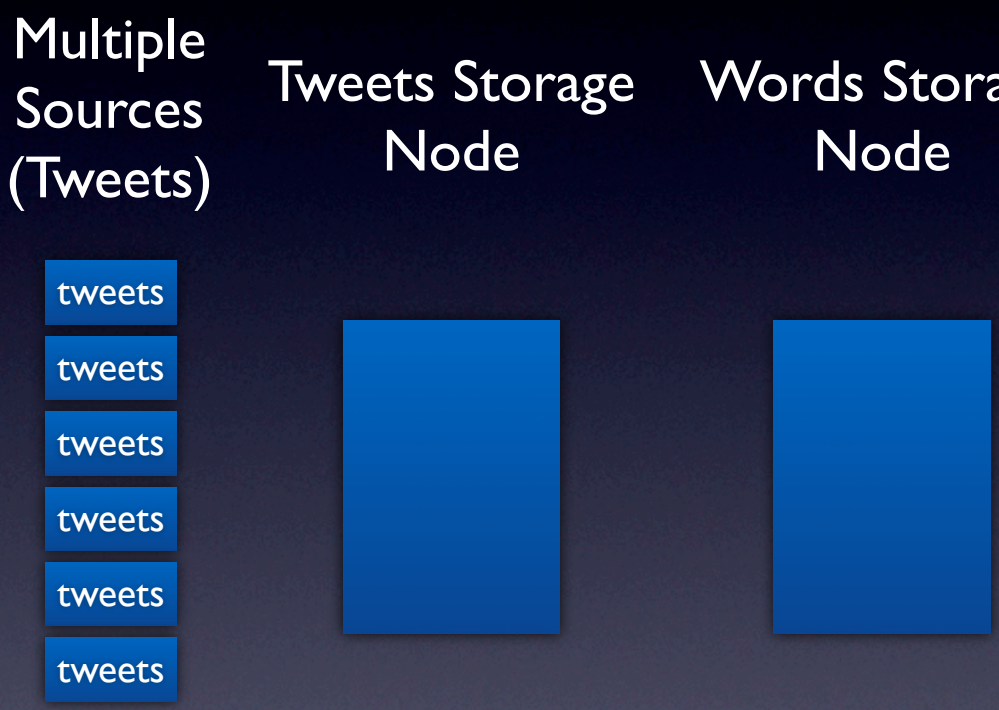
Computations

Multiple Sources (Tweets)

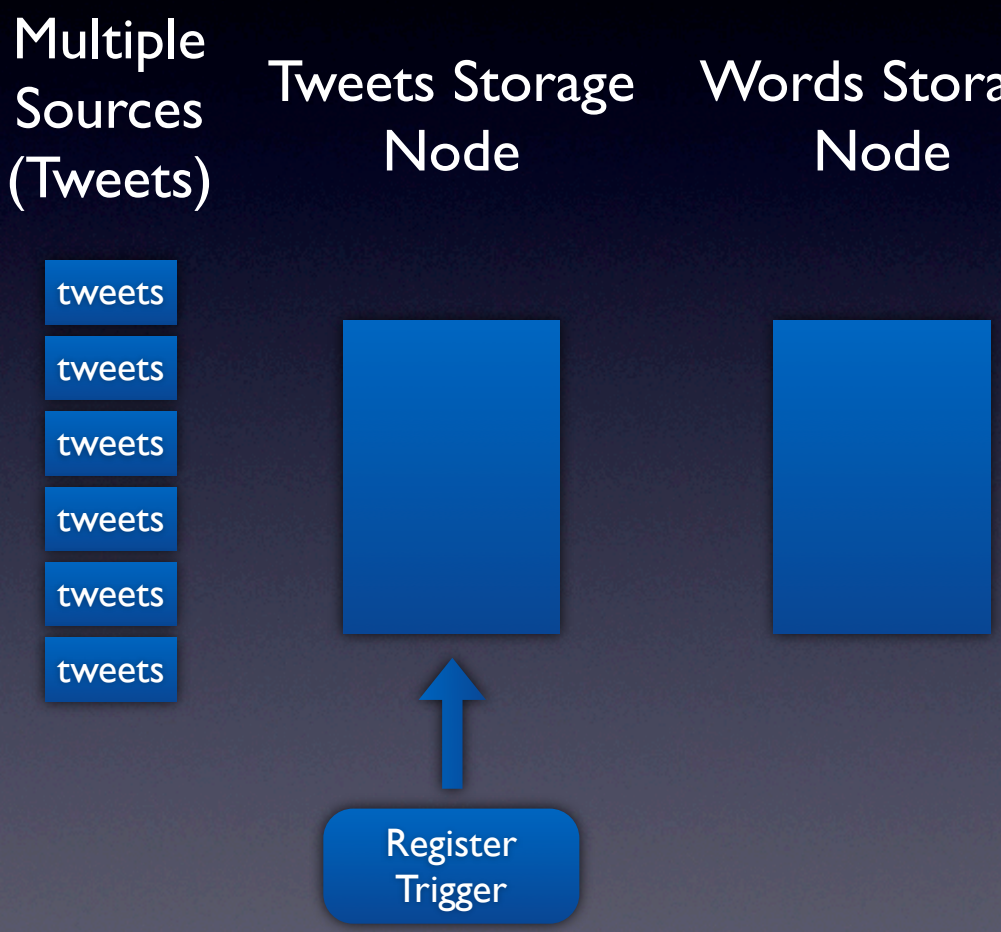
Computations



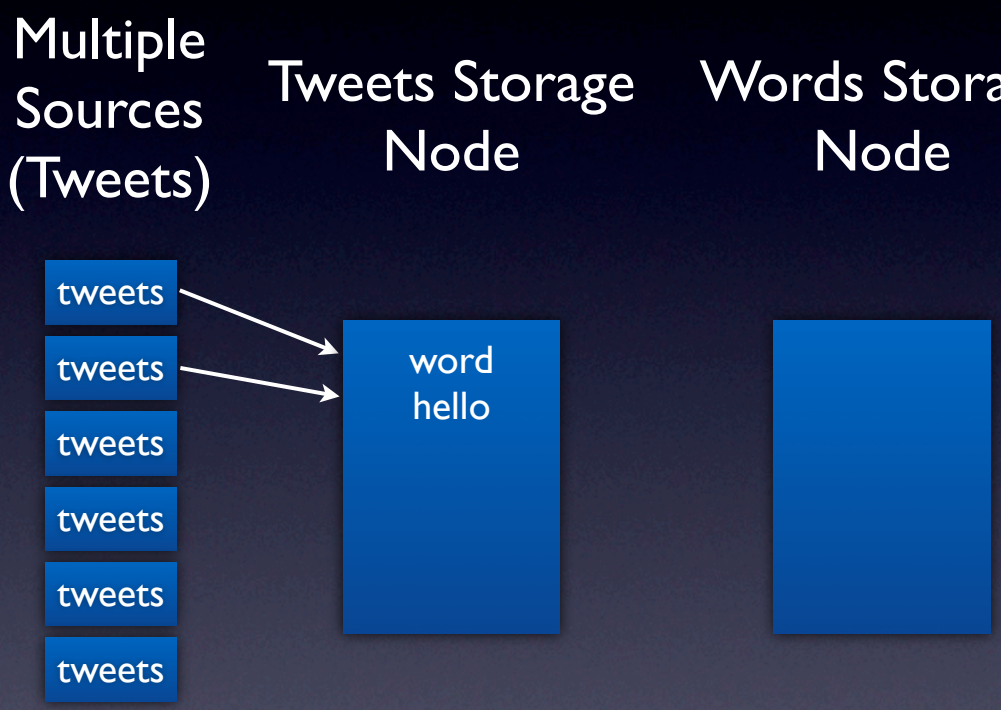
Computations



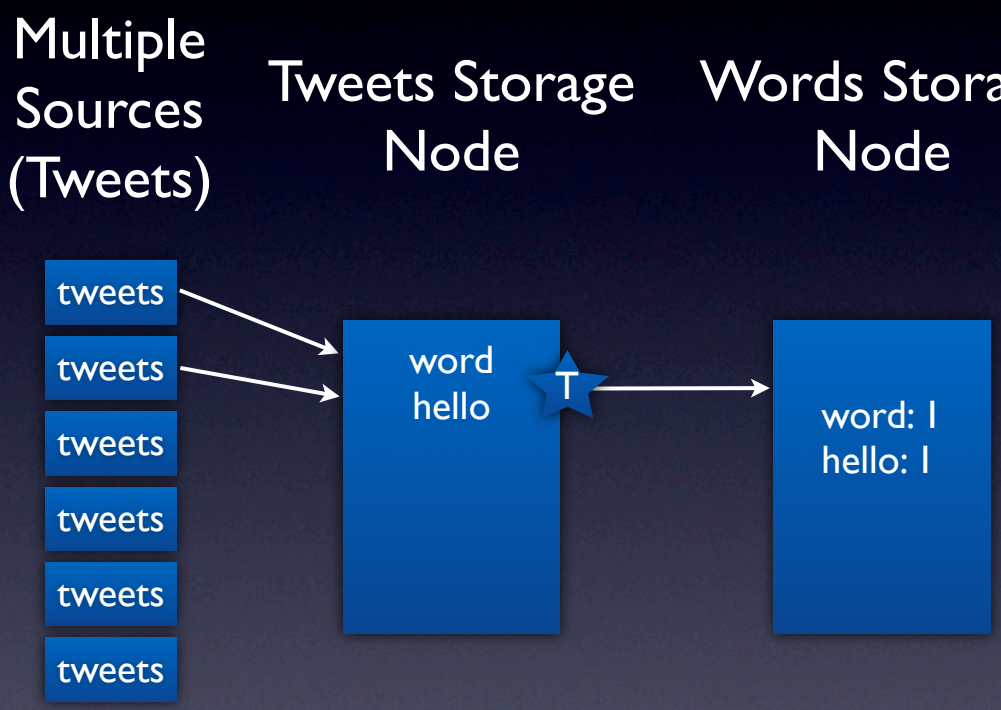
Computations



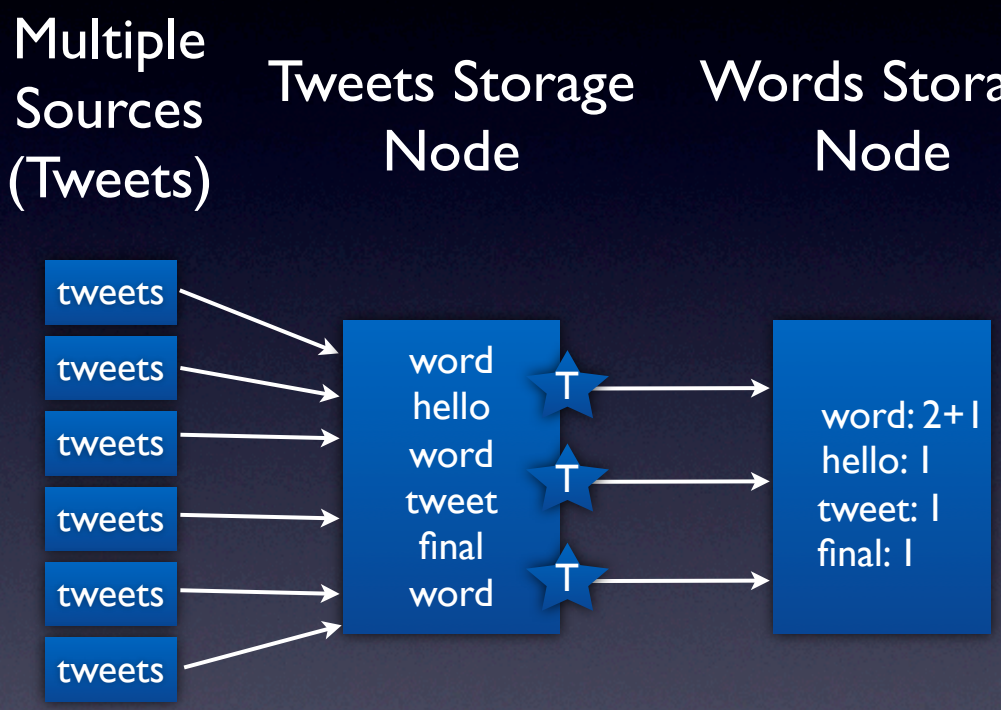
Computations



Computations



Computations



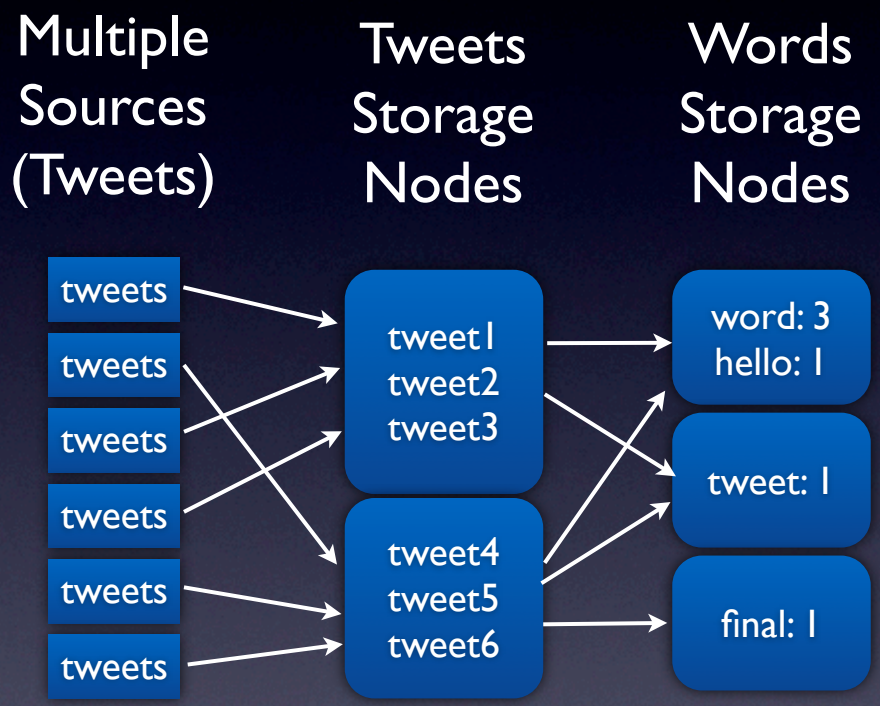
Scaling the Computations

- Massive amounts of data
  - Computation Partitioning
    - Map-Reduce like approach

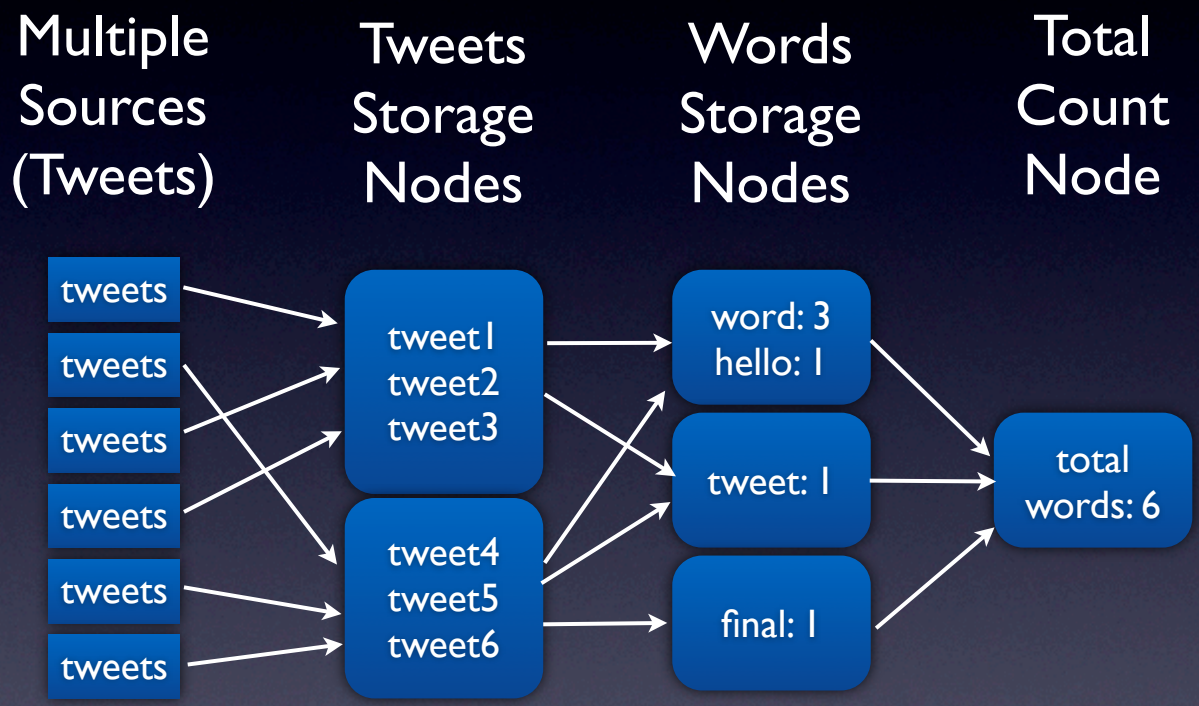
Partitioning

Multiple Sources (Tweets)

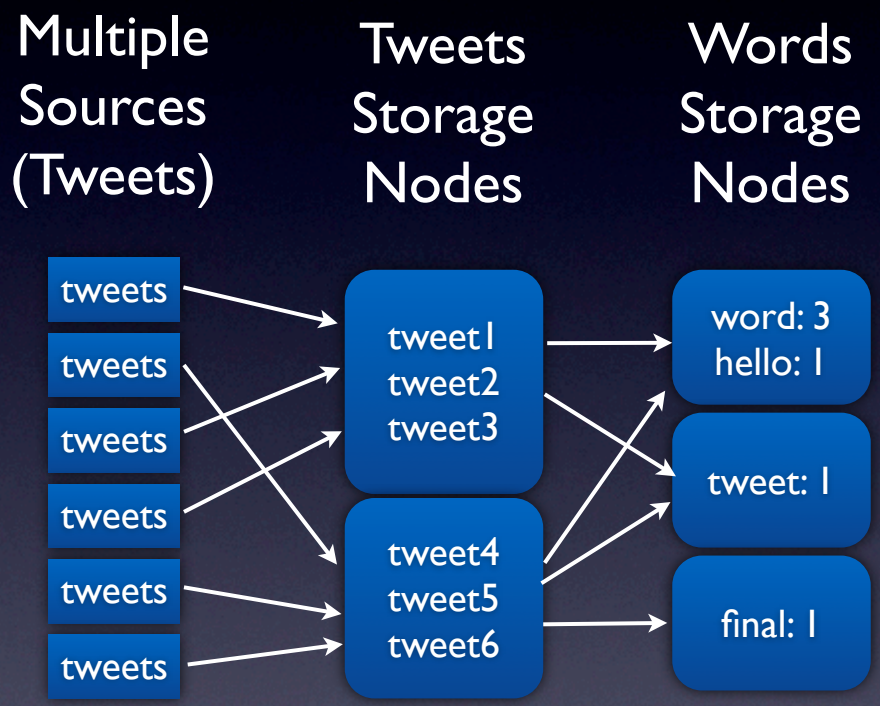
Partitioning



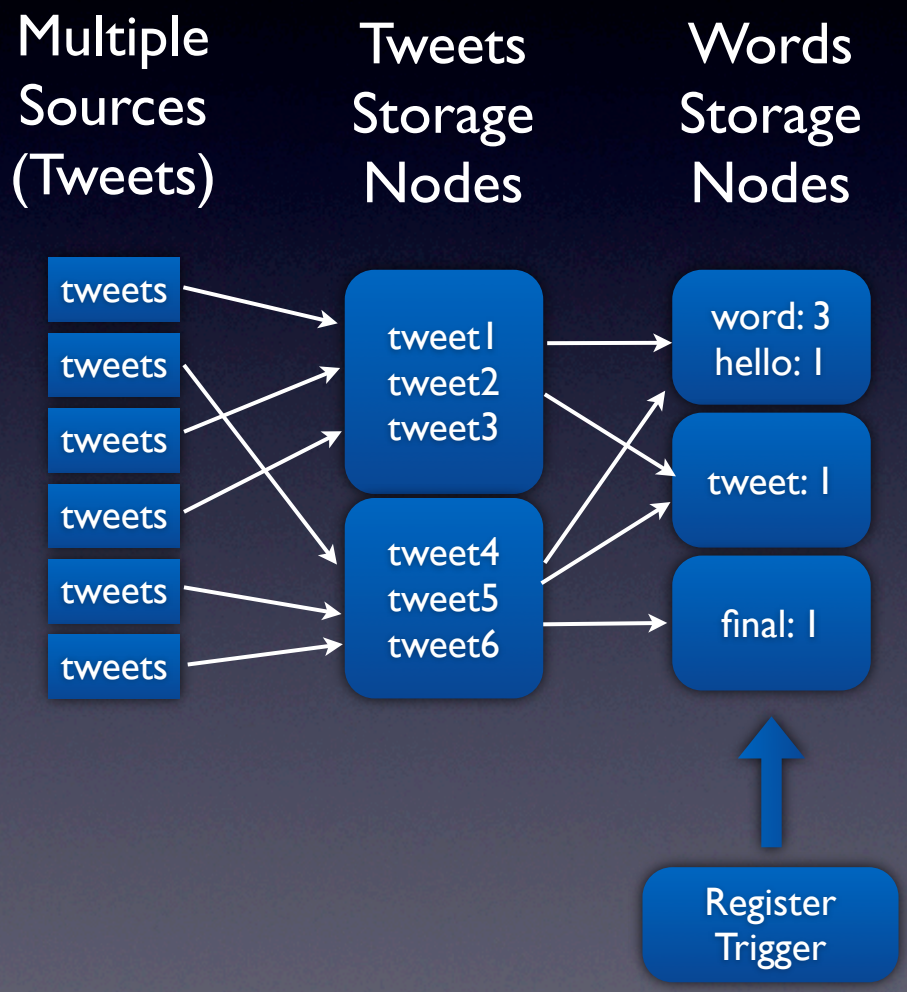
Partitioning



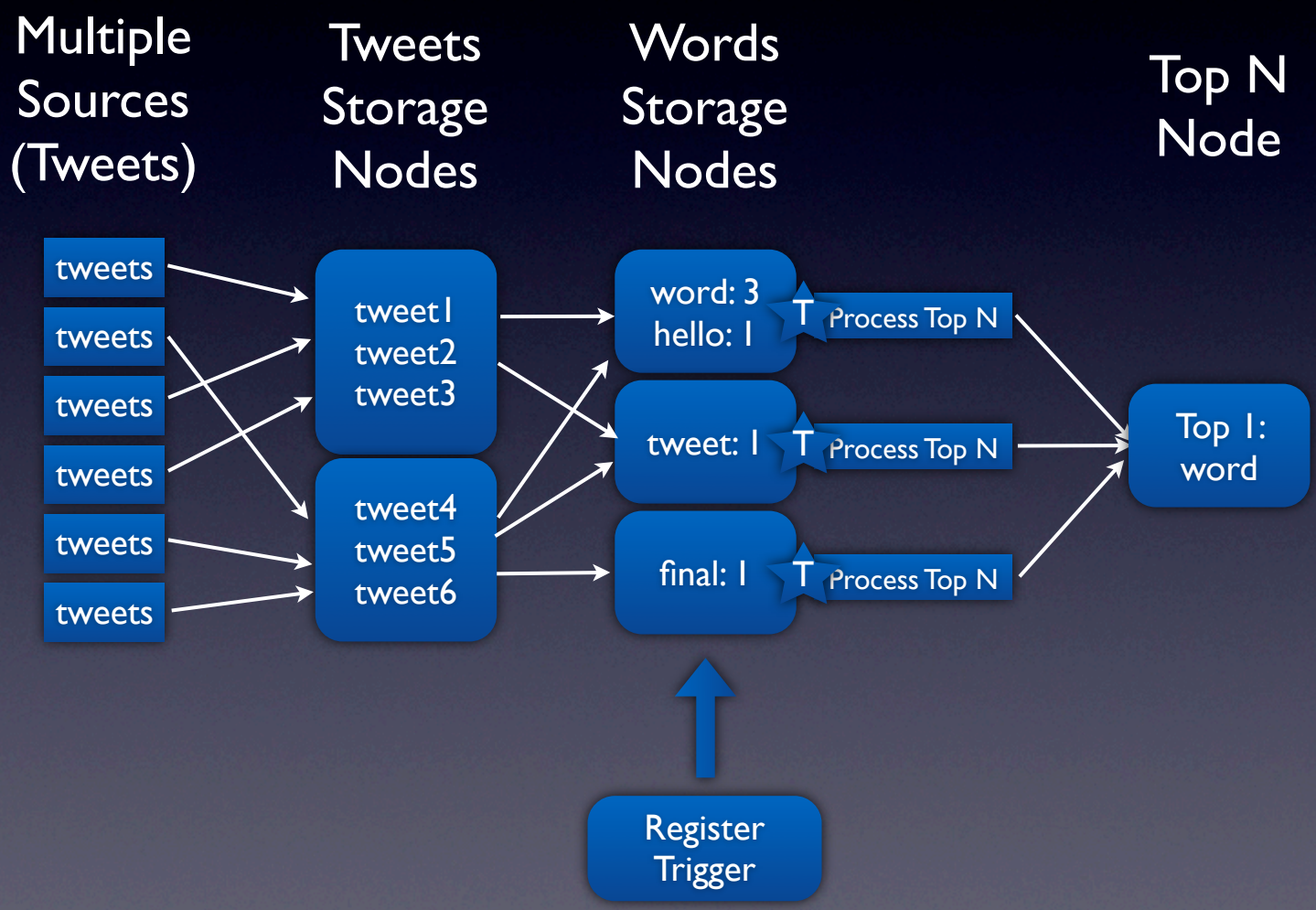
Top N


Multiple Sources (Tweets)
Tweets Storage Nodes
Words Storage Nodes
tweets
tweets
tweets
tweets
tweets
tweets
tweet1
tweet2
tweet3
tweet4
tweet5
tweet6
word: 3
hello: 1
tweet: 1
final: 1

Top N


Multiple Sources (Tweets)
Tweets Storage Nodes
Words Storage Nodes
tweets
tweets
tweets
tweets
tweets
tweets
tweet1
tweet2
tweet3
tweet4
tweet5
tweet6
word: 3
hello: 1
tweet: 1
final: 1
Register Trigger

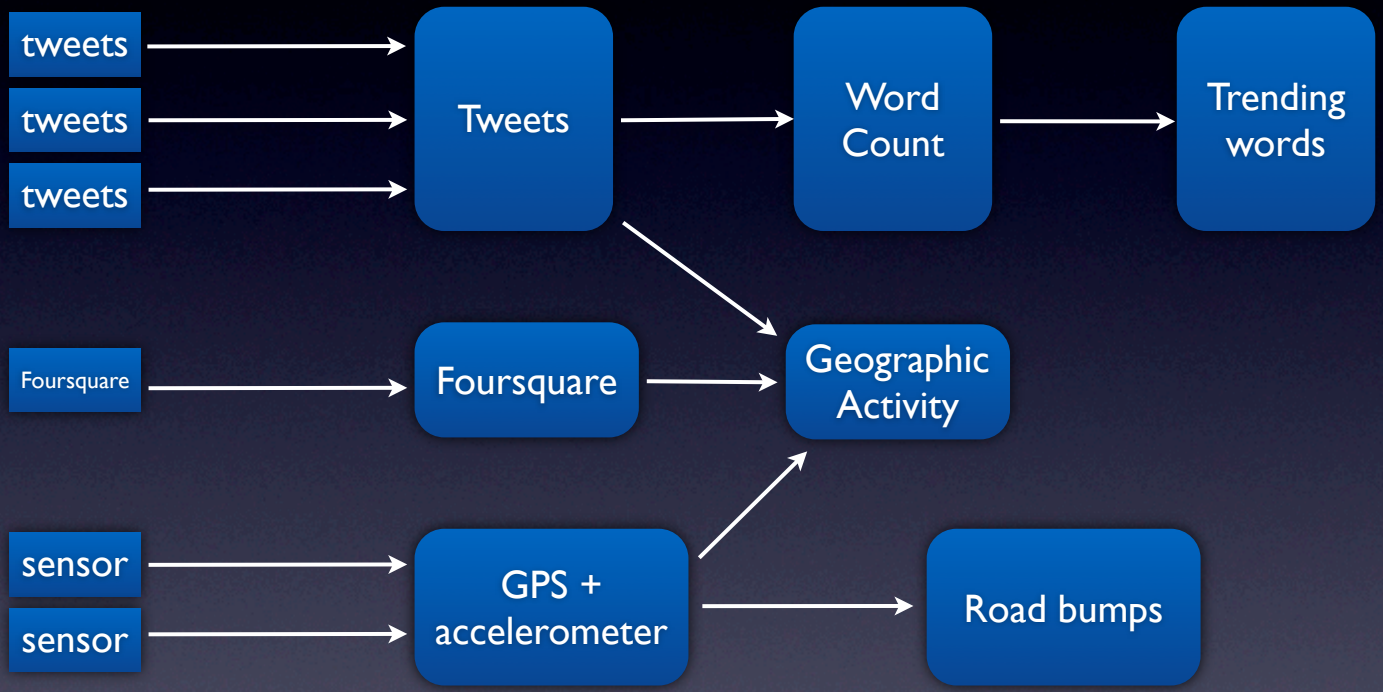
Top N


Multiple Sources (Tweets)
Tweets Storage Nodes
Words Storage Nodes
Top N Node
tweets
tweets
tweets
tweets
tweets
tweets
tweet1
tweet2
tweet3
tweet4
tweet5
tweet6
word: 3
hello: 1
tweet: 1
final: 1
T Process Top N
T Process Top N
T Process Top N
Top 1: word
Register Trigger

Conclusions

- Titan is a system intended to provide support for the creation and testing of Participatory Sensing applications
- It takes advantage of the highly commutative nature of PS applications
  - CRDTs are extensively used in the system to provide both storage and computation

Conclusions



Future Work

- Addressing Multiple keys
  - e.g., Time-based computations
- Automatic partitioning
- Fail/Stop recovery

Thank you