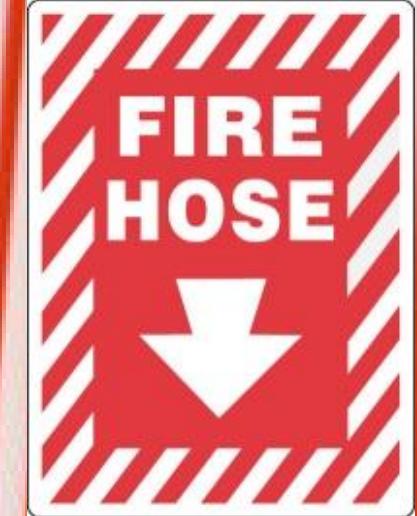




# Straw

Ryan Walker



Drink safely!

A platform for real-time search on streaming text data

<https://github.com/rwalk333/straw>



**Let's see the demo**



documents

documents



queries

queries

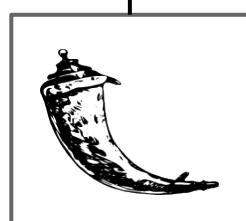
documents

queries

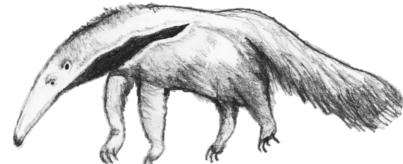
matches



elasticsearch



STORM



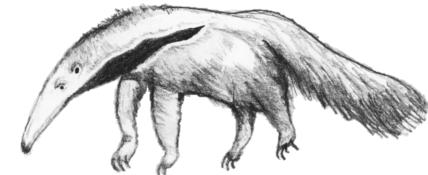
matches



redis

publish results to  
subscribers

# Streaming Text Search

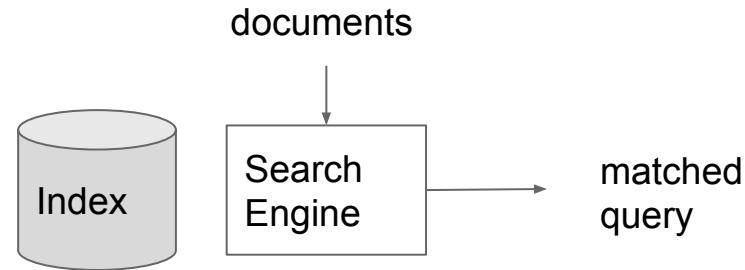


## Traditional Text Search



Inverted index holds  
documents -- slow to build  
but fast to query

## Streaming Search



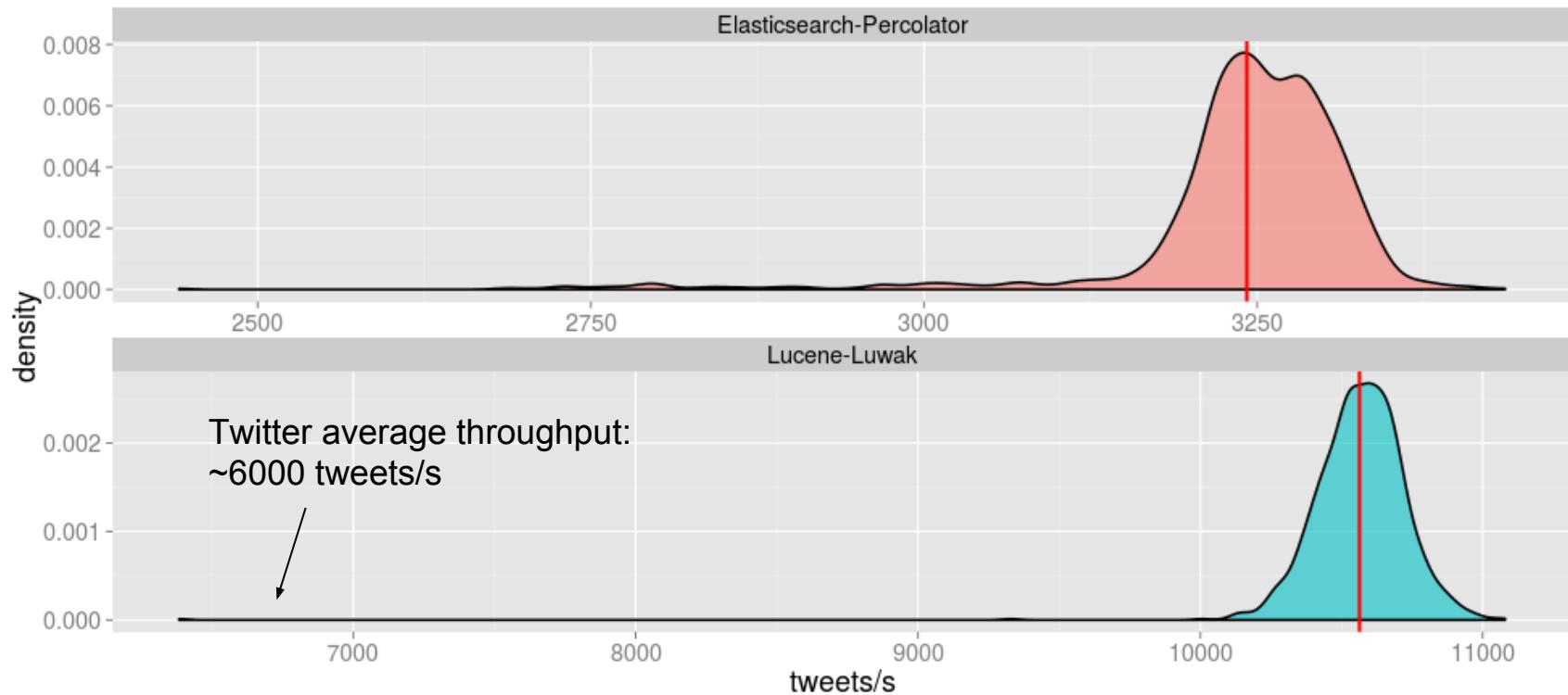
Inverted index holds the queries --  
documents stream through the  
search engine

- Implementations:
- Elasticsearch Percolators
  - Lucene-Luwak

# Some Benchmarks



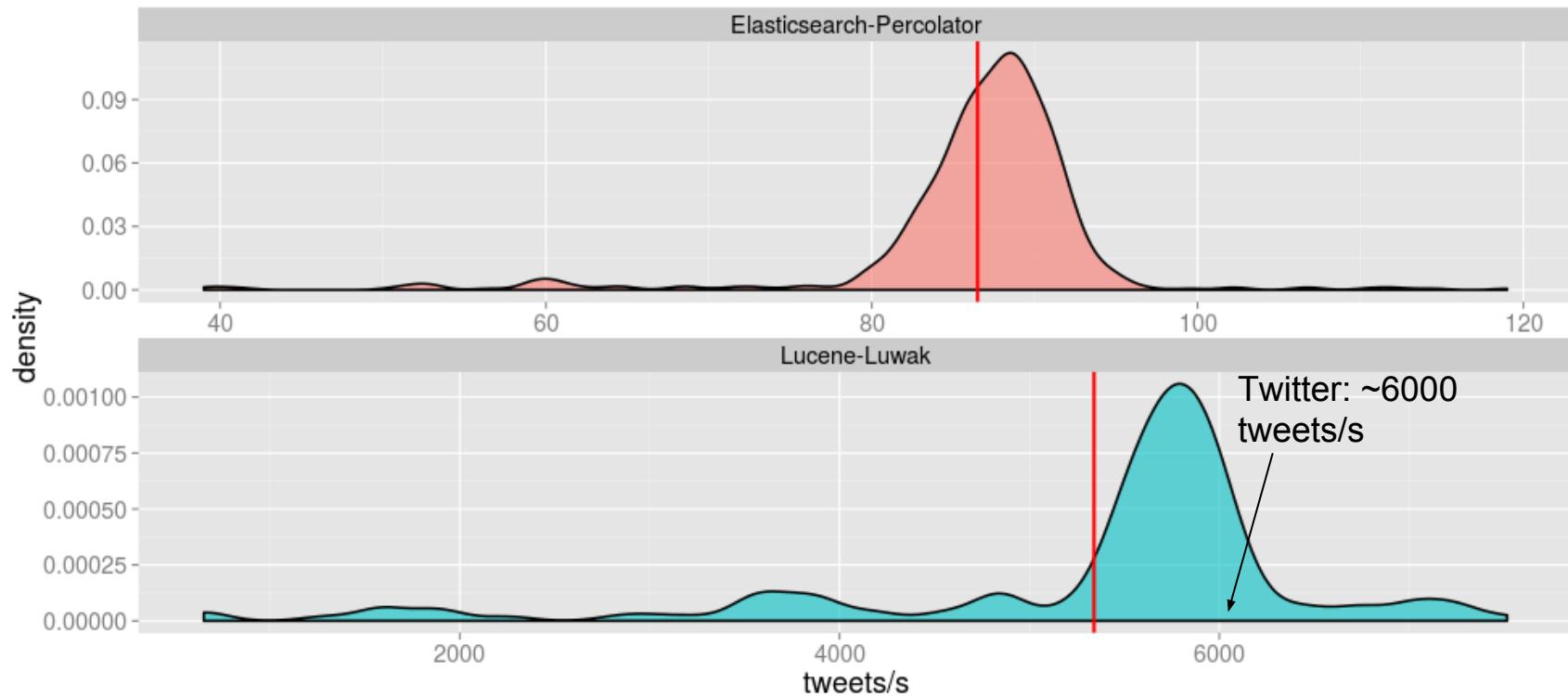
Throughput per second with ~500 registered queries on a 5 bolt Storm cluster of m4.xlarge instances



# Some Benchmarks



Throughput per second with ~100000 registered queries on a 5 bolt Storm cluster of m4.xlarge instances



# Ryan Walker



## Education:

- PhD Mathematics, University of Kentucky

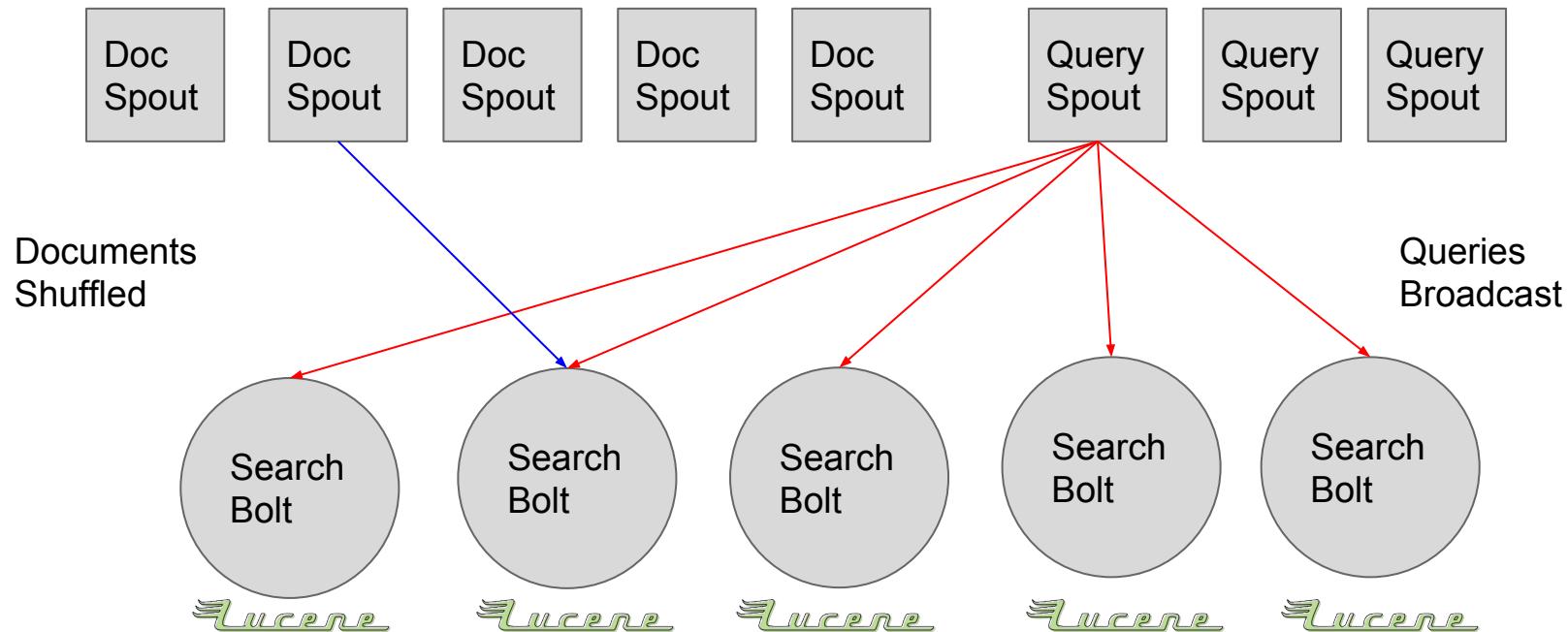
## Experience:

- Data science at major health insurer (2 years)
- Platform/Machine Learning engineer at NLP startup (2 years)

blue  of california



# Storm Topology



# Scaling for Queries



STORM

