

# Limmat: Analytics for the Real-Time Web

Maxim Grinev

Maria Grineva

Martin Hentschel

Donald Kossmann



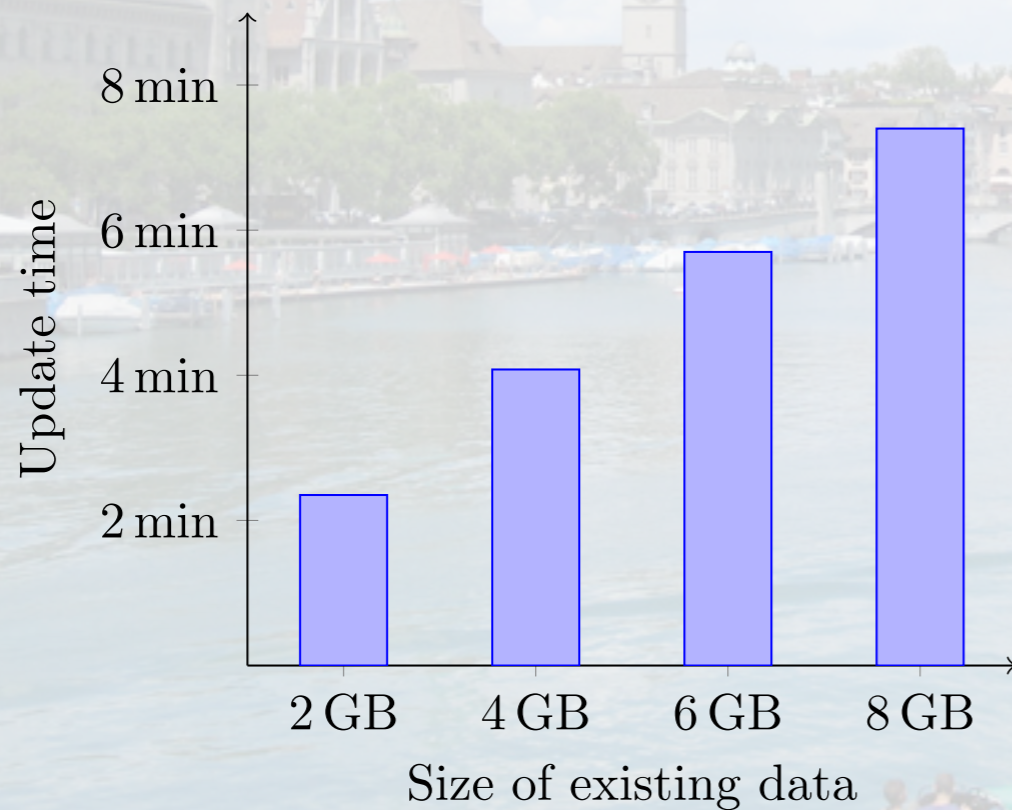
# MapReduce super cool!

- ▶ Simple programming model
- ▶ Fault tolerance
- ▶ Scalability

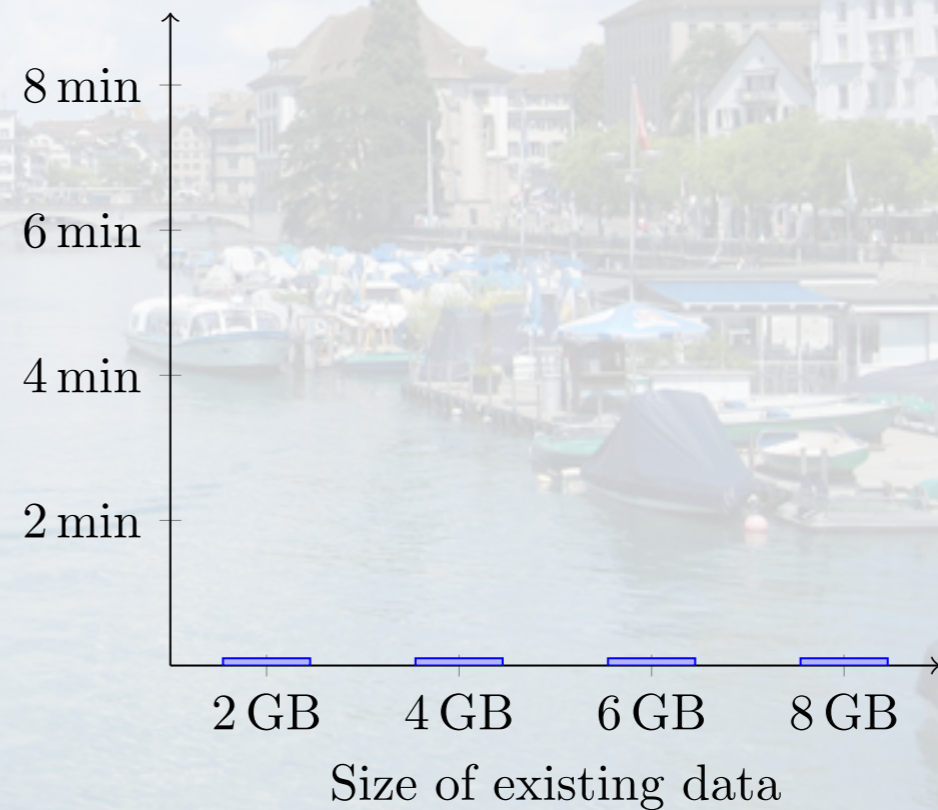
But: Data freshness crappy!

# Data freshness experiment: Update existing data with 2 MB of new data

## Hadoop



## Limmat



# Say Hello to Limmat

- ▶ MapReduce-like programming model

$map : (k_1, v_1) \rightarrow list(k_2, v_2)$

$reduce' : (k_2, v_2, agg_{old}) \rightarrow (k_2, agg_{new})$

- ▶ Fault tolerance
- ▶ Scalability
- ▶ Awesome data freshness!

# Demo

- ▶ Live web application: analyzes 1% Twitter stream  
<http://limmat.ethz.ch>
- ▶ Scalability: analyze 100% Twitter stream with 14 machines
- ▶ Fault tolerance: kill any node