



# Fast Data with the SMACK stack\_

---

data2day  
October 7th 2016

# Who am I?

## Matthias Niehoff

IT Consultant

codecentric AG  
Solingen

Cassandra and Spark  
Streaming

 @matthiasniehoff



# Who am I?

## Florian Troßbach

IT Consultant

codecentric AG  
Karlsruhe

Plain Old Java Dev  
Fast Data  
SMACK stack

 @ftrde83



Who are you?

---

## Tell us about yourselves.

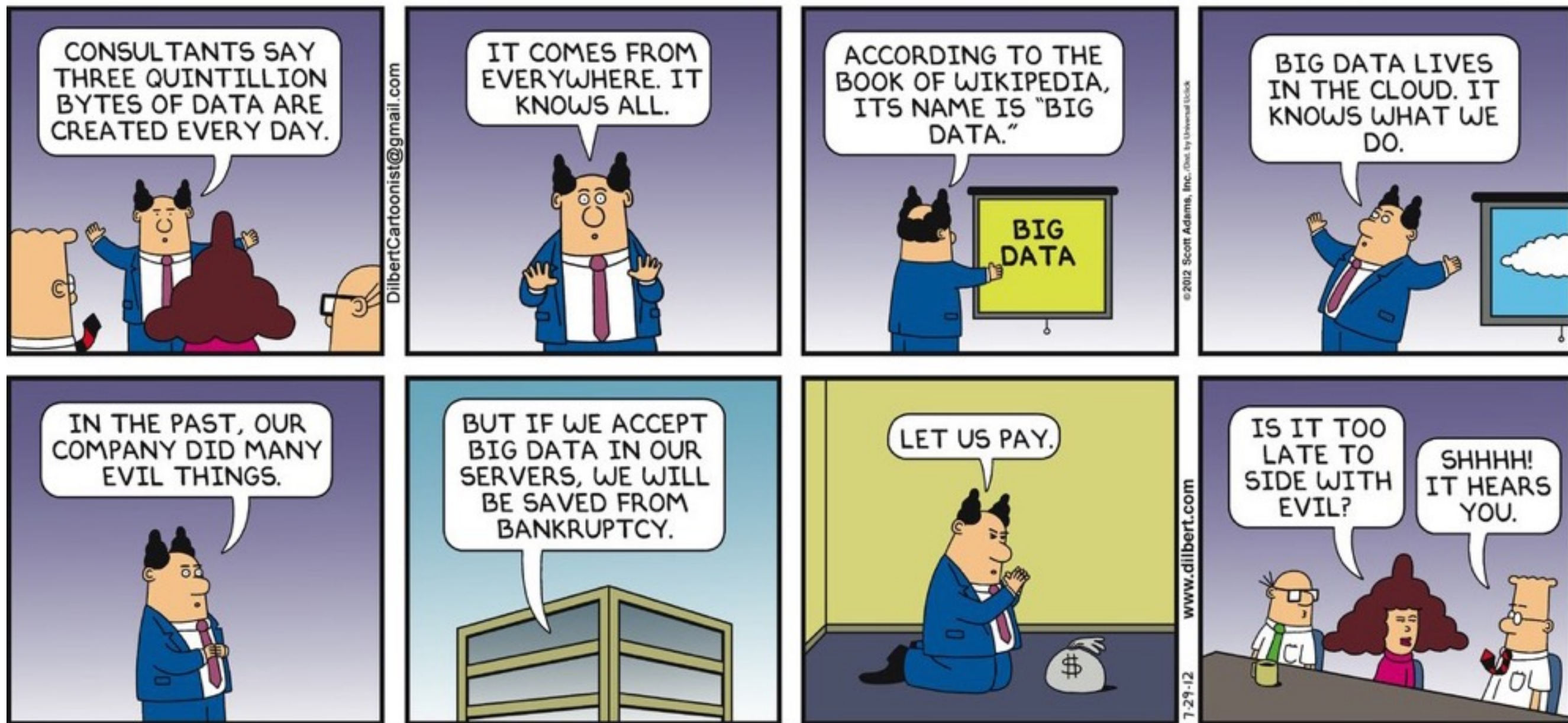
- What do you expect?
- Do you have prior experience with the stack?



# What is Fast Data?



# Big Data\_



## Traditional data processing does not work

- Data too large
- Data too complex
- Data possibly unstructured

# Hadoop

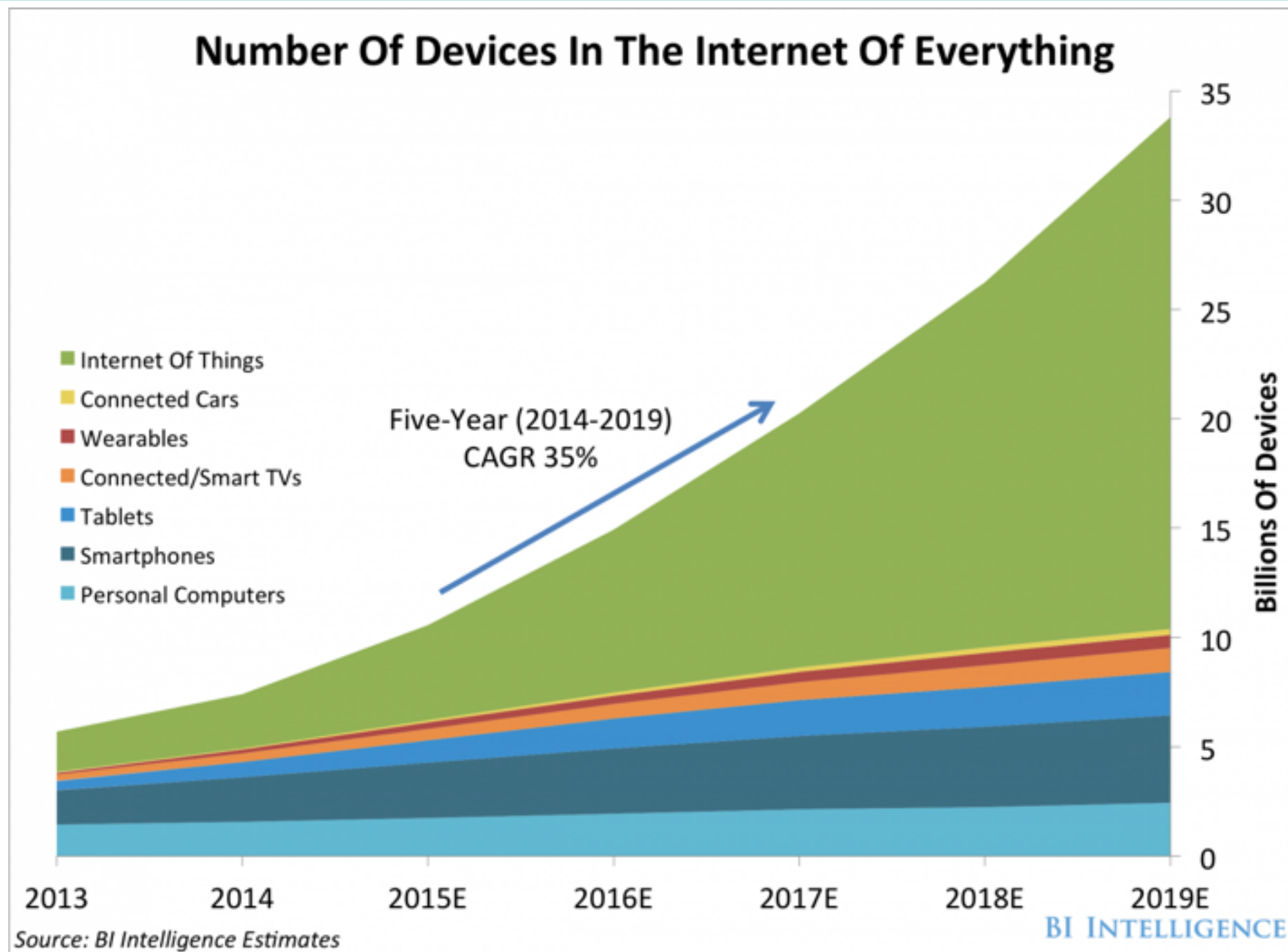
- MapReduce
- HDFS





Works well, doesn't it?

# Internet of Everything\_



<http://static1.businessinsider.com/image/54a4207d69beddbd357d04ad-1200/number-of-devices-in-the-internet-of-everything.png>





Image by <https://www.flickr.com/photos/joiseyshowaa/>  
<https://creativecommons.org/licenses/by-sa/2.0/>



## The case for Fast Data

- Volume of Data is only rising
- “Real-time” capability desired
- Platforms must be able to react to variable load

## Technologies that complement each other well

- Apache Spark for analytics
- Apache Mesos for basic infrastructure and Scalability
- Akka for data ingestion (and possibly extraction)
- Apache Cassandra for (hot) storage
- Apache Kafka as a buffer to deal with load peaks

# The use case for today



## We're going to analyse Spotify data

- Running on Mesos in AWS
- Using Akka to read data from a Websocket
- Write the data to Kafka
- Analyse it with Spark
- Store it in Cassandra

