

# Big Data Tools: Fantastic Beasts and What Do They Do

Pasha Finkelshteyn,  for Big Data 

# **What is Big Data?**

## **Nobody knows**

# What is Big Data?

- Doesn't fit a single node
- Scales without rearchitecturing when grows (3V)
  - Volume
  - Velocity
  - Variety
- Enough data to make reliable business decisions

# Who handles this data? Data Engineers!

# Responsibilities: Pipelines

Pipelines transfer data from sources  
to sinks

Sometimes pipelines leak

Data engineer is pipeliner and  
plumber™ in one



# Responsibilities: architecture

Architecture of storage

Architecture of pipelines

Tool selection

