BIG DATA STATE OF THE ART

Dra. Liana Napalkova
Senior Data Scientist
Eurecat
http://eurecat.org
@eurecat_news

Big Data Mining: Roadmap to a Winning Strategy

Liana Napalkova, PhD
Senior Data Scientist
Big Data Analytics Unit - Eurecat

www.eurecat.org
October 2016

eurecat
Centre Tecnològic de Catalunya

eurecat
Centre Tecnològic de Catalunya

Big Myths about Big Data

EXPECTATIONS **REALITY**

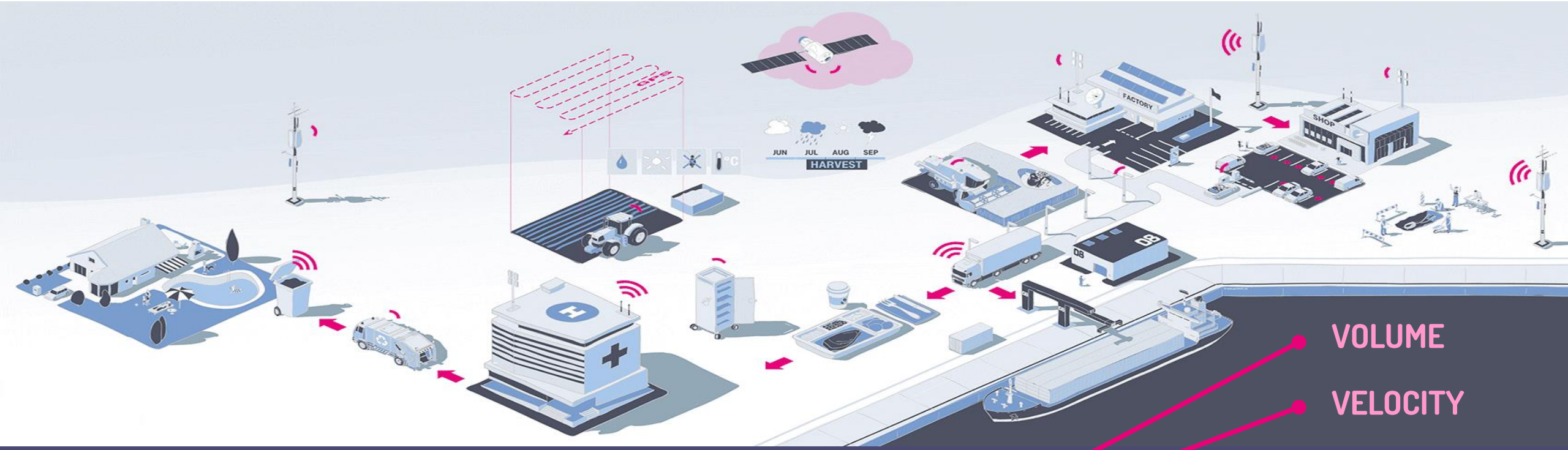
WHY DOES IT HAPPEN?


ONE SIZE FITS ALL


IT'S ALL ABOUT TECHNOLOGY


DATA EQUALS INFORMATION

Big Data requires New Vision



DATA IS EVERYWHERE.

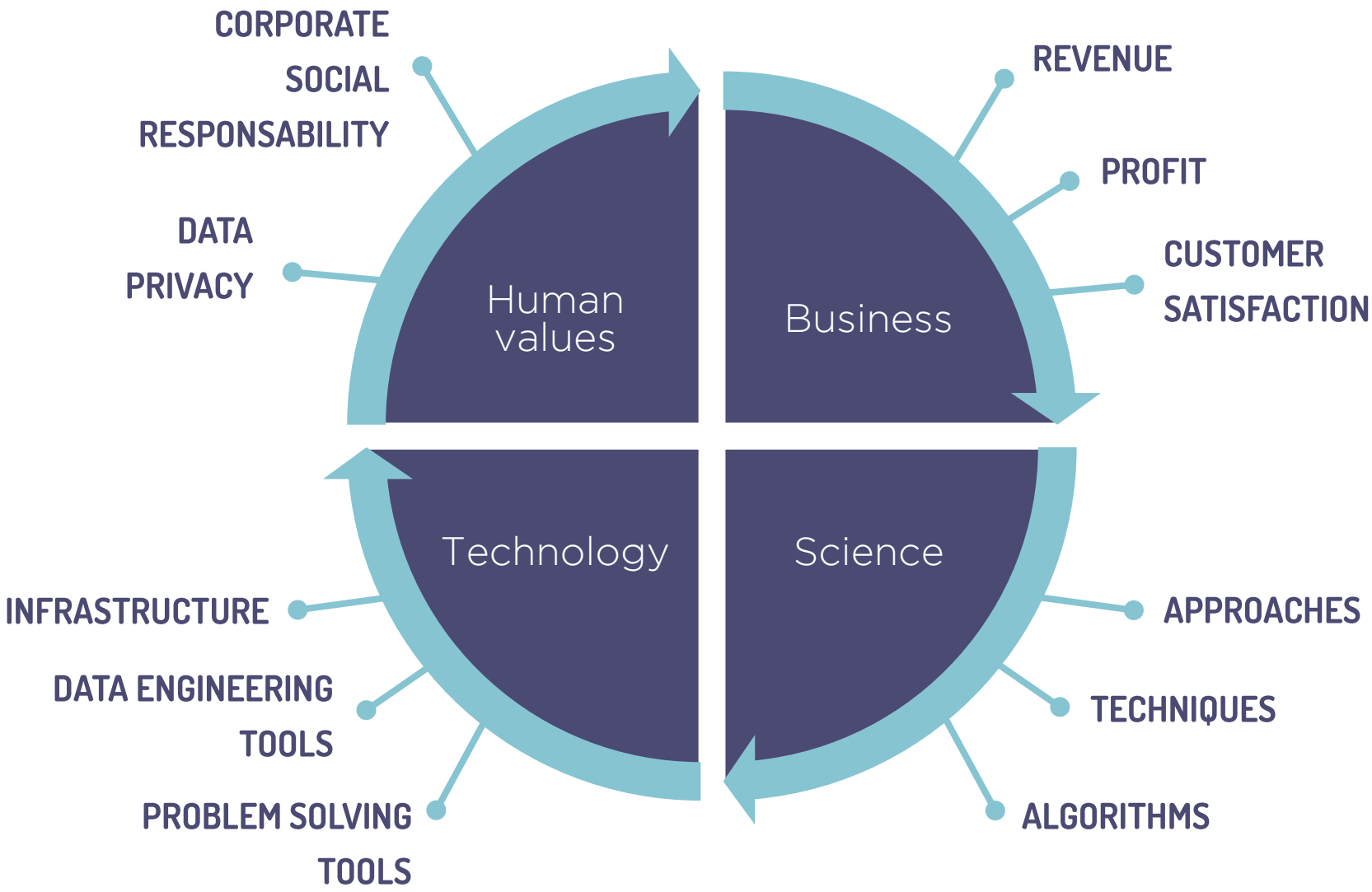
BUT WHAT DOES IT MAKE "BIG"?

BIG DATA 6V

- VOLUME
- VELOCITY
- VARIETY
- VERACITY
- VIABILITY
- VALUE

eurecat
Centre Tecnològic de Catalunya

The Big Deal With Big Data Isn't (Just) the Data



eurecat
Centre Tecnològic de Catalunya

A ROADMAP

1. WHICH ARE THE HIGH-LEVEL GOALS?

- S - SPECIFIC
- M - MEASURABLE
- A - ATTAINABLE
- R - RELEVANT
- T – TIME-BASED

2. WHICH TYPE OF QUESTIONS SHOULD BE ANSWERED?

- WHAT HAPPENED?
- WHY DID IT HAPPEN?
- WHAT WILL HAPPEN?
- HOW CAN WE MAKE IT HAPPEN?

3. WHICH ARE CORRESPONDING APPROACHES?

- DESCRIPTIVE ANALYTICS
- DIAGNOSTIC ANALYTICS
- PREDICTIVE ANALYTICS
- PRESCRIPTIVE ANALYTICS

4. ANALYTICAL SANDBOX (SCIENCE PERSPECTIVE)

- CLASSIFICATION
- CLUSTERING
- DIMENSIONALITY REDUCTION
- MODEL SELECTION
- REGRESSION
- RECOMMENDATION
- PREPROCESSING
- OPTIMISATION

5. TECHNOLOGICAL STACK (ENGINEERING PERSPECTIVE)

PIG, APACHE STORM™ Distributed · Resilient · Real-time, APACHE hadoop, APACHE Spark™, HIVE, TensorFlow, sqoop, mahout, OOZIE, HUE

Topics to be Considered in the Future

eurecat
Centre Tecnològic de Catalunya