

Introduction to IoT Data Stream Mining

Albert Bifet (@abifet)



Paris, 25 November 2016
albert.bifet@telecom-paristech.fr

Who are We

- ▶ Jesse Read
 - ▶ Associate Professor at École Polytechnique
 - ▶ MultiLabel Learning, Data stream mining and Deep Learning
 - ▶ MEKA: Multilabel Learning
 - ▶ Molearn: A Multi-label/multi-target framework in Python
 - ▶ MOA: Massive Online Analytics
- ▶ Albert Bifet
 - ▶ Associate Professor at Télécom ParisTech
 - ▶ Data stream mining algorithms and systems
 - ▶ MOA: Massive Online Analytics
 - ▶ Apache SAMOA: Scalable Advanced Massive Online Analytics

IoT Data Stream Mining

Outline

1. Introduction
2. Stream Algorithmics
3. Classification in Multi-output Data Streams
4. Concept Drift
5. Multi-output Learning
6. Ensemble Methods
7. Regression
8. Clustering
9. Frequent Pattern Mining

IoT Data Stream Mining

Assessment

10% Lab Assignments

30% Project

60% Test

Classes

25/11 and 2/12 Fridays at 9:00

29/11 and 6/12 Tuesdays at 9:00

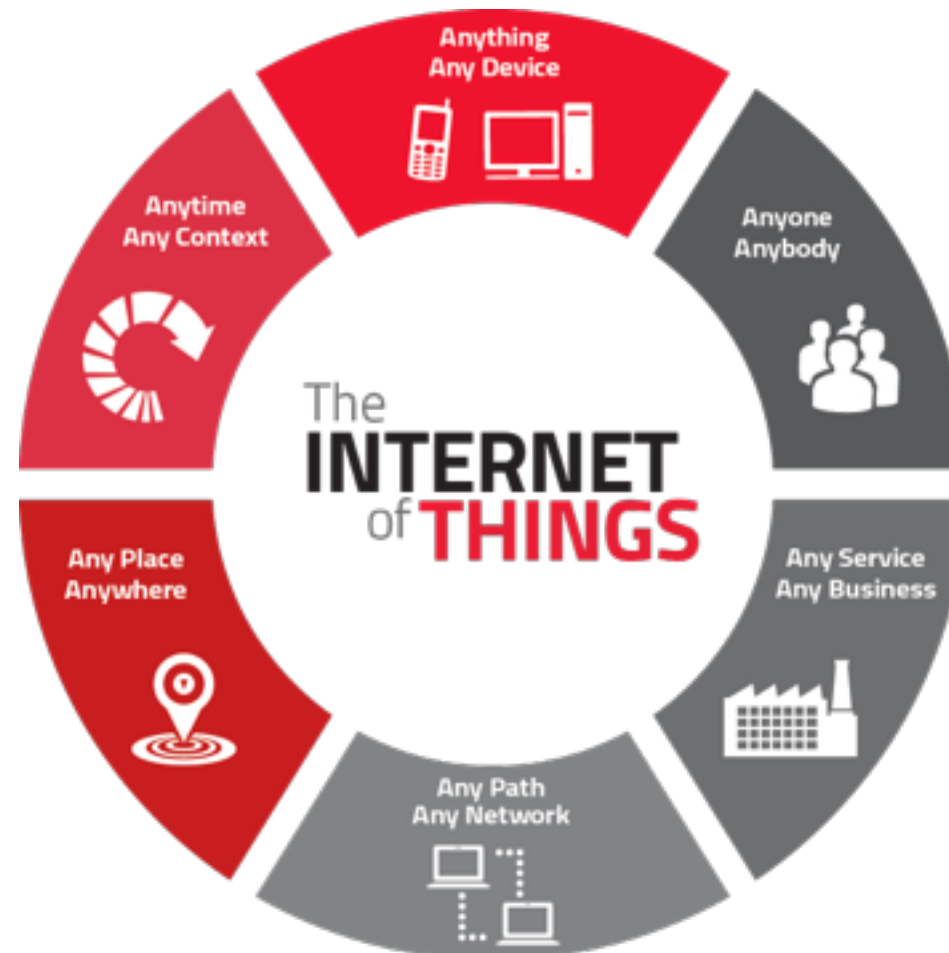
Session Labs: 9/12 and 16/12 Fridays at 9:00

Important Dates

Project Presentation: January 20

Test: February 3

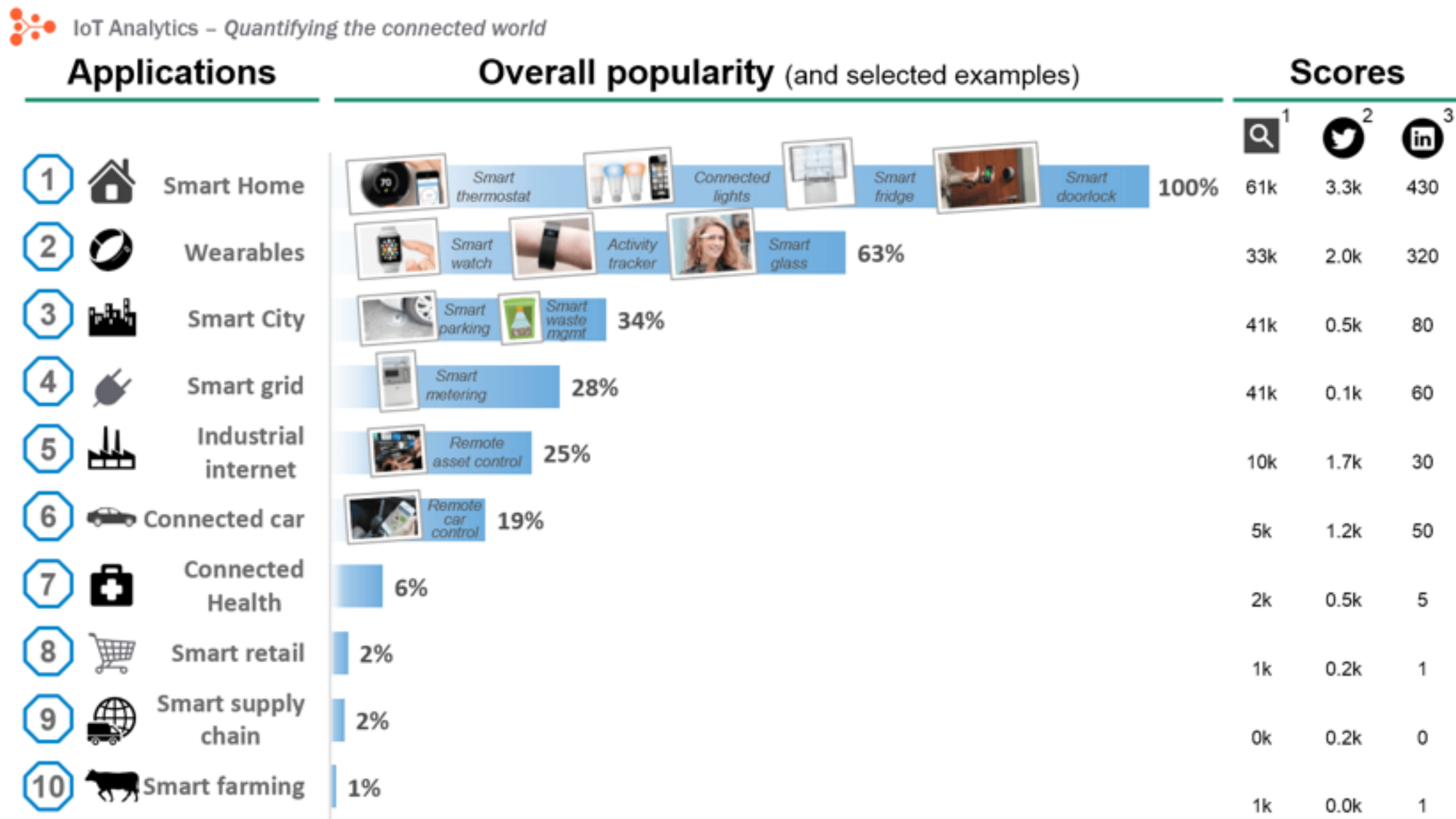
INTERNET OF THINGS



IoT: sensors and actuators connected by networks to computing systems.

- Gartner predicts 20.8 billion IoT devices by 2020.
- IDC projects 32 billion IoT devices by 2020

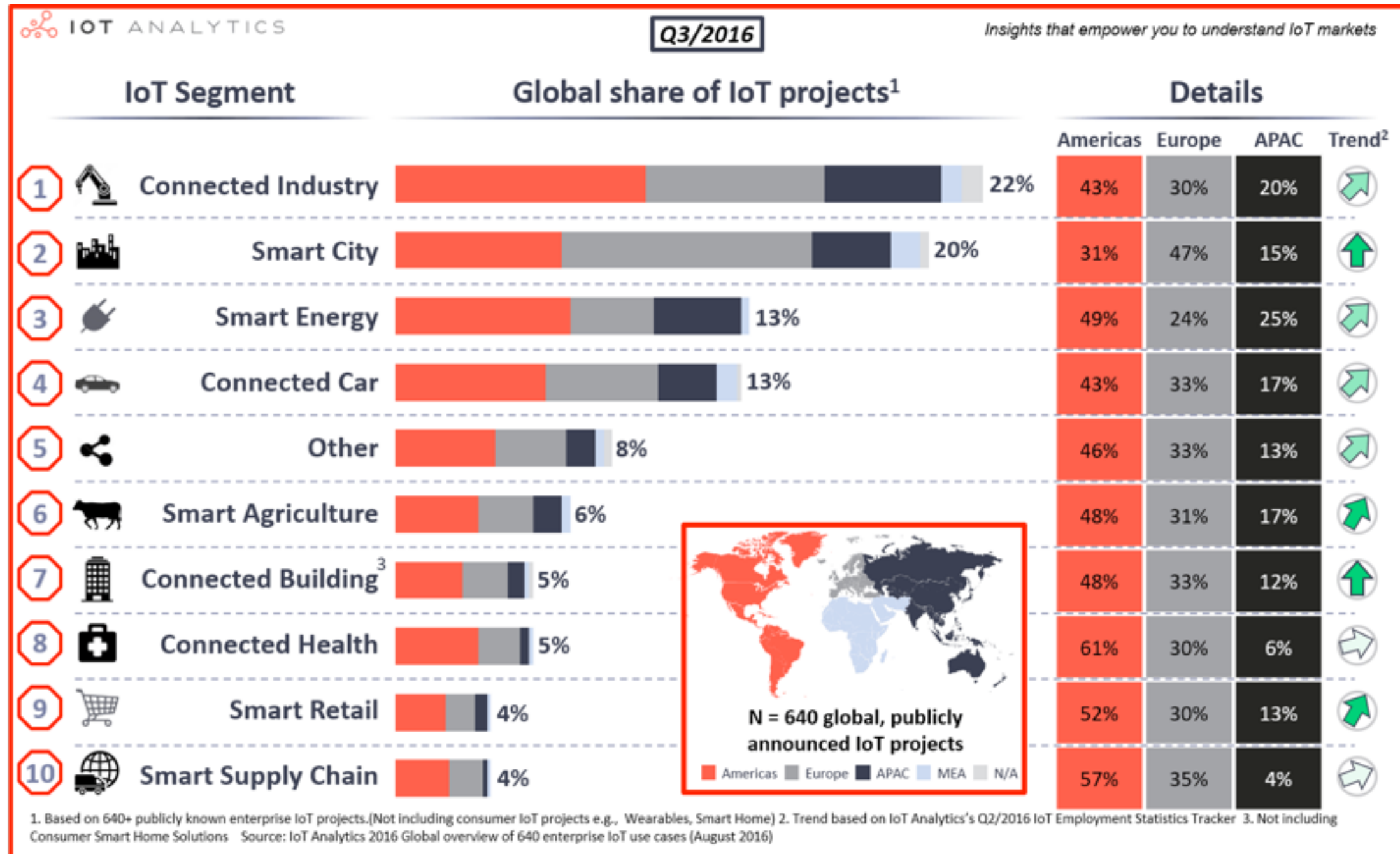
Applications IoT Analytics



1. Monthly worldwide Google searches for the application 2. Monthly Tweets containing the application name and #IOT 3. Monthly LinkedIn Posts that include the application name. All metrics valid for Q4/2014.

Sources: Google, Twitter, LinkedIn, IoT Analytics

Applications IoT Analytics

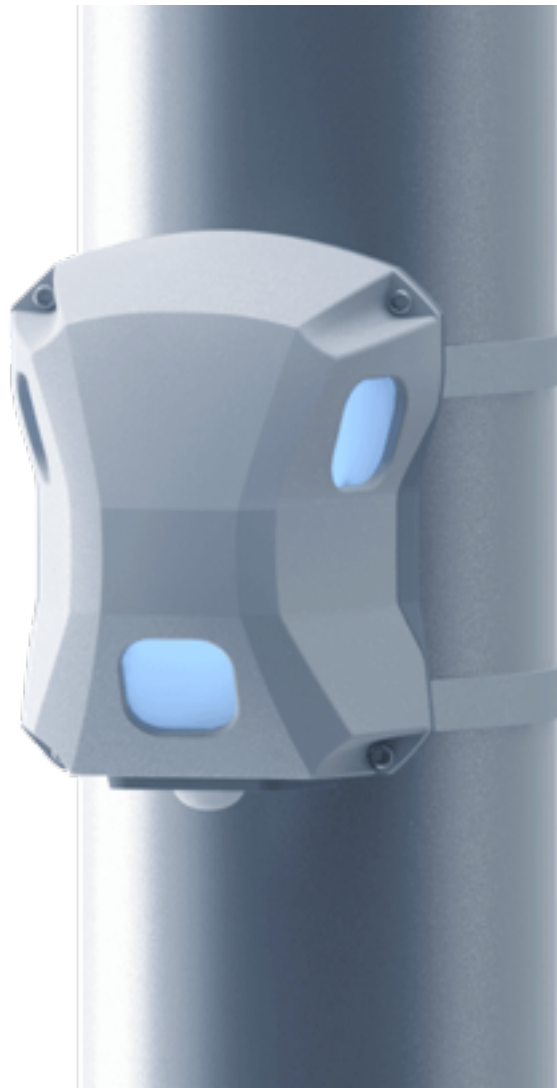




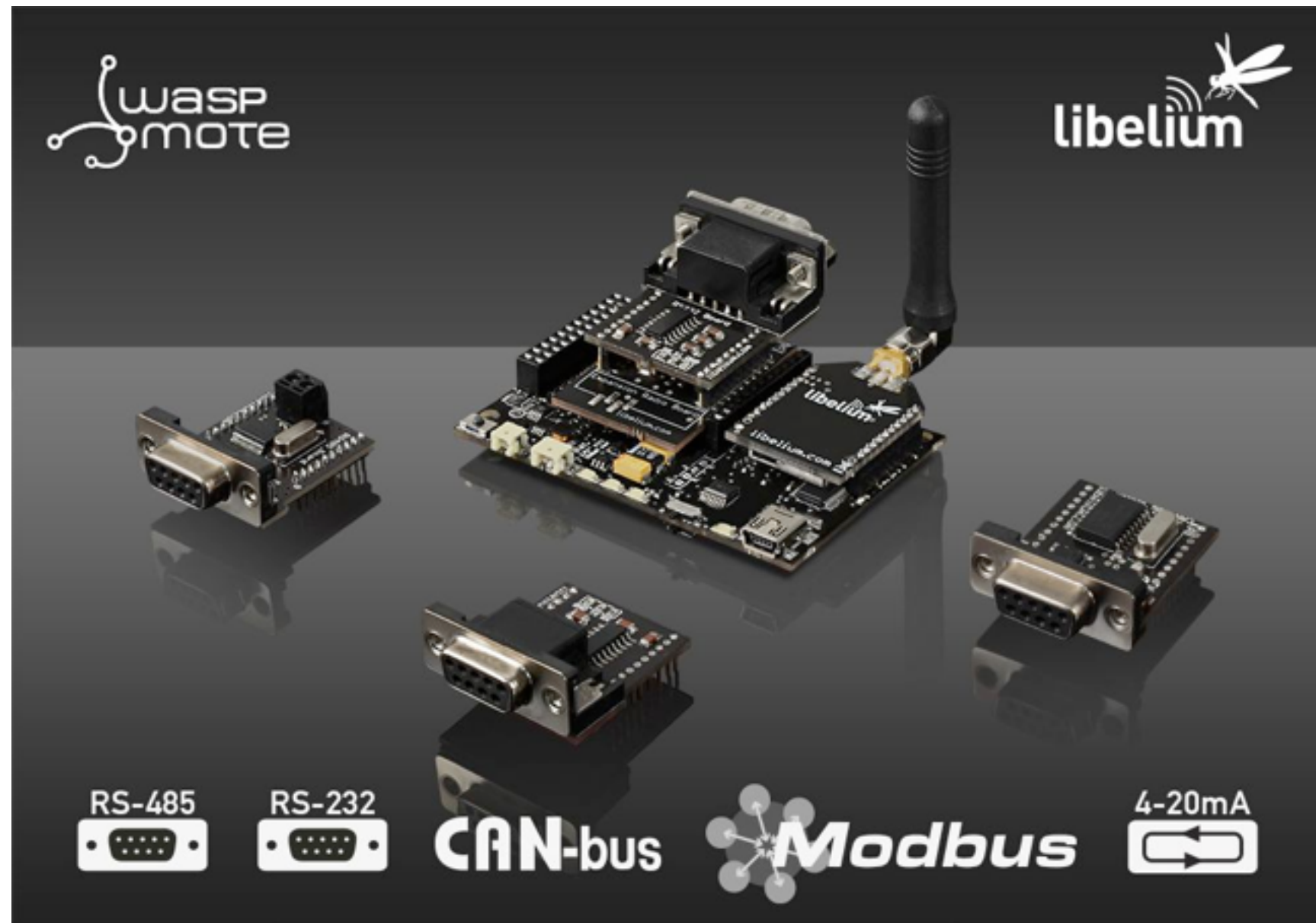
IoT Applications For Energy Management



IoT Applications For Connected/Smart Home



IoT Applications For Smart Cities



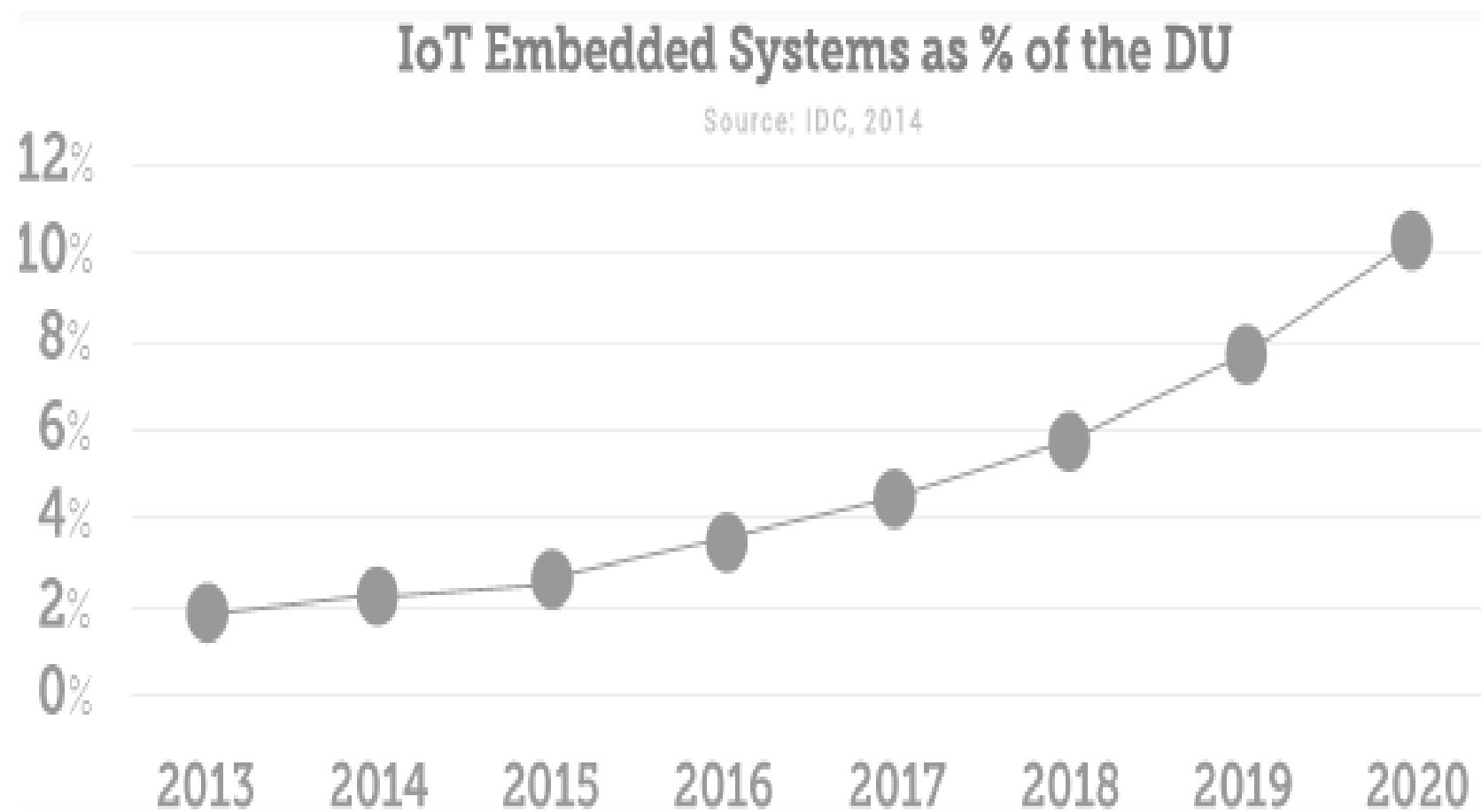
IoT Applications For Industrial Automation

IOT AND INDUSTRY 4.0



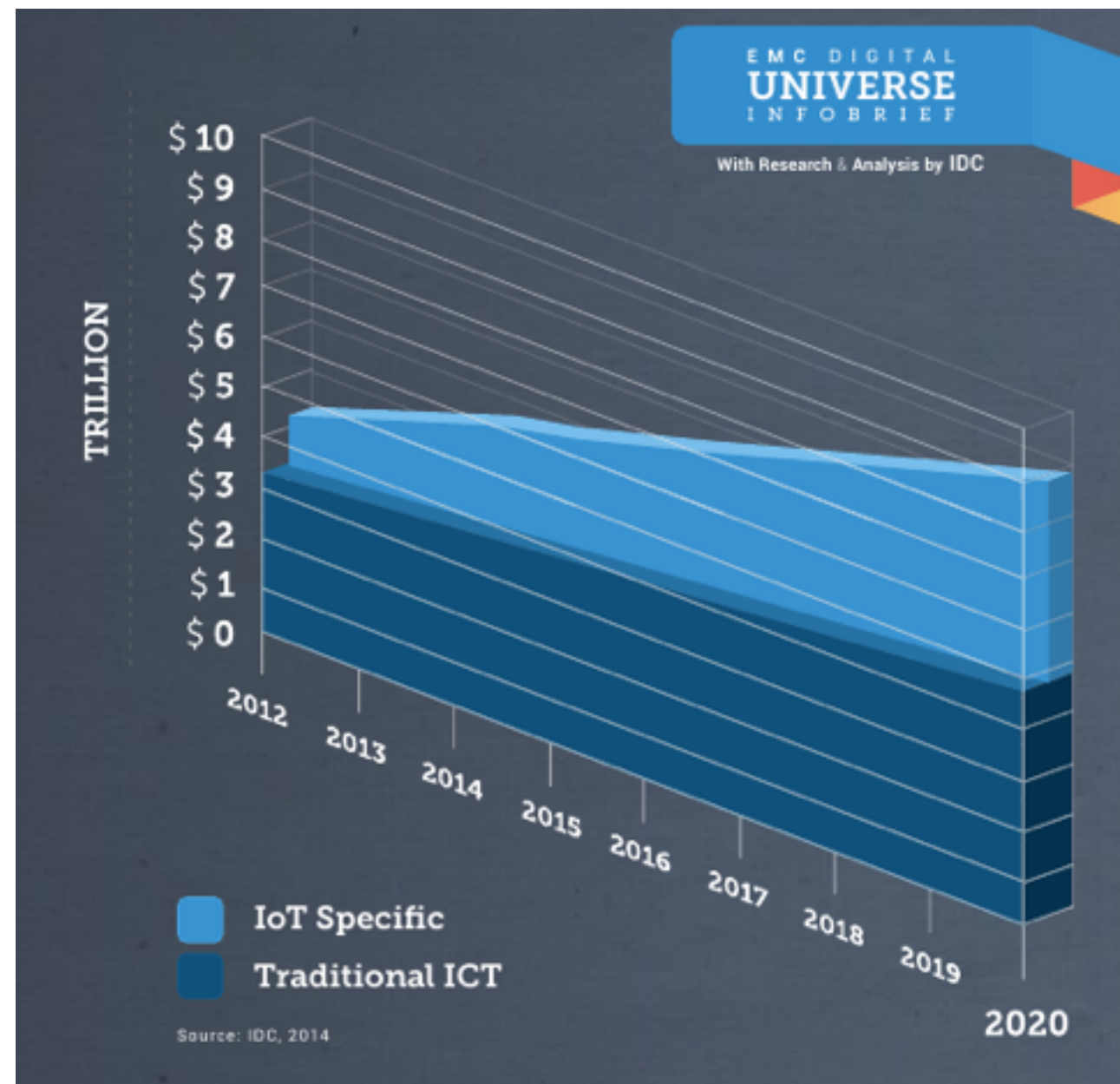
- Interoperability: IoT
- Information transparency: virtual copy of the physical world
- Technical assistance: support human decisions
- Decentralized decisions: make decisions on their own

INTERNET OF THINGS



- EMC Digital Universe, 2014

INTERNET OF THINGS



- EMC Digital Universe, 2014

IOT (MC KINSEY)

Exhibit 5

Sized applications of the Internet of Things could have direct economic impact of \$2.7 trillion to \$6.2 trillion per year in 2025



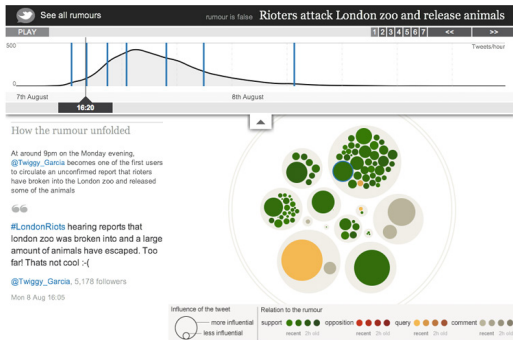
Sized applications	Potential economic impact of sized applications in 2025 \$ trillion, annually	Estimated scope in 2025	Estimated potential reach in 2025	Potential productivity or value gains in 2025
Health care	1.1–2.5	<ul style="list-style-type: none"> \$15.5 trillion cost of treating chronic diseases \$400 billion cost of counterfeit drugs, 40% addressable with sensors 50 million nurses for inpatient monitoring <ul style="list-style-type: none"> – Developed world: \$30 per hour – Developing: \$15 per hour 	<ul style="list-style-type: none"> 70–80% mobile penetration in patients who account for bulk of health-care spending Counterfeit drug tracking <ul style="list-style-type: none"> – Developed world: 50–80% – Developing world: 20–50% Inpatient monitoring <ul style="list-style-type: none"> – Developed world: 75–100% – Developing: 0–50% 	<ul style="list-style-type: none"> 10–20% cost reduction in chronic disease treatment through remote health monitoring 80–100% reduction in drug counterfeiting 0.5–1.0 hour time saved per day by nurses
Manufacturing	0.9–2.3	<ul style="list-style-type: none"> \$47 trillion in global manufacturing operating costs 	<ul style="list-style-type: none"> 80–100% of all manufacturing 	<ul style="list-style-type: none"> 2.5–5.0% saving in operating costs, including maintenance and input efficiencies
Electricity	0.2–0.5	<ul style="list-style-type: none"> 27,000–31,000 TWh global electricity consumption \$200 billion spending on transmission lines 300 billion consumer minutes outage 	<ul style="list-style-type: none"> 25–50% of consumers could adopt energy management 25–50% of grid monitored through sensors 50–100% of consumer meters automated 	<ul style="list-style-type: none"> 2–4% reduction in demand peaks in the grid Reduction of total load on grid Operating/maintenance savings; shorter outage time through automated meters
Urban infrastructure	0.1–0.3			
Security	0.1–0.2			

Streaming Data



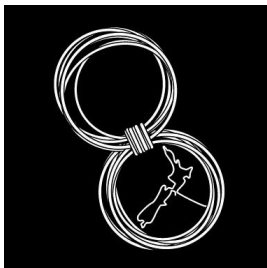
Big Data & Real Time

Real time analytics



We want to analyze what is happening **now**.

Time and Memory



Number 8 Wire Mentality

Time and memory are the resource dimensions of the process.

Time and Memory



Time and memory are the resource dimensions of the process.

Algorithms



Classification, Regression, Clustering, Frequent
Pattern Mining.

Applications

- ▶ sensor data: industry, cities
- ▶ telecomm data
- ▶ social networks: twitter, facebook, yahoo
- ▶ marketing: sales business

Data may come from: humans, sensors, or machines.



Big Data & Real Time