



**It's a streaming world**

Emanuele Della Valle

DEIB - Politecnico di Milano

<http://emanueledellavalle.org> - @manudellavalle



This work is licensed under the Creative Commons Attribution 3.0 Unported License.

## Your are free:



**to Share** — to copy, distribute and transmit the work



**to Remix** — to adapt the work

## Under the following conditions



**Attribution** — You must attribute the work by inserting "by E. Della Valle – <http://emanueledellavalle.org> - @manudellavalle" at the end of each reused slide

To view a copy of this license, visit

<http://creativecommons.org/licenses/by/3.0/>

# It's a streaming world ...

- Off-shore oil operations



- Smart Cities



- Global Contact Center



- Social networks



- Generate data streams!

E. Della Valle, S. Ceri, F. van Harmelen, D. Fensel **It's a Streaming World! Reasoning upon Rapidly Changing Information.** IEEE Intelligent Systems 24(6): 83-89 (2009)

# ... looking for reactive answers ...

- What is the expected time to failure when that turbine's barring starts to vibrate as detected in the last 10 minutes?



- Is public transportation where the people are?



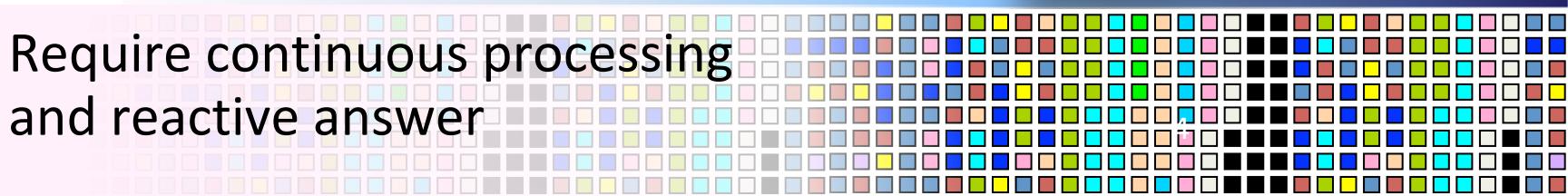
- Who are the best available agents to route all these unexpected contacts about the tariff plan launched yesterday?



- Who is driving the discussion about the top 10 emerging topics ?



- Require continuous processing and reactive answer





# ...with conflicting requirements 1/8

A system able to answer those queries must be able to

- handle **massive datasets**

- A typical oil production platform is equipped with about **400.000 sensors**



- Telecom data is the most pervasive data source in urban areas, in Milano there are **1.8 million mobile users**



- A global contact centre of a Telecom operator counts **500 millions of clients**



- Facebook alone has **1.1 billion** of active **users**



# ...with conflicting requirements 2/8

A system able to answer those queries must be able to

- process **data streams** on the fly
  - The sensors on typical oil production platform generates **10,000** observations per minute with **peaks of 100,000 o/m**



- The mobile users in Milano generates **20,000** call/sms/data connections per minute with **peaks of 80,000 c/m**



- A global contact centre receives **10,000** contacts per minute with **peaks of 30,000 c/m**



- Facebook, as of May 2013, observes **3 millions "I like"** per minute



# ...with conflicting requirements 3/8

A system able to answer those queries must be able to

- cope with **heterogeneous dataset**
  - The sensors on typical oil production have been deployed over 10 years by **10s of different producers**



- **Tens of data sources** are normally needed to make sense of an urban phenomena



- A global contact centre consists in **100s of offices** owned by different subsidiary companies **engaged yearly**



- Each social network has **its own data model, APIs, ...**



# ...with conflicting requirements 4/8

A system able to answer those queries must be able to

- cope with **incomplete data**
  - 10s of **sensors** and networking links **broke down** daily



- **Coverage is incomplete**



- Only standard cases are covered by fully machine processable data records  
**100s of contacts per minute** are managed ad-hoc



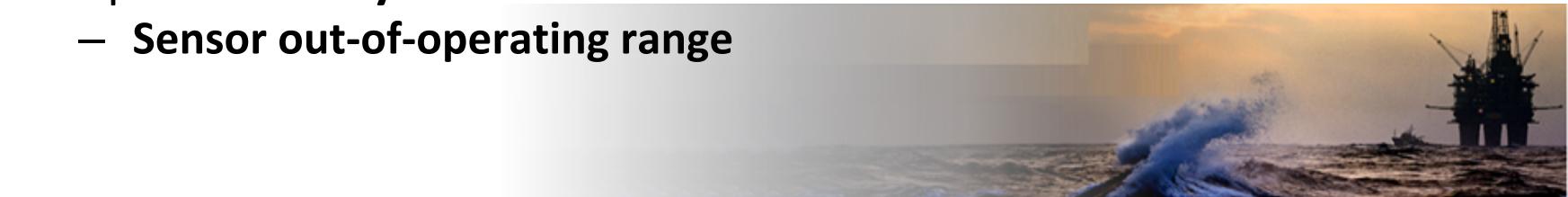
- **Conversations happen outside the social networks**, too!



# ...with conflicting requirements 5/8

A system able to answer those queries must be able to

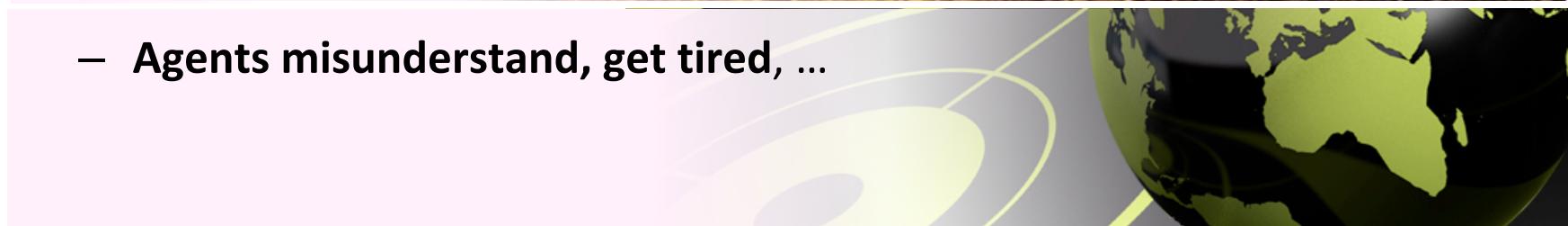
- cope with **noisy data**
  - Sensor out-of-operating range



- Faulty sensors



- Agents misunderstand, get tired, ...



- Irony, sarcasm, ...



# ...with conflicting requirements 6/8

A system able to answer those queries must be able to

- provide **reactive answers**
  - detection of dangerous situations must occur within **minutes**

- 
- recommendations to citizens must be performed in **few seconds**

- 
- routing a contact through each step of the decision tree must take less than a **second**

- 
- Search autocompleting may need to be updated every **few minutes**



10

# ...with conflicting requirements 7/8

A system able to answer those queries must be able to

- support **fine-grained information access**
  - Identify a **turbine** among thousands



- Locate a **bus** among thousands



- Contact an **agent** among thousands



- Identify an **opinion maker** among thousands of influencers for a topic



# ...with conflicting requirements 8/8

A system able to answer those queries must be able to

- integrate **complex domain models** of
  - operational and control process



- various **city aspects**



- contact management, **contract types**,  
**agent skills**, **contactor profiles**, ...



- **topics**, **user profiles**, ...



12



# Challenges

A system able to answer those queries must be able to

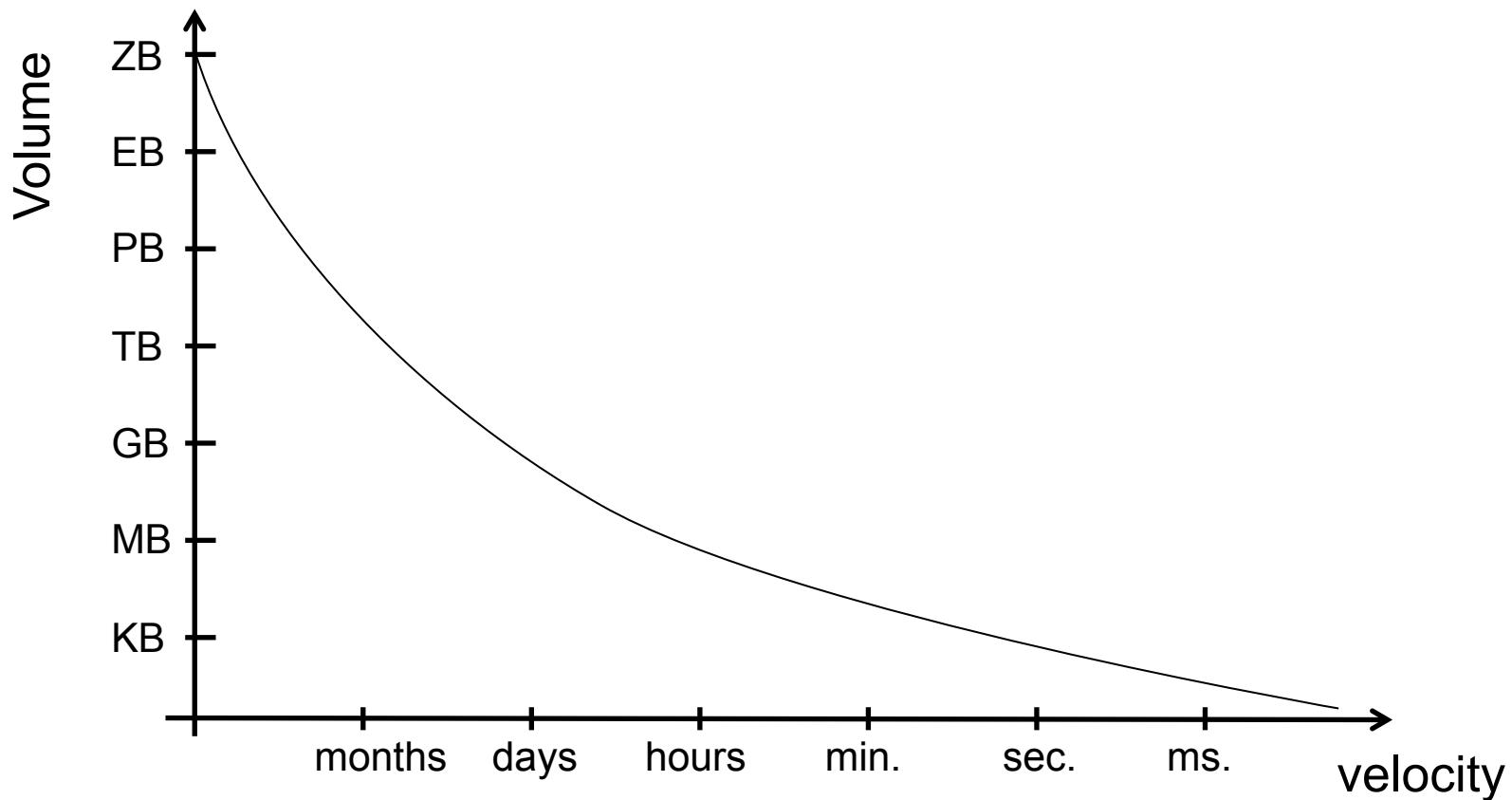
- handle **massive datasets**
- process **data streams** on the fly
- cope with **heterogeneous datasets**
- cope with **incomplete data**
- cope with **noisy data**
- provide **reactive answers**
- support **fine-grained access**
- integrate **complex domain models**

In **Big Data** terms →

X			
	X		
		X	
		X	X
			X
		X	
X	X		
		X	
Volume	Velocity	Variety	Veracity

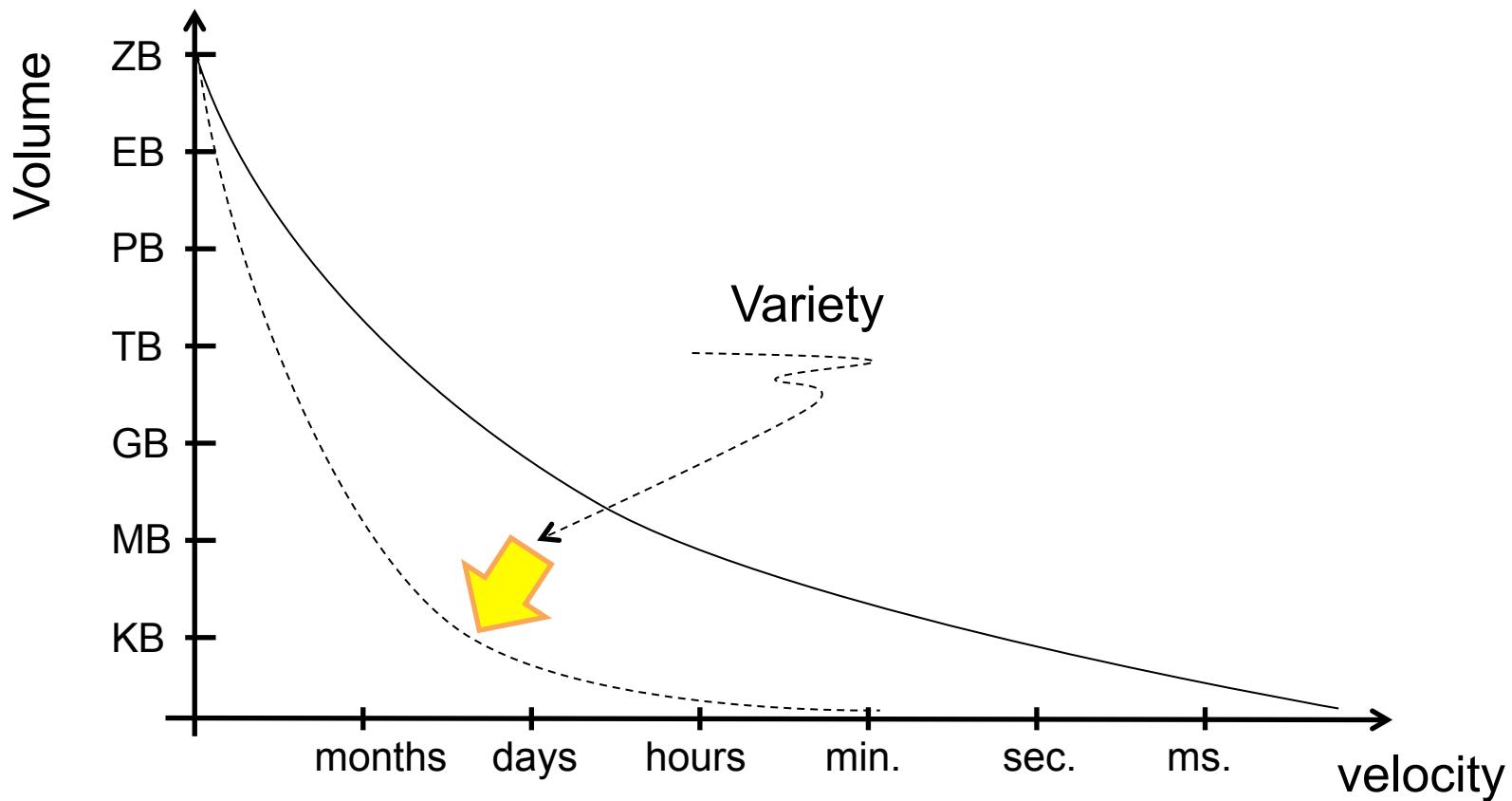
# Grand challenge

- Volume + Velocity = hard deal



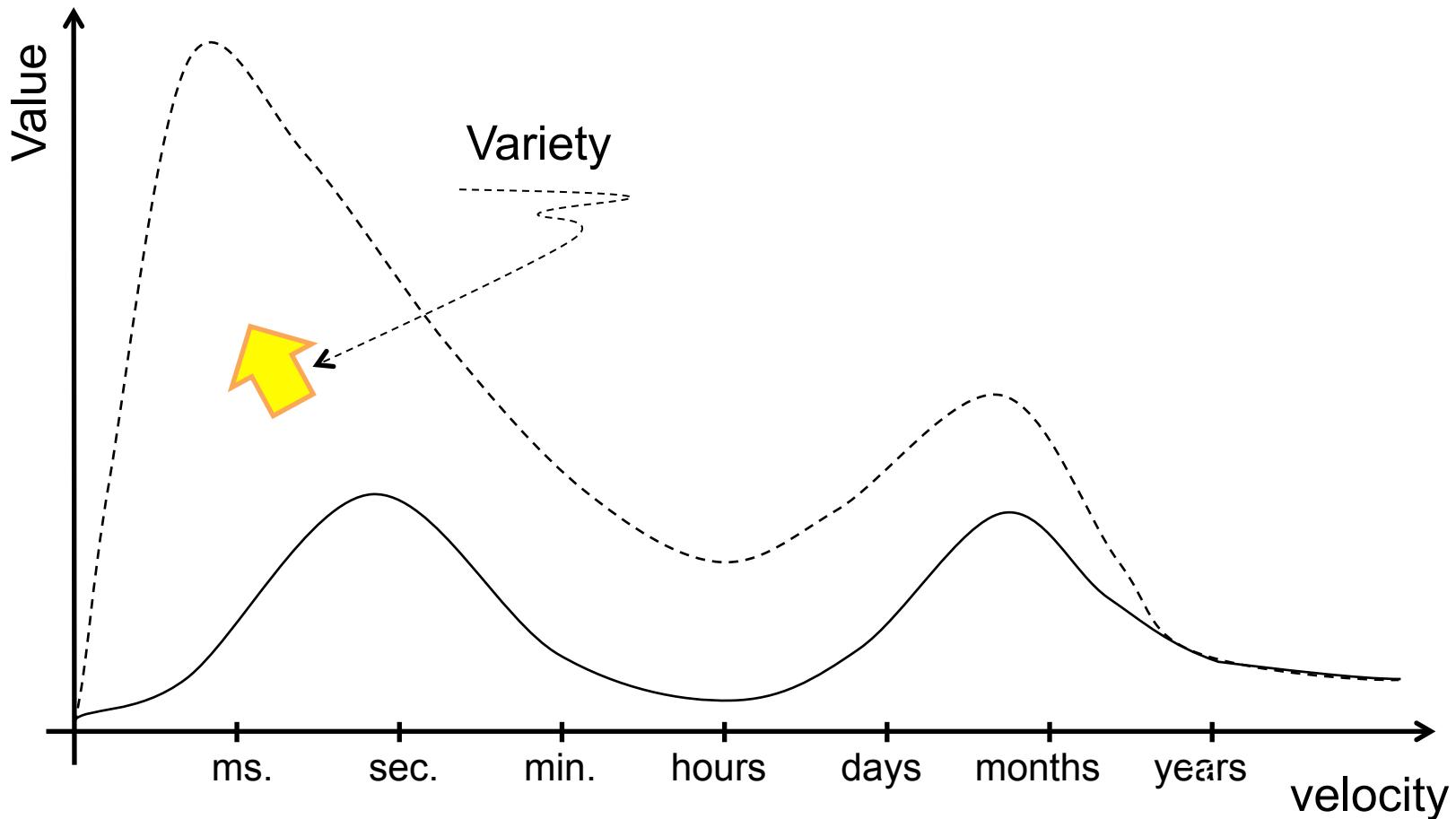
# Grand challenge ++

- Volume + Velocity + Variety = extremely hard deal



# A reason to embrace velocity+variety!

- ++ Variety → ++ value (exp. with velocity)





**It's a streaming world**

Emanuele Della Valle

DEIB - Politecnico di Milano

<http://emanueledellavalle.org> - @manudellavalle