



Roadmap for European Technologies
in Hardware and Networking for Big Data

The road to the Roadmap

Adrián Cristal

(Barcelona Supercomputing Center)

Madrid, 18 Sep 2014

www.rethinkbig-project.eu



Why are we here?

🗣️ Why a Roadmap for HW and Networking for Big Data?

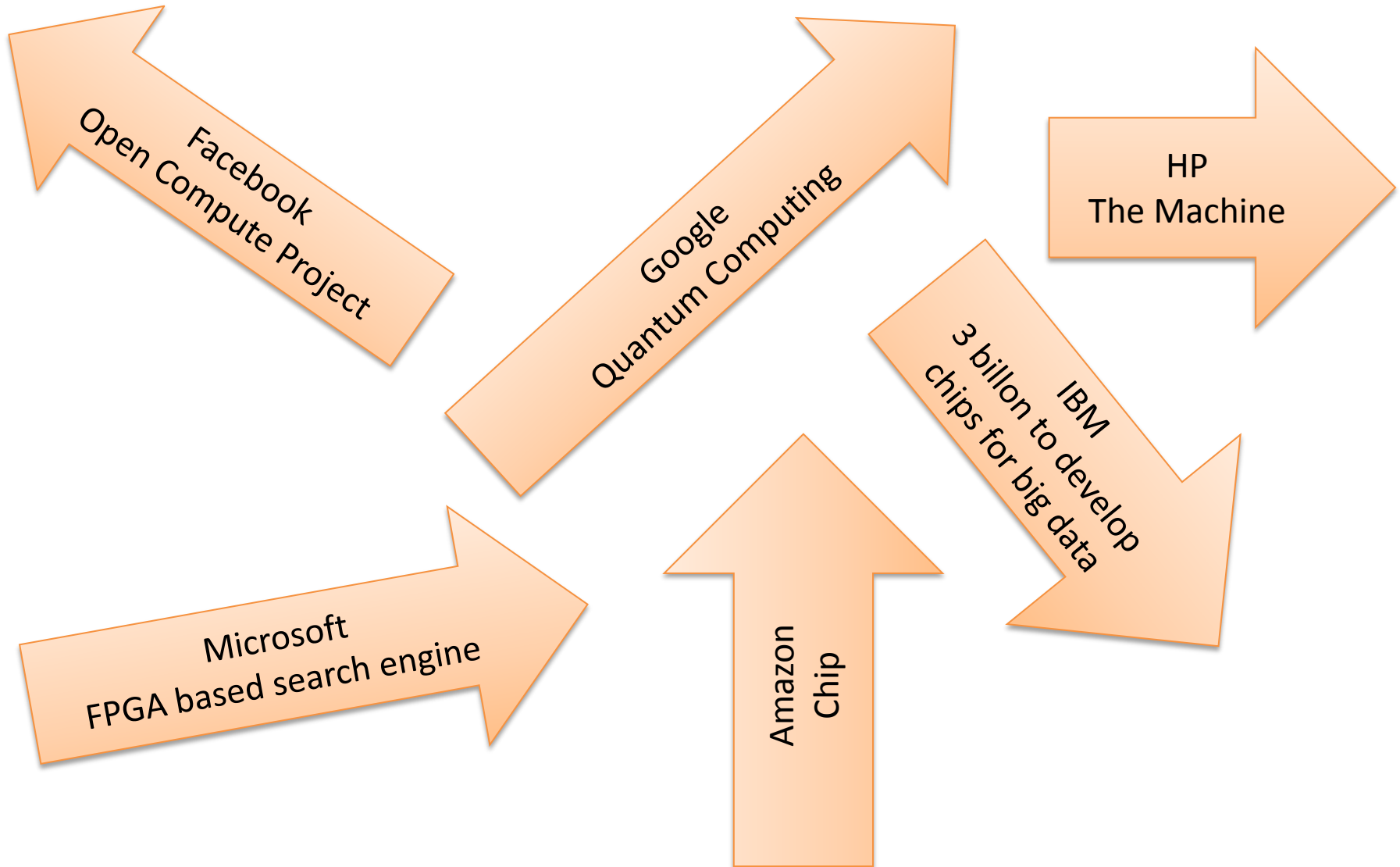
🗣️ We share a common goal:

- Make the EU a leader in the area of Big Data
- Make Big Data one of the key drivers of European Economy
- Improve European quality of life
- Make us all lots of €€€!

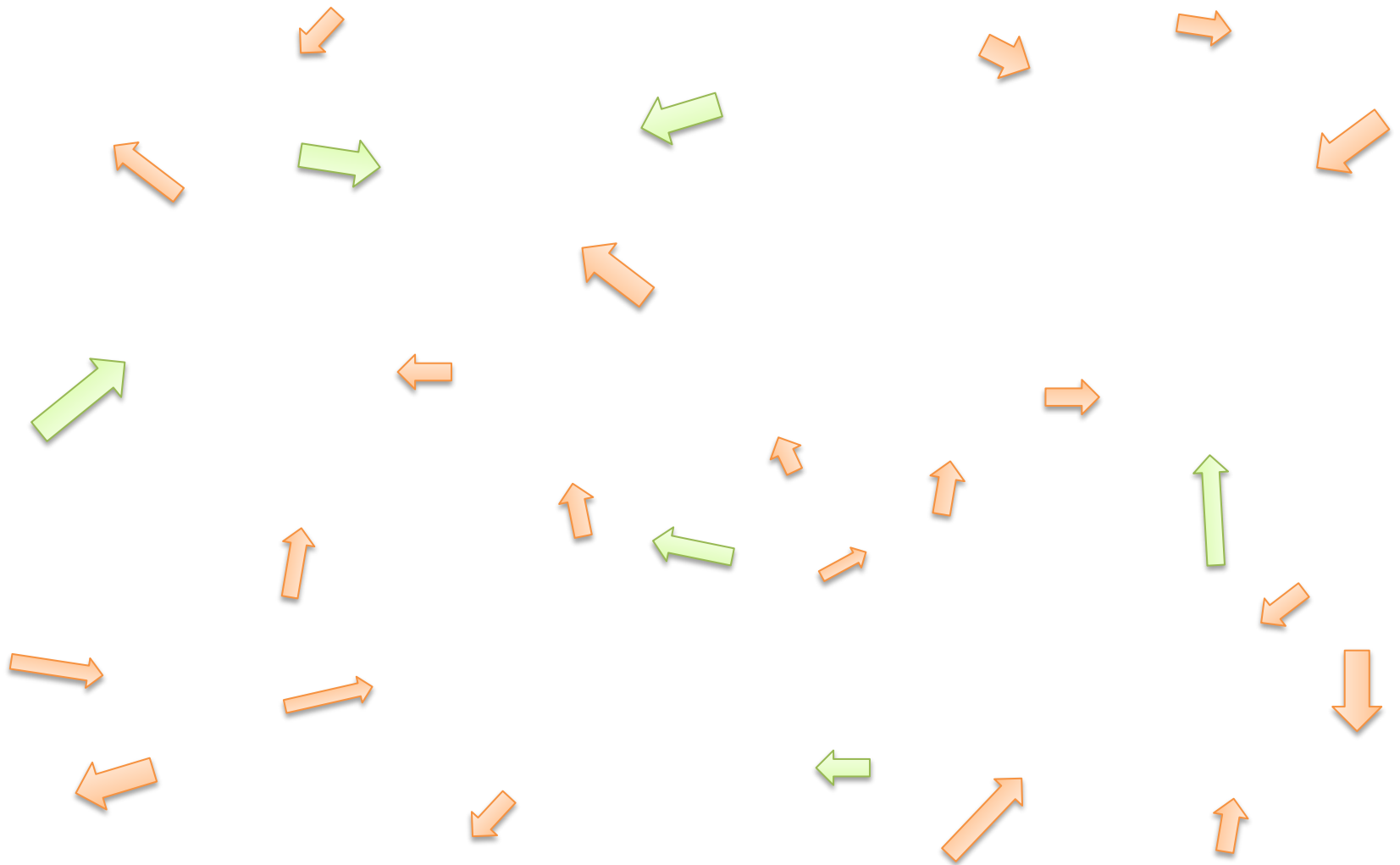
What is a Roadmap?

- 🗨️ **Roadmap: High level plan to achieve one or more goals under a set of conditions**
- 🗨️ **A roadmap that would be "irrational" not to follow...**
- 🗨️ **Strategy: (game theory) a complete algorithm for playing and winning the game, taking into consideration the moves of others**

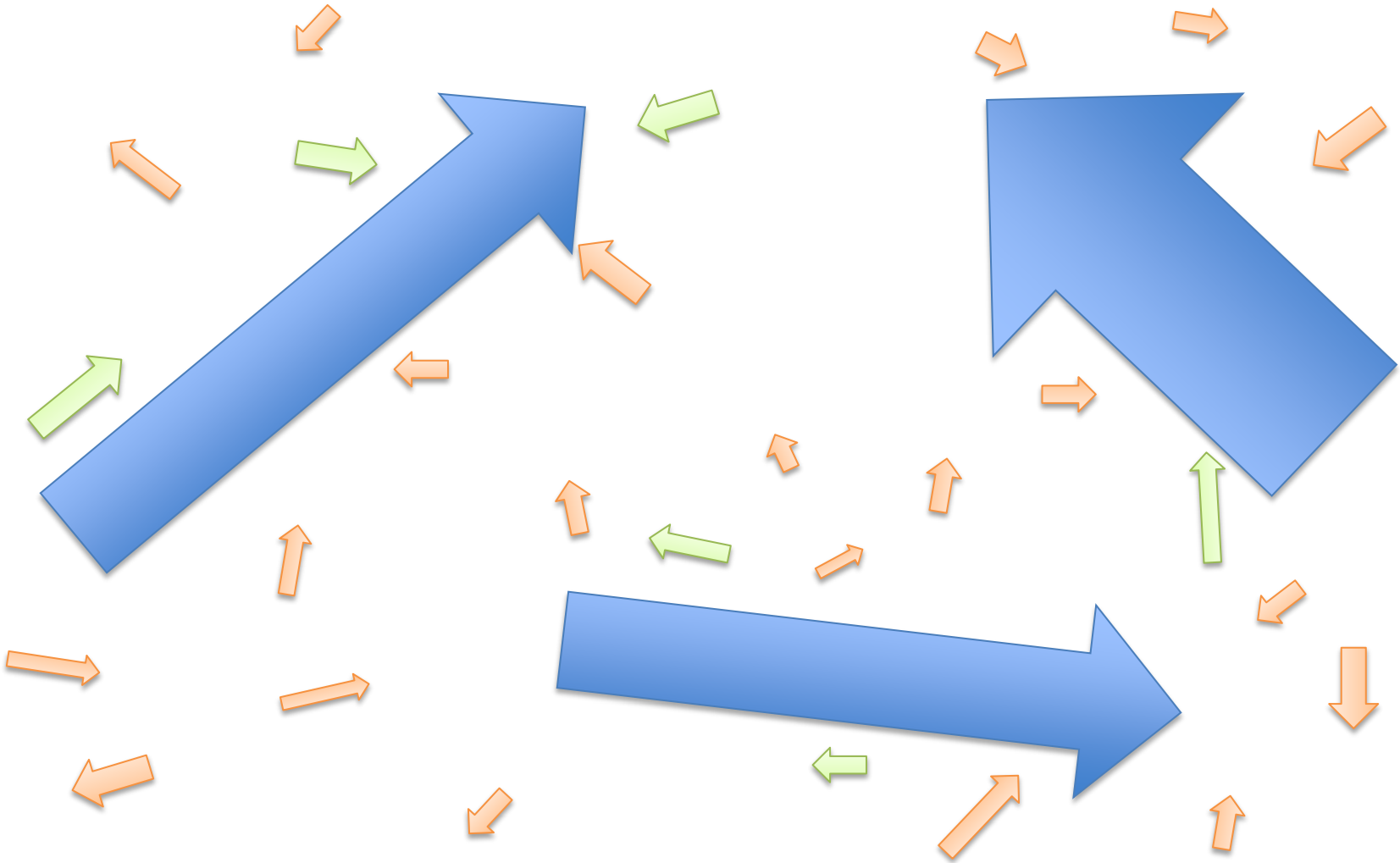
U.S. Big Data Vision



EU Vision



EU Strategy



EU H2020 Framework

- 🕒 Contractual Public Private Partnership (cPPP) on Big Data
- 🕒 An association that regulates EC spending for Big Data-related activities
- 🕒 EU Big Data Value Ecosystem
 - BIG Project
 - Byte Project
 - NESSI
 - RETHINK big Project

Why hardware matters

🕒 Terasort contest: sorting 100TB data

- Number 1: Hadoop
 - 2100 nodes, 12 cores per node, 64 Gb per node
 - ❖ 24.000 cores
 - ❖ 134 Tb memory
 - Time: 4300 secs
 - Cost in Amazon: \$ 8.800
- Number 2: Tritonsort
 - 52 nodes, 8 cores per node, 24 Gb
 - ❖ 416 cores
 - ❖ 1,2 Tb memory
 - Time: 8300 secs and 6400 secs
 - Cost in Amazon: \$ 294 and 226
- Hadoop (2013) is easy to program, but needs 57X more cores, 100X more memory, and only gets 2X performance

Meeting Objectives

- 🗣️ **To establish a common language, bringing a gap between SW, HW and Networking experts**
- 🗣️ **To understand the potential for new technologies to solve the problems of today and tomorrow**
- 🗣️ **To develop a clear understanding of common problems in Big Data**