Cefriel
POLITECNICO DI MILANO

BIG DATA

EMANUELE DELLA VALLE

18 MAGGIO 2018

Horizontal vs Vertical Scalability

Emanuele Della Valle
@manudellavalle
Prof. @polimi & Founder @fluxedo_

Introduction

- "Traditional" SQL system scale vertically:
  - Adding data to a "traditional" SQL system may degrade its performances
  - When the machine, where the SQL system runs, no longer performs as required, the solution is to buy a better machine (with more RAM, more cores and more disk)
- Big Data solutions scale horizontally
  - Adding data to a Big Data solution may degrade its performances
  - When the machines, where the big data solution runs, no longer performs as required, the solution is to add another machine

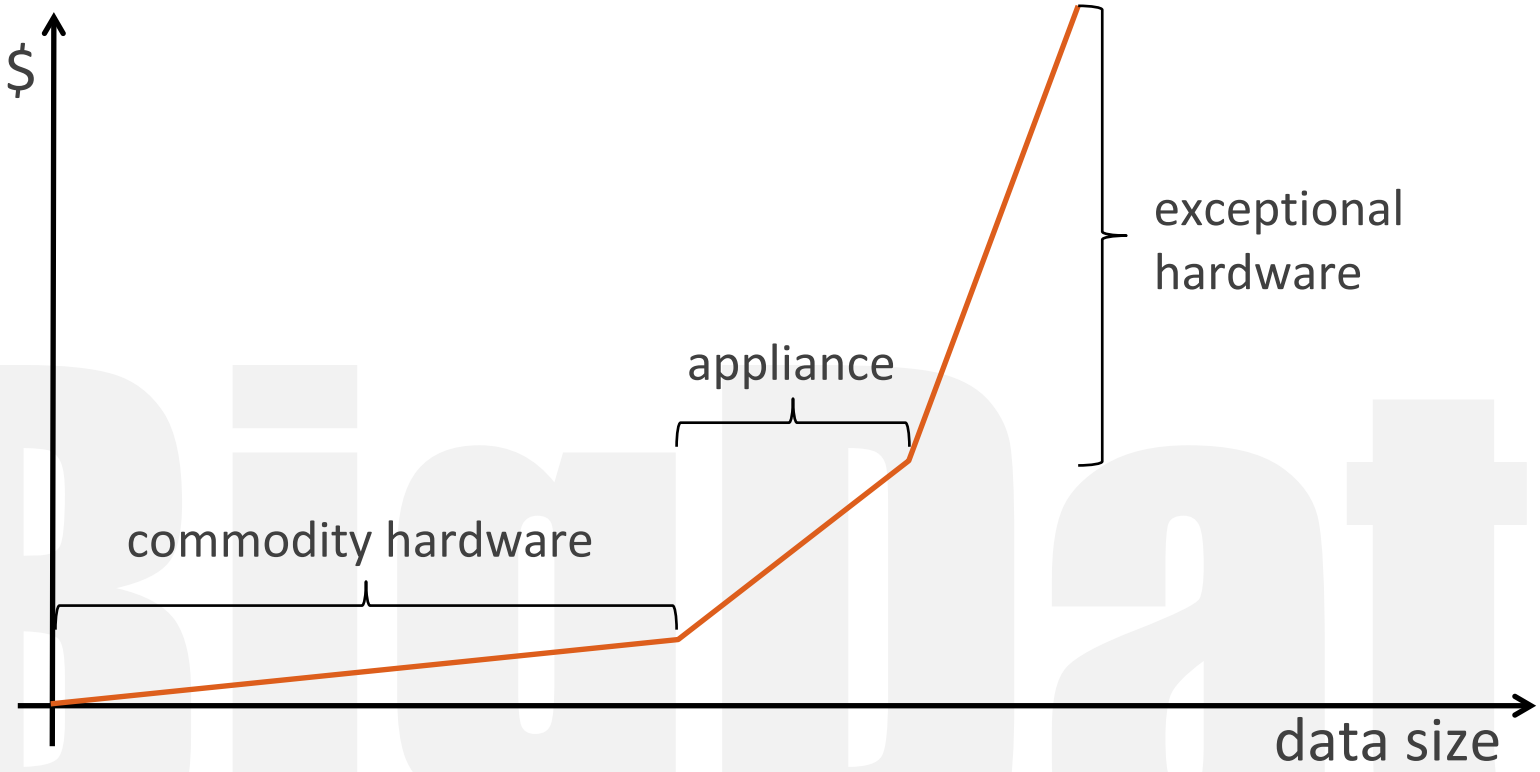
Commodity hardware

- CPU: 8-32 cores
- RAM: 16-64 GB
- Disk: 1-3 TB
- Network: 10 GE

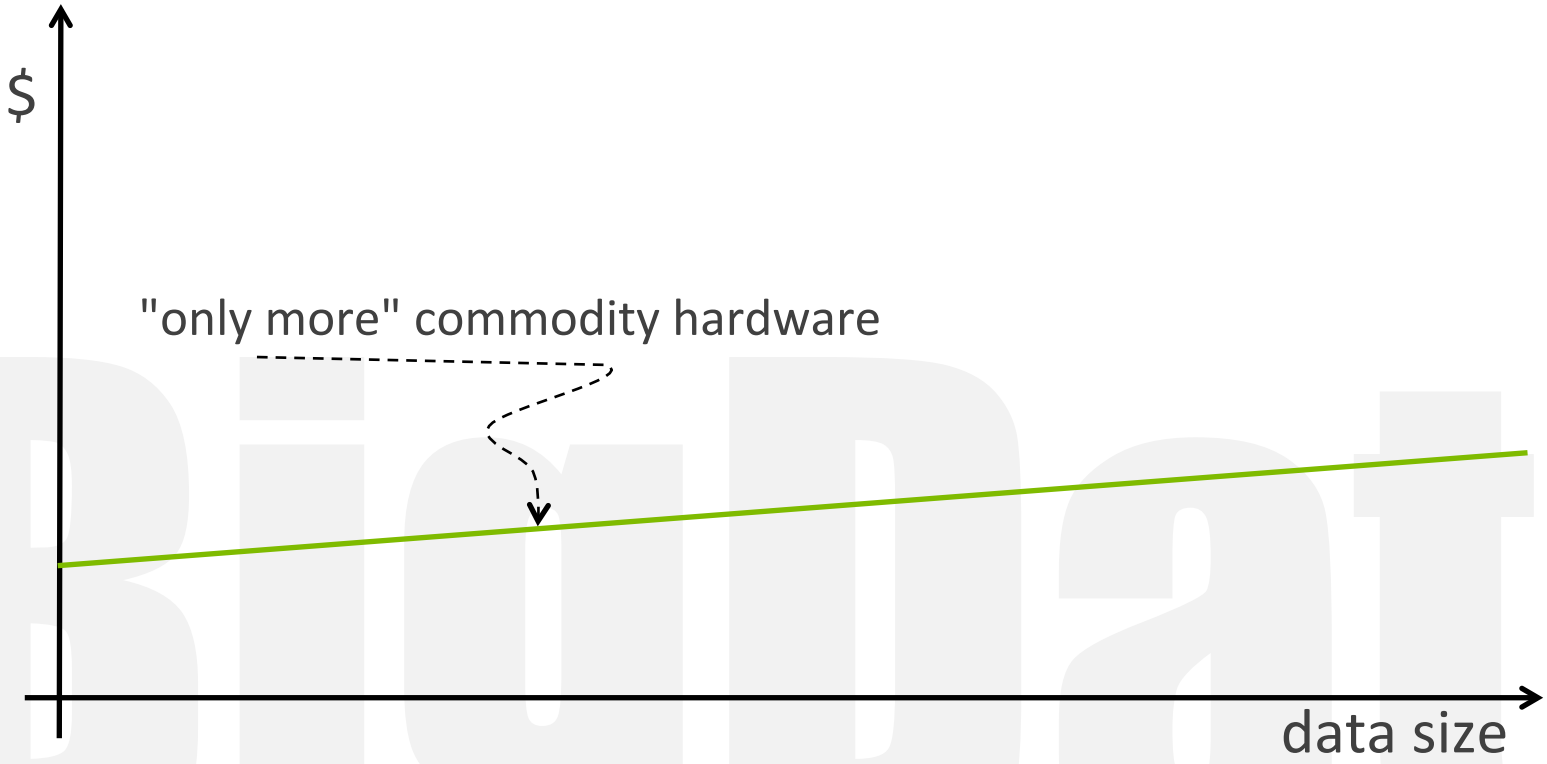
Appliance

- e.g. ORACLE EXADATA DATABASE MACHINE X6-8
- CPU: 576 cores
- RAM: 24TB
- Disk: 360TB of Flash Storage per rack
- Network: 40 Gb/second InfiniBand

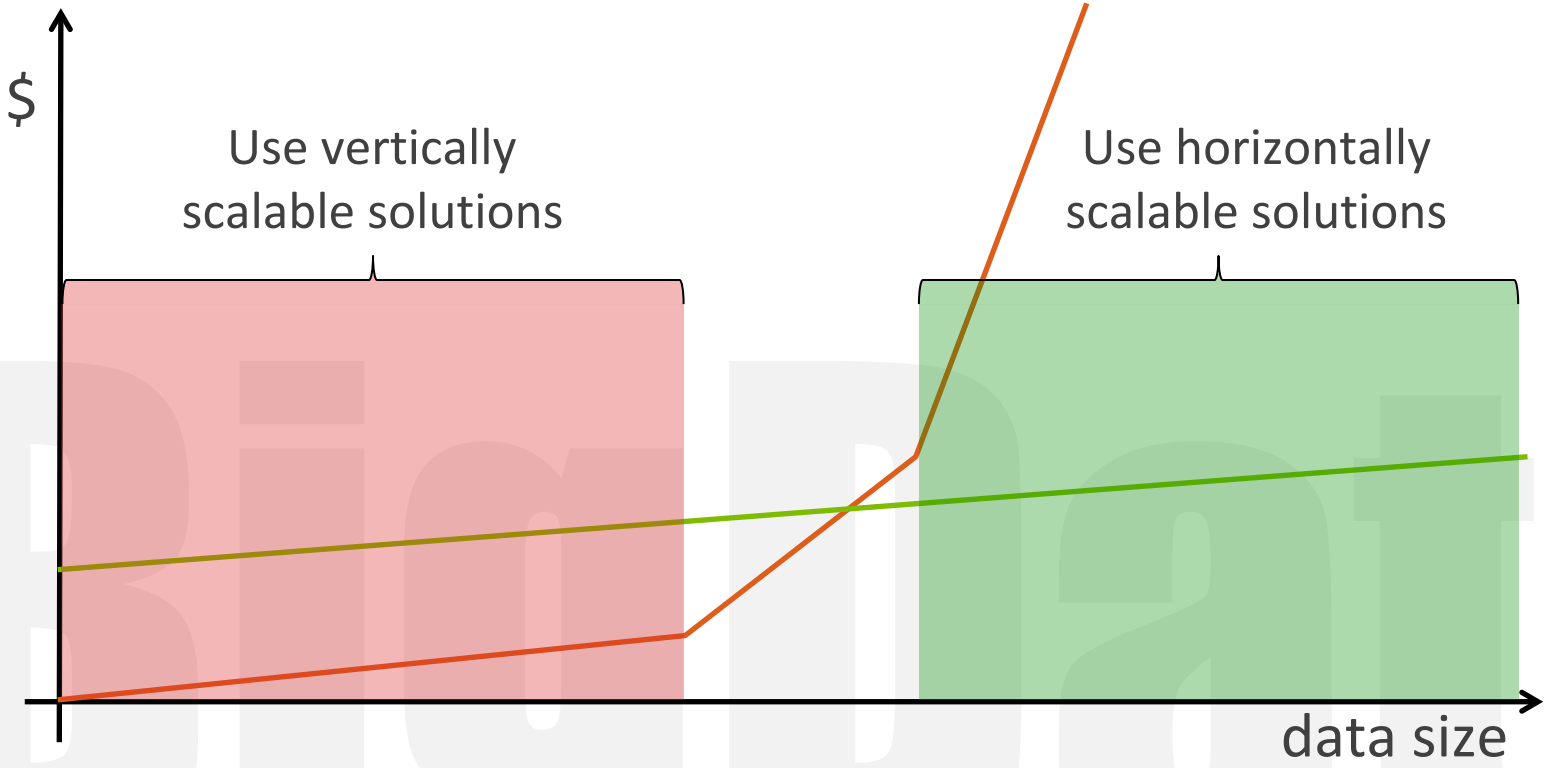
Vertical scalability



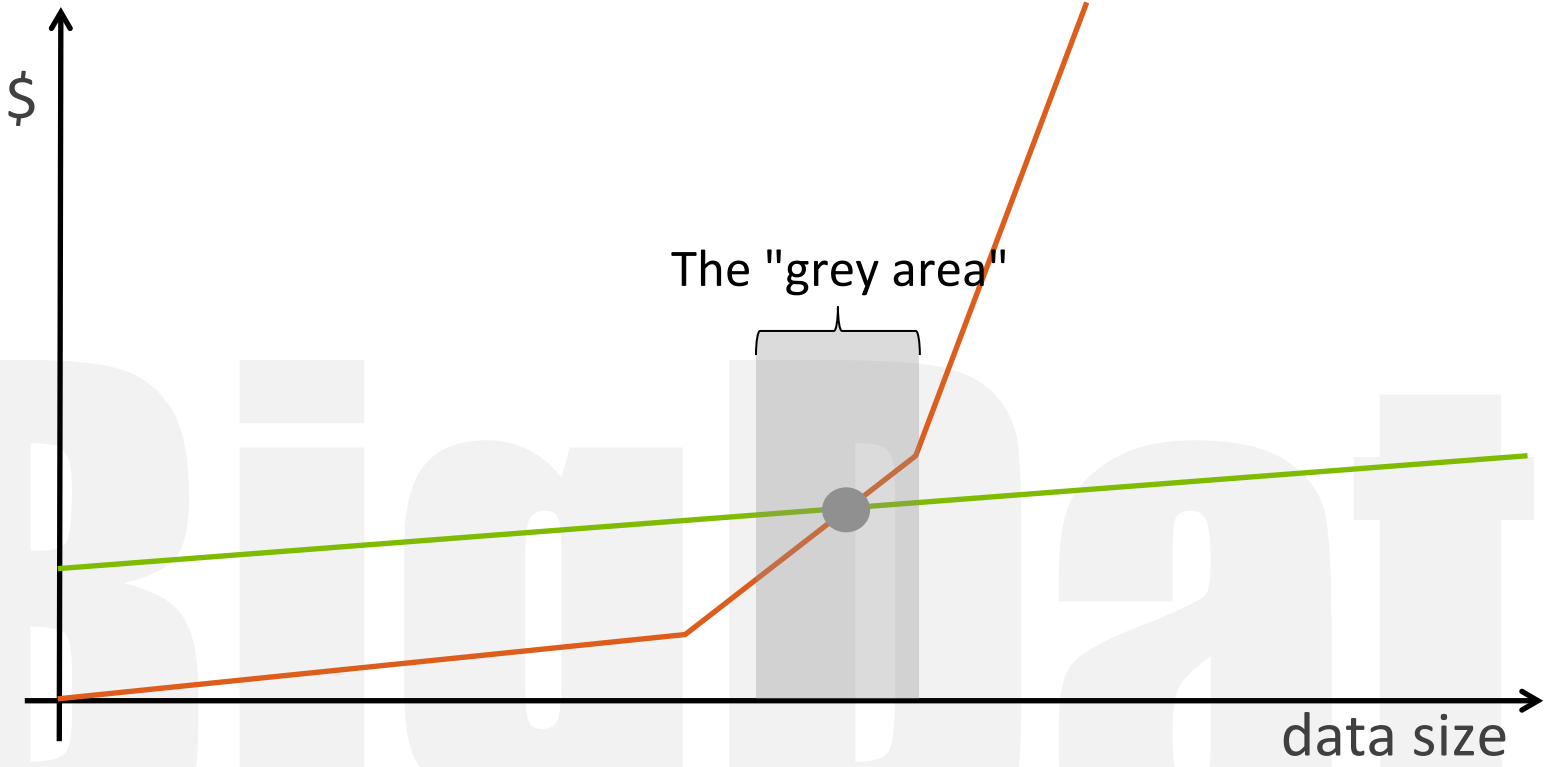
Horizontal scalability



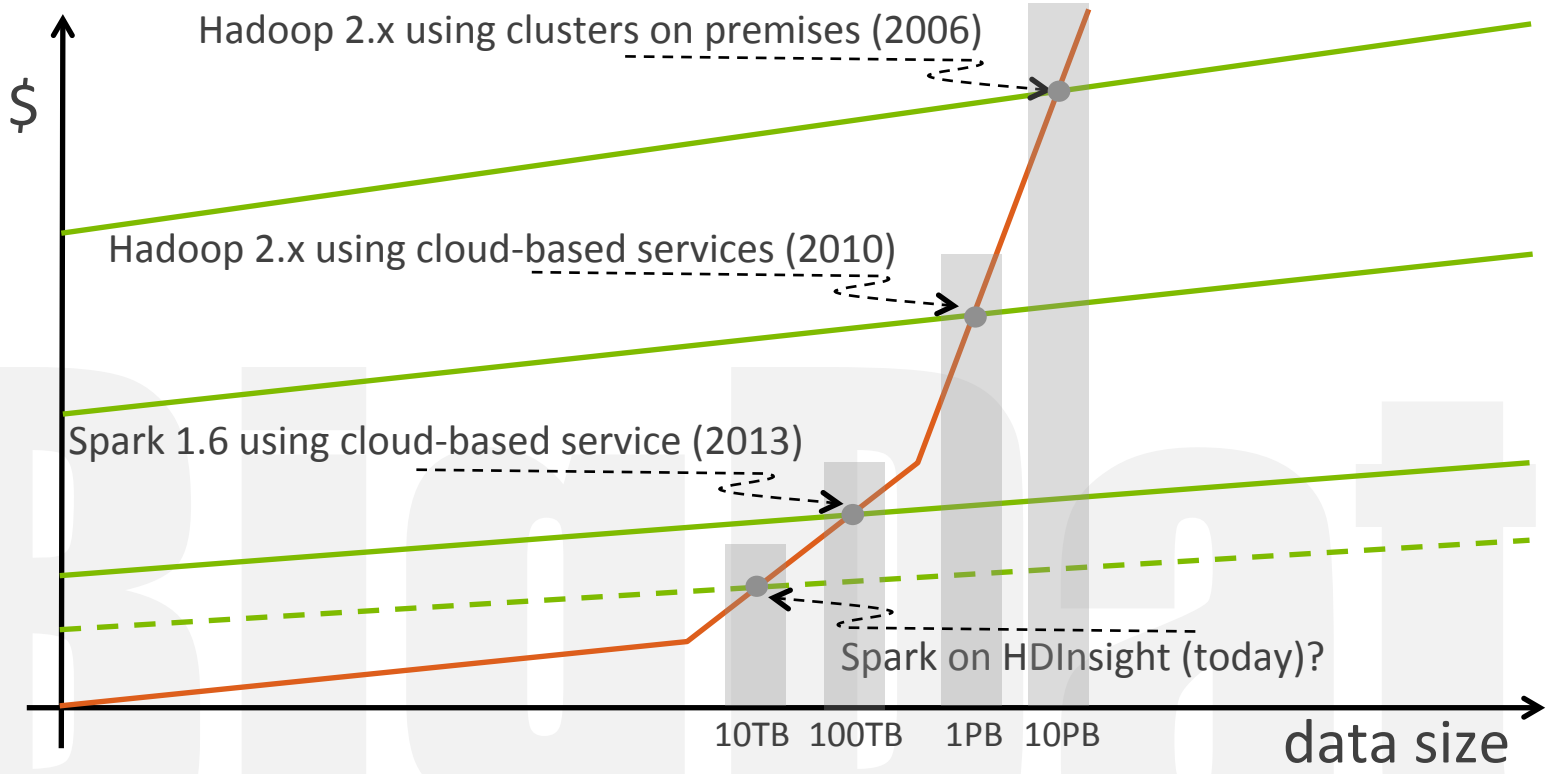
Vertical vs. Horizontal scalability



Vertical vs. Horizontal scalability



The "grey area" is time dependent



The "grey area" in the spark 2.x era (2017)

- there is not comprehensive study or answers in the literature and in the technical world at the moment
- we can determine it continuing the benchmarking work that we started