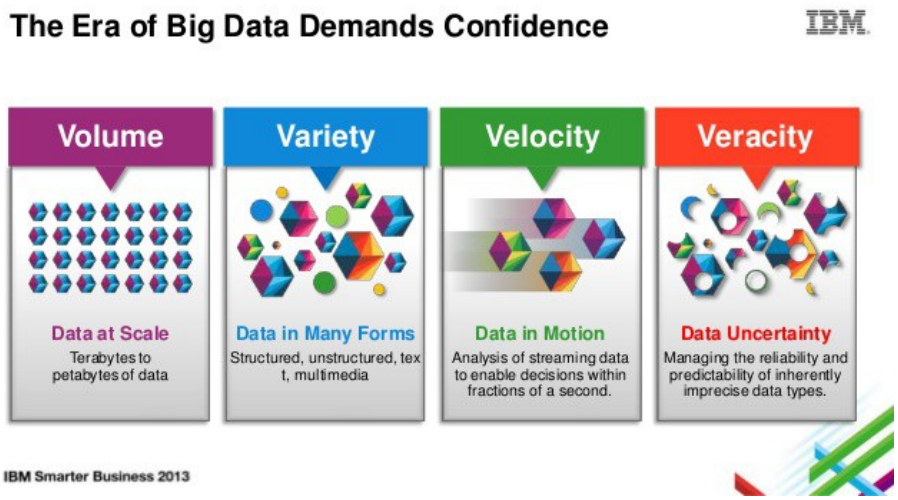
Towards Data-driven Software-Defined Infrastructures

Pedro Garcia Lopez
Universitat Rovira I Virgili

Table of Contents

Motivation

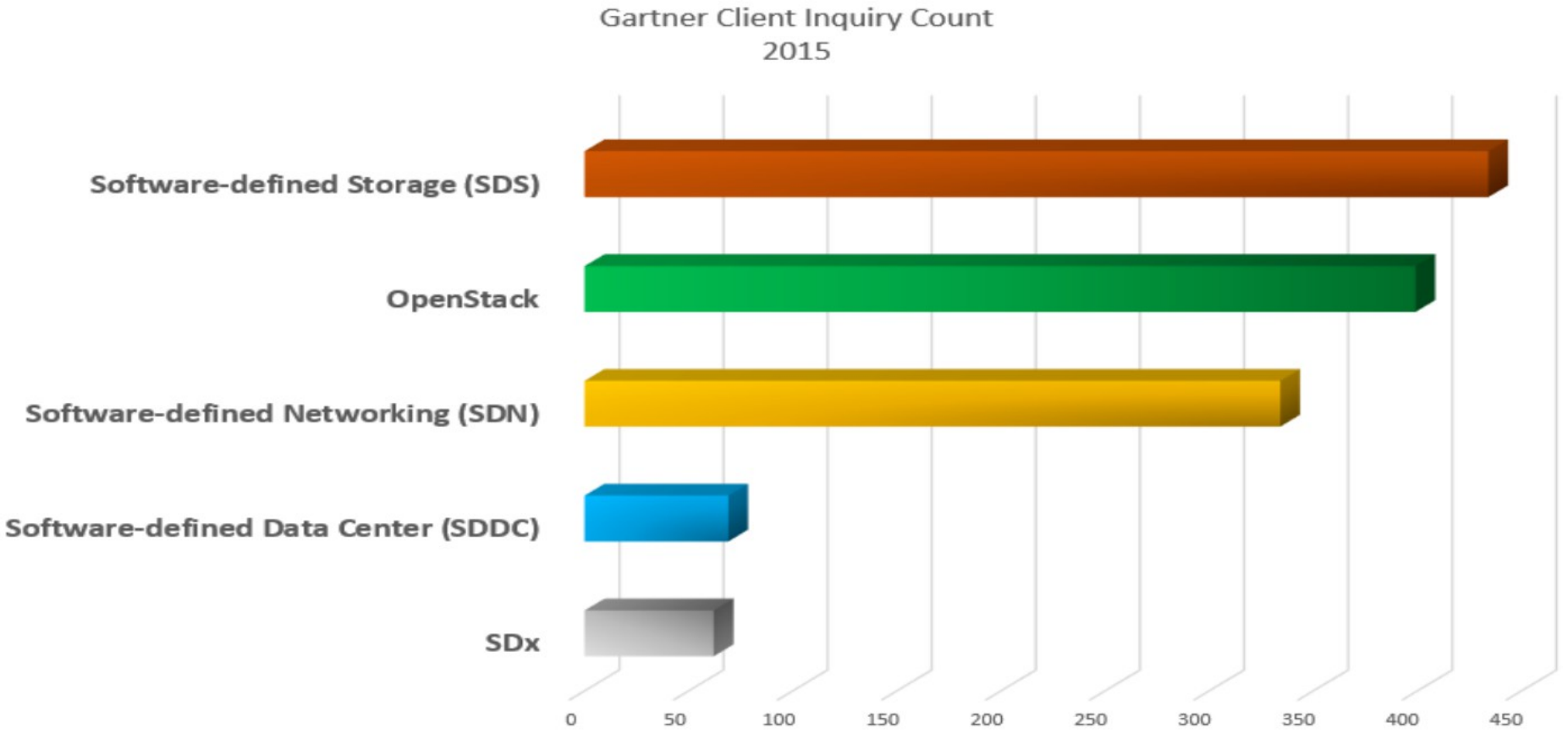


The main premise of the IOStack project is that **Software-defined Storage (SDS)** is a key enabler that can drastically reduce the costs of Big Data management in the Cloud.

Motivation

Gartner®

2015 Gartner Client Inquiry Volume – By Topic

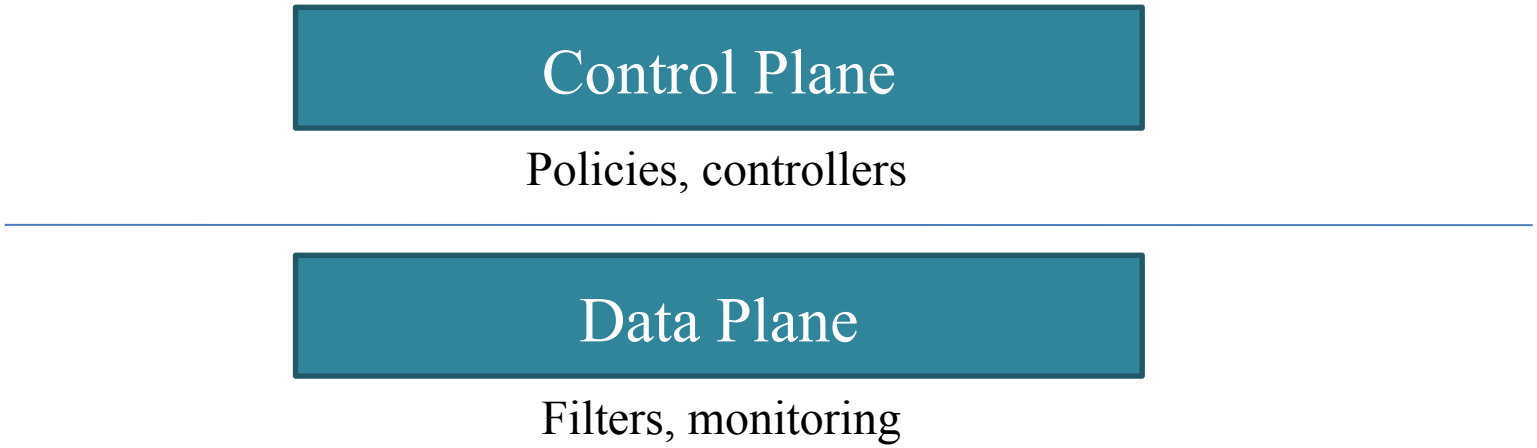


What is Software-Defined Storage?

Hey! First, what does Software-Defined Storage **mean**?

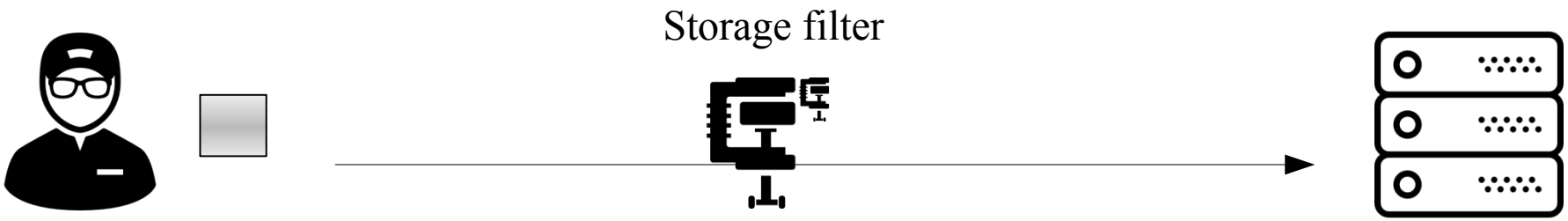
> "***Software-defined storage*** *(SDS) is a new term for computer data* ***storage software*** *to manage* ***policy-based provisioning and management*** *of data storage* ***independent*** *of the underlying hardware.*"
>
> (Wikipedia)

- A basic **principle of IOStack**: decouple control/data planes



Design Entities of IOStack

- **Filter**: Data transformation executed on data flows.
  - E.g., compression, caching, encryption, bandwidth allocation…



- **Monitoring metric**: Information of a particular aspect of the system operation at runtime.
  - E.g., bandwidth, requests/second, CPU load,…

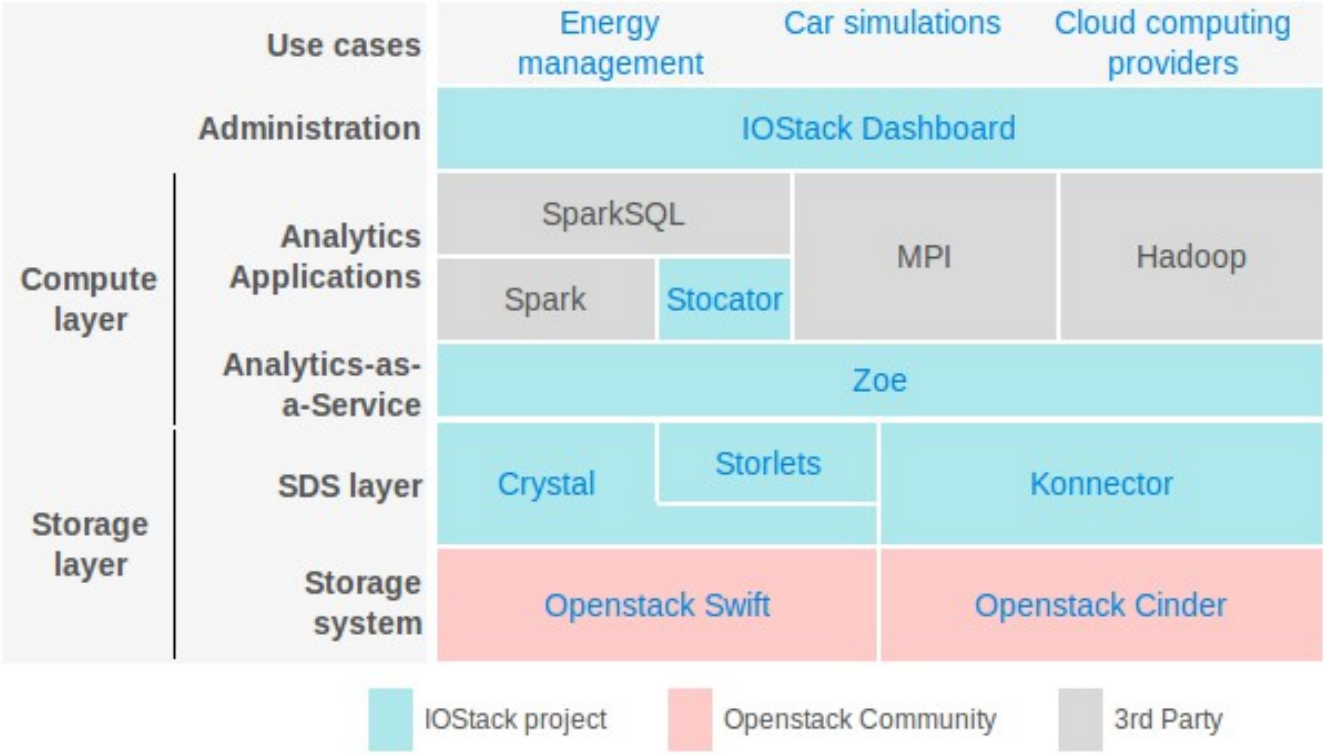
Design Entities of IOStack (II)

- **Policy**: Contract with the SDS system to provision a service/resource to a tenant.
  - E.g.: Apply data compression on tenant T1's request
  - E.g.: Ensure that T1 gets at least 50MBps of bandwidth (SLO)

- **Controller**: Algorithm that receives as input workload metrics to manage the execution of filters.

Do compression!
More bandwidth to this tenant!!
Stop caching!

A Software Stack for Big Data



Headlines

- **Performance**: Gridpocket's Spark experiments were optimized by IBM using IOSTACK technologies.
  - Tested in the RackSpace OSIC cluster (65 machines)
  - We obtained up to 25 speedup gains (from days to hours)

GRIDPOCKET
PERSONAL SMARTGRID SOLUTIONS

- **Cost reduction**: Idiada's increasing costs in storage were reduced up to 50% thanks to data reduction and flexible data management.

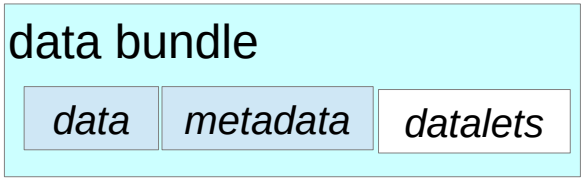
Applus⊕
IDIADA

Data-driven challenges

1. **Data-centric approach**: Beyond objects, blocks, files or rows
2. **User-centric approach**
3. **Programmability and extensibility**
4. **Interoperability**

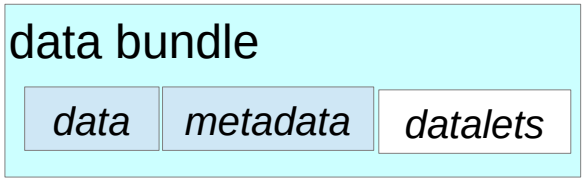
DATALETS

What is a datalet ?



- It is a new programming abstraction designed to cover the entire life-cycle of data
- Data bundles as self-contained entities including both data, meta-data and software (datalets)
- Tasks:
  - Data protection
  - Data management
  - Data manipulation
  - Data placement

Virtualizing data

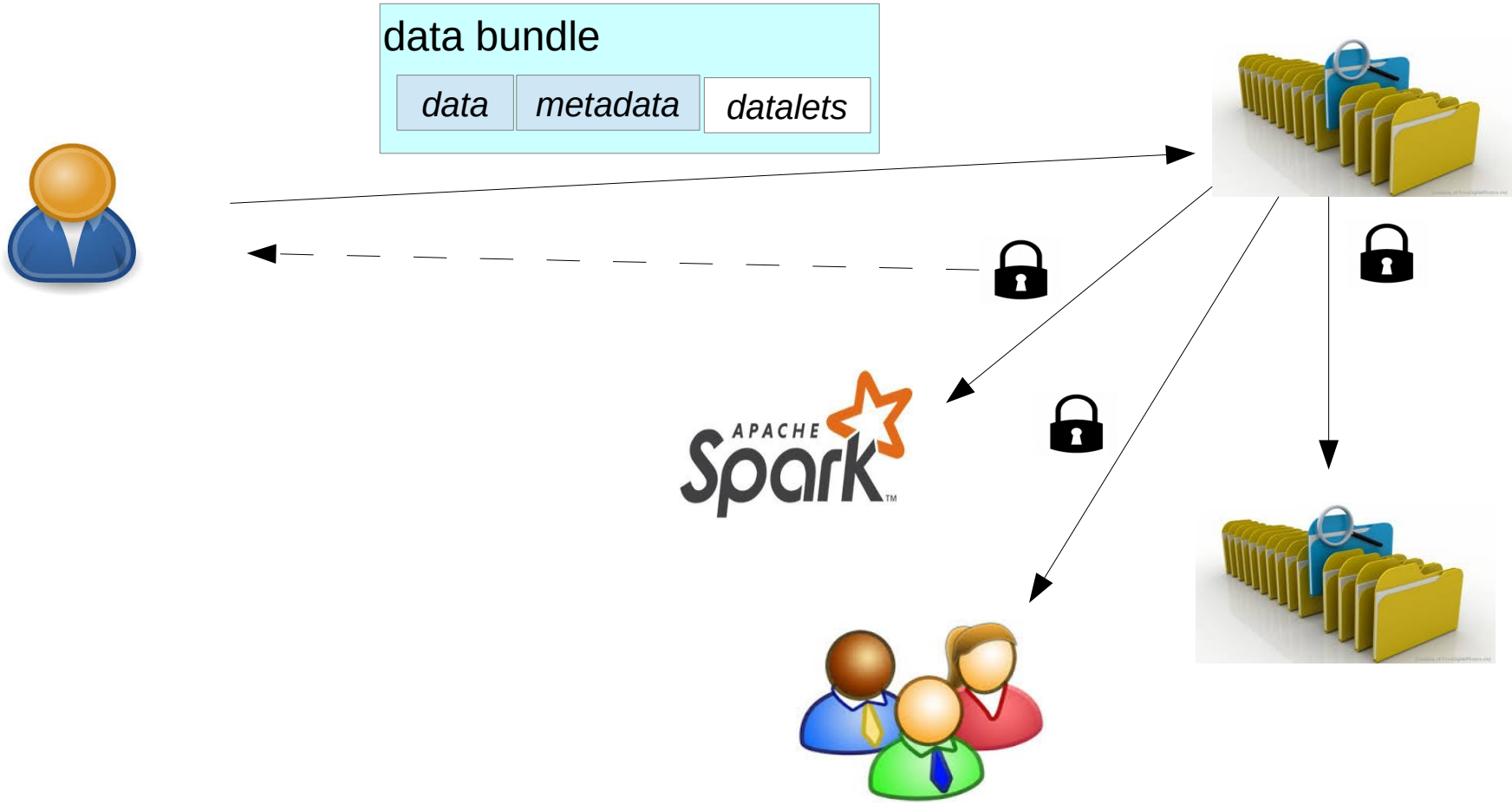


- Decoupling data from the underlying storage system
- Data transformations to rejuvenate data
- Datalets as micro-controllers managed by data owners

Software Defined Protection

- **Data-centric approach**: Secure code execution environment in the storage layer (data bundles)
- **User-centric approach**: secure micro-controllers managed by data owners
- **Programmability**: datalets as extensible software entities that participate in the entire data life-cycle
- **Interoperability** : Standards are required to avoid incompatible sandboxes for datalets. Data should not be tied now to specific software

Software Defined Protection



Conclusions

Conclusions

- Software-Defined technologies are appropriate for increasing **programmability** and flexibility
- SD technologies may "**virtualize data**" and decouple it from specific storage systems.
- **Datalets** as high-level abstraction entities designed to manipulate data
- Evolution of Software Defined Storage concepts: **from filters to datalets**

IOSTACK

Software-defined storage for big data

Universitat Rovira i Virgili

ARCTUR

BSC

EURECOM
Sophia Antipolis

GRIDPOCKET
PERSONAL SMARTGRID SOLUTIONS

IBM

Applus IDIADA

MPSTOR