STORAGE & DATA MANAGEMENT

Review of server platforms

Maciej Brzeźniak, PSNC

1st SIG-CISS meeting in SurfSARA,

Sep. 25-26th 2017

SERVERS FOR SDS – DECISION CRITERIA

- **Price / TB:**
  - Replication? (2x-3x?) -> need HDDs
  - Erasure coding -> then need CPU

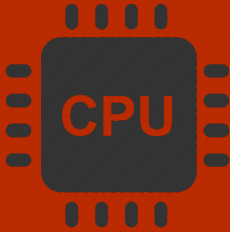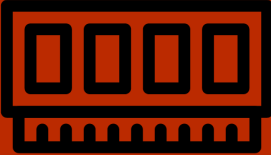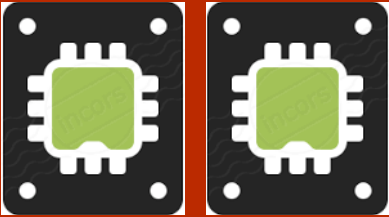- **Price / performance:**
  - The thinner the server the faster I/O (more cores / OSD/drive)
  - SSD vs HDDs

- **Price vs reliability?**
  - Avoid some vendors? we had bad experience with some
  - Max node size = Maximum number of drives * capacity of drives:
    - *what is the biggest volume of data than can be reconstructed in case of a server failure while not impacting the cluster? – experience?*

EXAMPLE ISSUE: STORAGE
HARDWARE PURCHASE COSTS

Component	COST [$]	% of CAPEX	REMARKS
CPU	500-1.000	10-20%	performance
(memory)	250-500	5-10%	data buffering / parallelism
(disks)	200-300 /disk 3.000-20.000 /server	50-80%	TB / IOPS
(SSD)	250-500 /disk 500-1.000 /server	5-20%	IOPS / latency

EXAMPLE ISSUE: STORAGE: TCO: ENERGY COSTS

server architecture	example server		TB/ rack (6TB disks)	energy/TCO 5 years
12 HDD/1U 48 HDD/4U **2 CPU / 12 HDD**			3840	**25%**
36 HDD/4U **2CPU / 36 HDD**			5760	**15%**
72 HDD/4U **2CPU / 78 HDD**			5760	**10%**

QuantaPlex T21P-4U:

- 4U server – 2 options:
  - 2x35HDD = 70 HDDs
  - 1x78HDD = 78 HDDs
  - 1x SATADoM
- CPUs & RAM:
  - 2 sockets, Intel E5 v3/4
  - 16 DIMM slots to up 512/1024GB
- Network:
  - 2x10GbE or 1x 40GbE now
- Planned usage:
  - Object storage (78 HDD option)
  - Object storage + RBD (2x 35 HDD)
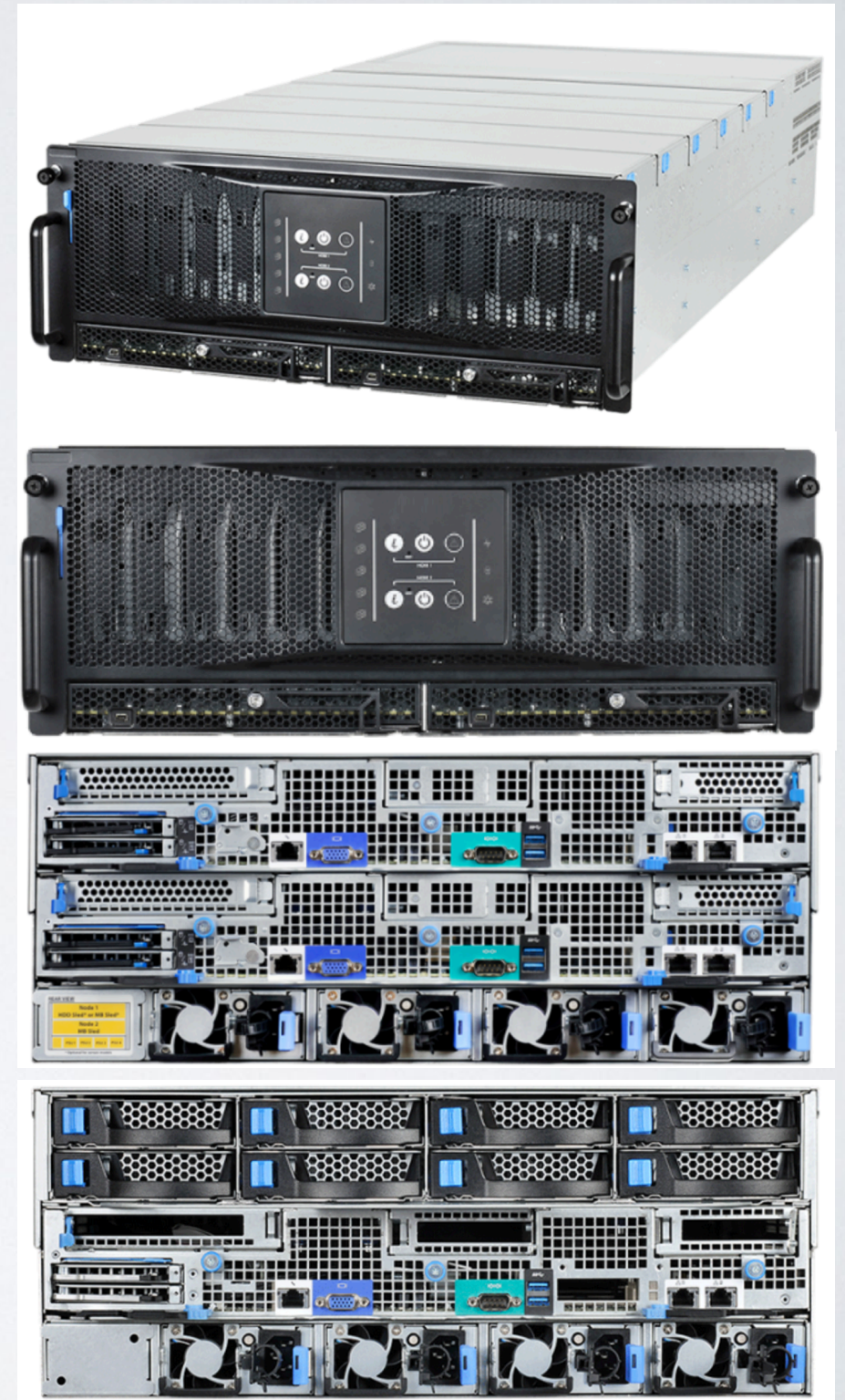
FUTURE PLANS
LARGE CAPACITY SERVERS

Say: QuantaPlex T21P-4U:

- Probably will go for:

  - 2x35HDD = 70 HDDs

  - *Not brave enough to use 1x78HDD*

- Sample on the way to PSNC

  - Want to see how the server is made

  - 1-node tests planned

- Possibly will purchase larger set in 201

Alternatives:

- *1U server + JBODs*
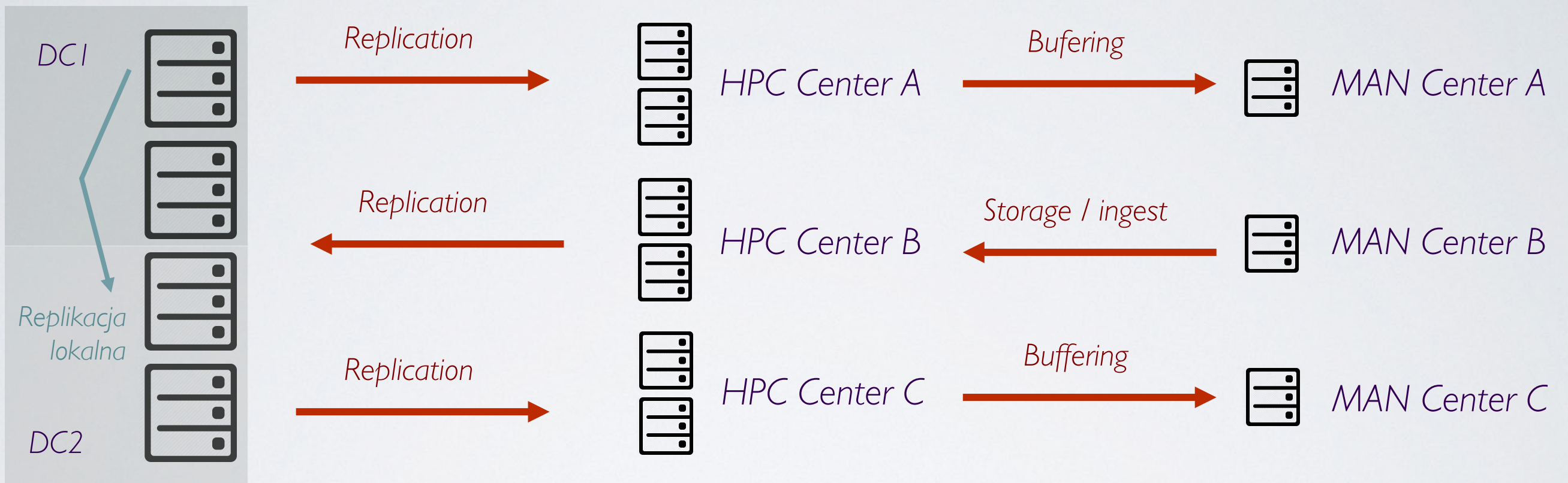
- *Servers >70 HDDs*

Poznań   *Other centres*

Long-term perspective (national data storage replacement)

Main (PSNC's) data centres	Geo-replicas and local 'buffers'	Local buffers & PoP of storage services

DC1

Replication → HPC Center A   Bufering → MAN Center A

← Replication   HPC Center B   ← Storage / ingest   MAN Center B

Replikacja lokalna

Replication → HPC Center C   Buffering → MAN Center C

DC2

Poznań

HPC center:

MAN centee:

20 servers
12 PB
10GB/s
2 racki

10 servers
4 PB
3GB/s
1 rack

5 servers
2 PB
3GB/s
1/2 racka

QuantaGrid D51PH-1ULH:

- 1U server – options:
  - 12xHDD 3,5'' – all hot swappable
  - 4x SSD 2,5'' – loadable from front
  - 1x SATADoM
- CPUs & RAM:
  - 2 socket mainboard
  - Intel E5 v3/4
  - 16x DIMM slots: up to 512GB
- Network:
  - 2x10GbE or 1x 40GbE now
- Usage:
  - Current: Ceph for RBD on HDDs
  - Future: ScaleIO on SSDs
    possibly in a 'hyper-convergent setup' – running OpenStack in the same time

SERVERS AT PSNC

40x QuantaGrid: 12 x 4TB drives = 48 RB /server = 1920 TB per 40U rack
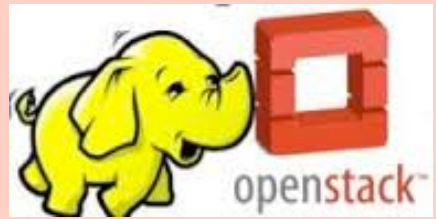
Usage: see below

Hadoop oD

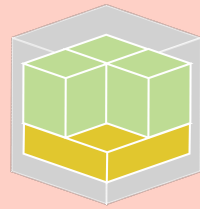**OpenStack Swift**

Ceph

BIG DATA ON DEMAND
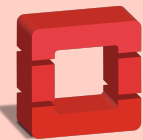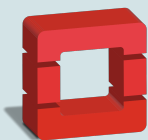

OpenStack Sahara

Big Data/
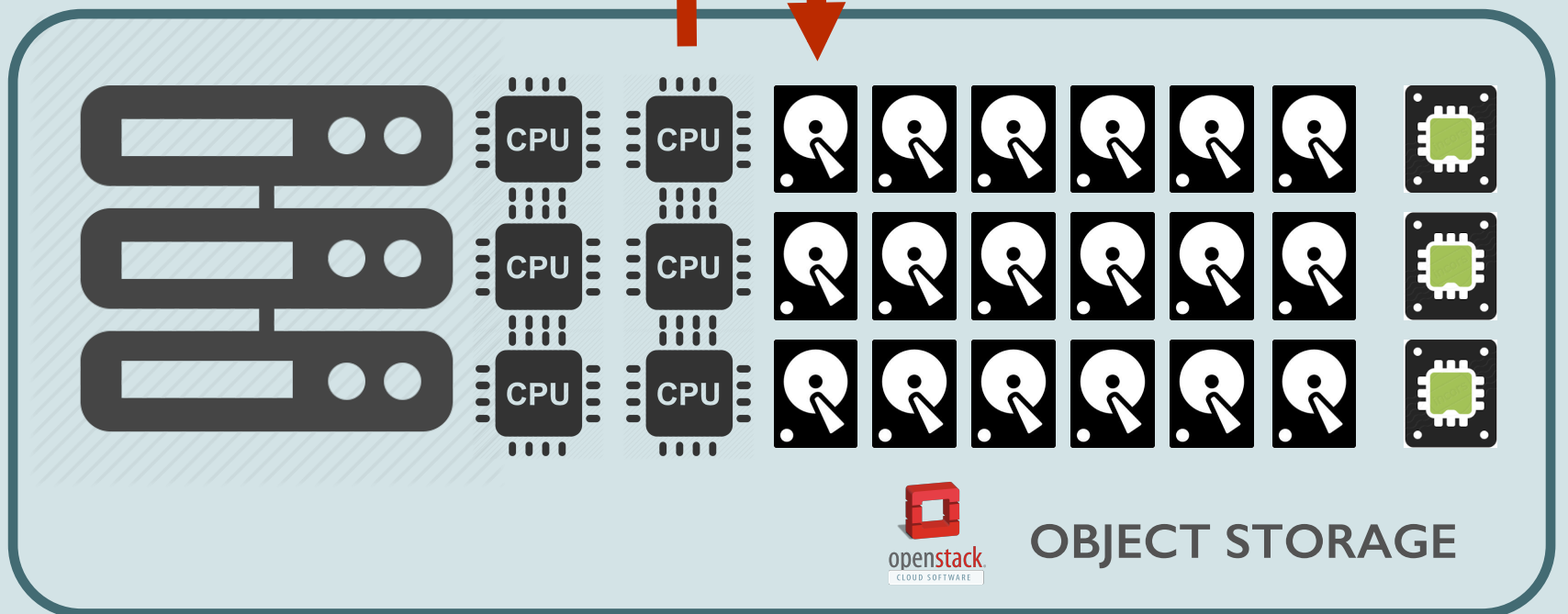Analytics
on demand

COMPUTE

openstack
CLOUD SOFTWARE

CPU CPU

CPU CPU

CPU CPU

Spark hadoop

Persistent
& efficient
storage
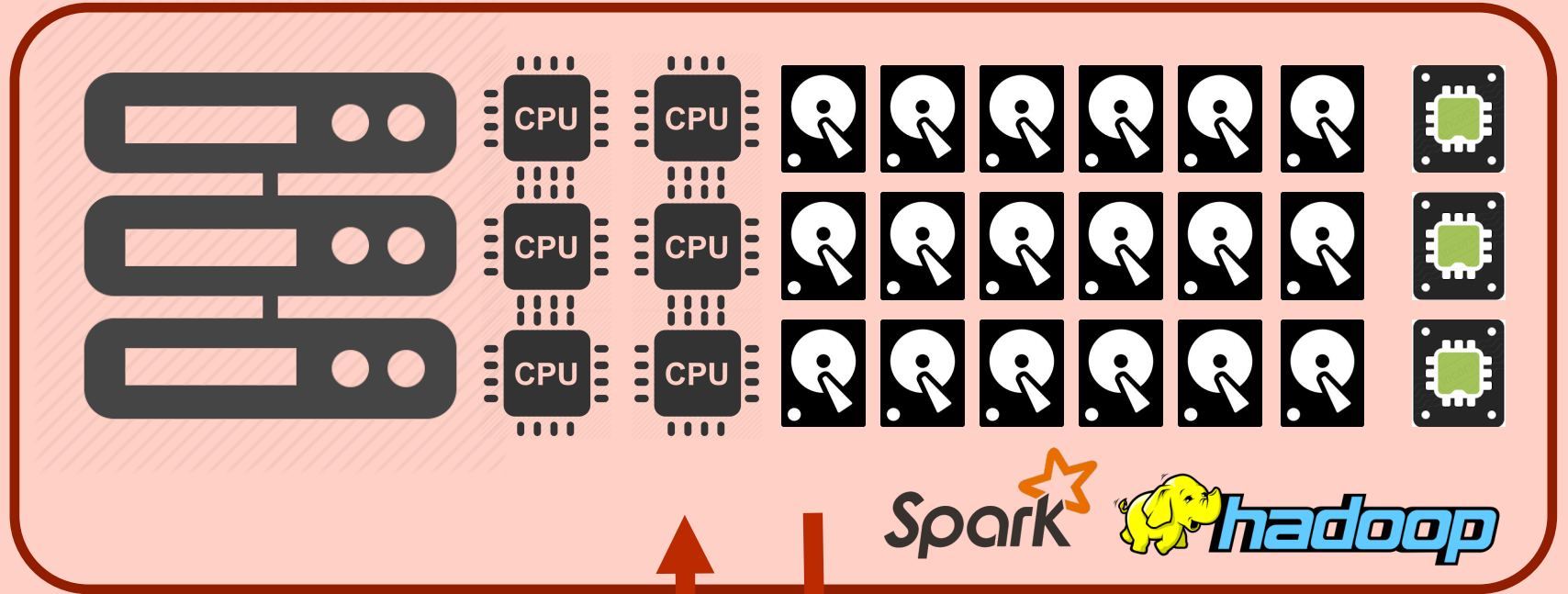
OBJECT STORAGE

openstack
CLOUD SOFTWARE

CPU CPU

CPU CPU

CPU CPU

openstack
CLOUD SOFTWARE

OBJECT STORAGE

QuantaGrid D51BP-1U:

- 1U server – options:
  - 10xHDD 2,5'' – all hot swap
  - 2 SATADoMs
- CPUs & RAM:
  - 2x Intel E5 v3/4
  - 16x DIMM slots: up to 512GB
- Network:
  - 2x10GbE
    or 1x 40GbE now
- Usage:
  - ScaleIO or
    Fusion STorage on SSDs
    possibly in a 'hyper-convergent setup' – running OpenStack in the same time

ScaleIO-read node from EMC:

- server – options:
  - 1U 10x SSD 2,5'' – all hot swap
  - 2U 24+2 x SSD 2,5'' (or 16x3,5'')
- CPUs & RAM:
  - 2x Intel E5 2600 v4
  - 24DIMM slots: up to 512GB
- Network:
  - 2x10GbE:
- Usage:
  - ScaleIO
  - Possibly hyper-converged

	HIGH-DENSITY	**HIGH-CAPACITY**
	Dell EMC PowerEdge R630	Dell EMC PowerEdge R730xd
	1U1N 1 node within a 1RU enclosure 10 drive slots	2U1N 1 node within a 2RU enclosure 24 drive slots
HYPER-CONVERGED	All-flash or hybrid (support designation DC200-F, DC200)	All-flash or hybrid (support designation CC200-F, CC200)
STORAGE-ONLY	All-flash, all-HDD, or hybrid (support designation DS200-F, DS200)	All-flash, all-HDD, or hybrid (support designation CS200-F, CS200)

STORAGE & DATA MANAGEMENT

THANK YOU!

Maciej Brzeźniak, PSNC

1st SIG-CISS meeting in SurfSARA,

Sep. 25-26th 2017