CSNSM
CENTRE DE SCIENCES NUCLÉAIRES ET DE SCIENCES DE LA MATIÈRE

IPN
INSTITUT DE PHYSIQUE NUCLÉAIRE
ORSAY

Distributed storage : Ceph Monitoring with Xymon

Jérémie Jacob / Patrick Le Jeannic

AGATA Week 2017

GPFS to Ceph : why ?

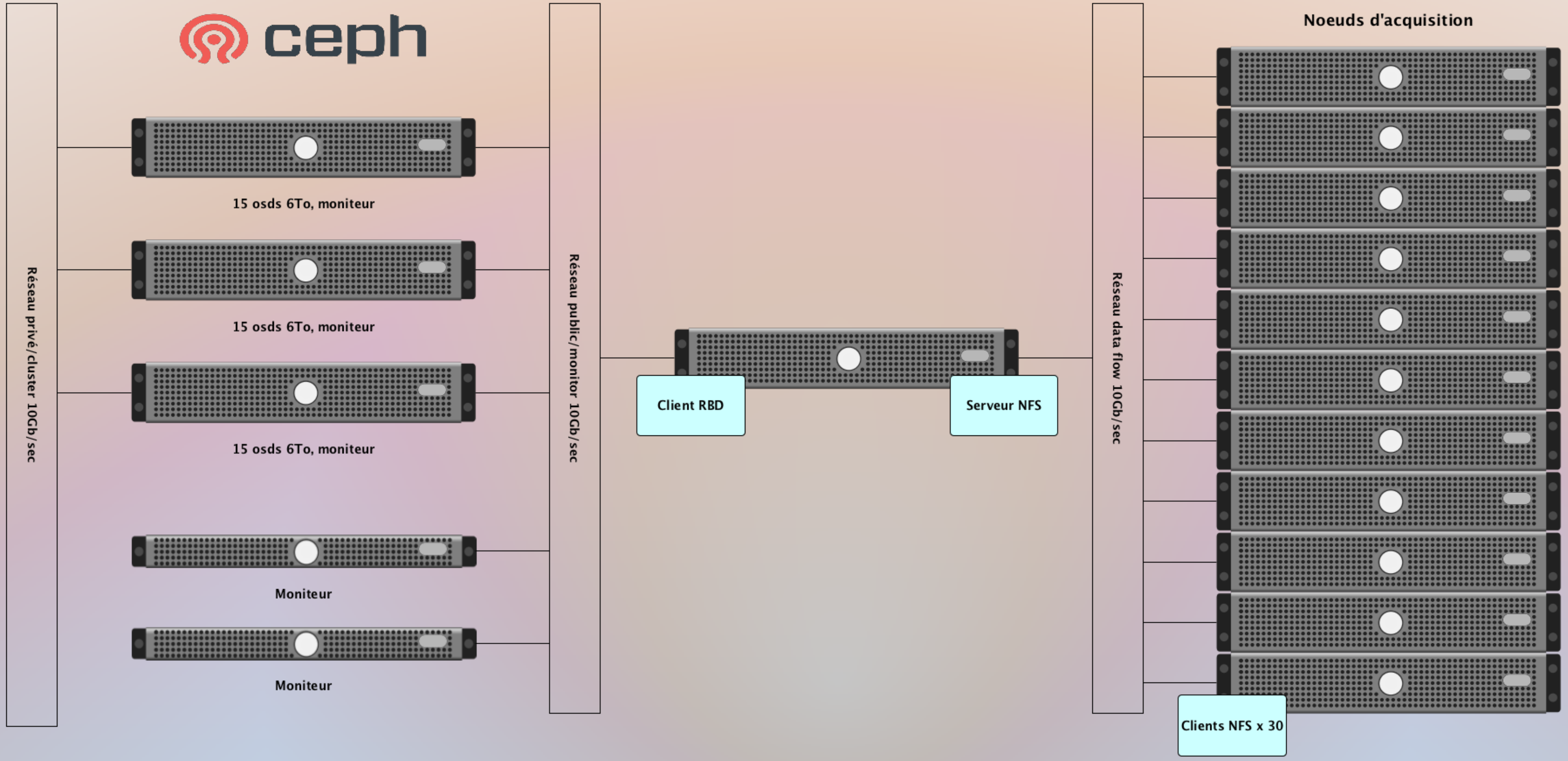
GPFS

- Limited bandwidth (1Gb/s), in regard to the increasing number of crystals
- Storagetek storage filers nearing end of warranty
- 1To internal HDD
- GPFS & Common Array Manager not supported in Debian
- Licences acquisition required for any upgrade/update

Ceph

- Common hardware use
- Open source solution, very active community, supported by RedHat
- Debian compatibility
- Scalable solution
- Positive community feedback

AGATA Ceph Architecture



Jérémie Jacob /Yann Aubert

Problems encountered

- Clock skew sensitivity

  → Ntp synchronisation of nodes

- Problem with the RBD client on reboot

  → Install a more recent kernel on the client

- Space freed by the filesystem but not freed by Ceph (thin provisionning)

  → Periodic fstrim, but strong augmentation of the latency

  → To avoid fstrim, use the discard option on mount

- During reconstruction, loss of RBD mount (the monitors could not see each other)

- Irregular filling of the disks => 80 % of the disk space usable

Conclusion

→ Complex system, but very efficient

- Common hardware
- Debian compatibility
- + 120To available (80 % usable)
- Bandwidth over NFS : 6Gb/s with 30 clients
- After stopping the GPFS cluster, physical space available in the racks of the server room

Evolutions

In progress

- Separation of the monitors / storage servers
- Backup server ready, in case the Ceph cluster becomes unavailable

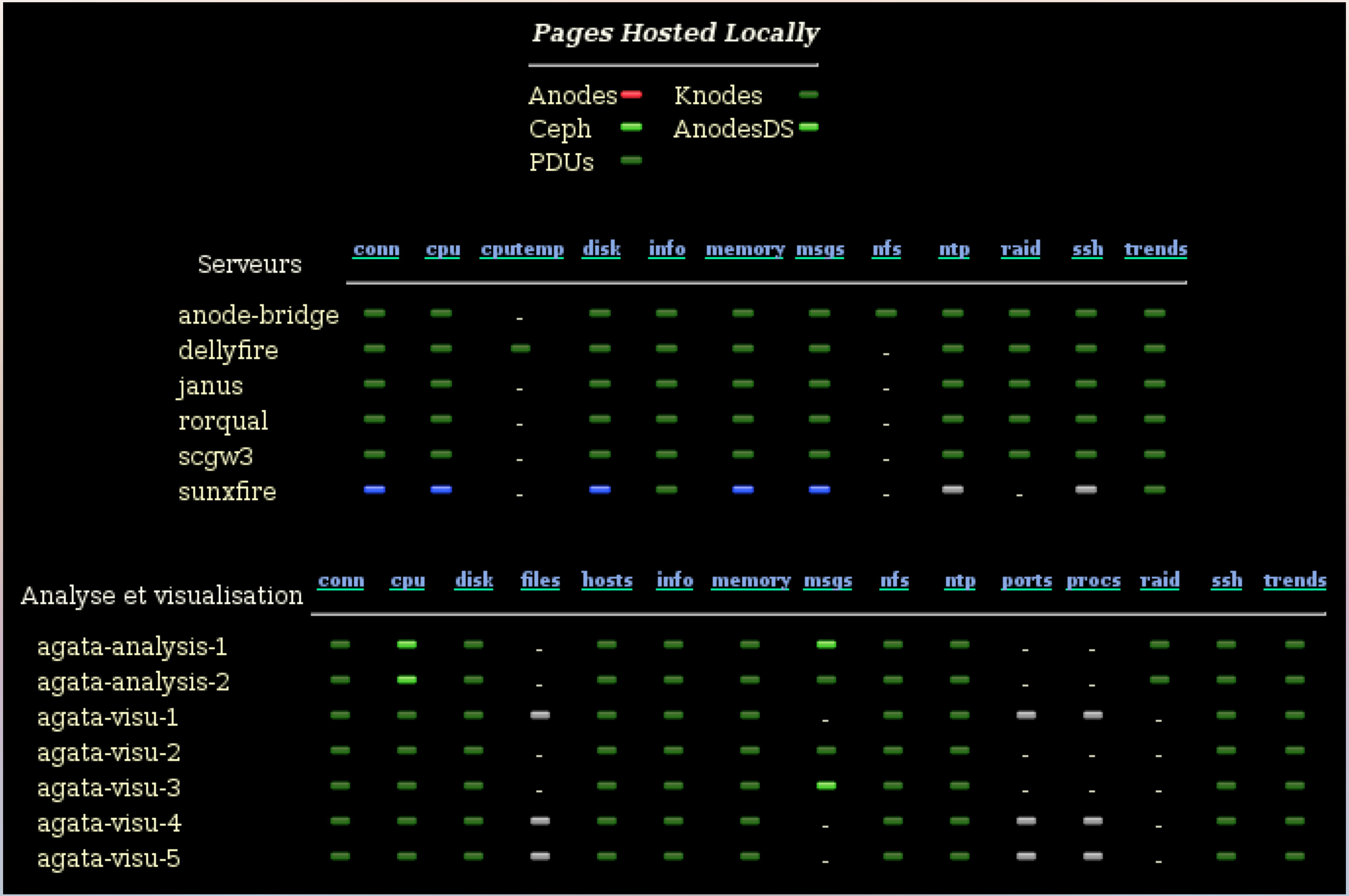
In the future / planned / forthcoming

- Add a 4th storage server to the cluster
- Upgrade to the latest LTS version available
- Add scripts to Xymon to monitor more precisely the cluster

Xymon monitoring

- Simple client/server system
- Web GUI : fast visualisation of the global state of the system
- Mail alerts, fast & simple to configure
- « home-made » tests : NFS mounts, XGGP, Xylinx, …
- History of the monitored services (graphs ...)

Aperçu de Xymon



Pages Hosted Locally

Anodes	🟥	Knodes	🟩
Ceph	🟩	AnodesDS	🟩
PDUs	🟩		

Serveurs	conn	cpu	cputemp	disk	info	memory	msgs	nfs	ntp	raid	ssh	trends
anode-bridge	🟩	🟩	-	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
dellyfire	🟩	🟩	🟩	🟩	🟩	🟩	🟩	-	🟩	🟩	🟩	🟩
janus	🟩	🟩	-	🟩	🟩	🟩	🟩	-	🟩	🟩	🟩	🟩
rorqual	🟩	🟩	-	🟩	🟩	🟩	🟩	-	🟩	🟩	🟩	🟩
scgw3	🟩	🟩	-	🟩	🟩	🟩	🟩	-	🟩	🟩	🟩	🟩
sunxfire	🟦	🟦	-	🟦	🟩	🟦	🟦	-	⬜	-	⬜	🟩

Analyse et visualisation	conn	cpu	disk	files	hosts	info	memory	msgs	nfs	ntp	ports	procs	raid	ssh	trends
agata-analysis-1	🟩	🟩	🟩	-	🟩	🟩	🟩	🟩	🟩	🟩	-	-	🟩	🟩	🟩
agata-analysis-2	🟩	🟩	🟩	-	🟩	🟩	🟩	🟩	🟩	🟩	-	-	🟩	🟩	🟩
agata-visu-1	🟩	🟩	🟩	⬜	🟩	🟩	🟩	-	🟩	🟩	⬜	⬜	-	🟩	🟩
agata-visu-2	🟩	🟩	🟩	-	🟩	🟩	🟩	🟩	🟩	🟩	-	-	-	🟩	🟩
agata-visu-3	🟩	🟩	🟩	-	🟩	🟩	🟩	🟩	🟩	🟩	-	-	-	🟩	🟩
agata-visu-4	🟩	🟩	🟩	⬜	🟩	🟩	🟩	-	🟩	🟩	⬜	⬜	-	🟩	🟩
agata-visu-5	🟩	🟩	🟩	⬜	🟩	🟩	🟩	-	🟩	🟩	⬜	⬜	-	🟩	🟩