

# Hadoop i Spark

Mariusz Rafało

[mrafalo@sgh.waw.pl](mailto:mrafalo@sgh.waw.pl)

<http://mariuszrafalo.pl>

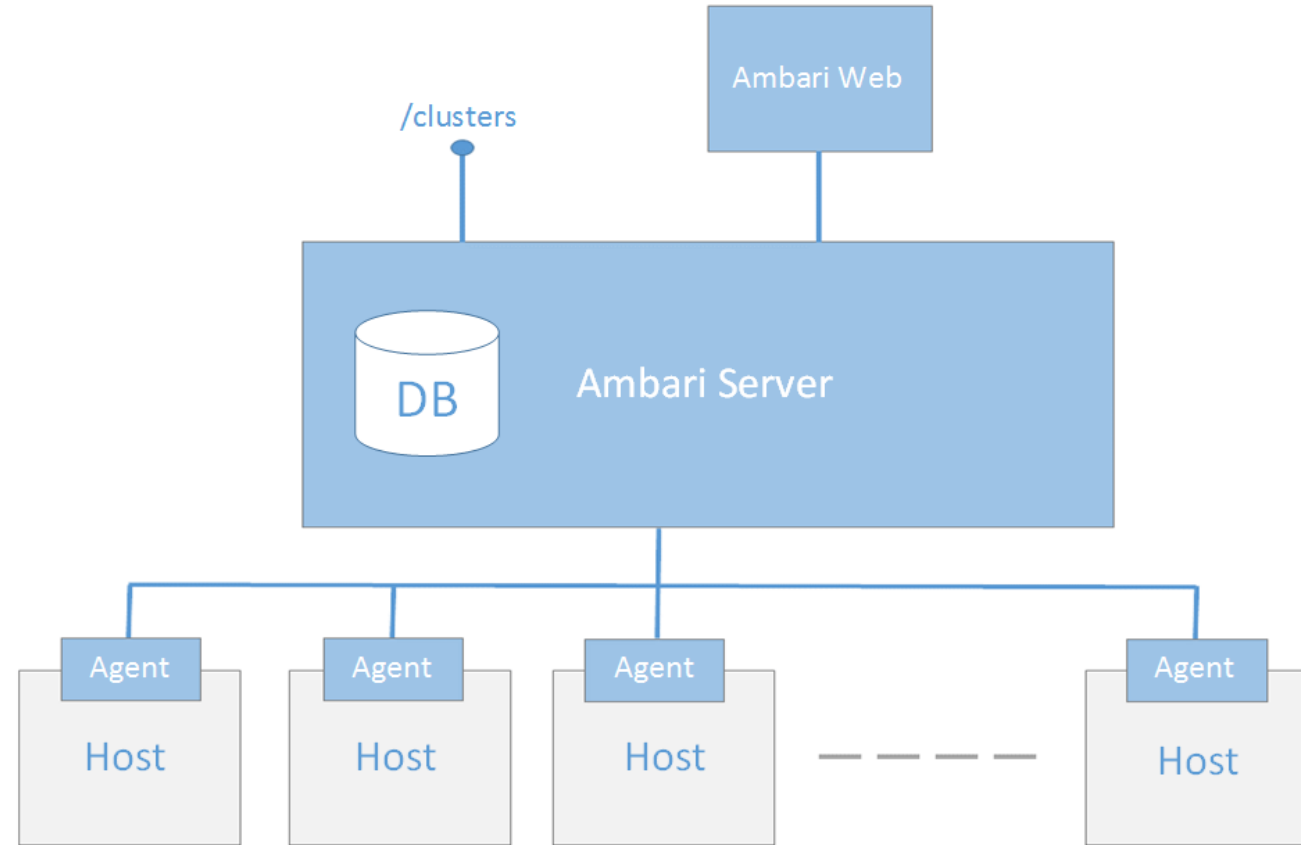
# ZARZĄDZANIE KLASTREM

# Zookeeper

- Scentralizowana usługa utrzymująca konfigurację klastra
- Zarządza metadanymi i konfiguracją klastra
- Zarządza konfiguracją rozproszoną, pozwala na automatyczne konfigurowanie nowego węzła (*data node*) lub nowej usługi na wielu węzłach

# Ambari

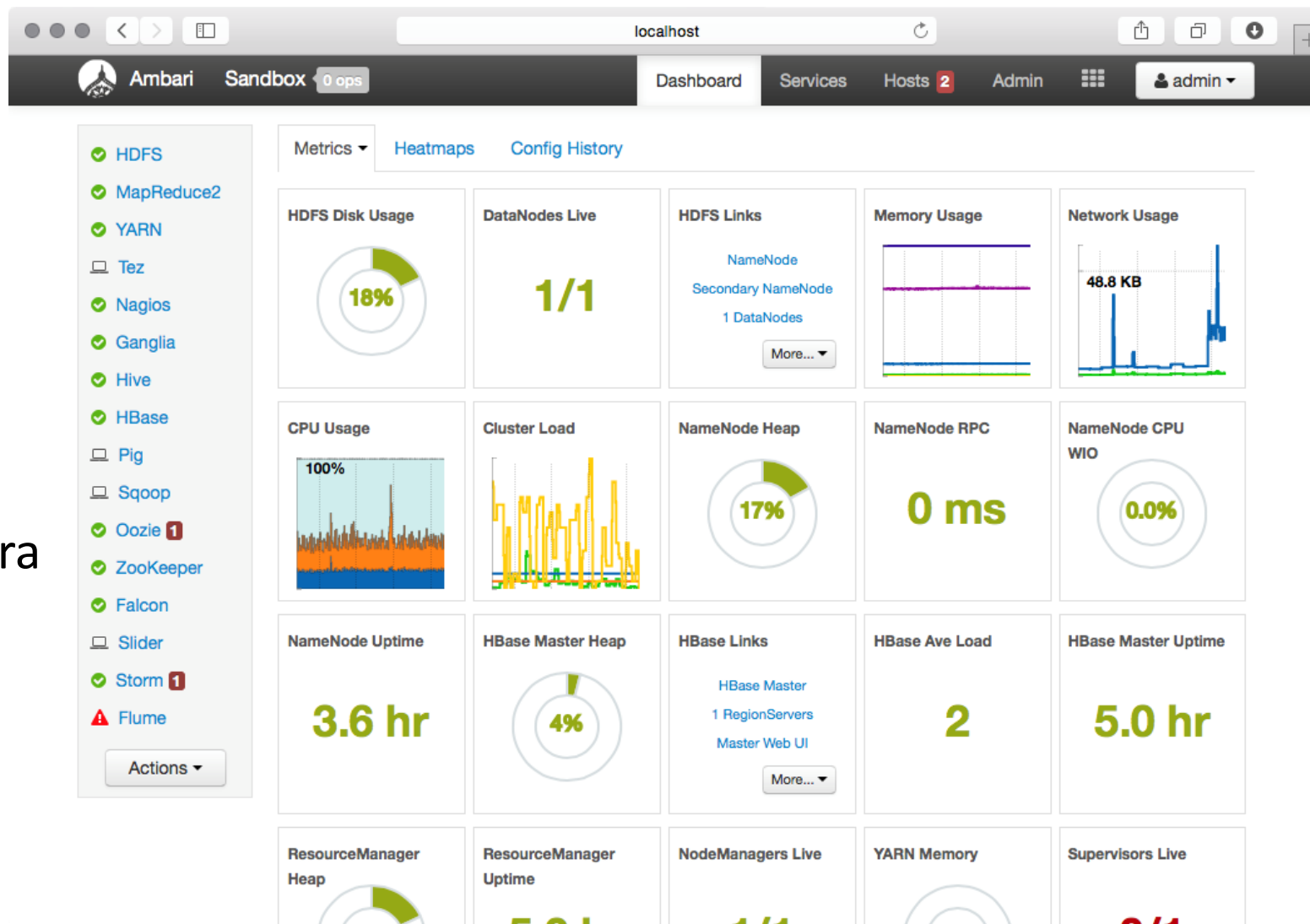
- Konsola do zarządzania klastrem oraz wszystkimi usługami
- Monitoruje zarówno zasoby klastra jak i poszczególne usługi



<https://www.e-zest.com/hadoop-cluster-management-with-apache-ambari>

# Ambari

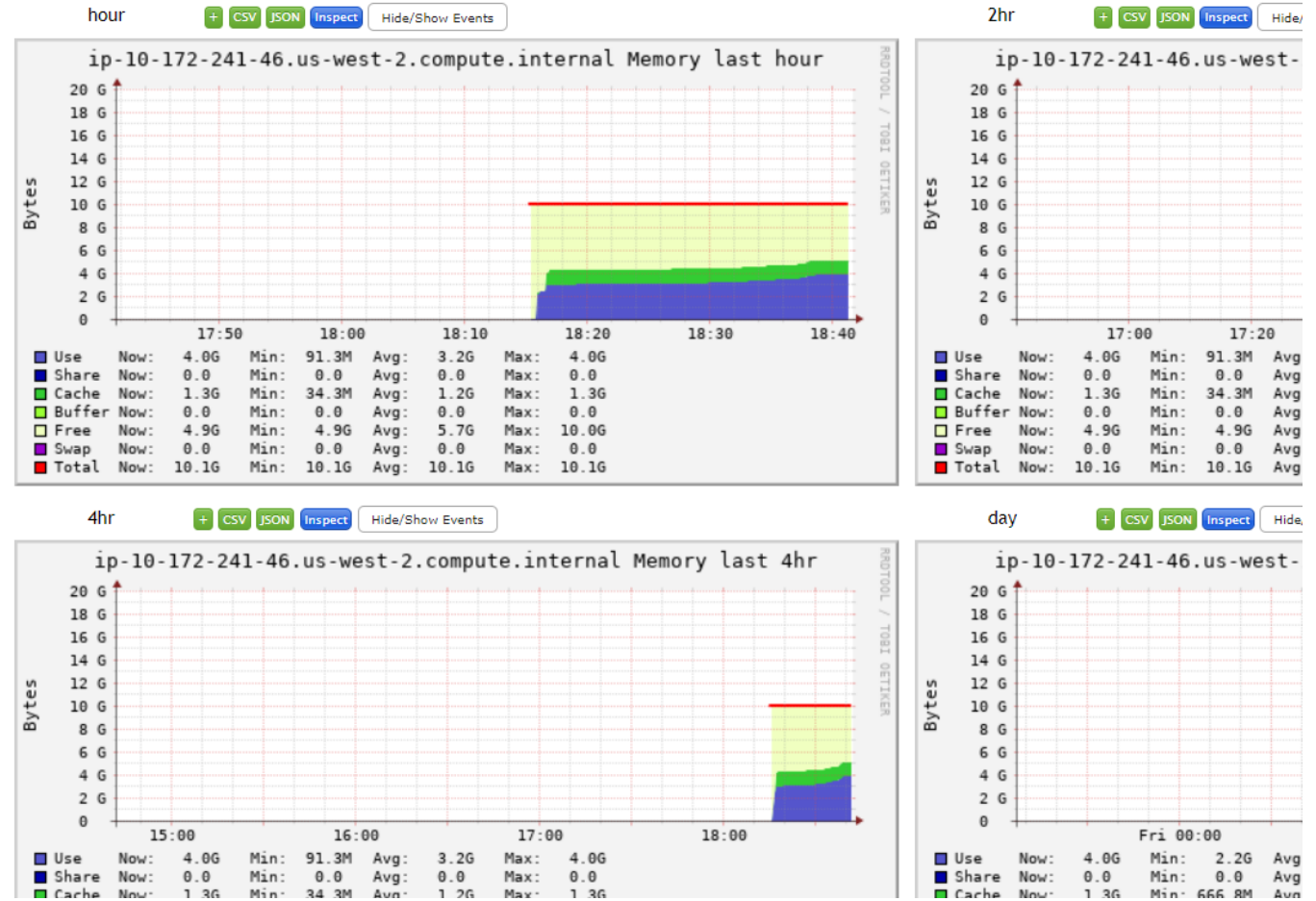
- Konsola do zarządzania klastrem oraz wszystkimi usługami
- Monitoruje zarówno zasoby klastra jak i poszczególne usługi



# Ganglia

- Monitorowanie zasobów klastra
- Alerty
- API

Host name: ip-10-172-241-46.us-west-2.compute.internal Graph: mem\_report



# Zarządzanie klastrem: zagadnienia

- Infrastruktura sprzętowa (jak monitorować?)
- Warstwa aplikacji (jak dbać o dostępność?)
- Przetwarzanie danych (jak ładować dane klaster?)
- Zarządzanie dostępem (bezpieczeństwo danych)
- Śledzenie i diagnostyka błędów w danych
- Wydajność

Dziękuję za uwagę