

# Hadoop High Availability al 100%

# Bla bla bla (sobre mi)

**Federico Leven**

- ➔ Skype: @federicol
- ➔ federico.leven@gmail.com
- ➔ Linkedin : <https://ar.linkedin.com/in/sojovi>

# Agenda

- ➔ Algunas definiciones iniciales
- ➔ Hadoop recién sacado de la caja
- ➔ HA Componente por componente
- ➔ Armando todo el rompecabezas
- ➔ Conclusiones

# Definiciones iniciales

DISPONIBILIDAD : El tiempo que un sistema se encuentra disponible para su uso en tareas productivas. Incluye el tiempo por mantenimiento programado.

ALTA DISPONIBILIDAD : Disponibilidad superior a la considerada disponibilidad normal. (No es lo mismo un Sistema de soporte de vida que un Sistema de un banco)

3 métricas importantes :

- 1) Tiempo promedio entre fallas
- 2) Tiempo promedio de reparación
- 3) Disponibilidad =  $(1) / ((1) + (2))$

ALTA DISPONIBILIDAD  $\geq 99.9\%$  (aprox 4hs 30mins al año)

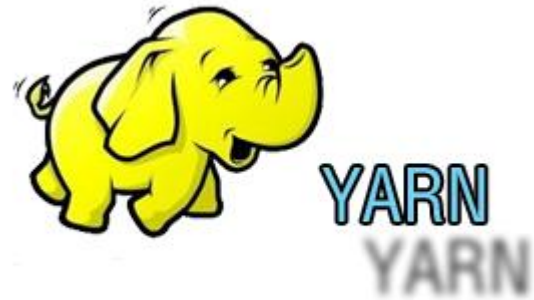
# Eventos que afectan la disponibilidad

- Fallos de hardware
- Fallos de red
- Errores en el Sistema Operativo
- Errores en la aplicación
- Errores humanos
- Sucesos ambientales (terremotos, inundaciones, etc)
- Sobrecargas del Sistema
- Ataques de seguridad

### 3 principios básicos de HA

- Eliminación de puntos únicos de falla -> Redundancia
- Detección temprana de fallas
- Recuperación automática

Hadoop recién sacado  
de la caja

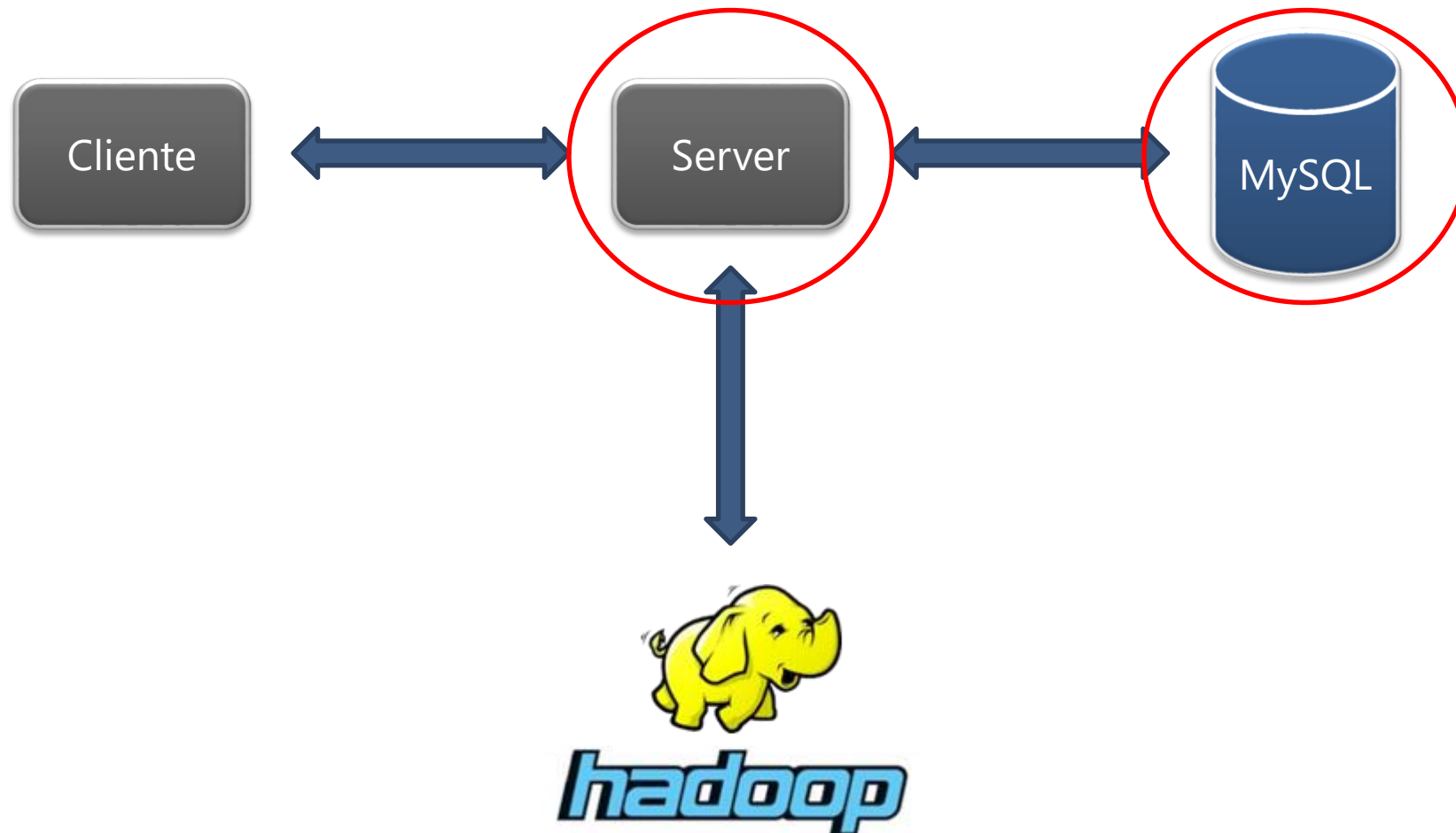


Ambari

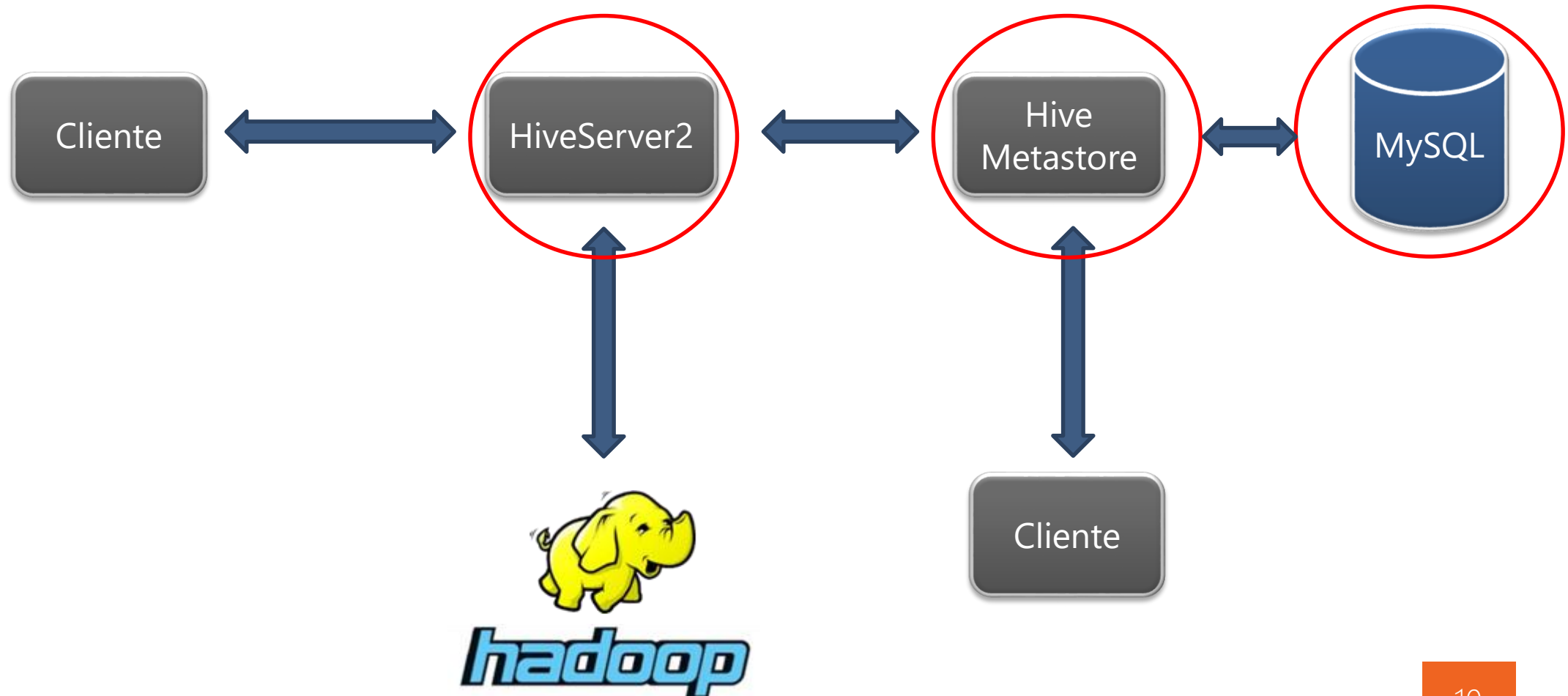




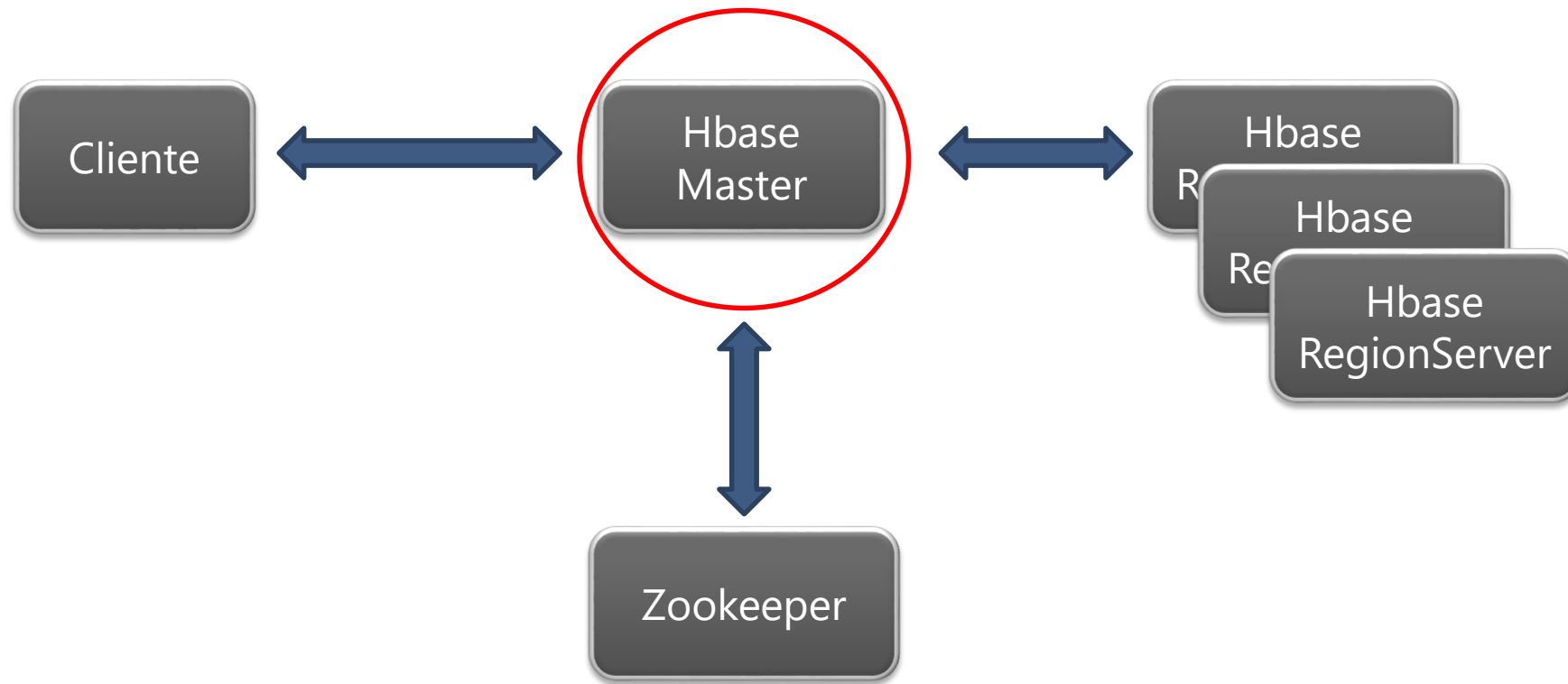
# Arquitectura de OOZIE



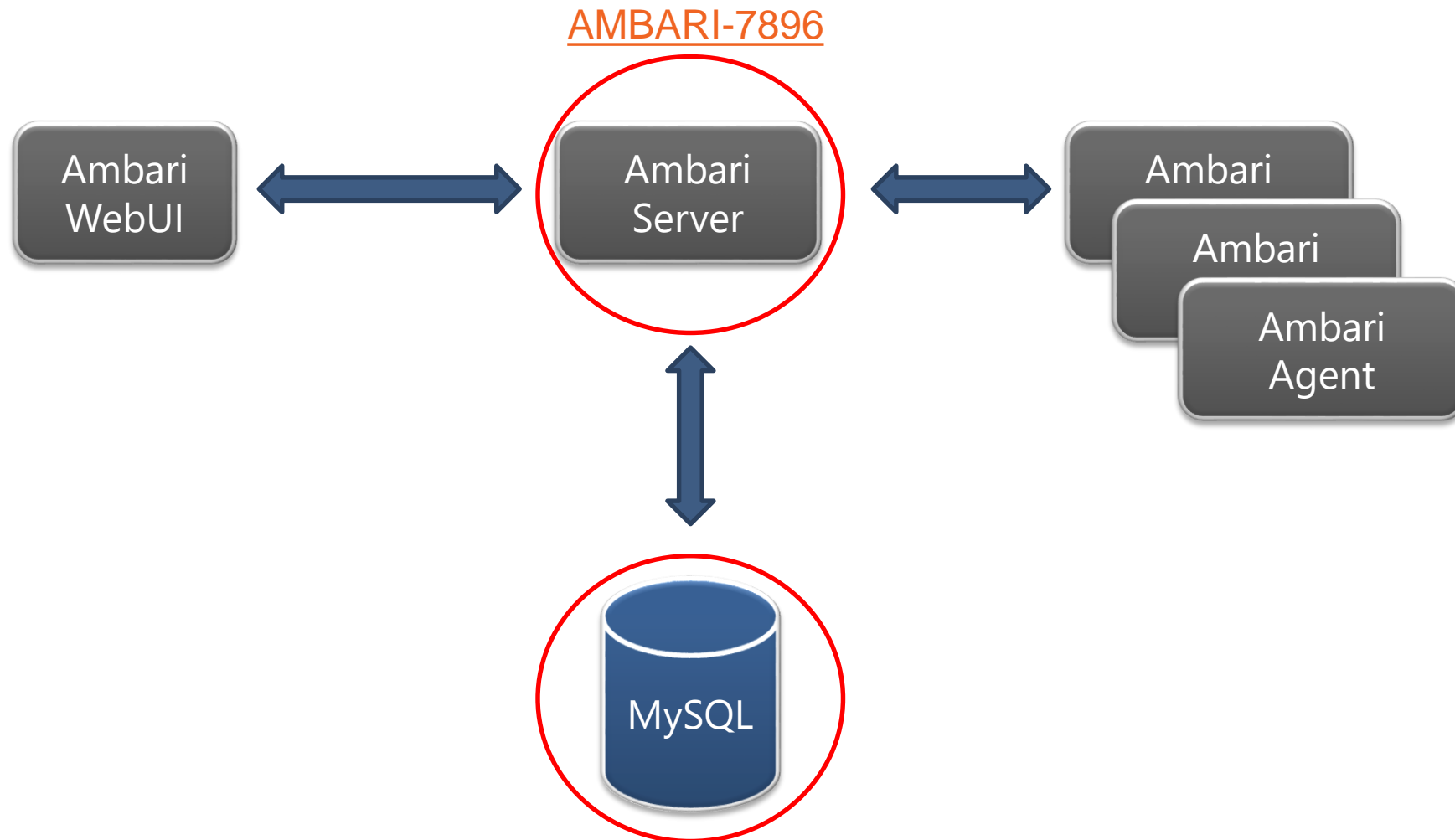
# Arquitectura de HIVE



# Arquitectura de HBASE



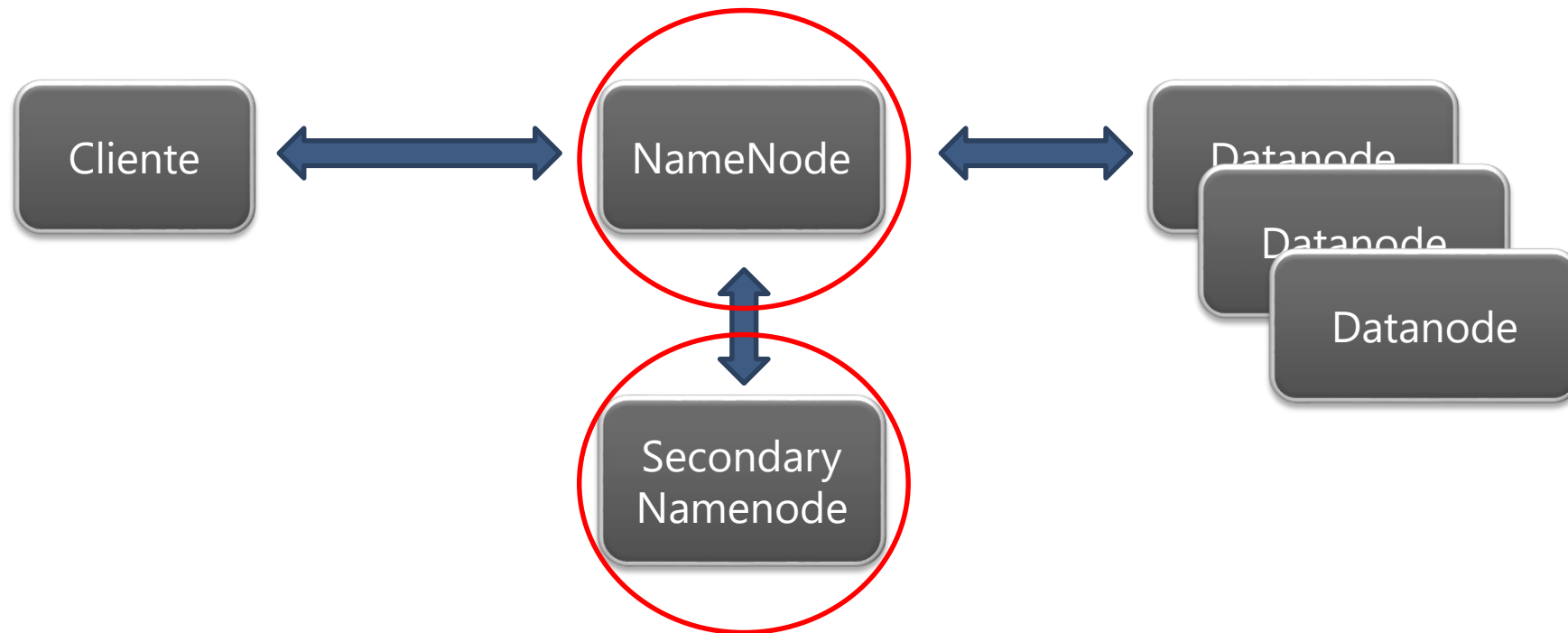
# Architecture de AMBARI



# Arquitectura de YARN

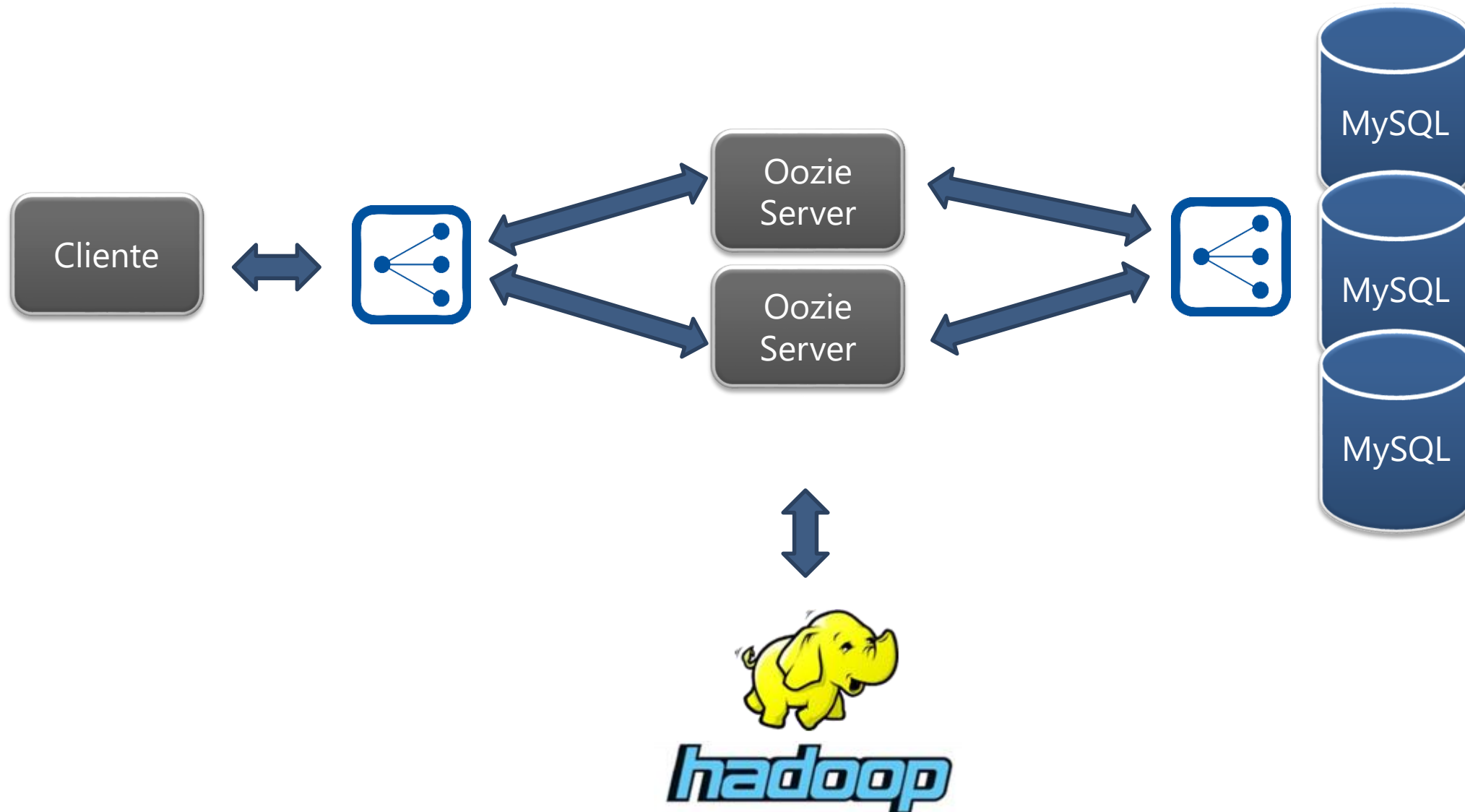


# Arquitectura de HDFS



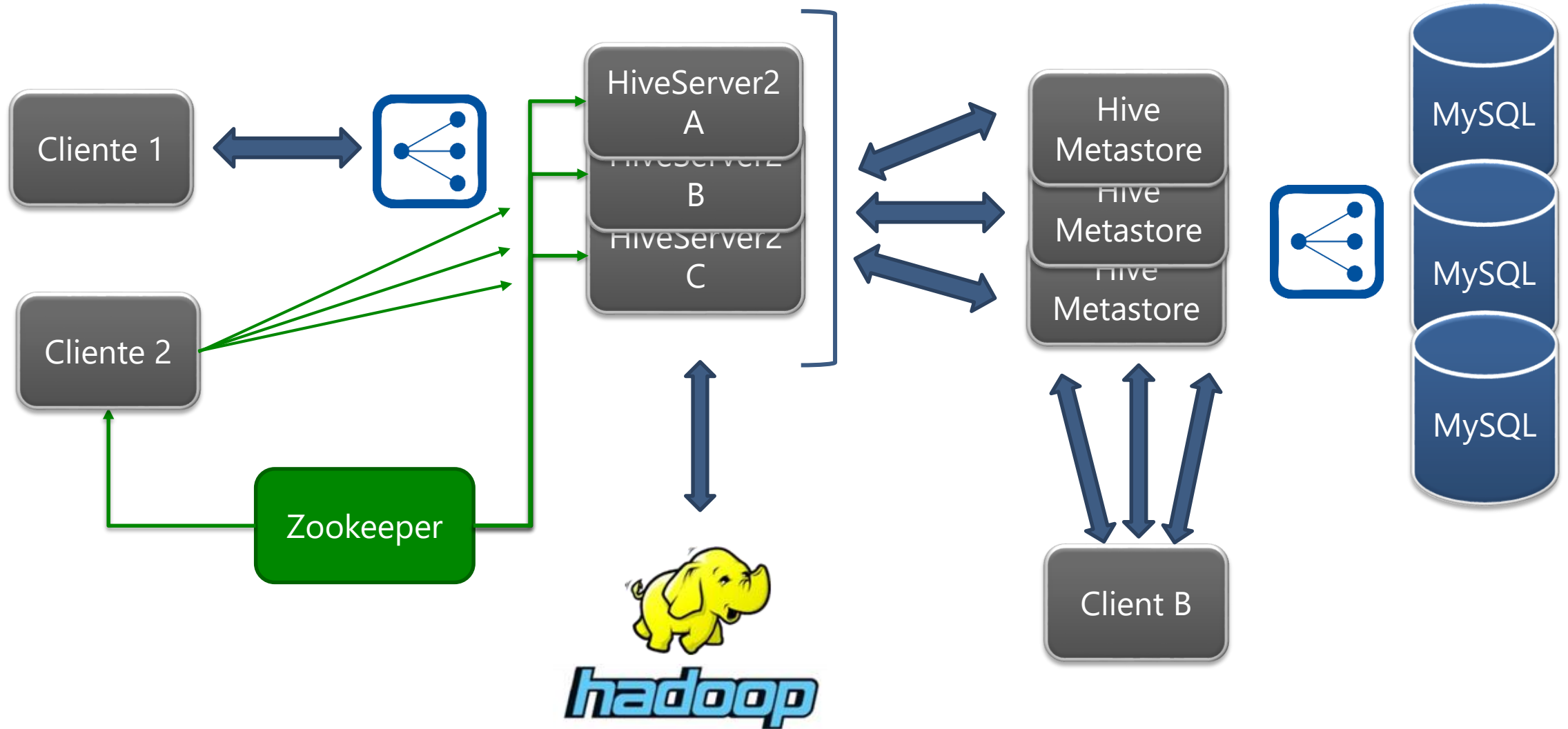
Armando todo el  
rompecabezas

# Arquitectura de OOZIE HA

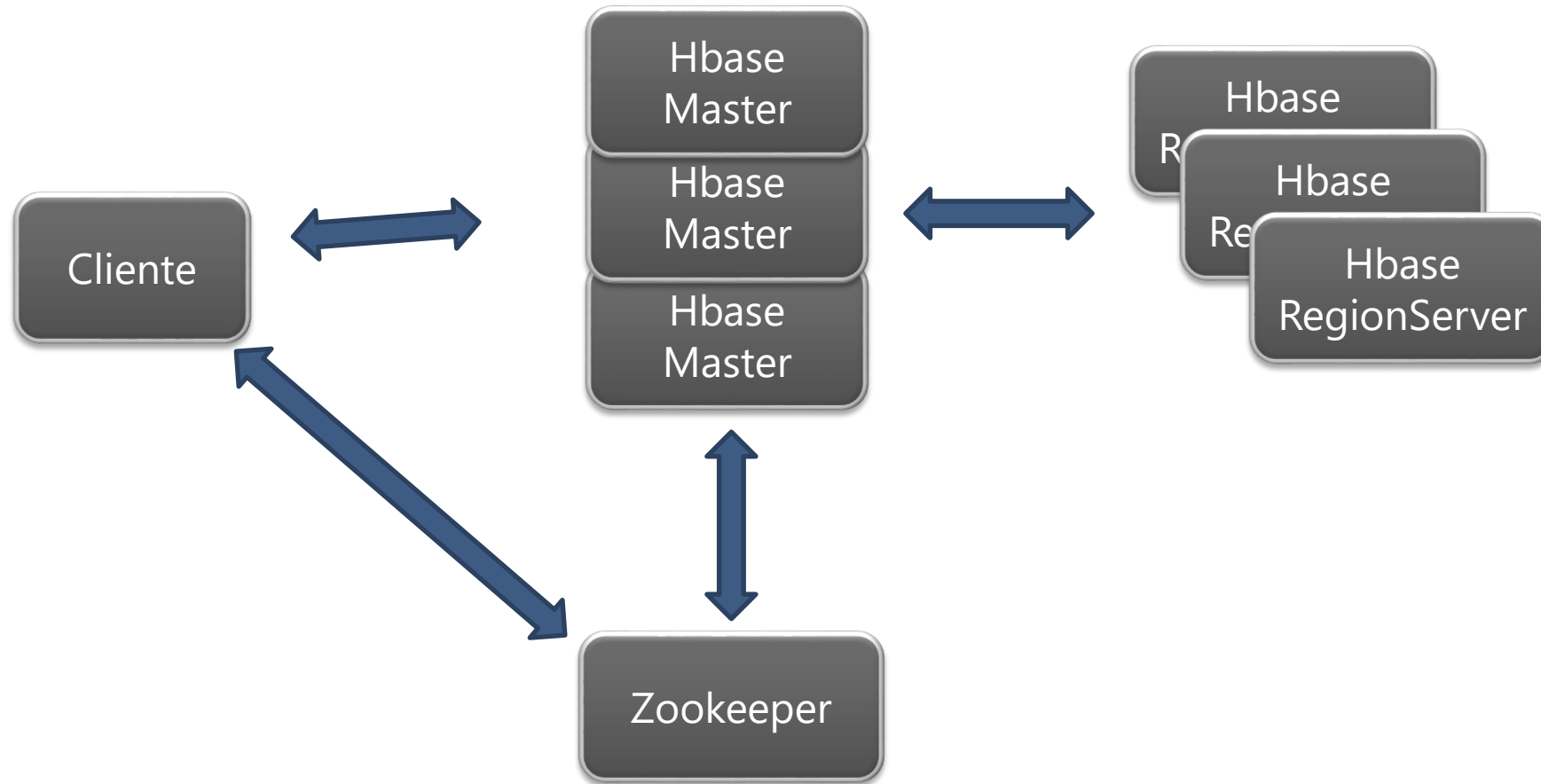




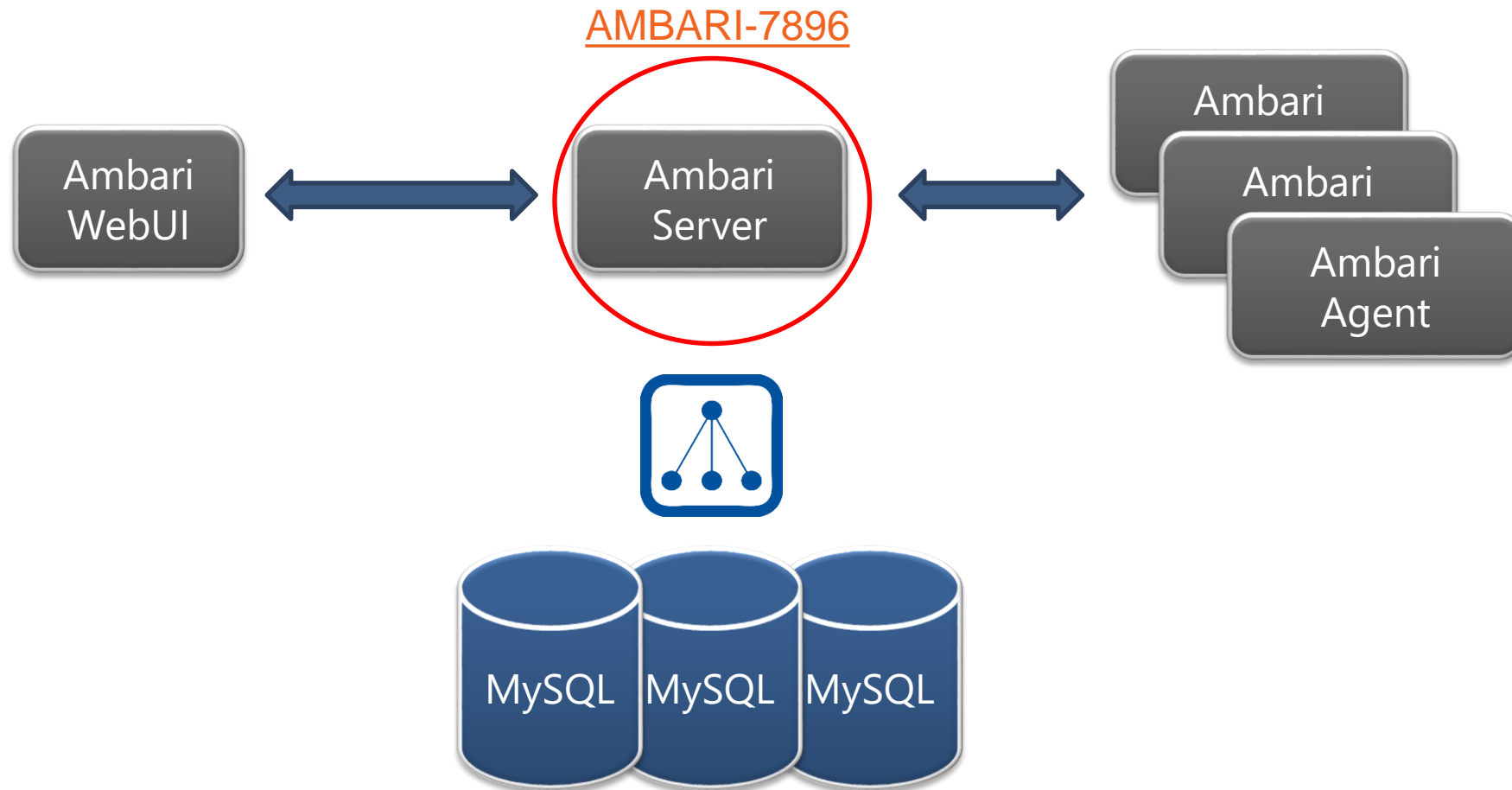
# Arquitectura de HIVE HA



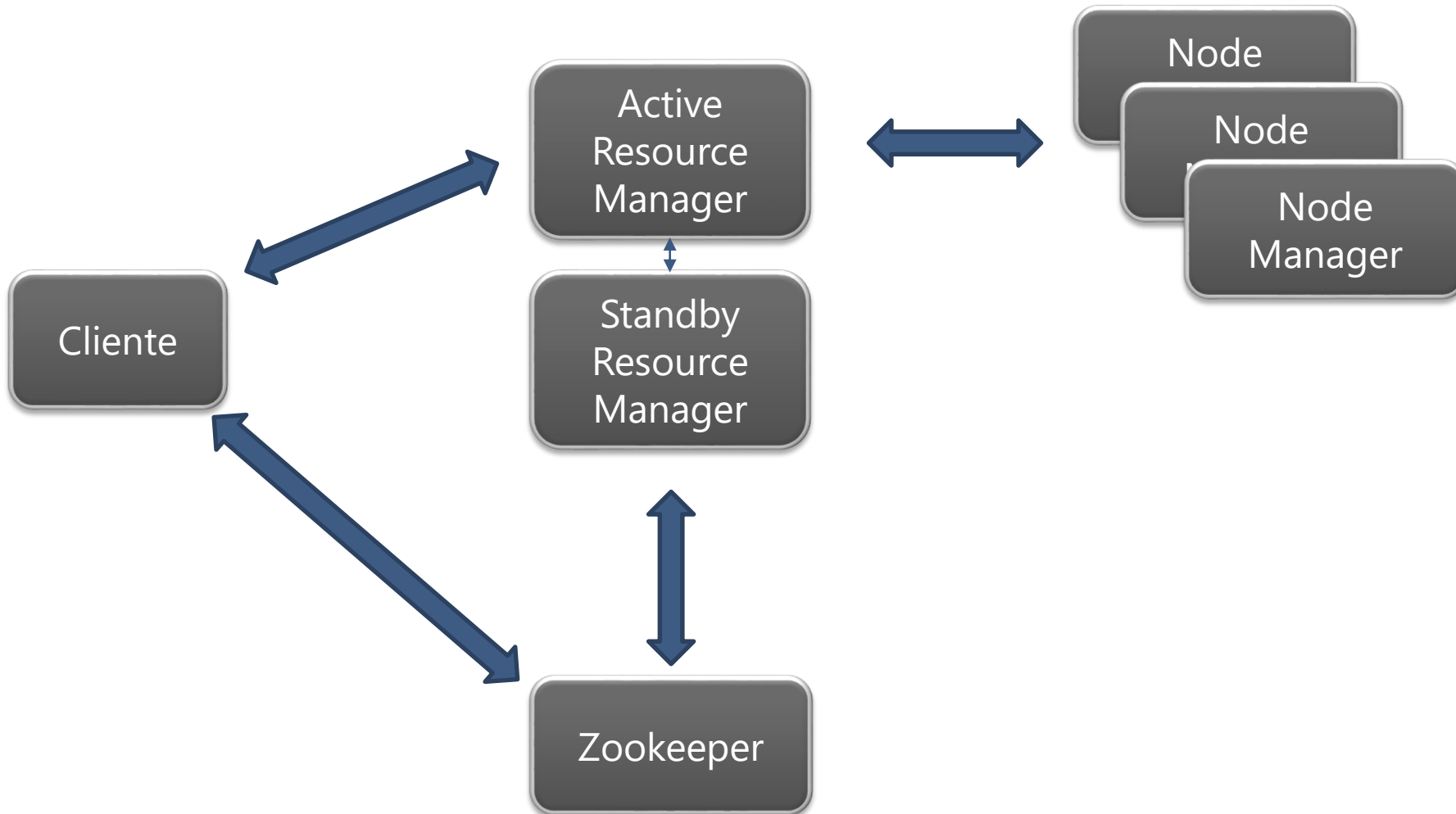
# Arquitectura de HBASE HA



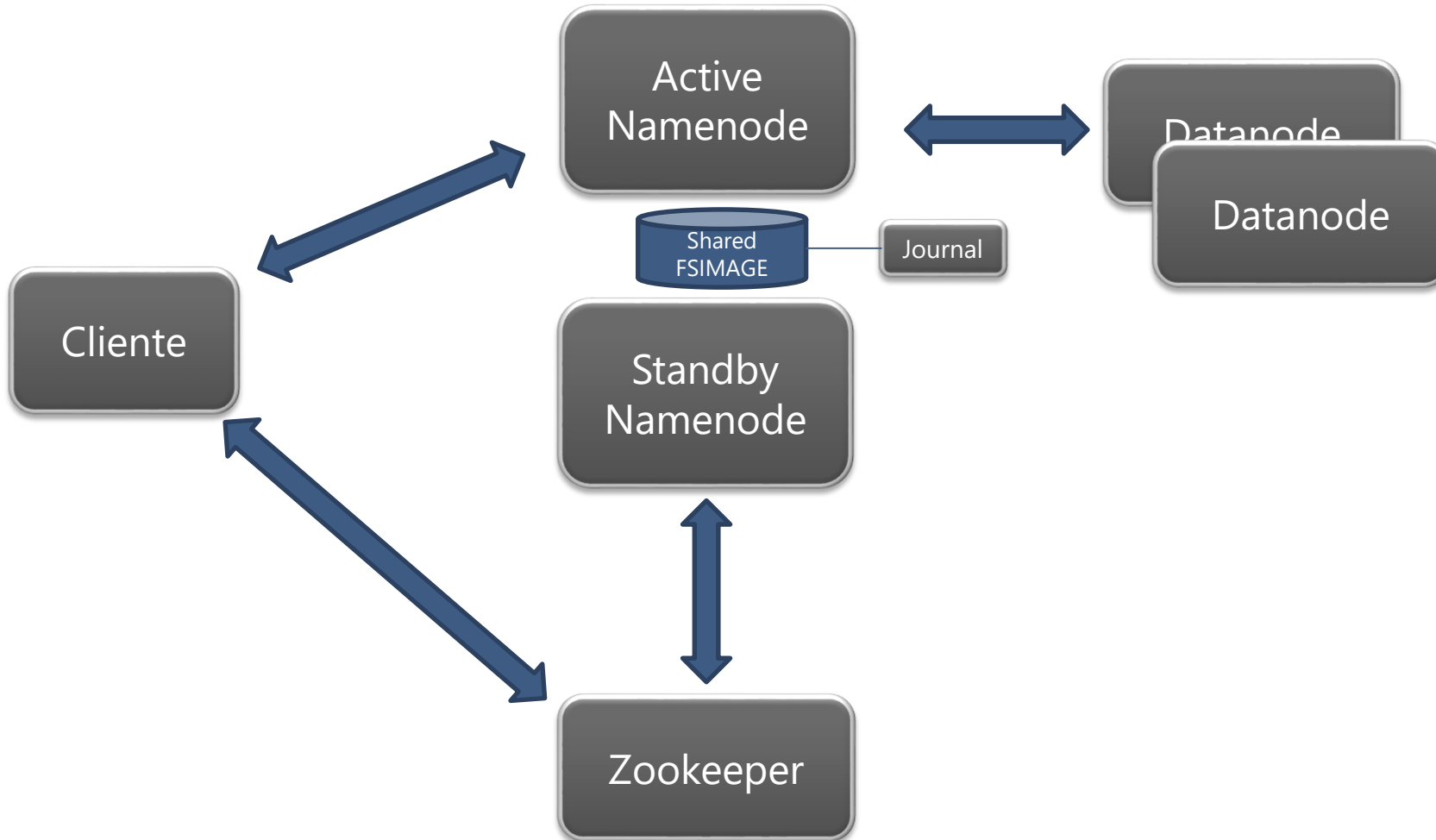
# Architecture de AMBARI HA



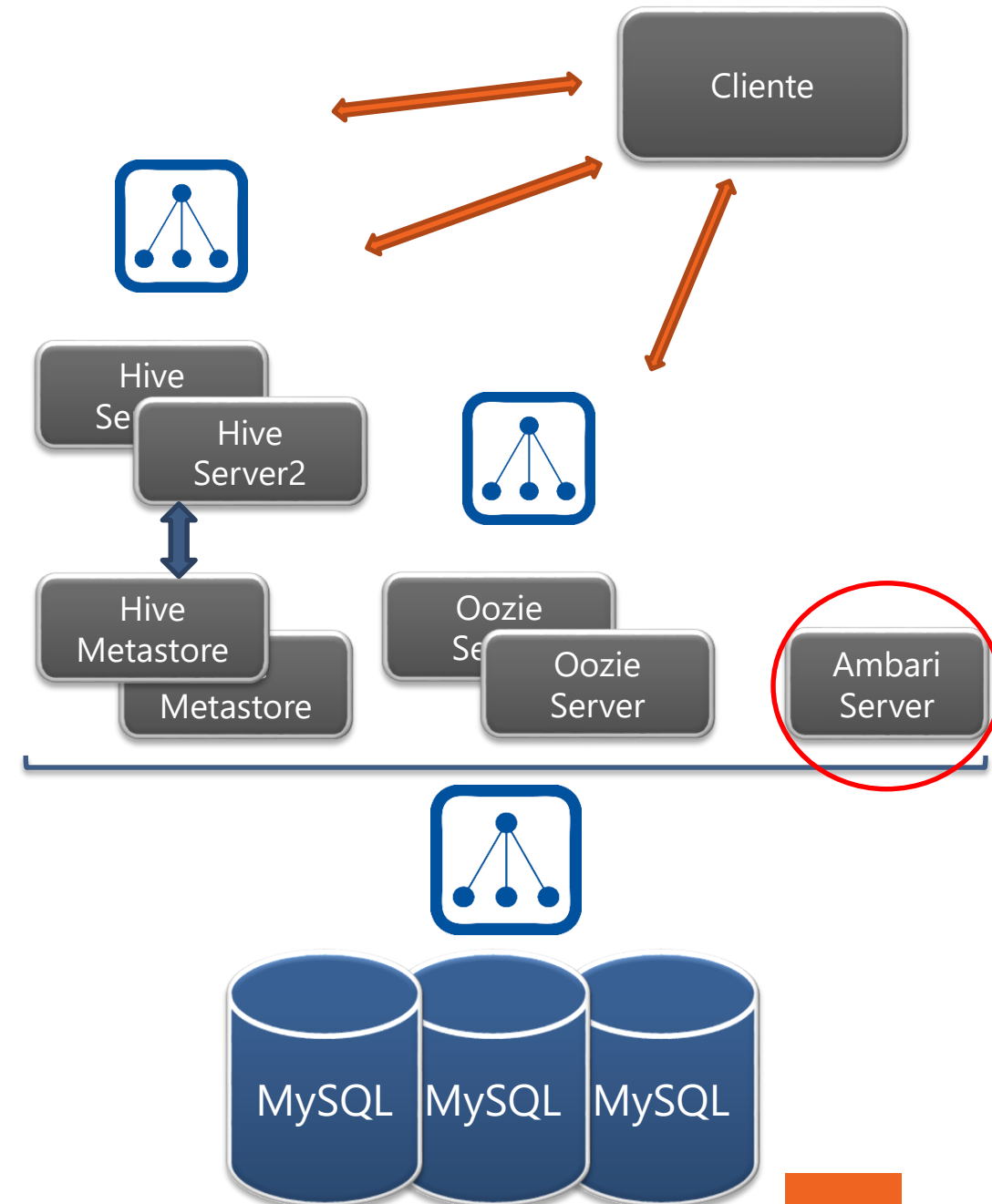
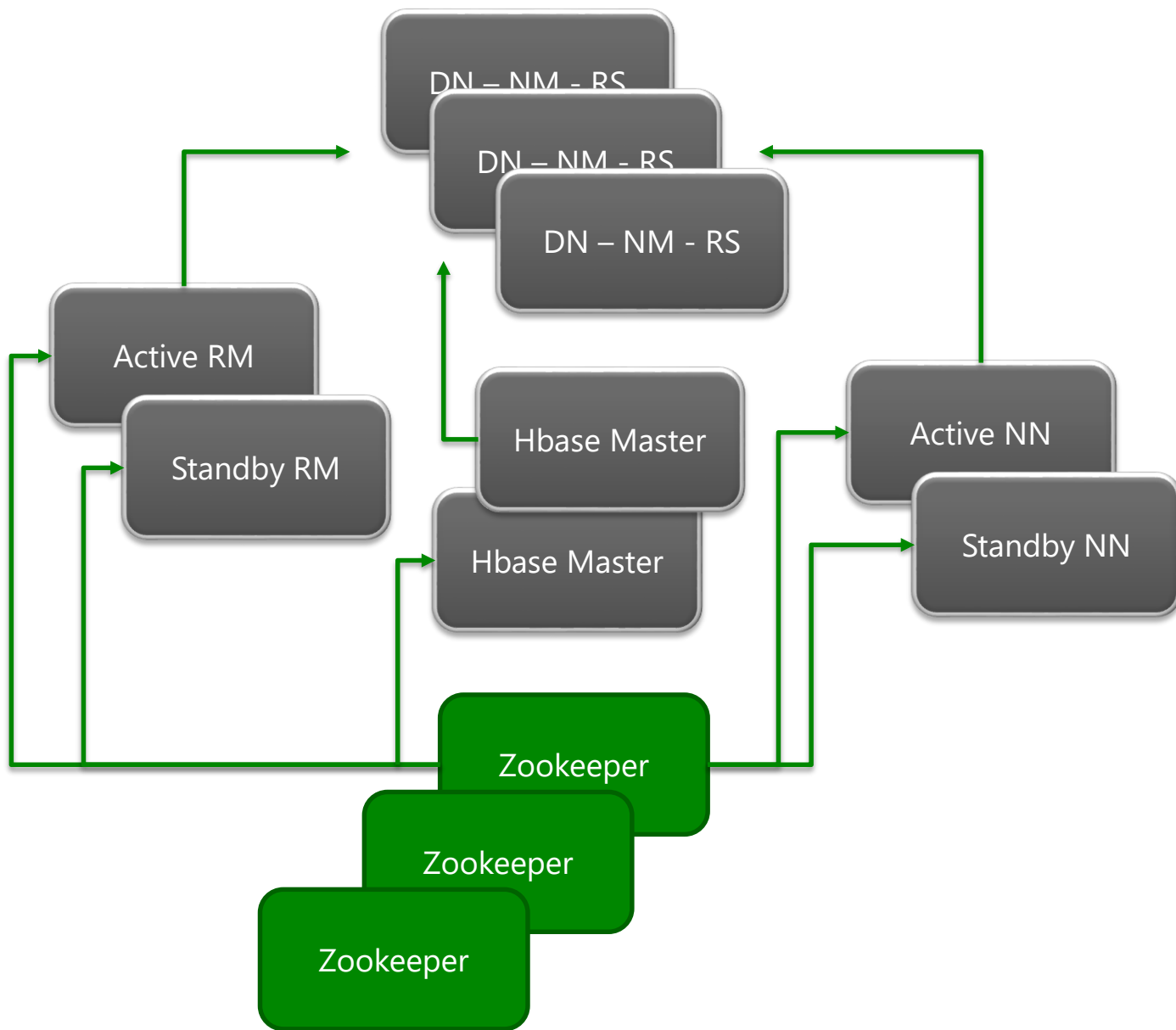
# Arquitectura de YARN HA



# Arquitectura de HDFS HA

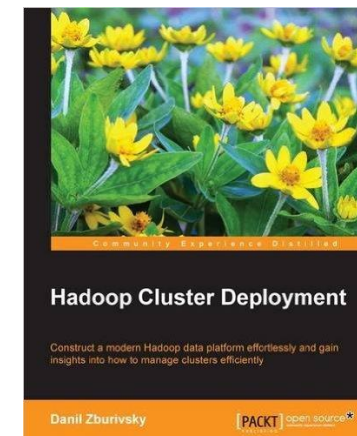
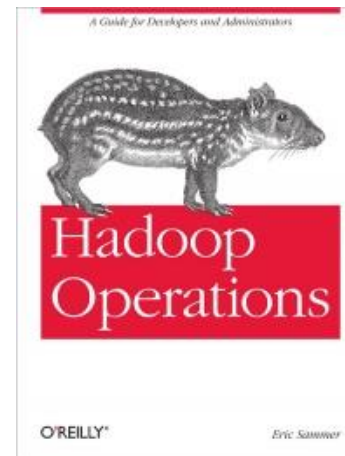
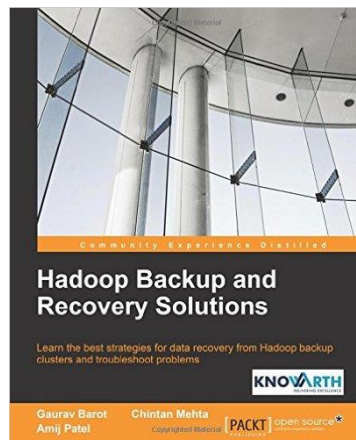
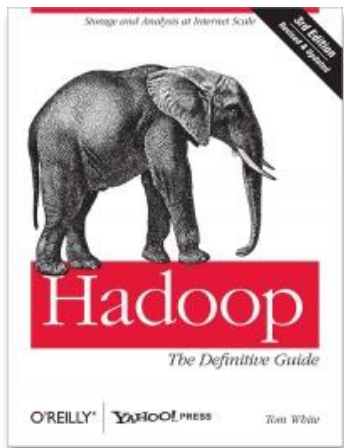


Armando todo el  
rompecabezas



# Bibliografia y Referencias

- ➔ <http://www.pearsonhighered.com/samplechapter/0130893552.pdf>
- ➔ <http://www.cisco.com/networkers/nw04/presos/docs/RST-2514.pdf>
- ➔ <https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HDFSHighAvailabilityWithQJM.html>
- ➔ <http://blog.cloudera.com/blog/2014/03/inside-apache-oozie-ha/>
- ➔ [http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.3.0/bk\\_hadoop-ha/content/ch\\_HA-ResourceManager.html](http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.3.0/bk_hadoop-ha/content/ch_HA-ResourceManager.html)





# Conclusiones... preguntas y respuestas

Presented by



Platinum Sponsor



Gold Sponsors



Silver Sponsors



Community Supporters



**Nos vemos en...**  
**[dataops.la/2016](http://dataops.la/2016)**