

# Open Source + AZURE

# Agenda

- Azure Nube Abierta
- Open Source en Azure
- ¿Qué es BigData?
- Demo: Ecosistema Hadoop

# Azure Nube Abierta

- Colección de servicios integrados en la nube
- Mayor productividad
- Una plataforma abierta y flexible
- Amplía la infraestructura de TI existente

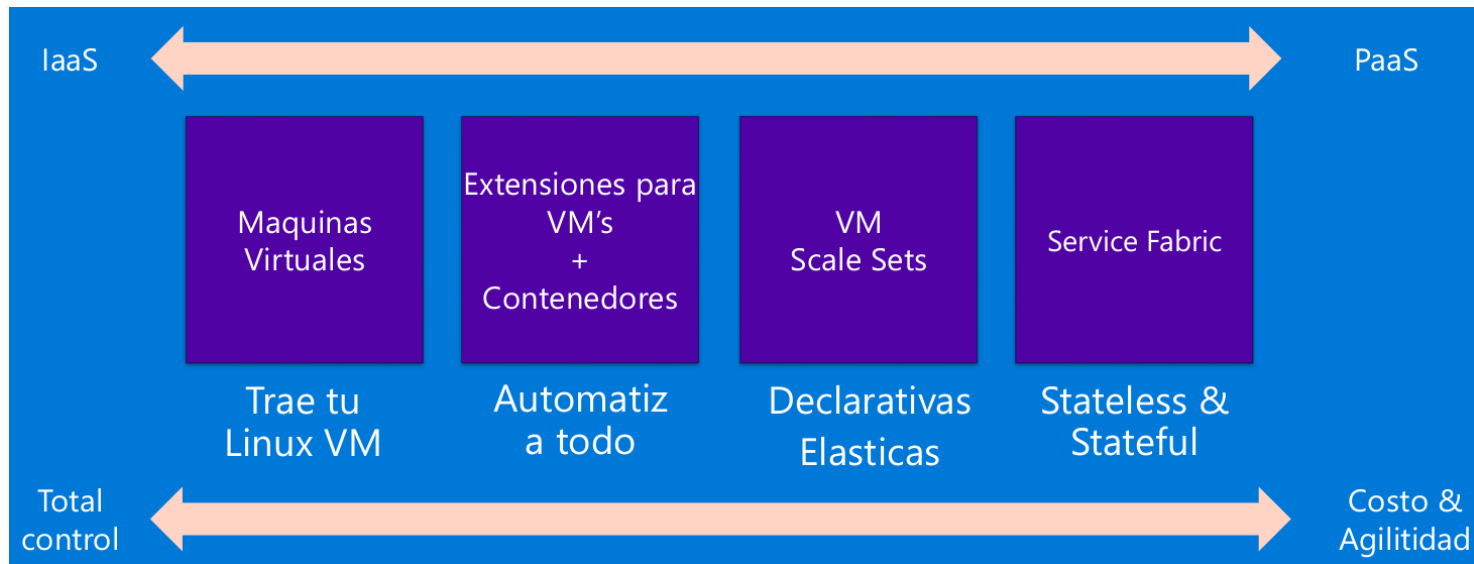


# Azure Nube Abierta

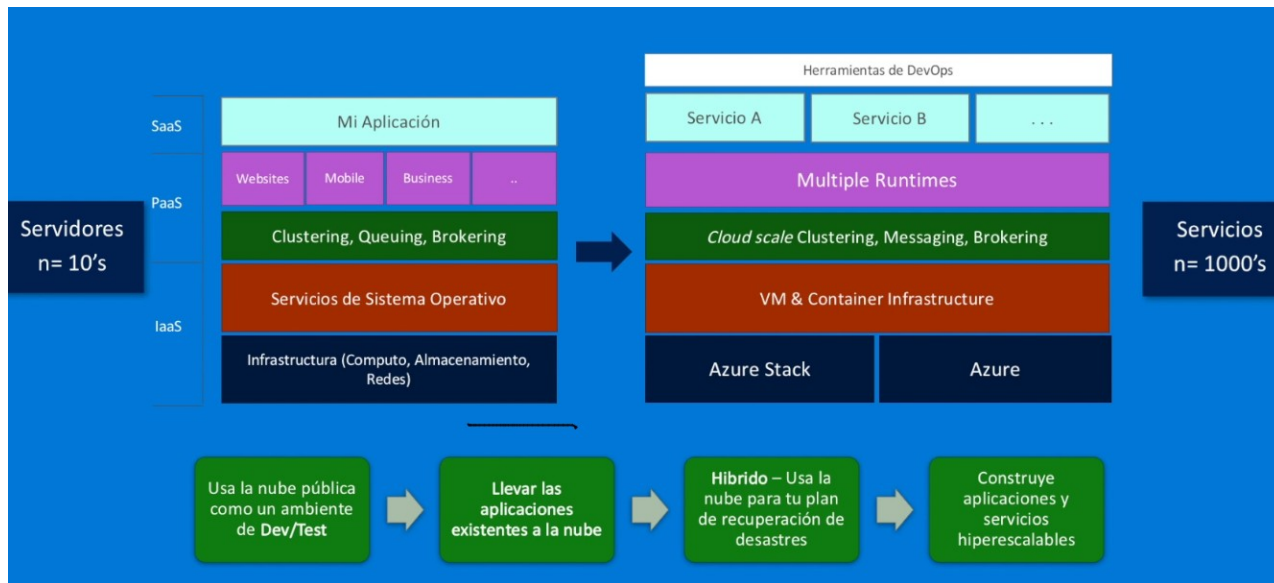
- Escalabilidad conforme se necesite y pago por uso
- Protección de datos
- Las aplicaciones pueden ejecutarse en cualquier lugar
- Toma de decisiones más inteligentes
- Un servicio en la nube de confianza



# Linux en Azure

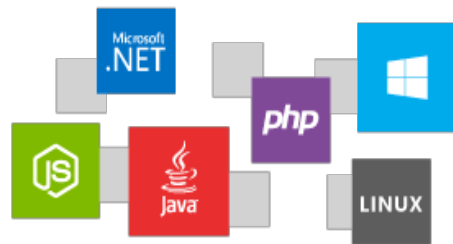


# Patrones de Empresas



# Open Source en Azure

- Distribuciones de Linux aprobadas
- Imágenes Open Source de la comunidad + DIY
- Integración de Docker
- Lenguajes de desarrollo



# Open Source en Azure

- DevOps CMTs
- Cargas de trabajo empresariales
- Soporte completo para Java





# Azure Nube Abierta

## Integrado por MS

Languages, Dev  
Tools & App  
Containers



CMS & Apps



Devices



Databases



Management



Operating  
systems



## Provisto por el ecosistema



**Jenkins**

Bring your own

Windows® **Web App Gallery**

Dozens of .NET & PHP CMS and Web applications



Via HTML/JS, cross-platform and native



**mongoDB**

DocDB  
DataStax



ANSIBLE



SALTSTACK

libcloud  
jclouds



**VM Depot**  
Microsoft Open Technologies



**FreeBSD**

Bring  
your own

# Open Source en Azure



- Ejemplos:
  - LAMP Servers
  - VM's + Contenedores
  - Moodle
  - Big Data en Azure

# LAMP Server

- Formas de desplegar
  - En una VM existente (clásico)
  - Extensión Custom Scripts y Azure CLI
  - Instalar imagen preconfigurada



# LAMP Server



- Desplegar con extensión Custom Scripts
  - Crear grupo de recursos y destino usando Azure CLI
  - Obtener el RM template de URL válida (github, Azure storage, etc)
  - Editar y desplegar el esquema

# LAMP Server



- Pre-requisitos
  - Instalar Azure CLI 2.0
  - Opcionales para editar RM template:
    - Credenciales, Storage accounts, DBs, VPCs

# LAMP Server



- Instalar imagen preconfigurada

# LAMP Server

Microsoft Azure Everything > Hardened Lamp on Centos 7.3

Everything

Filter

LAMP

Results

NAME	PUBLISHER	CATEGORY
LAMP	Bitnami	Compute
Hardened Lamp on Centos 7.3	Cognosys Inc.	Compute
Hardened LAMP on Ubuntu 14.04 LTS- Prof LIC	Cognosys Inc.	Compute
Hardened LAMP on Ubuntu 16.04 LTS-Professional-LIC	Cognosys Inc.	Compute
Xenofile for Azure - Ubuntu 14.04 LAMP stack	Your Shop Online	Compute
Hardened Lamp on Centos 7.3-Enterprise-LIC	Cognosys Inc.	Compute

**Hardened Lamp on Centos 7.3**  
Cognosys Inc.

This Image is made specially for Enterprise Customers who are looking for deploying a secured **Lamp** installation instead of just putting up a vanilla install.

**Disclaimer:**The respective trademarks mentioned in the offering are owned by the respective companies.

**For Stack specific support :** [Contact Stack Developer Team](#)(Since we do not own the IP for the stack,Stack developers should be contacted for any support)

**For Image related support:** [Contact SecureAnyCloud Team](#)

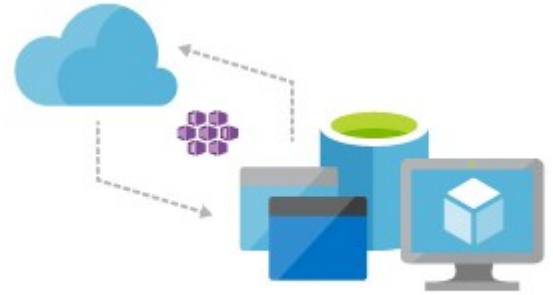
**Installation Instructions for CentOS**

- 1.Application URL:** Access the application via a browser at <http://PublicDNS>
- 2.Default installation path:** will be on your web root folder "/var/www/html".
- 3.Default ports:**
  - Linux Machines: SSH Port – 22
  - Http: 80
  - Https: 443
  - Mysql ports: By default these are not open on Public Endpoints. Mysql :3306
- 4.Database Login Details :**

**MYSQL Username :** root || **MYSQL Password :** Passw@rd123

# VM's + Contenedores

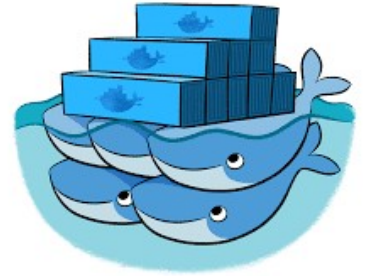
- Servicio de Contenedores de Azure
- Escalar y orquestar
- Usa herramientas abiertas que ya conoces
- Migrar la carga de contenedores dentro y fuera de Azure



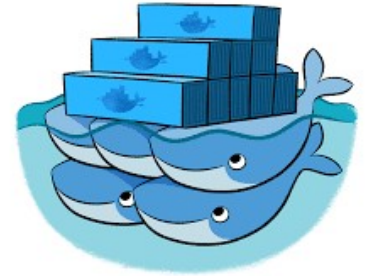


# VM's + Contenedores

- Extensión Docker VM de Azure
  - Docker Compose
  - Docker Machine
  - Docker Swarm Cluster



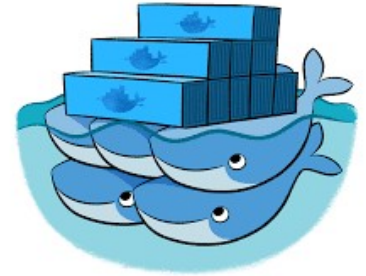
# VM's + Contenedores





- Formas de desplegar
  - Desde el portal con **Azure Container Service**
  - Usando el ACS desde el CLI de Azure
  - Imágenes del Market: Docker EE



# VM's + Contenedores











- Desde el portal con Azure Container Service





Microsoft Azure New > Azure Container Service > Create Azure Container Service > Summary


Create Azure Container Ser...  


Summary  




1 Basics  
Done 

2 Master configuration  
Done 

3 Agent configuration  
Done 

4 Summary  
Azure Container Service 

 Validation passed

Basics

Subscription	Azure Pass
Resource group	nlabsdockerdemo
Location	East US 2
Orchestrator	Swarm

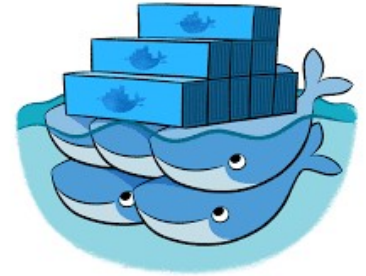
Master configuration

DNS name prefix	nlabsdockerdemo
User name	dockeruser
SSH public key	ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQACxM1UMIrNkgr...
Master count	1
VM diagnostics	Enabled

Agent configuration

Agent count	1
Agent virtual machine size	Standard DS2

# VM's + Contenedores



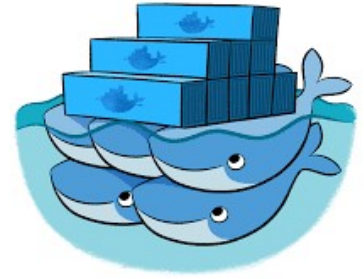
- Usando el ACS desde AZ CLI

1. Obtener uno de las plantillas disponibles en github
2. Logearse en el Azure CLI y crear el Resource Group
3. Editar `azuredeploy.JSON` , `deployparameters.JSON` y desplegar

```
> az group deployment create -g RESOURCE_GROUP -n DEPLOYMENT_NAME --template-uri  
    TEMPLATE_URI --parameters @azuredeploy.parameters.json
```

# VM's + Contenedores

- Desplegar desde el Market Place



[Products](#) > [Docker EE for Azure \(Basic\)](#)[GET IT NOW](#)**Categories**[Compute](#)[Developer tools](#)**Legal**[License Agreement](#)[Privacy Policy](#)

# Docker EE for Azure (Basic)

Docker, Inc.

[Overview](#)[Plans](#)

An integrated, easy-to-deploy environment for building, assembling, and shipping applications

- Docker EE for Azure (Basic) bootstraps all of the recommended infrastructure to start using Docker on Azure automatically so users don't need to think about rolling own instances or setup security groups or Load balancer

**Highlight Features**

- Docker optimized for Azure IaaS services
- One-click to deploy Docker on Azure
- Easily upgrade to Docker Datacenter or extend and integrate into the operational tooling of your choice to manage, monitor, gain visibility and control the resources of your app containers.
- Quick and Easy setup of Docker on Azure. Docker for Azure installs a Swarm of Docker Engines secured end to end with TLS by default, and is integrated with Azure VM Scale Sets.

**Achieve cloud portability for your apps: run them on any cloud provider or on-premises**

**Docker EE for Azure (Basic) will be offered on hourly billing**

# Moodle en Azure



- **M**odular **O**bject **O**riented **D**ynamic **L**earning **E**nvironment
- Licencia Pública General GNU
- La plataforma de aprendizaje más usada en el mundo
  - 79 millones de usuarios, 223 países, 71.600 sitios
  - Proyectos pequeños y grandes: Universidades, Shell, Microsoft



# Moodle en Azure



- Características principales:
  - Gestión de cursos y actividades
  - Perfiles de estudiantes y profesores
  - Variedad de plugins
  - Potentes herramientas de edición
  - Moodle Mobile

# Moodle en Azure



- Requisitos de Moodle:
  - Linux es recomendado
  - PHP 5.4 o superior
  - BD: MySQL, Postgres, MariaDB, MSSQL u Oracle

# Moodle en Azure



- Despliegue
  - Linux VM, LAMP, Imagen de VM Depot
- Almacenamiento
  - En discos de las VM, limitado.
  - Compartido: NFS, Glusterfs, SAMBA, rsync, btsync
  - Unidades propias

# Moodle en Azure



- Generar versiones propias
  - Crear VMs y configurarlas desde la interfaz de Azure
- Clonación
  - Se pueden clonar los VHD con o sin SO
- Seguridad de los datos
  - Redundancia local o global
  - Sistema de réplica y despliegue mediante agente

# Moodle en Azure



- Alta disponibilidad
  - VPCs y cross-region replication
- Escalabilidad
  - Definir scale sets para escalar a demanda
- Validar Performance
  - Herramientas de monitoreo y auditoría de Azure

# Moodle en Azure

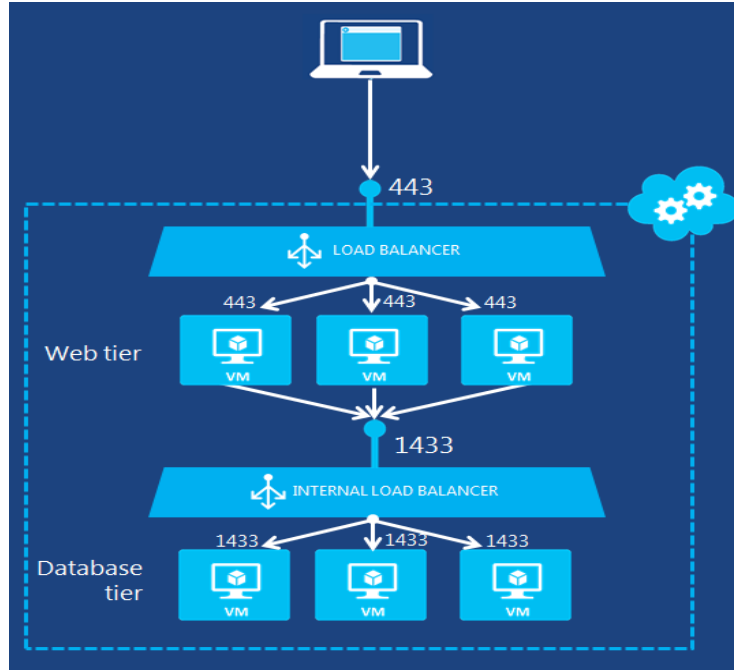


- Plugin Office 356 y acceso por Azure AD
  - Integrado por Microsoft
- Clonación
  - Se pueden clonar los VHD con o sin SO
- Seguridad de los datos
  - Redundancia local o través de regiones distintas
  - Sistema de réplica y despliegue mediante agente

# Moodle en Azure



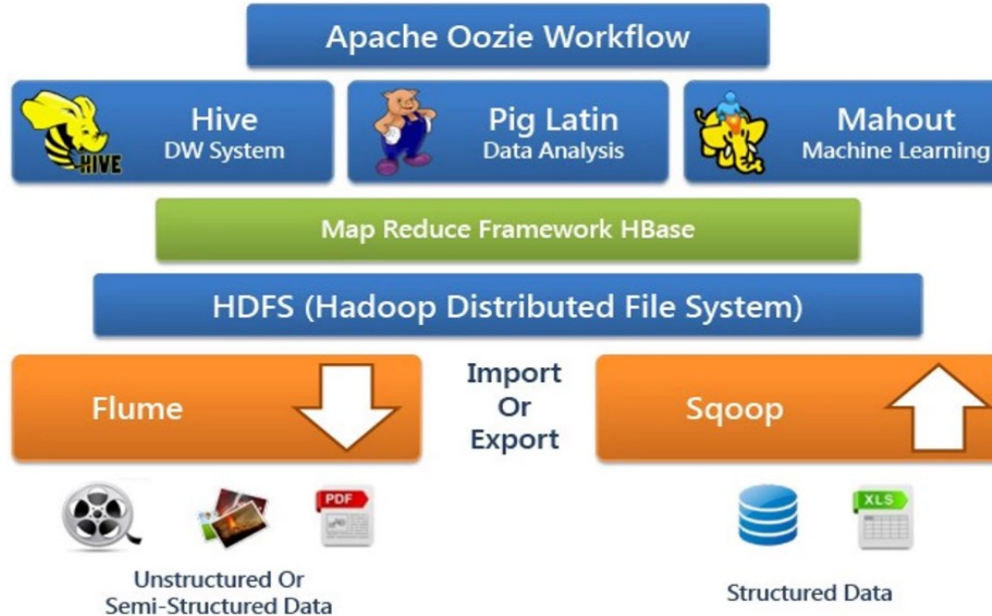
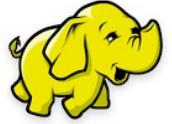
Arquitectura Cloud:  
Ignacio Escudero  
Open Sistema,  
España.



# Demo: Introducción al Ecosistema Hadoop en Azure

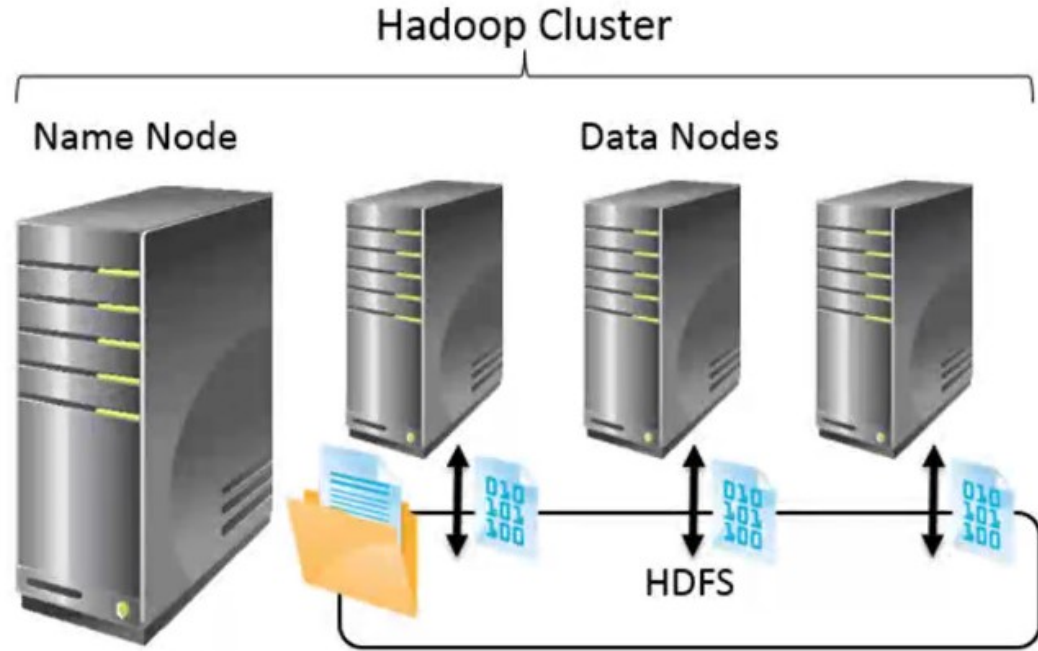
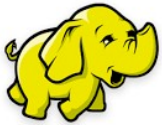


# Ecosistema Hadoop

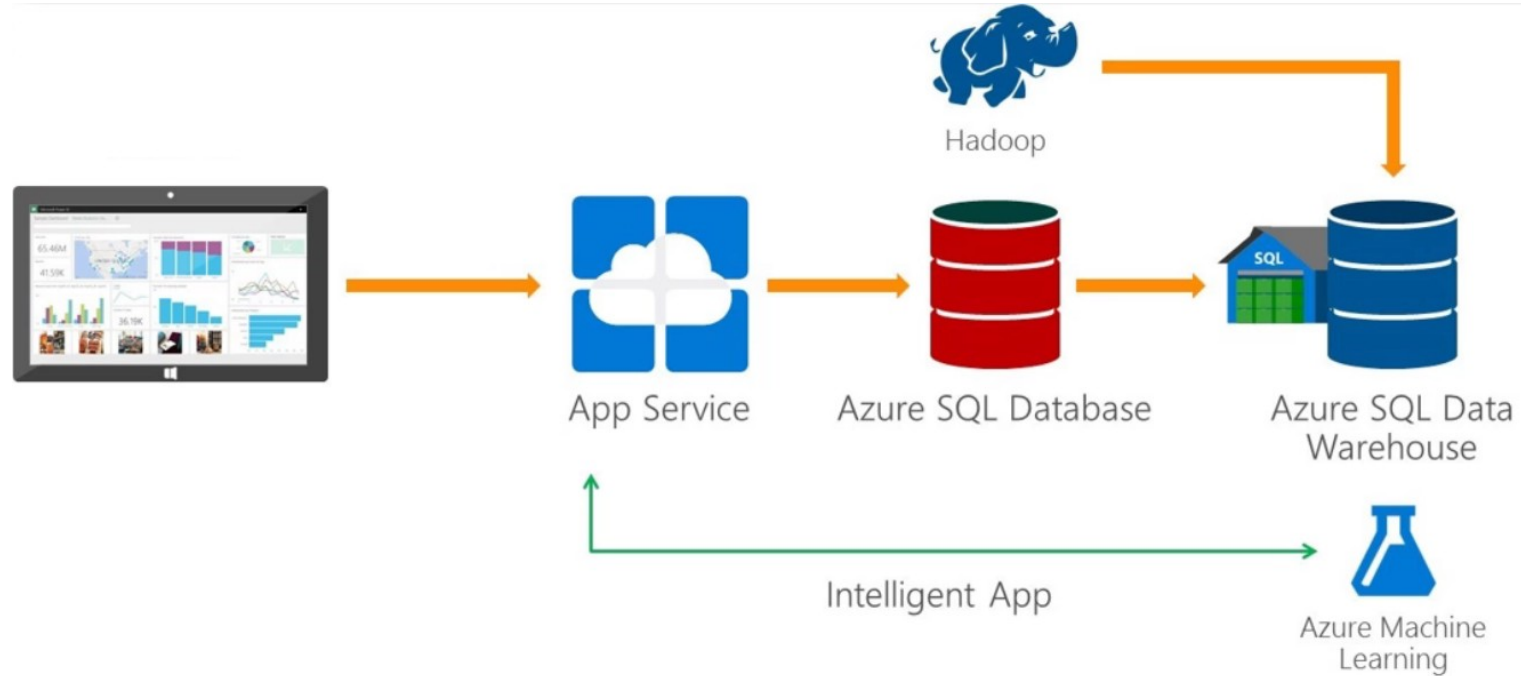


Crédito imagen: <http://jennyxiao Zhang.com/6-things-you-need-to-know-about-hadoop/>

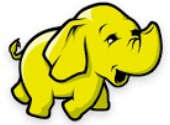
# Big Data en Azu



# Big Data en Azure



# Big Data Storage



DATA WAREHOUSE	vs.	DATA LAKE
structured, processed	<b>DATA</b>	structured / semi-structured / unstructured, raw
schema-on-write	<b>PROCESSING</b>	schema-on-read
expensive for large data volumes	<b>STORAGE</b>	designed for low-cost storage
less agile, fixed configuration	<b>AGILITY</b>	highly agile, configure and reconfigure as needed
mature	<b>SECURITY</b>	maturing
business professionals	<b>USERS</b>	data scientists et. al.

# Azure HDInsight

- Hadoop de la distribución de **HDP**
  - Implementa clústeres administrados
  - Confiabilidad y disponibilidad
  - Seguridad y gobierno de nivel empresarial con Active Directory.

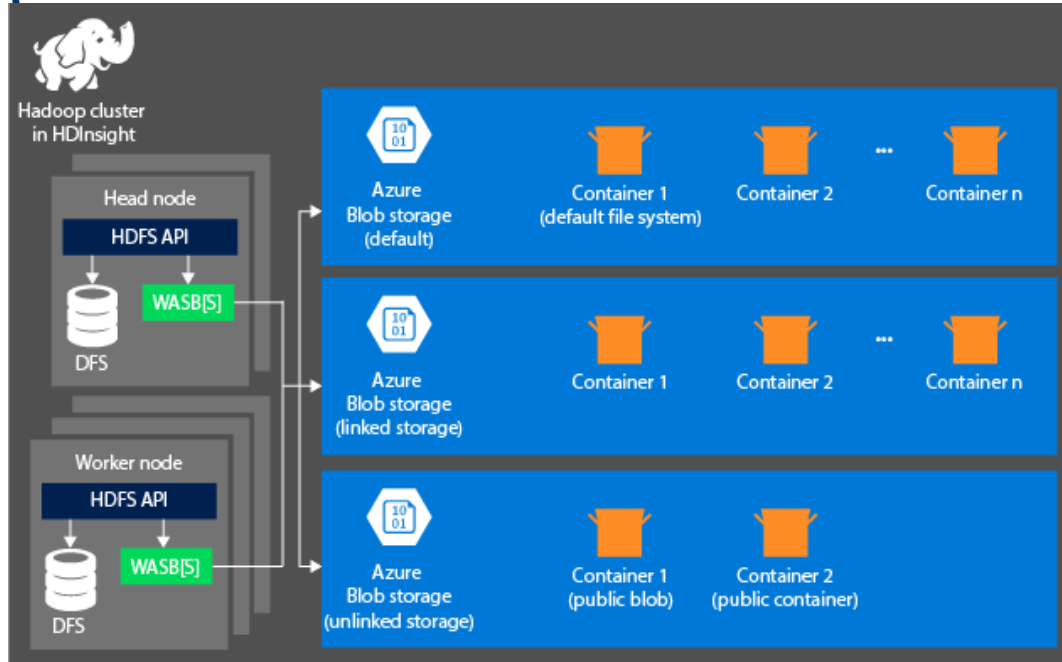


# Azure HDInsight

- Tipos de clúster:
  - Hadoop
  - Spark
  - Hbase
  - R Server
  - Storm



# HDInsight



# HDInsight Demo

Categoría	Hadoop en Linux
SO del clúster	Ubuntu 12.04 Long Term Support (LTS)
Tipo de clúster	Hadoop, Spark, HBase, Storm
Implementación	Portal de Azure, CLI de Azure, Azure PowerShell
Interfaz de usuario del clúster	Ambari
Acceso remoto	Shell seguro (SSH), API de REST, ODBC, JDBC



---

# Gracias

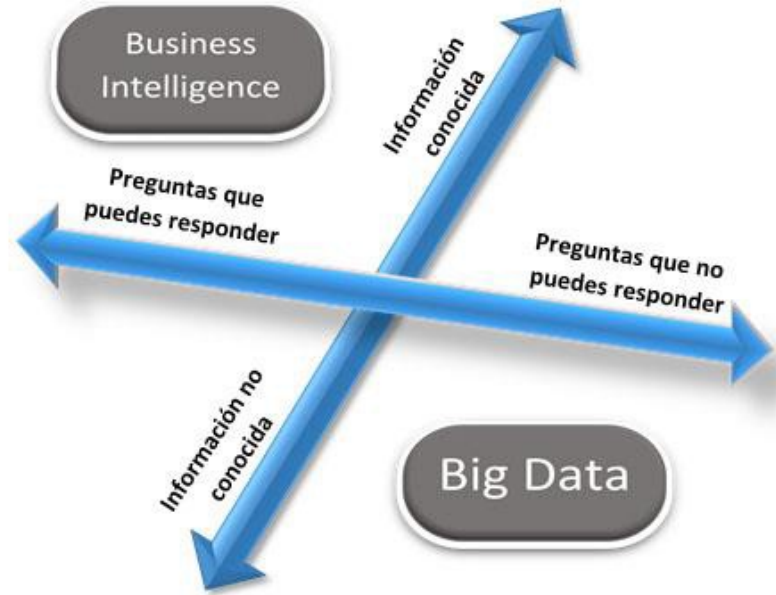
# BUSINESS INTELLIGENCE

- Objetivo: búsqueda de respuestas a preguntas conocidas para la toma de decisiones.
- Muestra el funcionamiento del negocio mediante el análisis de datos estructurados (Datawarehouse).
- Información centralizada en un servidor central.

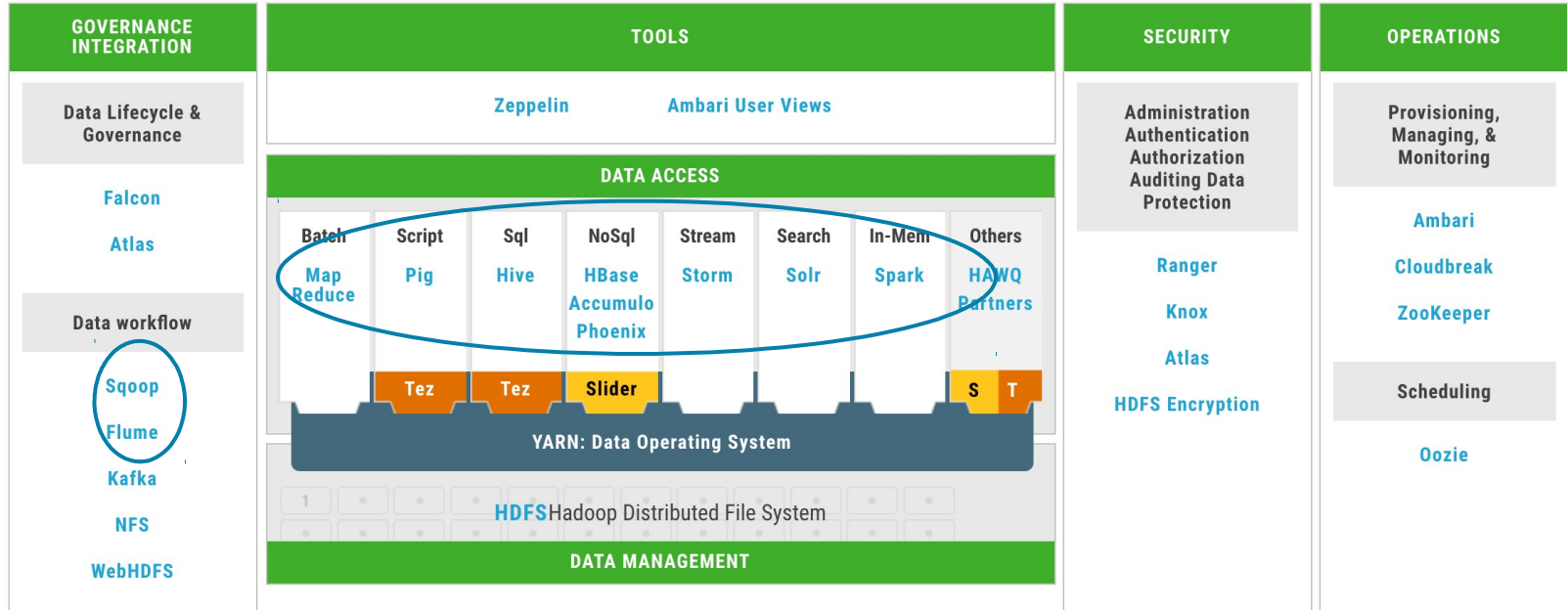
# DIFERENCIAS

- Almacenamiento distribuido.
- Análisis centrado en torno al dato.
- Fuentes y formato de los datos.
- Datos históricos o en tiempo real.
- Procesamiento en paralelo.

**BI complementa a  
BD**



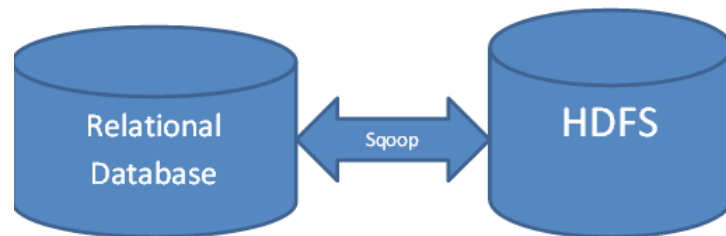
# TECNOLOGÍAS



# SQOOP

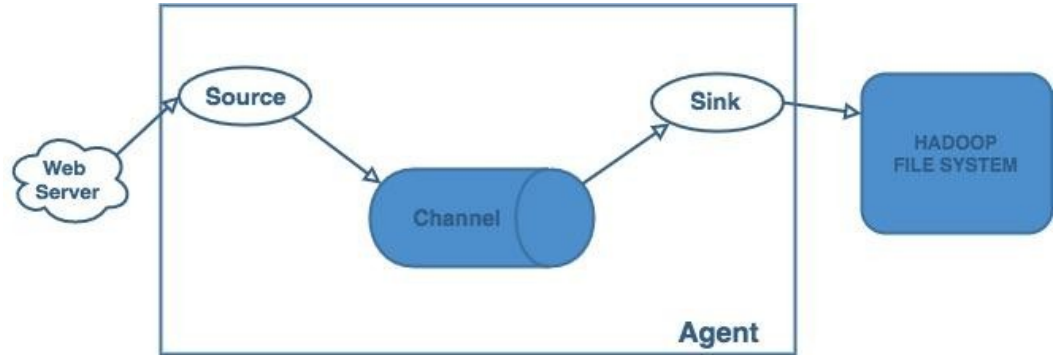


- Importar/Exportar datos
  - Hive
  - HDFS
  - HBase
- Crear **jobs** para ejecución posterior.



# FLUME

- **Event** : payload + atributos
- **Agent**: proceso JVM
  - **Source**: recibe los eventos externos.
  - **Channel**: almacena los eventos.
  - **Sink**: envía eventos al hdfs.



# PIG



- **Pig Latin:** Lenguaje de scripting
- Modo de ejecución: local, tez, mapreduce
- Funciones customizadas: **UDFs** (User Defined Functions) programadas en Java, Python, JavaScript, Ruby, Groovy.

# HIVE



- Infraestructura de data warehouse basado en Apache Hadoop.
- **HiveQL:** lenguaje para realizar consultas estilo SQL.
- **UDFs:** Extensión de las funciones definidas por el usuario.
- Almacenamiento
  - **Particiones:** Cada tabla puede se sub-divide en particiones.
  - **Buckets:** A su vez una partición puede dividirse agrupando por los valores de las columnas.
  - **External/Managed table**



# HBASE

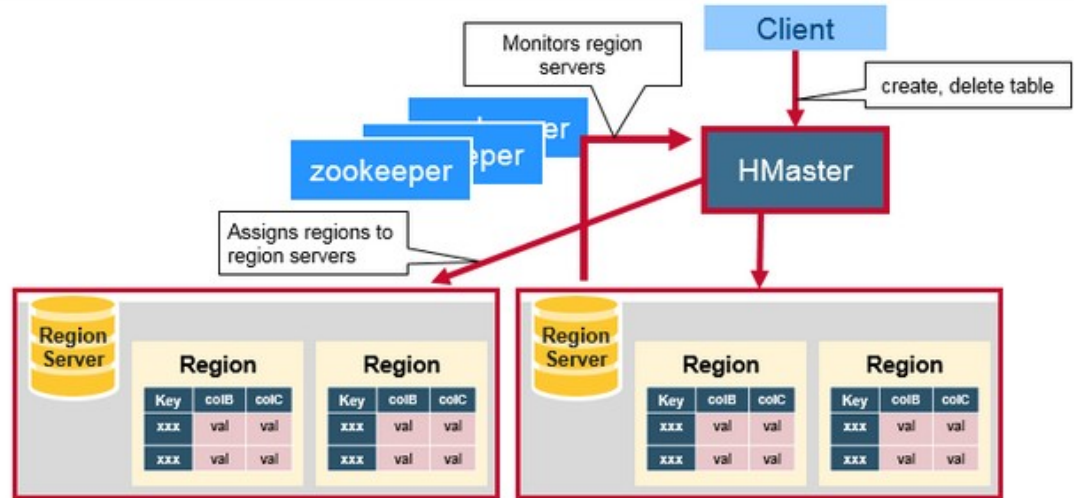


- Base de datos no relacional escalable horizontalmente y distribuida, es orientada a columnas.
- Está basada sobre Big table.
- Divididas por regiones en familias de columnas.
- Son almacenadas como archivos en el HDFS.

# HBASE



- HBase Master:
  - Asignación de región.
  - DDL (create, delete tables)
  - Balancea la carga.
- Region servers:
  - Operaciones reads/writes.
- Zookeeper: mantiene el estado de los nodos.



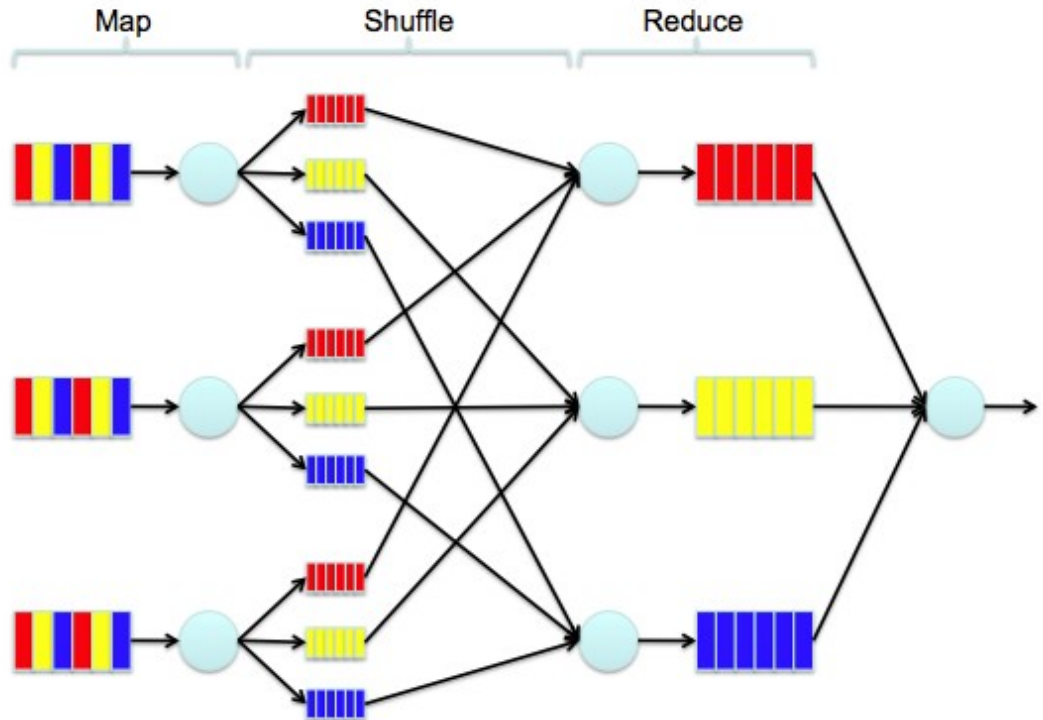
# SOLR

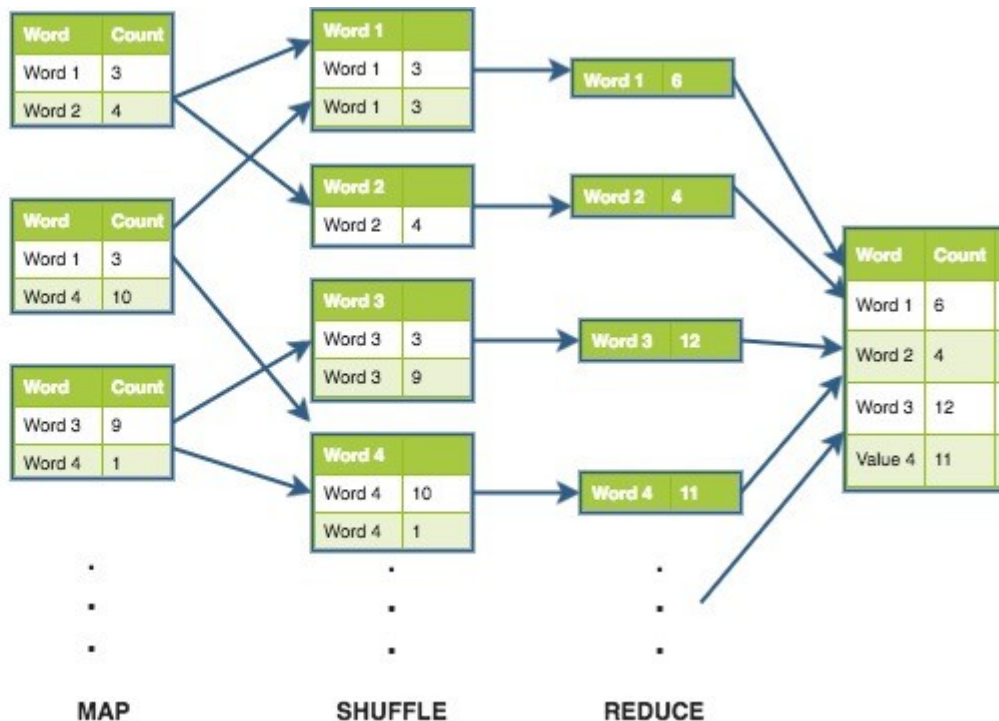
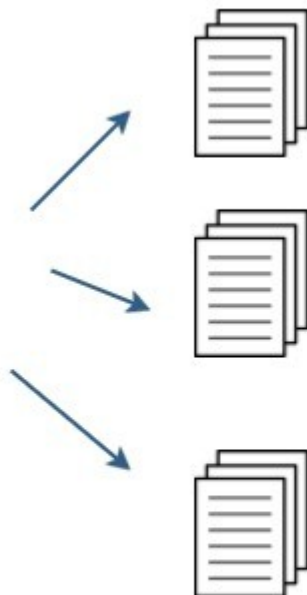


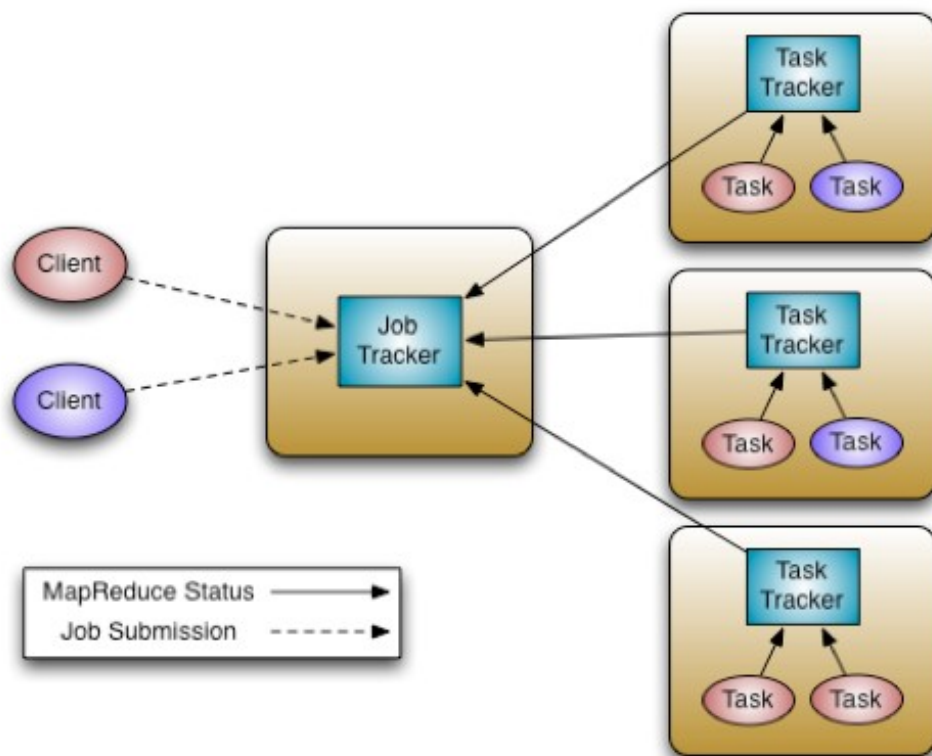
- **Búsquedas** sobre gran cantidad de datos en real time.
- Los tipos de documentos que se pueden **indexar** son XML, JSON, CSV.
- API HTTP para consultas.
- Es escalable y tolerante a fallas.

# MAPREDUCE

- **Map:** Aplica una función a los valores de entrada.
- **Shuffle:** Agrupa por clave.
- **Reduce:** Procesa la salida de los mappers y retorna el valor final.



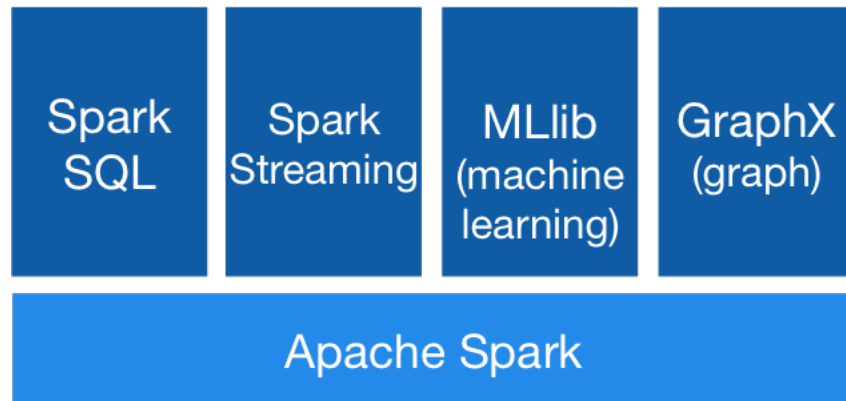




# SPARK

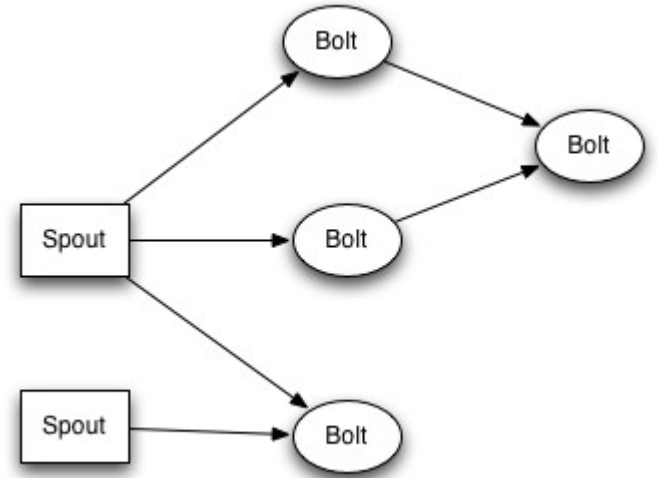


- Utiliza operaciones en memoria.
- Provee una API en Java, Scala, Python o R.
- **RDDs:** Resilient Distributed Datasets.  
Conjunto de datos tolerantes a fallos que son procesados en paralelo.
- Operaciones: Transformaciones o acciones.



# STORM

- Procesamiento de datos en real time.
- **No termina** hasta que se mata el proceso.
- Monitoreo, detección de fraude.



TOPOLOGY



# HANDS-ON

1. MapReduce: Word Count
2. Multas de tránsito de la Intendencia
  - a. cantidad de multas por esquina.
  - b. cantidad de multas por rangos horarios.
  - c. total de dinero recaudado.

Fuente de datos: <http://www.impo.com.uy/cgi-bin/bases/consultaBasesBS.cgi>

