

Big Data

APLICACIONES: TECNOLOGÍA E
INFRAESTRUCTURA

DEFINICIÓN



La arquitectura de Big Data consiste en el diseño de sistemas y modelos para el tratamiento de grandes volúmenes de datos de diferentes tipos y orígenes con el fin de transformarlos en información que permita la mejor toma de decisiones.

Es un proceso que consiste en analizar métodos de análisis no convencional para volúmenes amplios de datos.

Para poder valorar semejantes bases de información se diseñan de forma personalizada esquemas de trabajo y estructuras de información en las que con el apoyo de diferentes herramientas de software, se puede comprender de mejor manera los procesos relacionados a almacenamiento, gestión y procesamiento de datos.

CARACTERÍSTICAS

Escalabilidad, aumentar fácilmente las capacidades de procesamiento y almacenamiento de datos

Tolerancia a fallos, garantizar la disponibilidad del sistema, aunque se produzcan fallos en algunas de las máquinas

Datos distribuidos, los datos están almacenados entre diferentes máquinas evitando así el problema de almacenar grandes volúmenes de datos.

Procesamiento distribuido, el tratamiento de los datos se realiza entre diferentes máquinas para mejorar los tiempos de ejecución y dotar al sistema de escalabilidad.

Localidad del dato, tanto los datos como los procesos deben estar cerca para evitar las transmisiones por red que añaden latencias y aumentan los tiempos de ejecución.

3 CAPAS

Primera capa: Está relacionada al análisis y visualización; se centra en la visualización de los datos para su exploración y análisis mediante técnicas estadísticas, algoritmos de análisis predictivo, aprendizaje automático, etc.



Segunda capa: Está relacionada a gestión de los datos; se centra en la integración, el gobierno y la seguridad de los datos. Se trata de elegir los datos adecuados y que van a permitir un procesamiento eficiente, que tengan la calidad requerida y protegerlos de manera adecuada minimizando los riesgos de seguridad.



Tercera capa: Está relacionado al almacenamiento y procesamiento; se centra en el almacenamiento de los datos obtenidos y su procesamiento eficaz y eficiente acorde con las necesidades que tenemos.

ARQUITECTURA LAMBDA



Apareció en el año 2012 y se le atribuye a Nathan Marz. Su principal objetivo es tener un buen sistema tolerante a posibles fallas, tanto humanos como de hardware, escalable y que permita realizar escrituras y lecturas con baja latencia.

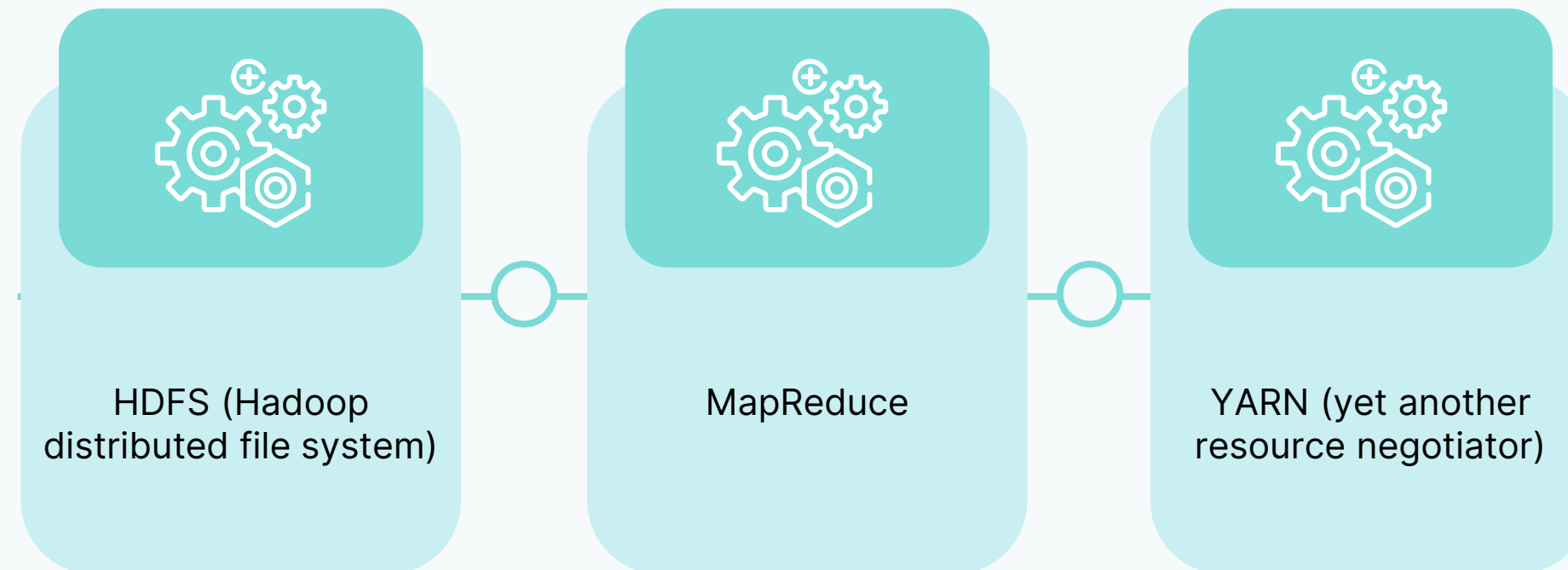
Algunas de sus principales características son las siguientes: la nueva información recogida se envía al speed layer, en la capa batch se gestiona la información sin modificar y los datos nuevos se añaden a los ya existentes.

ARQUITECTURA KAPPA



Aparece en el 2014 por Jay Kreps. Esta se centra en procesar datos exclusivamente como flujo, los datos de entrada fluyen por una capa en tiempo real y los resultados se colocan en la capa de servicio para consultas.

COMPONENTES



TECNOLOGÍAS INVOLUCRADAS



Hive



Spark



HBase



Phoenix



Sqoop



Kafka

TECNOLOGÍAS INVOLUCRADAS

Flume



Flume



Storm



Solr



Ignite



Hue



Mahout

DEFINICIÓN



Se refiere a la computación en la nube (cloud computing), es una tecnología que permite acceso remoto a softwares, almacenamiento de archivos y procesamiento de datos por medio de Internet, siendo así, una alternativa a la ejecución en una computadora personal o servidor local.

En el modelo de nube, no hay necesidad de instalar aplicaciones localmente en computadoras. La computación en la nube ofrece a los individuos y a las empresas la capacidad de un pool de recursos de computación con buen mantenimiento, seguro, de fácil acceso y bajo demanda.

BENEFICIOS

DISPENSAR EL HARDWARE

La computación en la nube dispensa el hardware. Cuando una empresa proporciona su propio software, debe ocuparse de los servidores. Estos servidores requieren un suministro de energía exclusivo y piezas de reemplazo. También es necesario configurarlos y supervisarlos en caso de que tengan problemas de rendimiento y requieran expertos de guardia para solucionarlos.

EXTREMADAMENTE SEGURA

Para lograr mayor seguridad, la computación en la nube es extremadamente segura, muchas veces superando los niveles de seguridad de la computación tradicional, permitiendo que las empresas atraigan y mantengan un equipo de seguridad cibernética de alta calidad. También permite implementar prácticas y tecnologías de seguridad de punta, orientadas por una visión más amplia de los patrones globales de amenazas en relación a aquellas de la mayoría de los gobiernos locales.

BENEFICIOS

CAPACIDAD DE UBICACIÓN

La computación en la nube tiene la capacidad de ubicar a todos en el mismo nivel. No importa si tienes decenas o millares de usuarios en la plataforma, la computación en la nube democratiza la aplicación de software corporativo. Con flexibilidad para aumentar o reducir lo que el cliente comparte en el pool rápidamente, algunas veces de modo automático, el usuario final, por lo general, no necesita saber lo que ocurre en la nube. El usuario necesita tan solo iniciar la sesión y trabajar en la tarea.



3 ROLES EN UN PROYECTO BIG DATA

DIRECTOR DEL PROYECTO

Responsable del mismo de principio a fin y aborda el ciclo completo del proyecto: organización, ejecución y control, cierre. Su objetivo final no es otro que alcanzar los objetivos del proyecto, en los plazos estimados, con la certificación de calidad y nivel de servicio requerido.

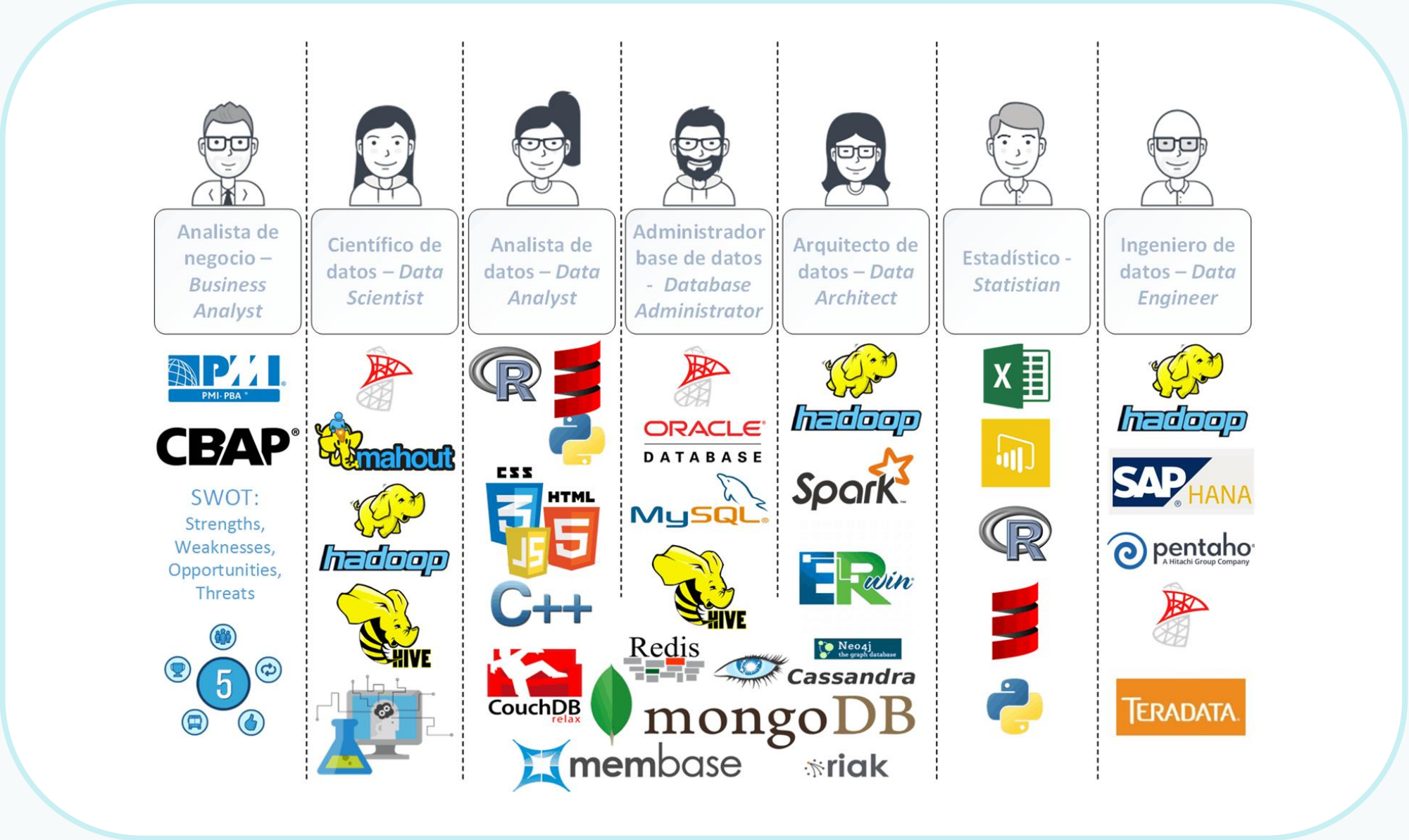
ARQUITECTO

El arquitecto, responsable de apoyar estratégicamente el diseño de la solución que vaya a implementarse dando recomendaciones para el planteamiento del proyecto.

CONSULTOR DE INTEGRACIÓN

Cumple la función de desarrollar y construir, el modelo y requerimientos de los clientes, que deben plasmarse en el proyecto. A su vez, es responsable del cumplimiento del plan, calidad y nivel de servicio requerido por el proyecto, así como de coordinar y solucionar los puntos críticos que afecten a lo planificado.

ROL Y PERFIL





Fundación Romero

www.fundacionromero.org.pe