

Big Data desde cero

Contexto, tecnologías y aplicación

ANTECEDENTES A BIG DATA

- La informática tradicional: de la calculadora a la computación distribuida
- Antecedentes en Big Data: limitaciones de la computación distribuida
- La clave de Big Data: la localización del dato

DE LA CALCULADORA A LA COMPUTACIÓN DISTRIBUIDA

- El **ábaco** como antecedente de la informática.



- De esos ámbitos, el **Tratamiento de Datos** será el foco.



- La informática evolucionó en **múltiples ámbitos y finalidades.**



DE LA CALCULADORA A LA COMPUTACIÓN DISTRIBUIDA

- Tratamiento de datos se puede entender como el proceso de:



- El dispositivo más sencillo que ilustra este proceso es la **calculadora**.

DE LA CALCULADORA A LA COMPUTACIÓN DISTRIBUIDA

- El Tratamiento de Datos es fundamental en el mundo empresarial.

- Ejemplo: **Analizar facturas**

1. Recolectar

2. Almacenar

3. Procesar

4. Visualizar



Facturas día anterior



Programa informático contabilidad



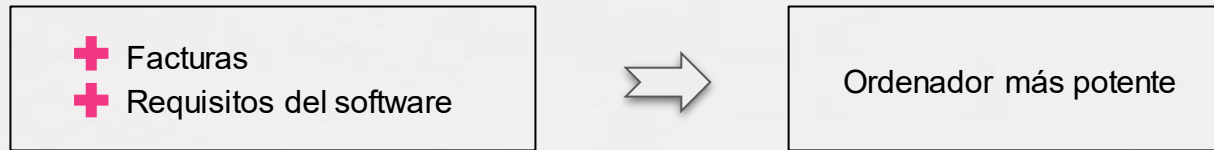
Indicación datos a calcular



Informes, indicadores, gráficos, ...

DE LA CALCULADORA A LA COMPUTACIÓN DISTRIBUIDA

- **PROBLEMA:** Aumentan los datos a tratar → aumentan los recursos necesarios



- **SOLUCIÓN:** “Crecimiento Vertical” (Escalabilidad Vertical):



Ampliación recursos

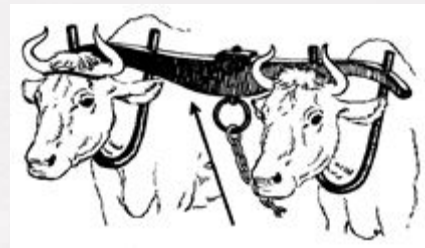
ó



Sustitución computador

DE LA CALCULADORA A LA COMPUTACIÓN DISTRIBUIDA

- El **Crecimiento Vertical** tiene limitaciones:
 - Computador no se puede ampliar más.
 - No exista computador tan grande.
- **SOLUCIÓN:** Grace Hopper y su parábola de los bueyes.



DE LA CALCULADORA A LA COMPUTACIÓN DISTRIBUIDA

- Aparece la “Computación Distribuida” y el “Crecimiento (escalado) Horizontal”

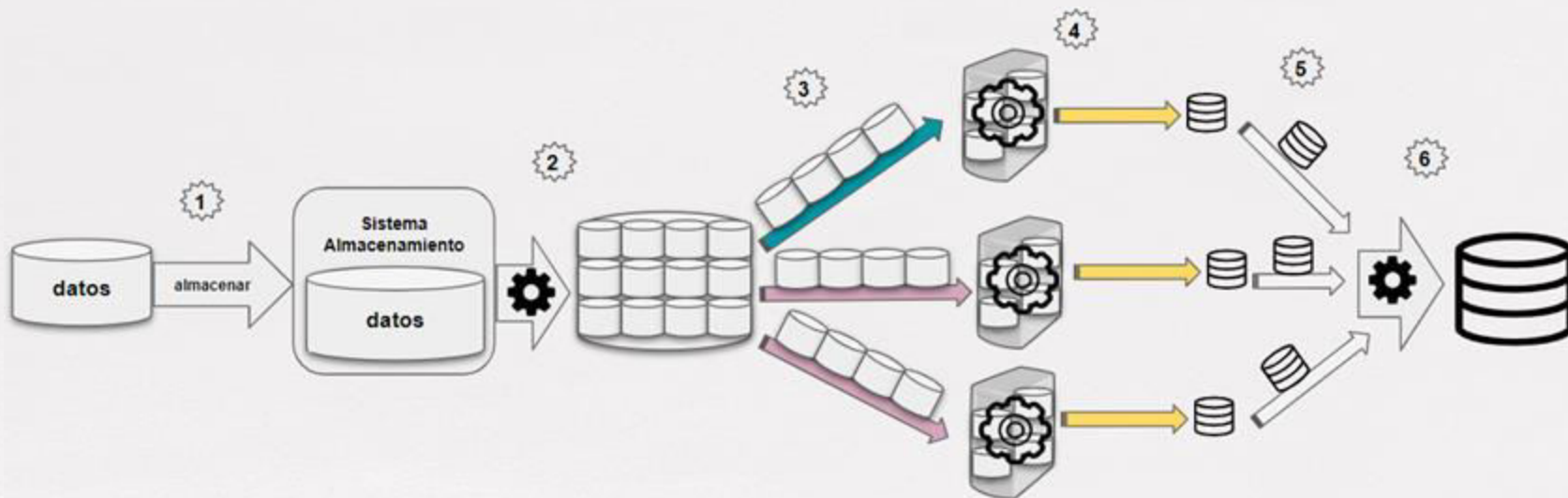


- Mayores necesidades → mayor número de computadores



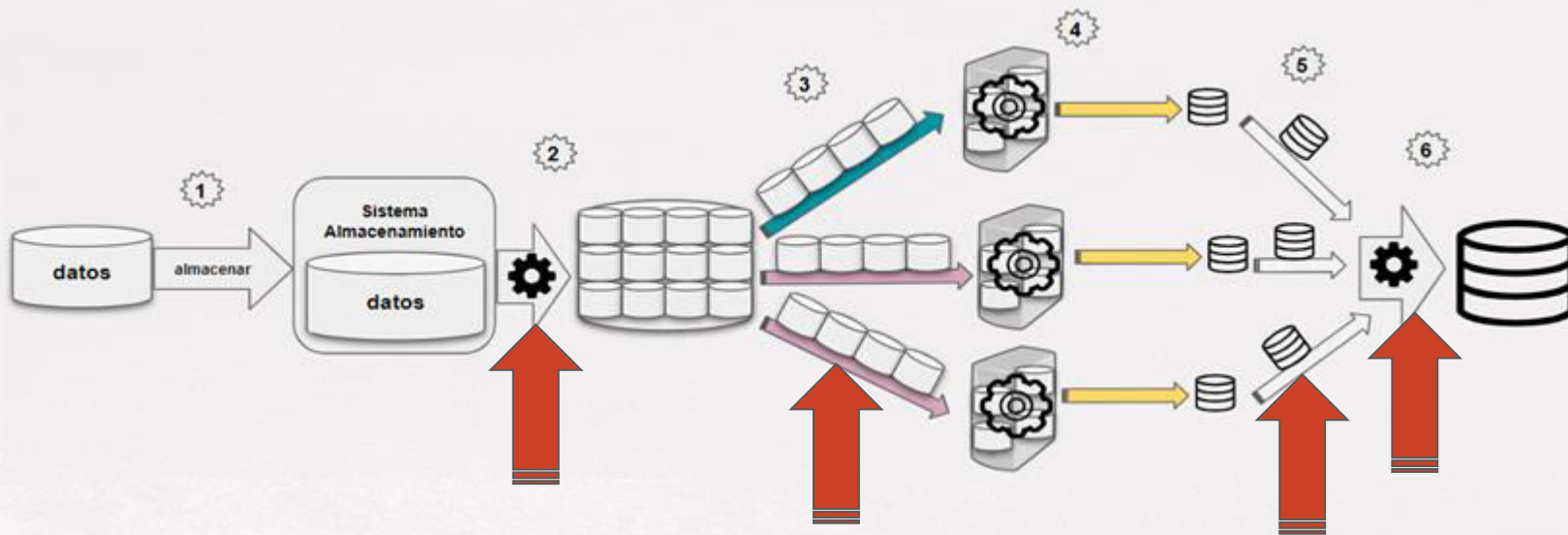
LIMITACIONES DE LA COMPUTACIÓN DISTRIBUIDA

- Tratamiento de Datos en “Computación Distribuida”:



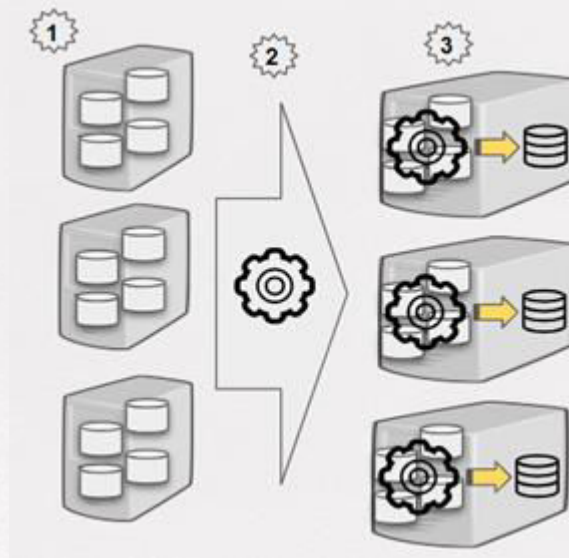
LIMITACIONES DE LA COMPUTACIÓN DISTRIBUIDA

- Crecimiento no puede ser infinito.
- Aumentan datos/computadores → Agrandan los **cuellos de botella**



LOCALIZACIÓN DEL DATO

- Big Data también es Computación Distribuida y “Almacenamiento Distribuido”



Big Data realiza el procesamiento allá donde está el dato → “Data

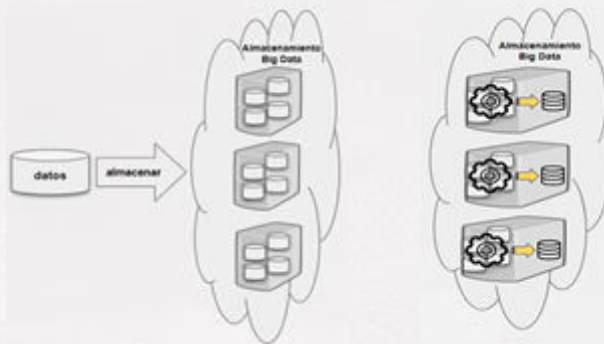
Locality”

LOCALIZACIÓN DEL DATO

- Al desaparecer fases, se eliminan los cuellos de botella



- En Big Data al almacenar se distribuyen los Datos



Data Locality

LOCALIZACIÓN DEL DATO

- Big Data aprovecha el “Crecimiento (escalado) Horizontal”



Mayor número
de computadoras



+ Almacenamiento
+ Procesamiento