



## Las cinco “V”s del **Big Data**:

- **Volumen:** Enorme cantidad de datos.
- **Velocidad:** Los datos se generan y se mueven a gran velocidad.
- **Variedad:** Los datos son de distintos tipos.
- **Veracidad:** El grado de confianza de los datos.
- **Valor:** La dificultad de explotación.

# Computación Distribuida

Velocidad en comunicaciones → Computación distribuida

Servidores y  
Supercomputadores



Sistemas Distribuidos



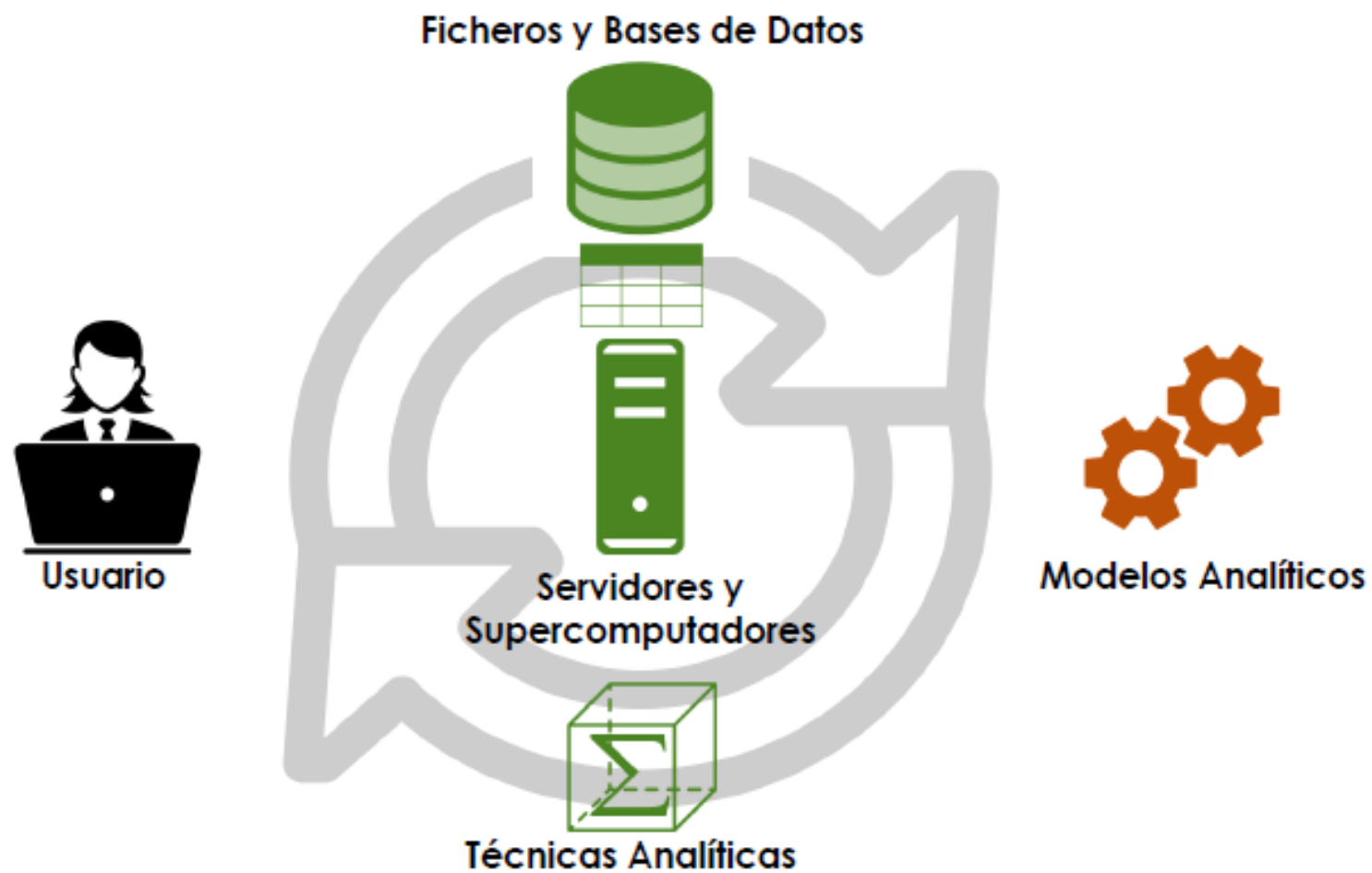
## Ventajas

- Alta disponibilidad
- Tolerancia a fallos
- Escalabilidad
- Bajo coste

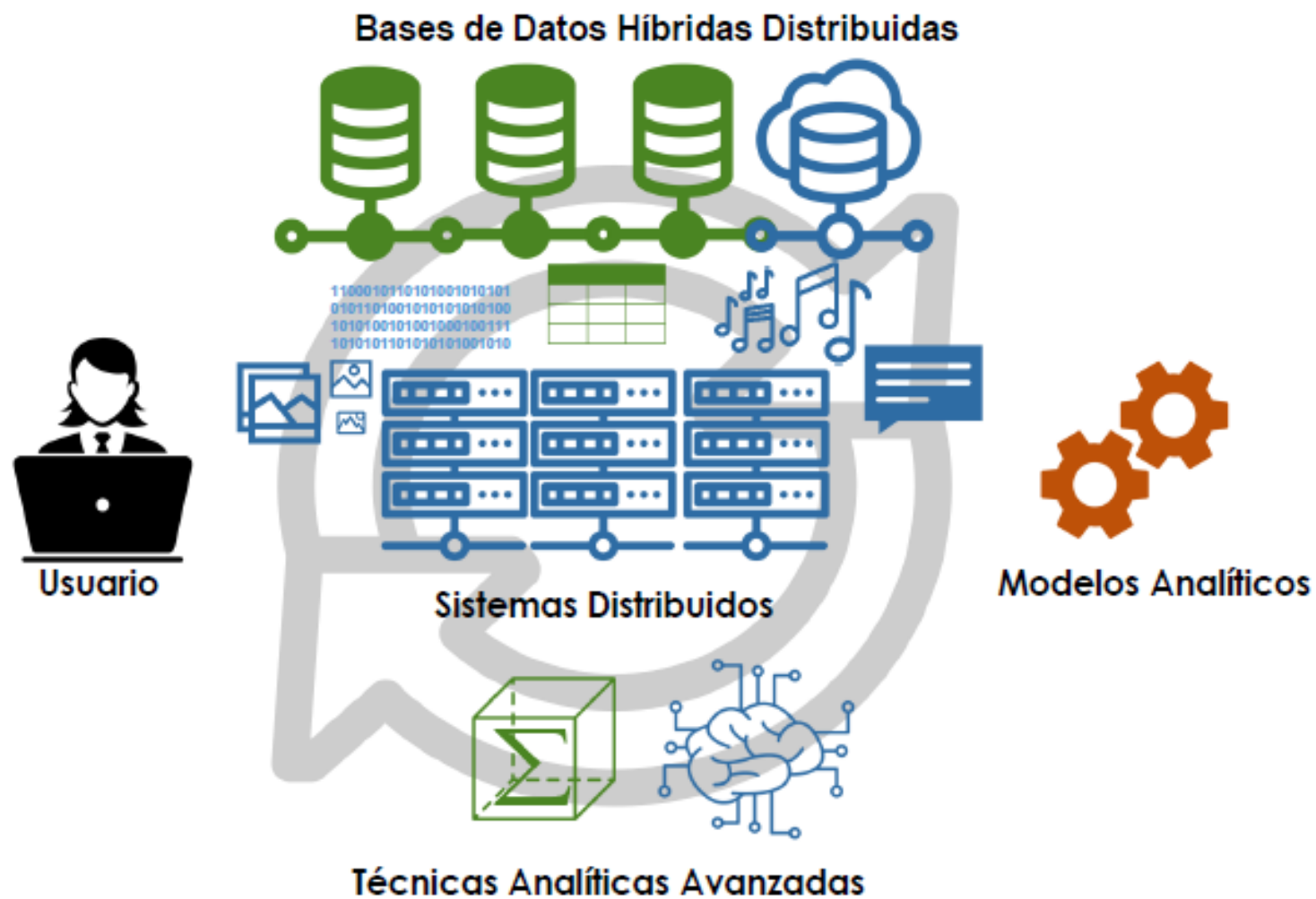
## Inconvenientes

- Seguridad: Más puntos de entrada
- Múltiples puntos de fallo
- Necesidad de software para computación distribuida

# Analítica Tradicional



# Analítica Avanzada



# Fuentes de Información: Identificación

Fuentes  
Internas



Redes  
Sociales

Fuentes  
Externas




Open  
Data

¿Dónde se encuentra la información?

En 2004 **Google** publicó un artículo describiendo la problemática para indexar la web y propuso una solución.

Esta solución se basaba en un sistema de ficheros (GFS) y en un modelo de programación denominado MapReduce

Doug Cutting vio el potencial de la solución y en 2005 **YAHOO!** lanzó una implementación basada en esta propuesta a la que llamó  **hadoop**.

Hadoop es desarrollado y mantenido por The Apache software Foundation.





