



**UAB**

Universitat Autònoma  
de Barcelona

Tomás Margalef

# Apache Flume

# Apache Flume



Fuentes de Datos



Adquisición de Datos  
*Data Ingestion*

Transporte de Datos  
*Data Collector*

MapReduce



YARN Yet Another Resource Negotiator



HDFS Hadoop Distributed File System

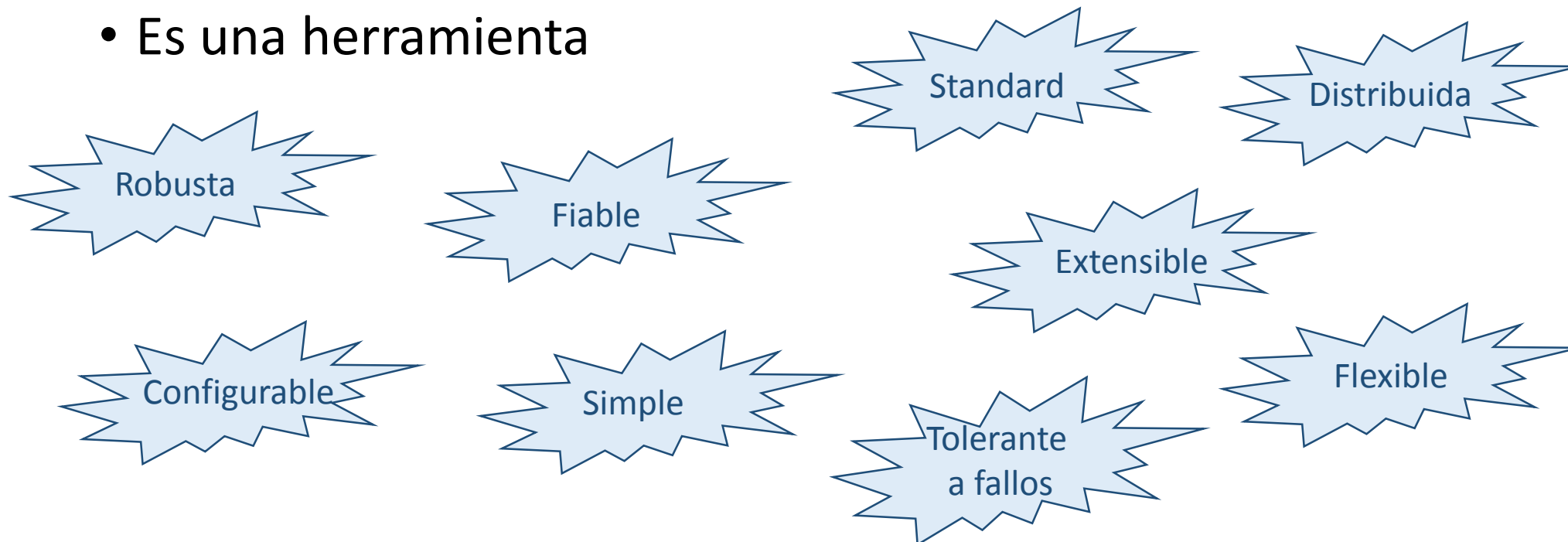


Hardware Básico

# Apache Flume



- Es una herramienta



para la adquisición de datos (*Data ingestion*) provenientes de diversos generadores de datos (webservers) en Hadoop.



# Apache Flume



Generadores  
de datos:  
Log/Eventos



Datos



Datos

Mediador:  
Flujo estable

Almacenamiento  
centralizado



# Apache Flume



Generadores  
de datos:  
Log/Eventos



Agente 1

Agente 2

Agente 3

Data collector



Almacenamiento  
centralizado





# Apache Flume



Generadores  
de datos:  
Log/Eventos



Agente 1

Agente 2

Agente 3

Header

Byte Payload

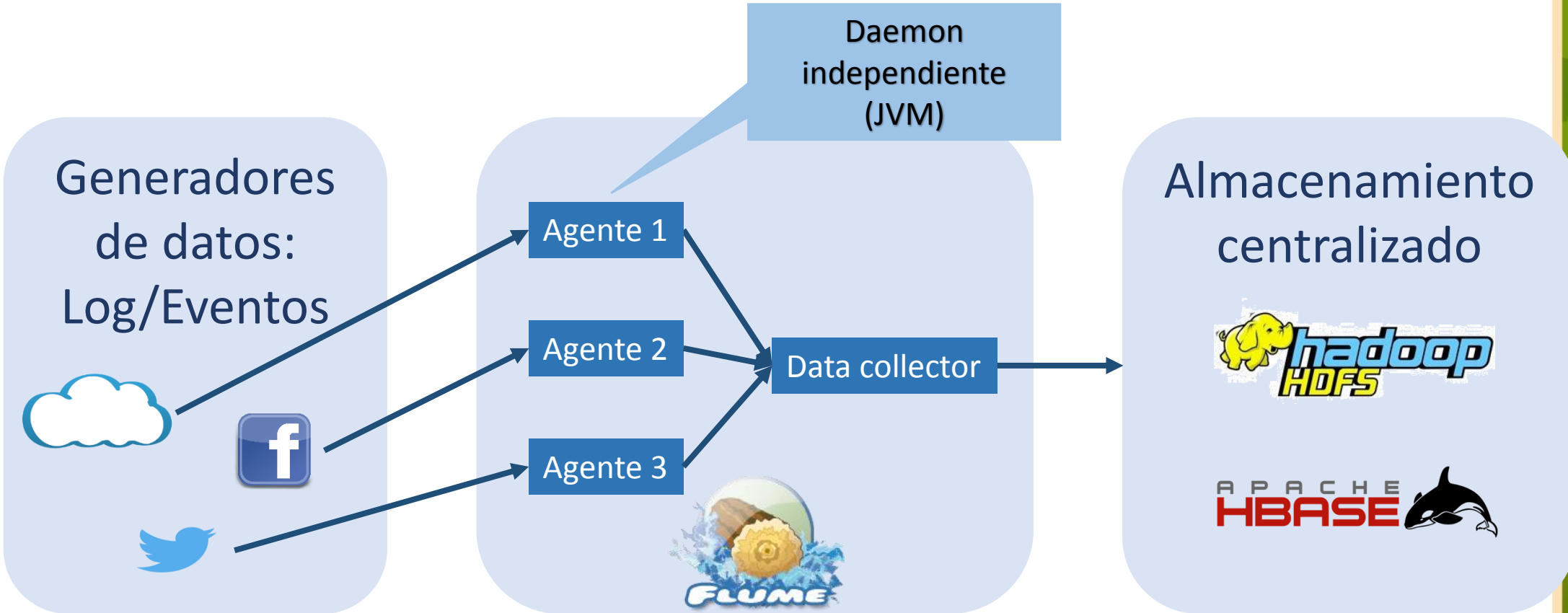
Data collector



Almacenamiento  
centralizado



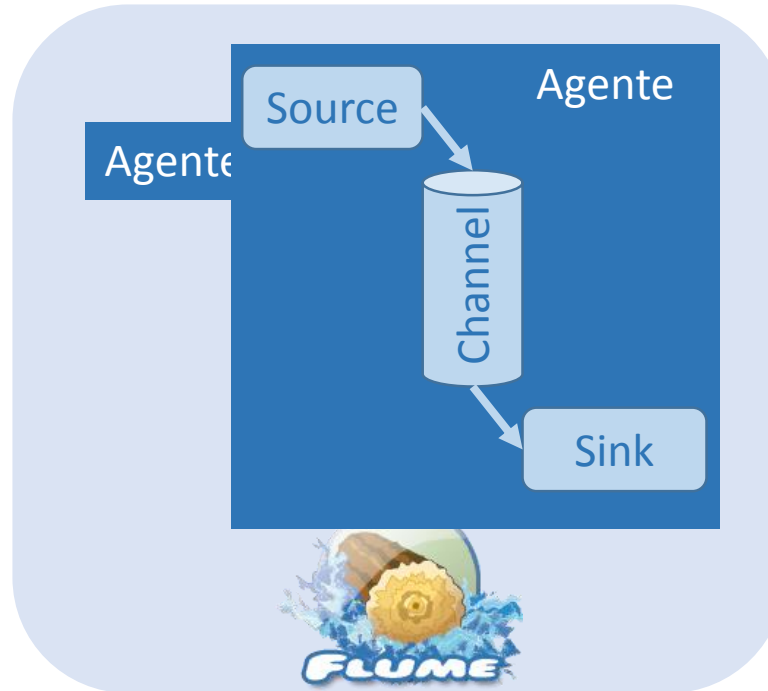
# Apache Flume



# Apache Flume



Generadores  
de datos:  
Log/Eventos



Almacenamiento  
centralizado

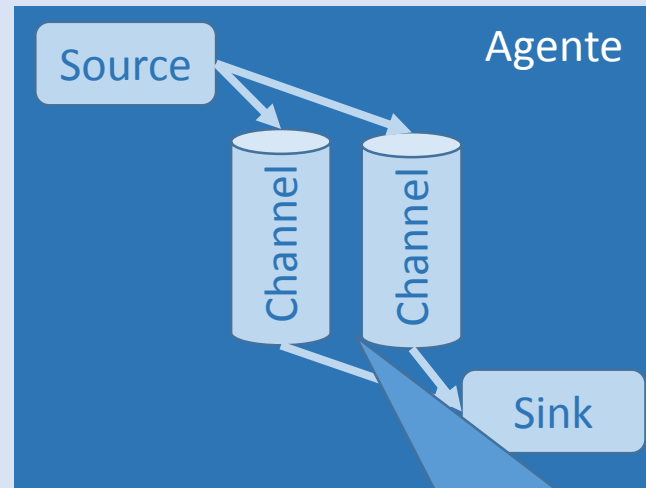




# Apache Flume



Generadores  
de datos:  
Log/Eventos



Fan-Out Flow:

- Replicating
- Multiplexing

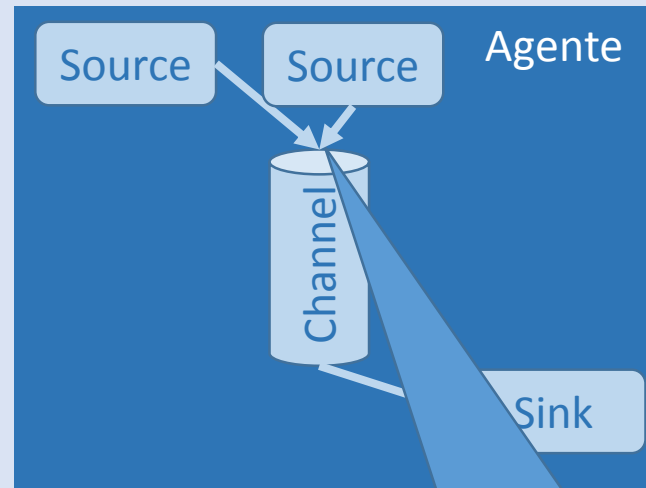
Almacenamiento  
centralizado



# Apache Flume



Generadores  
de datos:  
Log/Eventos



Fan-In Flow

Almacenamiento  
centralizado



A close-up of a laptop screen displaying the word "MOOC" in a bold, sans-serif font. The screen is framed by a thick black border. The background of the image is a solid light green color.

**MOOC**

# Apache Flume

A close-up of a person's hands typing on a laptop keyboard. The laptop screen is visible in the background, displaying the UAB logo and the text "MOOC Escola de Postgrau". The image has a green tint.

**UAB**  
Universitat Autònoma de Barcelona

**MOOC**  
Escola de  
Postgrau

**UAB**

Universitat Autònoma  
de Barcelona