



Curso de:

Hadoop

Benjamín Casazza

Data to BSN



- Recolección

- Análisis

- Custodia-trazabilidad

- "Convertir los datos en negocio"

Científico y arquitecto de datos

Apasionado por la tecnología

Data to BSN

Profesor de asignatura



@benjamin.casazza



uracilo

¿Qué aprenderás?

- ¿Qué es Hadoop?

- Diferenciar Hadoop** con otras herramientas de big data

- Identificar la **caja de herramientas** de Hadoop



Requisitos



Big data y ciencia de datos



Bases de datos SQL y No SQL



Docker

1 Introducción a Hadoop

1.1 Fundamentos

- ¿Por qué Hadoop?

- ¿Cómo surge Hadoop?

- ¿Qué es Hadoop?

¿Por qué Hadoop?



Big Data



**Análisis y
almacenamiento**



**Mercado
de los datos**



**Análisis y
procesamiento**

¿Cómo surge Hadoop?



— Doug Cutting

— Motor de búsqueda
Nutch 2002

¿Qué elementos tiene Hadoop?



YARN



MapReduce



HDFS

¿Qué es Hadoop?



└ Software libre
└ Escalable

¿Qué es Hadoop?



└─ Distribuido
└─ Confiable

| 1.2 Usar Hadoop en lo profesional

Introducción a Hadoop

SETI@home





Puestos actuales



- ✓ Ingeniero de datos
- ✓ Científico de datos
- ✓ Analista de datos
- ✓ Arquitectos
- ✓ Etc.

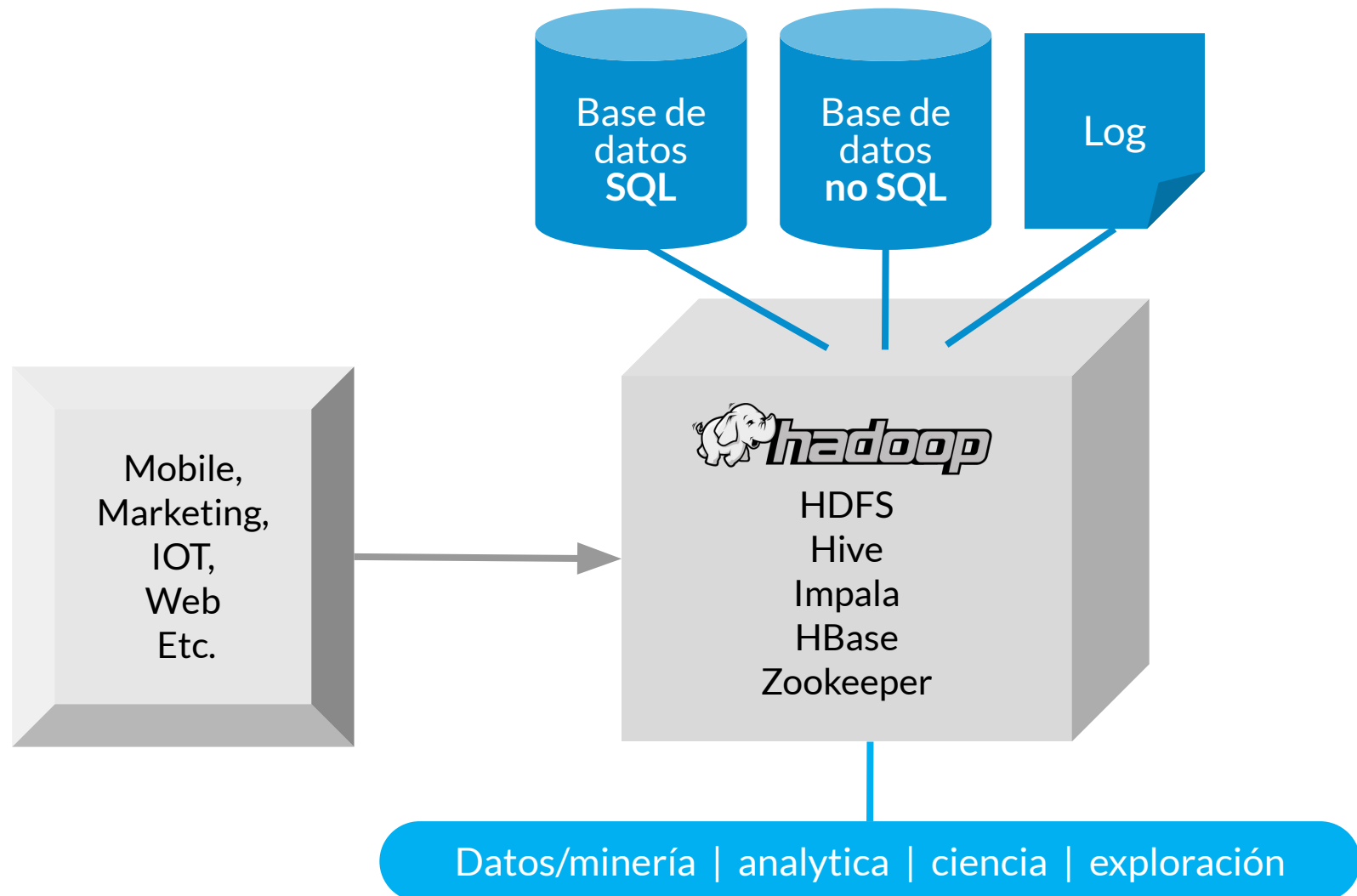
Configurando nuestro entorno con Hadoop

Crear tu propio dockerfile para tu entorno de pruebas con Hadoop

Comparar otras herramientas frente a Hadoop

Reconocer y diseñar flujo de datos

Arquitectura



Aprendiendo sobre los datos en Hadoop

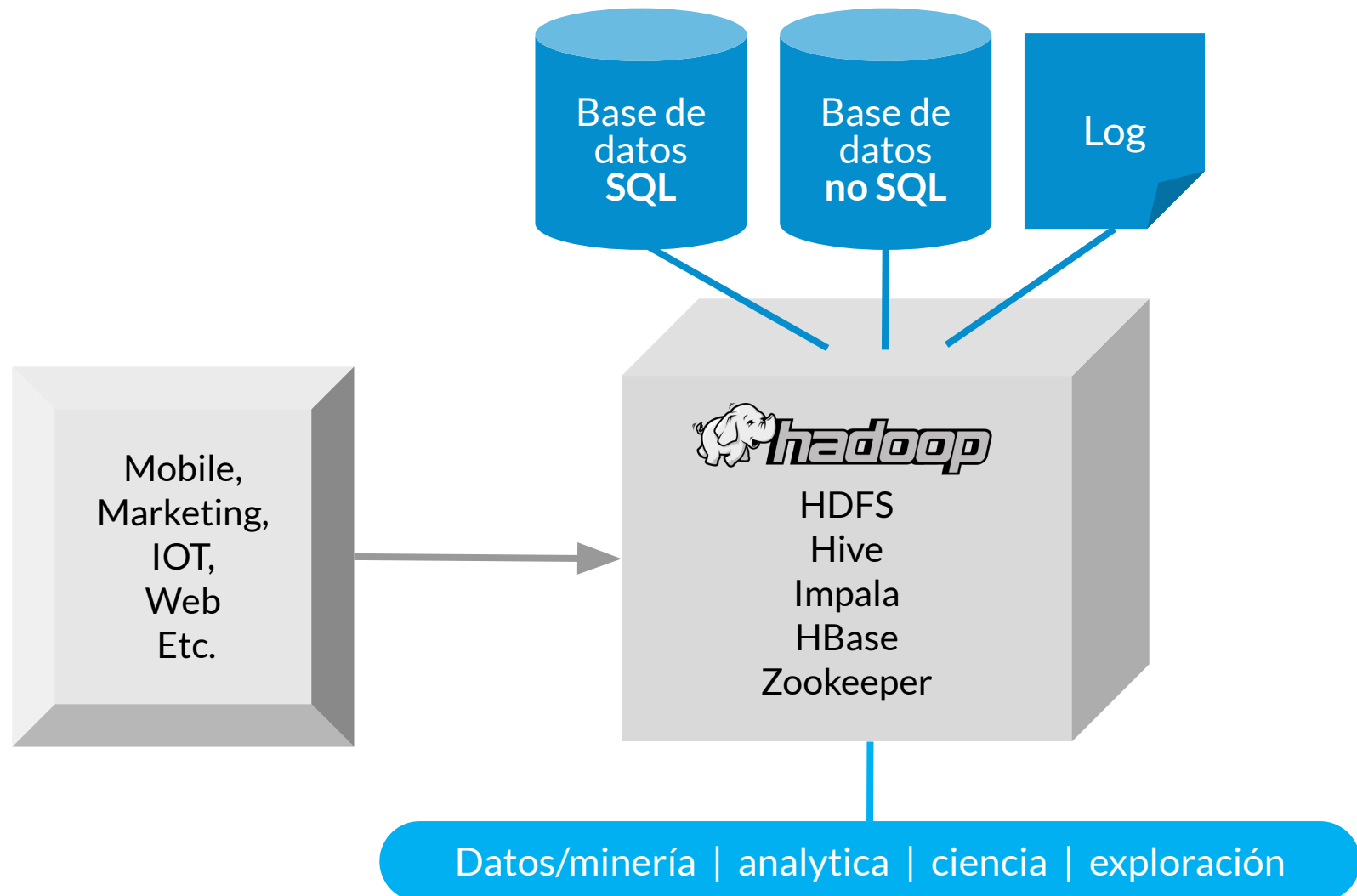
Comprender el funcionamiento de MapReduce

2 Elementos de Hadoop

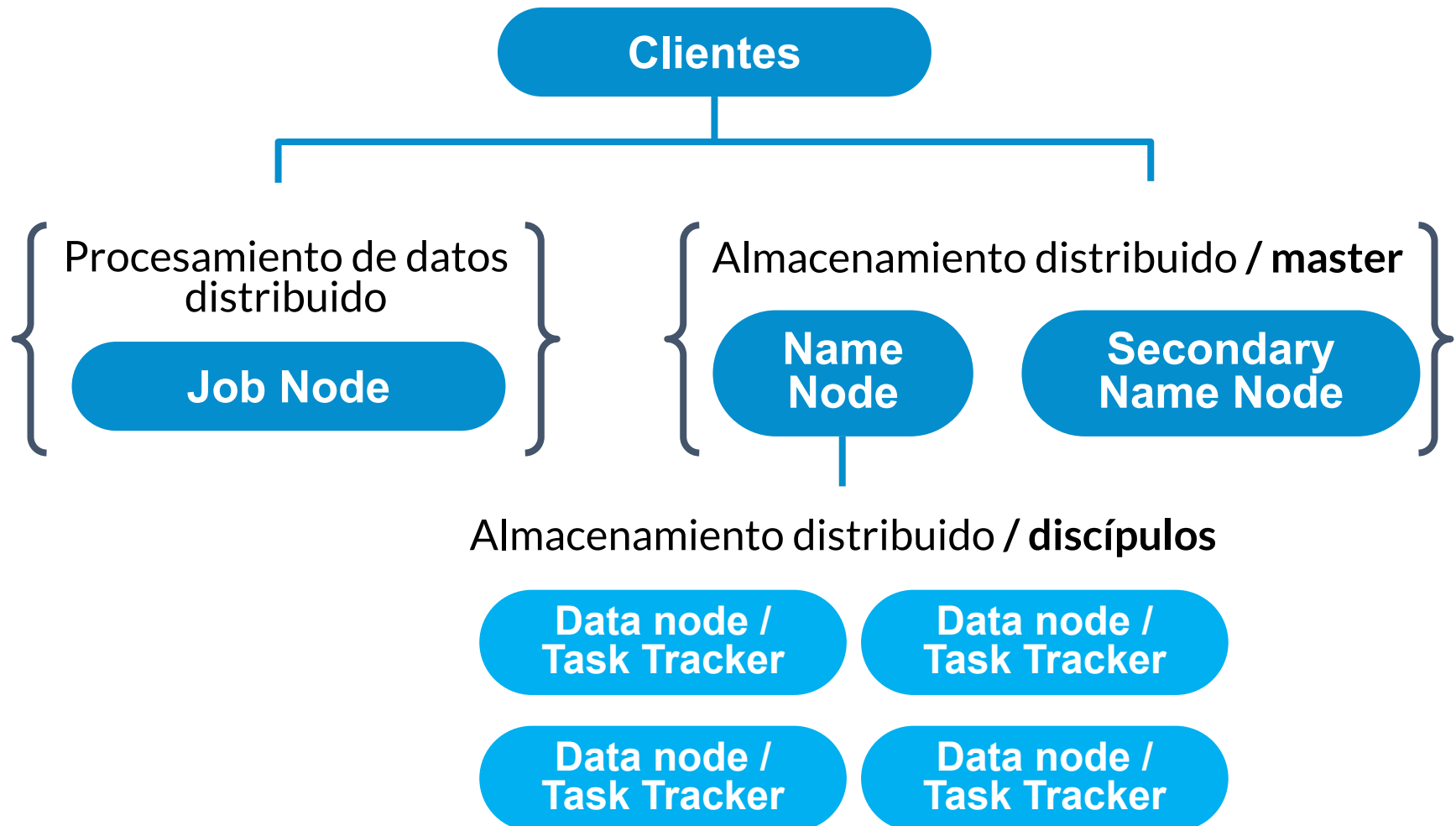
| 2.1 HDFS

Elementos de Hadoop

Arquitectura



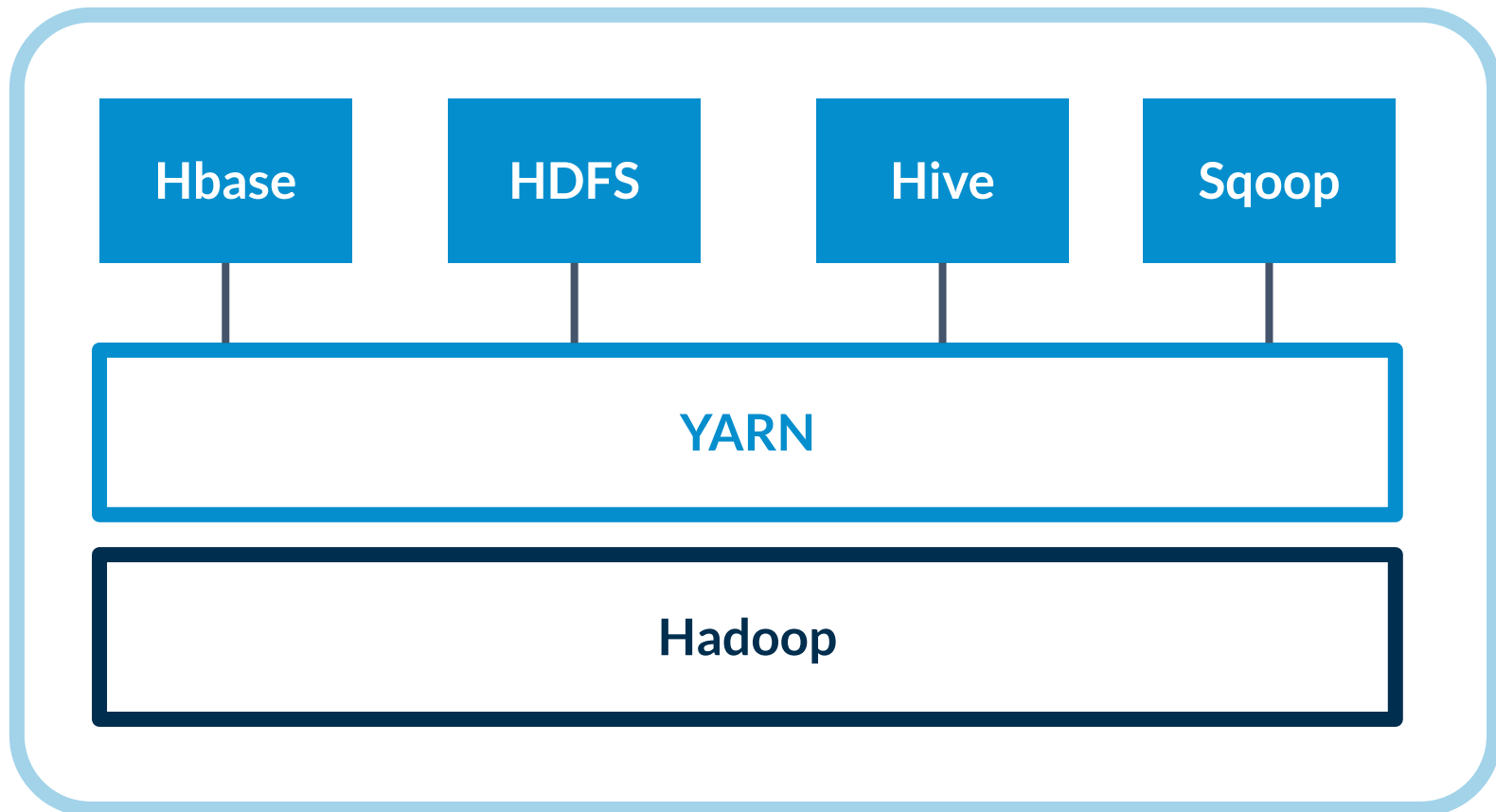
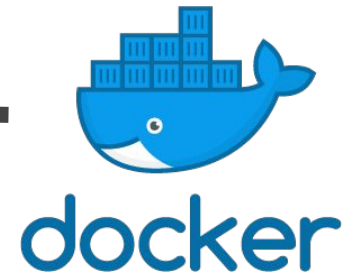
Hadoop distributed file system



2.2 YARN

Elementos de Hadoop

Arquitectura YARN



3 MapReduce

| 3.1 Framework

MapReduce



MapReduce

Input

Big data
usa Hadoop
Big data
es una
tecnología
poderosa

Splitting

Mapping

Shuffling Sort

Reducing

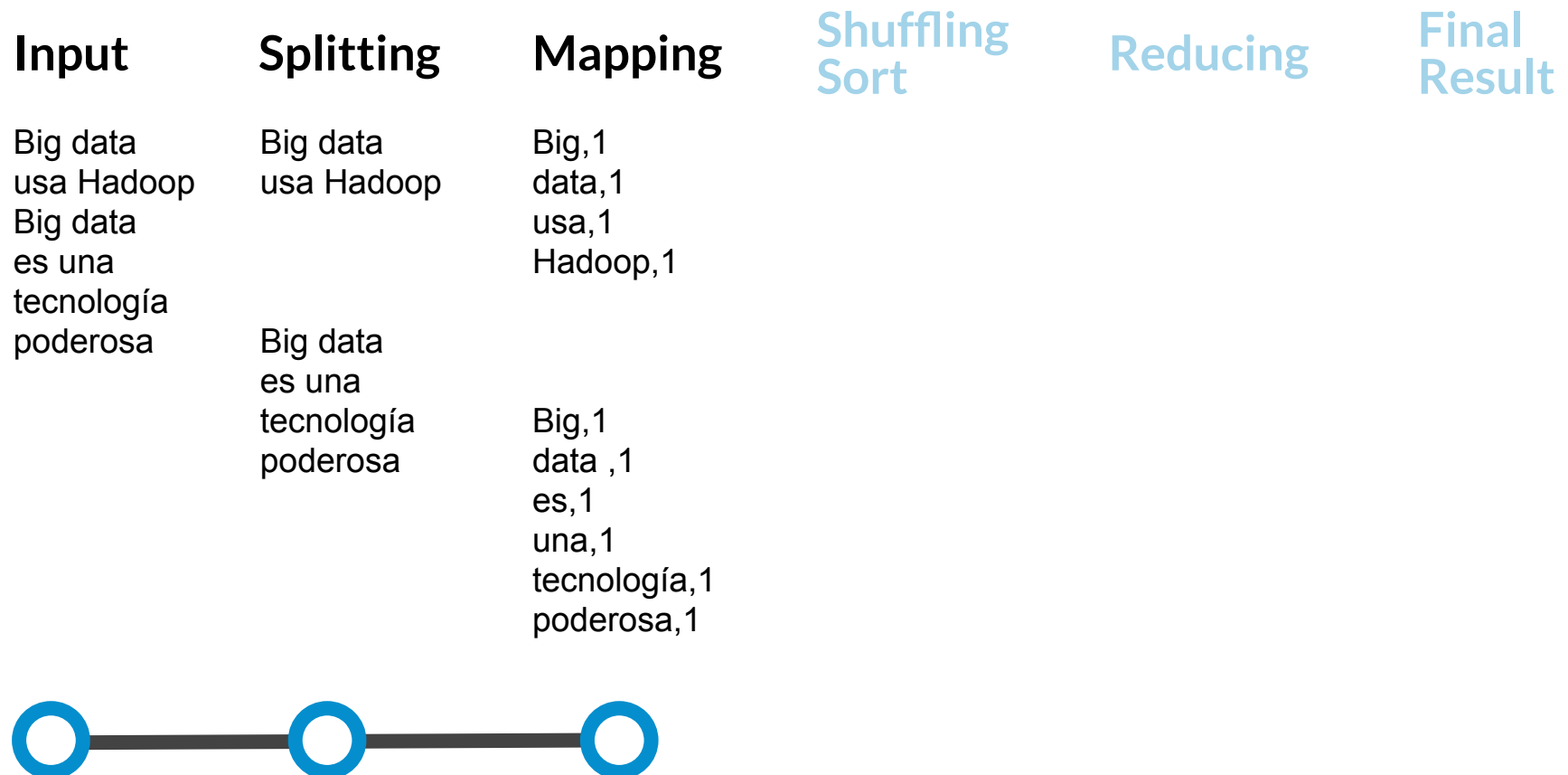
Final Result



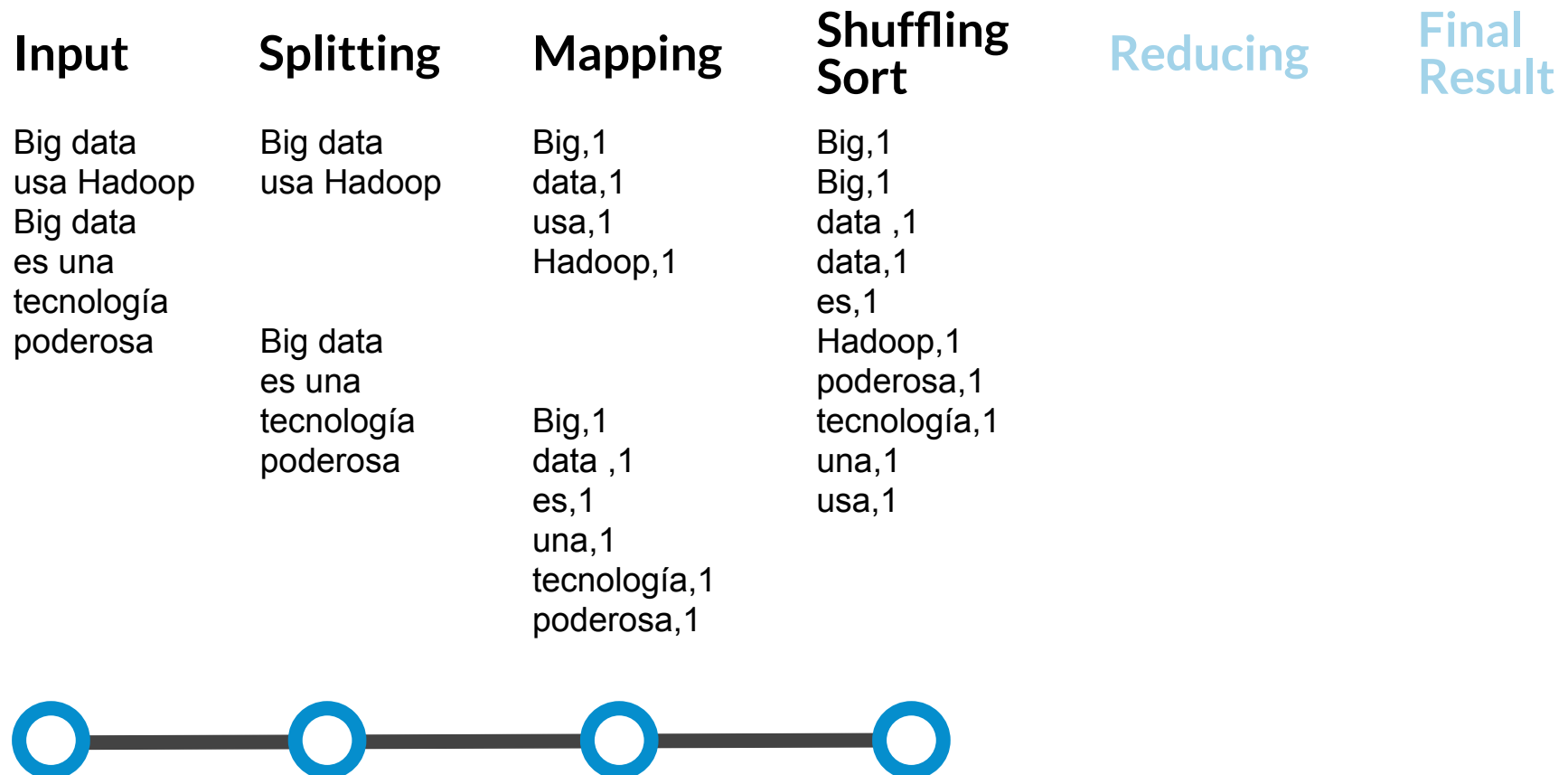
MapReduce



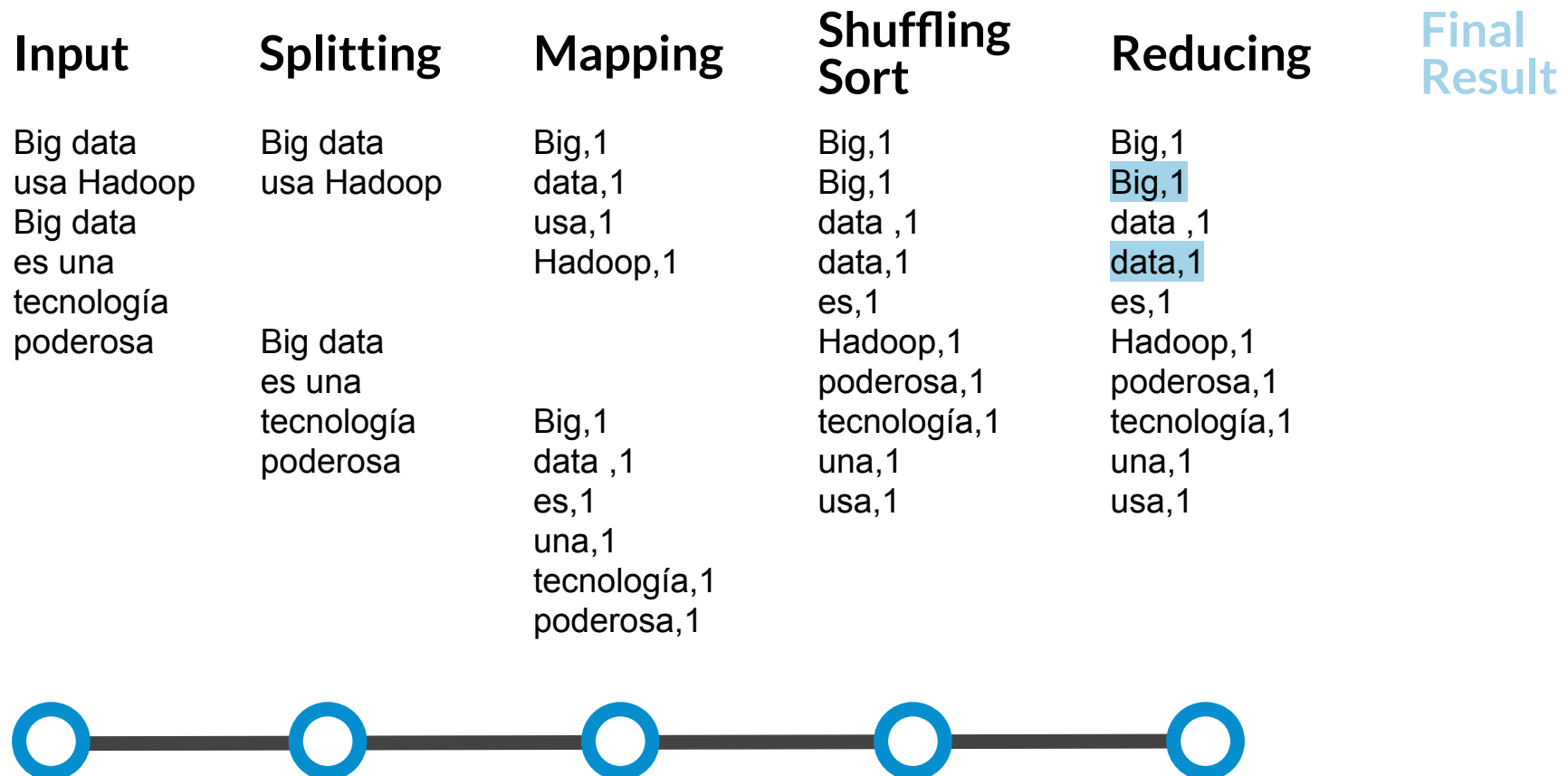
MapReduce



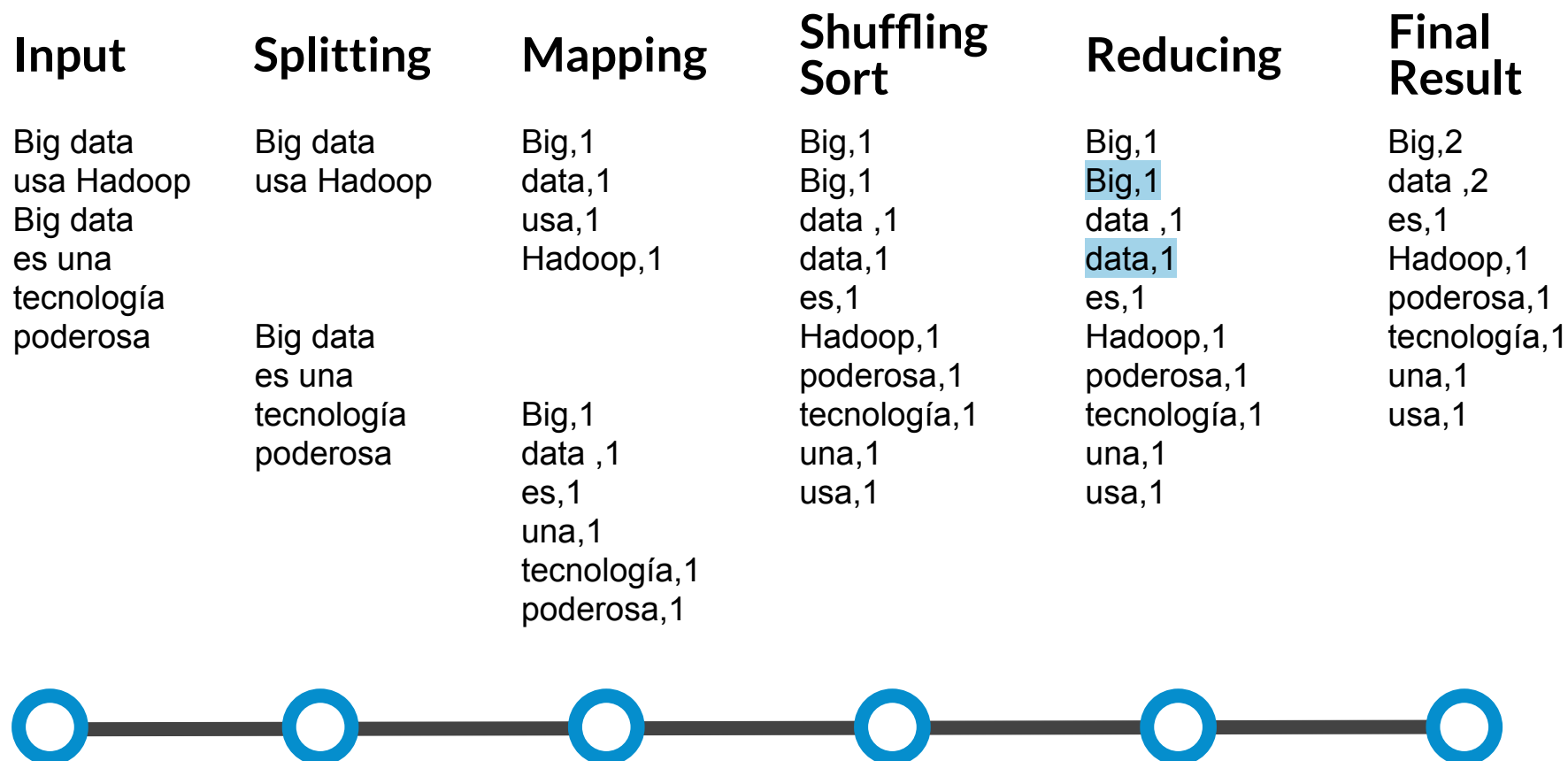
MapReduce



MapReduce



MapReduce



3.2 Identificar tipos de datos en herramientas de Big Data

MapReduce

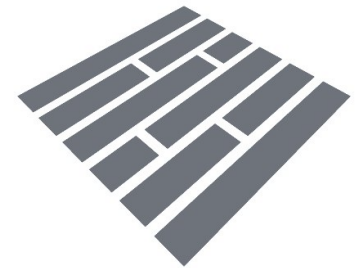
Datos en Hadoop



JSON



Avro



Parquet

Parquet

Desarrollar un clúster de Hadoop



Conocer Flume

Conocer Sqoop e integrar Hbase

Conocer Parquet y Avro

Generando
archivos tipo
Parquet



Conocer Hive

Administrar un clúster de Hadoop



Zookeeper



¡Felicidades!
Haz completado el curso de

Hadoop

Benjamín Casazza

¿Qué pasos siguen?

