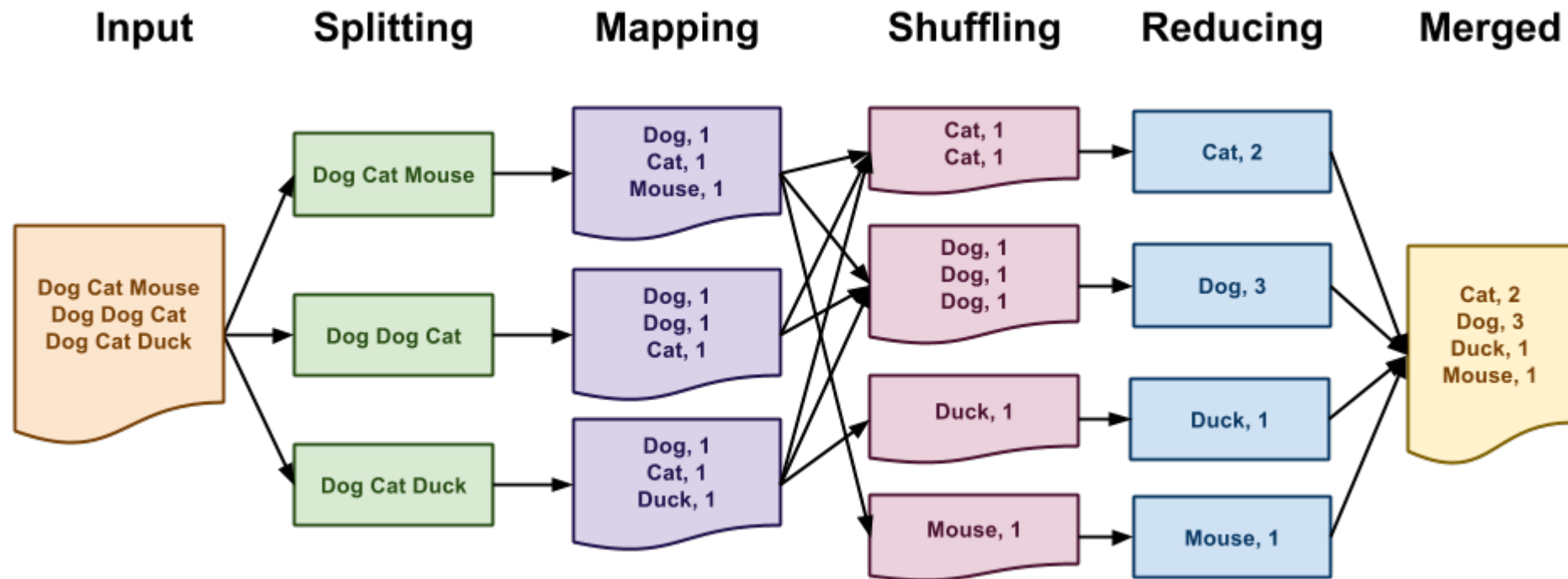


MapReduce, Hive, and Sqoop

MapReduce

- The very first big data technique
- Map – Turn only required data into key-value pairs
- Reduce – Aggregate the keys

An Extended MapReduce Example



Modern MapReduce

- Programs integrated the concepts
 - Hive
 - Spark
 - Pig

Hive

- SQL queries on big data

Comparing Versions

MySQL

- CRUD
- View is a temporary table accessed later
- Schema on write: structured data

HiveQL

- No CRUD
- View is a fleeting construct
- Schema on read: unstructured data

Sqoop

- SQL + Hadoop
- Run queries in MySQL on Hadoop

Questions?