



Big Data

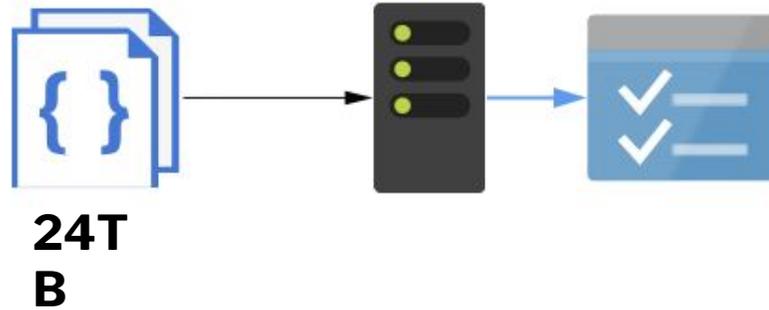
Эксплуатация больших данных

Руслан Габбасов

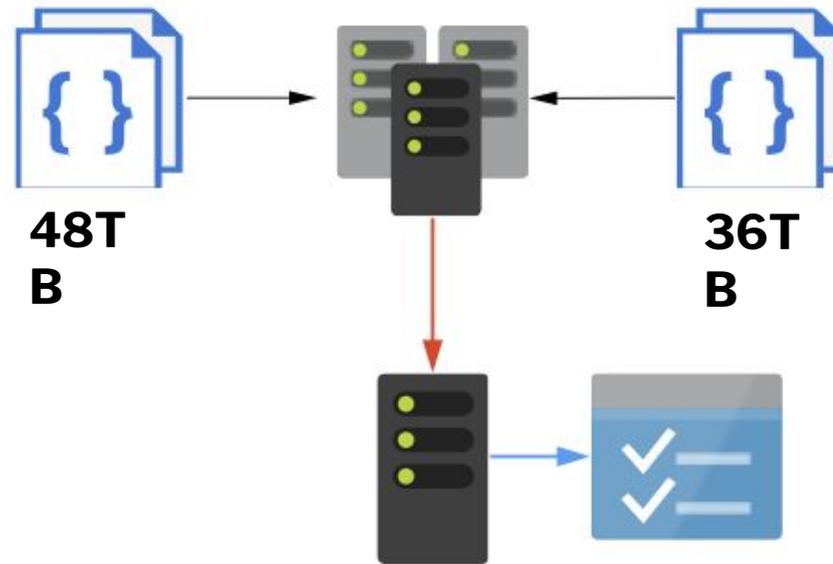


DIRECTUM

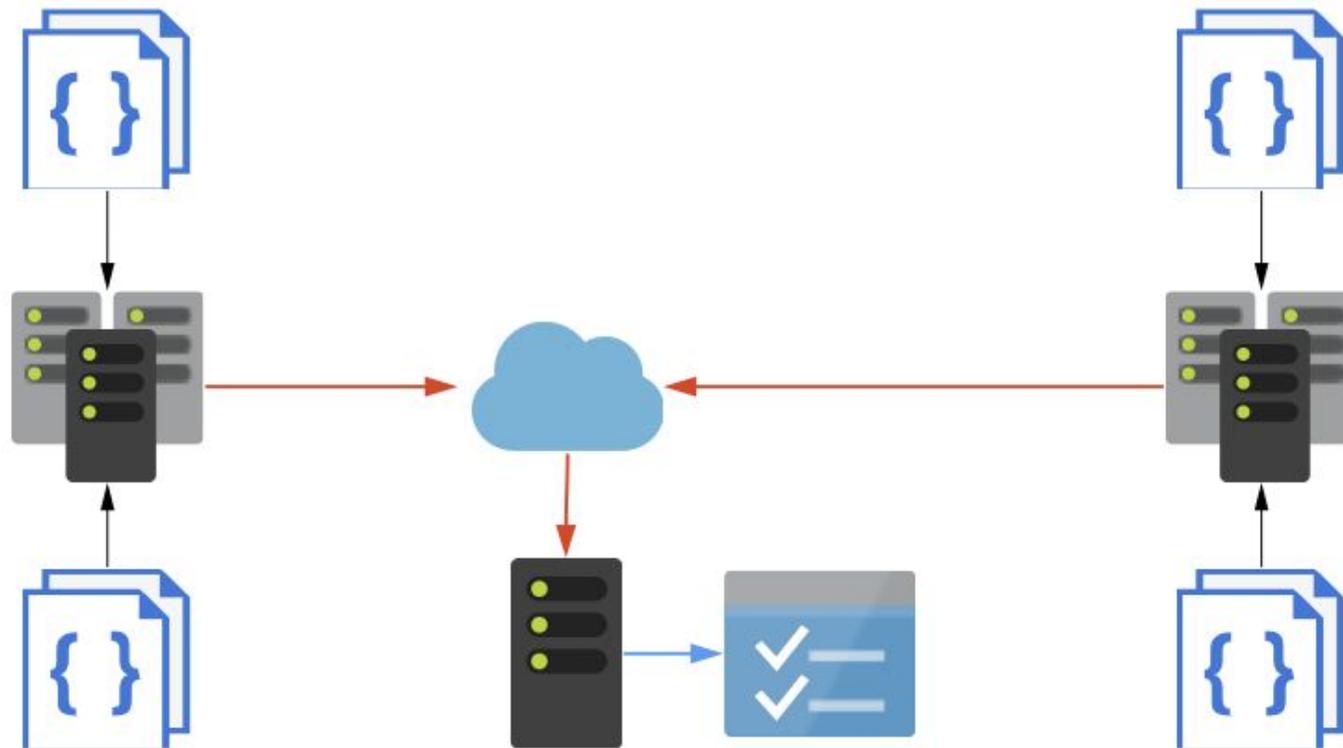
Что такое большие данные?



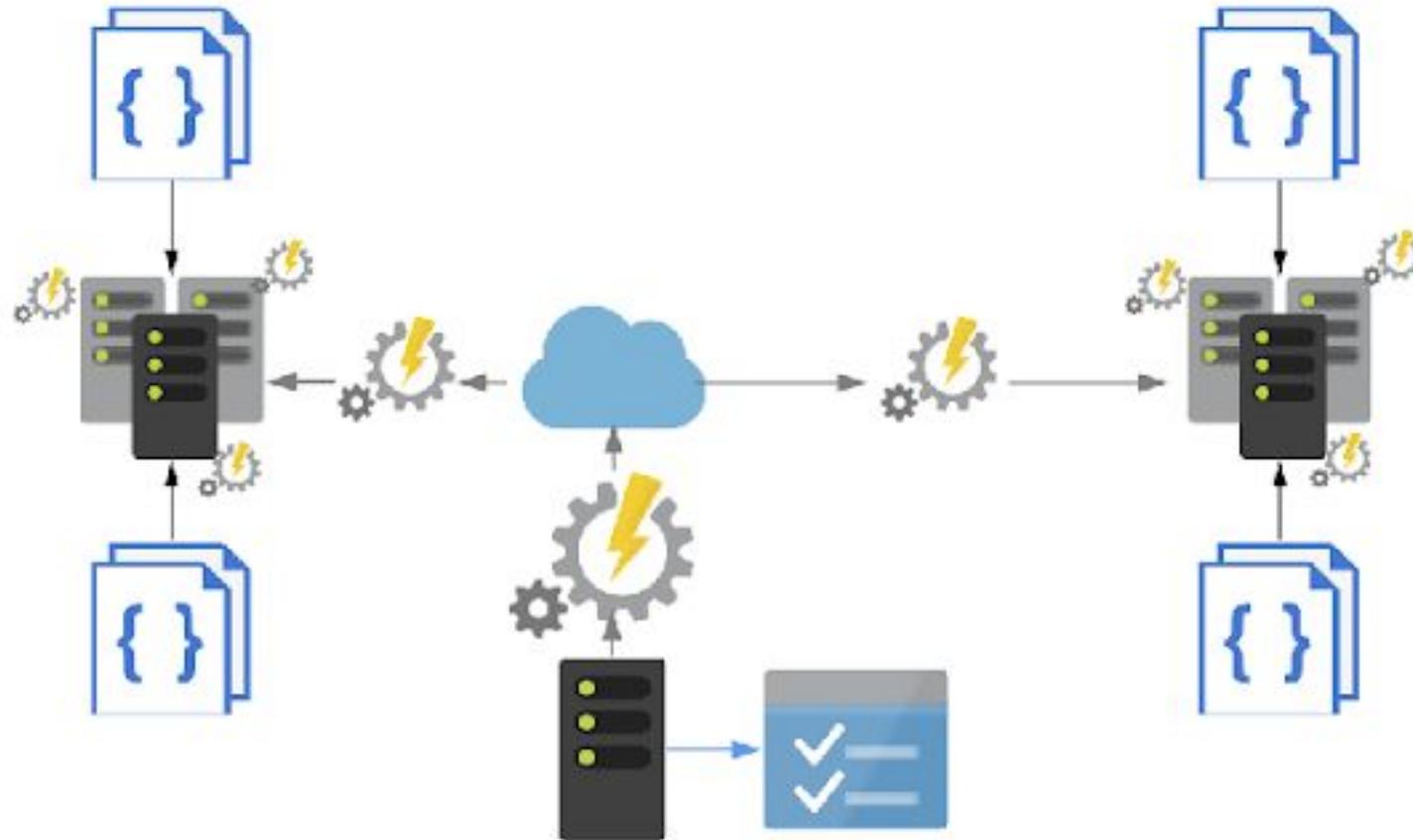
Что такое большие данные?



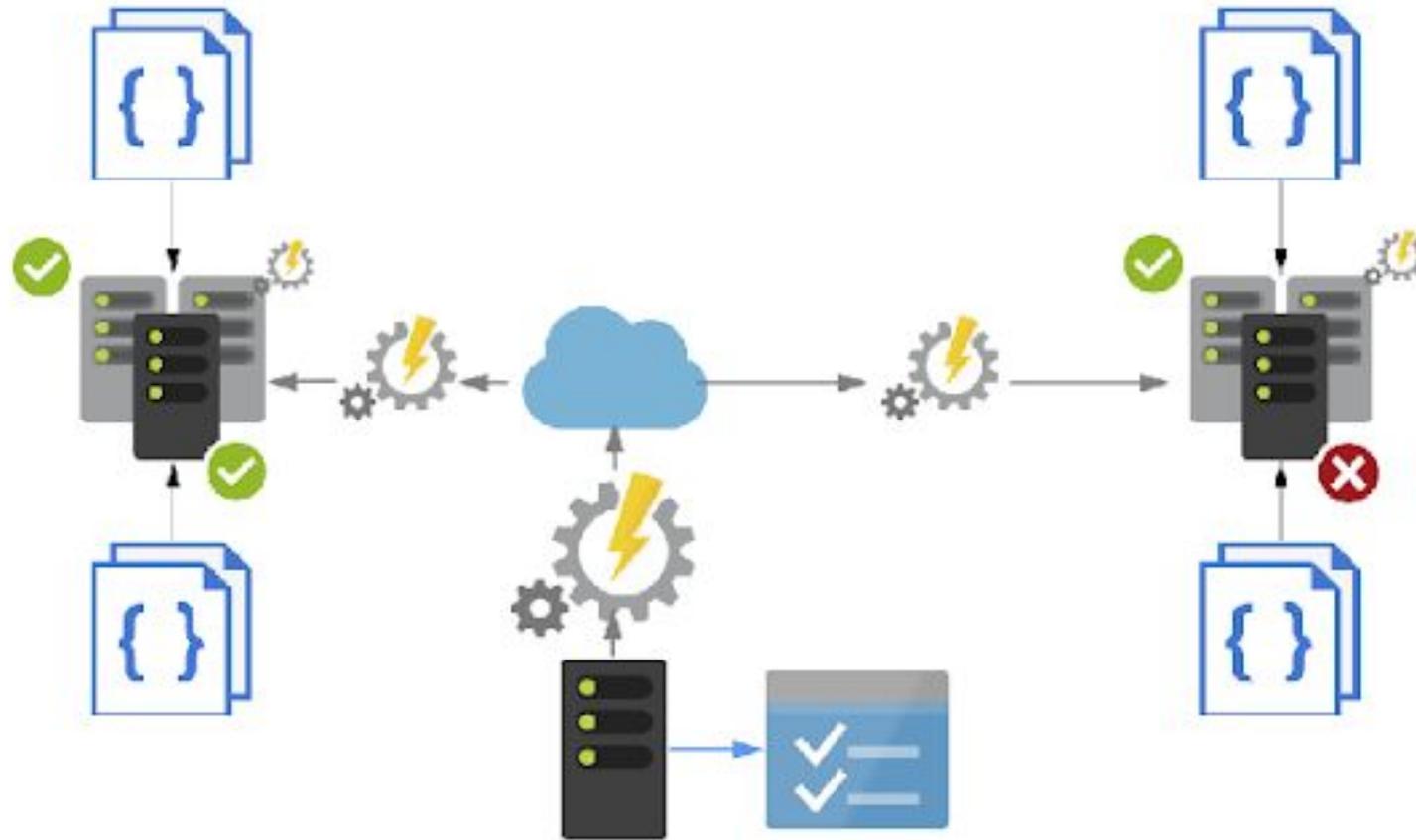
Что такое большие данные?



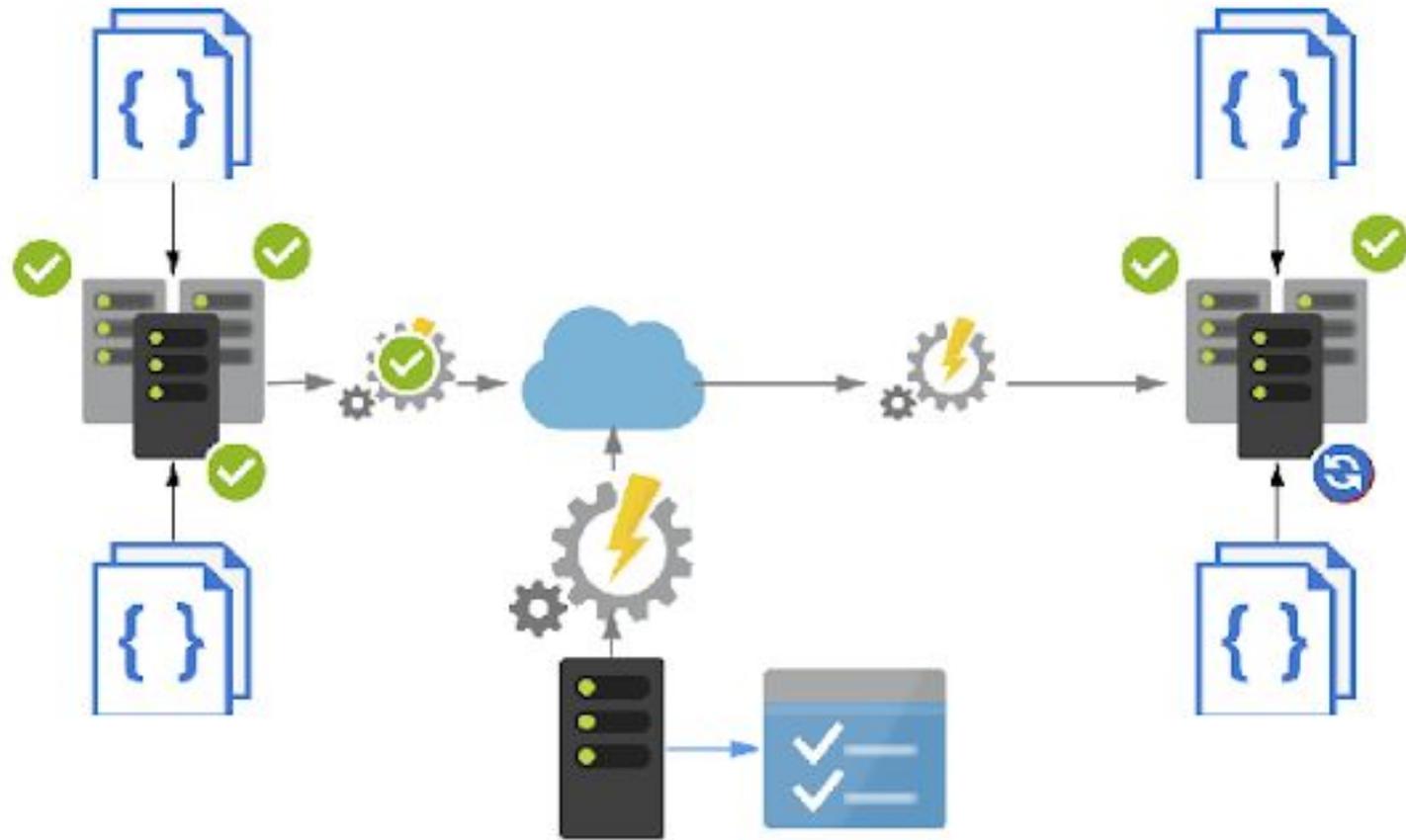
Это когда проще отправить код к данным



Это когда проще отправить код к данным



Это когда проще отправить код к данным



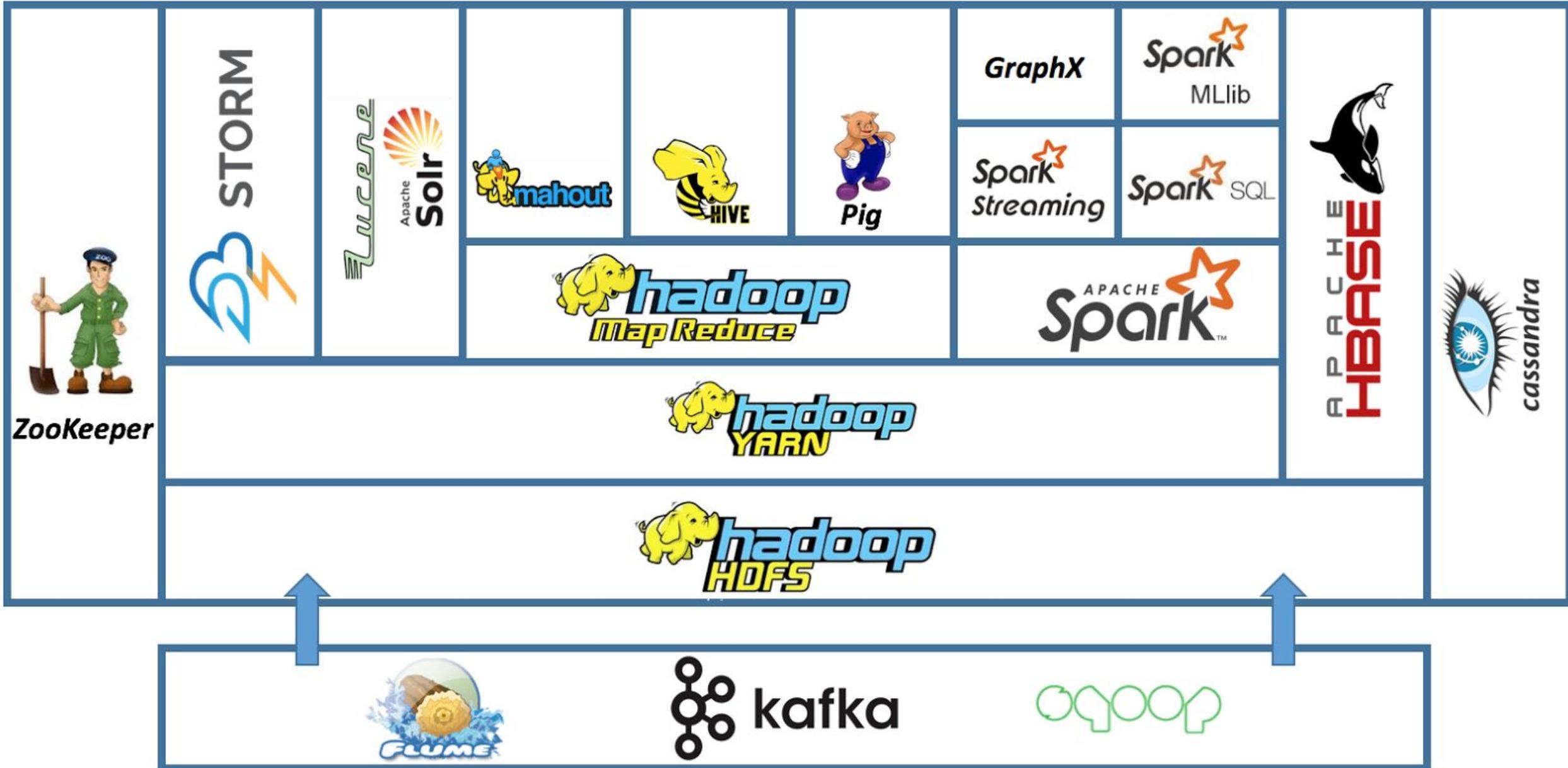
Это когда проще отправить код к данным





Apache Zookeeper

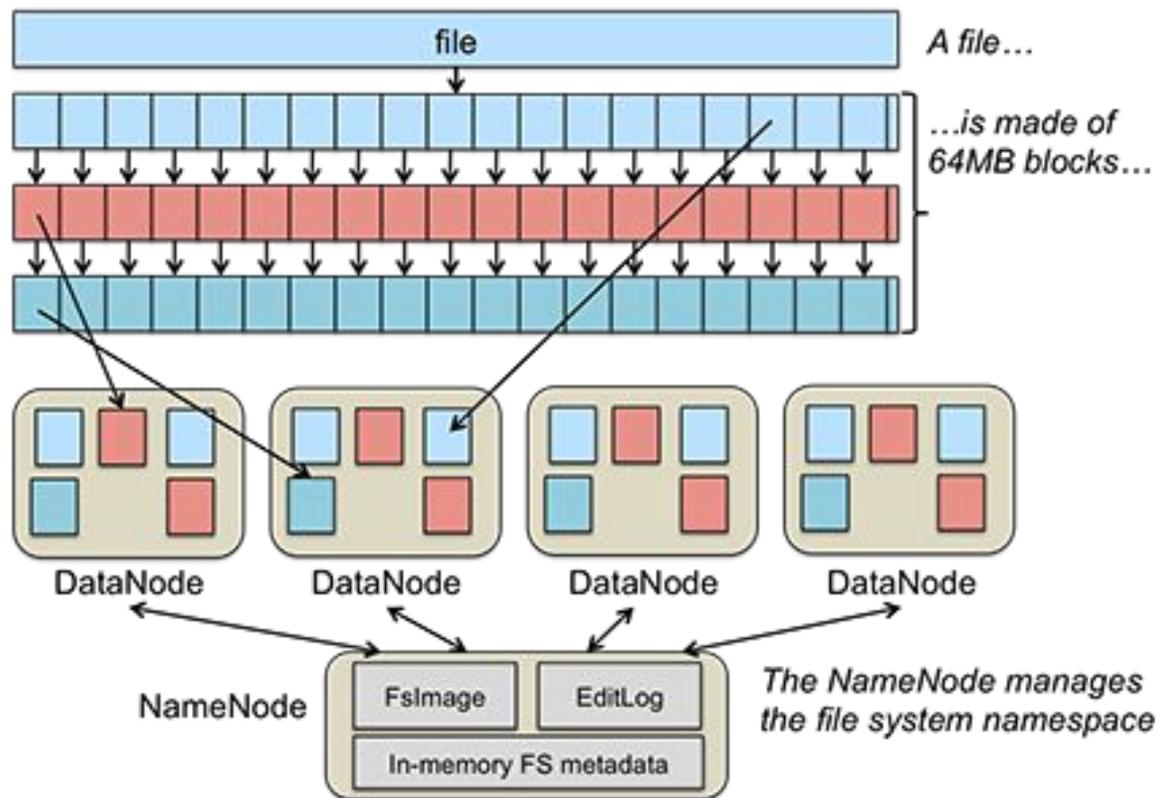
Экосистема Hadoop



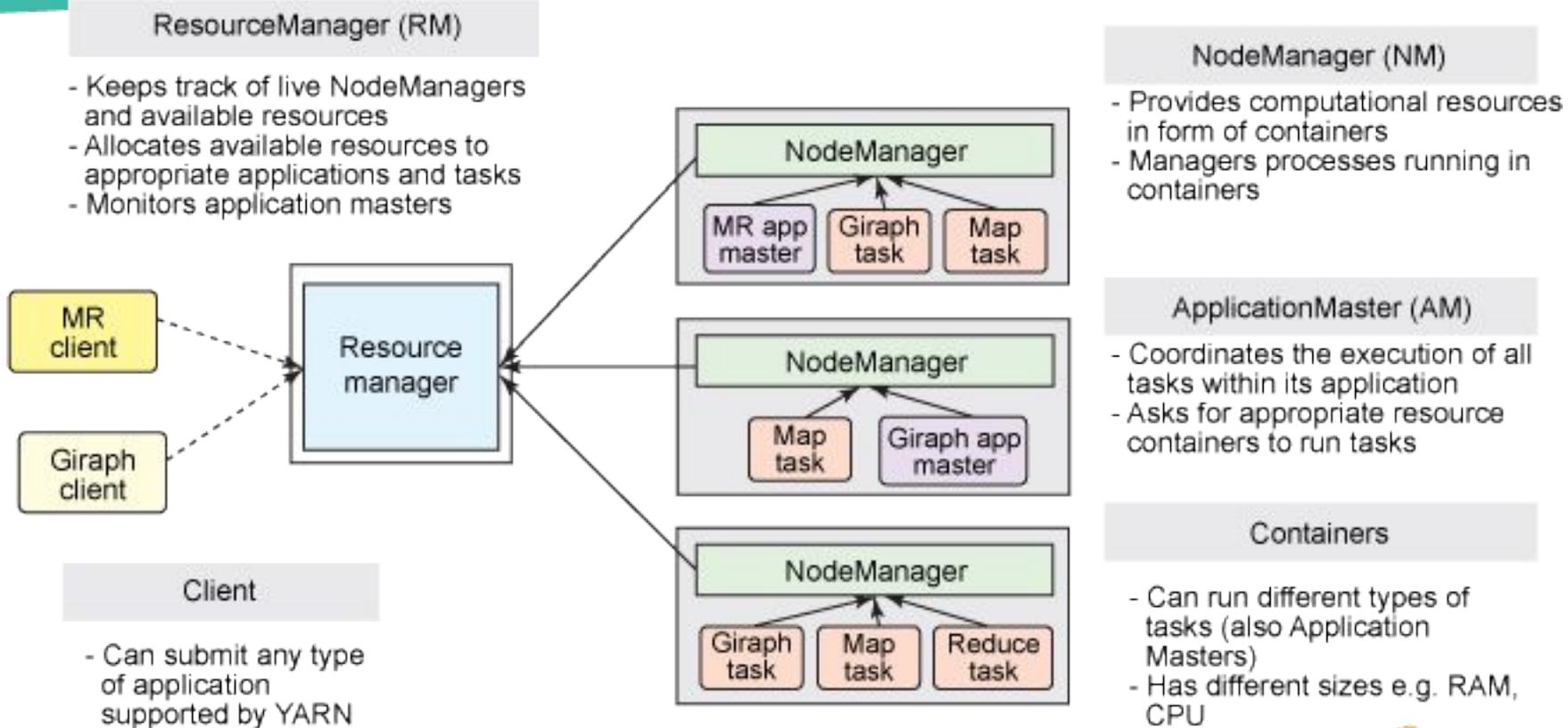
Big Data — зоопарк: слоны, свиньи и т.д.

- Hadoop — экосистема, набор библиотек и утилит для работы с большими распределенными данными
- HDFS — файловая система для блочного хранения больших файлов с поддержкой репликации
- YARN — утилита для планирования, координации и управления вычислительными ресурсами кластера
- Spark — фреймворк для распределенной обработки неструктурированных и слабоструктурированных данных

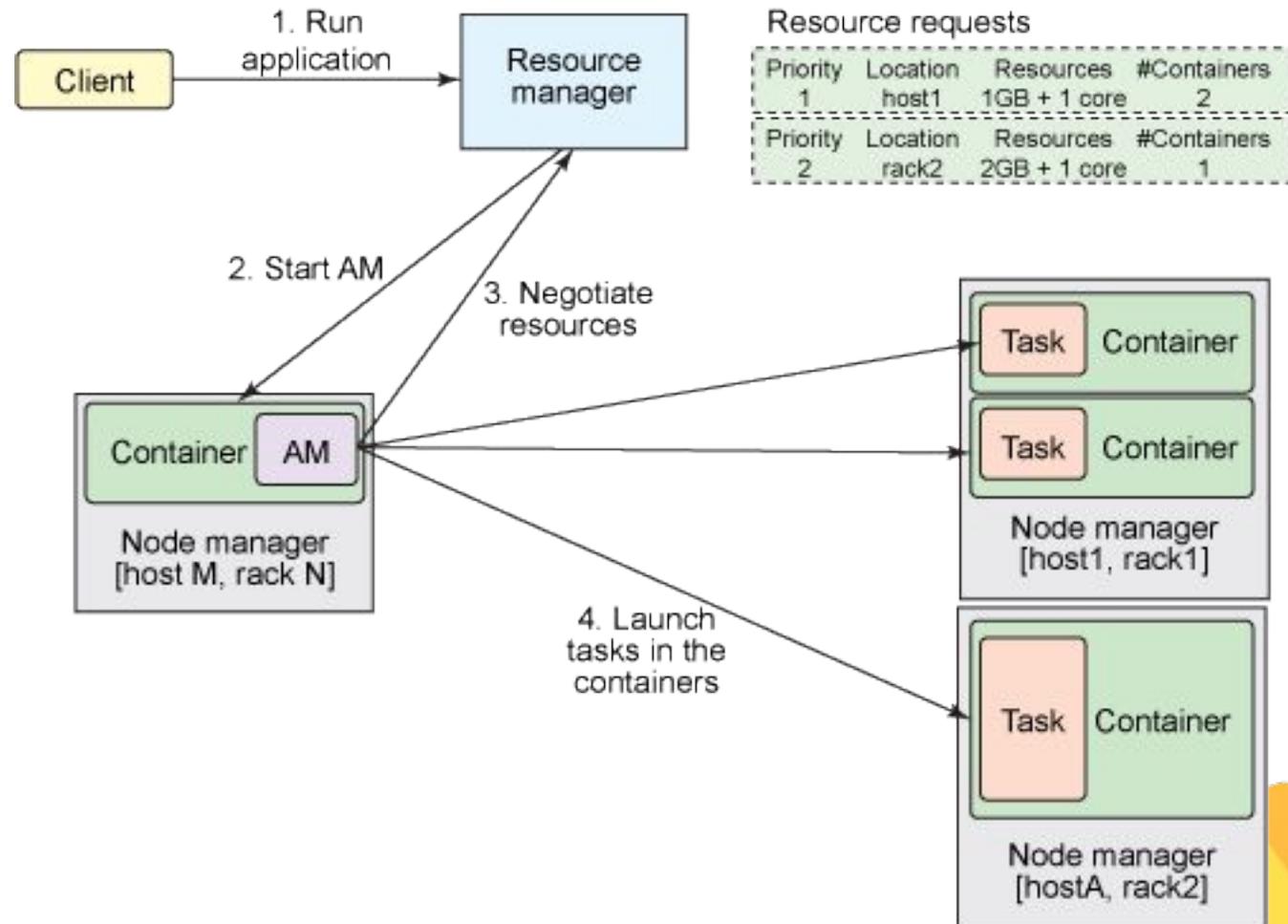
HDFS



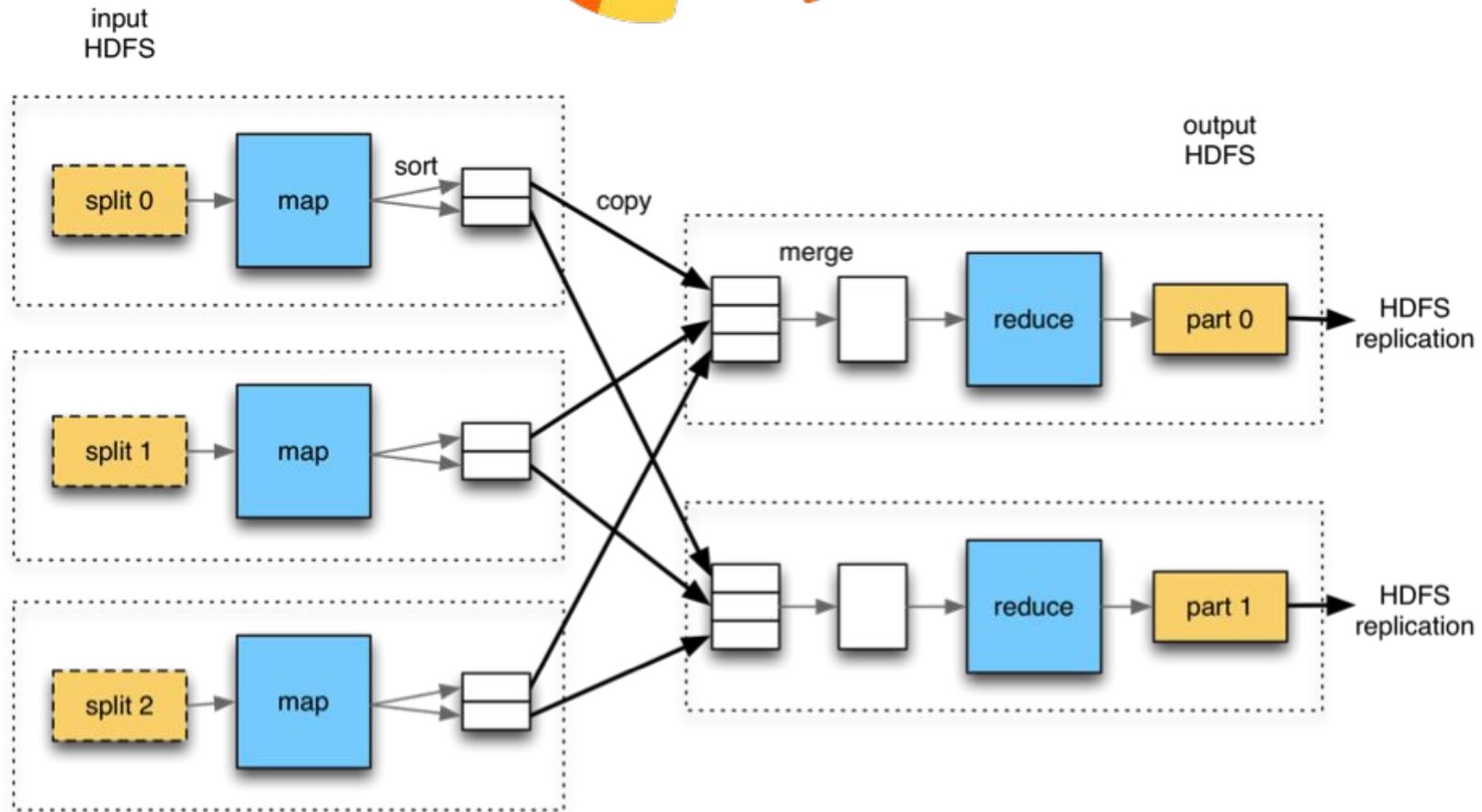
YARN



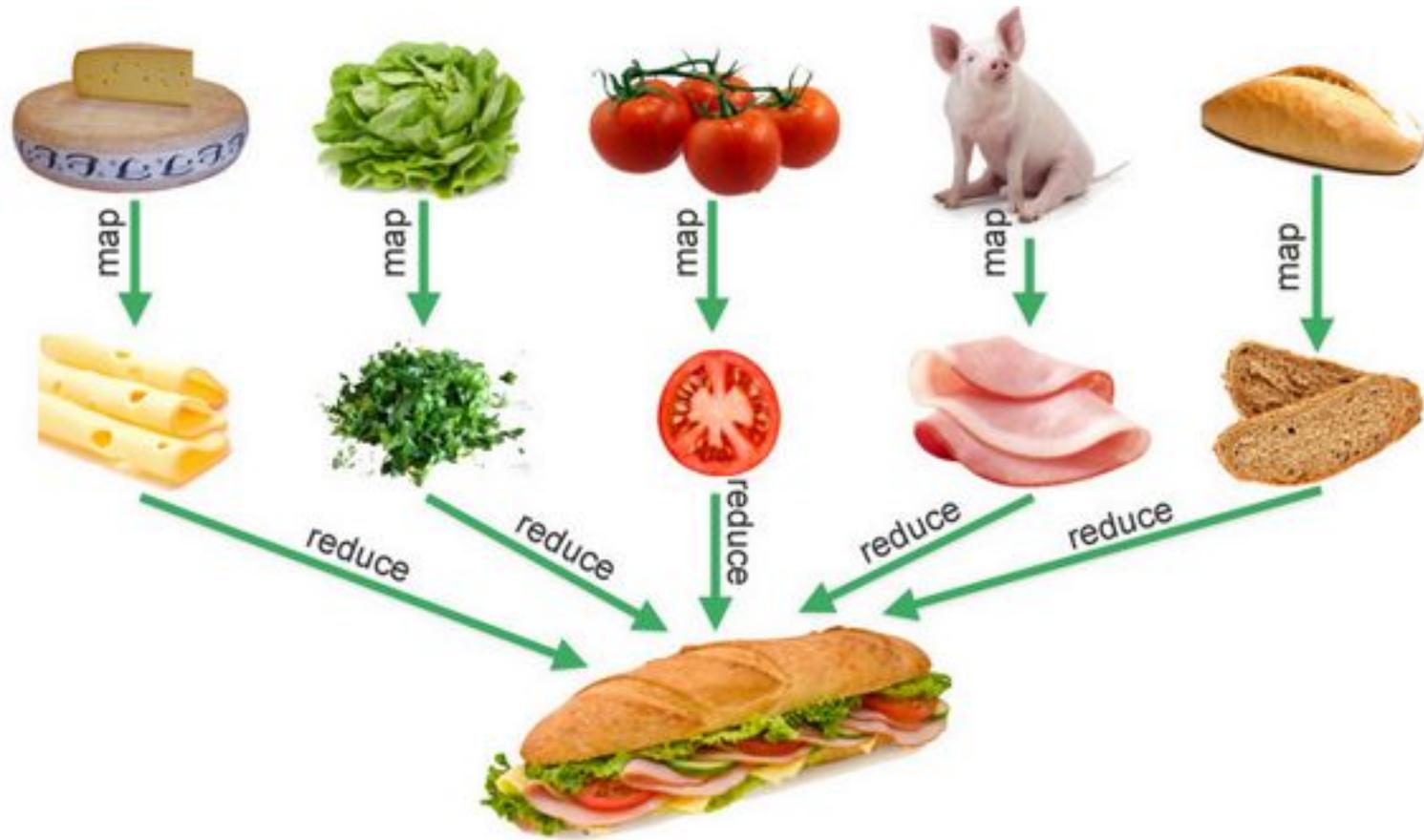
YARN



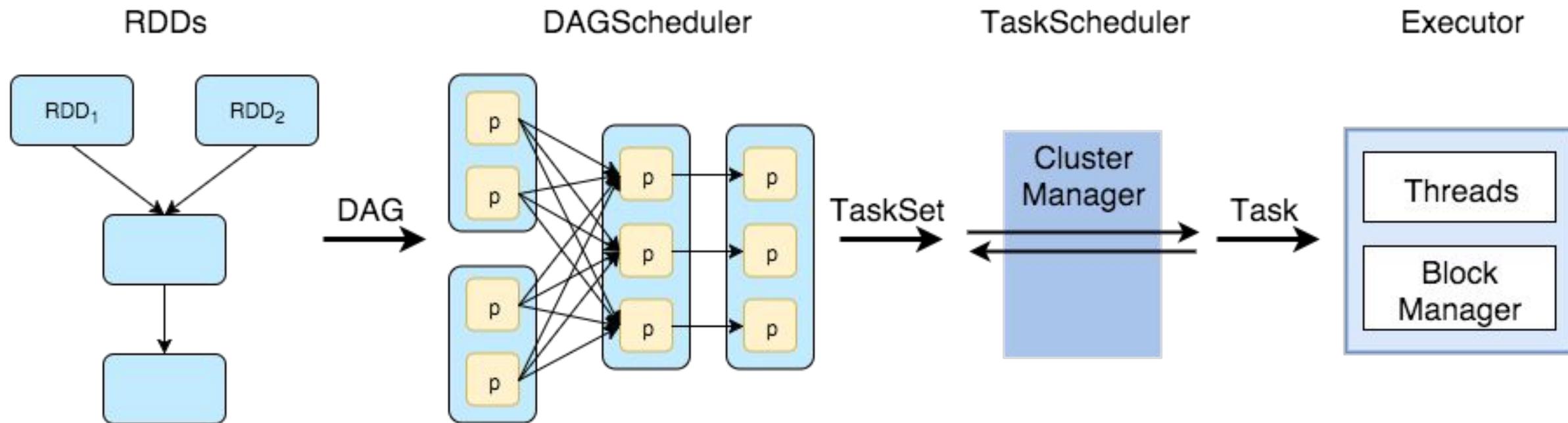
MapReduce



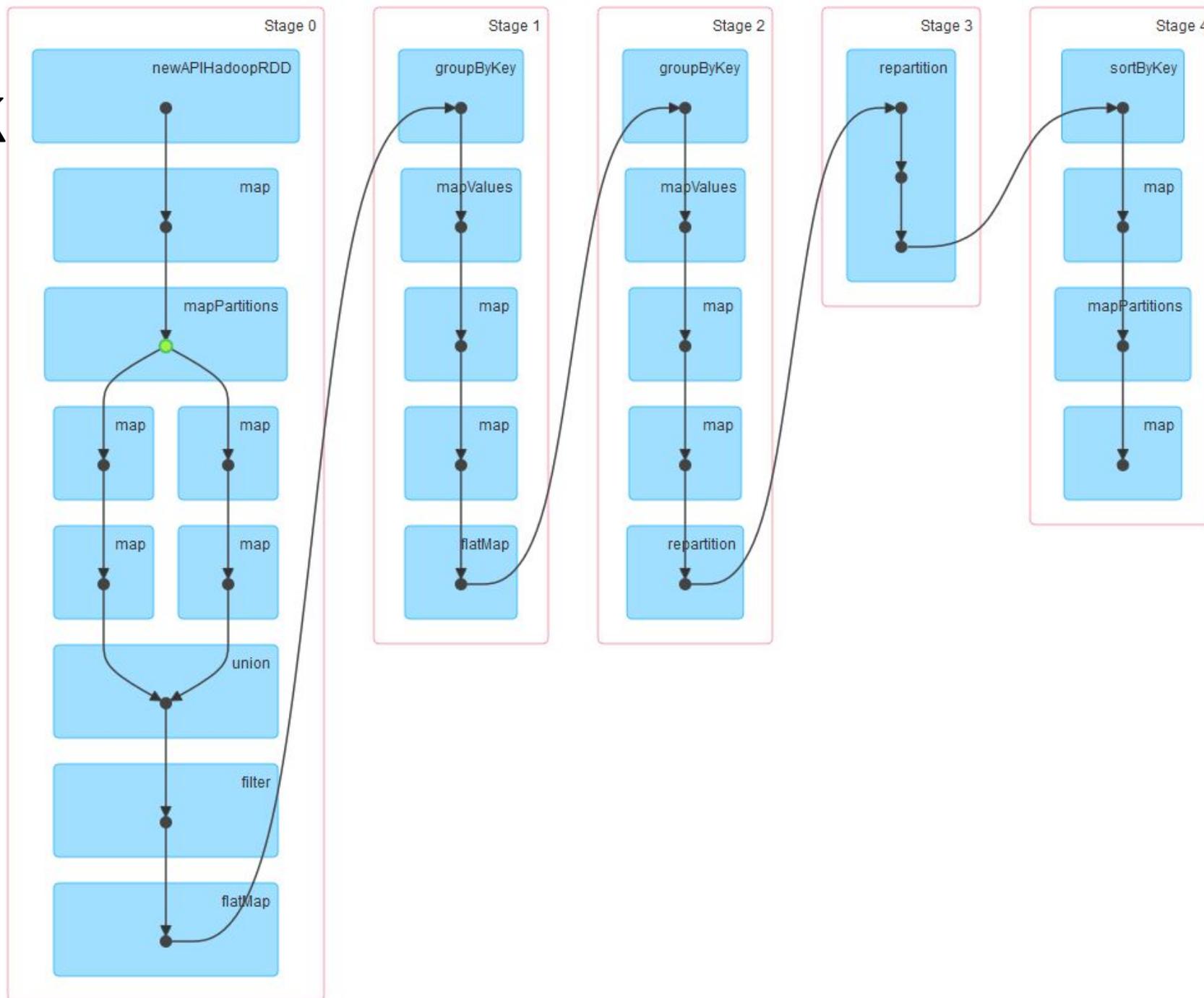
MapReduce



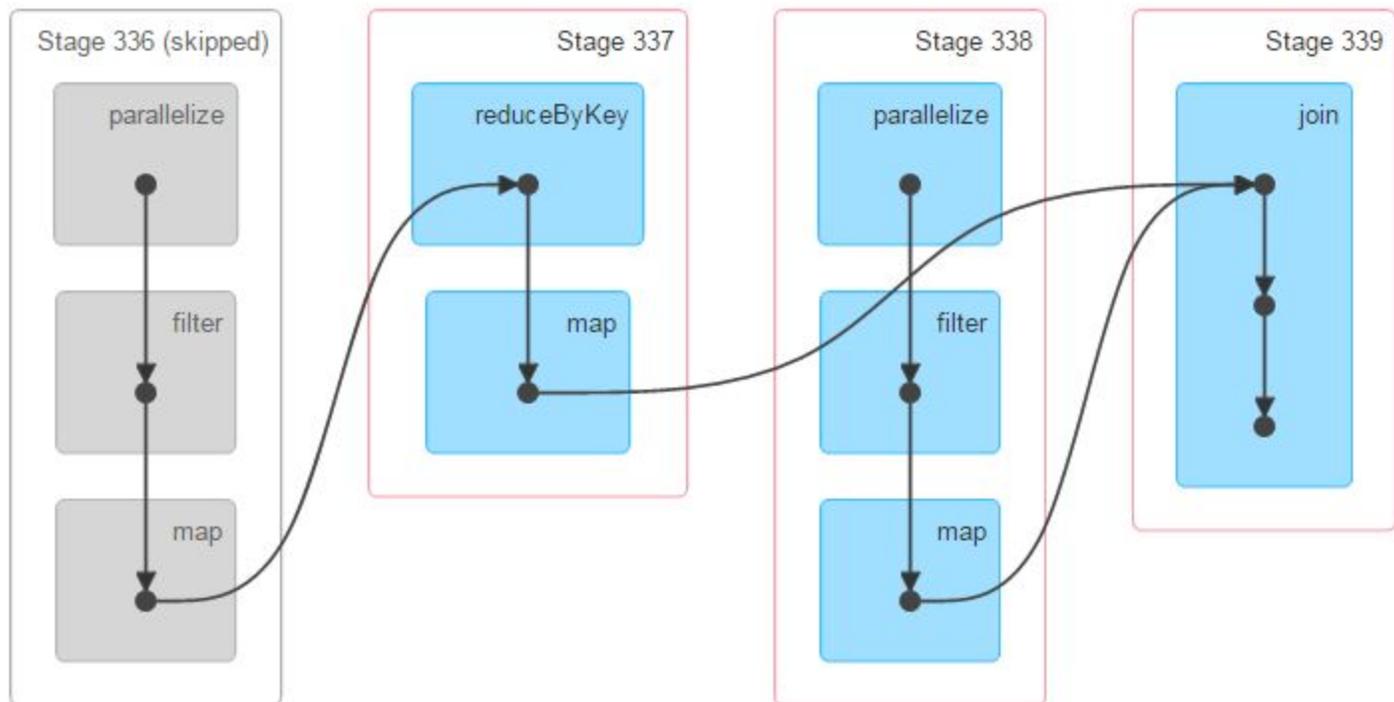
Spark



Spark



Spark





Руслан Габбасов

Руководитель центра разработки



DIRECTUM в г. Уфа

Gabbasov_RI@directum.ru

