

Putting it All Together

USER ID	MID	R	T
196	242	3	881250949
186	302	3	891717742
196	377	1	878887116
244	51	2	880606923
166	346	1	886397596
186	474	4	884182806
186	265	2	881171488

↓
MAPPER

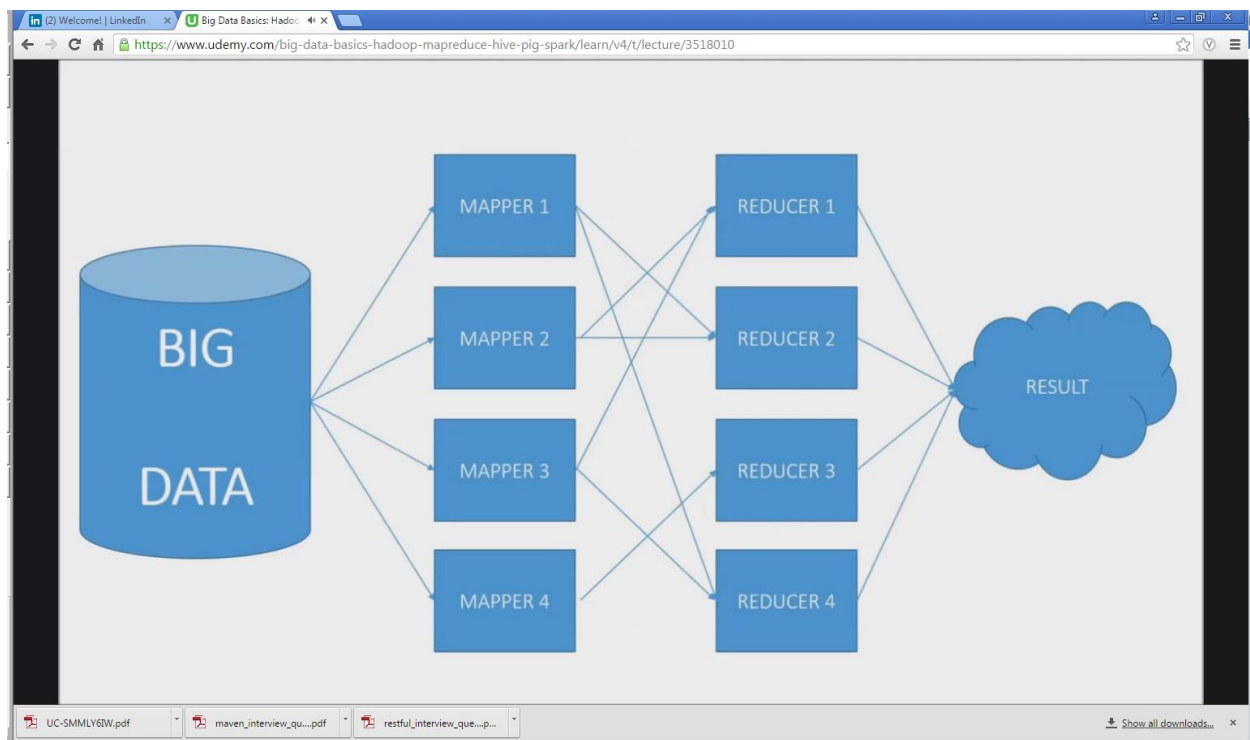
196:242 186:302 196:377 244:51 166:346 186:274 186:265

↓

166:346 186:302,274,265 196:242,377 244:51

↓
REDUCER

166:1 186:3 196:2 244:1



in Welcome! | LinkedIn x efc Banks still hiring in London x Big Data Basics: Hadoop x

← → ↻ 🏠 <https://www.udemy.com/big-data-basics-hadoop-mapreduce-hive-pig-spark/learn/v4/t/lecture/3518030> ☆ 🔍 ☰

Apache Hive

Where does your data come from?



- Hive is a data warehouse that works nicely with Hadoop and MapReduce.
- It uses Hadoop to distribute its storage for very large data sets.
- It allows SQL-like queries (via QL)
- And it also allows queries via MapReduce
- It can work with Amazon Elastic MapReduce
- Initially developed by Facebook

UC-SMMLY6W.pdf maven_interview_qu...pdf restful_interview_que...p... Show all downloads...

in Welcome! | LinkedIn x Big Data Basics: Hadoop x

← → ↻ 🏠 <https://www.udemy.com/big-data-basics-hadoop-mapreduce-hive-pig-spark/learn/v4/t/lecture/3518044> ☆ 🔍 ☰

What is Spark?

An alternative to MapReduce that's gaining popularity

It still uses Hadoop under the hood

Used by Yelp, Washington Post, Yahoo, and reportedly Twitter and other big web companies.

Includes built-in libraries for machine learning (MLlib), SQL, Streaming, and Graph Analysis



UC-SMMLY6W.pdf maven_interview_qu...pdf restful_interview_que...p... Show all downloads...