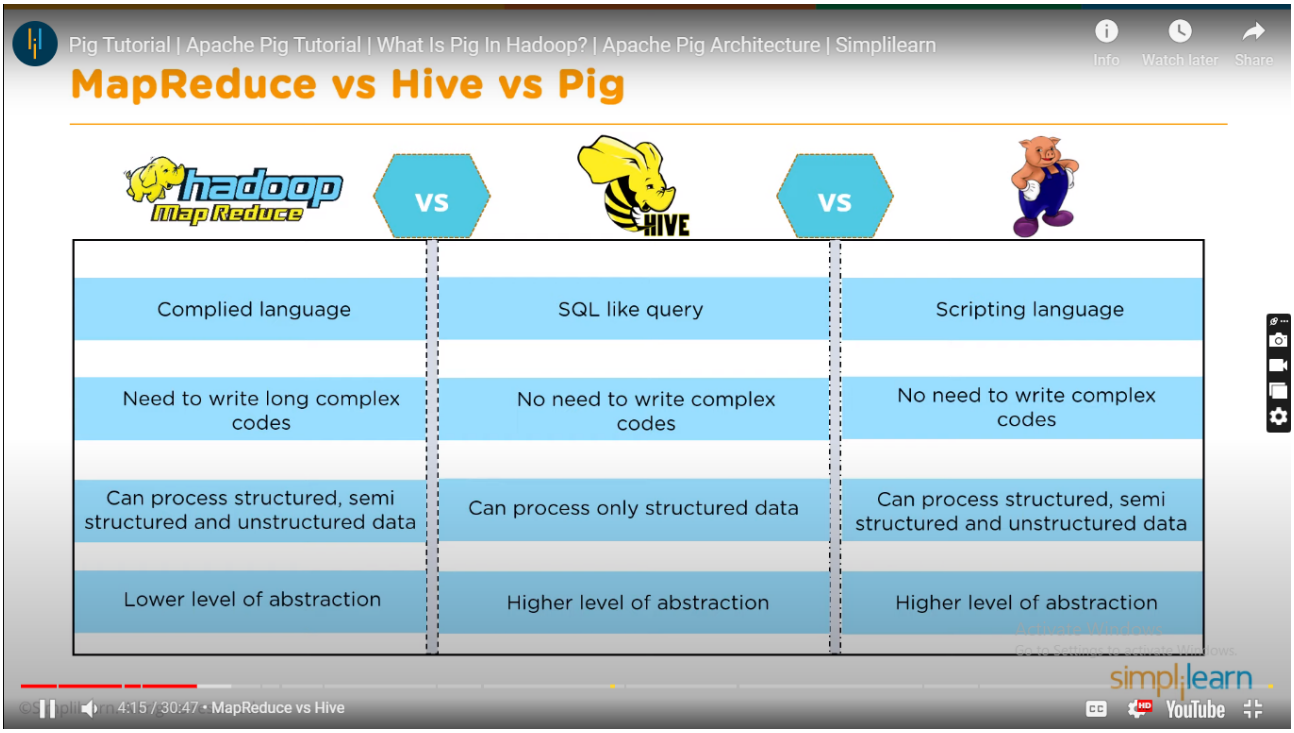
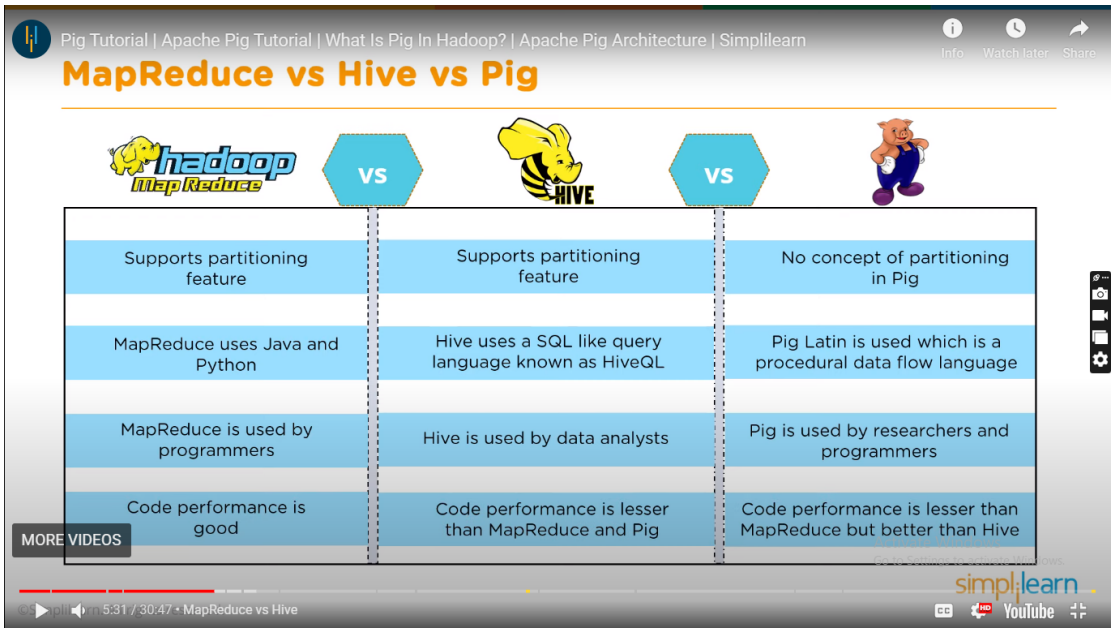
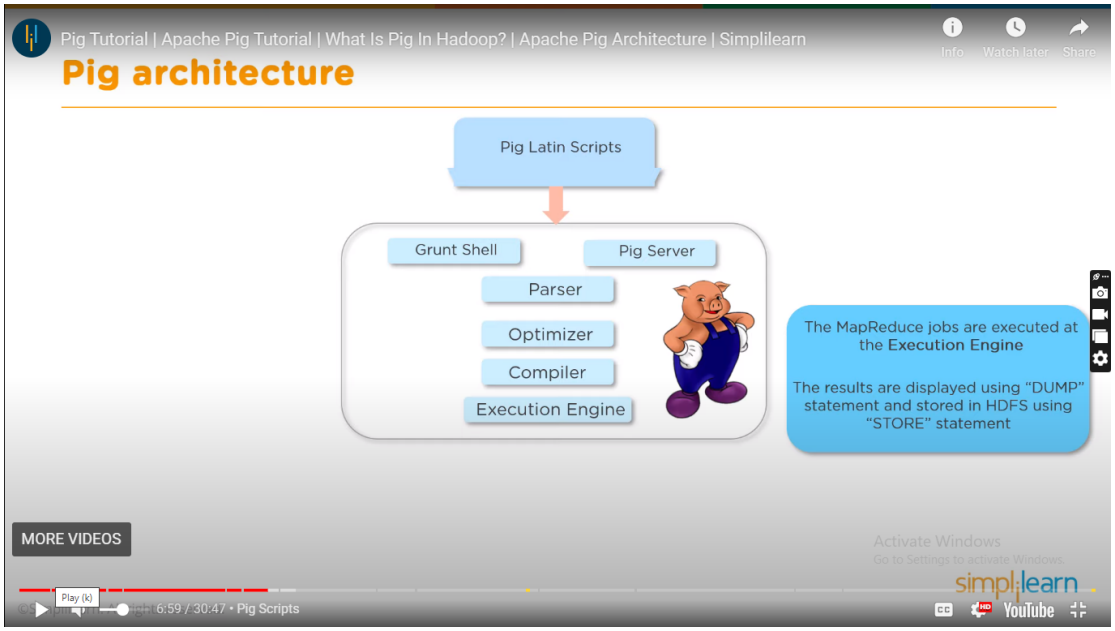
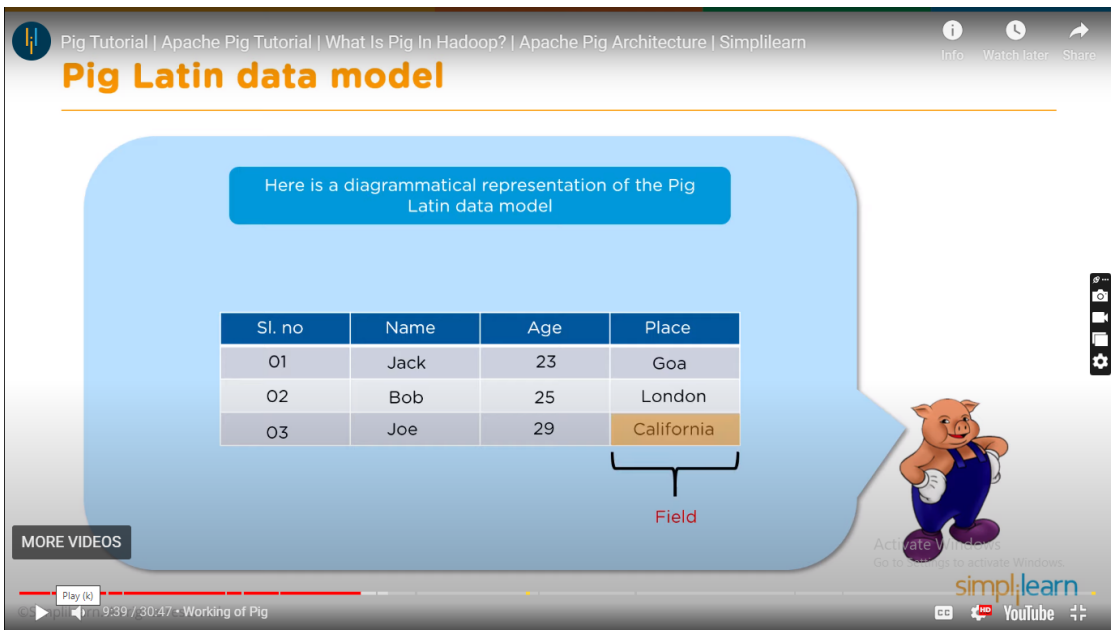
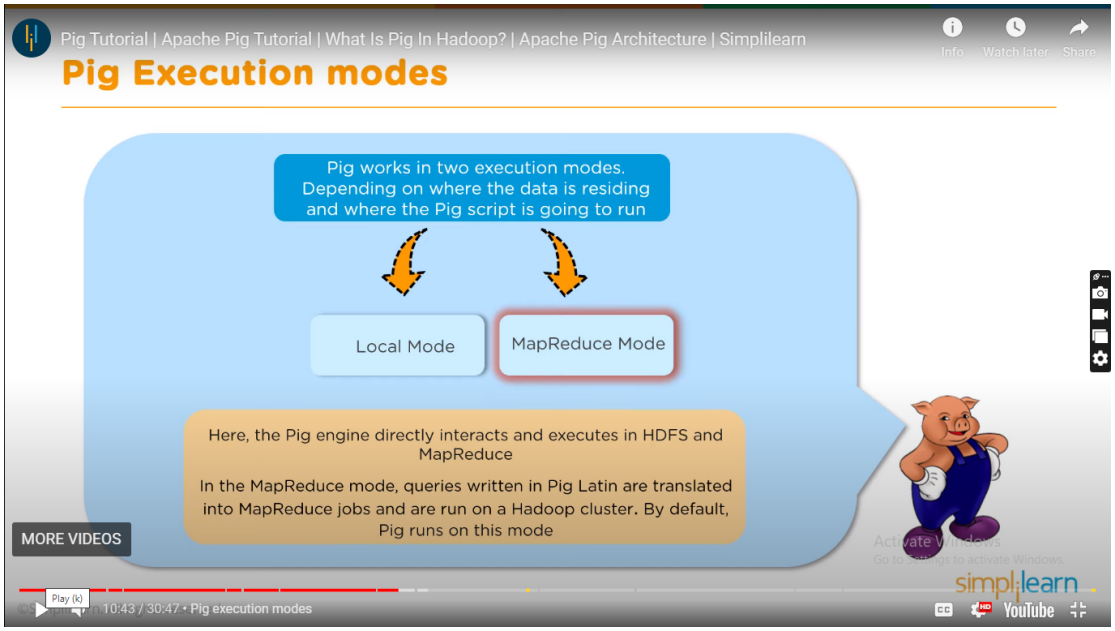
MAP REDUCE VS HIVE VS PIG

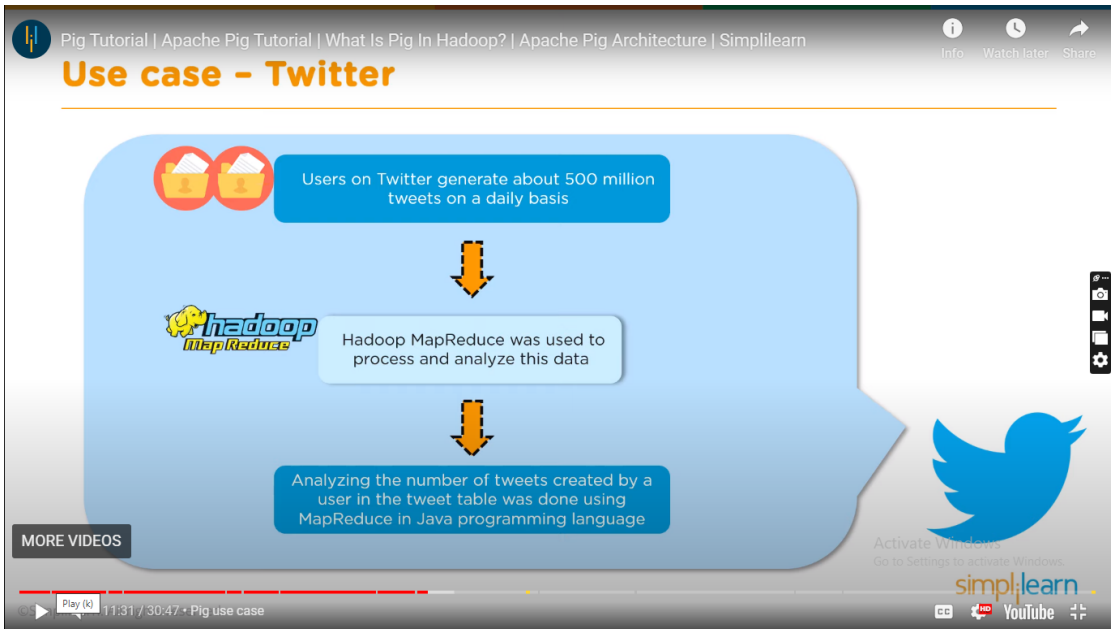


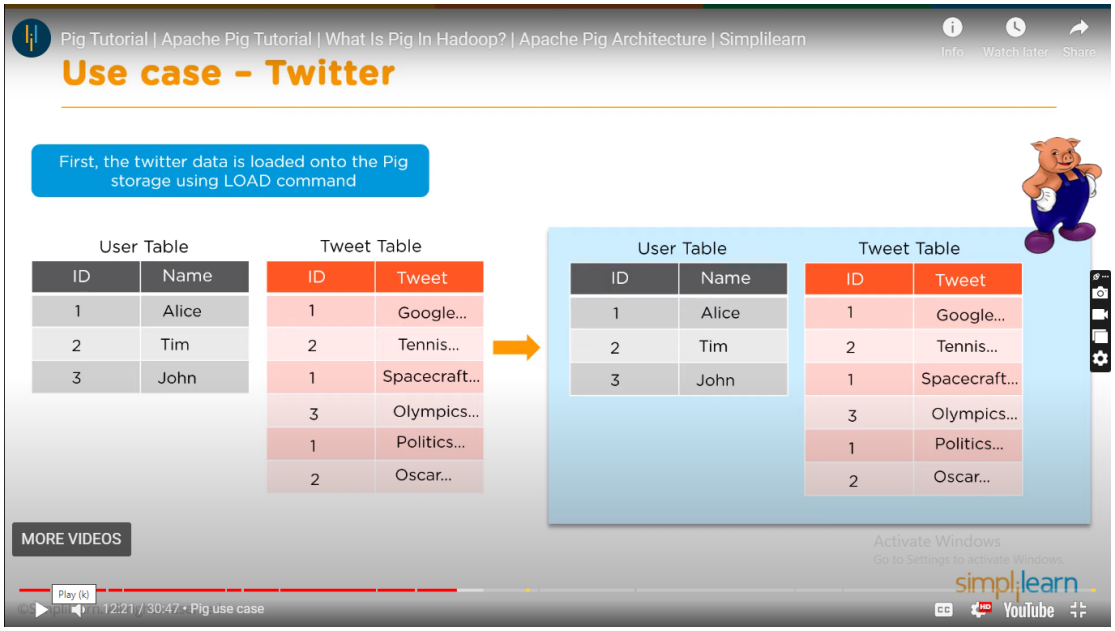


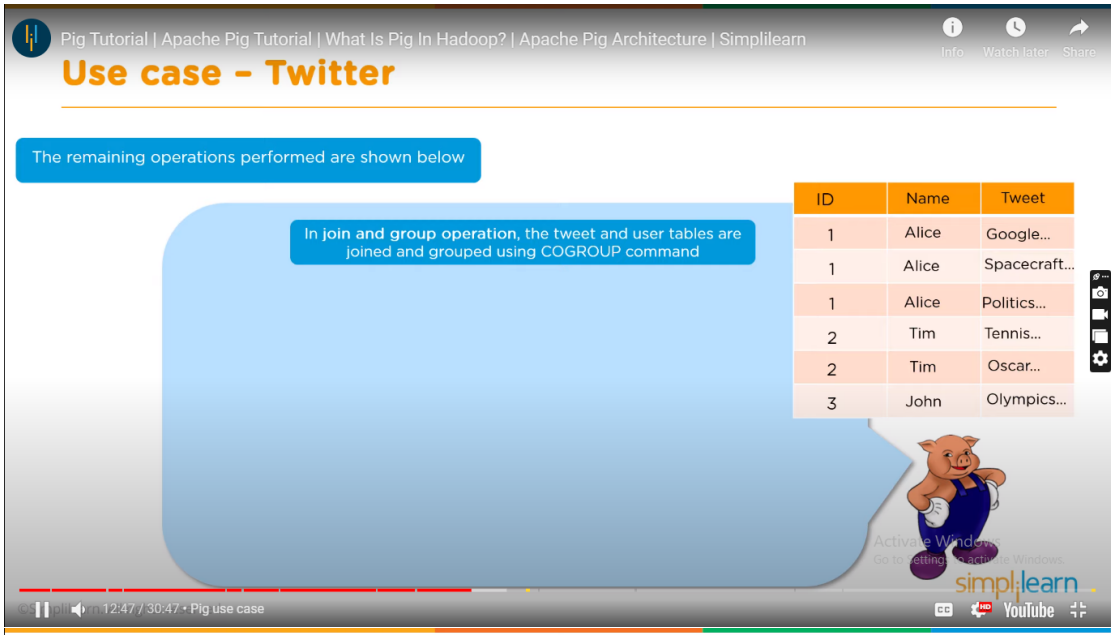

Pig Tutorial | Apache Pig Tutorial | What Is Pig In Hadoop? | Apache Pig Architecture | Simplilearn
Info
Watch later
Share
Pig architecture
Pig Latin Scripts
Grunt Shell
Pig Server
Parser
Optimizer
Compiler
Execution Engine
The MapReduce jobs are executed at the Execution Engine
The results are displayed using "DUMP" statement and stored in HDFS using "STORE" statement
MORE VIDEOS
Activate Windows
Go to Settings to activate Windows.
Play (k)
6:59 / 30:47 • Pig Scripts
simplilearn
YouTube


Pig Tutorial | Apache Pig Tutorial | What Is Pig In Hadoop? | Apache Pig Architecture | Simplilearn
Info
Watch later
Share
Pig Latin data model
Here is a diagrammatical representation of the Pig Latin data model
Sl. no
Name
Age
Place
01
Jack
23
Goa
02
Bob
25
London
03
Joe
29
California
Field
MORE VIDEOS
Activate Windows
Play (k)
9:39 / 30:47 • Working of Pig
simplilearn
YouTube


Pig Tutorial | Apache Pig Tutorial | What Is Pig In Hadoop? | Apache Pig Architecture | Simplilearn
Info
Watch later
Share
Pig Execution modes
Pig works in two execution modes. Depending on where the data is residing and where the Pig script is going to run
Local Mode
MapReduce Mode
Here, the Pig engine directly interacts and executes in HDFS and MapReduce
In the MapReduce mode, queries written in Pig Latin are translated into MapReduce jobs and are run on a Hadoop cluster. By default, Pig runs on this mode
MORE VIDEOS
Play (k)
10:43 / 30:47 • Pig execution modes
simplilearn
YouTube


Pig Tutorial | Apache Pig Tutorial | What Is Pig In Hadoop? | Apache Pig Architecture | Simplilearn
Info
Watch later
Share
Use case – Twitter
Users on Twitter generate about 500 million tweets on a daily basis
hadoop MapReduce
Hadoop MapReduce was used to process and analyze this data
Analyzing the number of tweets created by a user in the tweet table was done using MapReduce in Java programming language
MORE VIDEOS
Play (k)
11:31 / 30:47 • Pig use case
simplilearn
YouTube


Pig Tutorial | Apache Pig Tutorial | What Is Pig In Hadoop? | Apache Pig Architecture | Simplilearn
Info
Watch later
Share
Use case – Twitter
First, the twitter data is loaded onto the Pig storage using LOAD command
User Table
ID
Name
1
Alice
2
Tim
3
John
Tweet Table
ID
Tweet
1
Google...
2
Tennis...
1
Spacecraft...
3
Olympics...
1
Politics...
2
Oscar...
User Table
ID
Name
1
Alice
2
Tim
3
John
Tweet Table
ID
Tweet
1
Google...
2
Tennis...
1
Spacecraft...
3
Olympics...
1
Politics...
2
Oscar...
MORE VIDEOS
Activate Windows
Go to Settings to activate Windows.
simplilearn
Play (k)
12:21 / 30:47 • Pig use case
YouTube



Use case – Twitter

The remaining operations performed are shown below

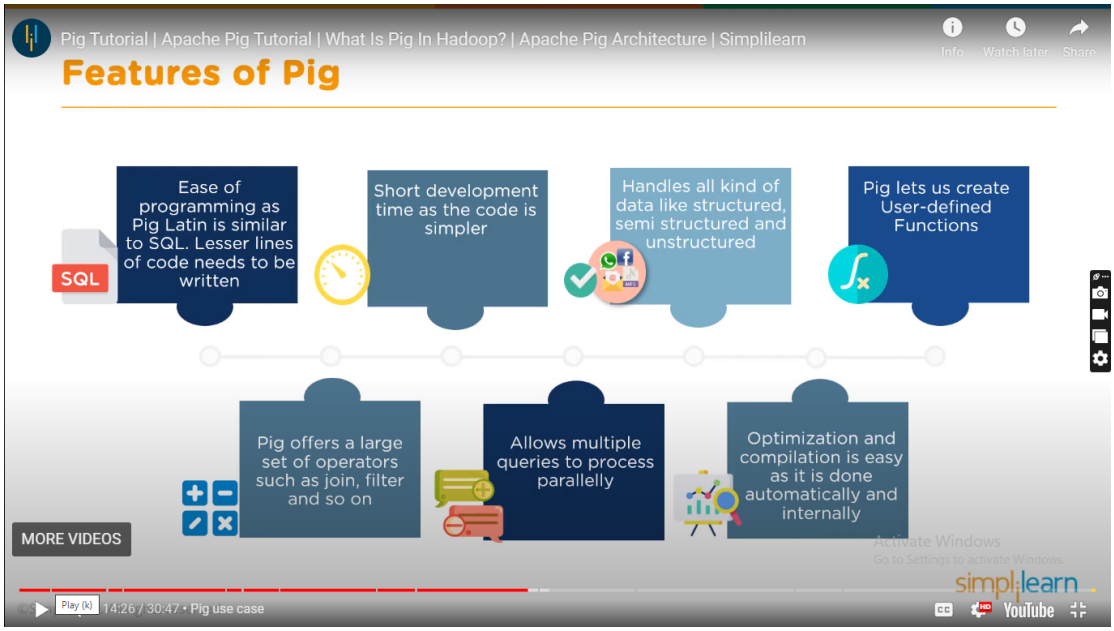
In **join and group operation**, the tweet and user tables are joined and grouped using COGROUP command

The next operation is the **aggregation**, the tweets are counted according to the names. The command used is COUNT

The result after the count operation is joined with the user table to find out the user name

ID	Name	Count
1	Alice	3
2	Tim	2
3	John	1

©Simplilearn. All rights reserved.


Pig Tutorial | Apache Pig Tutorial | What Is Pig In Hadoop? | Apache Pig Architecture | Simplilearn
Info
Watch later
Share
Features of Pig
Ease of programming as Pig Latin is similar to SQL. Lesser lines of code needs to be written
SQL
Short development time as the code is simpler
Handles all kind of data like structured, semi structured and unstructured
Pig lets us create User-defined Functions
Pig offers a large set of operators such as join, filter and so on
Allows multiple queries to process parallelly
Optimization and compilation is easy as it is done automatically and internally
MORE VIDEOS
Activate Windows
Go to Settings to activate Windows.
Play (k)
14:26 / 30:47 • Pig use case
simplilearn
YouTube