

DATA SCIENCE TOOLKIT

@learn.machinelearning



What tools to use for what



IDE tools



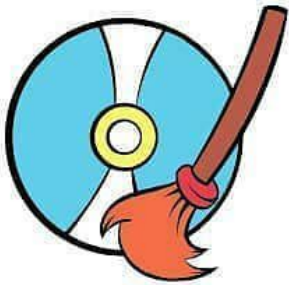
ML Tools



What tools to use for what

| | |
|------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  <p>Real - Time Tools</p> |  kafka  Spark Streaming  STORM  nifi   Flink |
|  <p>data Warehouse tools</p> |  Spark SQL  Google BigQuery  snowflake  S3  HIVE  Greenplum |

What tools to use for what



**Data
Cleaning**

 **cascade**

 **talend**



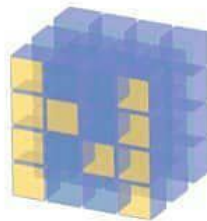
Apache Sqoop



Apache
Airflow



Statistics



NumPy



pandas



julia



GNU Octave



• learn.machinelearning



What tools to use for what

| | |
|--------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  <p>Data Store</p> |  elasticsearch  riak  mongoDB  CouchDB  APACHE HBASE  Cassandra |
|  <p>Data Visualization</p> |  Data-Driven Documents  Bokeh  plotly  tableau  R Studio Shiny  matplotlib |