

DATA SCIENCE TOOLKIT

@learn.machinelearning



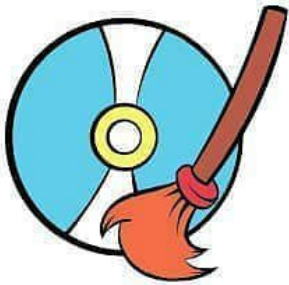
What tools to use for what

 <p>IDE tools</p>	    
 <p>ML Tools</p>	       

What tools to use for what

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

What tools to use for what



Data
Cleaning



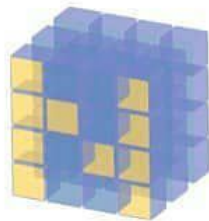
Apache Sqoop



Apache
Airflow



Statistics



NumPy



julia



GNU Octave

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  tableau  plotly  R Studio Shiny  matplotlib