From the tools perspective, a Data Scientist **must** know the following

* **Data Fetching Tools:** In most of the cases, the data is captured and stored in databases and the DS should be able to pull the data to do her analysis. So knowledge on **SQL** is needed.
* **Data Analysis Tools :** Once the data is fetched, the DS should be able to do all the necessary preprocessing / cleaning on the data and then to carry out her analysis / modeling. So knowledge on atleast one of the **data analysis tools like Python, R** is must.
* **Data Visualization Tools :**This is often overlooked but this is as important as the previous two. Once the analysis / modeling is done and the results are ready, they needed to be communicated in an effective way to the business stakeholders. I have seen plots / visualizations to be more effective in conveying the results rather than words / writings. So knowledge of **visualization tools like tableau, qlik**or knowledge on **plotting libraries like ggplot, matplotlib** is also important.

In addition to this, some other good to have tools are

* If the DS has to deal with **huge datasets**, then tools like **Hadoop, Spark** will be very useful.
* If the DS is doing analysis / models on **unstructured data** like image / text / voice, then knowledge on **deep learning tools like TensorFlow, Torch** will be useful.
* Also there will be times, when the DS has to fetch the data from web or APIs, so knowledge on some **basic programming / scripting** will be a plus.
* One another important tool without which the answer will not be complete is **Spreadsheets**

**------------------------------------------------------------------------------------------------------**

**Tools for data pulling & pre-processing**

**a. SQL**

This is a must skill for all data scientists, regardless of whether you are using structured or unstructured data. Companies are using latest SQL engines like Apache Hive, Spark-SQL, Flink-SQL, Impala, etc.

**b. Big Data Technologies**

This is the must out of the Skills Needed to Become a Data Scientist. The data scientist needs to know about different Big Data technologies like Hadoop and its ecosystem, Spark and Flink if possible.

**c. UNIX**

As most raw data is stored on a UNIX or Linux server before it’s put in a data-store so it’s nice to be able to access the raw data without the dependency of a database. So Unix knowledge is good for Data Scientists.

**d. Python**

Python is a most popular language for the data scientist. Python is an interpreted, object-oriented programming language with dynamic semantics. It is a high-level language with dynamic binding and typing.

**Tools for Data Analysis & pattern matching**

This depends on your level of statistical knowledge. Some tools are used for more advanced statistics and some for more basic statistics.

**a. SAS**

Lots of companies use SAS, so some basic SAS understanding is good. You can manipulate equations easily.

**b. R**

Ris most popular in the statistical world. R is an open-source tool and language that is object oriented, so you can use that anywhere. It is the first choice of any data scientist as most things are implemented in R.

**c. Machine Leaning**

Machine learning is the most demanding and most useful tool the data scientists must have. There are lots of machine learning tools are available in the market like weka, nltk, etc. but machine learning tools on top of big data technologies are grabbing industry attention like Mahout, MLlib, FlinkML.

**Tools for Visualization**

**a. Tableau**

It is a popular tool, especially in Silicon Valley.Apart from above-mentioned tools following tools are also popular – JasperSoft, SAP BI, QlikView, MicroStrategy, etc.

**Non-Technical Skills**

**a. Business Acumen**

**b. Communication Skills**

Companies are searching for data scientists who can clearly and confidently translate their insights on the data to other teammates. A data scientist arms them with quantified insights.

**c. Analytical Problem-Solving**

Analytical problem-solving skill is highly demanding for Data Scientist so that the right approach can be used to get maximum output in available time and resources.