

Tools of a Data Scientist

Data scientist use a variety of tools to obtain and manipulate data. Below are a list of tools and a description of how they are used by data scientists.

- **Measurement:** Measurement is meant as a tool to guide policies and insights. Whatever the goal, it must be measurable in some way. Data scientist use data to identify (or measure) the success or failure of a policy.
- **Programming Languages:** These are useful tools of data scientist because it enables them to use massive sets of data that otherwise would be too large to draw insights from without such languages.
- **Web Scraping:** An important tool for obtaining data
- **Handling Data:** Because of the massive size of some data, it cannot all store on one device. Data scientist use Resilient Distributed Datasets (RDDs) to distribute data across multiple devices. SQL is also helpful when obtaining sections of data from a large database.
- **Visualization:** Helpful in conveying data visually to others
- **Modeling:** Using the data to develop statistical models that will provide insight about the data