**Big Data Tools for Analysis**

Details of few tools used for big data analysis are listed below

1. **Microsoft HDInsight**

Azure HDInsight is a Spark and Hadoop service in the cloud. It provides big data cloud offerings in two categories, Standard and Premium. It provides an enterprise-scale cluster for the organization to run their big data workloads. Azure HDInsight is an easy, cost-effective, enterprise-grade service for open source analytics that enables customers to easily run popular Apache open source frameworks including Apache Hadoop, Spark, Kafka, and others. The service is available in 27 public regions and Azure Government Clouds in the US and Germany.

Azure HDInsight powers mission critical applications in a wide variety of sectors and enables a wide range of use cases including ETL, streaming, and interactive querying.

# Features OF HDInsight

* Reliable analytics with an industry-leading SLA
* It offers enterprise-grade security and monitoring
* Protect data assets and extend on-premises security and governance controls to the cloud
* High-productivity platform for developers and scientists
* Integration with leading productivity applications
* Deploy Hadoop in the cloud without purchasing new hardware or paying other up-front costs

Note: To download <https://azure.microsoft.com/en-in/free/>

More details <https://azure.microsoft.com/en-us/services/hdinsight/>

1. **SkyTree**

Skytree is a big data analytics tool that empowers data scientists to build more accurate models faster. It offers accurate predictive machine learning models that are easy to use.

# Features of SkyTree

* Highly Scalable Algorithms
* Artificial Intelligence for Data Scientists
* It allows data scientists to visualize and understand the logic behind ML decisions
* SkyTree via the easy-to-adopt GUI or programmatically in Java
* Model Interpretability
* It is designed to solve robust predictive problems with data preparation capabilities
* Programmatic and GUI Access

Note: To download <http://www.skytree.net/>

1. **Talend**

Talend is a big data tool that simplifies and automates big data integration. Its graphical wizard generates native code. It also allows big data integration, master data management and checks data quality.

# Features of Talend

* Accelerate time to value for big data projects
* Simplify ETL & ELT for big data
* Talend Big Data Platform simplifies using MapReduce and Spark by generating native code
* Smarter data quality with machine learning and natural language processing
* Agile DevOps to speed up big data projects
* Streamline all the DevOps processes

Note: To download  <https://www.talend.com/download/>

1. **Splice Machine**

Splice Machine is a big data analytic tool. Their architecture is portable across public clouds such as AWS, Azure, and Google.

# Features of Splice Machine

* It can dynamically scale from a few to thousands of nodes to enable applications at every scale
* The Splice Machine optimizer automatically evaluates every query to the distributed HBase regions
* Reduce management, deploy faster, and reduce risk
* Consume fast streaming data, develop, test and deploy machine learning models

**Note: To download** [https://www.splicemachine.com](https://www.splicemachine.com/)

### **Spark:**

Apache Spark is a powerful open source big data analytics tool. It offers over 80 high-level operators that make it easy to build parallel apps. It is used at a wide range of organizations to process large datasets.

**Features of Spark:**

* It helps to run an application in Hadoop cluster, up to 100 times faster in memory, and ten times faster on disk
* It offers lighting Fast Processing
* Support for Sophisticated Analytics
* Ability to Integrate with Hadoop and Existing Hadoop Data
* It provides built-in APIs in Java, Scala, or Python

**Note: To download** <https://spark.apache.org/downloads.html>

### **Plotly**:

Plotly is an analytics tool that lets users create charts and dashboards to share online.

**Features:**

* Easily turn any data into eye-catching and informative graphics
* It provides audited industries with fine-grained information on data provenance
* Plotly offers unlimited public file hosting through its free community plan

**Download link:** <https://plot.ly/>

### **Apache SAMOA:**

Apache SAMOA is a big data analytics tool. It enables development of new ML algorithms. It provides a collection of distributed algorithms for common data mining and machine learning tasks.

**Download link:** <https://samoa.incubator.apache.org/>

### **Lumify**:

Lumify is a big data fusion, analysis, and visualization platform. It helps users to discover connections and explore relationships in their data via a suite of analytic options.

**Features:**

* It provides both 2D and 3D graph visualizations with a variety of automatic layouts
* It provides a variety of options for analyzing the links between entities on the graph
* It comes with specific ingest processing and interface elements for textual content, images, and videos
* It spaces feature allows you to organize work into a set of projects, or workspaces
* It is built on proven, scalable big data technologies

**Download link:** <http://www.altamiracorp.com/index.php/lumify/>

### **Elasticsearch**:

Elasticsearch is a JSON-based Big data search and analytics engine. It is a distributed, RESTful search and analytics engine for solving numbers of use cases. It offers horizontal scalability, maximum reliability, and easy management.

**Features Of Elasticsearch:**

* It allows combine many types of searches such as structured, unstructured, geo, metric, etc
* Intuitive APIs for monitoring and management give complete visibility and control
* It uses standard RESTful APIs and JSON. It also builds and maintains clients in many languages like Java, Python, NET, and Groovy
* Real-time search and analytics feature to work big data by using the Elasticsearch-Hadoop
* It gives an enhanced experience with security, monitoring, reporting, and machine learning features

**Note: To Download:** <https://www.elastic.co/downloads/elasticsearch>

### **R-Programming:**

R is a language for statistical computing and graphics. It also used for big data analysis. It provides a wide variety of statistical tests.

**Features Of R:**

* Effective data handling and storage facility,
* It provides a suite of operators for calculations on arrays, in particular, matrices,
* It provides coherent, integrated collection of big data tools for data analysis
* It provides graphical facilities for data analysis which display either on-screen or on hardcopy

**Note: To Download:** <https://www.r-project.org/>

### 11) IBM SPSS Modeler:

IBM SPSS Modeler is a predictive big data analytics platform. It offers predictive models and delivers to individuals, groups, systems and the enterprise. It has a range of advanced algorithms and analysis techniques.

**Features:**

* Discover insights and solve problems faster by analyzing structured and unstructured data
* Use an intuitive interface for everyone to learn
* You can select from on-premises, cloud and hybrid deployment options
* Quickly choose the best performing algorithm based on model performance

**Download link:** [https://www.ibm.com/us-en/marketplace/spss-modeler/purchase#product-header-top](https://www.ibm.com/us-en/marketplace/spss-modeler/purchase)

Few more opensource tools for big data