Big Data Analytics

Advanced Analytics

Machine Learning

Deep Learning

Natural Language Processing

Computer Vision

# Uses of Big Data and Analytics in Telecom

Problems with Telecom Industry

1. Decreasing Margins
2. Congested Networks

85% respondents indicate Use of Big Data & Analytics is creating competitive advantage

Analysis of Network Traffic

Analysis of Call Data Records

Improve Quality of Service

Efficient Routing of Traffic

Fraud Identification & action by analyzing Call Data Records in Real time

Marketing can tailor campaign to better target Customers

Use insights to develop new products and services

Marketing

Mobile Analytics

5 Skill You need to know to become a Big-Data Analyst

|  |  |
| --- | --- |
| Skill | Description |
| Programming | **R**, **Python, Java**, C++, Ruby, SQL, Hive, SAS, SPSS, MATLAB, Weka, Julia, **Scala** |
| Data Warehousing | Relational – MySQL, Oracle, DB2  Non-Relational – Hbase, HDFS, MongoDB, CouchDB, Cassandra, Teradata |
| Computational Frameworks | Apache Spark, Apache Storm, Apache Samza, Apache Flink, Classic MapReduce and Hadoop |
| Quantitative Aptitude and Statistics | Statistics, Linear Algebra, Summary Statistics, Probability Distributions, Random Variables, Hypotheses Testing  Machine Learning |
| Business Knowledge | Good Knowledge of Domain |

How much data is handled by Relational Databases, Non-Relational Databases, Data Warehouses,

Big Data Analytics course (edX) – University of Adelaide

Big Data Analytic Products/Solutions Architecture and Development

<https://www.linkedin.com/in/harsha-s-adiga-08b1792/>

<https://www.linkedin.com/in/manoj-karanth-2026485/>

Languages: R, Python, Java, Scala

Hadoop Eco-System: Hadoop, MapReduce, Pig, Hive, Sqoop, Oozie, HBase, Zookeeper, Yarn

Spark Eco-system: Spark

BDA in Cloud: AWS

Machine Learning – Linear Regression, Logistic Regression, Decision Trees, Clustering, Support Vector Machine, pandas, PySpark, Scikit-learn,

Real-Time Streaming Analytics: Apache Kafka, Flume,

Search: Solr

Hadoop Distributions: Cloudera, Hortonworks

Data Visualization: Pentaho, D3,

No-SQL DB: Cassandra, MongoDB

<https://www.analyticsvidhya.com/blog/2018/05/24-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/>

1) what **is** the file size you’ve used? Reason Behind your answer.

2) How **long** does it take to run your script **in** production cluster? How did you optimized the timings. Challenges you have faced.

3) what was the file size **for** production environment?

4) Are you planning **for** anything to improve the performance?

5) what size of file **do** you **use** **for** Development?

6) what did you **do** to increase the performance(Hive,pig)?

7) what **is** your cluster size? Reason Behind your answer.

8) what are the challenges you have faced **in** your project? Give 2 examples?

9) How to debug production issue?(logs, script counters, JVM)

10) how **do** you **select** the eco system tools **for** your project?

11) How many nodes you are **using** currently?

12) what **is** the job scheduler you **use** **in** production cluster?

<https://acadgild.com/blog/understanding-big-data>

Steps Involved in Big Data Analysis

1. Identification of suitable storage for Big Data
2. Data Ingestion
3. Data Cleaning & Processing (Exploratory Data Analysis)
4. Visualization of the Data
5. Learning/Analysis - Apply Machine learning Algorithms (if required)