

## Trainer Profile – Arya Awasthi

---

Total Exp: 10+ years

Training Exp: 5+ years

Training Area: BigData/Hadoop, Spark, Storm, Lambda Architecture, Machine Learning

### Summary:

- Working experience with companies like Yahoo, Amazon, IBM and StartUps.
- Strong hold on Data Structure, Algorithm and Machine Learning and BigData Technologies.
- 6+ yrs of exp of Analysis and Optimization of BigData using various Bigdata tools and methodologies like Hadoop, Storm, Spark, NoSQL.
- Also experience in designing and deploying of efficient Cloud Architecture (VPC) on AWS.
- Last Project – Duplicate detection of news articles using Spark MLib(SVM), Spark Streaming, Hbase with Meso and Hadoop(YARN) Cluster.

### Training Summary:

- Successfully delivered more than 60 Customized Corporate Trainings and Workshops in India and Abroad of BigData/Hadoop, Storm, Spark, Lambda Architecture, NoSQL domain in companies like Dell, Toshiba R&D, Xerox, IBM, Volvo, Accenture, Wipro, Adobe, Intel, Nokia, Caterpillar, Oracle, HP, Deloitte, Global Logic, Yodlee and many Startups.
- Also assisted and helped many companies for development of their Bigdata Project & POC.
- Conducted many Regular BigData/Hadoop Developer classroom/online training programs through many leading Training Companies including official Cloudera -Data Analyst Training: Using Pig, Hive, & Impala with Hadoop and -Spark & Hadoop Developer Training.
- Working as Subject Matter Expert and also conducted many BigData/Hadoop Interviews.
- Recently Installed & Configured 200 Nodes Hadoop Cluster for a Startup Company using 10 Racks each having 20 Nodes.

### Technologies:

- **BigData/Hadoop, Storm, Spark**
  - Expert Knowledge in Hadoop(0.X,1.X,2.X) and its Ecosystem Components i.e
  - HDFS, MapReduce, YARN, PIG, HIVE, Apache Tez, Impala, Apache Drill,
  - Flume, Fluentd, Kafka, Oozie, Azkaban, Zookeeper, Kafka NiFi
  - Spark : –RDD, Shark/Spark SQL, MLib, Realtime data analysis using Spark Streaming, Lambda Architecture Development (kafka+Spark+NoSql)
  - Storm : - API development(Storm+Kafka+Cassandra)
  - NoSQL : HBase, MongoDB, Neo4j, Solr, Cassandra, Yahoo's Cluster
  - Commercial Framework Distribution: CDH(3.X,4.X, 5.X), HortonWorks, Map R
  - Certification : Cloudera Certified Hadoop Developer
- **Cloud Computing & Virtualization**
  - Expert Knowledge in AWS( EC2, S3, EMR, RDS....) & Private Cluster.
- **Languages**: Scala, Java, Python, Shell (bash), C/C++
- **Web**: Restful web services
- **Databases**: MySQL with Lucene, Solr, PostgreSQL, NoSQL
- **OS & ID** : Linux/MacOS/Windows, Eclipse, GIT