nayna patel

Big Data consultant

Phone: +44 7931747121

Email: nayna.patel@appexperts.me.uk

**SKILLS**

Operating System

Windows, Linux

Big Data Technology

Hadoop, Kafka, Spark, Zookeeper, Sqoop, Flume, Spark SQL, Cloudera Horton Work HDP platform

Other Technology

Microsoft Word, Excel, Outlook, Power Point, Oracle virtual box/manager, VM ware virtual manager,

Azure, API’s

Data Structures/Storage

Text files, CSV files, SSMS TSQL, MySQL, Hadoop Distributed File system (HDFS) , Azure Data Lake, AWS S3, Hive, Hbase, MongoDB, Basic Avro

Programming Languages

Python, MySQL SQL, Hive SQL, Hbase SQL, Hadoop MapReduce, Hadoop Distributed File system, Spark Resilient distributed dataset

Command scripting/formula

Excel, Windows, Linux, Ubuntu

Business Applications

Human Resources, Recruitment, FX Trading, Clients Reporting, On Boarding, Audits, Procurement, Insurance, Accounting, Reinsurance

Strengths and Soft skills

Stake Holder Management, Communications, requirements/information gathering, data collecting cleaning and processing

EXPERIENCE

My experiences include working with various types of data and have used data in setting up a database, developing programs/algorithms for business applications, data cleaning (correction of corruption, error, loss and missing values) , data extraction (selecting data sample and feature extraction), data preparation.

My first experience with data started after completing my degree in computer science in 1984. As an analyst/programmer within mainframe technologies I gained experience in waterfall methods from application design to application system testing.

Due to diversification in the market, my last 15 years has been in the business management administration/analyst area where I have collected, cleaned and prepared data using Microsoft technologies for visualization which is used to make decision in business.

**Big Data Technology experiences**

* Setting up Virtual Machines and Instances using Oracle/VM Virtual Manager, AWS, Azure
* Installing software on virtual machines and operating system scripting and updating of configuration files and jar files
* Hadoop single node cluster setup in pseudo-distributed mode (name node and data node on same virtual machine) Hadoop
* Hadoop multi node cluster setup. Three virtual machines were used for one master and two slaves. All machines had a data node set.
* Extraction of data from SQL database and text files and loading in Hadoop HDFS
* Using data from files to manipulate and print data using PySpark (Python Spark Libraries)
* Using data from files with python Kafka libraries to transfer data from producer to consumer (using two servers Kafka and zookeeper)
* Streaming real time tweets data for selected input from Twitter API to Kafka Producer and then using Spark Consumer to create a RDD to split the data files and load it on the data frame using Spark Context and Spark Session
* Search stored information from Spotify API , selected and formatted required information and passed via Kafka Producer to Spark Consumer to create a RDD to process and count data and created a schema and loaded it to the Spark data frame using SparkContext
* Used twitter agent to retrieve tweets in avro format from flume to HDFS using the configuration file
* Created and imported data into hive database from csv file
* Created and input data into an Hbase database. Integrated Hbase with Hive to reflect automated updates from one table type to another.
* Setting up Hive and HBase databases on Cloudera Horton Works HDP platform and creating an integration

## EDUCATION

## BSc Computer Science

1981-1984 University of Reading, UK