Januka Pandey Gautam

Big Data Engineer

Januka.pandey456@gmail.com | 929-251-4811
https://www.linkedin.com/in/januka-pandey

Professional Experience:

- **Over 5+** years of IT experience as a Developer, Designer & quality Tester with cross platform integration experience using **Hadoop development and Admin.**
- Firsthand experience in installing, configuring, and using **Hadoop Ecosystem** - **HDFS**, **MapReduce**, **Pig**, **Hive**, **Oozie**, **Flume**, **HBase**, **Spark**, **Sqoop**, **Flume** and **Oozie**.
- Strong understanding of various **Hadoop services**, **MapReduce, and YARN** architecture.
- Responsible for writing **Map Reduce** programs.
- Experienced in importing-exporting data into **HDFS** using **SQOOP**.
- Experience loading data to **Hive partitions** and creating buckets in **Hive.**
- Developed Map Reduce jobs to automate transfer the data from **HBase**.
- Expertise in analysis using **PIG, HIVE** and **MapReduce.**
- Experience in **HDFS** data storage and support for running map-reduce jobs.
- Experience in big data technologies: **Hadoop HDFS**, **Map-reduce**, **Pig**, **Hive**, **Oozie**, **Sqoop**, **Zookeeper** and **NoSQL**.
- Responsible for the Provisioning, installing, configuring, monitoring, and maintaining **HDFS, Yarn, HBase, Flume, Sqoop, Oozie, Pig, Hive, Ranger, Falcon, Smart sense, Storm, Kafka**.
- Experience in **AWS CloudFront**, including creating and managing distributions to provide access to S3 bucket or **HTTP** server running on **EC2** instances.
- Experience in gathering and defining functional and user interface requirements for software applications.
- Experience in real time analytics with **Apache Spark** (**RDD**, **Data Frames** and **Streaming API**).
- Used **Spark Data** Frames API over Cloudera platform to perform analytics on **Hive** data.
- Experience in integrating Hadoop with Kafka. Expertise in uploading Click stream data from **Kafka** to **HDFS**.
- Expert in utilizing **Kafka** for messaging and publishing subscribe messaging system.

Core Competencies:

Languages	Python 2.7.x and Python 3.x, SQL, PL/SQL, Shell Scripting, Storm 1.0, JSP, Servlets, Scala, Python, Java, R, JavaScript
Big Data Technologies:	Hadoop, HDFS, Map Reduce, HBase, Apache Pig, Hive, Sqoop, Apache Impala, Oozie, Yarn, Apache Flume, Kafka, Zookeeper
Databases:	SQL, Spark SQL, My SQL, MS Access, HDFS, HBase, Oracle 12c/11g
Project Execution Methodologies:	Kimball data warehousing methodology, Agile Scrum Methodology
Regression:	Linear Regression, Ridge Regression, Polynomial Regression, Lasso Regression, Elastic Net
Clustering:	k-Means, Hierarchical Clustering, Latent Dirichlet Allocation (LDA)
Cloud Platform:	Amazon Web Services, Microsoft Azure
Version Control:	GIT, SVN, CVS

Education Details

Capella University
Bachelors in Information Technology

Career Experiences:

AT&T, New York, NY **Jun 2020 - Present**
Big Data Engineer

Responsibilities:

- Responsible for modeling complex Institute problems, discovering insights and identifying opportunities with statistical, algorithmic, mining and visualization techniques.

- Proficient at integrating and preparing large, varied datasets, designing specialized database and computing environments, and communicating results.
- Developed the new **Spark** jobs using **Scala**, to run on the **HDP** clusters, which provided significant gains on the completion times. Design development of **Spark SQL Scripts** based on Functional Specifications
- Responsible for **Spark Streaming** configuration based on type of Input Source
- Apache **Kafka Streaming API** for ingesting the data to **Spark Streams** and also publish to Kafka Topics for publishing the anomaly
- Importing and exporting data into **HDFS** and **HIVE**, **PIG** using **Spark** and **Arttunity** and **Sqoop**
- Used **Dynamic** partition for **Hive** when loading data. Implemented External and Hive managed Tables for significant performance gains.
- Involved in creating **Hive** Tables, loading with data and writing Hive queries which will invoke and run **Spark jobs** in the backend.
- Writing **Spark** (**Hadoop**) programs to convert text files into **AVRO** and loading into Hive (**Hadoop**) tables
- Implemented the workflows using **Apache Oozie Spark** library to automate tasks.
- Worked with **NoSQL** databases like **HBase**, **MongoDB** in creating **HBase** tables to load large sets of semi structured data coming from various sources.
- Used Hive and created **Hive tables** and involved in data loading and writing Hive **UDFs**.
- Responsible for spooling data from **DB2** sources to **HDFS** using **sqoop**.
- Created **HIVE** tables and provided analytical **queries** for business user analysis
- Extensive knowledge on **PIG** scripts using bags and tuples.
- Created tables in **HIVE** by partitioning and bucketing for granularity and optimization of **HIVEQL.**

Northern Trust, Chicago , IL **Apr 2017 – Mar 2020**
Big Data Engineer

Responsibilities:
- Installed and configured **Hadoop MapReduce**, **HDFS**, Developed multiple **MapReduce** jobs in java for data cleaning and processing.
- Importing and exporting data into **HDFS**, **Pig**, **Hive** and **HBase** using **Sqoop**.
- Managing and reviewing **Hadoop** log files.
- Worked on loading and transformation of large sets of structured, semi structured and unstructured data into **Hadoop** system.
- Responsible to manage data coming from different data sources.
- Developed simple and complex **MapReduce** programs for Data Analysis.
- Load data from various data sources into **HDFS** using **Flume.**
- Implemented Partitioning, Dynamic Partitions, Buckets in **HIVE**.
- Developed **Java** MapReduce programs for the analysis of sample log file stored in cluster.
- Involved in identifying job dependencies to design workflow for Oozie and resource management for **YARN**
- Capturing data from existing databases that provide **SQL** interfaces using **Sqoop**.
- Efficient in building **pig**, **hive** and map-reduce scripts.
- Cluster coordination services through **Zoo Keeper**.
- Involved in loading data from **UNIX** file system to **HDFS**.
- Installed and configured **Pig**, **Hive** and also written **Pig** and Hive **UDFs.**
- Automated all the jobs, for pulling data from **FTP** server to load data into Hive tables, using Oozie workflows.
- Involved in creating **Hive** tables, loading with data and writing **hive queries** which will run internally in **map** way.
- Exported analyzed data to relational databases using **Sqoop** for visualization to generate reports for the **BI** team.

References Available Upon Request