Piyush Patel

9 Stonecrop Lane **Cell:**+1-(647) 533 - 4349

Etobicoke, ON, M9V 4J6 **E-mail:** piyushpatel.ee@hotmail.com

Java/Scala/Hadoop Engineer with over 11 years’ industry experience with various cutting-edge technologies like Spark, Kafka, Hadoop, Hive, HBase, Flume, Spring Boot, Angular, Scala, Java, etc.

**TECHNICAL SKILLS**

*Big Data, Hadoop, MapReduce Apache Spark, Kafka, Flume*

*Scala, Java, Python, Javascript, R Sqoop, Oozie, Hue, Streamsets*

*SQL, MySQL, Hive HBase, MongoDB, BigQuery*

*Spring Boot, Hibernate, JSP NodeJs, ExpressJs, jQuery*

**PROFILE SUMMARY**

* Proven experience with Spark streaming with Apache Kafka
* Experience building RESTful services using Spring, Spring Boot
* Full stack web development experience in Java and NodeJS backends
* Experience with Cloudera platform as well as Hortonworks platform
* Highly skilled to perform under pressure in Agile environment
* Experience with Git, CI/CD, Jenkins, Confluence and JIRA
* Development experience with major cloud service providers (AWS, MS Azure, GCP)
* Worked on different cloud products EMR, S3, EC2, Lambda, BigQuery, Cloud Storage, etc.
* Skilled in building data pipelines in NiFi and Streamsets
* Very good understanding of OOP as well as functional programming paradigm
* Optimization of various Spark jobs for ingesting from files or Kafka topic.
* Experience with Java, Spring, JavaScript, Python, Scala, SQL for web development projects
* Experience with reliable and fault-tolerant messaging system, Kafka
* Proficient with various build tools (SBT, Maven) and IDEs
* Highly adaptable to new technologies and always eager to learn

**EXPERIENCE**

**Bell Mobility, Canada (Contract)** Nov 2018 – Present

(Specialist, Big Data Hadoop)

* Developed highly efficient Stream processing pipeline with Apache Kafka and Spark Streaming
* Optimized streaming job to ingest nearly 10 million messages per minute.
* Developed pipeline to ingest data from Google Bigquery by utilizing Google cloud storage using Java and Scala programming on GCP
* Developed shell scripts to ingest CSV data
* Developed complete end-to-end pipeline to ingest inappropriate XML files with nearly 8TB raw files everyday.
* Optimized yarn resources to develop highly efficient Spark jobs
* Guided and mentored new developers for developing ingestion pipelines.
* Developed Oozie workflows and coordinators to parameterize properties
* Created CI/CD pipelines to deploy automatic pipelines using YAML files.
* Debugged production issues and rectified workflows.
* Created Parquet tables from one cluster data to new production cluster using Hadoop Distcp
* Develop shell scripts to automate the data ingestion process into Hive warehouse
* Worked with Apache Streamsets and developed custom processor in Jython for ingesting data from Cloud platform.
* Managed small file problem on Namenode by creating pipelines with large file sizes.
* Worked on ML related projects which involved deploying using Docker and Kubernetes with Helm charts within internal private registry
* Reviewed code written by peer developers
* Used Maven and Jenkins to automate pipeline deployment
* Designed, developed and architected data pipeline to migrate existing Informatica workflow to Hadoop infrastructure.

**Tech Mahindra at Rogers Communications** Jun 2018 – Nov 2018

(Sr. Software Developer)

* Developed Hive scripts for data warehousing and data analysis
* Developed REST services using Spring application framework for internal CMS system
* Wrote test cases for Junit and Mockito to validate new features
* Developed shell scripts to automate the data ingestion process into Hive warehouse
* Worked with Apache NiFi to develop automated data flow pipelines to ingest data into Hadoop ecosystem
* Worked on Spark ingestion jobs for maintaining projects
* Fixed the features for data ingestion as required by the client.
* Developed Software solutions for data ingestion along with data modelling for efficient data retrieval by analytics team.
* Developed tables for reporting using Hive functionalities like RANK, ROWNUM over certain partitions to create less granular data.
* Worked on software solutions using Scala, Spark, NiFi and various enterprise softwares.
* Manage and verify automation scripts using Control M
* Optimized Tez queries by setting various configuration parameters like vectorization and shuffling.

**Zone TV, Canada** Nov 2017–Jun 2018

(Java/Scala Developer)

* Developedtelevision application REST API using Scala/Play framework
* Worked on backend development with Play framework using Slick ORM library.
* Attended Daily Scrum meetings to understand the project requirements and follow up with quality code
* Working on Angular 4 UI bug fixes and product development
* Worked on POC on Amazon EMR cluster to deploy and test Hadoop job.
* Fixed bugs for their front-end UI to optimize Angular code.
* Created and enhanced few REST API end points written in Java Spring framework.
* Implemented new end points to fetch video contents and enhance search facility using Spring Boot 1.4, Spring Data with Maven and Jenkins.
* Extended identity verification and multi-tenant solution in NodeJS REST API.
* Explored Ooyala documentation, validated and managed interactions with Vendor representatives.
* Managed authentication API written in Python Django framework.

**Meditab Software Inc** Oct 2015 –Sep 2017

(Senior Hadoop/Spark Engineer)

* Developed Data Ingestion jobs using Scala as a programming language with RDDs
* Developed RESTful webservices for online gift shop and IMS systems.
* Developed 3-tier architecture applications with Spring and JPA
* Used RDDs/ DataFrames to transform and store the data into HDFS
* Tested program execution flow in interactive spark shell
* Developed Oozie workflow or coordinator jobs for automation
* Loaded data from various sources like Teradata using Sqoop/Kafka into HDFS
* Performed analysis using Hive queries and provided reports to Data Science team
* Participated in design work flow of Big Data extraction process
* Prepared deployment instructions for testing and deployment team
* Test the complete execution of the process on development and pre-production cluster
* Wrote complex Hive queries to retrieve data using staging tables.
* Worked on few projects involving web development with Spring framework and Hibernate
* Developed few Pig scripts to facilitate transformation and ingestion of data for clients.

**Envale Systems Pvt Ltd.** Jan 2015 – Oct 2015

(Big Data Engineer)

* Sqoop data from RDBMS like Oracle 12c server to HDFS system
* Transformed the data based on the client requirements
* Extracted and loaded streaming data using Apache Flume and make them available for analysis
* Partiticipated in design of workflows for Big Data ETL projects
* Prepared documents with instructions for cluster deployment for the admin team
* Tested the product on pre-production cluster with desired outputs
* Wrote Hive and Pig scripts to extract, transform and load data into/from Hadoop file system
* Wrote MapReduce programs using Java programming

**GeekMindz Solutions** Mar 2012 – Jan 2015

(Software Developer)

* Developed backend software solutions using Java, Spring, Hibernate and JavaScript solutions
* Participated in architecture discussions and provided helpful insights into architecture decisions.
* Developed full stack applications including UI as well as backend using JSP, Thymeleaf, Spring MVC, JPA, HIbernate
* Worked on NodeJS framework (Express) for developing REST APIs with clean code
* Participated in code reviews of the software solution with team leaders and architects
* Developed schemas for MongoDB database using Mongoose
* Learned new cutting-edge technologies with best practices
* Attended Scrum planning sessions and review meetings along with daily standup meetings
* Validated user inputs through interactive messages and animations using AngularJS front-end client.
* Worked on Grunt scripts to automate build process of the project.
* Experience with Maven for deployment on Tomcat server.

**EITL** Oct 2009 – Mar 2012

(Java Programmer)

* Analyzed and prepared total system architecture for different web projects.
* Developed applications using Servlets, JSP and Spring framework
* Wrote JSP web pages for dynamic UI pages
* Developed web applications using Spring, Hibernate, Tomcat and Glassfish servers.
* Worked on various projects with Java web frameworks like Struts and Servlets
* Helped analyze functional and non-functional requirements.
* Coordinated with customer to understand their requirements.

**EDUCATION**

**Bachelor of Electrical Engineering, India** (**GPA:** 3.8 out of 4.0)

**ITC – Diploma in Java Development**

**Coursera – Data Engineering on Google Cloud Platform**

**Lightbend – Akka Framework courses (Akka, Akka Streams)**