



# Pranav Malik

Software Engineer (II), Nagarro COE division

## Experience summary

Pranav has 25 months of experience in the software industry. He has experience in providing Big Data and Web Based solutions using technologies such as Java -Spring, Spark, Python-Django, Scala, etc. He has skills in working on various domains such as Big Data, analytics, Machine Learning, Web Development via Java Spring and Angular/ReactJS . Currently working in SAAS

## Contact Details:

Email :pranav2134malik@gmail.com

Phone: 9999318323

## Trainings and certifications

- Java and Data structures, Python and Data Structures Coding Ninjas
- Apache Spark 3 with Scala, Udemy
- Apache Spark Streaming ,Big Data Udemy
- Taming Big Data with Spark Streaming and Scala, Udemy
- Data Analyst Nanodegree, Udacity(Cleared Level 1,2,3)
- Node JS,Angular,Vue.Js Professional course, Udemy
- Taming Big Data with Spark Streaming ,Udemy
- Scala for Big Data ,Udemy
- AWS certified cloud solutions Engineer - Associate

## Skill set

<b>Programming and scripting languages</b>	Python, Scala, R, Java, Pixi.js, JavaScript, Angular, React Native, UnrealScript, C++, TypeScript, OpenGL, Vue.js, Vanilla.js, Sqoop, PySpark, Java Hybris,Java Spring, Spring Boot, Spring Cloud,Hibernate,Python Django.
<b>Frameworks, tools, and libraries</b>	Splunk, Hadoop, Spark, Storm, Cassandra, RapidMiner, Unity 3D, Unreal Game Engine, Phasor, Neural Networks, OpenCV, TensorFlow, NumPy, Keras, Panda, Hive, MapReduce, Amazon AWS-EC2,Lambda,Athena,RedShift,S3,Glue,Kinesis,Quicksight,ColdStorage.

<b>Servers and platforms</b>	REST Services, JSON, XML, e-commerce server backend using SQL Server, StreamSets, Kubernetes-Docker.
<b>Databases and BI</b>	SQLite, MongoDB, Cassandra, MongoDB, Neo4j Graph Data Bases
<b>Devices and OS</b>	Windows OS, Oculus Rift, IoT, Arduino, Raspberry Pi, Linux

## Education

- Bachelor of Technology, Computer Science, University School of Information and Communication Technology, Guru Gobind Singh Indraprastha University, Delhi, India (Grade- 8.32/10) [2015-2019] {AIR -1109 in IPU-CET}, {AIR-13,481 in JEE Mains} [2015]

## Work Experience

### 1. Price Master Product (Indus Valley Partners-Noida), (Role- Software Engineer- I)

<b>Company Description</b>	A product based company with a client base consisting of the most successful hedge funds and private equity firms in the world, specializing in Fintech Products and Trading.
<b>Project overview</b>	<p>Project involved developing an OCR using deep learning algorithms and natural language processing to analyze client invoices and adding a submodule in price-master product using Django-Python framework. It involved:</p> <ul style="list-style-type: none"> <li>– Optimizing the existing search pattern token files of price master using state of the art solution in React and made suitable API for its deployment</li> <li>– Developing IVP transport application, providing additional functionalities for the existing transport application, and receiving appreciation from the Manager for this project</li> <li>–</li> </ul> <p>The project involved developing recommendation system with long term and short-term profile on Hadoop and OLAP DB for recommendation system. It involved forming multi node Hadoop cluster for handling and storing huge amount of data for data analytics. The task required integrating Sqoop to transfer terabytes of data from Vertica to maintain historical data delivering 300 odd reports in worlds different calendar in a Critical Dataware housing project. The project also required the following:</p> <ul style="list-style-type: none"> <li>– Performance tuning of the jobs (ETL) and DB Scripts with many scenarios such as manual projection design, query profiling, DBD etc.</li> <li>– Developing complex framework which handled inter-related query, waiting for each other processes completion (interdependent query mechanism)</li> </ul>

	<ul style="list-style-type: none"> <li>– Integrated Sqoop to transfer data from source</li> <li>– Developed complex merge process to merge output from different system</li> <li>– Profiling and creating mapping document</li> <li>– Developing Hive script to generate internal and external table at runtime</li> </ul>
<b>Technologies</b>	Python, Dailogflow, Neural Networks, Django, MongoDB, Angular 6, Java, Scala, R, SPSS, Data Miner, Git, Graph Databases, Cassandra
<b>Responsibilities</b>	<ul style="list-style-type: none"> <li>• Involved in analysis of required functionality, task estimation, and allocation to the team</li> <li>• Designed application architecture and database</li> <li>• Worked on implementation of key modules and UI creation</li> </ul>

## 2.Nagarro,COE division(Role- Software Engineer- II(Big Data and Backend))

<b>Project 1.</b>	<b>Client Confidential is the Largest Customs Service Provider in Europe, through a wide range of services, Client optimizes your trade and customs management – helping improve trade performance, operational efficiency, risk management and customs duty control.</b>
<b>Project overview</b>	<p>Project involves robust mailing system capable of handling Millions of mails per client per year with extensive search capabilities on raw text as well as attachments based on multiple search criteria and logical operations on search criteria.</p> <p>Applied mail chain deduplication for discarding duplicate mails, attachments and extracting unique and useful data.</p> <p>implemented analyzers, tokenizers customized query processing using Elasticsearch, Enabled fast searching.</p> <p>Implemented saga pattern for distributed transactions on Elasticsearch and Mongo DB.</p> <p>used OData API along with rest API.</p> <p>A java, Spring boot based Web App implementing design principal and patterns for a Flexible and extensible notification management System and Employee online Portal for employees to send types of Notifications (SMS, Email, portal Notification).</p> <p>Developed Multiple POC on Big Data as a service within different modules to improve performance, job scheduling , Automatic Data Extraction</p>
<b>Technologies</b>	Java, Spring Boot, Mongo DB, Elasticsearch, Spark
<b>Responsibilities</b>	<ul style="list-style-type: none"> <li>• Involved in analysis of required functionality, task estimation, and allocation to the team</li> <li>• Designed application architecture and database</li> </ul>

	<ul style="list-style-type: none"> <li>• Worked on implementation of key modules and UI creation</li> </ul>
--	---

## - Big Data Engineering COE(Nagarro)

<b>Project 2: Team Description</b>	<i>Nagarro COE is the Product based Division [Internal-Subsidiary] of Nagarro Services Pvt Ltd.</i>
<b>Projects</b>	<p><b>1. Client Confidential is 2nd Largest Network Provider in Middle East.</b></p> <ul style="list-style-type: none"> <li>– Anomaly Detection in Telecommunication Systems, using Markov Chain+Ann model , Working in Geospatial Data using OpenLayer,Hadoop and D3.js, Neo4j Graph Databases)</li> <li>– Elastic Search feature was used to ingest data from producers connected via Kafka Message Queue.</li> <li>– Hadoop was used in Data Lake feature of the pipeline where a data generator was scheduled via Airflow to dump data at a Particular Time Interval.</li> <li>– Data Dumping onto open layer was done by PostgreSQL and Python Script</li> <li>– Job Scheduling was done via Airflow</li> <li>– Optimized Anomaly Detection Process by adding Forecasting power using Supervised Learning using LSTM-Autoencoders</li> <li>– Modified the WMS Open layer to WFS which enable querying, non-static data loading, Auto-Refresh Feature, Dynamic Data Loading</li> <li>– Data Modeling was done in D3.js.</li> <li>– Visualization was done Via Kibana Charts using Dynamain Data uploaded by Hadoop file system.</li> <li>– Data Pipelines were made via Stream sets.</li> <li>– Used Shading-Load Balancing Concept for better utilization of Clusters for memory and resource allocation.</li> </ul> <p><b>2. Client Confidential is the Largest American multi-level marketing company that sells health, beauty, and home care products. (SAAS Product for different regions in world)</b></p> <p><b>2.1 Work carried out in India Data lake, Tech Stack (Python, Golang,AWS,Terraforms,Spark,Django Framework)</b></p> <ul style="list-style-type: none"> <li>- Was responsible for both the lambda functions written in Golang and Python</li> <li>- Developed New end to end Data pipelines coupled with Microservices team to provide data in a Streaming fashion.</li> <li>-Wrote Integration Scripts with 3rd Party platforms like RP,KPMG in both Python and Golang.</li> </ul>

- Responsible for Writing Terraform Scripts to initialize AWS resources with custom policies.
- Built a recommender system POC to provide recommendations to n number of users in choice of Products to buy.
- Wrote Utility functions in Python for- Multithreading (via Semaphores), Flattening of NO-SQL data , Shading of data along custom sets.
- Wrote Utility Function in Spark and a Subsequent glue Job for removing duplicate records
- Wrote Spark Scripts to Efficiently apply Business Logic on approx. 1 million records, Developed Checkpointing and Fault tolerance evaluator functions.
- Optimized the I/O operations by using Mutual Exclusion and Semaphores in Python
- Developed Custom Encryption/Decryption API on AES algorithm for implementing Business Logic in Golang.
- Optimized the Ram Management Function for Local Debugging, bringing chunks (Actual Data) of data and Storing it on Dequeue acting as A Cache with LRU policy.
- Athena Query Development for Business Logic Implementation and Reconciliation Activity.
- POC development for Data Analysis and providing Indexing to NoSQL data stored in Data lake.
- Deployment Related Terraform Script Development and respective env set up
- Glue Job Development for Updating Same Order Id grouped Orders in Data lake.

**2.2 Work carried out in Russia Data lake, Tech Stack (Python, Golang,AWS,Terraforms,Spark)-**

- Resolved Tickets related to enhancements in Lambda Functions in Python and Golang
- Provided Support for Knowledge Transition Purpose
- Involved in Code Review ,Debugging Sessions with Russia Team.
- Provided Sessions Related to Knowledge Transfer , Business Logic, Implementation to Russia Data lake Team.
- Was involved in Python Microservices for SMS engine, Address Resolution , Implementation and Local Debugging ,Testing.

	<p>-Involved In Implementation and Debugging of MLOS Reconciliation Activity.</p> <p><b>2.3 Work carried out in SEA Data lake, Tech Stack (Python, Golang,AWS,Terraforms,Spark)-</b></p> <ul style="list-style-type: none"> <li>- Involved in Vietnam, Malaysia, Philippines region for Quick sight Report Development, Learnt Quick sight in 2 weeks.</li> <li>- Responsible for 150+ reports, resolved more than 30+ incidents in all the three markets.</li> <li>- Responsible for End to End story implementation, Requirement Gathering from client, First Point of Contact.</li> <li>- Cleared Certification as AWS-Quick Sight Professional.</li> <li>- Responsible for End to End Order Payload Optimization ,Flattening S3 Stores.</li> <li>- Responsible for Presentation , Reporting to Clients for future requirements, Analysis.</li> <li>- Closely working with Hybris Teams ,POS teams, Suggesting improvements ,Code Reviews.</li> </ul> <p><b>2.4Work Carried out in BIR Reports SEA-POS TEAM (TechStack- Python,Django,AWS)</b></p> <ul style="list-style-type: none"> <li>- Responsible for BIR, Python reporting Microservices module.</li> <li>- Responsible for Requirement Gathering, Analysis Phase.</li> <li>- Responsible for End to End Development for Microservices module.</li> </ul>
<b>Technologies</b>	PySpark,StreamSets,Kafka,Cassandra,Hadoop , Machine Learning ,ELK stack, Airflow
<b>Responsibilities</b>	<ul style="list-style-type: none"> <li>• Involved in analysis of required functionality, task estimation, and allocation to the team</li> <li>• Designed application architecture and database</li> <li>• Worked on implementation of key modules and UI creation</li> <li>• Individually Handled the Machine Learning Part and Hadoop Cluster Set in StreamSets</li> </ul>

## Personal Projects

### Smart India I4C-NCL Pune (Deloitte)- Smart India Hackathon 3<sup>rd</sup> Prize

<b>Client</b>	<i>NCL PUNE is a Research and Development and Consulting organization with a focus on chemistry and Chemical Engineering . It was an Joint Venture of Deloitte and SIH i4c-NCL Pune</i>
<b>Project overview</b>	<p>Project- Risk assessment of fire hazards in coal mines using data analysis and machine learning</p> <ul style="list-style-type: none"> <li>– A watch was made as a final project with the help of National Chemical Laboratory Pune(NCL-Pune) which is fitted with the technology to detect early fires before its occurrence A special care has been taken with respect to the issue of false alarms</li> <li>– A recommendation system written in Cassandra and Java, with front-end being in Angular 8, with basic functionality of elastic search</li> <li>– Exposed Machine Learning Model via a Flask Application</li> </ul>
<b>Technologies</b>	Python ,IOT,Aduino ,Django Framework, Cassandra,Java,Angular 8,Flask
<b>Responsibilities</b>	<ul style="list-style-type: none"> <li>• Involved in analysis of required functionality, task estimation, and allocation to the team</li> <li>• Designed application architecture and database</li> <li>• Worked on implementation of key modules and UI creation</li> <li>• An award of INR 1,00,000 was awarded.</li> </ul>