

Niraj Kumar

Data Engineer

Bangalore India 7899272929 nirajkmr007@gmail.com



<https://www.linkedin.com/in/niraj-kumar-0530324a/>



<https://github.com/nirajkmr007>



<https://stackoverflow.com/users/8581421/niraj-kumar>



<http://www.eworldcomputer.com/niraj/>

Summary

Innovative and versatile **Data Engineer** offering nearly **5 years** of comprehensive experience as a Software professional with **3+ years of relevant experience as a Data Engineer** known for optimism, high energy, self-organised and detailed approach leading to award winning performances

Work experience

Brillio (Senior Engineer)

2017-Dec - Present

Working as a Data Engineer and BigData Developer with components like Spark, Python, AWS(Athena, EMR, S3 etc)

Marlabs (Software Engineer)

2014-June - 2017-Nov

Worked as a Trainee Engineer and Software Engineer with components like, AWS, HADOOP, Data Warehouseing IOT, etc

Project Summary

Current Project : Similar-Homes Recommendation Engine

Application: Realtor.com

Feb 2017 to Present

Role : Developer

Team size : 9

Domain : Real State

Realtor.com is a real estate listings website operated by the News Corporation subsidiary Move, Inc. and based in Santa Clara, California. The company's business model focuses on selling advertising and leads to agents and brokers

Responsibilities:

Developed algorithm and spark application for similar-home recommendation

Helped in designing highly scalable, end to end pipeline with 80 node cluster

Dealt with very high volume and variety of data

Optimized code to run in optimal way to save cost and resources.

Written data transformation script for relational to complex nested json.

Technology used : Pyspark, Python, Jenkins, github, cloudformation, data-pipeline, EMR, S3, dynamoDB.

Project : 5 - Data Migration Framework :

Dec 2017 to Feb 2018

Application: Realtor.com

Domain : Real State

Role : Developer

Team size : 5

Responsibilities:

I was responsible to Engineer and design of data migration framework based on pyspark and AWS services.

which can easily and efficiently move huge amount of data from sql server and upload to s3 datalake in incremental way.

Framework also maps and transform data as per business logic.

Skills

Spark, Spark SQL

Hadoop, HDFS, Map Reduce

AWS, EMR, Athena, S3 etc

Python, Core Java, SQL

Linux

POCs

Chat Bot : using microsoft bot framework, skype and luis.

Smart parking : using Raspberry PI, ESP08, Ultrasonic Sensor

Education

B.TECH (CSE) 72% 2009-13

Haldia Institute Of Technology
West Bengal

Personal Details

DOB : 20-Dec-1990

Nationality : Indian

Gender : Male

Awards / Wins



Techgig Coding Competition.



Robotics Competitions



Chess Competitions.

Interests

Cricket
Dance
Social work
Adventure
Machine Learning
Teaching
Travel
Chess
IOT

Project : 4

June 2017 to Dec 2017

Application : M-Advisor

Domain : Data Analysis

Role : Developer

Team size : 16

Marlabs' M-Advisor cuts down time-to-insight to under a month. Being an automated analytics and AI platform, it brings speed to the process of turning enterprise data into insight.

Link - <https://www.marlabs.com/madvisor/>

Responsibilities:

Worked in a data-mashup and some other modules.

Written python code to dynamically mashup data frames based on user action.

Improved the performance of join on very large dataset.

Created multiple function to manipulate data coming from various sources

Project : 3

Client : Gate group

Domain : Airline Industry

Duration : Feb 2016 to May 2017

Role : Hadoop Developer

Team size : 9

Egate Solutions is a leading global provider of products, services and solutions related to a passenger's onboard experience. The project is to maintain the base of more than 25 airlines and the market only end-to-end passenger galley planning, Operation management, Supply chain planning and Forecasting.

Responsibilities:

Involved in exploring Hadoop MapReduce programming and its ecosystem.

Imported data into HDFS from databases using Sqoop and loaded data into HDFS from flat files.

Implemented Hive tables and HQL Queries on transactional data to gather specific data for generating the reports.

Analyzed log data generated per day using hive.

Enabled speedy reviews and first mover advantages by using Oozie to automate data loading into the Hadoop

Distributed File System.

Project : 2

Client : Laureate Inc

Domain : Education Industry

Duration : Sep 2014 to Jan 2016

Role : Development & Maintenance

Team size : 5

Technology used : Talend 5.4, PostgreSQL, Amazon Web Services

Laureate is a network of more than 70 campus-based and online universities in 25 countries offers undergraduate and graduate degree programs to over one million students worldwide. We have done analysis on Client's data as per requirement and provided the results and reports back to client.

Responsibilities:

Developed ETL jobs and Packages, Created Metadata and Repository.

Created, Monitored and troubleshoot the jobs for the process through TAC.

Involved in transforming the data from staging to target tables.

Taking Care of AWS Infrastructure like health of EC2 instances and RDS , starting and stopping instances, Managing files on S3,Creating AWS alerts and notification.

Project : 1

Client : MarLabs (Internal)

Domain : Parking

Duration : Oct 2016 to Dec 2016

Role : Development and Research

Team size : 3

Technology used : IOT, Rest, Python, Lua,

Devices used : Raspberry PI, ESP08, Ultrasonic Sensor(HC-SR04), Aurdino

Smart Parking System was an internal project for Marlabs. Objective of this project was to implement IOT concepts to make available status of parking space over internet which can be accessible through browser as well as mobile, anytime from anywhere.

Responsibilities:

Done POCs and Research on IOT basic components.Created python script to read digital value from sensor.

Exposed APIs for application to get data.Designed the circuit for device.Burned the code to device and tested it for daily use.