VIPUL RAI

vipulrai8891@gmail.com
Phone: +91-8892598819
https://in.linkedin.com/in/vipulrai

Engineering professional with substantial experience designing and executing solutions for complex business problems involving large scale data analysis, real-time analytics. Knows how to use the right tools when and where they make sense enabling organizations effectively analyze and process TB/PB of structured and unstructured data.
6+ years of experience working on Big Data Ecosystem.
DataBricks Certified Spark Developer (August-2015 , Certification No 1.1.0-0170)

SKILLS

- <u>Frameworks</u> – Hadoop , Spark
- <u>Big Data Ecosystem Tools :</u> PySpark , Scala- Spark , Data Frames API , Spark SQL, Spark Streaming , Kafka , Map Reduce , Hive , HDFS ,AWS , S3, Ubuntu , Centos, Solr, YARN, SQL
- <u>ML:</u> Classification Models , Regression Models , DL Basics, LSTM

Programming Languages	Python , Scala
IDE's / Build Tools/Version Control	Jupyter , Intellij Idea ,SBT , Maven , Git , GitHub , JIRA

EXPERIENCE

Smartnomad - Product Developer   October 2017 - Present

Objective	To build a product which enables users to book personalized travel itinerary based on their preferences.
Client	Product
Platform	Python , AWS Lambda, Mysql , EC2 ,S3
Process	- Built the entire pipeline beginning from data aggregation to the final itinerary. - Wrote proprietary algorithm which calculates the scores for each poi - Wrote code which adds functionality on top of original solver - Built the backend of the core application which provides personalized intelligent itinerary custom made for each user. - Built automated framework which tests the effectiveness, feasibility of the itinerary.
Role played	Design and development of the product from Scratch
Duration	2 years - ongoing

Independent Big Data Consultant   May 2017 - October 2017
Affine Analytics – Consultant   November 2015 - May 2017

Travel Portal - Optimization

Objective	To reduce the number of unqualified searches on portal's meta channel
Client	Fortune 100 Travel Portal
Platform	Pyspark , Jupyter
Process	▪ Gave an optimized solution for a leading online travel brand to reduce the number of unqualified searches it is getting from meta channels, to improve the business KPIs like ROI/profitability/efficiency. ▪ Wrote the Optimization Program using lpsolve from Pulp package (Pyspark) and Random Forest from MlLib.
Role played	Design and development of the solution
Duration	3 months

Retail Store – Revenue Optimization

Objective	To identify optimal promo/clearance sale prices which would result in maximum sales.
Client	American chain of departmental stores
Platform	Spark- Scala , Hadoop , HDFS , AWS ,EC2
Process	▪ The pricing team was tasked with the challenge to identify optimal promo/clearance sale prices which would result in maximum sales. ▪ Identified multiple factors which impact unit sales during promotions such as Price Elasticity, Seasonality, % sales in each Day-of-the-Week, effect of circulars/promotions, breakage due to unavailability of different sized SKUs, impact of public holidays, etc.
Role played	Designed the Architecture, Developer
Duration	3 Months

Travel Portal - Email Marketing and Campaigns

Objective	To process huge unstructured data and help the marketing team to gain insights about the current promotions and future strategy.
Client	Fortune 100 Travel Portal
Platform	Pyspark, Scala-Spark , Hive, AWS , S3 , HDFS, Qubole
Process	▪ Data was being generated at about 2GB per hour in json format. ▪ Converted data to parquet format for faster retrieval and less storage. ▪ Used client API to map the data and convert into standardized format. ▪ Used Qubole for accessing the spark cluster and running the queries. ▪ Wrote program in Pyspark and Scala Spark to select different campaigns for various categories of customers based on Business Rules ▪ Hands on experience working on AWS and S3
Role played	Solution designer and developer
Duration	9 months

Gaming Analytics Sciences

Objective	To process huge game logs to enable publisher find churn analysis, make future marketing strategies, find sentiments of players about the current version of game
Client	Top 3 Game Publishing Company
Platform	Pyspark, Scala-Spark , Hive, AWS , S3 , HDFS, Qubole , JIRA
Process	- Raw data was being generated in json / feeds on Hive - Converted data to parquet format for faster retrieval and less storage. - Creating automated QA module on Spark, which did basic QC of data such as calculating min, max , avg etc for each day's fresh feed thereby removing the necessity of manual intervention - Wrote an automated Pyspark program which calculated user stats on daily, weekly and monthly level - Created input data set for cross sell model using stats such as available PS Plus points, weekly spend on PS points - Created model to detect fraud and cheat during gameplay
Role played	Solution designer and developer
Duration	6 months

AIG Data Solutions - Big Data Developer   February 2015 - November 2015

Insurance Log Analyzer

Objective	To parse data incoming from 8 different types of servers to be consumed by Power BI
Client	Fortune 100 Insurance Company
Platform	Scala-Spark, Tableau, Kafka
Process	- Logs from 8 different types of devices were being generated simultaneously. - Used Kafka and Pyspark to generate insights such as - which user was accessing the websites which were blocked etc. - The final result was being used by Power BI for reporting
Role played	Designed the Architecture, Developer
Duration	6 months

Blue Star Infotech – Software Engineer   August 2013 – February 2015

ACADEMICS

Qualification	Stream	Year	Grade	Board
B. Tech	Computer Science and Engineering	2008-2012	67.31%	JNTU-K
12th	MPC-IP	2008	66.80%	CBSE
10th	Science	2006	79.80%	CBSE

CERTIFICATION / INDEPENDENT COURSEWORK

➢ Databricks O'Reilly Developer Certification for Apache Spark
➢ R Programming - Coursera
➢ edX Honor Code Certificate for Introduction to Linux
➢ Big Data - Spark Fundamentals - IBM
➢ Hadoop Fundamentals I - Version 3 (bigdatauniversity.com)
➢ Accessing Hadoop Data Using Hive V2 (bigdatauniversity.com)
➢ Moving Data into Hadoop (bigdatauniversity.com)
➢ Hadoop Developer Day (bigdatauniversity.com)
➢ Hadoop and the Amazon Cloud (bigdatauniversity.com)

AWARDS AND RECOGNITION

❏ Affine - Transformer of the month thrice and transformer of the quarter once
❏ AIG - Employee of the month for Spark POC
❏ Blue Star Infotech - Twice - Employee of the month

PERSONAL DETAILS

✓ Nationality: Indian
✓ Languages Known: English, Hindi
✓ Local Address : Koramangala , 8th Block , Bengaluru - 560095