KAMBALA YASWANTH
kambaalayashwanth@outlook.com
Mobile:- +91-7396873676
LinkedIn :- https://www.linkedin.com/in/yashwanth-kambala-2a2b5626/
http://yaswanth.surge.sh | https://github.com/yashwanth2804

APACHE SPARK DEVELOPER

Professional Summary:

- **Over** 2+ years of Professional experience in IT Industry in Developing, Implementing, configuring, Java, **J2EE**, **Big Data Technologies, Django.** 16 months of experience in the design and development of Big Data Analytics using Hadoop Ecosystem tools, Spark**-RDD, DataFrames, Datasets.**
- Excellent understanding/knowledge of Hadoop architecture and various components such as **HDFS, YARN**
- Hands on with Dataframes. Datasets and SQL Queries in SPARK SQL.
- Worked on Spark cluster and **standalone modes, YARN.**
- Worked on NoSQL database, MongoDB and spark job-server.
- Working experience on **SQL,** and **CRUD** operations.
- Experience with **CSV**, **JSON**, **Parquet** formats.
- Contributed towards creating web-based applications using **Java and J2EE** technologies.
- Have fair knowledge on multi-threading concepts, client-server and micro service architecture.
- Expertise in Java for Enterprise business applications using Java and J2EE technologies like **Servlets, JSP, JavaFX.**
- Experience in handling different file formats like parquet, JSON, CSV, flat files.
- Strong knowledge in NoSQL database, MongoDB and firebase.
- **Alteryx Certified Professional** from Alteryx partner program.
- Worked as a full stack web developer using Django framework.
- Published blogs on spark and Django in company's internal learning resources.

Technical Skills

Programming Languages	C, Java, python
Big Data Technologies	Spark-Core, Spark-Sql , Alteryx , HDFS, YARN
Databases	MySQL, MongoDB, Firebase
Web Technologies	JSP, Servlets, Django, CSS, HTML5, JavaScript, jQuery
Operating Systems	Windows, UNIX, Linux distributions (Ubuntu)
Other Tools	Eclipse, GitHub, Maven, NetBeans, Docker, Zeppelin, databricks notebook,Slack

EXPERIENCE:

EXAFLUENCE PRIVATE LIMITED

June, 2016 – Aug,2018

EXAFLUNCE is technology Solution Company focused on solving complex problems leveraging Big Data and Analytics utilizing new technologies and frameworks. Our accelerators and solutions use advanced technologies in IOT, Big Data and Machine learning to provide actionable insights. We develop expedited solutions built on micro services architecture using industry standard models and KPI repositories for Life sciences, financial services and Health care verticals

Project Name: ExaHealth
Role: Spark Developer, Apr 18 – Aug 18

Description:

NpiAlerts, is the module dedicated to track the changes of NPIs records from CMS.gov (*Centers for Medicare & Medicaid Services*) database. This project provides the updated NPIs records for every hour, whereas CMS will provide weekly in weekly updates section. This project also provides history of changes that a particular NPI record had gone through.

Role & Responsibilities:

- Used Java FutureTask to call the CMS API endpoint to get NPI records in parallel, hence reduced the download time of 120 million records (240 GB) from days to 53 minutes.
- Spark used to load the individual downloaded JSON files to a dataset and applied business logic on it.
- Used HDFS for storage and YARN for the cluster manager.
- Able to setup multi-node Hadoop cluster with 3 (t.xlarge) AWS instances.
- Applied UDFs (User defined Function) to format on particular columns to meet the required logic.
- Used spark *persist* method on dataset to perform fast analytics in multiple iterations.
- Provided schema externally for 350 columns instead of *inferSchema* option on JSON, optimized performance to 2x.
- Monitored Spark-UI for any straggler task and slated keys in-case while performing join operations
- Made an application in two modes, on demand and micro-batch mode. On-Demand mode lets thespark job to execute by sending a request from web-app using Spark-Job-Server. The micro-batch mode will execute in continuous fashion
- Using mongodb as the database, stored the updated set of NPI records for every iteration
- UI/UX accomplished with JSP, CSS as frontend and Java as a backend.

Environment: Java SE 8, Spark-core, Spark-SQL, JSP, Java, MongoDB, Spark-Job-Server, Maven

Project Name: Authorization of Pended Claims
Role: Spark Developer, Oct 17 – Mar 18

Project: This project provides the details of Pended claims details in Health care in USA, for the particular claim status, shows the entire insurance amount and due amount.

Responsibilities:

- Used Spark 2.2 with Java 8 to get data from different Oracle DB sources.
- Using *OracleJDBC* driver connected the spark application to the oracle database url
- Implemented *Mappartion* over Map while writing data to the database.
- Created UDF (User Defined Function) for converting the zip code, date-format to required
- Used complex Spark-SQL JOINS queries on dataset to generate the report.
- Fine-tuned the partition level to avoid uneven partitions.
- Preferred spark-sql built-in function to UDF because of performance optimization
- Created Schema externally for the dataset using *StructType*.

Environment: Java SE 8, Python, Spark-core, Spark-SQL, MongoDB, Maven

Product Name: LABLINKS BIOTECH, India

Role: Java J2EE Developer, May 17 – Sep 17

Description: This manufacturer is a supplier and exporter of specialized equipment for biotechnology laboratories and industry such as bio-reactors, custom built roller bottle equipment with built in incubators, disposable bio-reactor systems of patented design, blotting manifolds, filtration systems. The projects at **LABLINKS BIOTECH** focused on developing a Dynamic website, so that obtained data from database will be used to perform the analysis or business logic on it using big-data technologies

Responsibilities:

- Developed User Interface for the project using JavaScript, JSP, HTML, CSS.
- Developed R code to predict the maximum yield of bacteria for a given set of parameters like temperature, humidity, RPM using liner regression.
- Employed *plumber,* to pass the obtained results from R to website via REST API call.
- Worked on MongoDB for implementing CRUD operations.
- Provided Offline sync option to remote database.
- Involved in bug fixing during the Unit testing, Integrated System testing and User acceptance testing.
- Deployed project in VPS hosting.
- Configured Tomcat 8 server and MongoDB as the windows service flawlessly

Environment: Java 8, R programming, HTML5, CSS, JavaScript, AJAX, JSON, MongoDB, Tomcat 8, R Studio, Plumber

Project Name: AnalyticsConnect
Role: Spark Developer, Feb 17 – Apr 17

Project: *analyticsconnect.info* is an online web-based job portal, only dedicated for big-data technology people. This projects at **Analytics Connect** focused on developing a framework which developed in **Java-Spark-Mongodb** for Regex-Matching of emails, phone numbers, required technologies, keywords in the bulk uploaded resumes, and the respective details and scores are stored to MongoDB.

Responsibilities:

- Worked towards the designing and deployment of Spark**, Mongo DB, java spark-mongo drivers.**
- Able to launch the Spark-jobs from Django application api using `*spark-jobserver*`
- Developed Java Code using spark RDD, and applied regex-match to grep the required information operation on each partition. Mapped users Id with respect to generated result and saves to the **MongoDB**.
- Experience in the installation & configuration of Apache Spark on GoDaddy VPS.
- Deployed project in **Spark-standalone mode**.

Environment: Java SE 8, Python, Spark-core, Spark-SQL, MongoDB.

Project Name: Analytics Connect
Role: Full Stack Web Developer, June 16 to Jan 17
Website: http://analyticsconnect.info/

Project: AnalyticsConnect is an online web-based job portal, only dedicated for big-data technology people. The projects at **Analytics Connect** focused on developing a Modern website using popularly used python, web framework Django and MongoDB as database

Responsibilities:

- Worked on front-end of the project with technologies like HTML, JavaScript, CSS, Twitter Bootstrap and jQuery.
- Used Django web framework and MongoDB 3.4 as Database engine
- Interacted with UI/UX designers for the appropriate Images and dimensions
- Deployed both the Django and MongoDB in Go-daddy VPS hosting, and solved few bugs appeared
- Can be able to successfully migrate the Django-models into database schema

Environment: Python, Django, MongoDB, HTML5, jQuery, Twitter Bootstrap, CSS

Learning:

- Spark Streaming & ML, Kafka
- Blockchain, EOS
- ReactJS, frontend framework

EDUCATION:

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY TIRUPATHI, INDIA
B.Tech, Electronics and Communication Engineering, Graduated with **70% ,** May 2016