Uday Kumar Mandava

Software Engineer(Big Data)

Phone: +91 9848414899| email: udaymandava@hotmail.com

Address: Dno:54-18-37/2, Radio colony-1, Near ITI Bus stop, Vijayawada-520008, AP.

**Overview:**

* 2.9+ years of experience as a Big Data Developer,3+ months as intern in designing & migrating back end applications, integrating with various Databases and as a Java & Python developer.
* Expertise in building stable Big Data pipelines, Cluster Monitoring, Cluster Administration and Documentation.
* Expertise in overall SDLC including requirement gathering, development, testing, debugging, deployment, documentation, production support.
* Facilitated data requirement/status meetings with business and technical teams.
* Familiarity and experience in the work environment consisting of Business analysts, Production Support teams, Subject Matter Experts, Administrators and Developers
* Worked on different Environments and many big data processing tools.
* Worked in multiple cloud Environments (AWS and Google Cloud)
* Exceptional communication and presentation skills.

**Competencies:**

* **Hadoop Distributions**: Cloudera, Hortonworks
* **Data processing engines**: Pig, Hive, Apache **Spark**, Apache **Beam**
* **Data Ingestion**: Sqoop
* **Streaming**: Kafka, Flume
* **Database**: MS SQL Server 2008, Amazon RDS DB, MySQL
* **NOSQL Databases**: MongoDB, BigQuery
* **Cloud Tools:** Amazon EMR, Amazon S3, Google SDK, App-Engine, Dataflow
* **Languages**: Java, Scala, Python
* **IDE’s:** NetBeans, Eclipse, Jupyter, Spyder
* **Others**: Cloudera Manager, Ambari, HUE, Zeppelin
* **Versioning:** SVN, git
* **Scheduling:** Windows task scheduler, Crontab, Oozie, PowerShell(For Event based)
* **WebScraping:** Selenium, BeautifulSoup

**Career Highlights**:

* **TEG Analytics** from Jul 2019 to Current as **Software Engineer(Big Data).**
* **Litmusworld** from Apr 2019 to July 2019 as **Big Data Engineer.**
* **TEG Analytics** from OCT 2017 to MAR 2019, as **Big Data Developer**.

**Key Projects**:

**TEG Analytics (JUL 2019 – CURRENT)**

**Project**: Clorox Ad-Efficiency

**Description**:

Clorox needs real time efficiency numbers of their Campaigns run via ads, online, TV and other streams based on their Brand, Region. Data comes in to Client’s Network on real time basis. A Event based scheduling script monitors network for any new files added, and pushes them to Google Storage temporary bucket(by calling python scripts). Java/Beam code is called every 30 mins by App-Engine Cron job(runs on dataflow) which takes files from temporary bucket and moves them to input bucket. Output from this job is appended to BigQuery final table, which is connected to Tableau Dashboards.

**Environment:** Google SDK, Dataflow

**Tools Used:** Dataflow, Storage, App-Engine, Eclipse, Big-Query, Cloud tools, Apache Beam

**Scheduling:** Windows Task Scheduler, PowerShell**,** App-Engine

**Language:** Java 1.8, Python

**Roles and Responsibilities:**

* Involved from Scratch to Deployment.
* Assist Data Science team in writing SQL Queries
* Migrate those queries to Beam SQL with JAVA.
* Testing those numbers from Beam SQL with Big Query.
* Set up weekly meetings with Client to update them.
* Automated the whole process with App-Engine Cron job so that real time processing is done.
* Designing and developing back-end pipeline in Beam.

**Litmusworld (APR 2019 – JUL 2019)**

**Project:** LW 2.0 Product

**Description:**

Litmusworld has their own product to provide NPS (Net Promoter Score) for all those clients who are into CX(Customer experience). End to End Pipeline has been built in Python with back end as Mongo DB on EC2 servers. This back-end pipeline was migrated to Big Data environment using Spark.

**Environment:** AWS EC2 instances with ubuntu

**Tools Used:** Hadoop, Spark, Mongo DB, Kafka, Python

**Responsibilities:**

* Design a POC and develop codes in PySpark for one brand showcasing big data capabilities
* Administering a Big Data cluster to meet their data requirements.
* Daily stand up meetings and provide every team with KT on how to work with Spark.
* Designing and developing back-end pipeline in Spark.

**TEG Analytics (OCT 2017 – MAR 2019)**

**Project:** Nike

**Description:**

Nike wants to monitor the day to day sales activities of all stores. Nike is not having the complete visibility of the stores data and the data for the stores not in ERP system. They won’t be able to get the store data on time causing delay in making the decisions at store level. We provide a single version of the truth for Sales, Merchandizing, and Finance team by monitoring key performance indicators in real time and allow Nike management and store owners to take immediate action. Create, analyze and maintain Sales Performance reports, recap and analyze actual sales results to plan, identify risks and opportunities to business and recommend strategies to achieve financial goals**.**

**Environment:** Hortonworks, Ubuntu

**Tools Used:** Hadoop, Spark, SSIS, SQL Server 2008, Hive

**Responsibilities:**

* Meet Project requirements and showcase Hadoop capability by understanding their data and deliver POC’s.
* Requirement gathering from each subsidiary, Understand each process, files, SQL queries etc.
* Design and co-ordinate between Backend and frontend team.
* Constructed a data ingestion pipeline to gather data from SFTP that is a single utility which will load data into Hadoop from the SFTP location.
* Execution of Sqoop script to export and import data from RDBMS to Hadoop (HDFS) and vice-versa
* Gather information and Prepare Weekly Status Report.
* Built Spark Scripts on Scala.
* Monitoring Development and Production Environment.
* Scheduled restart of Cluster and Administering the Cluster.
* Writing Core Business Logics.
* QC on Refresh, Report Query outputs from Spark & SSIS, SQL.

**Project**: Clorox IIW, Clorox DDW

**Description**:

Clorox needs their sales report and other reports based on their DMA region and in timely manner (bimonthly, monthly, weekly). There were many manual steps involved like getting data from their portal, clean up, refresh and arrange data accordingly. These manual steps were automated when they migrated to Big Data Environment.

**Environment:** Hortonworks, Ubuntu

**Tools Used:** Hadoop, Spark, AWS EMR, AWS S3, SSIS, SQL Server 2008

**Roles and Responsibilities:**

* Understanding technical aspects of requirement, its coverage and entry and exit criteria.
* Design Migration with Data Science team.
* Work closely with Clorox team to prepare them for EMR Environment.
* Built Spark Scripts on Scala, Python.
* Monitoring Development Environment and Deploying EMR Cluster on AWS.
* Executing Spark Refresh/Report Queries on EMR and Store output to Amazon S3.
* Creating and Editing Unix Shell Scripts for downloading/uploading data to EMR.
* Scheduled restart of Development Cluster and Administering the Cluster.
* Migration of SAS & SSIS Packages to Big Data Environment.
* Automation of Manual Steps they had in SAS Environment.

**Academia:**

* **Bachelors of Technology** in Mechanical Engineering (Aggregate 7.8/10) from **VIT University, Vellore, India.**
* **Masters of Science in** Computer Science (Aggregate 3.55/4) from **Silicon Valley University, California, USA.**

**Declaration:**

I hereby declare that the above information and facts are true to the best of my knowledge and belief.

DATE:

PLACE: Bangalore. (MANDAVA UDAY KUMAR)