Thangarajan

Data Engineer with 5+ years of experience in Big data and building data products. Implemented multiple data engineering project across many domains such as Retail, Market Research and Healthcare. Deep knowledge on Hadoop, Spark, Flume, Hive, NiFi, Kafka, Solr, ELK (Elasticsearch, Logstash, Kibana), MongoDB, Neo4j. Acquired programming knowledge in Python, SQL, Java. Having hands-on dirty data to curated data. Having good understanding on NLP and having experience in deployment.

# EMPLOYMENT

**Feb 2018 - present** Senior data engineer – **Pathfinder Global FZCO, Chennai**

* Designed and developed complex data pipelines and maintained the data quality to support a rapidly growing business
* Utilize Apache Spark with Python, Apache Drill & MongoDB.
* Building ETL Pipeline in Apache Spark and Python.
* Built data pipeline for collecting information multiple heterogeneous systems.
* Worked closely with the Analytics team, API team, Web content team, internal Operations, external Hadoop Operations and other teams to integrate and operate data services across the company
* Provided suggestions and solutions to client IT needs.
* Handling all Research and Development projects with 5-8 team members.

**Mar 2017 – Feb 2018** Senior Executive – **AC Nielsen, Chennai**

* Built an OSRM (Open Source Routing Mechanism) for Auditors travel plan.
* Build a text search engine using Apache Solr and Python
* Built data pipeline for many KPI and developed executive dashboard using Spotfire.
* Initiated a lot of fine-tuning mechanisms to tune the database as well as the queries to complete a set of given jobs or tasks in optimal time
* Wrote production level code for logging, querying, and media retrieval

**Dec 2013 – Feb 2017** Senior Software engineer – **Hexaware Technologies, Chennai**

* Build 11 node Hadoop cluster (Cloudera) for CoE Environment
* Built streaming analytics with Apache Spark and Apache NiFi
* Built reports using Tableau\Qlikview\Spotfire
* Hive execution tuning using Apache Tez
* Wrote SQL and PL/SQL programs to identify and correct data anomalies at request of SME's
* Built interactive Data explorer tool using Tableau

# Skills

**Platforms:** Cloudera and Hortonworks

**Big Data**: Hadoop, Sqoop, Hive, Spark with Python, Apache NiFi, Kafka,

Flume, Apache Drill,

**Visualization Tools**: Tableau, Qlikview and Sisense.

**Programming**: Python and Java.

**Database**: MySQL, SQL Server.

**NoSQL**: Apache Solr, Elasticsearch, MongoDB and Neo4j.

# EDUCATION

**2010 – 2013** Master of computer applications– Madurai Kamaraj University

**2007 - 2010** bachelor of computer science– Madurai Kamaraj University

# Certification

**2019** Neo4j Certified Professionalfrom **Neo4j**

**2016** MongoDB for Developers from MongoDB

**2015** IBM Big Data Fundamentals Technical Mastery Test v1 and V2

**2014** MapReduce Programming, Moving Data into Hadoop, Hadoop Fundamentals I - Version 3

# AWARDS

**2015** **Star of the Unit 2015**

For exemplary contribution on Business Intelligence and Business Analytics

**2013 Top Performer 2014**

Came up with out of box ideas to leverage Big Data Analytics

# Data products

**OCR Engine**

Extract content from PoS (Point of Sales) system Bill and parse product name and totals.

**Big Leap**

A data platform to move data from major RDBMS to Hadoop and to maintain the Hadoop cluster.