Thangarajan Pannerselvam

Synopsis: Possess over 5 year experience in BigData analytics landscape having extensive exposure to the latest Big Data technologies like Hadoop, Spark, Hive, MapReduce ,Pig, Flume, SQOOP, Apache NiFi ,NoSQL's and also have knowledge on Image and Video Processing and Text Mining coupled with big data programming languages like **Python,** Scala and Java to unearth true value of big data to clients in large scale enterprise systems.

EXPERIENCE SUMMARY

- Expertise in Bigdata technologies and Hadoop configuration, BigData administration, Storage, Processing and Retrieval.
- Expertise in Text Mining methodologies such as **NLP**, Topic Modeling and Sentiment Analysis using LDA, Guided LDA
- Expertise in Image Processing methodologies such as Object Detection using Tensorflow and YOLO V3, OCR using Tesseract, also developed People count solution for Retail and Vehicle Classification and Count solution for Governments
- Have strong knowledge in Hadoop architecture and Eco-System
- Expert in Data Migration and Data Ingestion from different data sources to HDFS and NoSQLs using Flume, SQOOP, Apache NiFi and WebHDFS Rest API.
- Strong knowledge of SQOOP, Hive, PIG, HDFS, Flume.
- Expertise in Spark using Scala and Python, Hadoop v2, MapReduce and HDFS
- Generating Automated SQOOP Scripts to migrate data from/to Hadoop from/to RDBMS
- Have good knowledge in NoSQL like Elasticsearch, MongoDB and Neo4j Graph DB.
- Expertise in Big data integration with Cognos and Informatica & Reporting with Qlikview and Tableau.
- Expertise in Hadoop ecosystem, Linux commands and Web Content Extraction using Python
- Having good knowledge in Business Intelligence concepts
- Works well in both team environment and Individual assignments.
- Successful in meeting new challenges, leading a team and finding solutions to meet the needs of the clients and have good communication skill
- Excellent turnaround time to meet the requirements of clients.

Areas of Experience

Industries
Retail
Healthcare, Insurance, Education, Social Media Analytics

Big Data Administration
→ Hadoop Configuration
→ Oozie
→ Ambari

Big Data Ingestion and Retrieval
→ SQOOP, Flume, Apache NiFi
→ Hive, Impala

NoSQL
→Elasticsearch, MongoDB

Big Data Programming Stack
→Spark, Map Reduce, Pig
→Scala, Python and Java

Analytics and ML
→NLP, Tensorflow,

Reporting Tools:
→Tableau, Qlikview
→Spotfire, Sisense

RDBMS
SQL Server 2008 R2 /2012, My SQL, Postgres

Operating Systems
UNIX, Ubuntu

QUALIFICATION

Year	Institute/University	Degree/Examination	Percentage
2010-2013	Madurai Kamaraj University, Madurai.	Master of Computer Applications	77.67
2007-2010	N.M.S.S.V.N. College Nagamalai, Madurai.	Bachelor of Computer Science	71.27

Certification	Institute	Year
Neo4j Certified Professional	Neo4j	2019
Introduction to Neo4j	Neo4j	2019
MongoDB for Developers	MongoDB, Inc.	2016
IBM Big Data Fundamentals Technical Mastery Test v1	IBM	2014
IBM InfoSphere BigInsights Technical Mastery Test v2	IBM	2014
R Programming	Coursera	2014
Introduction to MapReduce Programming, Moving Data into Hadoop, Hadoop Fundamentals I - Version 3	IBM Big Data University	2014

Company	Designation	Period
Hexaware Technologies	Retainer	December 2013 to May 2014
Hexaware Technologies	Software Engineer	June 2014 to Feb 2017
AC Nielsen	Senior Executive	March 2017 to February 2018
Pathfinder Global FZCO	Senior Data Engineer	February 2018 to till date

Award	Description
Star of the Unit 2015	For exemplary contribution on Business Intelligence and Business Analytics
Top Performer 2014	Came up with out of box ideas to leverage Big Data Analytics

Project Name	:	Third Eye – People and Vehicle count solution
Role	:	Team Lead
Environment	:	Ubuntu
Duration	:	Feb 2018 to till date
Tools Used	:	Python 3.5, Tensorflow, YOLO V3

Description:

Third eye is a Face Recognition based solution to monitor Footfall in malls and airports. It is capable of finding face demographic information's like Gender, Age group and Emotions for the people entering the premises. Crime prevention, vehicle classification and license plate recognition and authorizing the staffs.

Roles and Responsibilities:

- As a team lead provide my extensive knowledge on Image and Video Processing and working with customers to understand algorithm requirements and deliver high-quality solutions
- Research, design, and implement algorithms in Deep Learning for Computer Vision and Image Analytics.
- Contribute to research projects that develop a variety of algorithms and systems in computer vision, image and video analysis.
- Fast prototyping, feasibility studies, specification and implementation of image and video analysis product components.

Project Name	:	Retail Index and Benchmarking
Role	:	Team Lead
Environment	:	Amazon EC2, MongoDB, AWS S3, Apache Kafka, Apache Spark
Duration	:	Feb 2018 to till date
Tools Used	:	Python 3.5, Apache Drill

Description

Involves combing the sales data of all shopping centres across the country into data lake and to see the trends and the direction in which the Retail industry is going. How is the growth of Retail industry in India? How is the trend in next 6 months? Is it upwards or downwards or flat? If so, at what rate? How is a category [Ex: Fashion] is performing in a market [Ex: Mumbai]?

Roles and Responsibilities:

- As a team lead finding the pain point of the customers on finding How generate more revenue on Malls
- Creating data pipeline using Apache Drill to access multiple MongoDB and file system at one place
- Creating KPI to clients so that they can easily find out where they need to change their business strategy

Project Name	:	OData 360° Catchment Analysis – New store location recommendation engine
Role	:	Team Lead
Environment	:	CentOS, MongoDB, GCP (Google Cloud Platform)
Duration	:	July 2018 to October 2018
Tools Used	:	Python 3.5, REST, Leaflet JS

Description:

OData 360° makes it easy to choose a best location to open a store and to identify potential area around it. It is based on the geo spatial concepts and advanced statistics which can make effective analysis on customer demographics, social economic, trading and competitors' factors and will help in analyzing the proper catchment areas. Draw flexible and actionable market boundaries for area up to the level of 5 sq.km. Prioritize the market based on the consumer demography, customer spent pattern, city's sales potential to perform the catchment analysis and to know the competitor's presence.

Roles and Responsibilities:

- As a team lead Identifying trusted data sources for New store location recommendation engine
- Developing data pipeline using Python to extract relevant data from Open Data and Google Place API
- Integrating and relating multiple data source to build the New store location recommendation engine and catchment analysis so the client can easily know which place they can pitch in.
- Instead of raw data points providing the client with a Map UI so the client can easily customize their need on the fly

Project Name	:	Planning and Route Optimization
Role	:	Team Member
Environment	:	CentOS, OSRM
Duration	:	August 2017 to Feb 2018
Tools Used	:	OSRM, Python, Oracle 11

Description:

Planning and Route Optimization is to create a travel schedule for auditors to visit nearby stores based on the Auditor Location. It includes to find the actual distance and best path to travel. And provide real-time routing.

Roles and Responsibilities:

- As a Team Member of the project provided an Open Source Solution for Route Optimization.
- Setting up OSRM (Open Street Routing Machine) in Centos
- Finding fastest path between two geo-points using OSRM
- Providing Adhoc routes to auditor in real-time with low time latency
- Implemented OSRM instead of commercial product as a cost saving

Project Name	:	Match and Merge on Global Store Data
Role	:	Team Member
Environment	:	CentOS
Duration	:	March 2017 to July 2017
Tools Used	:	Apache Solr and Python

Description:

Match and Merge Algorithm enables GSD (Global Store Data) to accept data from multiple data sources and perform matching in such a way that only one golden copy of the store in retained. Performs the matching across various attributes and gives match score for all the matches displayed. Perform both exact and inexact matches based on the threshold defined by the users. Also provide results with less relevance to the actual query.

Roles and Responsibilities:

- As a Data Engineer of the project provided an Open Source Solution
- Setting up Centos instance setup
- Setting up Apache Solr environment for Development and Production.
- Deploying, managing and configuring Apache Solr.
- Building data migration framework using Python to copy data from Oracle to Apache Solr.
- Creating python code for processing data using Pandas.
- Implemented Levenshtein distance, Metaphone, Double Metaphone, Fuzzy Search and Jaro-Winkler

Project Name	:	Text analytics project for leading international educational firm
Role	:	Data Engineer
Environment	:	Hortonworks, CentOS, Apache Hadoop, AWS
Duration	:	June 2015 to Till Date.
Tools Used	:	Apache Tika, HDFS, Spark, Scala, Hive, SQOOP, Qlikview

Description: The purpose of project is to create application which can detect plagiarism in the assignments submitted by students through online. This involved extraction of text content form .DOC, DOCX and PDF using Apache Tika and load it into Hadoop .Using MapReduce processing engine we create N-Grams and using Jaccard co-efficient to check similarity between the documents. And the reporting has been done in Qlikview.

Roles and Responsibilities:

- Setting up Java/MySQL instance setup
- Setting up Apache Hadoop environment for development and Hortonworks for production.
- Deploying, managing and configuring HDP with Ambari.
- Building content extraction framework using Apache Tika.
- Creating map reduce programs for n-gram assignment to documents for document similarity
- Creating external tables in HIVE for reporting and creating Hive Queries for reporting.
- Performance improvement through Apache Tez for Hive.
- Creating file watcher for student assignments to move files to HDFS using Apache Flume
- Moving structural data to MySQL using SQOOP for reporting

Project Name	:	Twitter Real-time Named Entity Extraction
Role	:	Data Engineer
Environment	:	Hortonworks, CentOS, Elasticsearch, Kibana and Neo4j
Duration	:	Jan 2015 to May 2015
Tools Used	:	Python, Flume, HDFS, Spark, Cypher Query, Java, JSP and Java script

Description: The purpose of project is to create application which can pull streaming Twitter data into Bigdata environment HDFS and Elasticsearch. The dashboard has been visualize using Kibana. Then named entity extraction done using spark and stored in Neo4j as a batch. Using Neo4j data an interactive network Visualization has been developed using Vis.js.

Roles and Responsibilities:

- Setting up Elasticsearch, Kibana and Neo4j instance.
- Building Named entity Extraction engine using Spark.
- Creating a real-time interactive visualization Dashboard in Kibana.
- Creating an interactive web application using JSP and Java script that tell user how the entity has related each other via Node and Edges.

Project Name	:	Detect and Report Web-Traffic Anomalies in Near Real-Time
Role	:	Hadoop Developer
Environment	:	Hortonworks, CentOS, Elasticsearch and Kibana
Duration	:	August 2014 to December 2014.
Tools Used	:	Hive, Pig, Flume, HDFS, Apache Tez, Hue

Description: The purpose of project is to create near real time log analytics application for analyzing huge logs of web applications servers and IIS servers of leading investment banking firm. The application will provide real time insight for different server metrics which will help manage their IT infrastructure in better fashion by detecting major bottlenecks. This involves real time data ingestion of web server logs into Elasticsearch and HDFS using Flume.

Roles and Responsibilities:

- ➢ Setting up Apache Hadoop for development environment and Hortonworks for production
- ➢ Configuring Flume agents and HDFS/ Elasticsearch sinks for capturing real time data for web server logs
- ➢ Extracting useful information of web server log like user info, time stamps, web page accessed, IP,
- ➢ Browser Information using Pig
- ➢ Creating HIVE external tables and Queries for reporting in Hive data warehouse
- ➢ Performance improvement through Apache Tez for pig scripts and HIVE
- ➢ Creating network chart visualization using JSP to find User Travel Path
- ➢ Providing Kibana GUI for exploration of web log data using Elasticsearch.

Project Name	:	Fraud detection for Auto Insurance Company
Role	:	Report Developer
Environment	:	Windows 7
Duration	:	May 2014 to July 2014
Tools Used	:	Microsoft SQL Server and Qlikview

Description

In the Insurance Company claims are submitted, out of which few claims can be fraud. Job involves gathering requirements and setting up a design for dataflow. Creating ETL mappings and loading data into the database for report generation. The idea of the work is to find the fraudulent claims and analyze the factors which normally makes a claim a fraudulent claim like high claim amount, number of parts damaged etc.

Roles and Responsibilities:

- ➢ Analyzing the existing EDM data model in the SQL Server.
- ➢ Capturing important metrics and dimensions.
- ➢ Determine Interface points – ETL and Reporting.
- ➢ Created Standard Data Model in MSSQL.
- ➢ Generate interactive dashboard such as displaying fraudulent claims. Created KPI in Qlikview.

Project Name	:	Customer and Employee Competitor Analysis
Role	:	Team Member
Environment	:	Windows 7
Duration	:	March 2014 to April 2014
Tools Used	:	Python, Microsoft SQL Server Tableau Desktop, Tableau Server and Qlikview

Description:

Job involves gathering requirements and setting up a design for dataflow. Creating ETL mappings and loading data into the database for report generation.

Responsibilities:

- Work in the PoC for existing as well as prospective clients on social media analytics
- Extracting data from social media sites like – Twitter, Facebook, Yammer and Glassdoor using Python
- Cleansing the data extracted from social media
- Generating reports on sentiment analysis for existing as well as prospective clients
- Generate interactive dashboard such as word cloud based from Glassdoor data's n-gram
- Participated in benchmark analysis between Ernst & Young and its competitors

Project Name	:	SAP BO Administration
Role	:	Team Member
Environment	:	Business objects XIR2, XI3.1, and AIX 6.1
Duration	:	Dec 2013 to Feb 2014
Tools Used	:	BO (BO XI R3 and XI R4), Web Sphere 7, Tomcat 6, AIX 6.1, Web Logic 10.3, IHS

Description: The Project involves maintaining Business object applications through AIX 6.1 operating system. Also involves installations, migrations, user management and resolving issues.

Responsibilities:

- Troubleshooting issues that the user faces while logging in, security issues and all access related issues.
- Scheduled BO migrations and upgrades across versions.
- Responsible for creating & maintaining various user & groups and their security using CMC
- Configuring webservers & deploying war files.
- Setting alert notification for report failures.
- Managing servers through CMC and UNIX box
- Installation of BO on AIX Servers, WAS cell creation
- Monitoring reports for failure and troubleshoot for successful scheduling.
- Server maintenance and space management.

Tools Built

Product Name	:	OCR Engine
Components	:	Extract and Parser
Environment	:	Ubuntu, NLP,
Tools Used	:	Python 3.5 Tesseract, NLP, OpenCV

Description:

OCR engine to extract content from PoS (Point of Sales) system Bill and parse product name and totals Using Python, Tesseract, NLP and Image Processing Techniques.

Product Name	:	Big Leap Platform
Components	:	Data Governance and Data Migration
Environment	:	Hotonworks Data Platform, CentOS, Elastisearch and Kibana
Tools Used	:	Hive, Sqoop, Flume, Python and PyWebHDFS

Description:

A component of “Big Leap Platform”, it’s a web-based application used by Data Admin to push data into HDFS from relational databases (MySQL, Microsoft SQL server etc.). Before pushing the data, Admin can do look up of sample data and restrict some of column from loading and can-do data type transformation. Extracting Yarn job information from Yarn REST API and Store it in Elasticsearch using Python. Extract HDFS metadata information using WebHDFS using python. Alert environment users on abnormal execution on Job based on previous job execution stats.

Awards and Recognition

Award	**Description**
Star of the Unit 2015	For exemplary contribution on Business Intelligence and Business Analytics
Top Performer 2014	Came up with out of box ideas to leverage Big Data Analytics

Personal Details:

Name as in Passport	:	THANGARAJAN P
Email id	:	thangam.rajan8@gmail.com
Mobile No	:	+91 8220829575
Date of Birth	:	11/01/1990
Passport No	:	N9683993
Passport Expiry Date	:	26/04/2026
Marital Status	:	Single
Nationality	:	Indian
Sex	:	Male