**SINCHAN DATTA**

Phone no: 9874878551

Email ID: sinchandatta90@gmail.com

|  |  |
| --- | --- |
| **EXECUTIVE SUMMARY** |  |

An IT professional with total 6 years and 4 months of working experience in various technologies which includes Hadoop Development, Data Science, Machine Learning, Python, AWS, Teradata, SQL and Unix.

|  |  |
| --- | --- |
| * Quickly Adapts to New Professional Settings * Fast Learner; Acquires Skills/Knowledge Easily * Persuasive, Influential, Effective Communicator | * Skilled in Research/Analysis/Statistical Functions * Extensive Technical and Analytical Skills * Strong Organizational and Time Management Skills |

Skills/Technical Expertise

* 6 years 4 months of Industry Experience in Big Data, Data Science, Python, AWS, SQL, Unix.
* Proficient in Apache Hadoop, HDFS, Hive, Pig, Oozie, Kafka, Oozie, Sqoop.
* Experience in data ingestion and data transformation from relational databases to Data Lake.
* Experience Spark with Scala, Java and Python.
* Experience in Python Numpy, Matplotlib and Pandas.
* Experience in AWS IAM, EC2, S3, EBS, ELB, RDS, DynamoDB, EMR, RedShift.
* Experience in HBase NoSQL Database.
* Experience in Machine Learning Algorithms K nearest Neighbour, Naive Bayes, Logistic Regression, Linear Regression, Support Vector Machines, Decision Trees, Ensemble Models.
* Experience in Natural Language Processing.
* Experience in Performance measurement of Machine Learning Models.
* Experience in Dimensionality Reduction and Visualization using PCA, t-SNE.
* Experience in Exploratory Data Analysis.
* Understanding of Linear Algebra, Probability and Statistics, Feature Engineering.
* Experience in Data Ingestion from Relational Databases to Hadoop in Data Lake.
* Exposure to Performance tuning in Teradata and Hadoop.
* Experience in HTML, CSS and JavaScript.
* Highly adaptable, quick learner, planning skills, honest and hardworking.
* A good team player with strong communication & interpersonal skills.

Area of Expertise

* **Big Data Ecosystems**: Hadoop, Hive, HDFS, Sqoop, Oozie, Pig, Spark.
* **Programming Languages**: Scala, Java, Python
* **Database Programming**: Teradata, Oracle.
* **Web Technologies**: HTML, CSS, JavaScript.
* **Scripting Languages**: UNIX.
* **Job Scheduling Tool**: Autosys.
* **Platforms**: Mac OS X.

Education

Post Graduate Diploma in Business Analytics (Executive) from IMT Ghaziabad Center of Distance Learning (2017-2018)

Bachelor of Technology (Computer Science &Engineering) from Guru Nanak Institute of Technology (2008-2012)

Professional Experience

Company Name: Cognizant Technology Solutions India Private Limited (Oct 2016 - Present)

**Project #1:**

Title: Data Ingestion and Data Transformation

Client: Leading Bank

Role: Big Data Developer

Project Summary:

This project is to ingest and transform the data sources from multiple relational data sources to Data Lake. This is a development project to implement transformation logic using Spark and Scala. Transformation are written in Scala and applied on data using Spark SQL. Implemented models using machine learning models and natural language processing for the application identifying use case along with Exploratory Data Analysis.

**Role and Responsibilities**:

* Coding
* Unit Testing
* Ensure Day to Day Seamless Delivery

Tools & Technology: HDFS, Hadoop, Hive, Shell Scripting, Scala, Spark, AWS, Python, Redshift, Data Lake, Python, Machine Learning, Natural Language Processing

**Project #2:**

Title: Data Ingestion and Transformation

Client: A leading Insurance Company

Role: Big Data Developer

Project Summary:

This project is to ingest and transform the data sources from multiple relational data sources to Data Lake. This is a development project to ingest raw data Ingested into HDFS system and transformation logic implemented on this using Spark. Finally, the processed data consumed by Analytics team for reporting purpose.­­

**Role and Responsibilities**:

* Coding
* Unit Testing
* Ensure Day to Day Seamless Delivery

Tools & Technology: HDFS, Hadoop, Hive, Pig, Shell Scripting, Core Java, Spark, AWS, Python, Redshift, Data Lake, Python

**Project #3:**

Title: Data Ingestion

Client: A leading Insurance Company

Role: Big Data Developer

Duration: November 2016 to April 2017

Project Summary:

This is a development project to ingest their sensitive and non-sensitive data into Hadoop ecosystem. The data source comprises of Oracle, DB2 and ADABAS files. Raw data Ingested into HDFS system and transformation logic implemented on this using Spark. Finally, the Processed data consumed by Analytics team for reporting purpose.

**Role and Responsibilities:**

* Coding (involves fetching metadata for various data sources like Oracle, DB2, ADABAS files using Jar. Sqoop the data based on that then basic cleansing with Pig. Put the data into Hive Landing and persistent zone)
* Use Autosys scheduler to make it as a Job
* Unit Testing
* Ensure Day to Day Seamless Delivery

Tools & Technology: HDFS, Hadoop, Hive, Pig, Shell Scripting, Core Java, Python

Company Name: Infosys Limited (February 2014- September 2016)

**Project #1**

Title: Reports and Charts Generation

Client: A leading product based company

Role: Big Data Developer

Duration: From June 2014 to Jan 2015

Project Summary:

Charts Reporting project provides daily sales extracts of music and related contents to external partners based on the data received from data warehouse.

**Role and Responsibilities**:

* Development and enhancement of reports by ETL technologies using Teradata and Hadoop.
* Involved in writing Teradata Stored Procedures, Macros and Complex SQL Queries.
* Involved in writing Hive queries and job scheduling through Oozie workflow.
* Performance Tuning of Hive Queries.
* Scheduling of jobs through Autosys Tool.
* Involved in implementation of relevant SDLC activities.
* Involved in Autosys Job creation and Configuration files.

**Tools & Technology**: HDFS, Hadoop, Hive, Teradata, UNIX, Sqoop, OOZIE

**Project #2**

Title: Analytics Project

Client: A leading product based company

Role: Big Data Developer

Duration: From Feb 2015 –July 2016

Project Summary: Analytics project provides various insights and thus a thorough analysis of business to the client which helps the client to take effective strategic and business decisions.

**Role and Responsibilities:**

* Development and enhancement of reports by ETL technologies using Teradata and Hadoop.
* Development of adhoc reporting requests using Hadoop and Hive.
* Experience in Sqoop to transfer data from RDMS to Hadoop and vice versa.
* Performance Tuning of Hive Queries.
* Involved in working with Teradata Utilities like Fastload, FastExport and Multiload.
* Involved in implementation of relevant SDLC activities.

**Tools & Technology**: HDFS, Hadoop, Hive, Teradata, UNIX, Sqoop

**Project #3:**

Title: Reports and Charts Generation

Client: A leading product based company

Role: Big Data Developer

Duration: From August 2016 to September 2016

Project Summary: Charts Reporting project provides daily sales extracts of music and related contents to external partners based on the data received from data warehouse.

**Role and Responsibilities:**

* Development and enhancement of reports by ETL technologies using Teradata and Hadoop.
* Involved in writing Teradata Stored Procedures, Macros and Complex SQL Queries.
* Involved in writing Hive queries and job scheduling through Oozie workflow.
* Performance Tuning of Hive Queries.
* Scheduling of jobs through Autosys Tool.
* Involved in implementation of relevant SDLC activities.
* Involved in Autosys Job creation and Config files.

**Tools & Technology:** HDFS, Hadoop, Hive, Teradata, Unix, Sqoop, OOZIE