Andalib Shadani Ansari

DOB- 02nd September, 1989

501 A Wing, La Bellezza, Rishivan, Near Siddharaj Colony,
Kajupada Road, Abhinav Nagar, Borivali East, Mumbai-400066
Mobile: +91 8105752007, Email: andalib.info@gmail.com

Summary

- I am currently working with **Freestone InfoTech Pvt Ltd, Mumbai,** from 2nd January, 2014 to till date, as a **Data Analyst.**
- I have **1.3 years** of experience in **Big Data (Hadoop, Hive, and Pig), Bash Scripting (UNIX), SQL, MySQL and Data Warehousing**.
- Previously I worked for **Trendwise Analytics** (a Big Data start-up in Bangalore) as an **intern.** There I worked on three different offshore projects as a team member and underwent various short term trainings on **Hadoop, Hive, Pig, Sqoop and NoSQL Technologies (MongoDB, Cassandra, HBase).**
- Good exposure of **Amazon's** Cloud based Services- **EC2, EMR, RDS, Redshift**

Work Experience

Current Assignment: Automation of ETL Pipeline for Data Aggregation and Credit Scoring

Tools and Technology Used: Amazon's EMR, Hive, Pig, S3, Customized Java Apps

Location: Mumbai, India

Role: - Team Member

Project Summary:

Cignifi Inc. provides data and credit scoring infrastructure for mobile operators to maximize the value of financial product offerings. Cignifi's credit scoring technology includes **Cignifi Risk Score**, a significant discriminator of default risk, and **Cignifi Response Score**, a significant score of customer take-up.

The aim of this project is to build an automated ETL Pipeline and Credit Scoring Infrastructure (Cloud Based) for the clients (Telecom Operators). The designed infrastructure is able to handle Big Data (Terabytes of Data) gathered from the clients. Once data is received or uploaded to Cignifi's FTP Servers, they go through multiple phases of Data Preparation, Data Aggregation and Credit Scoring Processes to build scores. We are also building different monitoring apps to track these automated tasks.

My role includes:
- Involved in the daily Scrum Call with the analytics team in the **US** and **Brazil** to understand each clients' requirement
- Writing different bash scripts for pre-processors, data validations etc.
- Writing Hive and Pig Scripts for Data Preparation, Aggregation and Credit Scoring
- Writing different Hive and Pig UDFs if needed
- Coordinating with the development team for different interrelated tasks
- Tuning different Hive and Pig processes to improve the performance at production
- Involved in benchmarking of different Hadoop components

Clients- Airtel (Ghana), Equifax (Mexico), IUSACELL (Mexico), TISA (Nicaragua and Peru) and J6W (Philippines)

Tools Used: Behavior based ETL Platform

Role: Data Analyst
Location: Mumbai, India

Project Summary: Data generated (Humongous Data) in the form of Call Data Records (CDR), Accounts and Recharge are collected together for Analysis, Predictive Modeling and Credit Scoring. The clients' raw data vary from **Gigabytes** to **Terabytes** in a single month. We are running our apps at Amazon's **EC2** for data processing**.**

My role included:
- Involved in daily **Scrum Call** with the analytics team in the **US and Brazil** to understand each clients' requirement
- Analyzing raw data collected from clients. I use Data Science of Unix to analyze, visualize, and doing preprocessing on raw data
- Doing data visualization using Python (Pandas, Matplotlib, NumPy etc.) to understand raw data more deeply

- Performing pre-processing and data cleaning on raw data using shell script (at **AWS EC2**).
- Generating input variables (behavior based) for different scoring models.
- Transforming raw data according to business need
- Performing aggregation on transformed data
- Doing Data validations.
- Managing MySQL Server at production.
- Monitoring Production System to check issues, and raising tickets if issues found.

Note: All above mentioned projects have been outsourced from **Cignifi Inc., USA**.

Product: Real TAM (Television Audience Measurement), Trendwise Analytics

Tools Used: NoSQL Databases (MongoDB, HBase, Cassandra)

Role: Team Member

Location: Bangalore, India

Project Summary: Trendwise Analytics is engaged in building a product that can provide a real-time analytics to Media Companies. The product can be connected to different data sources and then can be used to visualize real-time data. Real TAM is based on Television viewership, big data gathered through digital set top box and social media using Apache Hive, Apache Sqoop and NoSQL data stores.

My role included:
- Analyzing architectures of **HBase, MongoDB and Cassandra**.
- Working with different data modeling techniques in NoSQL database environment, and learning how data can be stored and retrieved from these databases.
- Testing with bulk data import and export in NoSQL Databases.

Client: MCN – Middle East Communication Network - Dubai

Tools Used: MS SQL Server 2008, MySQL 5.0

Role: - Team Member

Location: Bangalore, India

Project Summary: MCN has seen a dramatic growth in its operations; the opportunities to develop new services and helping the customers with reporting facility have seen the Company expand in a number of directions.

 This project contains two modules.

1. Administration – Maintaining the Data Warehouse for the new data, and also the cube.
2. Reports – Generating the reports based on data available.

My role included:
- Interviewing MCN team and analyzing requirements
- Documenting all requirements in the User Requirement Specification
- Designing Data Model
- Writing relevant section of design document
- Designing and building Data Warehouse
- Designing and building Cube using MS SQL Server 2008 (Analysis Services)
- Writing SQL scripts and PL/SQL procedures.

Client: Brilliant @ Group, South Africa

Tools Used: MySQL 5.1, PHP Reporting

Role: - Team Member

Location: Bangalore, India

Project Summary: Due to increasing number of customers and new services the Brilliant @ Group has decided to provide better reporting facility to their customers. Currently they are having two databases in different regions to serve their customer, but now they are planning to merge these databases into a central database. The existing databases need to be modified and normalized for their new requirements.

The Project contains four modules:

1. Merging and Normalizing the existing databases
2. Administrator Reports Design
3. Corporate Reports Design
4. Data Suppliers Report Design

My role included:
- Interviewing Brilliant @ Group team and analyzing requirements
- Documenting all requirements in the User Requirement Specification
- Designing Data Model for Centralized Database
- Writing relevant section of design document
- Designing and building Data Warehouse for New Reporting System
- Writing SQL scripts for all different reports.

Big Data Skills

- Good exposure of **Hadoop Ecosystems** and **NoSQL Technologies** (MongoDB, HBase and Cassandra)
- **Hive:** good understanding of hive architecture and Hive Queries. I was involved in a POC to analyze telecom's data with the power of Hadoop and Hive.
- **Pig:** good understanding of pig architecture and ETL using pig scripts. I was involved in a POC to process telecom's data using PIG to move them into HDFS for further use.
- **Sqoop:** good understanding of **Import, Incremental Import, Export and Free-form Query.**
- Good Understanding of **Google's BigQuery Architecture** and good working knowledge on **Big Query platform.**
- **Hortonworks HDP 2.1 , Amazon's EMR, EC2, RDS, Redshift**

Business Intelligence Skills

- **OBIEE 10g:** good understanding of **Repository Design (RPD), Dimension Hierarchy Building, Answers, Creating Reports and Dashboards**
- **Cube Design using MS SQL Server 2008/2012**

Computer Skills

1. Database

- **SQL** (Strong knowledge of DDL, DML, DCL, Advanced SQL Tuning)
- **PL/SQL**: Knowledge of all basics functionality of PL/SQL
- **Oracle 10g**: knowledge of Database Configuration Assistant: Instance creation, Knowledge of Database Connectivity.
- **Data Warehousing**: knowledge of DW Schema, Data Modeling

2. Programming Language: Python (NumPy, Pandas, Matplotlib)
3. Platforms: Windows 2000/XP/Vista/7, UNIX (**good working knowledge of data science in UNIX)**

Certifications

- Certified Developer in **MongoDB for Developers using Python**
- Certified Developer in **MongoDB for DBA**

Education

Year	Degree	Institute	CGPA/Percentage
2009-2013	B.Tech-CSE	Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad	9.53/10
2007	10+2, CBSE	Ishan International Public School, Patna	53%
2005	10th, CBSE	Millia Convent English School, Purnia	71%

- Awarded **Merit Certificate** for Securing 9th position (among top 10 outstanding performers) in **Department of Computer Science and Engineering** of **Faculty of Engineering and Technology.**

Academic Projects

Summer Training: Demonstrating OBIEE 10g Capability

Intern, HSG Technologies Pvt. Ltd., Noida
Duration: 4th June, 2012 to 27th July, 2012

BI Tool: Oracle Business Intelligence Enterprise Edition (OBIEE 10g)

Database: Oracle 10g

My Role: Working with **RPD**, **Dimension Hierarchy Building**, Working with **Answers, Creating Reports** and **Dashboards.**

Description:
OBIEE is one of the leading products by Oracle in **BI (Business Intelligence)** market. The idea behind the project was to build a sample reports and dashboards to showcase OBIEE capabilities and its feature. This application can be hosted on any OBIEE sever and it can be viewed on web browser (best supported by IE).This demo application will be used by organization's sales person who will demonstrate OBIEE's feature to prospect clients/business. This BI application has been designed using OBIEE and Oracle 10g and MS Excel as a data source.

References

- **Ali Ansari,** Senior Consultant, Ernst and Young, Gurgaon, +91 9999973024, ali.ansari@in.ey.com
- **Atul Singh,** Senior Consultant, Deloitte, Gurgaon, +91 991 1963 771, atsingh@deloitte.com
- **Dr. W. Jeberson,** Head, Dept. of Computer Science and Information Technology, SHIATS, Allahabad,+91 9452248375, jeberson@rediffmail.com

-Andalib Shadani Ansari