Manish Khobragade

manieshkhobragade@gmail.com

+91-7972043775

Professional Experience

• Over 7.5 years of experience in the field of Data Engineering, have great understanding on Hadoop Ecosystem, Spark, Cloudera, Cloud as AWS , AWS technologies such as S3 storage, Redshift Warehouse, Athena, Teradata etc.

• **4+ years** of working experience in Big Data Ecosystem.

• Started career in **Teradata Inc.** in 2012, now working **HP Inc.** from 2018

• Experienced in Big Data Hadoop Ecosystem tools **like HDFS, Sqoop, Hive, Spark 2.0, AWS EMR, Redshift, Amazon S3, Amazon Athena, Databricks and programming in SQL, Scala, Java, and Python.**

• Expertise in designing ETL flow with batch jobs on different Big Data and Warehousing projects.

• Developed Spark applications using Java/Scala/Python and pushed to central repositories using **GIT**.

• Well versed with **SCRUM** agile development methodologies.

• Have cross industry development experience such as Telecom, Retail etc.

• Contributor on **StackOverFlow**

(https://stackoverflow.com/users/9569498/manish**)**

• Experience on requirements gathering, documentation, Business/Functional Analysis

• A Team Player with excellent communication, analytical, verbal and writing skills.

• Proficient in grasping new technical concepts quickly and utilizing the same in a productive manner.

Skills Summary

Languages Known : Scala, Java, Python, SQL Platform : Apache Hadoop (Cloudera), Unix/Linux/Windows Framework : Spark Framework, AWS, Databricks-cloud Tools : Eclipse, Intellij, Databricks-Notebook, Git Hadoop Ecosystem : Spark, Hive, HDFS, AWS EMR, AWS S3, AWS Athena,

AWS Redshift SQL Solutions : Hive, Redshift, Teradata, Oracle, MySQL No SQL Solutions : Dynamo DB File Formats : Csv, Parquet, Text, Avro. Agile Framework : Microsoft VSO for SCRUM and agile development and GIT for version control

Project # 1

Duration: June 2019 to Till
Now

Company **: T-Systems** Project : Halo (Event Based Business
Efficiency) Domain : Telecommunications Platform :
Hortonworks Technologies : PySpark, JupyterHub, Hue

**Project
Overview:**

Halo is a analytics platform which is a hub for analytics application which are developed as asset and all other teams that require these application they can just download and deploy the code, saving massive redundant development time. Event Based Business Efficiency project identifies likely churner based occurrence of various events.

**Work and
Responsibilities:**

- • Development of Spark Jobs on Hortonwork in Python and maintaining output tables in Hive.
- • Maintaining the code in Git Repository.
- • Involved in testing of code in Dev and Stage environment before sending to Production.
- • Supporting the deployed code in Production with Dev-Ops.

Project # 2

Duration: March 2018 to May 2019

Company **: Hewlett Packard – R&D Department.** Project : DaaS
(Device as a service) Domain : Device Management Platform :
AWS EMR, Databricks Cloud Technologies : Spark, Hive, Athena,
Redshift, Databricks, S3

**Project
Overview:**

Being part of HP's R&D center in Pune, I am involved with one of service of HP, Device as a service popularly known as DaaS in market. DaaS as used to Optimize IT assets and resources with HP Device as a Service (DaaS), a complete solution that combines hardware, insightful analytics, proactive management, and services for every stage of the device lifecycle.

Work and

Responsibilities:

• Involved in designing project architecture, decision making on tools which are cost
effective and best suited for project and creation of technical specifications.
• Development of Spark Jobs on AWS EMR and Databricks Cloud in Java/Scala/Python and
maintaining output tables in Redshift/DyanmoDB.
• Maintaining the code in Git Repository.
• Involved in testing of code in Stable and Stage environment before sending to Production.
• Supporting the deployed code in Production with Dev-Ops.

Project # 3

Duration: April 2017 to Feb 2018

Company **: Teradata Incorporated.** Project :
Nordstrom US Domain : Retail Platform :
Cloudera Hadoop Technologies : Spark, Hive,
Sqoop, Git

**Project
Overview:**

Nordstrom Inc. is an American chain of luxury department stores, also operating in Canada and
headquartered in Seattle, Washington. Nordstrom wanted to provide a customized personal
experience with the vast amounts of data they collect. Apart from the data from their website or
the Point-of-Sales data, they generate lots of data from likes on Facebook, followers on Twitter.
In addition they generate vast amounts of data from their Fashion Rewards Program as
customers that want to enjoy the large amounts of benefits provided in this program will have to
use a Nordstrom credit card that tracks shopper spending and reward points.

**Work and
Responsibilities:**

• Involved in designing project architecture, decision making on tools which are cost
effective and best suited for project and creation of technical specifications.
• Development of Spark Jobs for specific reporting.
• Created a framework for capturing the receiving flat files and processing those files using
Hive.
• Created Hive tables to store the raw, staging and core data in tabular format.
• Written SQOOP jobs for migrating data from Teradata and MySQL databases into HDFS.
• Analyzed data in HDFS with tools such SPARK using Scala and Hive tables.
• Achieved performance gains in existing scripts by tweaking and analyzing existing jobs.

• Supporting the deployed code in Production.

Project # 4

Duration: October 2015 to March 2017

Company **: Teradata Incorporated.** Project :
Vodafone Australia Domain : Telecom
Platform : MapR Hadoop Technologies :
Spark, Hive, Sqoop, Git

**Project
Overview:**

Vodafone Australia is a mobile telecommunications company and Internet service provider that
operates the Vodafone brand in Australia. Vodafone Australia wanted to gain insight into their
customers and maintain analysis through end-to-end platform that offers answers to business
questions through the use of Big Data and telecom data analysis based on customer behavior.

Work and Responsibilities:

• Involved in designing project architecture, decision making on tools which are cost
effective and best suited for project and creation of technical specifications.
• Development of Spark Jobs for specific reporting.
• Created a framework for capturing the receiving flat files and processing those files using
Hive.
• Created Hive tables to store the raw, staging and core data in tabular format.
• Written SQOOP jobs for migrating data from Teradata and MySQL databases into HDFS.
• Analyzed data in HDFS with tools such SPARK using Scala and Hive tables.
• Achieved performance gains in existing scripts by tweaking and analyzing existing jobs.
• Supporting the deployed code in Production.

Project # 5

Duration: September 2014 to September 2015

Company : **Teradata Incorporated.** Project :

Volvo Sweden Domain : Manufacturing
Platform : Teradata Technologies : SQL,
Informatica

**Project
Overview:**
The Volvo Group is a Swedish multinational manufacturing company. They had a datawarehouse with lot of datamarts which were used for reporting. **Work and Responsibilities:**

• Wrote SQL stored procedures and triggers for implementing business rules and transformations.

• Developed transformation logic as per the requirement, created mappings and loaded data into respective targets.

• Created Source and Target Definitions in the repository using Informatica Designer – Source Analyzer and Warehouse Designer.

• Worked extensively on different types of transformations like Source qualifier, Expression, Filter

• Aggregator, Rank, Lookup, Stored procedure, Sequence generator.

• Used Mapping Designer to create mappings.

• Replicated operational tables into staging tables, to transform and load data into the enterprise data warehouse using Informatica.

Project # 6

Duration: May 2013 to August 2014

Company : **Teradata Incorporated.** Project : RBI
(Raiffeisen Bank International) Domain : Banking
Platform : Teradata Technologies : SQL, Informatica

**Project
Overview:**

Raiffeisen Bank International A.G. **(RBI) is an Austrian banking group,** RBI had their own warehouse to store new and different reporting projects. The data required for these project was produced by our team.

**Work and
Responsibilities:**

• Wrote SQL stored procedures and triggers for implementing business rules and transformations.

• Developed transformation logic as per the requirement, created mappings and loaded data into respective targets.

• Created Source and Target Definitions in the repository using Informatica Designer – Source Analyzer and Warehouse Designer.

• Worked extensively on different types of transformations like Source qualifier, Expression, Filter

• Aggregator, Rank, Lookup, Stored procedure, Sequence generator.

• Used Mapping Designer to create mappings.

• Replicated operational tables into staging tables, to transform and load data into the enterprise data warehouse using Informatica.