

# SACHIN SHARMA

(682)-256-5153 | sachin.sharma@tcu.edu

## Principal Data Engineer, MS Azure and Oracle Certified Professional

Results-driven oracle certified professional with 9 years of experience specializing in **data engineering, machine learning and technology consulting** within finance domain. I can implement effective **IT strategies** at local and global levels. My greatest strength is **global business experience** and awareness, which enables me to streamline infrastructure and applications. Striving to leverage my technical & management skills at upcoming opportunity.

### Expertise

- Big Data Queries and Interpretation
- Machine Learning Algorithms
- Forecasting and Data Modelling
- Research, Reports and Forecasts
- Data Mining and Visualization Tools
- Business Intelligence
- KPI Dashboards
- Team Management
- Problem Solver
- Servant Leader
- Cross-functional Communication

### Technical Skills

|                                    |  |
|------------------------------------|--|
| <b>Programming</b>                 | Scala, Python, PySpark, Databricks, AWS, PL/SQL, T-SQL, SQL  |
| <b>Databases</b>                   | <u>RDBMS</u> : Oracle, PostgreSQL, Couchbase, SQL Server, IBM DB2, MS SQL, MYSQL, Oracle 12c,11g, 10g, 9i, SQL Server Management Studio<br><u>NoSQL Database</u> : MongoDB<br><u>Cloud Storage</u> : Amazon EC2, Amazon S3, Microsoft Azure, Snowflake |
| <b>Machine Learning Algorithms</b> | <u>Classification</u> : Logistic Regression, Support vector machines (SVMs), Decision Tree, Random Forest<br><u>Regression</u> : Linear Regression, Multiple Linear Regression<br><u>Clustering</u> : K-Means Algorithm                                |
| <b>Tools/IDEs</b>                  | PyCharm, Eclipse, Spring tool suite (STS), Gradle, SQL*Plus, Toad, SQL Loader, SSAS, SSIS, SQL Navigator, Visual Studio, Toad, Putty, Jira, Exadata, SQL Developer, Explain Plan, TKPROF, Microsoft Excel, Microsoft PowerPoint                        |
| <b>Big Data Tools</b>              | Hadoop, MapReduce, Spark   |
| <b>Data Visualization</b>          | Tableau, Microsoft Excel, VBA, R   |
| <b>Other</b>                       | Deep Mining, Text Mining, Google Analytics, Mailchimp, AS400   |
| <b>Operating Systems</b>           | Linux (Ubuntu/Fedora/Kali), Windows, Mac   |

### Professional Experience

**NativeML**, Minneapolis, MN

Jul 2018 – To present

*NativeML is a global consulting firm for cloud-native data products and machine learning operations.*

#### Senior Data Engineer, Data Science and Engineering Division

- Worked closely with CTO and senior solution architect in developing and implementing data ingestion pipelines using tools and frameworks in Databricks and Hadoop ecosystem.
- Directed on-prem to cloud data migration efforts for clients. Worked with Attunity, MS Azure, Blob Storage, Databricks, Snowflake and raw data RDBMS storage platforms.

- Successfully implemented data staging in cloud environment for analysis. Utilized logistics regression, k-means clustering machine learning algorithms to draw intelligence from staged data.
- Developed automation and data collection frameworks.
- Highly experienced in mapping business needs to engineering systems.

**Environment: Scala, Spark, Amazon S3, EC2, Snowflake, PySpark, AWS, RedShift, S3, EC2, Machine learning, Oracle, SQL, PLSQL, IBM DB2.**

**U.S. BANK**, Minneapolis, MN

Apr 2016 – Jun 2018

*U.S. Bank, 7<sup>th</sup> largest bank in the USA, provides banking, investment, mortgage, trust, and payment services products.*

**Sr. Data Engineer (Contractor), Management Reporting Group**

Project summary: To identify and predict who can be the defaulters of the credit loan based on the applications received. Automate Basel I, II and III calculations as well as credit risk modelling.

To reach this goal, we used our previous customer's information such as their demographics, work pattern, residence status, credit to debt ratio, and previously defaulted or not, etc. We've also processed big datasets to calculate risk drivers such as PD, LGD and EAD and to build robust credit risk model for BASEL calculations.

Responsibilities:

- Led team of 3 contractors and worked with credit risk management and wealth management teams to gather the requirements while uncovering and defining multiple dimensions.
- Utilized pandas, numpy, sklearn and matplotlib.pyplot to process risk rating data and develop machine learning model. Also used sklearn preprocessing library to encode categorical rating data.
- Used logistic regression to obtain the probabilities of non-defaulters and defaulters.
- Developed segmentation tress (optimization, pruning, modelling) to find out high risk segment of population.
- Involved in writing data migration scripts/create table scripts to load data successfully from old schema to new schema.
- Directed daily ledger posting project. Developed package and assisted team members in writing logic for posting transactions in balance sheet.
- Involved in development and maintenance of Oracle database using PLSQL and backend development using Python and Core Java.
- Extensively worked on PLSQL stored procedures, functions, materialized views, Python scripts and cursors.
- Executed what if scenario analysis to discover effective, implementable ways of reducing loan defaults.
- Performed k-means clustering to understand customer backgrounds and segment the customers based on the customer transaction behavior information for customized product offering.

**Environment: Python, MS Azure, Snowflake, Hadoop, Decision tree, REST API, Git, Jira, PLSQL, SQL, Oracle 12c, SQL Developer, Toad.**

**Novartis - Alcon Division**, Fort Worth, TX

Jul 2014 – Mar 2016

*Alcon is a division of Novartis and the global leader in the pharmaceutical industry.*

**Lead Date Engineer, Enterprise Analytics Division**

- Re-designed and developed a critical ingestion pipeline to process over 300 TB of data.
- Developed Spark programs using Python API's to compare the performance of Spark with Hive and SQL.
- Used Spark-SQL to load JSON data and create schema RDD. Loaded it into Hive Tables and handled structured data using SparkSQL.
- Developed and implemented production unit stage tool (Java, Maven, Spring MVC, Git, SQL).
- Led team of 4 members and worked with the business to gather the requirement and identify the application development timeframe.
- Perform analysis of large data sets using components from the Hadoop ecosystem
- Developed applications and database objects like views, triggers, functions, procedures, and packages.
- Created various python and java scripts for verification of the required functionalities and everyday changes after various process
- Worked on day to day operational support tasks, handled member inquiries, verified daily data refresh from different sources related to membership system and provide update to business.
- Created Records, Tables, Objects, Collections (*Nested Tables and VARRAYS*), Table Partitioning and Error Handling, Primary Key, Composite Key.
- Performed data modeling, data source analysis, data sourcing specifications, ETL design & specification.

**Environment:** Spark, Azure, Snowflake, PySpark, Python(*Tensorflow, Scikit-learn*), Tableau, Hadoop, NoSQL, MongoDB, Decision tree, Core Java, REST API, Git, Spring MVC, Maven.

**ACCENTURE SERVICES, LTD.**, Mumbai, India

Feb 2010 – Apr 2014

*Accenture is a global management consulting, technology services and outsourcing company.*

**Lead Database Developer, FinTech Division (Client: Lufthansa Airlines, Germany)**

- Based at the Lufthansa headquarters in Germany, responsible for Accenture's day to day performance in processing client's financial data. Identified problems, developed solutions and managed the coordination with Accenture's IT staff in India.
- Led backend development efforts and provided REST API's, for client's internal financial tool, to front end team.
- Independently develop Oracle database objects includes tables, views, materialized views, indexes, triggers, functions, procedures, packages, etc.
- Worked on optimizing procedures, functions and views using Explain Plan.
- Used various kind of complex SQL queries involving Correlated Sub queries, union, Inner and Outer Joins.
- Cooperated with BA, SME to collect and document database design requirements and do data modeling with DB architect using ERWIN.
- Assisted DBA for job scheduling, data loading, and performance tuning using SQL tuning/access advisor, hints, explain plan and SQL trace.
- Write complicated queries using a lot of aggregate functions, joins, analytical functions, subqueries, etc. to provide real-time data from Oracle DB for client and UI development supporting.
- Checking execution plan using explain plan together with SQL trace with TKPROF using to realize end-to-end application tracing.
- Add optimization hints into high-load and top SQL statements to change the optimization goal, access method, join method, join order, and parallelization, etc.

**Java/SQL Developer, FinTech Division (Client: Dixons Carphone, UK)**

- Implemented a web-based library of common technical issues. Reduced assistance requests and general inquiries by 43%.
- Analyzed financial transactions within the Oracle business environment; identified trends and enhancement opportunities. Presented recommendations to client.
- Job scheduling using Autosys. Cooperated with QA team for debugging and regression testing.
- Involved in development of APTOS tool, client's internal financial transactions posting application (Java, Big Data, Oracle, SQL, PLSQL).

**Environment:** Java, PLSQL, SQL, Couchbase, Oracle 11g, Toad, C, Tableau, Stored procedures, IBM DB2, Big Data, Git, Jira, Spring MVC, Maven.

---

**Education**

**University of St. Thomas**, St. Paul, Minneapolis

**M.S. in Data Science and Artificial Intelligence**

**UPTU University**, India

**Bachelor of Technology** (Major: Computer Science & Information Technology, Minor: Analytics)

---

**Additional Information**

**Certifications:**

- |            |  |
|------------|--|
| • Jul 2019 | SnowPro Certified Professional,                        |
| • Dec 2018 | AZ-103 Microsoft Azure Administrator,                  |
| • Mar 2017 | Lean Six-Sigma Green Belt,                             |
| • Feb 2016 | Oracle Database 11g Administrator Certified Associate. |

**Awards:**

- |            |  |
|------------|--|
| • Aug 2017 | Winner of Hackathon, Europe.                                     |
| • Dec 2015 | MBA - Full Tuition Scholarship. Perfect GMAT quantitative score. |
| • Dec 2014 | Accenture Employee of the year.                                  |