

# Tanisha Sinha

## Senior Big Data Engineer

+61- 405 431 065

@ tanishasinha21@gmail.com

<https://www.linkedin.com/in/tanishasinha21>

Sydney, Australia



## Summary

- **Senior Big Data Engineer** with 9+ years of extensive experience in **end to end design and development** across various phases, processes and concepts of **Big data** and **web-based applications**.
- 4+ years of experience as a developer using Big Data Technologies like **Databricks/Spark** and **Hadoop Ecosystems**.
- Experience on **Migrating SQL database to Azure data lake**, Azure SQL Database and Databricks and migrating on premise databases to Azure Data Lake store using **Azure Data Factory**.
- Hands on experience Databricks Workspace User Interface, Managing **Databricks Notebooks, Delta lake** with **Python** and **Spark SQL**.
- Good understanding of **Spark architecture** including Spark Core, Spark SQL, Data Frames, **Spark Streaming** along with Databricks.
- Have good knowledge of **Azure Synapse analytics** to ingest, analyse and export large amount of data into/from Azure Data Lake Storage.
- Setting up Azure with Databricks and managing the clusters in databricks.
- Experienced in leading, coordinating and mentoring onshore/offshore clients and team members to assign and distribute day to day tasks for short or long term deliverables.
- .Strong troubleshooting skills with ability to work quickly with an eye towards writing clean code that is efficient and reusable.
- Self-motivated, result oriented professional who can seamlessly integrate into any working environment.

## Experience Summary

### Senior Big Data Engineer

#### Dipole Consulting Group Pty Ltd(Optus)

June, 2021 – Ongoing Sydney, Australia

Singtel Optus Pty Limited is a leading Australian telecommunications company.

- Experience in developing Spark applications using Spark-SQL in Databricks for data extraction, transformation and aggregation from multiple file formats for Analysing and transforming the data to uncover insights into the customer usage patterns.
- Extract Load and Transform data from various types of sources to Azure Data Lake Storage services using Azure Data Factory and Spark-SQL.
- Data ingestion to Azure Data Lake Storage and processing the data in Azure Databricks.
- Create Pipelines in ADF using Linked Services/Datasets/Pipeline to Extract, Transform and load data from different sources like Azure SQL, Blob Storage and RDBMS(Oracle/Teradata)
- To meet specific business requirements wrote UDF's in Scala and Pyspark.

## Technical Skills

### Hadoop Eco system

HDFS, Spark, Hive, Kafka, Sqoop, Impala, Zookeeper

### Cloud

Cloudera, Azure

### Data Engineering Tools

Databricks, Pyspark, Azure Data Factory, Spark-Scala, Azure Synapse Analytics

### Relational & NoSQL Databases

Oracle, MySQL, MS SQL, Elastic Search, Azure SQL Database

### Web Technologies

SOAP, REST, HTML5, CSS, Spring Boot, Angular

### Operating Systems

Linux(RedHat), Windows, Mac

### Others

Control M, Informatica MDM

## Programming Languages

Java.....	●●●●●●●●●●
Python.....	●●●●●●●●●●
Scala.....	●●●●●●●●●●
Shell Scripting.....	●●●●●●●●●●
SQL.....	●●●●●●●●●●
HTLM5/CSS.....	●●●●●●●●●●
Type script.....	●●●●●●●●●●

- Designed and implemented scalable infrastructure and platform to offload data from S3 buckets and upload data files into S3 buckets.
- Worked on end to end ingestion of file ingestion to hadoop ecosystem from various sources, then transforming the data into different use cases and extracting and sending the data to the downstream for use cases like campaigning and other analysis.
- Experience working in CI/CD environment utilizing tools like Jira, Confluence and Jenkins.
- Performed various performance optimizations like using distributed cache for small datasets, Partition, Bucketing and broadcasting in hive and spark.
- Worked on design, code, configurations, and documentation for components that manage data ingestion, real time streaming, batch processing, data extraction, transformation, loading and scheduling across various platform.

## Technical Lead

### Wipro Limited

 January, 2019 – Ongoing  Sydney, Australia

Wipro Limited is an Indian multinational corporation that provides information technology, consulting and business process services.

- Designed and implemented scalable infrastructure and platform for large amounts of data ingestion, aggregation, integration and analytics in Hadoop.
- Developed Spark programs in Scala to perform data transformation, creating Data Frames and running Spark SQL and Spark Streaming application in Scala.
- Design, build and manage Elasticsearch cluster for search functionalities for the App for data usage.
- Loaded large sets of structured, semi-structured, and unstructured data with Kafka and Sqoop.
- Developed Kafka producer to publish real time data to Kafka and Kafka consumer to receive the data and store it in real time to Cloudera.
- Written sqoop scripts to import, export and update the data between HDFS/Hive and Relational databases.
- Worked on design, code, configurations, and documentation for components that manage data ingestion, real time streaming, batch processing, data extraction, transformation, loading and scheduling across various platform.
- Worked on Control-M for scheduling various applications in big data.

## Senior Software Developer

### Wipro Limited

 August 2013 – December 2018  Bangalore, India

Wipro Limited is an Indian multinational corporation that provides information technology, consulting and business process services.

- Worked on E2E solution and development of web based application using Angular and Spring boot.
- Developed REST APIs for integration with front end web applications.
- Implemented web services to migrate batch process to real time for better customer experience.
- Used Hibernate framework in persistent layer for mapping an object oriented domain model to database.
- Used Maven to build and deploy application to development environment.
- Developed SOAP web service integrating with Informatica Master Database Management using SIF APIs.
- Worked as Database Administrator and managed Oracle 12cR2. Handling of database and schema creation.
- Developed schema objects like functions, procedures, triggers, packages, etc.

## Certifications

- Databricks Associate Data Engineer
- Databricks Lakehouse Fundamentals
- Databricks Platform Administrator

## Industry Expertise

[Telecommunication](#) | [English](#) | [Hindi](#) |

## References

- Available on request