

Andrew Joseph

Big Data Engineer

✉ andrewjoseph04@gmail.com

☎ +91-9652851825

📍 Hyderabad, India

Professional Summary:

- 7 years of overall experience in the industry with 3 years of experience working on **Hadoop Ecosystem** in developing applications and the rest as a Data Support Engineer.
- Exposure in Designing and development of solutions for Big Data using Hadoop eco system technologies like **HDFS, Hive, Sqoop, HBase**.
- Experienced in designing and developing applications in **Spark** and its ecosystem: **Spark-core, Spark-SQL, and Spark Data Frames with Scala** using Cloudera distribution.
- Having experience in processing **Structured, Semi-Structured** and **complex data** using Spark.
- Knowledge on incremental imports, partitioning and bucketing concepts in HIVE and Spark SQL required for optimization.
- Hands on experience with Data ingestion tool like Sqoop.
- Worked extensively on Messaging queue tool **Kafka** and **No SQL** Database like **HBase**.
- Worked in Agile Methodologies.
- Good Knowledge of **ITIL** Standard process of operations.

Work Experience

Dell Technologies Aug 2017 – Aug 2022 5 Years

Big Data Engineer Nov 2019 – Aug 2022 2.9 Years

- Performing import and Export of data into HDFS and Hive using Sqoop and managed data within the environment streaming data into Kafka through Spark streaming.
- Collected JSON data from HTTP Source and developed code using Spark APIs that inserts and updates data into Hive tables.
- Wrote HIVE Queries for transformation of data to be used by Downstream models.
- Implemented performance tuning mechanisms such as partitioning, bucketing and indexing on various Hive tables to improve query performance.
- Wrote Spark applications for data validation, cleansing, transformation, custom aggregation.
- Worked on **Complex Data Processing** and Generation
- Loaded and transformed various file formats like JSON, XML, Avro, ORC and Parquet.
- Capable of processing large sets of Structured, Semi-structured and Streaming data
- Interactive analysis of Hive tables through various data frame operations using SparkSQL.
- Working in Agile methodology with Daily standup calls, issue tracking list and maintaining process status mails at the end of day.

ETL Support Aug 2017 - March 2019 2.3 Years

- Worked as an ETL Support Engineer monitoring and supporting production jobs for Dell global data warehouse ecosystem
- Worked with UNIX to validate files involved with our ETL loads, check the scripts used for file movements with in our ETL Server and even to Monitor the ETL loads. In some cases, used to update the existing shell scripts used for file creations or validations.
- Driving transition calls with Development teams ensuring they follow our standard internal practices and giving signoff on behalf of support team.
- Always available on call Support for Production job failures and analyse the issues on the fly and in most cases come up with an RCA or to provide an ETA for production fix ASAP.
- Engaging and Driving Major Incident Management calls (MIM) for P1 and P2 issues to ensure a quick turn around and resolution of business-critical issues.

Accenture

Data Analyst, Sep 2015 – July 2017 2 years

- Worked in Level 2 Production Support environment, enhanced with on call Support for any Production job failures by following ITIL standards.
- Fixing all production failures such as Duplicate Records, Broken Records, Null values in Not Null fields, Storage issues, Access issues and others
- Be the application owner, regularly communicating the status of the jobs & Platform services its related failure and/or any issues through appropriate communication channels.

Technical Skills:

Data Ecosystem	Spark, Hadoop, Sqoop, Hive, HBase
Schedulers	Control-M
ETL Tools	Streamsets, Informatica
RDBMS Databases	Greenplum, Teradata
No SQL Databases	HBase
Languages	Scala, SQL

Education Background:

Andhra University

M. Tech, Computer Science 2012-14

Malla Reddy Engineering College

B. Tech, Information Technology 2008 -12