**ANIL KUMAR PALWAI** 

**E-mail:** anil.palvai@gmail.com

 **Phone:** +65-84032906

www.linkedin.com/in/anilpalwai

|  |
| --- |
| **Summary:** |

• A Bigdata engineer with 7 years of experience in creating data lake, Data Platform, and Data Ingestion Pipelines.

• A Certified in AWS certified solutions architect Associate.

• A Certified Developer of Confluent certified developer of Apache Kafka.

• Experienced in building and design of scalable and highly available data platform for a leading bank in Singapore.

• Experienced in the end-to-end life cycle of big data applications i.e. design, development, deployment, and monitoring.

• Experienced in the development and migration of data ingestion pipelines using Spark with Scala and Hive.

• Experienced in realtime data streaming pipelines with Apache Kafka and Spark streaming.

• Trained and experienced in deploying and developing AWS services such as S3, VPC creation, EC2 instance launching with Autoscaling, and load balancing.

• Experienced in writing, scheduling, and debugging Airflow DAGs and custom operators.

• Experienced in working and integration of Alluxio with compute cluster and S3 data lake.

• Experienced in working a container engine i.e. Docker

• Experienced in implementing CI/CD best practices for the Data engineering platform using Jenkins, CARA, and JIRA tools.

• Experienced in design and visualizing the Dashboards using Grafana.

• Capturing data from existing databases that provide SQL interfaces using Sqoop.

• Experienced Agile scrum methodology to achieve the deliverables.

|  |
| --- |
| **Degree and Certifications:** |

• Experienced with both Git and SVN as Version Control System for application development.

* Master of Science in Data Science and Data visualization from JNTUH in collaboration with Carnegie Mellon University (CMU), Pittsburgh, USA.
* Confluent Certified Developer for Apache Kafka
* AWS certified solutions architect Associate.
* All in India 2500 Rank in GATE-2012 in Computer science Stream.

|  |
| --- |
| **Professional Skills** |

* Hadoop Eco System : Apache Spark, Hive, HBase, Sqoop, and MapReduce.
* File system : Hdfs, S3, and Alluxio.
* Streaming applications : Kafka and Spark Streaming.
* Public Cloud : AWS (S3, EC2, VPC, and Cloudformation)
* Container Engine : Docker
* CI/CD : Jenkins and CARA
* Scripting Languages : Linux
* Programming Languages : Java, Scala
* Version Control : Git, Svn
* Repositories : Bitbucket, Nexus.

|  |
| --- |
| Employment History: |

**August 2017 – Present: ADA-Data platform.**

|  |  |
| --- | --- |
| **Company** | Accenture AI, Singapore. |
| **Client** | Confidential (Leading bank in Asia) |
| **Environment** | Hadoop, Spark-Scala, Kafka, Hbase, Alluxio, Airflow, Graphna, Hive, Sqoop, Docker, Jenkins, CARA, JIRA, and Unix. |
| **Role** | Senior Bigdata Engineer |

**Roles and Responsibilities:**

* Involved in the design and implementation of highly scalable and highly available Data Lake using S3 as UFS
* Development and debugging of Data ingestion applications on the top of the Data platform.
* Integration and optimizing of Alluxio as virtual data layer between compute applications such as Spark with scala and S3 as the storage
* Implemented Airflow DAGs for integration and ingestion jobs.
* Implemented log aggregation on Airflow DAGS and monitored on Grafana dashboards
* Automation and integration activities to make applications selfheal at runtime.
* Implemented Alluxio wrappers to protect the data from accidental deletion.
* Development and debugging activities of Kafka streaming applications.
* Designed and developed deployment pipelines for applications on Data platform using Jenkins, JIRA, and CARA agents.
* Code review and CR activities for application development.
* Creation and launching of AWS services for research activities.
* Using an Agile Scrum methodology to achieve the deliverables.

**Feb 2017 – July 2017: Nielsen Retail Intelligence: Aggregation Platform**

|  |  |
| --- | --- |
| **Company** | KOGENTIX Inc, Hyderabad, India. |
| **Client** | Nielsen |
| **Environment** | Hadoop, Spark, MapReduce Hive, Sqoop, Java, Unix, and MySQL. |
| **Role** | Bigdata Engineer |

**Roles and Responsibilities:**

• Implementation of customer data analytics that could help improve the performance of a global Retail management company.

• Developed applications for aggregation engine of Retail Intelligence with Apache Spark and MapReduce.

• Using Agile scrum methodology to achieve the deliverables.

• Use Sqoop to transfer the data from the relational database to HDFS.

• Worked on complex Hive queries which include UDF and UDTF to compare aggregation engine data against traditional data.

• Developed applications that can handle Parquet format files.

• Extensively using Git as a Version Control System for application development.

**March 2014 – January 2017: IMS-Web of Science.**

|  |  |
| --- | --- |
| **Company** | Virtusa Corporation, Hyderabad, India. |
| **Client** | Thomson Reuters |
| **Environment** | Hadoop, Hive, Sqoop, Hbase, Java, Unix, and MySQL. |
| **Role** | Hadoop Developer |

**Roles and Responsibilities:**

• Implementation of the next generation of Web of Science application based upon the principles of Citation Connections.

• Extensively using the Hadoop stack for application development (Mapreduce, Spark, HDFS, Hive, Sqoop, Hbase).

• Worked on parsing XML to ingest the data into the Hadoop cluster.

• Worked on complex Hive queries to create data reports upon client requirements.

• Developed UDF and UDTF function in Hive as a part of generating country and institution reports.

• Written Shell Scripts to support and schedule cluster activities.

• Using the Agile Scrum methodology to achieve the deliverables.

• Using SVN as Version Control System for application development.

• Using JIRA to plan, track the development process.

• Using Sqoop to transfer the data from Hadoop Store to the relational database.

|  |
| --- |
| Personality Traits: |

* Excellent at Strategic planning and analytical skills.
* Good at visualizing End product.

|  |
| --- |
| Personal Details |

* Nationality: Indian
* Current Location: Singapore
* Visa type: Employment Pass