

## DATA ENGINEER

### Professional Summary

Energetic Data Engineer in developing robust code for high-volume businesses. Strong decision-maker with 6 years of experience in Data engineering and help firms in designing and executing solutions for complex business problems involving large scale data warehousing, real-time analytics and reporting solutions. Ability to translate business questions and concerns into specific quantitative questions that can be answered with available data using sound methodologies.

### Skills

- Python 3.x, R, SQL
- Hadoop, Apache Spark
- Hive, Pig, Kafka, Sqoop, Oozie
- Teradata, Snowflake
- Amazon S3, EMR, Lambda
- Git, Jenkins, Splunk
- MS Office
- Microsoft Visual C#.NET

### Work History

Data Engineer , 04/2020 to Current

Avanade – Bangor , ME

- Combining data from multiple source systems (Profile, Systematics, etc..) and multiple platforms (Snowflake, OneLake, Hubs) and computing canonical gold-star metrics –once and for all– to cut operational costs.
- Develop Spark jobs to transform data and apply business transformation rules to load/process data across enterprise and application specific layers.
- Experience in building/operating/maintaining fault tolerant and scalable data processing integrations using AWS.
- Configured S3 buckets with various life cycle policies to archive the infrequently accessed data based on requirement.
- Good working experience on submitting the Spark jobs which shows the metrics of the data which is used for Data Quality Checking.
- Working on building efficient data pipelines that transform high volume data into a format used for analytical, fraud prevention and ML use cases.
- Extensively used Splunk Search Processing Language (SPL) queries, Reports, Alerts and Dashboards.
- Excellent knowledge of source control management concepts such as Branching, Merging, Labeling/Tagging and Integration, with tool like Git.
- Performing Data Quality checks like row count, schema validation, Hash key validation for all data movement between applications.

Data Engineer , 08/2019 to 04/2020

Avanade – Burlington , NC

- Responsible for designing and developing various analytical solutions for gaining analytical insights into large data sets by ingesting and transforming these datasets in the Big Data environment using technologies like Spark, Sqoop, Oozie, HIVE.
- Scheduling jobs to automate the process for regular executing jobs worked on using Oozie.
- Developed Oozie workflow schedulers to run multiple Hive and Pig jobs that run independently with time and data availability.
- Familiar with data architecture including data ingestion pipeline design, Hadoop information architecture, data modeling and data mining, machine learning and advanced data processing.
- All the projects which I have worked for are Open Source Projects and has been tracked using JIRA.
- Parallel copying of files between various clusters using Distcp, Kafka in Hadoop.

Data Scientist Intern , 09/2018 to 12/2018

Ascend Learning – Memphis , TN

- Acquire, clean, integrate, analyze and interpret disparate datasets using a variety of statistical data analysis and data visualization methodologies, reporting and authoring findings where appropriate.
- Developed Linear Mixed Effects Models for Boston Ed-fi dataset to estimate the teacher contribution to the student test scores.
- Random intercept model which utilized a pre-post-test design and included fixed effects for student demographics and a growth model where students were nested by time (random slope and intercept) and teacher was tested.
- Generated various Clustering models for entire West Virginia department of education and evaluated cluster's performance.

Big Data Engineer , 01/2014 to 07/2017

Verizon

- Identify customer's digital analytical needs and engage with customer and principal architects daily, understand the business requirements for big data analytical solutions and break down large scale requirements into detailed system specifications.
- Moving data to HDFS framework using SQOOP from Teradata, SQL Server.
- Good working experience on Hadoop tools related to Data warehousing like Hive, Pig and also involved in extracting the data from these tools on to the cluster using Sqoop.
- Solved performance issues in Hive with understanding of joins, groups, bucketing, partitions and working on them using HiveQL.
- Deployed programs written in PySpark to run Spark MLlib, for analytics and reduced customers churn rate by 25%.

### Education

Master of Science : Business Analytics , 04/2020

The University Of Texas At Dallas - City , State

- <https://www.linkedin.com/in/keerthidanda>
- <http://www.github.com/keerthidanda>

## Skills

- Python 3.x, R, SQL
- Hadoop, Apache Spark
- Hive, Pig, Kafka, Sqoop, Oozie
- Teradata, Snowflake
- Amazon S3, EMR, Lambda
- Git, Jenkins, Splunk
- MS Office
- Microsoft Visual C#.NET

## Work History

Data Engineer , 04/2020 to Current

Company Name " Livingston , State

- Combining data from multiple source systems (Profile, Systematics, etc..) and multiple platforms (Snowflake, OneLake, Hubs) and computing canonical gold-star metrics once and for all to cut operational costs.
- Develop Spark jobs to transform data and apply business transformation rules to load/process data across enterprise and application specific layers.
- Experience in building/operating/maintaining fault tolerant and scalable data processing integrations using AWS.
- Configured S3 buckets with various life cycle policies to archive the infrequently accessed data based on requirement.
- Good working experience on submitting the Spark jobs which shows the metrics of the data which is used for Data Quality Checking.
- Working on building efficient data pipelines that transform high volume data into a format used for analytical, fraud prevention and ML use cases.
- Extensively used Splunk Search Processing Language (SPL) queries, Reports, Alerts and Dashboards.
- Excellent knowledge of source control management concepts such as Branching, Merging, Labeling/Tagging and Integration, with tool like Git.
- Performing Data Quality checks like row count, schema validation, Hash key validation for all data movement between applications.

Data Engineer , 08/2019 to 04/2020

Company Name " City , State

- Responsible for designing and developing various analytical solutions for gaining analytical insights into large data sets by ingesting and transforming these datasets in the Big Data environment using technologies like Spark, Sqoop, Oozie, HIVE.
- Scheduling jobs to automate the process for regular executing jobs worked on using Oozie.
- Developed Oozie workflow schedulers to run multiple Hive and Pig jobs that run independently with time and data availability.
- Familiar with data architecture including data ingestion pipeline design, Hadoop information architecture, data modeling and data mining, machine learning and advanced data processing.
- All the projects which I have worked for are Open Source Projects and has been tracked using JIRA.
- Parallel copying of files between various clusters using Distcp, Kafka in Hadoop.

Data Scientist Intern , 09/2018 to 12/2018

Company Name " City , State

- Acquire, clean, integrate, analyze and interpret disparate datasets using a variety of statistical data analysis and data visualization methodologies, reporting and authoring findings where appropriate.
- Developed Linear Mixed Effects Models for Boston Ed-fi dataset to estimate the teacher contribution to the student test scores.
- Random intercept model which utilized a pre-post-test design and included fixed effects for student demographics and a growth model where students were nested by time (random slope and intercept) and teacher was tested.
- Generated various Clustering models for entire West Virginia department of education and evaluated cluster's performance.

Big Data Engineer , 01/2014 to 07/2017

Company Name

- Identify customer's digital analytical needs and engage with customer and principal architects daily, understand the business requirements for big data analytical solutions and break down large scale requirements into detailed system specifications.
- Moving data to HDFS framework using SQOOP from Teradata, SQL Server.
- Good working experience on Hadoop tools related to Data warehousing like Hive, Pig and also involved in extracting the data from these tools on to the cluster using Sqoop.
- Solved performance issues in Hive with understanding of joins, groups, bucketing, partitions and working on them using HiveQL.
- Deployed programs written in PySpark to run Spark MLlib, for analytics and reduced customers churn rate by 25%.