

PROFESSIONAL SUMMARY:

- Around 7 years of experience as a **Data Engineer** with expertise in the Hadoop Ecosystem, including Azure and Snowflake.
- Extensive experience deploying cloud-based applications using Amazon Web Services such as Amazon EC2, S3, RDS, IAM, Auto Scaling, Cloud Watch, SNS, Athena, Glue, Kinesis, Lambda, EMR, Redshift, and Dynamo DB.
- Worked on ETL Migration services by developing and deploying AWS Lambda functions for generating a server less data pipeline which can be written to Glue Catalog and can be queried from Athena.
- Proven expertise in deploying major software solutions for various high-end clients meeting the business requirements such as big data Processing, Ingestion, Analytics and Cloud Migration from On-prem to AWS Cloud using AWS EMR, S3, Dynamo DB
- Hands-on experience with various Cloud Services GCP, AWS (EC2, S3, RDS, ECS, Elastic Load Balancer, EBS, RDS, IAM, Route 53, Auto Scaling Groups)
- Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like Azure SQL, Blob storage, Azure SQL **Data warehouse**, write-back tool and backwards.
- Experience with migrating SQL databases to Azure Data Lake, Azure Data Lake Analytics, Azure SQL Database, Data Bricks, and Azure SQL **Data warehouse**, as well as controlling and giving database access and migrating on-premises databases to Azure Data Lake stores using Azure Data Factory.
- Developed ETL pipelines in and out of **data warehouse** using combination of Python and Snowflakes SnowSQL Writing SQL queries against Snowflake.
- Demonstrated understanding of the Fact/Dimension **data warehouse** design model, including star and snowflake design methods.
- Experienced in building Snow pipe and In-depth knowledge of Data Sharing in Snowflake and Snowflake Database, Schema and Table structures.
- Designed and developed logical and physical data models that utilize concepts such as Star Schema, Snowflake Schema and Slowly Changing Dimensions.
- Hands on experience across Hadoop Ecosystem that includes extensive experience in Big Data technologies like HDFS, Map Reduce, YARN, Apache Cassandra, HBase, Hive, Oozie, Impala, Pig, Zookeeper and Flume, Kafka, Sqoop, Spark.
- Built real time data pipelines by developing Kafka producers and Spark streaming applications for consuming.
- Experienced with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data frame API, Spark Streaming, Pair RDD's and worked explicitly on PySpark.
- Familiar with data processing performance optimization techniques such as dynamic partitioning, bucketing, file compression, and cache management in Hive, Impala, and Spark.
- Excellent understanding and knowledge in handling database issues and connections with SQL and NOSQL databases like Mongo DB, Cassandra, HBase and SQL server.
- Experience in Dimension Data modeling concepts like Star Join Schema Modeling, Snow-Flake Modeling, FACT and Dimensions Tables, Physical and Logical Data Modeling.
- Created Partitions and Bucketing Hive tables in Parquet File Formats with Snappy compression and then loaded data into Parquet hive tables from Avro hive tables.
- Experience in designing and implementing RDBMS Tables, Views, User Generated Data Types, Indexes, Stored Procedures, Cursors, Triggers, and Transactions.
- Experience in creating and maintaining reporting and analytics infrastructure for internal business clients utilizing AWS services such as Athena, Redshift, Spectrum, EMR, and Quick Sight.

Skills:

Big Data Technologies: HDFS, Yarn, Map Reduce, Pig, Hive, HBase, Cassandra, Zookeeper, Oozie, Sqoop, Flume, H Catalog, Apache Spark, Scala, Impala, Kafka, Storm, Tez, Ganglia, Nagios, Splunk, Elastic Search, Kibana

Hadoop Distributions: Apache Hadoop 2.x/1.x, Cloudera CDP, Hortonworks HDP, Amazon EMR

Visualization Tools: JMP, Microsoft Excel, SAS text miner, Tableau, Microsoft Visio, Microsoft Project, Power BI

Programming Skills: Python (NumPy, Pandas, Matplotlib), R, Java

Analytical Tools: Python, R, SAS JMP, Microsoft Excel, SAS

Database and cloud skills: Oracle SQL, Microsoft SQL, GCP, Big Query, HDFS, Hive, Map-reduce

PROFESSIONAL EXPERIENCE

Azure Data Engineer

Sept 22 to till Date

First Republic Bank, New York, NY

Responsibilities:

- Contributed to the development of PySpark Data Frames in Azure Data bricks to read data from Data Lake or Blob storage and utilize Spark Sql context for transformation.
- Experience in Creating, developing, and deploying high-performance ETL pipelines with PySpark and Azure Data Factory.
- Responsible for estimating the cluster size, monitoring, and troubleshooting of the Spark data bricks cluster.
- worked on an Azure copy to load data from an on-premises SQL server to an Azure SQL Data warehouse.
- Worked on redesigning the existing architecture and implementing it on Azure SQL.
- Experience with Azure SQL database configuration and tuning automation, vulnerability assessment, auditing, and threat detection.
- Integration of data storage solutions in spark – especially with Azure Data Lake storage and Blob snowflake storage.
- Managed and monitored the Services which are project related in GCP.
- Deployed code and maintained services and closely worked with DEV teams in GCP (Google Cloud Platform).
- Developed ETL pipelines in and out of data warehouse using a combination of Python and Snowflakes SnowSQL Writing SQL queries against Snowflake.
- Involved in Migrating Objects from Teradata to Snowflake and created Snow pipe for continuous data load.
- Improving the performance of Hive and Spark tasks.
- Knowledge with Kimball data modeling and dimensional modeling techniques.
- Worked on cloud point to identify the best cloud vendor based on a set of strict success criteria.
- Used Hive queries to analyze huge data sets of structured, unstructured, and semi-structured data.
- Created Hive scripts from Teradata SQL scripts for data processing on Hadoop.
- Developed Hive tables to hold processed findings, as well as Hive scripts to convert and aggregate heterogeneous data.
- Created and utilized sophisticated data types for storing and retrieving data in Hive using HQL.
- Used structured data in Hive to enhance performance using sophisticated techniques including as bucketing, partitioning, and optimizing self joins.
- Created a series of technology demos utilizing the Confidential Edison Arduino shield, Azure Event Hub, and Stream Analytics, to show the possibilities of Azure Stream Analytics.

Environment: Apache Big Data, Spark (Python/Scala), Hive, GCP, Kafka, Spark Streaming, Docker Containers, PostgreSQL, Rabbit MQ, Celery, Flask, ELK Stack, MS-Azure, Azure SQL Database, Azure functions Apps, Azure Data Factory(V2), Azure Data bricks, Python, SSIS

Data Engineer

June 19 – Dec 21

Mindtree, Hyderabad, India

Responsibilities:

- Architect and implement ETL and data movement solutions using Azure Data Factory, SSIS
- Understand Business requirements, analysis and translate into Application and operational requirements.
- Designed a one-time load strategy for moving large databases to Azure SQL DWH.
- Extract Transform and Load data from Sources Systems to Azure Data Storage services using Azure Data Factory
- and HDInsight.
- Created a framework to do data profiling, cleansing, automatic restart ability of batch pipeline and handling rollback strategy.
- Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL
- Lead a team of six developers to migrate the application.
- Implemented masking and encryption techniques to protect sensitive data.

- Implemented SSIS IR to run SSIS packages from ADF.
- Developed mapping document to map columns from source to target.
- Created azure data factory (ADF pipelines) using Azure blob.
- Performed ETL using Azure Data Bricks. Migrated on-premises Oracle ETL process to Azure Synapse Analytics.
- Worked on python scripting to automate generation of scripts. Data curation done using azure data bricks.
- Worked on azure data bricks, PySpark, HDInsight, Azure ADW and hive used to load and transform data.
- Implemented and Developing Hive Bucketing and Partitioning.
- Implemented Kafka, spark structured streaming for real time data ingestion.
- Used Azure Data Lake as Source and pulled data using Azure blob.
- Used stored procedure, lookup, execute pipeline, data flow, copy data, azure function features in ADF.
- Worked on creating a star schema for drilling data. Created PySpark procedures, functions, packages to load data.
- Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics.
- Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Data bricks.
- Responsible for estimating the cluster size, monitoring, and troubleshooting of the Spark data bricks cluster.
- Creating Data bricks notebooks using SQL, Python and automated notebooks using jobs.
- Creating Spark clusters and configuring high concurrency clusters using Azure Databricks to speed up the preparation of high-quality data.
- Create and maintain optimal data pipeline architecture in cloud Microsoft Azure using Data Factory and Azure Data bricks.

Environment: Hadoop, Hive, Impala, Oracle, Spark, Pig, Sqoop, Oozie, Map Reduce, Teradata, SQL, Abinitio, (S3,RedShift, CFT, EMR, CloudWatch), Kafka, Zookeeper, PySpark.

Cyient, Hyderabad, India

Jan 16 – May 19

Data Analyst

Responsibilities:

- Developed stored procedures in MS SQL to fetch the data from different servers using FTP and processed these files to update the tables.
- Responsible for Designing Logical and Physical data modeling for various data sources on Confidential Redshift.
- Performed logical data modeling, physical Data modeling (including reverse engineering) using the Erwin Data modeling tool.
- Created dimensional model for the reporting system by identifying required dimensions and facts using Erwin.
- Designed and Developed ETL jobs to extract data from Salesforce replica and load it in data mart in Redshift.
- Involved in performance tuning, stored procedures, views, triggers, cursors, pivot, unpivot functions, CTE's
- Developed and delivered dynamic reporting solutions using SSRS.
- Extensively used Erwin for Data modeling. Created Staging and Target Models for the Enterprise **Data Warehouse**.
- Involved in Normalization / De normalization techniques for optimum performance in relational and dimensional database environments.
- Resolved the data type inconsistencies between the source systems and the target system using the Mapping Documents and analyzing the database using SQL queries.
- Worked on ETL testing, and used SSIS tester automated tool for unit and integration testing.
- Designed and created SSIS/ETL framework from ground up.
- Created new Tables, Sequences, Views, Procedure, Cursors and Triggers for database development.
- Created ETL Pipeline using Spark and Hive for ingest data from multiple sources.
- Involved in using SAP and transactions done in SAP - SD Module for handling customers of the client and generating the sales reports.
- Creating reports using SQL Reporting Services (SSRS) for customized and ad-hoc Queries
- Coordinated with clients directly to get data from different databases.
- Worked on MS SQL Server, including SSRS, SSIS, and T-SQL.
- Designed and developed schema data models.

- Documented business workflows for stakeholder review.

Environment: ER Studio, SQL Server 2008, SSIS, Oracle, Business Objects XI, Rational Rose, Data stage, MS Visio, SQL, Crystal Reports 9

Vedhasree Kakkunuri

- Santa Clara, CA, US

Contact Information

- jph-rlb-8ea@mail.dice.com
- 4086419367

Summary

I strongly believe I am an efficient **data engineer** with over 7 years of experience and a...

Work History

Total Work Experience: 8 years

- **Azure Data Engineer** First Republic
Sep 01, 2022
- **Data Engineer** Mindtree
Jun 01, 2019
- **Data Analyst** Cyient
Jan 01, 2016

Education

- **Masters** | Western Illinois University

Skills

- **apache hive** - 7 years
- **apache spark** - 7 years

- **data warehouse** - 7 years
- **etl** - 7 years
- **microsoft ssis** - 7 years
- **sql** - 7 years
- **data modeling** - 4 years
- **dimensional modeling** - 4 years
- **microsoft sql server** - 4 years
- **microsoft windows azure** - 3 years
- **amazon redshift** - 6 years
- **database** - 6 years
- **framework** - 6 years
- **mapping** - 6 years
- **oracle** - 6 years
- **procedure** - 6 years
- **qa** - 6 years
- **stored procedures** - 6 years
- **transact-sql** - 6 years
- **cloudera impala** - 3 years

Work Preferences

- Desired Work Settings: Remote or On-Site or Hybrid
- Likely to Switch: False
- Willing to Relocate: False
- Travel Preference: 0%
- Work Authorization:
 - US
- Work Documents:
 - Employment Auth. Document
- Security Clearance: False
- Third Party: False
- Employment Type:
 - Contract - Corp-to-Corp
 - Contract - Independent
 - Contract - W2
 - Contract to Hire - Corp-to-Corp
 - Contract to Hire - Independent
 - Contract to Hire - W2

Profile Sources

- linkedin: <https://www.linkedin.com/in/vedhasree-kakkunuri-181489147>
- Dice:
<https://www.dice.com/employer/talent/profile/0c988e09ab52f28ea11796ddf7d94999>