Manvendra Singh Khati

Cloud Operations Analyst/Big data Developer
Email: manvendrakhati@gmail.com

Mobile: 91- 9627380472 / 7219084472

<u>Professional Summary:</u>

- 6 years of experience in application development, maintenance and support in IT.

- 2+ years of experience in Big data technology. 1 Year of experience in AWS and QlikView reporting.

- Currently working on Cloud Optimization Services to reduce the cost of cloud systems in AWS and Azure by generating recommendations.

- Experience in Hadoop Technology stack which includes ecosystems like Hive, Sqoop, SPARK, KAFKA.

- 3 Years Working experience in ETL tools like Infomatica PowerCenter.

- Knowledge of integrating KAFKA with SPARK streaming application for near real time data ingestion.

- Used CloudHealth, CloudApptio, CloudYn tools to analyze cloud spend analysis,forecasting,Trend Analysis etc.

- Worked with various data formats like csv, json, orc, parquet and compressions like GZIP, SNAPPY. Used generic as well as SQOOP metastore logics to ingest data from SQOOP.

- Worked on ServiceNow Integration witht the existing system programatically.

- Exposure and experience on different RDBMS like Oracle 11g,Mysql.

- Experience in capturing Hadoop/spark application logs and failure alerts for production running jobs.

- Knowledge of ML algorithms like Random Forest Classifier and xgboost used in project. Knowledge on working in Advanced Excel.

- Knowledge on implementation of Data Warehouse and DataLake systems.

- Worked in Banking and Financial Services projects.

Technical Skills:

Big data Technologies	SPARK, Hadoop (HortonWorks, Pivotal), Hive, Pig, Sqoop, Kafka,
Programming Languages	Python, Core Java
ETL Tools	Infomatica Power Center
CD tool	Bamboo,Perforce
Build Tools/IDE	Eclipse
Databases	Oracle,Hbase
Cloud	AWS,Azure(Baiscs),Cloudyn,CloudAppitio etc.
Reporting	QlikView

Accenture

Projects Profile:

Project #1

Role   :   Cloud Operations Analyst
Duration   :   May 2020 – Present

Currently working in Internal Project to reduce the cost of the Cloud System in AWS and Azure by recommending optimizations for instances, storage, migration etc.

- Clients are of two types BYOC(Bring Your Own Cloud) and Prime Clients which use ACP provided serivces for discount
- Billing data, Utilization data and Storage Metrics is extracted using PowerShell and Python scripts.
- Cost Allocation Reports showing Amortized cost, Rightsizing Recommendation,Sizing Optimization ,PowerScheduling Recommendations etc have been created in Qikview to transform the data and reduce the cloud spent.
- Innovation of new features in cloud to enhance optimal services use etc.

Project #2

Role   :   Big data Developer
Duration   :   Jan 2019 – April 2020

Client is one of the largest bank in USA. We maintain a Hadoop data lake containing data from different source systems. Our data lake provides information about the Customer financial transactions, Card Payment,account related information fraud detection

- Data is ingested through various sources like SAS, Genays through SFTP ,Oracle and other upstreams through Kafka.
- Sqoop jobs are created to ingest data from Oracle. For PII data, Pig is used for processing into structured format.
- Hive and Spark are used for processing the data in hdfs.
- Protegrity is used for encryption and tokenization of the data.
- Kafka jobs loads customer and IVR data to Hbase and is monitored through Grafanna
- Once the data is processed it is loaded to either Hbase or Gemifre for fast query fetching by different API's.
- API's are deployed in PCF(Pivotal Cloud Foundry) and New Relic is used to monitor them.
- Splunk is used to Monitor logs for hbase and Gemfire issues.
- AutoSys is used as for scheduling jobs and datacopy scripts from PROD to DR are scheduled in Crontab.

Utilities created:

Below are few utilities created for helping team and client maintaining healthy and validated data lake –

- For loading incremental/historcial data from various databases like CDCI ,RAMS into protected folders sqoop jobs are created. Once data is loaded into hdfs Pig used todetokenize it using protegrity functions. Hive and Spark are used for processing the data and use windowing and dedup functions to store the latest records.

- For POS(Point Of Sale) Authorization files are loaded in local directory through upstream and then copied to protected folder in hdfs. Spark job is used to load the data in hbase(phoenix is used on top of hbase).

- SYFid ,client id for specyfing all the customers uniquely and includes insertion updation deletion operation.

- Kafka jobs are both running 24*7 and scheduled in AutoSys loads the data in hbase , offset is also stored in hbase.

- FIDE (Fraud Ingestion and Decision Engine) is created using the transactional data and data from SAS and creating decision tree on top of that data to identify fraud transactions.

Responsibilities:

- Involved in analyzing system and business requirements.

- Worked in creation of sqoop jobs for loading the data int hdfs from Oracle.

- Created Hive and Spark Jobs to load maintain the daily tasks.

- Involved in Integration among Pig,Hive and Hbase.

- Understanding of Kafka Jobs and ML algorithms.

- Monitored the jobs once deployed in Production.

Infosys

Project #3

Role   :   Big data Developer
Duration   :   From Jan 2018 to Dec 2018

Client is one of the largest bank in USA. The objective is to maintain a datalake containing data from different source systems.

Technologies involved:

- SQOOP and SFTP from various upstream for data ingestion.

- Hive for processing the data in hdfs.

- Bamboo integration with bitbucket for continuous delivery.

- Monitored the jobs once deployed in Production.

Responsibilities:

- Involved in analyzing system and business requirements.

- Worked in creation of sqoop jobs for loading the data int hdfs from Oracle.

- Created Hive Jobs to load maintain the daily tasks.

Project #4

Role   :   Informatica Senior Developer
Duration   :   From March 2017 to Dec 2018

- The objective is to migrate the account existing and new accounts in PCS into another application CRD

- Real Time and Batch Pipeline were in the architecture and Informatica was used to build different models such as Asset Allocation,Risk Relationship and UMA etc.

- Informatica was used to create views and a Hierarchical Structure was maintained in models to maintain allocation of assets.

- Unix was used to unzip the files and perform validations on data. Once the data is processed and loaded in CRD.

- Archiving and cleanup is done. Scheduling of jobs is done through AutoSys.

-
Technologies involved:

- Informatica PowerCenter, Oracle, Unix, Flat Files, AutoSys

Responsibilities:

- Requirement gathering and involved in various phases of SDLC.

- Involved in creation of different models through Informatica.

- Development of Real Time and Batch Pipeline for Account Conversion and Onboarding.

- Involved in creation of scripts for archive and cleanup jobs.

- Involved in creation of jobs for Exception handling

- Involved in deployment of code through BitBucket.

Project #5

Role   :   PLSQL Developer

Duration   :   From Feb 2016 to Feb 2017

- Fixed Income Investment and Portfolio Data Management involves maintenance and enhancement of Fixed Income side of CGCs Transactions, Reference and Historical Databases.
- Creation of a Portfolio manager attribution system close interface with various internal applications and Third party file transfer handling.
- These applications are used for trade processing, portfolio accounting, managing both positions and holding and maintenance of history records for all of them

Technologies involved:

- Informatica PowerCenter, Oracle, Unix, Flat Files, AutoSys

Responsibilities:

- Requirement gathering and involved in various phases of SDLC.

- Creation of PLSQL scripts to load the data from staging table to final table using merge statements, recursive cte, cursors etc.

- Involved in creation of scripts for archive and cleanup jobs.

- Involved in creation of reusable scripts for automation process.

Project #6

Role   :   Informatica Developer.

Duration   :   From June 2015 to Jan 2016

- PPMIS is related to maintenance of Partner Payment for AMEX cards.

- Creation of Payment interface with various applications of Third Party file handling.

- These applications are used for payment calculation, new card information , maintenance of history records for all of them etc.

Technologies involved:

- Informatica PowerCenter,Informatica PowerExchange, Oracle, Unix, Flat Files, AutoSys

Responsibilities:

- Requirement gathering and involved in various phases of SDLC.

- Involved in extracting data from sources like Mainframe and Oracle.

- Used execution plan to load data from OLTP tables to corresponding OLAP tables.

- Involved in creation of scripts for archive and cleanup jobs.

Personal Information

Name	:	Manvendra Singh Khati
Date of Birth	:	14-04-1992
Nationality	:	India
Languages Known :		**Hindi and English**