**Parth. P**

**Senior AWS Data Engineer**

**SUMMARY:**

* An astute professional **with 8 years of experience in IT industry, comprising of build release management, software configuration, design, development, and cloud implementation.**
* Configuring, Automating and Deploying Chef, Puppet and Ansible for configuration management to existing Infrastructure.
* Used Ansible and Ansible Tower as Configuration management tool, to automate repetitive tasks, quickly deploys critical applications, and proactively manages change.
* Worked on upgrading Open stack releases from Kilo/Juno to Liberty and Mitaka releases.
* Worked on deploying Cloud both automatically using Fuel, Open stack Ansible and manual deployment.
* Developed in Pentaho, Informatica, Talend for Big Data projects, Dashboard Designs, and Data Visualizations. Utilized with Pentaho (PDI) Kettle, Informatica IDQ, PowerCenter, MDM, data management, Electronic Data Interchange (EDI) data integration/quality, and data governance.
* AWS Analytics services, Azure, Machine Learning, Internet of Things, IOT, DevOps with Terraform services, GCP, Big Table, Directory, Scripting, Automation, Data Lake, S3, Blob Storage, Salesforce, Azure Data Factory, ESB(Enterprise Service Bus), Logic Apps, API, EMR, Hive, SQOOP
* Intensive experience in AWS Cloud Environment, MS- Azure Services, and Hadoop ecosystem at designing, deploying, and operating highly available, scalable, and fault-tolerant systems.
* Extensive experience in leading multiple Big Data and Data transformation implementations in the Agriculture domain, Banking and Financial Services (BFS), Retail, and Fast-Moving Consumer Goods (FMCG) sector.
* Worked closely with Clients, Stockholders, Beneficiaries, Directors, Managers, SMEs in handling end-to-end data engineering, building data analytics, and transforming the business needs into viable products.
* Installed and configured WebSphere Application Server, Portal Server
* Administered the WAS Servers on WebSphere, wrote jack scripts to install application, wrote J2EE complaint java servlets to perform image transforms.
* Develop, Maintain and support Continuous Integration framework based on Jenkins
* Automation experience with tools such as Jenkins, Selenium, Junit, or equivalent

**TECHNICAL SKILLS:**

**Amazon Web Services (AWS):** EC2, S3, ELB, Auto scaling Servers, Glacier, Storage Lifecycle rules, Elastic Beanstalk, Cloud Front, Elastic cache, RDS, VPC, EBS, Route 53, Cloud watch, Cloud trail, Ops work, IAM & Roles, SNS subscription service, SQS, SNS, Code Commit, Redshift, Dynamo Db, Lambda, Code Deploy, EFS.

**Web Technologies**: HTML, CSS, Java Script, JQuery, Bootstrap, XML, JSON, XSD, XSL, XPATH Application ServersTomcat, Apache, Web logic, WebSphere, IIS 7.0 and JBoss.

**Languages/Scripts**: Java/J2EE, C, C++, Shell, Perl, Ruby, Python, JavaScript.

CI Tools Hudson, Jenkins, Bamboo

**Deployment Tools**: Chef, Puppet, Ansible, Docker.

**Version Control Tools:** Subversion (SVN), GIT, Perforce.

**Tracking Tools:** Jira, Remedy, and ClearQuest.

**Database**: Oracle, SQL Server, My SQL, DB2

**EDUCATION**: Master’s In information computer science – University of Texas at Arlington- USA.

Bachelor’s in computer science Texas Wesleyan university – USA – Texas

**PROFESSIONAL EXPERIENCE:**

**Client: SAFEWAY, PLEASANTON CA Sep 2019 to Present**

**Role: Sr. AWS Azure Cloud Engineer**

**Responsibilities:**

* Built on an end-to-end analytical landscape solution, involving PowerBI dashboards connected to the
* back-end SQL Server 2016 system on Azure to enable a government infrastructure agency to analyze/detect. farm loans and market facilitation. Included conceptualizing/streamlining various agency architectures.

into a single Govt reach out point.

* Built a data discovery platform for a large system integrator using Azure HDInsight components. Used
* Azure data factory and data catalog to ingest and maintain data sources.
* Created an end-to-end Azure cloud-based analytics dashboard for a federal government for showing.
* real-time updates for the various loan programs, market facilitation programs, and catastrophic relief
* funds.
* Successfully strategized and executed multiple enterprise cloud transformation programs in areas of
* Big Data and Enterprise Applications in multiple Elastic Map Reduce (EMR) clusters on Hadoop.
* MapReduce applications for enabling multiple PoC's.
* Utilized the Azure poly-based utility to run SQL queries on external data in Hadoop and as well as to import.
* and export data from Azure blob storage.
* Implemented microservices, data lake ingestion of data, and BI functions using the power U-SQL scripting
* language provided by Azure and python.
* Created real-time streaming Operational Data Store environment with Azure event hubs and stream.
* analytics which directly streamed into MS Power BI for corporate reporting
* Spun up HDInsight clusters and used Hadoop ecosystem tools like Kafka, Spark, and data bricks for
* real-time analytics streaming, Sqoop, hive, and Cosmos DB for batch jobs.
* Enabled multiple Big Data use cases for customers ranging from Security as a service to Customer.
* sentiment analysis
* In process of ramping up on Confidential Azure to enable the creation and management of Big Data
* architectures using MS Azure.

**Environment:** Azure Data Factory, Azure Data Lake Storage, Azure Stream Analytics, HDInsight, Databricks,

ML Studio, HDFS, MapReduce, Spark, HIVE, YARN, MongoDB, Kafka, Sqoop, Flume, Oozie, Python, Azure

SQL, PowerBI, Salesforce, and Zookeeper

**Client: Halliburton, Houston, TX. Dec 2018 to Sep 2019**

**Role: Sr. AWS Data Engineer.**

**Responsibilities:**

* Developed and implemented software release management strategies for various applications as per the agile process.
* Worked extensively with AWS services like EC2, S3, VPC, ELB, Autoscaling Groups, Route 53, IAM,
* CloudTrail, CloudWatch, CloudFormation, CloudFront, SNS, and RDS.
* Gained good experience by working with configuration management tool Ansible and CI/CD tool Jenkins.
* Managed Amazon Redshift clusters such as launching the cluster by specifying the nodes and performing the data analysis queries.
* Worked on AWS Elastic Beanstalk for fast deploying of various applications developed with Java, PHP,
* Node.js, Python, Ruby, and Docker on familiar servers such as Apache and IIS.
* With the help of IAM created roles, users, and groups and attached policies to provide minimum access to the resources.
* Worked on the databases of the Amazon RDS and carried out functionalities for creating instances as per the requirements.
* Design, develop and deploy ETL processes and performance tune ETL programs/scripts.
* FTP scripts and Extraction of Unstructured Data from various customers Data formats using Perl/Shell Scripts.
* Extensively worked on Data Masking transformation to mask data in DEV/QA/SIT environments.
* Worked on Performance Tuning to optimize the Session performance by utilizing, Partitioning, Push down.
* optimization, Index Cache, Data Cache, and incremental aggregation.
* Worked on different databases such as Oracle, DB2, and Teradata.
* Created Technical Design Document / BRD along with Data Discovery sheet.
* Closely worked with SME's Application owners and end-users to get details on their apps and requirements.
* Designed Java API to connect the Amazon S3 service to store and retrieve the media files.
* Implemented Amazon RDS multi-AZ for automatic failover and high availability at the database tier.
* Created CloudFront distributions to serve content from edge locations to users so as to minimize the load on the front end servers.
* Configured AWS CLI and performed necessary actions on the AWS services using shell scripting.
* Implemented CloudTrail in order to capture the events related to API calls made to AWS infrastructure.
* Implemented Ansible to manage all existing servers and automate the build/configuration of new servers.
* Wrote Ansible Playbooks with Python SSH as the Wrapper to Manage Configurations of AWS nodes and
* Tested Playbooks on AWS instances using Python.
* Involved in scrum meetings, product backlog, and other scrum activities and artifacts in collaboration with the team.

**Environment:** S3, Redshift, EC2, ELB, Autoscaling Groups, CloudTrail, CloudFormation, CloudWatch,

CloudFront, IAM, ANS, RDS, Jenkins, Ansible, Shell/Bash scripting, Python, JIRA, GIT, Elastic Cache,

Redshift, Glue, Lambda, Snowflake and CI/CD.

**Client: Hudson’s Bay Company, New York, NY. Aug 2017 to Nov 2018**

**Role: Data Engineer.**

**Responsibilities:**

* Spark, Sqoop, PySpark, Datameer, Tableau, Airflow tool
* Worked on Utilization platform where Data from multiple sources is ingested into AWS S3/ Redshift and
* transformed using Spark/ Scala scripts in Data Bricks/Glue.
* Developed python code for different tasks, dependencies for each job for workflow management and
* automation using Air flow tool.
* Designed and implemented a snowflake -schema Data Warehouse in SQL Server based on the existing data

warehouse.

* Experience in Data warehouse architecture and designing star schema, Snowflake Schema – FACT and
* Dimension Tables/Physical and Logical Data modeling.
* Built User Acceptance Test plans, End-End test cases, activity diagrams, user stories.
* Designed high-level views of the current state CISCO refurbished data files, leads, and website activity using tableau.
* Explained technical considerations at team meetings and delivered demo sessions on code, data model.
* architectures, and ETL transformations, including those with internal clients and other team members.
* Drafted SQL Queries for ETL and transformations based on dependency graphs for source tables using.
* Datameer. Created HIVE tables and worked on them using HiveQL.
* Imported and exported data into S3 buckets from the oracle Siebel database and vice versa using Sqoop.
* Partner with internal clients to improve understanding of business functions and informational needs.
* Migrated legacy reports from Business Objects servers to Tableau servers.
* Developed Python-based API (RESTful Web Service) to track sales and perform sales analysis using Flask, PostgreSQL.
* Converted SQL scripts into the hive to get better performance.
* Developed a web crawler code to obtain the raw data of product review and performed data cleansing in

Python.

* Implemented Custom File loader for Pig so that we can query directly on the large Data files such as build logs
* Improved the Performance by tuning HIVE and MapReduce.
* Created connection through JDBC.
* Created and maintained Technical documentation for launching HADOOP Clusters and for executing
* Hive queries, Pig Scripts, Sqoop job.
* Prepare a multi-cluster test harness to exercise the system for performance and failover.
* Exported summary tables into MySQL for reporting purposes using Sqoop
* Extracted data from the oracle database to HDFS using Sqoop.
* Handled importing of data from machine logs using Flume.
* Handled incremental data loads from Teradata into HDFS.

**Environment**: Hadoop, MapReduce, HDFS, pig, Hive, Sqoop, flume, oozie, Mango DB, Hadoop Distributions

(Cloudera, Hortonworks, data tax, Teradata), Kafka, PyCharm, GitHub, SVN, CVS, MS office

**Client: Toyota financial, TX. Sep 2016 to July 2017**

**Role: AWS engineer.**

**Responsibilities:**

* Worked as Azure, design and implementation on-prem and hybrid cloud solutions utilizing Azure and AWS.
* Experienced with DevOps, CI/CD pipeline tools Jenkins, Azure Cosmos DB, Data Bricks, Event Hubs, RLM, Bitbucket.
* Utilized Azure services and automation tools including Azure Resource Manager, Puppet, Chef, Ansible to implement cloud operating model to enable Environment-as-a-Service and DevOps capabilities
* Strong in MS Azure Databricks with Database and ADLS (Python, Spark SQL, using Parquet files)
* Worked as a Databricks Engineer with focus on Data Warehousing ETL development in Azure Databricks.
* MS Azure Databricks with Database and ADLS (Python, Spark SQL, using Parquet files)
* Strong in MS Azure Data Factory with Databricks MS Azure Databricks with Database and ADLS (Python, Spark SQL, using Parquet files)
* Worked in developing and designing Microsoft Azure SQL Data Warehouses; and, developing and maintaining SQL Server Analysis Services (SSAS)/SSIS/SSRS.
* Administered/Set up new networks in Mainframes, identified and solve potential program/data problems ensuring resolutions.
* Developed programs and queries to run analytics for the compensation program.
* Worked to form a cloud based big data engineering team to deliver platform automation and security.
* Built data workflows by using GCP, HBase, Big, Table, Big Query, AWS EMR, Spark, Spark SQL, Scala, and Python
* Worked with Azure Data Factory (ADF) to compose and orchestrate Azure data services.
* Working knowledge of Azure DevOps and its interaction with Databricks and Data Factory
* Good/Working Knowledge of Databricks Delta Lake
* Working knowledge of Azure DevOps and its interaction with Databricks and Data Factory
* Good understanding of Dimensional Modeling, SQL & DBMS concepts, ETL & Data Warehousing concepts
* Worked with Automation Tools Chef and Puppet, Data Cloud Architect, Azure, Ansible, Jenkins, Docker, Kubernetes, DevOps, Automation, CI/CD, Azure Services Compute, Networking, Storage, and Public Cloud platform.
* Worked with Visual Studio 2015, Team Foundation Server, .NET 4.5, SSRS, SQL Server Reporting Services, Web Security.
* Utilized Azure Data Factory to create, schedule and manage data pipelines, Data Cloud Architect, Azure, Ansible, Jenkins, Docker, Kubernetes, DevOps

**Client: PNC Bank, PA February 2015 to July 2016**

**Role: Jr. Data Engineer.**

**Responsibilities:**

* Participated in data acquisition with the Data Engineer team to extract historical and real-time data.
* Conducted Exploratory Data Analysis using R and carried out visualizations with MS-Excel reporting.
* Designed and implemented cross-validation and continuous statistical tests.
* Migration Services and AWS Schema Conversion Tool
* Coaches each Chef de Cuisine so that they are able to operate independently and creatively within their own profit centers.
* Installation & Fix Pack Upgrades of WebSphere Application Server instances
* Troubleshoot build issues in Jenkins, performance and generating metrics on master's performance along with jobs usage!
* Created reports and dashboards to explain and communicate data insights, significant features, model
* scores, and performance of new recommendation system to both technical and business teams.
* Market Assessment (External Data Analytics) - Retailer vs. Competitor/ Market - Share & trend
* analysis, Channel Assessment, Retail format-based analysis, Pricing Comparison between regions, Event
* comparisons, Category mix comparison, Private label/ National brand analysis, Promotion mix, Customer
* buying behavior.
* Customer Analytics - Demographic and purchase behavior segmentation, Customer
* Churn-Acquisition-Retention, Market Basket Analysis, Loyalty based analysis, RFM Scoring, Campaign
* Analysis, Customer Concentration Analysis, and Customer Purchase Behavior Analysis.
* Operations Analytics comprises store Scorecard, Cluster-based analysis, Store and Resource productivity
* analysis, Growth-Trend analysis, Like-for-Like Store analysis.

**Client: Broadridge- INDIA Jan 2013 – Dec -2014**

**Jr. AWS ENGINEER**

**Responsibilities:**

* Implement systems that are highly available, scalable, and self-healing on the AWS platform.
* Support production issues and troubleshooting for both on premise DBs as well as AWS RDS
* Private Cloud Environment - Leveraging AWS and Puppet to rapidly provision internal computer systems for various clients
* Installs and configures Windows server hardware and software, to include physical and virtual servers, legacy Linux, Windows Server, Active Directory, IIS, VMware ESX, Virtual Center, VMotion and HBSS
* Strong hands-on skills in Windows Server 2003/2008/2012 and virtualization (VMware)
* Used Ansible to document all infrastructures into version control.
* Implemented Ansible to manage all existing servers and automate the build/configuration of new servers.
* Deploying, managing, and operating scalable, highly available, and fault tolerant systems on AWS
* Frequently switched between positions as Cold Food prep, Dessert Station and Relief Chef mid shift to support changing needs of the kitchen, developed popular daily specials with personally sourced ingredients for broiler and saute stations.
* Application Server - IBM WebSphere Application Server Network Deployment, WebSphere Commerce, WebSphere Portal (WWCM / RCI), Apache Tomcat, JBoss
* Acts as administrator for legacy Linux, Windows, and ESX server operations. Responsible for building, installation and configuration of all legacy Linux, Windows, and ESX servers. Applies patches, fixes, hardware and software upgrades and physical to virtual migrations. Plans upgrades and coordinates with users
* Worked in all areas of Jenkins setting up CI for new branches, build automation, plugin management and securing Jenkins and setting up master/slave configurations.
* Worked on administering Jenkins, Git, NuGet, Puppet, Microsoft Visual Studio, WiX Toolset, PowerShell
* Amazon AWS Cloud Architect Certification or Microsoft Equivalent
* Manage the global Windows Server infrastructure, both directly (hands-on) and indirectly (through vendors)
* Worked on troubleshooting Openstack issues from fault detection, fault isolation and resolving many key issues.
* Worked as an Openstack administrator in creating, managing users, roles, groups, domains, services, endpoints; managing quotas, instances, snapshots, metadata, creating new images; managing and creating networks, routers, subnets, allocation pools; managing swift objects, time stamps, cinder volumes; using heat templates and troubleshooting.
* Migrate existing databases from on premise to AWS using various tools including but not limited to AWS Data

**Environment**: Microsoft Azure (Azure Virtual Machines,VMSS,Azure Storage Disk(Block Blob),VPN Gateway,Application Insights, AZURE Portal), AWS Cloud services IasS(EC2,S3,VPC,CloudWatch), Pivotal PasS, HDP 2.5/2.3,HDFS,MapReduce,Tez,YARN,Pig,Hive,HBase,Sqoop,Oozie,Zookeeper