Shishir
Email Id:manushishir008@gmail.com

Mob. No: +91 8743912122

Experience Summary

- 11+ years of experience in **Designs, Development, Maintenance of Cloud Infrastructure, and Big Data Solution.**
- Good understanding and exposure in solution designing and implementation of **big data application using AWS services.**
- More then 5 years of strong experience on Big data solution using AWS and Azure services like - **EC2, EBS, S3, Lambda functions, CloudWatch, RDS, Redshift, Glue, Kinesis** , **EMR**, **Athena**, **Snowflake, ADLS, Databricks, VMs, Notebook** etc. Also worked on Hadoop different distribution like Cloudera and Hortonworks.
- Strong experience in cost management and cost saving.
- An experienced and passionate Big data professional with extensive knowledge of Big Data ecosystem and tools like **Hadoop, MapReduce, Hive, Spark, Kafka, Airflow, Yarn, Sqoop, Hbase & cloud services(AWS and Azure)**. Accumulated deep experience in Enterprise System development.
- Good understanding and exposure of DevOps ecosystem like **terraform, Jenkins, Ansible, GitHub, bitbucket** etc.
- Developing optimized and standardized solutions for Data Integration (From upstream systems), data cleansing, and data Transformation and Data delivery to Downstream Systems.
- Strong exposure and hands-on in AWS Solution Architecture, Systems Design, Disaster Recovery, Security, Storage Administration, Data Center technologies & Linux/Unix/Windows Operating systems.
- Developed the ingestion framework for ingesting data from different in-house database sources by configuring the Metadata information using Spark and ScalaConsuming JSON and XML data and creating structure using Hadoop in the form of Hive tables on S3 and HDFS.
- Manage conflicts and able to prioritize deliveries in a resource constrained situation.
- Build the production-grade prototypes to demonstrate the technical depth in the new technologies.
- Knowledgeable in data modelling/architecture of NoSQL stores like **MongoDB/HBase/Snowflake/redshift** and contributing in data modelling as needed in the product.
- **Good experience to work in UK (European customer).I worked in UK location around 18 months.**
- **I have good experience to communicate with data scientists team in terms of understanding view and data quality. Currently coordination with BCG(Boston consulting Group, Germany)team to provide data quality for research work**

Expertise in following Technologies/Tools:

Bigdata and Hadoop	Hive, Sqoop, Kafka, HBase, Elasticsearch, YARN, Spark
Database	Snowflake, DynamoDB,SQL,Redshift
DevOps Tool	Jenking, Docker, BitBucket,CI/CD pipeline, Terraform, Git, GitLab, Branching strategy ,Deployment strategy etc
Language	Core java, Python
Visualization	Tableau 10.5
Scripting language	Unix, shell-script, Javascript, Jquery
ETL Tool	Talend Open Studio

Workflow tool	***Apache Airflow***
AWS Cloud	AWS console, EC2, EMR, S3, Glue, Athena, Redshift, Kinesis, Cloud watch, lambda, API gateway, VPC, ALB, ELB, Redis etc.
Azure Cloud	Databricks, ADLS, ADF, Web App, Notebook, Azure DevOps, VMs, Azure keyvault, Resource group ,Azure CDN,ACS etc.
Data Security	Ranger, encryption/decryption (Python)
OCR	Tesseract, Tikka
Agile Tool	Jira/Azure DevOps agile board
Web Tool	JSP, JDBC API,JPA,Jquery, XSLT,XML,Java

EXPERIENCE SUMMARY:

Currently working in IBM India, as a Tech Lead—Data Engg. since December 2020 to till now.
Worked with kloud9 Pvt. Ltd. as a Tech Lead—Data Engg. from 26th May 2020 to December 2020.
Worked with Sonata software Ltd as a Big Data Technical Lead from 23rd Oct 2018 to 23rd May 2020 deputed in UK.
Worked with Wipro Technology as a Big Data Technical Lead from 18th August 2017 to 18th October 2018.
Worked with HCL as software consultant from 10th Feb, 2014 to 18th August 2017.

Significant Project Experience:

1. **Project:** **Caspian data Lake**

   Client: **Nike(USA)**
   Team size: 15
   Tools & Technologies: **AWS,EMR,Athena,Glue,Spark,Python,Airflow,Redshift,Hive,Bitbucket,terraform**
   Role: Team Lead

Responsibilities:

- Worked with Client Nike for migration of on-premises applications to AWS.
- Analysis Data and It's Size.
- Data Transformation as per business.
- Design ETL pipe line for Data Ingestion from end to end
- Design complex workflow for pipeline from end to end.
- Creating Jenkins CI/CD pipeline
- Creating AWS infrastructure using terraform.
- Coordination with Product owner and understanding requirement.
- Coordination with DevOps team and sharing development team code repository stuff and working CI/CD automation.
- Designing and Implementation Data and Networking architecture, designing data security solutions for AWS services. Determine when to use AWS Storage encryption, AWS Disk Encryption, AWS SQL Database security capabilities

- Solved various performance related problems being faced using architecture modification in Existing infra setup

2. Project: **Next generation analytics platform**

Client: **TUI(UK)**

Team size: 12

Tools & Technologies: **AWS,EMR,Athena,Glue,Spark,Python,Airflow,Redshift,Hive,Bitbucket,terraform**

Role: Team Lead

Responsibilities:

- Improving, Designing and Creating ETL jobs performance.
- Used CloudFormation and ansible for deployment and configurations of EC2 instances and other AWS services.
- Work with prospects and Product owner for understanding requirement and creating SOW.
- Lead planning and design of public cloud-base architectures, solutions, configurations, integrations, and migrations for mobility service platform.
- Keep stakeholders up-to-date on project status, risks and challenges and proactively resolve risks and challenges
- Setting up CI/CD pipeline using Jenkins to code deployment.
- Setup and secured environment based on AWS best practices

3. Project: **Sentiment Analysis (Machine Learning)**

Client: **Rajasthan Govt.**

Team Size: 6

Tools & Technologies: AWS, Python, Spark ,hive, Tablue,Hbase

Role: Team Lead

Responsibilities:

- Interacted with client to understand requirement, planning with team to achieve goal
- Creating Requirement based story and assigning to team after discussion with Product owner
- Estimation of cluster Node and understanding of data size and its behavior
- Worked on AWS EC2, IAM, S3, EBS, Elastic Load balancer (ELB), auto scaling group services.
- Created a best practice Build environment using Jenkins, immutable instances, and AWS.
- Worked on bucket policies, S3 storage like Glacier, maintained SNS and CloudWatch monitoring services.
- Utilized AWS CLI to automate backups of ephemeral data-stores to S3 buckets, EBS and create nightly AMIs for mission critical production servers as backups.
- Manage AWS EC2 instances utilizing Auto Scaling, Elastic Load Balancing and Glacier for our QA and UAT environments as well as infrastructure servers for GIT.
- Build, manage, and continuously improved the build infrastructure for global software development engineering teams including implementation of build scripts, continuous integration infrastructure and deployment tools.
- Created JIRA issues to prioritize what's important and stayed up to date with what's going on around the project.
- Troubleshooting alerts triggered from various monitoring tools.
- Configured Cloudwatch alerts and created Cloudwatch Dashboards for monitoring and reporting.
- Deployed ETL Jon and monitoring its performance
- Created SOPs for various task for Team members

4. **Project:** **Digital Library**

Client: **Rajasthan Govt.**

Team Size: 8

Tools & Technologies: **Hbase, java, Solr, Sqoop , Kafka ,Tesseract ,Tikka , pyspark**

Role: Team Lead

Responsibilities:

Digital Library is kind of Content Management System (CMS) for Rajasthan Govt. Technically here we are inserting any format of documents in HBase except Audio and Video and in Hbase table ,we have enabled HBase Lily Indexer which basically index metadata along with contents of document in Solr. It is real time searching and as when documents get indexed in Solr immediately we can search the same on search page.
To get the documents content and metadata we have implemented the Apache Tika in Java but it cannot get the contents of images so for that we are using tesseract OCR engine as well.
Here we have developed two Tabs in Java as details are mentioned below.

- Upload Document: Using this tab we upload the document in Solr by using Java API (Apache -Tikka).
- Search Document: Using this tab we search the document by using Solr.

5 **Project:** **Face Recognition(Machine learning)**

Client: **Orissa govt.**

Team Size: 10

Tools & Technologies: **Hbase, java, Solr, pyspark,OpenCV**

Role: Team Lead

Responsibilities:

This Face Recognition use case is related to find matching faces from stored images in Oracle DB. Images stored in Oracle in form of Blob. We use algorithms to compare a digital image captured through a camera and stored in Oracle DB in terms of Blob. We used HOG (Histogram of Oriented Gradients) method very effectively for object recognition and thus suitable facial for face recognition also.
We use following steps as mentioned below:

1. Image Source
2. Face detection
3. Face normalization
4. Feature extraction (Vector)
5. DB based feature matching
6. Face ID

6 **Project:** **UIDAI**

Client: **Govt of India.**

Team Size: 4

Tools & Technologies: **hive,java,Hadoop,hbase**

Role: Technical Consultant

Responsibilities:

Aadhaar database CIDR (Central ID repository) is hosted on a central system powered by data centers. This data is used to serve Aadhaar project's core objectives such as:

(1) **Enrolment application:** is used for receiving new client enrolment requests and capturing new data. After verifying the uniqueness of the request, the Registrars enroll the data that is received in magnetic media from various logistic providers. This data is then uploaded to Aadhaar database post-validation. The Registrars include (but are not restricted to) ministries and departments of state and central governments, banks and other financial institutions, telephone companies, etc. Once this is done, the Aadhaar number is generated for the request.

(2) **Authentication application:** will conduct online authentication of identity (demographic and biometric information) done by querying the Aadhaar database that responds to such queries in the form of Valid/Invalid type of response. Also, de-duplication of biometric data is done by assigning a scaled data fusion score to each duplicate record.

Thanks,

Shishir