VAMSHI KARANAM

**SUMMARY**

* Solution & Data Architecture: Vamsee has 20 years of overall IT experience and 10 years as a cloud solution architect. Skills vary from analyzing the infrastructure needs, Cloud vendor assessment, application platform choice, ETL tool assessments and reporting platform assessments.
* Performing proof of concepts, building the various components and interfaces, interacting with various teams both onsite and offshore, ramping up the support team are various activities of my daily routine.
* Azure Cloud Experience (7 years): Azure certified Solution Architect with 7 years of experience in Azure cloud. Architected Systems for various clients in Azure cloud leveraging Azure SaaS and PaaS. Right from the blueprint phase I played a crucial role in building this platform from the ground up.
* The services I used include APIM, Logic Apps, Azure Data Lake storage, Functions, Azure Data Factory, Synapse, Databricks, AKS, CosmosDB, Azure SQL managed instances, Event Grid and Azure DevOps.
* AWS Cloud Experience (5+ years): 5+ years of experience in the AWS cloud. Services used are EC2, S3, EMR, Athena, Glue, Kafka, ActiveMQ, DynamoDB, Talend, RDS(Oracle) and Snowflake.
* Generative AI Experience (1 year): Exploring Gen-AI capabilities on Azure OpenAI platform for Safety reporting and Regulatory Submissions. Experience with Langchain, Microsoft’s Autogen AI agent, generating embeddings, storing in vector databases like ChromaDB and Azure Search.
* ETL Experience: 10+ years of ETL and Data experience using tools like Synapse, Databricks, Talend DI & ESB, Informatica, Azure Data factory, AWS Glue, AWS EMR, Cloudera Hadoop. Proficient in SQL and Python.
* CRM Experience: A decade experience in CRM space as Siebel CRM architect and developer. Salesforce certified developer.
* Project Management: Project estimation using Agile project estimations, Project plan with phase wise development, licensing review with help of stewards, deployment schedules, release management, stepping into other footsteps to resolve conflicts, convincing others of a proposal with detailed analysis.
* Techno-functional expertise with a blend of strong interpersonal skills is what you can expect out of me. I work alongside business and IT leadership to create alignment and sustainable execution process.
* Clients: IQVIA, Merck pharmaceuticals, Janssen pharmaceuticals, Celgene, Blue Cross and Blue Shield of North Carolina, One Call medical and Lowes.

**SKILLS**

Azure, AWS, Databricks, Unity Catalog, Delta Live Tables, Synapse, Kusto, Talend, Informatica, Oracle, EMR, Glue, Azure Data factory, Snowflake, Salesforce, Cosmos DB, Neo4j, Couchbase, Postgres, Siebel, AKS, AWS Athena, Hadoop, Cloudera, DynamoDB, MongoDB, MuleSoft, Logic Apps, Azure API Management, Kafka, ActiveMQ, SQL, JAVA, Python, Microservices, Datalake, Azure DevOps, Azure OpenAI, Vector Databases

**SKILLS, TOOLS & TECHNOLOGIES**

Azure Databricks Delta Live Tables, Unity Catalog, Spark StreamingADLS, Synapse, Kusto, APIM, Logic Apps, Event Grid, KeyVault, AKS, Functions, Cosmos DB, Azure Monitor, Azure DevOps

ETL Talend, Data Factory, Synapse, Databricks, Informatica, Glue, MuleSoft

AWS EC2, S3, Glue, EMR, DynamoDB, KMS, RDS, AthenaEstimation and Gap analysis

SQL Databases Oracle, Postgres, Snowflake

NoSQL Databases Cosmos DB, Couchbase, MongoDB, DynamoDB, Neo4j, Redis, Data Modeling

**ETL EXPERIENCE**

**Azure Databricks**

* Designed and built the lake house (medallion architecture) using Databricks and Azure Data Factory
* Data from various source systems had to be extracted using ADF pipelines and ingested to ADLS bronze layer.
* Configured pipelines in Databricks using jobs and notebooks to read the data from bronze, perform the transformations in pyspark and write it to ADLS (silver – refined)
* Configured Delta-live tables for near real-time processing of the IoT datasets.
* Configured Unity Catalog per Databricks standards.
* Expertise in performance tuning the notebooks and jobs using various spark configurations and code.
* Built CICD pipelines in Azure DevOps to build and deploy resources across various environments using PowerShell scripts.

**Azure Synapse Analytics**

* Built pipelines using Dataflows and copy activity to copy over the raw data into the data lake.
* Leveraged Synapse notebooks in pipelines to perform lookups and transformations using Pyspark.
* Populated data in ADX/Kusto (Analytics Store/ Data Explorer) in batch mode. PowerBI would run reports and generate dashboards out of this data.

**Other ETL tools & Experience**

* Built various pipelines in ETL tools like ADF, Talend, AWS Glue and MuleSoft.
* Worked on Informatica (v9) to load the Clinical Trial Management System’s data into a warehouse.

**CLOUD EXPERIENCE**

**Azure Experience:**

* Designed the network topology in Azure using VNets, Subnets, VPN Gateway, Express Route and palo alto firewall (F5 networks).
* Building the various 3rd party integrations using Logic apps, Functions, Event Grid, Azure SQL Managed Instances, Cosmos DB and exposing them using APIM.
* Integrated the IHP (Integrated Health Platform) with a third-party telehealth system which provide video conferencing features for the patient and the provider. Azure APIM for REST APIs, logic apps and Cosmos DB were leveleraged for this feature.
* Integrated IHP with a IBMCD (EDC system) in batch mode using Databricks.
* Integrated IHP with 1UPHealth using REST APIs. 1UpHealth connects to various EDC systems and sites to fetch the EHRs, convert them to FHIR format and expose the data via REST APIs in both bulk and individual resource types. MuleSoft, ADLS, Logic Apps were used to achieve this functionality.
* Developed Databricks notebooks & jobs to access the data in ADLS, build delta tables in the delta lake and populate datamarts using this refined data.
* Worked with complex nested JSON structures and CSV format datasets to generate parquet files in ADLS.
* Orchestrated Databricks jobs using Azure DataFactory for triggering the jobs on certain events in the EventHub.
* Performance tuning the notebooks and jobs using various spark configurations and code.
* Built an Eventing framework using Azure Event Grid, logic apps, Data Factory and SendGrid(email and SMS). This also includes Error handling framework.
* Used pyspark heavily with Synapse to build notebooks, orchestrate them in pipelines for processing huge amounts of data.
* Setup Azure DevOps pipelines for Synapse. Configured the connection to the repo, created the YAML files for buddy, PullRequest and Official release. Setup the connection service, release pipelines to deploy the notebooks, pipelines, dataflows and linked services to higher environments with the respective parameterized configurations.
* Populated data in Azure Kusto (Analytics Store/ Data Explorer) in batch mode. PowerBI would run reports and generate dashboards out of this data.

**AWS Experience**:

* Implemented the registry platform in AWS. Was heavily involved in the exploration phase of the various AWS PaaS and SaaS tools evaluation and pricing.
* Leveraged Cloud formation templates to automate instantiation of EC2 instances (redhat linux) and deploy the registry application (java, spring mvc, reactjs) backed by Oracle in RDS (multi zone for HA).
* Designed and built various integrations leveraging ETL tools and AWS services, some of which included EC2, S3 storage, Talend, Kafka managed service, RDS (Oracle), EMR (spark), AWS Lambda, AWS glue and Snowflake.
* Architected and built an interface for Cancer registry for which data from file uploads would have to be extracted and pushed to Kafka topics (AWS MSK). Spark jobs on EMR would consume the data from the topics and load to Snowflake DWH.
* Built a reconciliation framework using Athena to find any missing data/updates between the registry application and the DWH.
* Was involved in the installation of Talend ETL job and ESB servers. Built Talend ETL jobs and routes to consume the feed from the registry application from ActiveMQ queues, parse the dataset and load it into Oracle database.
* Built various ETL jobs in Talend to load data for migrating the on prem data to RDS (oracle). These were run on the Talend job servers.
* Leveraged Splunk as a centralized logging service. AppDynamics for monitoring the performance of the application and databases. Bluestripe for network performance and issues. Oracle OEM for database issues and tuning.

Solution & Data Architect Apr 2010 – present

IQVIA | Durham, NC

https://www.iqvia.com

CRM Consultant

Lowes | Mooresville, NC 02/2010 - 04/2010

https://www.lowes.com

CRM Consultant

Celgene Pharmaceuticals (BMS) | Warren, NJ 05/2009 - 01/2010

www.bms.com

https://www.celgene.com.au

CRM Consultant

Blue Cross Blue Shield of North Carolina | Durham, NC 10/2008 – 04/2009

https://bluecrossnc.com

CRM Integration Consultant

Federal Student Aid (Perot Systems Govt Services) | Fairfax, VA 01/2008 – 09/2008

https://studentaid.gov

CRM Consultant

Merck Pharmaceuticals | West Point, PA 08/2006 – 12/2007

https://merck.com

Janssen Pharmaceutical Companies of J&J | Titusville, NJ 07/2005 - 07/2006

https://janssen.com/us

CRM Integration Consultant

One Call Care Medical | Parsippany, NJ 04/2005 – 06/2005

https://onecallcm.com

Web Developer

Sify India Pvt. Limited | Chennai, Tamil Nadu, India 11/2002 - 11/2004

<https://sifytechnologies.com>