**A blue and white logo

Description automatically generated with low confidence**

**Ashish Saugar**

**(515) 360 2833**

[**ashishsaugar@gmail.com**](mailto:ashishsaugar@gmail.com)

**PROFESSIONAL SUMMARY:**

* Azure Data and Solution Architect with 15 years of experience in IT, Automobile, Healthcare, High-Tech, Ecommerce, Power, Retail domain. Established best practice standards and guidelines for design & development, deployment of innovative business solutions for data analytics and cloud solution for various clients with strong business acumen and technical expertise into Azure, AWS, GCP and Big Data space.
* Results oriented, decisive leadership in all Big Data, Azure technologies that combines with Big Data and cloud consulting, strategy and implementations, CoE setup, designing, presales, and revenue optimization.

**Areas of expertise include:**

* End to end data migration setup for Big data and Azure
* Data operations, engineering involves end to end pipeline deployment
* Azure Infra and DevOps engineering
* Data Warehouse - Azure Synapse Analytics, HDInsight/Hive, AWS, GCP
* Language – Python, Java, SQL, JSON and PowerShell Scripting etc.
* Application - Power BI, Tableau, QlikView, Azure Data Factory
* Azure and Big data governance
* Legacy Modernization
* NoSQL, Low Latency
* Big Data/Hadoop solution Architecture, Design & Development
* Agile & Waterfall Methodologies
* Social Media, Community & Viral Marketing
* Big Data strategy & solution design
* BI Consulting
* Presales
* Visual and real time Analytics
* Data ware Housing and Cloud

**Specialties:**

* Operations in Azure & Big Data space, Azure-Big Data implementations, Administration, NoSQL/HBase/Cassandra, EDW, RFP/RFI/Defence proposal, cloud & Big data strategy budgeting
* Experience in different Hadoop distributions like Cloudera, Hortonworks, and Azure HDInsight.
* Expertise in deploying Hadoop on cloud like Azure HDInsight, ADF, Pipeline, Linked Services, ADW, ADL, Cosmos DB, .Net, Java, Spring Boot, PowerShell, Azure Quantum computing, Azure HPC, Azure Arc.
* Experience in automate the deployment of HDInsight cluster using PowerShell scripting and to install various component on HDI cluster with different Hadoop distributions.
* Implementation of security cert (Service Principal) and enablement of security/firewall on ADL and storage accounts.
* Involved in RFP, defence pitch, Big Data sales pitch and strategies engage with customers and provide solution design, consulting, and recommendations.
* Provide consulting to customers in identifying Big Data use cases and then guiding them towards implementation of use cases.
* Big Data strategic planning, technology roadmap, talent acquisitions and mentor team for cutting edge technology competitiveness such as Hadoop, Hive, HBase, Cassandra, AWS, GCP, Cloud computing, Tableau, QlikView, Spark, R, social graph, big data analytics, data visualizations.
* Expertise in installing, sizing, configuring, Commissioning, Decommissioning, performance tuning and upgrading HDInsight cluster with size of more than 100 nodes.
* Expertise in deploying Hadoop on cloud like Amazon, HDP and Microsoft Azure, HDInsight
* Involved deploying code multiple times using PCF.
* Followed product development methodology based on Agile as Scrum master
* Involved in creating and analysing software requirement specifications (SRS) and Functional specification document (FSD). Strong knowledge of Software Development Life Cycle (SDLC)
* Involved in presales data solution.
* MSBI, COSMOS, .Net, Azure VMs, CICD, GIT, VSO

**PoC Implementation:**

* Real Time Fraud Detection (RFD) by using Flume, Storm, HBase
* Big-data Aggregator by using Storm, HBase and Twitter API
* ETL on Hadoop: Mainframe to Hadoop Migration using Talend ETL, Hive, Impala, Tableau
* Augmented DWH on top of Hive using DMX-h, Hive, Impala, QlikView
* Augmented ADWH on top of Azure

**TECHNICAL SKILLS:**

* Designing and setting up ADO CICD pipeline and automate using Bamboo, Bitbucket, Ctrl-M, Docker
* HDFS, MapReduce, Pig, Hive, SQOOP, Oozie, Zookeeper, Flume and HBase, Spark, Strom, Azure (ADF, ADLS, HDInsight, Synapse analytics, ADW), Amazon EC2, S3 and RedShift), GCP, Spark, Storm, Impala, Elasticsearch, Talend 6, DMX-h, Docker, Bamboo, Bit bucket, Jira
* Experience in MS SQL DB (on-perm & Azure) and most of the Azure cloud and Bigdata based technologies.
* BI (Business Intelligence) end to end pipeline deployments and consulting
* Experience Big data/Azure combine solution & design proposal
* Experience developing real-time Big Data solutions on Windows/Linux.
* Experience in various file formats JSON, Parquet, Avro, and ORC.
* Experience in configuring HDInsight, Spark, HBase, Apache Storm, Firewall, Ranger for HDP cluster.
* Working experience in creating Automated deployment of Azure HDInsight Hadoop cluster using PowerShell with ARM template
* Good experience & knowledge working in Azure PowerShell.
* Experience in scale-up and scale down the HDInsight (HDP, CDH) cluster
* Rich hands-on experience in various Hadoop distributions (Horton Works, Cloudera)
* Troubleshooting, performance optimization techniques and data visual analytics jobs and scheduling jobs/workflows.
* Good experience in versioning tool like Git Hub, Bit bucket, SVN.
* Experience in internet of things.
* Experience in multitier application on web and ETL/ELT.
* Experience in most of the Hadoop ecosystem (YARN, HDFS, Livy-spark rest API service, Zookeeper, Hive, Spark SQL, Impala, SQOOP, Hue, Ambari, Kafka, flume, Kerberos, KNOX, HBase etc.)
* Experience in Azure/HDInsight admin
* Experience in Azure Storage systems like BOLB Storage, ADL, ADW, Primary/Secondary storage
* Apart from that experience in ADF, Linked services, e2e data pipeline setup
* Specialization in Legacy modernization techniques, Quantum Computing.
* Experience in OS image cloning and migration/updating
* Classified - Azure active Dir, Azure security, Azure firewall, ARM template Cosmos DB-Azure, Load balancer, storage accounts/Containers, Power BI, Azure Marketplace,
* VMs, visual studio on cloud, Function App, Event Hub, U-SQL, COSMOS script, T-SQL, Automation accounts, Recovery Service Vault, DevOps activity, Cassandra-SQL.
* Project methodology: SAFe Agile and Waterfall

**PROFESSIONAL EXPERIENCE:**

**Consultant -** Big Data Operations | DevOps  **May 2022 - Present**

**Mercedes Benz | Atlanta**

**Product Name:** Vehicle Sale

**Description**: A legacy modernization requirement, which involves data movement from Mainframe and SAP to Azure Synapse Analytics.

**Responsibilities:**

* Collaborate with stakeholders & business users, design for data extraction, transform, load and provision workloads.
* Designed end-to-end data processing platform with high availability, optimization techniques.
* Ingestion and troubleshooting in data migration process from different data sources to Azure ADLS and Synapse.
* Collaborate with business users/stakeholders to understand data lineage, governance and data correctness in Synapse ADW and articulate the information technically for development team.
* Resolution of ADF issues related to pipeline failures and incremental data unavailability.
* Designed and integrates various cloud and on-prim tech in the stack for data flow.
* Review Azure Databricks code with technical team offshore, onsite.
* Involved in designing DevOps CICD for different work areas.
* Involved in designing & RFP for new data operations and DevOps proposal.
* Involved in generating Power BI reports for business users and UAT.
* Worked on RFP proposal for Microsoft Fabric design.

**Environment:** Microsoft Azure, ADLS, ADF, Databricks, Synapse, Microsoft Fabric-Power BI, Terraform, ARM Template, PowerShell, CICD, ADO, GitHub, Azure NSG, Azure VM, HPC, Agile, Confluence, Aecorsoft, SAP, DB2.

**Azure Data Architect Dec 2018 – Apr 2022**

**Microsoft Corporation**

**Product Name:** 1P SRE & D365 – CXP Business Portal

**Description:** Dynamic 365 –CXP BP portal provides Business Intelligence reporting to CXP team and various stakeholders inside Microsoft, CXP team has a Data Mart, which consumes data from multiple sources along with analytical data to provide stakeholders a better insight for all D 365 workloads/product’s businesses.

**Responsibilities:**

* Big Data, Azure DevOps, CI, CD, CT, CS, Azure, PySpark, PowerShell, solution designing.
* Ingestion and troubleshooting data migration from different on-premises and online data sources to Azure VM, ADLS.
* Designing ADO CICD automation using Terraform and Helm.
* Created pipeline for PySpark jobs for data ingestion and code refresh
* Automate creating Azure instances and HDInsight cluster using PowerShell and Azure Automation.
* Managed Azure Infrastructure, Azure web roles, Worker roles, VM Role, Azure SQL, Azure Storage, Azure AD, VM Backup and Recover from a Recovery services vault using Azure PowerShell and Azure portal.
* Virtual Machine Backup and Recover from Azure Site Recovery, vault using Azure PowerShell and Azure Runbook Automation account
* Involved in windows server OS Patching.
* Involved in Power BI reporting setup for MSBI on Azure portal and generated reports for Microsoft stake holders and leadership
* Data ingestion and recovery on top of Azure VMs, Blob Storage and ADL using ADF, Event Hub from sources like MSSQL, COSMOS DB and SQL servers.
* Implemented CICD pipelines for Development, Test and Production environments with GIT repos.
* Automate the creation of different instances and resource like ARM, Service Principle, Blob Storage accounts, HDInsight with ADL, SQL servers for Hive and Oozie
* Data pull from COSMOS, MS SQL Server and SharePoint online and store it into target Data Mart (Azure VM), using Scope scripting, SQL queries.
* Communicate to various Microsoft stakeholders for their data requirement.
* Handling technical issues tickets from client like Azure server, ADL Storage, host/port.
* Expertise in Azure Data Lake, Azure Data Factory, MSSQL, Power BI, HBase, Hive & performance tuning, troubleshooting and optimization for Azure HDInsight.
* Involved in application and code migration from Azure Data Bricks to Synapse.
* Extensively worked with Scheduling, deploying, managing container replicas onto a node cluster using Kubernetes and experienced in creating Kubernetes clusters work with frameworks running on the same cluster resources.
* Created and maintain HDInsight cluster and fine-tune the cluster components.
* Installed ACR and downloaded images and from Docker Hub and created
* Docker files to automate the process of capturing and using the images.
* Developed Networking policies for Docker Containers. Worked on Open source development tools like Docker Containers
* Responsible for gathering business requirements for various workloads/products of Office 365
* Provide Microsoft Tenant related data to different stakeholders.
* Managing leading offshore team provides them daily assignments
* Involved in daily scrum calls in order to address various issues to the project team, product owners and leadership to discuss resolutions/mitigation
* Communicated daily development reports to the client using VSO/SPO

**Environment:** Azure, Big Data, ADL, ADF, COSMOSDB, Dynamic 365, Cloud networking, security, Jarvis, Synapse, HPC, Azure Data Bricks, VSTS, AKS, Terraform, Helm, YAML, Docker, Power BI, Agile, SharePoint-online, Git, CICD, Visual Studio, VSO.

**Technical Team Lead - Cloud May 2018 - Nov 2018**

**Exelon Corporation Oakbrook, IL**

**Product Name:** BIDA Grid – Data Analytics Platform (DAP**)**

* The BIDA Grid – DAP is a Data Lake for Electrical Network Connectivity related data and is intended to analyse or identify any incorrect/implausible mapping of meters to transformer through analysis of variance of different parameters like Consumption, Voltage, Geospatial location constraints and outage information. The DAP is expected to support the analytical application to geospatial represent Exelon’s connectivity model.
* Involved in solution design for legacy migration from Oracle to Hadoop.
* Participated extensively in system Integration and Designing
* Involve in data processing on top of Hive and stored into data lake
* Setting up Dev, Test and UAT Environment for all the teams.
* Created Various ETL jobs in ODI for historical data
* Involved creating Azure Data Lake and end to end pipeline activities.
* Created Azure blobs and containers for history and raw data store
* Develop various MR jobs/algorithms and Spark SQL for certain value adds
* Import and export data using from on-premises and cloud data sources to HDInsight and vice versa
* Involved in Commissioning and Decommissioning and performance tuning of nodes in WHDFS
* Use networking and platform switching tools
* Configured active directory, Kerberos, and key vaults.
* Created and analysed Software Requirement Specifications (SRS) and Functional requirement and the pain areas of clients existing system
* Functional Requirement Document (FRD)

**Environment:** Azure, HDInsight-CDH, Hive, ADL, ADF, ODI, PySpark, Elasticsearch, GCP, Spring Boot, ADW, CICD/GIT, Confluence, Agile.

**Technical Team Lead – Big Data | Azure Oct 2017 - Apr 2018**

**Anthem Inc., Richmond, VA**

**Product Name: Large Group Rating System**

* The LGRS is the DataMart piece of this project operates in an Agile Methodology and the development for the DataMart development work will be done in sprints and program increments.
* LGRS is a strategic asset; therefore, it is processed on the Strategic Asset on top of Big Data Platform.
* Responsible to pull data from Teradata appliances using TDH method and store that to Azure CDH HDInsight
* Automating the HDInsight cluster deployment for seven different teams by using ARM template and PowerShell.
* Responsible for setting up end- to- end ETL data pipelines and linked services on ADF Pipeline, ADL and ADF.
* Involved in setting up and configured Spark cluster and run data pipeline on ADF
* Created Linked Services using azure portal & PowerShell to define the connectivity and data flow using PowerShell script for on premise and Azure based data sources.
* Provided solution design for the data migration from Teradata to Hadoop (HDI)
* Created end-to-end PowerShell script to create HDInsight cluster by deploying ARM template and to create Linked Services to execute hive and spark jobs on top of HDInsight.
* Monitored HDInsight cluster and optimize performance of Hive and Spark.
* Bug Fixing and incident reporting on call from admin team testing team and coordinating with other stockholders.
* Managed cloud security and identity management, AADS implementation.
* Written various hive and Spark SQL scripts for various operations.
* Analyse existing HDInsight clusters and optimize Hive jobs/queries

**Environment:** Azure Portal, Azure Data Lake, ADF, Blob storage accounts, AD FS, HDInsight, Hive, Teradata appliances, Hive, Load Balancer, PowerShell, python, visual Studio, Bamboo (CICD), Bit bucket, Control-M, Docker, Jira, BI Reporting, Confluence, Agile.

**Sr. Consultant - Cloud Mar 2017 - Aug 2017**

**Microsoft Corporation, Redmond, WA**

**Product Name: SCBT-BI Max (Supply Chain Business Technology)**

* SCBT-BI Max is a modern BI platform that have a common store for data of all sizes, shapes, and temperatures (azure data lake), made available to people, platforms, and processes with as little friction as possible.
* Analyse and determine existing system architecture requirements to achieve business goals.
* Optimized Azure HDInsight HDP cluster for better performance.
* Gather detailed from existing SCBT system requirements and use them to create automation script to deploy ARM template to create HDInsight Hadoop/Spark cluster using PowerShell.
* Responsible for end- to- end ETL process by creating, Event Hubs, ADF Pipeline, Linked Services,
* Deployed HDInsight cluster, ADL and ADF.
* Involved in setting up and configured Spark cluster and run pipeline on ADF
* Designing solution for migration of data and reports from on-prim MSSQL server to HDInsight.
* Created Linked Services using PowerShell to define the connectivity and data flow using PowerShell
* Involved in windows server OS patching.
* Installed HDInsight, Spark, HBase individual clusters and administrating, optimizing, and troubleshooting and setting up security.
* Created Blob storage to store various Hive scripts and data directories.
* Created end-to-end PowerShell script to create HDInsight cluster by deploying ARM template and to create Linked Services to run hive jobs on HDInsight.
* Monitored HDInsight cluster and optimize performance.
* Bug Fixing and incident reporting on call from dev team testing team and coordinating with other stockholders.
* Written various hive and Spark SQL scripts for various operations.
* Analyse existing HDInsight clusters and optimize Hive jobs/queries
* Involved data verification, validation, and data cleaning to manage data quality.

**Environment:** Azure Portal, Azure Data Lake, ADF, Storage Account (Blob storage), Azure (AD), HDInsight/HDP 2.6, Hive, Spark, HBase, Wind. Server 2012, Spark SQL, Cosmos DB, Power BI, Net, Load Balancer, PowerShell, Ambari, AAD, Ranger, Kerberos, python, visual Studio, Power BI, BI Reporting, SharePoint, Agile.

**Sr. Bigdata Consultant – Cloud | Big Data Sep 2016 - Feb 2017**

**Dexcom Inc, San Diego, CA**

**Product Name: Victoria Data Lake**

* Dexcom Inc. is a medical device company that develops, manufactures, and distributes continuous glucose monitoring systems for diabetes management. It can show a new glucose reading every 5 minutes for up to seven days of uninterrupted wear-time to show where glucose levels are, where they have been and where they are going. The system is made up of three technologies: the sensor, transmitter, and receiver.
* Analyse and gather info to perform development tasks.
* Written various hive script to insert records in Data warehouse.
* Created various hive tables and worked closely with data scientists.
* Provided design recommendations and thought leadership to sponsors/stakeholders that improved review processes and resolved technical problems.
* Managed and reviewed Hadoop log files.
* Tested raw data and executed performance scripts.
* Shared responsibility for administration of Hadoop, Hive
* Created hive tables to store unable process data with their error msg.
* Created Spark SQL scripts to test the data quality and count matrices.
* Involved in providing inputs to create system architecture up-gradation.
* Monitoring Hadoop cluster and optimizing hive queries.

**Environment**: HDInsight/HDP 2.5, Spark, Spark SQL SQOOP, Hive, Kafka, Oozie, Azure Data Lake, Load Balancer, ADF, Blob storage, Azure (AD), Docker Container, Zeppelin, Cassandra, Wind. Server 2012, Rest API, J-Meter, python automation Scripting, SBT, Confluence, Agile, Ranger, bamboo.

**Bigdata Consultant June 2016 - Aug 2016**

**Concat System Inc., Urbandale, IA**

**Project: AVA (Audio Video Automation)**

* Concat Systems Inc. is a pioneering video technology company, specializing in creating, managing, and distributing dynamic and cost-effective online and mobile video products on the world’s largest content delivery network. Each video is designed to help our clients reach and engage their customers with richer content and greater impact that results in increased sales and ROI. Data, photos, inventory, audio, music, and special offers are automatically combined to create compelling multimedia video solutions that are fast, easy, and affordable.
* Involved in requirement analysis and gathering
* Involved in project planning and implementation
* Worked on to process image files into Sequence File and store it into the HDFS using HIPI.
* Written Spark SQL queries to manipulate Hive table data.
* Implemented Compression while storing data into Hive.
* Involved in writing various SQOOP scripts multi table, single or part table loading
* Involved in proposal and building architecture for this application.
* Created Oozie workflows and patch into the project build for scheduling all jobs.
* Manage Hadoop Cluster Size is 5 nodes.
* Involved in unit testing and SIT to maintain quality of the image data
* Involved in setting up GIT for versioning
* Used to update status and maintain documents on SharePoint

**Environment**: Azure HDInsight/HDP, ADF-Pipelines/Linked Services/Datasets, Hive2.0, Spark SQL, Cassandra, GIT, python automation Scripting, Talend 6.0.

**Sr. Bigdata Consultant Mar 2015 - Mar 2016**

**The Emirates Group, Dubai, UAE**

**Project: Helix - Emirates Customer Genome**

* Helix is an intended to be the single, reliable Data Lake repository of all real-time data relevant to the Emirates Group and its customers. The Emirates Customer Genome (EmCG) within Helix is the holistic digital and transactional footprint of each individual consumer in the Emirates Ecosystem.
* Helix follows the Lambda architecture so, there are two ways to process the data on speed layer and Batch layer
* Analyse, define client’s business strategy, and determine system architecture requirements to achieve business goals.
* Formulate strategic plans for component development to sustain future project objectives.
* Gather detailed business requirements, use cases related to Emirates Airlines customer’s record, and translate technical specifications into product requirements.
* Involved in implementing Lambda Architecture by using Big Data components as top-level project in Emirates on their existing data storage system to avail customer related data for visual analytics to generate revenues.
* Involved setting up and configured Spark Streaming and Spark SQL for Speed layer
* Develop data architecture design to facilitate targeted customer analysis.
* Implemented CDC using Hive script.
* Implemented Oozie workflows.

**Environment:** Azure CDH 5.4 HDInsight, Hive, Impala, Pig, SQOOP, Elastic Search, TIBCO, Oozie workflows, HBase, Spark SQL, Spring Boot, Kafka, Angular JS, Agile/Scrum, Micro Strategy and SharePoint QlikView.

**Sr. Analyst Programmer May 2013 - Feb 2015**

**Employer: Atos - Syntel LTD Mumbai, India.**

**Client: Daimler Trucks North America (DTNA)**

**Project: Vehicle Specs & Shop Broadcast**

* Daimler Truck North America (DTNA) is the largest heavy-duty truck manufacturer in North America and a leading producer of medium-duty trucks and specialized commercial vehicles. Daimler Trucks North America is also a leading provider of heavy-and medium-duty diesel engines and other components.
* Shop Broadcast is a DTNA’s one of the crucial applications based on Mainframe DB2 by which they track on each truck assembly line manufacturing under the future orders and have various warehouse applications like SCD-CDC, they use Tableau to generate reports on daily basis and monitor change data capture. Due to some data, storage issues in Mainframe and cost effectiveness they wanted to migrate this application on Hadoop (Hive Warehouse).
* Worked to understand DTNA existing system and architecture for data storage and report analysis.
* Involved in to gather information about where the client facing trouble with their existing system of Mainframe to get proper report generation.
* Worked on to design solution architecture for Mainframe DB2 to HDFS, Hive
* Setup Talend for ETL from DB2 to Hive platform
* Responsible for installation of SQOOP, Hive, Impala and involved in CDH4.2 installation.
* Implemented Compression while storing data in hive tables using Snappy.
* Monitored and troubleshoot issues raised while extracting data from DB2 to Hadoop using SQOOP

**Environment:** CDH, HDFS, Hive, Pig, HDFS, Impala, Java, Flume, Storm, HBase, Twitter API, JDBC, SQOOP, SQL, DB2, Impala Tableau, Putty, Talend5.x, DMX-h ETL, Eclipse

**Sr. Software Engineer Aug 2011 - Dec 2012**

**Employer: Doyen InfoSolutions Pvt Ltd.**

**Client: Brain2, Belgium**

**Project: Nike Data**

* End client is a multinational corporation that is engaged in designing, manufacturing and worldwide marketing and sales of footwear, apparel, equipment, accessories, and services. The project is about to extracted billing data from their operational systems and process using Hadoop Map reduce and store it into HBase and generate reports on top of Web UI for sentiment analysis and revenue generation and for other analytical purposes.
* Environment: Java, MapReduce mapping, HBase, Hive, SVN, Zookeeper, CDH3, Flume, Eclipse, CentOS 5.5, XML, Confluence.

**Sr. Software Developer Dec 2010 - Aug 2011**

**Employer: Icicle Technologies PVT LTD.**

**Client: CC Avenues Ltd., Mumbai, India**

**Project: Multi Currency Payment Gateway**

* CC Avenue is a payment gateway provider and provide services that authorizes payments for e-businesses, and multi-currency payment gateway, enables payment acceptance and processing in 27 global currencies.
* Environment: Java/J2EE, Servlet, EJB, JSP, Struts1.3 (MVC2), Hibernate3.0 (ORM), HTML, XML, CSS, J-Query Ajax, UI-Development, Linux, SVN Repository, MSSQL 2008, Oracle 11g, MyEclipse 7.1, JBOSS 5.x

**Software Developer Sept 2009 - Dec 2010**

**Employer: Viveka Solutions, India**

**Project: Mine Resource Management System**

* MRMS the objective of this software is to generate automated reports based on the inputs provided by the Mines data to submit as monthly/quarterly/yearly returns to various Government Mining Authorities.
* **Environment:** Java/J2EE, Servlet2.3, JSP, Struts1.2 (MVC2), EJB2, HTML, Applet, XML, CSS, Ajax, SVN, Google Maps, Linux, XAMPP, PHP-My-admin, MVC, MySQL 6.0, Edit Plus, Eclipse, Apache Tomcat6.0.1.

**EDUCATION:**

* Bachelor of Technology | Sept 2009 | Yashwantrao Chavan Maharashtra University (YCMOU)

**CERTIFICATIONS:**

* 05/2023 Microsoft Certified Azure Solution Architect Expert