

# MUKUL TANWAR

Linkedin: <https://www.linkedin.com/in/mukul-tanwar-02573852>

Skype: Mukul Tanwar (live:.cid.dca530742a2f25d9)

**Title:** Sr. Azure Data Architect/Databricks Architect  
**Location:** Carmel, IN  
**Languages:** English



## Experience Summary

- **Over 14+ Years of IT experience comprising Data Architect, Big Data Architect, Data Engineering, Design & Develop Applications, Business Intelligence, ETL, Databricks Architect, Data Governance and Data Engineer Pipeline.**
- **Architect** processes and data flow for Data Management Platform (**DMP**), Consumer Data Mart (**CDM**), Consumer Data Platform (**CDP**), Data Mart(**DM**), Analytic Hubs (Content, Taxonomy and Audience), Delta Lake House and Data Warehouse.
- **Create conceptual, logical and physical plan** for data ingress and egress for DMP.
- **Educate and enlighten** the teams on how to **design and process** data in **distributed computation** while **reducing the cost despite increasing data load**.
- Conducted POC on stacks for **better and faster performance and usability**.
- **Architected data pipelines for processing streaming and batch data** for data size varying from **~100 Gigabyte to ~500 Gigabyte** compressed per hour and per day process **~50 Terabytes** of data.
- Analytics pipelines designed to process data in **Petabytes** scale of data.
- **Architect the designs to reduce the cost** of the entire tech stack on cloud.
- Expertise in **designing and coding ETL process** on enterprise level.
- Follow agile practices in development life cycle.
- Hands on experience in **relational and non-relational databases**.
- Architected processed for **semi-structured, structured and unstructured data** for huge media houses processing data from multiple sources.
- Create/Implement/Design data models (conceptual, physical and logical), data governance (data sanity, data lineage, data lifecycle management, data policies, data dictionaries), delta lake house, data marts and data warehouses.
- **Compliance** under HIPAA, PCI and **Risk Management** applying Data Governance to maintain data policies, data dictionaries, data lineage, data life cycle, data access policies, data sharing and schema enforcement.
- Comply with the data privacy policies (**GDPR, CCPA**).
- Designed and coded **applications for sharing** data across partners.
- Implement Data Ware housing concepts by removing data silos on **implementing feature stores**.
- Work with **Data Scientist, Data Analysts and Data Engineers** closely on data flow and design principles.
- Design tools and components that are used for data sharing across regions.
- Multi Zone Architecture to tackle Disastrous Recovery and Zone failure problems.
- Have extensive experience in **Databricks** where leveraged **databricks** for processing data on streaming data and batch data.
- Implemented data governance using **databricks unity catalog**.

Mobile :: 6827069397

Mail :: mukultanwarcs@gmail.com

- Developed and Designed the entire **ETL pipeline on databricks on Azure Cloud Solution.**

#### KEY SKILLS:

Database	DWH SQL Server, MySQL, ClickHouse, PostgreSQL, No SQL, HBase, Cassandra
Data Processing	Spark, HDInsight, HDFS, Virtual Machines, HQL, <b>Databricks(Spark Cluster, Notebooks, Jars)</b>
Operating Systems	Linux(Ubuntu, Centos), Mac OS, Windows
Frameworks	Spring, Spring boot
Cloud	<b>Azure HDInsights, Spark Cluster, Azure Redis, Azure Event Hub, Azure MySQL, Azure Synapse, Azure Blob Storage, Azure Admin, Azure CosmosDB, Azure Kubernetes, AWS S3, AWS Kinesis, AWS Glue</b>
Web/Application Server	Apache Tomcat
Programing Language	Python, Scala, Java, PySpark
Big Data Tools	Hadoop(Hortonworks and Cloudera), Hive, Ozzie, Flume, Scoop, Zookeeper, Ambari, Yarn, Tez
	JUnit, Mockito, Log4j, Maven
Workflow Management Tools	Service Now
CICD	Jenkins, Terraform
BI Reporting Tools	Tableau, Business Objects, Xcelsius, BIRT
Orchestration Tools	Apache Nifi, Ozzie, WorkFlow
Scripting Language	Shell Scripts, Java Scripts, Python
Others	Redis, Aerospike, Memcached, Kafka, <b>Databricks, Spark, Workflows, Unity Catalog, Admin, Cluster Configurations, Notebooks</b> , Jira, Looker

## Experience

**01 Type of Industry:** Vyts, INC  
**1 Client:** Verizon  
**Project Names:** Network Team

**Role/Title:** Lead Data Architect/Engineer

**Project Duration:** 6 Months July`23 to December`23

**Responsibilities:** Data Architect, Data Engineering, Application Development, Rest Base API, UI. Educate the team on optimization techniques where how should we process data with efficiency leveraging Flink,Pulsar,Kafka with Java Spring Boot Applications.

- Architect, design, build and implement data pipelines for processing data from multiple sources.
- Optimised pipelines for running on huge volume of data and propose new architecture for the entire flow.
- Design and implemented Applications on Flink, Pulsar, Kafka for streaming data and consuming data where the downstream pipelines consumed data and used in model training.
- Implemented the process and pipeline to consume data from streaming engine(Kafka/Pulsar) and push notification on mobiles for registered users on any breakage or activity planned events.
- Implemented the pipelines for data science project where had designed the process for ML.
- Developed the process to get the predictions from model and rerun the model on the learnt feedback from users.
- Developed/Coded services to process streams of data in Java Spring Boot services the shared this data with other microservices as per their business case.
- Developed/Coded streaming pipelines using Java Spring Boot to process and transform data from one ecosystem to other.
- Architected and Implemented best practices for development and packaging the solutions.
- Kubnernetes for deploying and scheduling jobs.
- Manage the master data in GCP .

Designed, Architected and implement their pipelines leveraging Java Spring Boot Application using all the components like **Kubernetes, Flink, Pulsar, Kafka and MLFlow**

**Special Tools/Env:** Java Spring Boot, Pulsar, Kafka, Flink, GCP, FCM, ArgoCD, Kubernetes, Docker, Looker, Druid.

2 **Client :** JPMC

**Project Names:** Databricks Architect (on AWS)

**Role/Title:** Sr. Data Architect/Databricks Architect

**Project Duration:** 1 Month April`23 to May`23

**Responsibilities:** Data Architect, Data Governance, Data Engineering, Data Warehousing, Delta Lake House, Application Development, Rest Base API, Databricks. Educate the team on migration hurdles and optimization techniques where how should we process data with efficiency leveraging databricks.

- Presented databricks performance comparison to the native system being used in the organization.
- Helped them in **tuning the jobs on Databricks.**
- Helped them in setting up **Clusters.**
- Educate them on **Ganglia graphs and how to use them.**

- Helped them in understanding the power of using **Unity Catalog**.
- Designed and architected their pipelines leveraging Databricks using all the components like **DBSQL, WorkFlow, Spark Clusters, Scheduling jobs from outside Databricks env, Jobs Cluster, External Tables and MLFlow**.

**Special Tools/Env:** Scala, Python, Databricks.

**02 Type of Industry:** **Media House (Mediacorp Pte LTD)**

**Project Description:** **Consumer Insights Analytics (CIA)** is a team responsible for proving intelligence on the clickstream data. Mediacorp has 8 sites and one OTT platform where articles, videos, audios and images are hosted and traffic lands from all part of world. This data generated by users either impressions or clicks flows into system as streams of messages or batch files on periodic interval. The data generated by these events were stored and processed to draw out intelligence to get user demographics and draw predictions for digital marketing team. This data was used further for decision making, future planning, forecasts and ad-targeting via campaigns.

**Project Names:** CDP(Consumer Data Platform), Ripple(Content Analytic Hub), Meid(Audience Analytics Hub), Merewards, CX360, Inbox Targeting, CIfEeds, Taxonomy, Majority Voter, Model Prediction, Benchmarking.

**Role/Title:** Sr. Azure Data Architect / Azure Data Architect

**Project Duration:** 3 years 4 months November`19 to February`23

**Responsibilities:** Data Architect, Data Modelling, Data Governance, Data Engineering, Data Warehousing, Delta Lake House, Application Development, Rest Base API, Databricks, Azure Cloud Technology, Design Patterns, Develop, enforce best practices, costing, explore new and advance technologies, planning and future road map.

- **Architect, design, build and implement** data pipelines for processing data from multiple sources.
- **Architect and Design** processes on **CosmosDB** to process CMS feeds for sites to run data science models on it.
- **Designed** processes to stream data into **EventHub** to run recommendation for user on **near real time**.
- Stored all the metrics on **Azure Synapse for Analytics hub**.
- Written application that processed data ingress-outgress on **Azure Redis, Azure SQL, Azure MySQL, Azure Postgres**.
- Designed and developed complex ETL and ELT jobs using **Spark, Databricks, HDInsight's & Java Based Applications**.
- Processed **structured, semi-structured, unstructured, clickstreams and clickstream batch** data.
- Rearchitect the processes to **reduce the processing time by 60-90%**.
- Rearchitect the application to reduce response time from **1.13 milliseconds to 0.09 milliseconds**.
- **Migration of technology from HDInsight's to Databricks**.

- **Designed/Created delta lake house** that saved too much **processing and storage** which reduced cost on both side processing and storage.
- Work with **Data Scientists in designing pipelines and model** deployment and help them to optimize their codes too.
- **Developed tools to transform data in warehouse** for data pipelines and process data in a short time.
- Constantly evaluate and redesign the codes to process data in **less time and efficiently** to avoid pipeline interrupts.
- Designed the processed to **auto heal themselves**.
- Developed/Designed the **entire pipelines on Scala/Python**.
- Worked on **Data Governance and Data Quality** to get alerts on data changes.
- Created **Data lineage and Data analysis** documents.
- Work on budgeting for the team.
- **Define appropriate Testing (Unit, SIT, Performance, and UAT), Quality Assurance, and Quality Control Strategy**.
- Develop test plans, test cases, test scripts, and test validation data sets for all the newly built interfaces.
- **Multi Zone Data replication** for keeping data on multiple zone for **DR and Zone Failure** issues.
- Follow scrum/agile methodologies for running team productivity.
- **Process clickstream data** batch and streaming data.
- **Architect/Design** process for feature stores for feature engineering for the models in the org.
- Introduce **Databricks into ecosystem** after driving POC, shared POV to get shifting processing from HDInsight's to Databricks.
- Worked on Data Science Pipelines using **PySpark**.
- Processed data in Kafka for some pre-aggregates.
- Worked on **Looker** in exploring tables and data.
- Handled migration from HDInsights processing to Databricks, migration from Azure Synapse to Clickhouse hosted in house and migration from MySQL to Cosmos DB to handle cost and performance.

**Special Tools/Env:**

Scala, Python, Java, Shell Scripts, Python Scripts, Azure Cloud Solutions(Blob, Redis, CosmosDB, VM, EventHub, HDInsights MySQL, DWH SQL, Azure AD Portal), DataBricks(Spark Clusters, Workflows, Unity Catalog, Admin, Computes, NoteBooks), Java Spring Boot, SQL ClickHouse, Apache Nifi, MySQL, Redis.

**03 Type of Industry**

**Media House (Times Internet Limited)**

**Project Description:**

**Data Management Platform (DMP)** is a platform for generating, processing and storing clickstream data. Further this data is used for drawing intelligence and serve to the Ad-engine. Audience creation, updation and maintenance and to serve to Ad-engine was responsibility to DMP. The data stored was used to merge data from partners. Unify users based on their previous interactions.

**Project Names:**

CDP(Consumer Data Platform), DMP Service, DMP Spark Streaming, Data Marting, User Unification, Pixel Manager, Auto Categorizer, PCR, Partner Data Processor, Data Sharer.

**Role/Title:**

Data Engineer/ Big Data Developer

<b>Project Duration:</b>	4 year 5 months	July`15 to November`19
<b>Responsibilities:</b>	<p>Data Architect, Data Modelling, Data Engineering, Data Warehousing, Data Lake, Application Development, Rest Base API, Big Data Architect, Design Patterns, Develop, enforce best practices, explore new and advanced technologies, planning and future road map.</p> <ul style="list-style-type: none"> <li>➤ <b>Architect, design, build and implement</b> data pipelines for processing data from multiple sources.</li> <li>➤ Designed and developed complex ETL and ELT jobs using <b>Spark, Hive, HBase, Cassandra, Aerospike, Redis, Memcached &amp; Java Based Applications.</b></li> <li>➤ Design/Architect the extraction process to <b>generate clickstreams from sites.</b></li> <li>➤ <b>Designed/Created data lake</b> to store all data structured and unstructured.</li> <li>➤ Developed/Designed products to capture user journey, unify users from various platforms, consume data from all platforms and process it for analytics.</li> <li>➤ <b>Architected/Designed a self-service</b> tool for creating pixel and get it deployed on the site.</li> <li>➤ Designed, built, and implemented data engineering solutions on HDFS, Cassandra, HBase, Aerospike and Redis.</li> <li>➤ Designed and developed complex ETL jobs for extracting, transforming, integrating, and loading data into on-prem HDFS Cluster.</li> <li>➤ Worked health care data under setting compliances.</li> <li>➤ The entire application was pushing data on Kafka and further my pipelines were processing the data using spark for transformations.</li> <li>➤ Migration from HBase to Cassandra and Hive jobs to Spark Jobs.</li> </ul>	
<b>Special Tools/ENV:</b>	Java, Spring Framework, Spring Boot, Java Spark Streaming API, HDFS, HQL, AQL, SQL, Hive, Spark, Cassandra, Solar, HBase, Aerospike, Redis, Kafka, Memcached, Java Scripts, Shell Programming, AWS S3, AWS Kinesis	

<b>04 Type of Industry</b>	<b>Travel Industry (Expedia)</b>	
<b>Project Description:</b>	<p><b>Enterprise Data Warehouse (EDW)</b> is a team responsible to develop and design new tools. The team was responsible to enhance the tools and develop new utilities in the tools to solve business problems. Explore and Develop new architectures to solve time complexity. The tools designed where for analytics and data processing.</p>	
<b>Project Names:</b>	Enterprise Data Warehouse, Consumer Data Mart(CDM), Service Health Monitoring.	
<b>Role/Title:</b>	Data Engineer	
<b>Project Duration:</b>	2 years 11 months	August`12 to July`15
<b>Responsibilities:</b>	<p>Data Engineering, Data Warehousing, Application Development, Big Data Developer, Develop, explore Cloudera and Hortonworks new products and feature.</p> <ul style="list-style-type: none"> <li>➤ <b>Develop, design, build, and implement ETL/ELT frameworks and solutions.</b> Codes the entire pipeline in <b>using Java and Hive HQLs.</b></li> </ul>	

- Coded a suit where **user transaction and behavior data was recorded and updated** to the latest timestamp of user interaction to help the customer care support team.
- **Develop and design the entire process of monitoring** the system health where we had all the **Virtual Machines, Java Applications, Windows Servers, SQL and other components that were used in the ecosystem.**
- Executed the **migration from relational database to Hadoop.**
- Worked on **migration from Cloudera to Hortonworks HDFS.**
- **Developed and Designed Systems to ETL** all the data for windows servers and SQL components to get data and record all the time series data and alert on stack health to the respective teams automatically.
- Developed and enhanced the **data processing tools** for developers.
- Built processes supporting **data transformation, data structures, metadata, dependency, and workload management.**
- Develop **stored procedures/views for loading Dimensions and Facts.**
- Wrote **complex SQL queries and stored procedures.**
- Help teams to schedule jobs on **ozzie**, mentor them on using **hue, hive, scoop, ambari**, etc.
- Prepared **technical design documents and source to target** mapping documents.

**Special Tools/ENV:** Java, Hadoop, Hive, Ozzie, Flume, Scoop, Splunk, MSSQL Server, MySQL, HBase, Java Scripts, HQL, SQL, Django Framework

**05 Type of Industry:** **Telecom and E-marketing (One97/Paytm)**

**Project Description:** **Data Warehousing (DWH)** is an team responsible to process data and store data. This data was used for analytics and revenue assurance. The objective of team was to reduce cost on processing data and do the ETL and make the data meaningful by storing it in relational tables and serve it business users. The data was published and was made visible to users by report tools like Business Objects, Tableau, BIRT and Xcelsius Dashboards.

**Project Names:** Data Warehouse, ETL Team, Revenue Assurance, Business Intelligence Team.

**Role/Title:** Data Engineer

**Project Duration:** 2 years 8 months December`09 to August`12

**Responsibilities:** Data Engineering, Data Warehousing, Application Development, SQL Procedure, Business Intelligence Developer, Develop and ETL Tools and Concepts.

- **Business Objects: Slicing and dicing, Designer for universe creation, Business Objects XI- Web Intelligence** for report creation and CMC for user maintenance for BO.
- Involved in **Universe creation.**
- Involved in **Web intelligence/Desktop Intelligence** reports creation.
- Scheduling reports.
- Involved in Coding and Unit Testing.
- Involved in Code review.

- Creating **backend summaries for Reports.**
- **Business Intelligence and Reporting Tool:** Creating rpt designs for Managements reports and Operator specific reports. Creating multiple data sources for creating reports from multiple databases from multiple servers and joining multiple datasets.
- Code stored procedure based datasets.
- **Enhancing and modifying python** code according to client requirements.

**Special Tools/ENV:**

Business Objects XI R3.1, Designer for universe creation, Web Intelligence for report creation and CMC for user maintenance for BO, MySQL, Business Intelligence and Reporting Tool, Python, SQL, Shell Scripts

## Education

---

Bachelors in Computer Science & Engineering in 2009 from Uttar Pradesh Techincal University, Uttar Pradesh, India.