Shishir Jaiswal

Cell# +1-647-674-7422
Email: shishirjaiswal@yahoo.com
Linked-In Profile: http://www.linkedin.com/in/jaiswalshishir
AWS Serverless Website: www.shishirjaiswal.com

PROFFESSIONAL SUMMARY

- Seasoned professional with rich experience (**17+ years**) in architecting, designing and developing high performance Big Data Engineering and Data Lake solutions with **Amazon AWS, Python, Apache Spark, Apache Kafka, Kubernetes, Apache Hive, Java, Apache Hadoop, Ab Initio, Informatica, Teradata, Oracle, DB2**.
- Certified **AWS Solutions Architect**
- **17+ years'** experience in data warehouse, **data engineering** and **data lake** solutions on Teradata and **Apache Hadoop** for Capital Markets, Personal Banking, Credit Cards domains.
- **12+ years'** experience in **Data Modeling** and deployment of multi-dimensional schema (**Star and 3NF Schemas**), **Conceptual/Logical/Physical design, Entity Relationship modeling**.
- Built own **AWS Serverless** website: www.shishirjaiswal.com
- Designed **real-time streaming** solution (including frontend, backend, middleware, analytics) for capital market regulatory trade reporting and analytics for Scotiabank
- Designed and developed Market Data Scenario Service on **Spark, Hadoop** with **compute burst on Spark on AWS**.
- Designed and developed **Risk Data Lake solution on Hadoop, HDFS, Spark, Hive and Airflow** for RBC Market Risk LOB.
- Designed and developed **metadata driven data lake ingestion framework** for JP Morgan Chase's.
- Designed **data quality controls and measurement processes for data lake**.
- Designed and developed Quantitative Analytics framework on **PySpark, Hadoop** for quants to **run analytical models directly on big data clusters**.
- Designed **streaming data lake using Kafka, Nifi, ElasticSearch, Spark on Kubernetes and MinIO** for Scotiabank.
- Previously headed Cognizant's Ab-Initio Center of Excellence in North America for Banking and Financial Services customers.
- Authored blogs on ETL performance tuning and best practices.
- Strong experienced in **agile project execution and DevOps** processes.
- Experienced in various process/**performance improvement** methodologies (six sigma, fish-bone analysis, pareto analysis etc.)
- Well versed with PCI DSS (Payment Card Industry Data Security Standards).
- Team worker with excellent interpersonal skills.

CERTIFICATIONS

AWS Certified Solutions Architect – Associate (Verify@ https://www.youracclaim.com/go/OnRGqJADJWITEzIwW0JhDQ)	01/2021
Azure Certified Fundamentals (Verify@ https://www.youracclaim.com/badges/e5ca54bf-41cc-4f2c-9b19-e3897b8cf9b3/public_url)	02/2021

SKILLSET

Skill Category	Skills	Exp. (Yrs.)
Cloud Technologies	AWS: EC2, S3, Kinesis, Lambda, Athena, Glue, DynamoDB, RDS, IAM, VPC, Security Groups Azure: VM, BLOB, SQL Server	1.5
Containerization	Kubernetes, Docker	2

Big Data	Apache Kafka, Spark (PySpark, SparkSQL, Dataframes), Hadoop, Hive, HDFS, HBase, Sqoop, Nifi, Airflow, ElasticSearch, MinIO, Parquet, Avro, ORC	5+
BI	Business Objects, Tableau	5+
DevOps	Jira, GitHub, Jenkins, Urban Code Deploy, SonarQube, SonarLint	5+
	AWS CodePipeline, AWS CodeBuild	1
Databases	Teradata, DB2, Oracle, Postgres, MySQL, Erwin Data Modeler	17
ETL Tools	Ab Initio, Talend, Informatica	15
Languages	Python	5+
	SQL, PL/SQL	17
	Java	2+
	Unix/Linux Shell Scripting	15
Integration	REST API	2+
OS	RedHat Linux, Centos, Ubuntu, Windows, Solaris	17

PROFESSIONAL EXPERIENCE

Senior Solutions Architect – Data and Analytics 06/2021-Till Date
Amazon Web Services

Solutions Architect 01/2020-05/2021
Scotiabank (Bank of Nova Scotia)

Principal Architect (Associate Director) 11/2010-12/2019
Cognizant Technology Solutions Canada

Programmer Analyst 05/2010-10/2010
Empower Technology Solutions Inc.

Systems Analyst 12/2003-04/2010
Satyam Computer Services Ltd.

Graduate Engineer Trainee 06/2003-11/2003
Motherson Sumi Systems Ltd.

PROJECTS

Scotiabank (Bank of Nova Scotia), Toronto, ON, Canada
Role: Solutions Architect 01/2020 – 05/2021

Tools: Python, Kafka, ElasticSearch, MinIO, Nifi, Apache Spark, Docker, Kubernetes, Java, Dremio, Postgres DB, Linux, Jira, Github, Jenkins, Accelerator CI/CD, Airflow

Regulatory and Compliance Technology Architecture

Scotiabank's Global Regulatory and Compliance Technology group is responsible for delivering Capital Markets regulatory and compliance reporting business capabilities for all jurisdictions across the globe.

Responsibility

- Lead Architect for Regulatory and Compliance IT projects.
- Designed and developed streaming data platform architecture on Kafka, ElasticSearch, MinIO, Nifi, Apache Spark on Kubernetes.
- Design end to end IT solutions (including front end, back end and data management systems) for complying to regulatory reporting requirements in globally distributed jurisdictions.
- Work with project teams to conceptualize, design, develop and deploy the solution.
- Present the solution architecture to senior management team (VPs, SVPs), Architecture Review Board.
- Define and standardize HA and DR strategy for the application.
- Conduct design and architecture reviews to ensure reusability, consistency and adherence to reference architecture, standards and conventions.

Client: Royal Bank of Canada, Toronto, ON, Canada

Role: Solutions Architect 09/2016 – 12/2019

Tools: AWS, Java, Python, Spark, Hadoop, Hive, Sonarqube, Docker, Jira, Github, Jenkins, Urban Code Deploy, MS SQL Server, Linux, AWS (S3, Lambda, IAM, SQS, Athena, Spark Standalone on EC2)

Project: Fundamental Review of Trading Book

FRTB is a set of regulations by OSFI as part of Basel framework for the next generation market risk regulatory capital rules for large, internationally active banks. The regulation requires the banks to calculate and report capital charge using different methodologies – Standardized Approach and Internal Models Approach. Market Data Scenario Service is a strategic store to source, store and enrich Risk Factors, Base Market Data and generate stress scenarios for Risk Analytics and Capital charge calculations.

Responsibility

- Architect the Market Data Strategic Store on big data platform using Spark as the computation engine.
- Designed PySpark based Quantitative Analytics Framework on Spark and Hadoop for quants to run analytic models directly on Hadoop cluster.
- Study the legacy and current platform and identify the short falls and design the new platform to overcome the same.
- Ensure optimal balance between configurability, scalability, maintainability of the application.
- Assess various design approaches for scalability, flexibility, maintainability, consistency and robustness and pick the best match for the solutions.
- Define and standardize HA and DR strategy for the application.
- Conduct design and architecture reviews to ensure reusability, consistency and adherence to reference architecture, standards and conventions.
- Lead the delivery team to develop, test and deploy the application.

Client: JP Morgan Chase, Columbus, OH, USA

Role: Solution Architect 05/2010 – 08/2016

Tools: Ab Initio, Informatica, Erwin, Hadoop HDFS, Hive, Python, Java. O/S: Linux, MPP, Control M

Project: CCB Data Ecosystem (Ab Initio, Hadoop)

CCB Data Ecosystem is the new platform built on Hadoop – Ab Initio cluster to host ETL applications. The platform is architected to deliver high performance, and be linearly scalable yet cost effective. The platform is also designed to seamlessly capture and integrate the end-to-end metadata in a single enterprise-wide repository.

Responsibility

- Design and develop metadata driven data lake ingestion framework.
- Work with IT teams to understand and define the technical requirements from the ETL platform.
- Study the legacy and current platform and identify the short falls and design the new platform to overcome the same.
- Define and standardize HA and DR strategy for the platform.
- Define and standardize platform governance processes.
- Define reusable components in Ab Initio, Hadoop ecosystem for ETL and other analytical processing.
- Design micro-services based DWH Controls Dashboard for SLA monitoring and automated diagnostics mining.
- As part of Chief Development Office, designed and established Data Warehouse design and coding standards and review processes.

Project: Integrated Customer Data Warehouse – Channel Data Marts

ICDW Channel Datamarts contain data related to interactions between the bank and the customers through various channels – Online, Mobile, Phone, IVR, Teller, ATM, POS transactions (monetary as well as non-monetary). The Channel data marts provide wealth of data for Marketing, Risk and Operations to perform analytics related to marketing campaigns, product development, fraud and operational readiness.

Project: Chase Wealth Management Datamart

Chase wealth management datamart is a central repository of all affluent customers and their investment portfolios. It provides a 360 degree view of all assets and holdings of each affluent customer. This enables efficient tracking of money in/out flow of the bank. Also, it is helpful in consolidating sales made by advisors and hence is used in determining advisor compensation and benefits.

Project: Online and Mobile Banking Datamart

Online and Mobile Banking data mart contains all the information related to online and mobile banking customers, their preferences and settings and transactions. The data mart is used by the business to perform analysis related to usage and popularity of online/mobile banking products and features. It is also used to determine effectiveness of certain marketing campaigns and to identify opportunities for improvement and introducing new products.

Responsibility

- Interact with business analysts and business users to understand business process flow and identify key attributes needed for reporting.

Define the overall application architecture, attribute naming convention, and create the conceptual, logical and physical data models to suit business reporting needs and efficient data retrieval.

- Conduct review sessions with architecture review board to obtain necessary approvals.
- Guide the development team and ensure the development happens as per the reference architecture.
- Design the end-to-end ETL application architecture.
- Design robust data audit framework (job audits, data capture audit, reject acceptability) to ensure accuracy of data.
- Tune the BO report SQL to use optimal query path.

Coordinate and guide the team during testing and interface with business during UAT.

Client: GE Money, Stamford, CT, USA

Role: Technical Lead 12/2003 – 04/2010

Tools: Abinitio 2.14, Business Objects XI R2, Database: Oracle 10g, O/S: Sun Solaris 10, MPP, Toad

Project: JC Penney Digital Cockpit

JC Penney Digital Cockpit is a extranet site that provides a dashboard of daily performance of Store Managers and Associates. The dashboard is used by JC Penney to track and Reward the Associates on a daily basis.

Project: Actimize Anti Money Laundering Application Data Feeds

Actimize is an anti money laundering application that tracks high value cash payments and credit balance refunds to and from a credit card account and raises money laundering alerts for suspicious transactions. GE Corporate uses this tool for compliance with Patriot Act regulations.

Project: Account Level Profitability Data Mart

Account Level Profitability datamart stores the profitability decile and percentile data for each account. The data is provided by Argus on a monthly basis. Business uses this data to determine profitability of accounts, portfolios and reward profitable accounts by waiving late fees etc.

Project: Strategy Optimization

Strategy Optimization process is used by Collections business to use predictive modeling techniques to optimize the collections dialer campaigns and strategies so as to collect maximum amount from delinquent customers. The aim of this project was to consolidate the variables required by predictive analysis tool from across different Datawarehouses and OLTP systems and provide them in a single repository.

Project: Mail Tracker Application

Mail Tracker application tracks the remit coupons mailed in by card holders along with their monthly payments and removes the delinquent accounts from automated dialers to avoid making calls to accounts that have already made a payment.

Project: Collections Data-Warehouse Reengineering

The Collections Datawarehouse was running on legacy platform using outdated ETL and reporting technologies. The aim of this project was to discover, performance tune and eventually reengineer the datawarehouse using abinitio as the ETL tool and Oracle 10g as the Database.

Phase I: Discovery Phase

- Study the existing system using six sigma methodology
- Collect the time and data volume statistics on the performance of ETL jobs.
- Perform statistical analysis (pareto and fishbone) on the collected data.
- Identify and prioritize the pain areas.

Phase II: Performance Tuning Phase

- Analyze the identified code/architecture for performance bottlenecks.
- Tune the SAS, PL/SQL Autosys code and improve the ETL architecture.
- Design, develop and test the code.
- Preapare the IT RDD, design Documents, test plan, test cases, FMEA, implementation plan and Oversee the quality procedures.
- Measure and verify the improvments.

Phase III: Reengineer the Datawarehouse using Abinitio.

- Creation of Abinitio ETL environment from ground up.
- Complete Re-Design ETL processing of 13 interfaces in Ab-Initio.
- Re-write and tune the logic from PL/SQL procedures into abinitio graphs.
- Work with DBAs to improve partitioning and indexing strategy.
- Preapare the IT RDD, design Documents, test plan, test cases, FMEA, implementation plan and Oversee the quality procedures.
- Estimation of UNIX and Database Space Estimates
- Perform end to end testing
- Implementation of the code on the production server.
- Deep dive presentation
- Worked closely with users during UAT
- Meta data management using Abinitio EME

Project: DB Controls ETL and Audit Reporting with Generic BO Architecture

The DB Controls project was intended to implement tight process audits in the ETL process as well as to gather the hardware performance metrics. The data from these audits was loaded into a separate repository from which daily audit reports were generated for IT team to analyze.

TRAININGS

S. No.	Title
1	Emerging Program Managers – Stanford University
2	TOGAF 9
3	Machine Learning and Data Science Training – Stanford University
4	Python Practitioner – University of Michigan
5	Confluent Kafka Training

TECHINICAL PAPERS/BLOGS

- Big Data Process Performance Tuning – A structured approach
- How to build a Serverless Resume website on AWS

EDUCATION

B.Tech.
GB Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India July 1999 to June 2003