

BISWAJIT BORUAH

Director, Data Engineering at Prudential | Ex VP at Goldman Sachs

LinkedIn - <https://www.linkedin.com/in/xanoids/>

201-889-3430 | xanoids@gmail.com

PROFESSIONAL SUMMARY

- ✓ **I am a visionary Data Engineering Technical Leader with 15 years of domain experience including 8 years of high visibility leadership roles.** Expertise spans across Data Architecture, Cloud Transformation and Strategic delivery of Data Engineering, Data Warehousing, Data Science, and Data Analytics projects. Experience across AWS Services and Eco system, Spark framework, Microsoft Fabric, Kafka, Snowflake, Databricks, ETL tools viz. Informatica Power Center, PL/SQL, Python programming, Hadoop, Unix Shell scripting etc. End-to-end responsibility of strategy, planning, design, development, data transformation, data quality and delivery of Data Engineering, Data Science and Data Analytics projects.
- ✓ Spearheaded the **strategic design and implementation** of Data Engineering pipelines for AI/ML and **Large Language Model (LLM)** applications, including data ingestion, curation, and feature engineering for **LLM training and inferencing**. This involved architecting scalable solutions for processing vast datasets to build and deploy chatbots and other generative AI initiatives, ensuring data quality, model performance, and efficient data flow within the AI/ML ecosystem.
- ✓ 14 years of extensive experience with Data Engineering Teams in financial orgs. like **Goldman Sachs, JP Morgan Chase, Morgan Stanley** while integrating, modelling, and developing data warehouses involving various data complexities to handle huge volumes of data to be used by Analytics and Business Intelligence purposes for strategic decisions in both on-prem and Cloud systems.
- ✓ Experience in uplifting code and data pipelines from **On-Prem Traditional Data Warehouse (Oracle/RDBMS and HDFS) to Cloud Native Framework (AWS S3 and Snowflake)** using AWS, Python, Spark, Kafka, Data Bricks, Snowflake etc. for better scalability and performance while ensuring continued availability of Data and Analytics to Business. This involved migration of 10 years' worth of historical data as well as development of delta pipelines for processing and loading on a go-forward basis.
- ✓ More than 8 years of experience in **leading teams Data and BI Engineers** located across multiple geographical regions. Actively involved in coding, designing ETL solutions, end-to-end framework, code reviews etc. along with the team.
- ✓ In-depth experience working with Traditional **ETL Tools like Informatica Power Center 8.6, 9.5, 10.1 etc.**, RDMS systems like Oracle, Teradata, Postgres etc. while designing and ingesting data from multiple data sources and ensuring better performance and data availability timelines.
- ✓ **Technical Product Ownership & Stakeholder Management** - Actively involved in **planning and discussion with Business Stake Holders** to understand the goals/commitments for the division. Monthly/Quarterly presentation to leadership showcasing the Data Engineering Ongoing efforts, plans on Optimization, scalability etc. This primarily focuses on benefits/gains with Cloud Migration, cost savings, user satisfaction, batch optimization, data availability and readiness metrics.
- ✓ Experience in **designing end to end solutions** from gathering business requirement, identifying the potential data sources, designing the flow of data ingestion, Data modelling, developing the solution by fetching data from the source systems and applying business logic, writing the output to storage layers in Cloud and generating Analytical reports using the refined data to be used by Business functional Teams.
- ✓ Worked in **Tech Risk and Regulatory Projects/deliverables** while ensuring data quality and adequate data controls are in place for all High Tier Data ingestions and pipelines. Experience in **handling PII data** involving different encryption methods while ingesting into Cloud to ensure data integrity.

PROFESSIONAL EXPERIENCE

Director, Data Engineering

Employer – Prudential Insurance Company of America
213 Washington Street, Newark, NJ 07095

Sep 23, 2024 – Current

I am currently working as a Director, Data Engineering at Prudential Insurance Company of America responsible for Planning, Strategy, Design and Architecture of Projects in Data Engineering and Data Science domains. I am responsible for leading a Team to build and optimizing data pipelines, ingestion logic with latest coding practices including modern design patterns and architectural principles, while removing technical impediments. Actively conduct complex data analysis and report on results, prepare data by combining raw information from different sources and aggregating the same to achieve the desired business deliverables. I develop and review high quality, well documented and efficient code adhering to appropriate standards using python, SQL, pyspark, shell scripting etc. I build data pipelines using AWS services to source data from multiple different source types and source systems, transforming, aggregating the results and finally loading into storage and reporting layer for use by different product and business teams. I oversee the strategic integration of large language models (LLMs) and advanced AI/ML capabilities into our data products, driving innovation in areas like natural language processing and intelligent automation. I am also involved in research and development of solutions to address the issues and problems identified by quality assurance teams as part of unit and integration testing. I provide technical directions to team members on relevant and emerging technologies to ensure innovation, efficiency in the product solutions and impact on future concepts, products and technologies. I actively collaborate with data analysts, data scientists and data architects on different projects to ensure optimal data architecture, efficient ingestion flows, data governance, transformation, data reliability and data quality.

Data Engineering Tech Lead

Employer – Goldman Sachs 200
West Street, New York, NY

Designation – Vice President

Feb 11, 2019 – Sep 22, 2024

I worked as the Data Engineering Tech Lead for Asset and Wealth Management (AWM) Division at Goldman Sachs. Responsibilities include Design, Data Modelling, Develop and manage Data Ingestion and transformation workflows, Data Analytics, Data Quality and Project Delivery. Lead a team of 12 Data and BI Engineers and guide them in successfully completing the project deliverables. This includes hands-on end-to-end development as well as design/code review for the Team. Technologies include Python, Spark, Informatica Power Center, PL/SQL, Cloud Technologies such as AWS Eco System, Kafka, Data Bricks, Snowflake etc. to onboard new datasets and enhance existing data flows for use by Business Analytics Teams. Data is sourced from multiple different source systems, which are then extracted and transformed as per business logic and finally loaded into Data warehouse for Analytics, BI and Reporting purposes. On Prem warehouse is in Oracle and Postgres, while Cloud data Storage is in AWS S3 (parquet/Iceberg File formats) and in Snowflake. Also design operational reporting leveraging PL/SQL and Snowflake to produce tactical operational reports for use by Business Teams. Additional responsibility includes Migration of existing on premises Data Warehouse to Cloud Data Lake by uplifting the code and storage to be Cloud Native.

Data Engineering Lead**Employer - JP Morgan Chase & Co.**

545 Washington Blvd, Jersey City, NJ,07310

Designation – Associate

April 3, 2017 – Feb 10, 2019

I worked as a Data Engineering Lead along with being an individual ETL Developer for JP Morgan Chase's Private Wealth Management Data Engineering Team. Job responsibilities include developing, designing, review code to extract data from various data sources, transforming as per business rules and loading into Investment Risk data marts in Oracle and Greenplum. It involves ETL Tool - Informatica Power Center, Oracle, Python, Spark, Hadoop, PL/SQL Scripts, Unix for transforming and formatting the data and finally loading it into the data marts.

Data Engineer and ETL Developer Employer -**Mphasis**

Jersey City, NJ

Designation – Module Lead Systems

June 20, 2016 – Apr 2, 2017

I worked as a Data Engineer for the Risk Analytics Project responsible for generating Analytical Risk Metrics based on the data loaded into the warehouse. The data pipelines read data from various internal/external data sources E.g., Oracle, Flat Files, Json etc. and load into Investment Risk data marts. XML files are then generated on top of these data mart tables incorporating additional business logic. These are then processed with external Risk Analytics vendor for capturing the market Risk factors. Once the results are received, the results are loaded into Investment Risk Final tables with additional logic to format the data as required which is then modelled and loaded into the Data Lake. This data is then made available to business users for analytical decisions.

Data Engineer and ETL Developer Employer -**Tata Consultancy Services** Pune, India & New

York, NY

Designation – IT Analyst

December 13, 2010 – June 19, 2016

I worked as a Data Engineer and ETL developer in Morgan Stanley Enterprise Data Warehouse (EDW) Wealth Management Project. Primary responsibility includes extracting data from various operational and transactional systems such as Flat Files, Oracle, Greenplum, Teradata, DB2, Sql Server etc., transforming them as per business requirement and loading them into the warehouse and data marts. This project uses business intelligence tools like Informatica for ETL suite, Unix as platform, Business Objects for reporting and Teradata for data warehouse – approximately 100 TB capacity.

EDUCATION**Degree - Bachelor of Engineering in Computer Science, 2010.**