

Data Architect/Engineer TOR

Innovation and partnership bond the five institutions of the World Bank Group (WBG): the International Bank for Reconstruction and Development (IBRD) and the International Development Association (IDA), which together form the World Bank; the International Finance Corporation (IFC); the Multilateral Investment Guarantee Agency (MIGA); and the International Centre for Settlement of Investment Disputes (ICSID). The World Bank Group is one of the world's largest sources of funding and knowledge for developing countries. It uses financial resources and extensive experience to help our client countries to reduce poverty, increase economic growth, and improve quality of life. To ensure that countries can access the best global expertise and help generate cutting-edge knowledge, the World Bank Group is constantly seeking to improve the way it works. Key priorities include delivering measurable results, promoting openness and transparency in development, and improving access to development information and data.

Information and Technology Solutions (ITS) enables the WBG to achieve its mission of ending extreme poverty by 2030 and boosting shared prosperity in a sustainable manner by delivering transformative information and technologies to its staff working in over 130 client countries.

ITS services range from: establishing the infrastructure to reach and connect staff and development stakeholders; providing the devices and agile technology and information applications to facilitate the science of delivery through decentralized services; creating and maintaining tools to integrate information across the World Bank Group, the clients we serve and the countries where we operate; and delivering the computing power staff need to analyze development challenges and identify solutions.

ITS is one of three VPUs that have been brought together as the World Bank Group Integrated Services (WBGIS), to provide enhanced corporate core services and enable the institution to operate as one strategic and coordinated entity.

As a unit within the WB Ops and Corporate (ITSOC), ITSDA provides state-of-art information and technology applications to support the operations of the World Bank Group. Functions provided ensure that the systems meet the business needs of users and external clients to manage business processes for the World Bank. The current application landscape encompasses the core technologies include Oracle, SQL Server, Business Objects, Tableau, Cisco Information Server (Composite), SAP BW/Hana, Informatica, .Net, HTML 5, CSS Frameworks, SharePoint and many others. Our plans are to migrate our on-prem data repositories and re-engineer based on new cloud architectures in the coming years.

ITSDA is seeking a Data engineer, with strong technical skills, to join a team of staff and consultants in the data management office. The data engineer will work closely with data architects (to determine what data management systems are appropriate) and data scientists (to determine which data are needed for analysis). The position requires a highly motivated and versatile candidate with strong technical knowledge, extensive data management skills on the latest tools and platforms to join the data team responsible for technical deliverables to provide business decision makers with information upon which they can make more accurate and effective decisions across multiple domains. The data engineer primary responsibility will be to use best practice data integration techniques to connect data architects/modelers and support the needs of data scientist and/or data intensive analytical applications.

Duties and Accountabilities:

The primary responsibilities of Data Engineer are the following:

- Design, construct, install, test and maintain highly scalable data management systems
- Employ a variety of languages and tools (e.g. scripting languages) to marry systems together
- Understand data scientist requirements and build scalable data pipeline to support those requirements
- Research opportunities for data acquisition and new uses for existing data
- Work together with data analyst to model data for staging into a data warehouse and architect data warehouse multidimensional models
- Develop data set processes for data modeling, mining and production
- Integrate new data management technologies and software engineering tools into existing structures
- Create custom software components (e.g. specialized UDFs) and analytics applications
- Recommend ways to improve data reliability, efficiency and quality
- Collaborate with data architects, modelers and IT team members on project goals
- Building data pipelines to collect data and move it into storage;
- Preparing the data as part of an ETL or ELT process;
- Work with complex technical data stores such as Hadoop, Amazon S3, relational databases, Azure BLOB, Azure Data Lake store, and MS Azure DW
- Stitch data together with scripting languages;
- Work with application DBA and modelers to construct data stores;
- Ensure data is ready for use by consuming application, analyst and scientist.
- Using frameworks and microservices to serve data.
- Design, develop and maintain services as part of enterprise service inventory, with focus on Service Normalization patterns. Ensure implementation of approaches to reduce redundancy in service logic and governance burden.
- Develop an in-depth understanding of the business processes and data and apply the knowledge towards the design and development of SOA applications.
- Perform analysis of system data to identify patterns, exceptions and erroneous data
- Ensure that data validation processes, data quality control standards, testing strategy and test plans for data services are applied to applications
- Work with the Project Lead to refine and tighten the security framework and access control for internal and external information/data sharing
- Work with the Project Lead to develop governance processes for managing services (version control), support services, service migration, monitoring and evaluation of services.
- Document definitions for all new data constructs and document the enhancements in a standard document.
- Contribute to the support and maintenance of an already existing body of data virtualization applications currently in production within the ITSDA application landscape
- Research new database technologies and recommend solutions in order to improve performance, scalability, & storage.

Selection Criteria

- Master's degree in Computer Science, Information Systems or a related technical field with a minimum of 5 years relevant experience or Bachelors Degree with a minimum of 7 years relevant experience.
- Minimum 2+ years' experience with schema design and dimensional data modeling
- Minimum 2+ years' experience in building enterprise data warehouses, data marts, operational data stores for large scale operations and at least 2 years of practical experience in the development of applications using Data Virtualization technologies.
- Experience with custom ETL design, implementation and maintenance, with expertise in error handling, performance tuning
- Experience in Cloud Technologies such as Microsoft Azure Data Lake Storage (ADLS) and Azure Datawarehouse (DW), working with an MPP system.
- Experience working with major database platforms such as Oracle and SQL Server
- Experience with building and maintaining physical and logical database models
- Experience in the integration of published views/services with tools like Power BI, Tableau, and Microsoft technologies such as Analysis Services
- Proficiency in UNIX environment and writing shell scripts.
- Experience in various development methodology like Agile/Scrum development methodology
- Excellent organizational and time management skills and a keen sense of priorities. Must be able to communicate effectively
- Excellent teamwork and interpersonal skills in operating collaboratively across organizational boundaries.

The selected candidate will be offered a 2 years term appointment.