|  |  |  |
| --- | --- | --- |
|  | Total 18+ Years of IT experience working as Data Architect from last 2 years.  Databricks Certified, Google certified, Oracle certified SQL, DATA, ETL Developer | **SUBRATA KUMAR PARIDA**  **7892633857**  **subratcall@yahoo.com** |

**PROFESSIONAL SUMMARY:**

Over **18+ years** of strong experience in the IT industry, from last 5 years mostly dealing with design and implementation of building Data Lake, Delta Lake, Lakehouse, Big data, Data warehouse and data analytics solutions.

Strong work experience working with various cloud providers like AWS, AZURE, GCS, OCI with various cloud data , analytics and AI technologies.

At present I am working on Azure data platform tools like Azure ADF, Azure synapse analytics, Azure data bricks, IICS (informatica) data integration. IICS MDM, Delta Lake tables.

.

* Over 8 **years** of experience in **Big Data platform** implementations, including **Architect, design, development**, and performance tuning using Spark, MongoDB, Kafka, S3, DMS, Glue implementations on AWS cloud platforms.
* Over 3 **years** of experience in **Big Data platform** implementations, including **Architect, design, development**, and performance tuning using Spark, Azure Storage, Azure SQL database , Azure data bricks , Azure ADF, Azure synapse analytics , Azure Datawarehouse
* Experience on understanding of **current state applications** and building **Future/To be state** architectures for data platforms including data ingestion, cloud (data lakes, delta Lakes, data warehousing), data transformation, and data governance.
* Strong experience in building a **data architect strategy** including data modeling, master data management, operational data management, data analytics and big data.
* **Design and develop data ingestion programs to process large data sets in Real Time, Near-Real Time and Batch data pipelines that bring together data from disparate sources** using **Apache Spark, Kafka, Spark Streaming, Scala & Databricks** and **managing resulting data stores in Data lake, Delta Lake and NoSQL Databases.**
* Strong understanding of underlying **AWS cloud Architectural concepts and distributed computing** paradigms for big data analytics.
* **6 years** of experience in **Data integration**, **Data modeling**, design, development and Support of Data Warehousing and business intelligence using **Erwin (Data modeling Tool), ETL, Oracle, Teradata** databases.
* **Extensive** experience of **Data Warehouse Architecture** and **Designing Star Schema, Snowflake Schema**, F**act and Dimensional Tables, Physical and Logical Data Modeling using Erwin.**
* **Experience working on Spark, Kafka & MongoDB integration with multiple Spark jobs to consume messages from multiple Kafka partitions and stored in MongoDB.**
* Hands on experience to build **data pipelines on real time** for data analytics using **Spark Streaming, Spark SQL, EventHub/Kafka** on AWS / Microsoft Azure.
* Strong experience in **Extraction, Transformation and Loading (ETL)** data from various data sources into Data Marts and Data Warehouse using Informatica components.
* **Good experience working in Agile development process & Software Development Life Cycle using Azure DevOps.**
* Database Modeling and designing OLTP, Analytic and Reporting Application.
* Writing efficient and scalable SQL queries for Application and reporting.
* Designing and writing PLSQL programs for various business and application needs.
* Writing UNIX shell scripts and Perl scripts for creating batches and backend jobs.

**Technical Skill:**

## *Big Data Ecosystems: HDFS, Kafka 3.0, Spark 3.0, Zookeeper 3.4, Databricks, Redshift.*

**NoSQL & RDBMS Databases**: Oracle, MySQL ,MongoDB ,PostgreSQL

## *Scripting Language: Python, Shell, Perl scripting.*

## *Operating System: Windows 10, Unix, Ubuntu 16.04 for Linux.*

**DevOps Experience:** Docker/ Kubernetes, Jenkins 2.0 & Git

## *Measure data Area: Data Lake, Delta Lake, Lakehouse, Datawarehouse and Data Analytics.*

## *Programming languages: Python, Pyspark, SQL & PL/SQL.*

## *Cloud Computing: AWS (EC2, EMR, S3, Redshift & Glue), Azure (ADF, Synapse Analytics*

## *DWH & BI AppTools: ETL Informatica , Power BI.*

## *Data Modelling Tools: Erwin 8x/ 2021 R1, TOAD Data Modeler, MS Visio*

**Business Domain Experience:**

Banking and insurance : 6.5 years

Retail and ecommerce : 8.5 years

Media and publication : 2.5 years

Travel and tourism : 1.5 year

**Career Profile:**

|  |  |  |  |
| --- | --- | --- | --- |
| Organization | **Dates** |  | **Designation** |
| Infocepts | Mar-2023 | Till date | Data architect |
| Epsilon India | 21-Nov-2016 | Mar-2023 | Lead Software Engineer /Associate Architect |
| MINDTREE Limited, Bangalore | Oct-2010 | Nov 2016 | Technical Lead |
| SOCIETE GENERALE Global Solution Centre, Bangalore | 23rd JUL 2007 | Oct-2010 | Senior software Engineer, Module Lead |
| EXILANT CONSULTING PVT. LTD Bangalore | 14th Mar 2006 | 10th July 2007 | System Analyst |
| SYSTEM LOGIC SOLUTION Ltd. Bangalore | 14th Nov 2005. | Mar 2006 | Software Engineer |

**Qualifications:**

|  |  |  |
| --- | --- | --- |
| Degree | Institute | Major and Specialization |
| BE in Electronics and Telecommunication Engineering. | IETE | Telecommunication |

**Certification**:

|  |  |
| --- | --- |
| Year and Month | Title |
| Jul 2015 | Oracle certified Expert SQL tuning 11.2 |
| Jul 2020 | Oracle cloud infrastructure foundation associate |
| Oct 2020 | Oracle cloud infrastructure associate architect |
| Oct 2023 | Databricks certified Data engineer, PySpark developer, Data analyst |
| Mar 2023 | Google certified associate engineer |
| May 2024 | Oracle cloud Generative AI Professional |

Current Assignments:

**Data Architect**, Sep 2023 to till date.

**Client: Luxury fashion brand USA and Europe**

**Organization:** Infocepts.

**Project Name:** Customer data platform Datahub.

**Technologies:** Data Lake, Delta Lake, Data Bricks, Spark, Azure ADLS, Azure ADF, IICS ,MDM

**Description:** Client is building an enterprise data platform for enterprise use and cater to its partners and customers as well. Data from nine different source system with different format and from different applications are the source for the CDP system. Different ETL logic applied to store the data in a common table. Many Masters data management logic applied to normalize the data for the downstream consumption.

**Responsibility:**

* Architectural design from current to future state which are highly available Data Fabric approach solutions for delivering scalable and with advanced analytics, data governance, data pipelines and data repositories.
* Work directly with cross functional teams including product managers and engineering to build and transform the customer system in a hybrid cloud environment.
* **Involved with understanding of current state applications and strangulate strategy implementations on future state.**
* **Complete analysis and design of Data Lake, NoSQL data modeling, performance tuning & Data Analytics.**
* **Build Data lakes and complete flow of how the data will be received, validated, transformed and then published.**
* **Architectural solution to suggest a strangulate pattern to move from monolithic applications to microservices solutions.**
* Develop and manage the data modelling best practices, patterns and standards. Support business and IT strategic planning (includes financial performance management).
* **Architect and designed the data flow for the collapse of 3 legacy data warehouses into an S3 Data / Delta Lake.**
* Generate and maintain scripts to create and update database schema s interactively over the course of the development phase of the project.
* **Comprehension of information management methodologies, advanced analytics, data integration, master data management and other data domains.**
* Responsible for defining the complete data architecture (i.e. Data modeling, ETL process, Data Marts, DW) and classifying the key business drivers for the data warehouse initiative.
* Collaborating with business users to understand GE capital & Corp. business requirements and translate them into process/technical solutions.
* Documented logical, physical, and dimensional data models (Data warehousing data modeling)
* Created entity relationship diagrams and multidimensional data models, reports and diagrams based on the requirements.

**Data Architect:** Mar 2023 to Aug 2023.

**Client:** Fashion Luxury brand

**Organization:** Infocepts.

**Project Name:** SAP data integration

**Technologies:**  AWS cloud, Amazon Glue, Amazon EMR Cluster , Redshift , Spark , Apache airflow

**Description:** Client uses SAP invoicing system, and those data needs to be integrated with EDW system for BI reporting purposes.

**Responsibility:**

* Responsible for understanding the business requirements and classifying the key business drivers for the data management.
* Collaborating with business users to understand business requirements and translate them into process/technical solutions.
* Documented logical, physical, and dimensional data models (data modeling)
* Created entity relationship diagrams and multidimensional data models, reports and diagrams based on the requirements.
* Develop the solution using Apache airflow and AWS.
* Developing Mappings, Mapplet, Sessions/Tasks, Worklets and Workflows using Informatica.
* Responsible for defining the testing procedures, test plans, error handling strategy and performance tuning for mappings, Jobs and interfaces.
* Performing code reviews for ETL mappings from a performance & Error handling perspective.

Older assignments:

|  |  |
| --- | --- |
| Project | **CRM & Loyalty** |
| Customer | Pepsi Co , Kellogg’s , SunTrust , Exxon Mobil |
| Period | Dec 2016 to Feb 2023 |
| Description | Worked for various client in CRM and Loyalty and marketing domain |
| Role | Oracle database in aws EDS, Redshift for warehouse |
| Solution Environment | Agile Process , DEVOPS method |
| Tools | AWS,Redshift,Oracle RDS , Custom Cloud platform |

Personal Details:

|  |  |
| --- | --- |
| Nationality | Indian |
| PAN | Yes |
| PASSPORT | Yes |