|  |  |
| --- | --- |
| **Asrith Sunkara** |  |
|  | ***Data Engineer*** |
| **Phone: +1(815)508-4217** |
| **E-mail: - Asrith.jobs@gmail.com** |

**Professional summary:**

* Highly motivated, Result - driven IT Professional with over **9+ years** of experience in Database Development, Design, System Analysis and Support, Business Intelligence applications, and Maintenance.
* Expert in all areas of SQL Server development including **T-SQL Queries**, Stored Procedures, User-defined Functions, Views, Triggers, Case expressions, Cursors, T-SQL relational database models and data integrity.
* Extensive working experience in **Normalization** and **De-Normalization** techniques for both **OLTP** and **OLAP** systems in creating Database Objects like tables, Constraints and Indexes using **DDL**, **DML**, **DCL** commands.
* Deep knowledge of different **DataWarehousing** methodologies and concepts, including **starschemas**, **snowflakes**, **ETL** processes, dimensional modeling and reporting tools.
* Experience in **Azure Data Factory**, **Azure Databricks**, **Qlik replication**, **Azure Data Lake gen2**, **Azure SQL Database**. **DeltaLake**, **AzureLogicApps**, SSIS, **Azure** Functions, Azure Synapse, **Pyspark**.
* Successfully Migrated On-premises Oracle databases to Azure Database using Azure Data Factory.
* Excellent knowledge of **ADF building blocks** – **Integration Runtime, Linked Services, Data Sets**, and **Pipelines Activities**.
* Hands on in error handling and debugging coding issues and troubleshooting of production problems.
* Highly qualified in **RDBMS** concepts such as Tables, User Defined Data Types, Indexes, Indexed Views, Functions, **CTE**, Table Variables and strong knowledge of SDLC principles.
* Skilled in performing data transfers between different **heterogeneous** systems (**Excel**, MS Access, Flat Files, XML file, Oracle, **JSON**, **Delta File**, **Parquet Files**) and SQL Server as a part of Data Migration, Data Maintenance and other ETL tasks.
* Professional in usage of **AzureControlFlow** items (for each Loop, execute package/SQL tasks, script task, send mail task) and **AzureDataFlow** items (Conditional Split, Data Conversion, Fuzzy lookup, Union all).
* Expertise in creating and managing Package Configurations, Logging system, Package deployment and User-defined Variables for **SSIS 2016/2012/2008 R2/2008**.
* High-skilled in writing Parameterized Queries for generating Tabular reports, Formatting report layout, Sub reports using Global Variables, Expressions, Functions, Sorting the data, Defining Data Source and subtotals for the reports using **SSRS 2016/2014/2012/2008 R2/2008.**
* Proficient in performing DBA tasks including installing and configuring SQL servers and backups/restore and recovery of databases.
* Worked in agile and waterfall environments, worked together with end users/customers to understand business requirements, recommended technical solutions, and documented functional requirements.
* Good team player, strong interpersonal and communication skills combined with self-motivation, and initiative and quick learner.
* Well versed with complete **SDLC** and **Agile** methodologies. Experience in every stage of software development life cycle (Analysis, Design, Development, Testing, deployment, and Support)
* Expertise in design and development of ETL solutions including Extraction, Transformation, and Loading data from heterogenous sources into Data Warehouses and Data Marts using **Informatica PowerCenter & SSIS**
* Expertise in creation of reports from Tableau and Power BI by connecting various enterprise Databases for data visualization.
* Experience in performance tuning in every stage of the ETL process and SQL Queries
* Strong knowledge on **UNIX** and shell scripting. Created Shell Scripts to automate various processes to save cost and time.
* Experience in Teradata, MS SQL, Oracle **19c**, **12c** &**PL/SQL** programming
* Hands on experiences on **Azure Data Factory, Azure Databricks, Azure SQL database, Azure SQL Datawarehouse (Synapse Analytics)**, **Scala**, **Python**, **Spark SQL**,**MSBI**
* Created customized reports in **Tableau, PowerBI**
* Worked on both WATERFALL model and AGILEmethodology.

## Expertise in release managementactivities.

* Working experience with front end and back-end products.
* Ability to quickly learn programming languages and otherapplications/software.
* Ability to organize, coordinate activities and motivate others through leadership and a positiveoutlook.
* Ability to work independently and in teamenvironments.

**Certifications & Achievements**

* **Certified in DP-203 (Azure Data Engineer)**

**Technical Summary:**

|  |  |
| --- | --- |
| **Skill** | **Details** |
| **DWH Technologies** | SSIS 2016, Informatica Power Center 10.X, Azure Data Factory 2023, Datastage 11.7, Informatica B2B and ab initio |
| **Reporting tools** | Tableau, Power BI, SSRS |
| **Hadoop** | Hadoop 3.X, Hive 3.13, PIG, HBase, Spark 3.X |
| **Languages** | MS-SQL 2019, PL/SQL, Python 3.X, Scala 3.X |
| **Database** | MS SQL 2018 & 2012, Teradata 12/13, ORACLE-19c/18c, DB2 13, Mongo DB 5.X, |
| **Database Tools** | TOAD, SQL\*PLUS, Teradata SQL Asst |
| **Cloud Technologies**  **(AZURE)** | MS Azure SQL, Cosmos DB, Azure Functions, Azure Stream Analytics, Azure Data Lake, Azure Databricks, Azure Kubernetes Services, Blob Storage, Storage Queues, Monitor, Service Bus, Active Directory, Resource Manager, HD-Insight, Virtual Network, Load Balancer |
| **Testing** | Functional, Regression, Performance, Adhoc, Integration, System, Compatibility, Automation, Data and Database Testing, Data Reconciliation Testing. |
| **Version Control** | SVN, GIT |
| **Others** | VSC, Jupyter Notebook, HUB, Data Bricks, Atlassian - Jira & Confluence, Typescript (basics), Apache Airflow, Apache Kafka |
| **Unix Scripting Tools** | SSH, Putty, Reflections, WinSCP, AWK |

**Professional Experience:**

|  |  |
| --- | --- |
| **DaVita | Denver, CO** | **2022 Nov – Present** |
| ***Senior Data Engineer*** |  |

The main goal of this data strategy initiatives is to understand the stakeholder’s data needs, collect data, transform data, and aggregate it into actionable data and deliver end to end solutions. Automated Data Quality checks and reconciliation data of email statistics for reporting and consumption of upstream team for their analysis.

* As a Data Analyst in the Platform Operations team, responsible for rolling out Corporate Medical Policies for 5 million customers across North America.
* Develop ETL infrastructure & automated data pipelines for writing scoring data files from **On-Prim to Azure Containers** using **ADF Pipelines**
* Perform Data Analysis on the analytical data present in **Azure storage Containers** using **SQL**, **Python**, **Spark**, and **Databricks**.
* Support the building the Data Marts on the Databricks platform using the analytical data received from Business.
* Measure the impact and performance of various customer engagement initiatives, including incremental lift, revenue, and business metrics across 3 communication channels Email, Direct mail, and Mobile.
* Handle a variety of file formats such as **Parquet**, **Avro**, and **JSON** for data analysis and debug using error logs fetched from **Azure storage Containers**.
* Help build internal data tools which can help the business analyst and broader team to detect anomalies and automate repetitive tasks.
* Manage the entirety of the campaign’s logic, including random segmentation of audience, exclusions, and assignment of offers and channels using **SQL Scripts**.
* Program corporate credit policy using scripting languages such as SQL, and Python and validate business intent for edge cases & corner cases analyzing output data of marketing campaigns. Create and manage versions of the scripts using **GitHub**.
* **Anaconda installations** on non - local (fiber-channel, iSCSI, NAS) devices in RHEL kickstart.
* Worked with **ETL operations** in **Databricks** by connecting to different relational databases using **Kafka** and used **AzureMonitor**&**Azure Functions** for creating, executing, and monitoring sessions and workflows.
* Ensured data quality and integrity of the data using **MSSQL server** and automated ETL deployment and operationalization.
* Developed spark programs using **Scala** APIs to compare the performance of spark with **HIVE**&**SQL**
* Performed Streaming of pipelines using **Spark Streaming** and **Azure Stream Analytics** to analyze the data from the data-driven workflows.
* Worked with consistent unification of Streaming, processed the data, and worked on ACID transactions using **Apache Spark**.
* Used Webpack to bundle **Typescript** and JavaScript modules.
* Used **Kendra's** advanced search capabilities, including natural language processing (**NLP**), relevance ranking, and support for multiple data sources.
* Designed and implemented **Resource Manager** to automate the provisioning of **Virtual Machine** instances and associated resources.
* Managed infrastructure as code for a multi-cloud environment, ensuring resource consistency and efficiency using **Resource Manager**.
* Worked with **Blob Storage** and developed the framework for the implementation of the huge volume of data and the system files.
* Access Controls/Permissions. Activity Dashboard. Audit Management. Audit Trail. Automated Scheduling. Collaboration Tools. Data Aggregation and Publishing using **MatillionETL.**
* Implemented of distributed stream processing platform with low latency and seamless integration, with data and analytics services inside and outside **Azure** to build your complete big data pipeline using **MatillionETL.**
* Worked on optimization of **Hive** queries using best practices and right parameters and using **Hadoop, YARN, Python,** and **PySpark**.
* Used **Sqoop** to extract the data from **Teradata** into **HDFS** and export the patterns analyzed back to Teradata.
* Worked on **Kafka** to bring the data from data sources and keep it in **HDFS** systems for filtering.
* Used Accumulators and Broadcast variables to tune the **Spark** applications and to monitor the created analytics and jobs.
* Involved in developing **Delta Lake** pipelines to manage data quality and governance.
* Tracked **Hadoop** cluster job performance and capacity planning and tuning Hadoop performance for high availability and Hadoop cluster recovery.
* Worked with **PowerBI** for generating reports and created Tableau dashboards, pie charts, and heat maps according to the business requirements.
* Worked with all phases of **Software Development Life Cycle** and used **Agile methodology** for development.
* Extensive knowledge using **Key-Vault, Active Directory**
* Automate the job schedule by calling **Azure Data Bricks** using **Azure Functions** written in **Python.**

**Environment:**Azure Data Factory, Azure Data Bricks (Python & Spark), Data Lake, Cosmos DB, Synapse, Scrum, Service Now, Git hub, Confluence, Azure message queues & event hubs, Power BI, Key vault, Active Directory, Hadoop, Hive, SQL, MS-SQL, Matillion ETL, Teradata, Resource Manager, Apache Kafka, Yarn, Azure Storage Containers, Kubernetes

|  |  |
| --- | --- |
| **Anthem|Norfolk, VA** | **Aug 2021 – Oct 2022** |
| ***Senior Data Engineer*** |  |

Data Hub is a strategic component which will advance our ability to integrate with current and future strategic systems and partners. Data Hub provides an elegant alternative to point-to-point interfaces which promotes data integration via a defined set of standard interfaces to move data between source and target systems. Reduced number of integration interfaces over time, as consumers start to subscribe to existing data standard publications (reuse). Defined architectural patterns and development standards on hub, promotes reusability in interfaces, transformations & routines, which brings simplicity and consistency in data integration. Support of various integration channels and technologies enables integration with any type of consumer easy.

* Requirement gathering from stakeholders and business analysts.
* Design solutions and provide a migration roadmap to business with the best **Azure tool set**.
* Developed and maintained end-to-end operations of ETL data and worked with large datasets and data migrations with **Azure Synapse Studio**.
* Developed Databricks notebooks using **Python** and**Pyspark** to load the raw files into **Cosmos DB**.
* Used event hubs to load policy data from the billing system to **Azure message queues** to publish files to the alfresco system.
* Configuring and loading **CDC** data into **Cosmos DB** tables.
* Working with the **Majesco** API team to resolve any data issues with the **Cosmos DB**.
* Peer reviews of **Azure** code developed by the team members.
* Working with the **DevOps** team to set up the code repositories for seamless code migration to higher environments.
* Setting up the Security and disaster recovery plans for the **Azure** environment.
* Loading data into **Cosmos DB** using **Databricks notebooks** with integration of **ADLS** data with message queues for downstream consumption.
* Setting up the best practices for **Azure** code development and implemented data masking for sensitive data in the **Synapse**.
* Developing Databricks notebooks **Python** (**Pyspark**) for complex data processing in **Azure** data integration.
* Performed the aggregation of log data from different servers and used them in downstream systems for analytics using **Apache Kafka**.
* Worked on designing and developing the **SSIS** Packages to import and export data from MS Excel, SQL Server, and Flat files.
* Collaborated with cross-functional teams to define **Resource Manager** best practices and establish coding standards for infrastructure deployments."
* Worked on **InformaticaData Integration** for extracting, transforming, and loading processes for the designed packages.
* Designed and deployed automated **ETL workflows** using **Azure Functions**, organized and cleansed the data in S3 buckets using **AzureDataBricks**, and processed the data using **Synapse Analytics**.
* Worked within the **ETL architecture** enhancements to increase the performance using query optimizer.
* Implemented the data that is extracted using **Spark**, **Hive**, and large data sets using **HDFS.**
* Worked on Streaming data transfer, data from different data sources into **HDFS**, **No SQL** databases.
* Created ETL Mapping with **Talend** Integration Suite to pull data from Source, apply transformations, and load data into the target database.
* Worked on scripting with **Python** in **Spark** for transforming the data from various files like Text files, CSV and JSON.
* Managed infrastructure as code for a multi-cloud environment, ensuring resource consistency and efficiency using **Resource Manager**.
* Loaded the data from different relational databases like **MySQL** and **Teradata** using **Sqoop** jobs.
* Worked on processing the data and testing using **Spark SQL** and on real-time processing by Spark Streaming and **Kafka** using **Python**.
* Scripted using **Python** and **PowerShell** for setting up baselines, branching, merging, and automation processes across the process using **Bit Bucket**.
* Worked with the implementation of the ETL architecture for enhancing the data and optimized workflows by building DAGs in **Apache Airflow** to schedule the ETL jobs and additional components in Apache Airflow like Pool, Executors, and multi-node functionality.
* Used various Transformations in **SSIS** Dataflow, Control Flow using for loop Containers and Fuzzy.
* Worked on creating **SSIS packages** for Data Conversion using data conversion transformation and producing advanced extensible reports using **SQL Server Reporting Services**.

**Environment:** Azure Data Factory, Azure Data Bricks, Airflow, Python, Pyspark, Synapse, PowerBI, Data Lake, SQL DB, Scrum, Service Now, Bit Bucket, Confluence, Apache Airflow, Apache Kafka, Python, PySpark, Teradata, Cosmos DB, Spark SQL, SSIS, Sqoop, Hadoop, Hive, MS-SQL, Informatica

|  |  |
| --- | --- |
| **AbbVie| Chicago, IL** | **2020 Jan – July 2021** |
| ***Data Engineer*** |  |

Strict regulatory data warehouse can help for AbbVie and storing data in a way that facilitates compliance with regulations and standards in the industry. Storing and analyzing data related to research and development activities, including drug discovery and clinical trials, can be crucial for making informed decisions and accelerating the development process. Managing the supply chain efficiently is critical in the pharmaceutical industry. A data warehouse can help in tracking and optimizing the supply chain by providing insights into inventory levels, demand forecasting, and distribution.

* Analyze, design, and build Modern data solutions using **Azure** PaaS service to support visualization of data. Understand current Production state of application and determine the impact of new implementation on existing business processes.
* Extract Transform and Load data from Sources Systems to **Azure Data Storage** services using a combination of **Azure Data Factory**, **T-SQL, Spark SQL**, and **Azure Data Lake Analytics**. Data Ingestion to one or more Azure Services - (**Azure Data Lake**, **Azure Storage**, **Azure SQL**, **Azure DW**) and processing the data in **Azure Databricks**.
* Created Pipelines in **ADF** using Linked **Services**/**Datasets**/**Pipeline**/ to Extract, Transform, and load data from different sources like **Azure SQL, Blob storage, Azure SQL Data warehouse**, write-back tool and backwards.
* Developed Spark applications using **Pyspark** and **Spark-SQL** for data extraction, transformation, and aggregation from multiple file formats for analyzing &amp; transforming the data to uncover insights into the customer usage patterns.
* Responsible for estimating the **cluster size**, **monitoring**, and **troubleshooting** of the **Spark DataBricks cluster**.
* Experienced in performance tuning of **Spark** Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
* To meet specific business requirements wrote **UDF’s** in **Scala** and **Pyspark**.
* Developed JSON Scripts for deploying the Pipeline in **Azure Data Factory (ADF)** that processes the data using the SQL Activity.
* Hands-on experience on developing SQL Scripts for automation purposes.
* Created Build and Release for multiple projects (modules) in production environment using **Visual Studio Team Services (VSTS).**
* **Analyzed** and **developedreports** using **customer transactional data** to create a multi-dimensional customer segmentation based on frequency of usage and product usage.
* Conducted analysis that linked customer sentiment with user behavior and product usage data.
* **Designed** and **automated** interactive **KPI dashboards** with sentiment scores in Pentaho + Python
* Analyzed survey data, marketing emails data from third party tools **Mail Chimp** using Excel and **PowerBI**.
* Prepared **KPI’s** and **Year Over Year comparison** reports as per the user requirement.
* Responsible for creating interactive dashboards to track metrics and scorecards.
* Created dashboard from scratch that included drill-through, **filters, Heat/ Tree Maps, Pie/Bar Charts, CircleViews, Line/Area Charts and Bullet Graphs.**
* Responsible for **PowerBI Server Administrative functions**, **upgrades**, **installations**, configurations, license management, backup of servers and load balancing techniques.
* Extensive **Power BI** Experience in Enterprise Environment and as Administrator including technical support, troubleshooting, report design and monitoring of system usage.
* Experience working with Extract & Live connections with high data volumes and connecting to multiple data sources like **Oracle**, **Excel**, and **My SQL** using **Data Blending**.

**Environment:** Azure Data Factory, Azure Data Bricks, Airflow, Python, Pyspark, Synapse, PowerBI, Data Lake, SQL DB, Scrum, Service Now, Bit Bucket, Confluence, MS-SQL, Data Governance, Data Validation, DAX Query, Power Query, Mail Chimp, Azure Data Factory, Azure Data Storage, Blob Storage, Visualizations

|  |  |
| --- | --- |
| **UHG | Minneapolis, MN** | **2018 Oct – 2019 Dec** |
| ***Data Engineer*** |  |

UHG ERP Integration is moving from current Oracle legacy systems to Azure Cloud. To support this milestone, several systems are currently interacting with their existing ERPs that need to be integrated. AN.100 defines the logic, rules, and key assumptions needed for the Export of Transactional Details from Oracle legacy systems to Azure Synapse.

* Responsible for analyzing legacy system data to store in **Azure data lake**.
* Requirement gathering from stakeholders and business analysts.
* Design solution and walkthrough with the business.
* Develop data migration **pipelines** from legacy to **Azure** using **Azure data factory** and data bricks.
* Developed Power BI data sets for the **Azure Data Lake** files.
* Working with the **DevOps** team to set up the code repositories for seamless code migration to higher environments.
* Configured the **Azure** failure alerts for the **Azure Data Factory**.
* Peer reviews of **Azure** code developed by the team members.
* Designed Hive repository with external tables, internal tables, buckets, partitions, **ACID property, UDF and ORC** compressions for incremental data load of parsed data for analytical & operational dashboards.
* Experienced in developing **Hive Queries** on different data formats like **Text file, CSV file, Log files** and leveraging time-based partitioning yields improvement in performance using HiveQL.
* Implemented functionality-based data modelling on Hive Tables and stored the resultants record sets into **HBASE via HBase storage Handler Column mapping.**
* For real time streaming and dynamic lookup of the data that is of **SCD1** performed windowing with domain specific grouping

**Environment:** Azure Data Factory, Azure Data Bricks, Airflow, Python, Pyspark, Synapse, Qlik replicate, Data Lake, SQL DB, Scrum, Service Now, Bit Bucket, Confluence, DevOps, SCD, HBase

|  |  |
| --- | --- |
| **PAREXEL | Durham, NC** | **2016 Jun – 2018 Sep** |
| ***Data Engineer*** |  |
|  |  |

PAREXEL is a CRO assisting leading pharmaceutical companies in conducting clinical trials by leveraging cutting edge technologies. Product focuses on building a data mart covering the necessary metrics to support the various client needs.

* Worked closely with the **Business Analysts/Engineers/Vendors** in gathering requirements and providing services as required.
* Responsible for analyzing Oracle source data and identifying potential issues while migrating the **Oracle** data into the **Azure Synapse database.**
* Configure Qlik replication tool to replicate on-premises Oracle transaction databases into Azure cloud.
* Worked on POC to explore Stream replication tool for real-time on-premises data sync up with Azure Cloud.
* Developed complex Azure Pipelines and stored procedures to automate the daily data sync process.
* Developed Azure Databricks notebooks using Pyspark to migrate the on-premises Oracle data into Azure SQL database.
* Developed spark jobs to load data from **ADLS gen2** to **Azure SQL server**.
* Responsible for analyzing and setting up the Azure database's disaster recovery process.
* Involved in critical production defect fixes.
* Developed **Azure failure alerts** for the data factory pipelines.
* Involved in Unit, Integration, System, and Performance testing.
* Responsible for enhancement and performance tuning of existing pipelines.

**Environment:***SQL Server 2016/2014, Visual Studio 2015/2013, SQL Server Reporting Services, SQL Server Integration Services, SQL Server Analysis services 2016/2014 T-SQL, Agile, TFS (Microsoft Team foundation Server), Tableau.*

|  |  |
| --- | --- |
| **InfoTech | Hyderabad, India** | **2014 June – 2016 May** |
| ***Big Data Engineer*** |  |
|  |  |

IData Express is an ELT tool which is used for loading of data after extracting from different data providers like Netezza, Flat file, PostgreSQL, Teradata which act as source databases and load the data into target data base Netezza only by using different transformations such as Single Joiner, Multi Joiner, Splitter, Filter, Roll up, Stage, Union, Minus and Expression.

* Responsible for design and execution of new or enhancements in existing code.
* Orchestrated robust data processing workflows utilizing Azure data bricks using Pyspark and spark frameworks.
* Developing data factory pipelines.
* Production defect fixes in **ADF pipelines**/**Databricks** notebooks.
* Responsible for scrum status calls daily basis as per schedule.
* Working as the technical lead for one of the teams.
* Working in Agile methodology - involving Epics, stories, tasks, and Demos at the end of each Iteration.
* Participate in Story grooming sessions.
* Involved in sprint and PI planning as per schedule.
* Daily updates about current sprint tasks in the rally.
* Discussions with **SME’s** daily to resolve most of the current sprint issues.

**Environment:***Azure data factory, Databricks, Airflow, Informatica, Oracle, Azure Cloud, Python, Pyspark & Spark, JAVA & Spring Boot.*

**Education Profile:**

Bachelor of Technology, India.