Prithu Kumar

Data Engineer

Professional Summary

Prithu has 7.5 years of **Data Engineering** experience implementing data and analytics solutions for cross sector clients with focus on problem solving using Data Warehousing technologies, ETL tools(**Informatica ,Talend**), big data technologies like **Spark , Hive ,** different databases like Oracle, Snowflake , BigQuery , MySql , Postgres . He has proved his mettle to create optimal compute and storage intensive Data Models to cater to the analytical needs.

- Hands-on experience in GCP components – **Dataproc , BigQuery, Storage Services , Cloud DataFlow**.
- Worked on Development , Modernization Initiatives , Data Lineage , Data Quality, Performance Tuning , Migration .
- Designed & Developed optimized pipelines in SPARK for ingesting large scale data from on-premise , transforming it and loading to Google Cloud Platform.
- Designed and implemented cloud solution on GCP to stitch various data sources to create an Integrated customer journey by tracking online activity

Work History Summary

Technology Consultant, Tredence Analytics Pvt Ltd
2021 May - Present , Tesco
Tech Stack - Spark, Python, Hive, Gcp(GCS, Dataproc, Big Query) , Unix
2020 Jan - 2021 May , Walmart
Tech Stack - Spark, Python, Hive, Gcp(GCS,

Contact

Address
Bengaluru , KA, 560067

Phone
919-901-305821

E-mail
prithu.kumar@tredence.com

LinkedIn
https://www.linkedin.com/in/prithu-kumar-62694a6b/

WWW
https://www.mindfarm-blogs.com/

Skills

- Spark
- Sql
- Hive
- Python
- DWH
- ETL
- Gcp
- Shell
- Retail
- Snowflake

Dataproc, Big Query) , Unix

2019 Sep - 2020 Jan , Tailored Brands

Tech Stack - Snowflake, Python, Power BI

Data Engineer, Tredence Analytics Pvt Ltd

2019 Jun - 2019 Oct , Kimberly Clark

Tech Stack - SAP Hana

2019 July - 2019 Aug , Ecolabs

Tech Stack - Snowflake, Power BI

2018 Oct - 2019 Jun , Barracuda

Tech Stack - Spark, Gcp, MySql, Salesforce , Talend , Python , Java

Data Engineer, Infosys

2016 Sep - 2018 Oct , Wells Fargo

Tech Stack - Unix, Python, Informatica, DWH , Oracle

2014 Dec - 2016 Aug , RBS

Tech Stack - Unix, Python, Informatica, DWH , Oracle

Work History

2021-05 - Current

Senior Data Engineer, Tredence

Client - Tesco

On-Shelf Availability

Tech Stack - Spark, Python, Hive, Gcp(GCS, Dataproc, Big Query) , Unix

The project aimed at predicting the onshelf availability of products at store level using Kalmann Filter Algorithm for the purpose of increased sales and reduced wastage.

- Implemented 3 modules out of 12 using PySpark and GCP
- Led the discussion to create the flow and data model supporting this
- Designed a parallel pipeline framework to execute two versions of the same algorithm in parallel

2020-01 - 2021-05

Technology Consultant, Tredence

Client - Walmart

Customer Journey

This project aims at creation of multiple analytical

Power BI

Certifications

Got the certificate for CCDSAP Foundation Exam for Codechef Certified Data Structures And Algorithms Programme on 18th March 2018.

Education

2010-08 - 2014-08

B. Tech: Computer Science Engineering

Dr. B. C. Roy Engineering College - Durgapur

GPA: 7.8/10.0

Accomplishments

- Got PAT ON THE BACK award two times within a year for performance in Walmart.
- Got PAT ON THE BACK award for performance for client Barracuda.
- Got INSTA award for performance in Data Hub.
- Secured 87% in Foundation training (Java, Database, and

datasets for many workstreams for GM(General Merchandise) and OG(Other Groceries) and W+ in GCP. Achieving the goal involves fetching data from multiple sources and create optimal data pipelines. Primary challenges include handling huge amount of data, treating different types of data, creating a technical solution consuming minimum storage, regular incremental and others.
· Responsible for design and development of various data pipelines in Spark to migrate data from various source systems
· Led the discussion with Onsite/Client to precisely understand the set of requirements
· Design and Develop Robust Data Models
· Guide the sub-ordinates in developing the scripts, pipelines
· Validate the pipelines and scripts created
· Framed the technology stack & architecture
· Carve out a project plan for the whole team to adhere to
· Perform RCA on failed/time consuming pipelines
· Design /Finalize the dashboard structure

Software Engineering) internally in Infosys.
- Got certifications on Informatica and SSIS internally in Infosys.

2019-09 - 2020-01

Lead Data Engineer

Client - Tailored Brands

Creating Analytical Datasets
This project aims at creation of multiple analytical datasets for marketing – facebook , cheetahmail ,SEM, SEO, NPS (and others) . Achieving the goal involves fetching data from multiple sources ,create optimal data pipelines, robust data models ,corresponding dashboards and provide valuable insights to the client. Primary challenges included connecting the dots between different datasets coming from different sources, identify the major KPIs, regular incremental and others. Another component of it included a Gap analysis of the current architecture and present alternative solutions.
• Responsible for design and development of various

data pipelines to integrate data from various source systems
• Led the discussion with the Client to precisely understand the set of requirements
• Design and develop the Robust Data Models
• Guiding the sub-ordinates in developing the scripts, pipelines
• Validate the pipelines and scripts created
• Frame the technology stack
• Carve out a project plan for whole team to adhere to
• Perform RCA on failed/time consuming pipelines
• Managing a team of 5 members
• Creating dashboards
• Did a Gap Analysis of the present DE architecture of the marketing division
• Presented alternative technical architecture to cover the flaws found through Gap Analysis

2019-06 - 2019-10

Lead Data Engineer , Tredence

Client - Kimberly Clark

Data Quality Framework
Tech Stack - SAP Hana, Python
Designed and implemented a Data Quality Framework, complemented by an alert system. Designed the dashboard for the presentation of the metrics.

2019-07 - 2019-08

Senior Data Engineer, Tredence

Client - Ecolabs, Bangalore, India

Snowflake Platform Evaluation POC
Evaluate Snowflake database as a potential option to replace Azure analysis services in the longer term for Institutional Sales Dashboard. Consider usability, maintainability, security, cost of ownership and future scalability while evaluating the Snowflake database.
· Evaluated Snowflake database platform to deliver the Performance baseline and comparative results (Live Connection and Import), Data Integration

approach (One Time and Incremental) and Platform Cost Estimates (extrapolated based on single report, query volume based on current usage and live / import in Power BI).
· Reverse Engineer ETLs and Semantic Layer.
· Data Integration for required objects.
· Design the data load strategy for the ETL pipelines.
· Integration testing of the data in the underlying tables.
· Design Semantic Layer in power BI
· Design and develop the report.
· Functional, Integration, concurrency testing of the report.
· Snowflake cost evaluation by running the report in different cluster sizes(S,M,L,XL)

2018-10 - 2019-12

Senior Data Engineer , Tredence

Client - Barracuda

Data Preparation
Tech Stack - PySpark, Talend, Python, Java , Azure Blob Storage, Postgresql , Salesforce , MySql
This project aimed at creation of a single source by consolidation of data from different sources, based on business rules, for many workstreams and upload on a newly created platform. A lot of sources like salesforce, pardot crm, Postgresql acted as input sources. Based on different algorithms, it was needed to tag the same element across different datasets/objects. Primary challenges included identification of same data, downloading huge amount of data, regular incrementals and others.
• Responsible for design and development of various data pipelines to migrate data from various source systems
• Worked across multiple modules of the project – contact, account, lead
• Developed ~40 ETL pipelines using Talend/PySpark
• Leveraged Azure Blob storage in attachments module

• Leveraged azure components in Talend
• Was responsible for designing and implementing the flow in contact and attachment. Had contributions in lead as well
• Designed and optimized sql scripts and stored procs
• Optimized flows related to sources like salesforce, pardot crm
• Automated the flows in different modules
• Wrote sql queries for validating logic across different workstreams

2016-09 - 2018-10

Senior Data Engineer, Infosys

Client - Wells Fargo, Bangalore, India

Data Hub

Tech Stack - PySpark , Python, Informatica, Unix , Oracle

This project aims at development of code for Operational Data Sources in the Oracle Exadata environment. New UNIX scripts, mappings, workflows, stored procedures are created according to the design documents. Sed, awk and pmcmd along with other commands have been used extensively in these scripts. SQL loader utility has been used to load data into many tables.

• Worked as a developer in team of 12 members at client location.
• Worked on performance tuning in Oracle
• Worked on performance tuning of mappings/workflows
• Developed mappings, sessions and workflows in Informatica PowerCenter for ETL process
• Created Pyspark pipelines,Unix scripts, Oracle queries
• Created Stored procedures and functions according to business requirements
• Worked on utility, SQL*Loader
• Worked on concepts like Exchange Partition

2014-12 - 2016-08

Data Engineer , Infosys

Client - RBS

Cards Bis

Tech Stack - Informatica, Oracle, Unix

It is an application, which deals with the development of data warehouse for the debit card and credit card information of customers of the newly created bank. Unix scripts, stored procs, mappings were created to meet the requirements.

- Worked on performance tuning in Oracle
- Developed mappings, sessions and workflows in Informatica PowerCenter for ETL process
- Created Unix scripts, Oracle queries
- Created Stored procedures and functions according to business requirements
- Developed stubs so that if inbound files are not available, they would do the purpose
- Analyzed CRs