Abhik Chakraborty

Senior Engineer– Data & BI

abhik.chakraborty92@gmail.com | +918587920073 | Gurugram, Haryana, India

linkedin.com/in/abhik-chakraborty-77bb1275/

https://abhik-chakraborty92.medium.com/

https://github.com/abhikchakraborty92

Senior Data and BI Engineer with more than 6.5 years of experience in Healthcare, ITSM, and Product analytics designing terabyte scale Data solutions in DSS, SSIS, SQL Server, Python, AWS & more, visualization reports in Tableau, PowerBI & Excel, RPA using UIPath & serverless analytics solutions in AWS Lambda

Work Experience

Data Engineer III

Expedia Group | (5-2021 – Present)

- Leading implementation as **a Tech lead for a team of 2 people**
- Constructed **data processing pipeline** for usage/quality logs of audio/video calls with over **billion** events monthly using **AWS S3, EMR, PySpark and Redshift**
- Successfully **migrated 10+ On-prem** team business databases with over 300 tables to **cloud AWS SQL Server** RDS using Python Boto3
- Designed **data warehouse** and its ingestion pipelines for change management data stream with AWS **DynamoDB, SQS and Lambda**
- **Enhanced application metadata of 60000+ applications** by designing a metadata correction service using **Github API, AWS Lambda and Redshift**

Senior Reporting Analyst II

Expedia Group | (3-2020 – 5-2021)

- Engineered domain performance evaluation of Expedia DPI, **improving average health by 15%** by developing **SQL Server data warehouse for 50+ KPIs**
- Designed and replaced a legacy revenue loss estimation system, **reducing estimation errors by 15%,** with fault tolerant **PySpark ETL pipelines** for terabyte scale bookings data
- Improved post COVID employee retention by **20%** by capturing employee stress data using **AWS Lambda** and **DynamoDB**
- Aided in license **cost reduction over 25%**, by implementing **SQL Server Data Warehouse** for product license & usage logs and analytics API endpoints with **AWS Lambda**

Senior Reporting Analyst I

Expedia Group | (5-2018 – 2-2020)

- **Architected entire performance reporting eco-system** for Expedia IT replacing ad-hoc data analysis with fully functional **data warehouse** on **SQL Server** database and **Tableau dashboards** with **40+ KPIs**
- Improved **Customer Satisfaction by 5%** by standardizing customer feedback data analysis, designing metrics data store using **Redshift and PySpark**
- Improved **SLA compliance by 10%** with live SLA monitoring **developing NRT database** from JIRA APIs to Redshift using **PySpark and EMR**
- Automated manual data reviews and data-based notifications by introducing **RPA ecosystem with UiPath**

Senior Business Technology Analyst

ZS Associates | (1-2018 – 5-2018)

- **Managed a team of 3 people** and drove pure data engineering projects for world's second biggest pharmaceutical company
- Standardized physician & patient profile information in Europe by developing **PySpark ETL pipelines** for **terabyte scale** data from client data sources in AWS **Redshift**, **S3, delimited files** to AWS **RDS SQL Server**

Business Technology Analyst

ZS Associates | (6-2015 – 12-2017)

- **Transformed** 10+ years historical **sales transactions for client vs their generic drugs** to generate sales insights after a few drugs' loss of exclusivity
- **Developed terabyte scale data warehouse** and related **ETL** pipelines using **Dataiku DSS** to capture sales data for client Rare Disease drugs in Europe region
- **Developed Tableau dashboards** to support client drug expansion in the Asia Pacific Region by tracking sales, distribution, and budgeting
- **Created fully automated Excel VBA** based summary reports to support client-side inventory logging of drugs in specific locations
- **Generated Tableau and Excel Dashboards** to gather market insights before the launch of **Onivyde** drug in US region
- **Maintained 3+ Tableau dashboards** and their **related data pipelines** to generate monthly insights for 6 drugs in Europe and USA

Education

Bachelor of Technology

Guru Gobind Singh Indraprastha University, Delhi
08/2011 – 07/2015   **80.4%**

Senior Secondary Examinations (CBSE)

Greenway Modern School, Dilshad Garden, Delhi
03/2009 – 03/2010   **82.8%**

Indian School Certificate Exams (ICSE)

Holy Angels' School, Rajendra Nagar, Ghaziabad
03/2007 – 03/2008   **86.8%**

Skills/Toolstack

Skillset

ETL | Data Modelling | Data Warehousing | Data Visualization | Agile | Serverless Design | Robotic Process Automation | Cloud Big Data Processing

Toolstack

Scripting

SQL | HiveQL | Python | Excel VBA | PySpark SQL

Databases

MS SQL Server | MySQL | AWS Redshift | PostgreSQL | DynamoDB | RDS

ETL Tools

Dataiku DSS | SSIS | EasyMorph | PowerQuery | Tableau Prep

Data Visualization Tools

Tableau | Power BI | MS Excel | Streamlit

Cloudstack

AWS

S3 | Lambda | EMR | API Gateway | SQS | SNS | Redshift | DynamoDB | RDS | Boto3 Python SDK

GCP

GCS | BigQuery | Cloud Data Fusion | Dataproc

Personal Projects

TXP: Fully mechanized **Tableau** controls with single line **python** commands for:

- Downloading data, images, or PDF of the dashboards
- Adding, refreshing, or deleting data sources
- Viewing/downloading personal usage data

Slack Analytics app with AWS Lambda: This implementation is:

- Single file python-based template to create Slack apps
- Sends **customized insights** to users in response to specific slack texts
- Uses **AWS Lambda**, **AWS HTTP Gateway** & and **AWS S3**
- Can handle **multiple requests** due to its serverless design

EMRAuto: Automated EMR Pipeline creator. It has following features:

- Automatic creation of **EMR Clusters** to set up analogous cloud Hadoop environment
- Schedules Python scripts as pipeline steps on EMR from local/On-prem
- Maintains **logs** of triggers
- **Automatic destruction** of clusters to save costs

Achievements

ECommerce Platform Champion Award: Designing entire performance management ecosystem for Expedia IT with 40+ KPIs, related data pipelines and performance notification bots

Way to Go! Award: Awarded for record time implementation of fully automated live Problem Management & SLA Monitoring Dashboard at Expedia

Published Author: Author of acclaimed mystery novel "Shadows of the Departed", exhibited in Cairo and Abu Dhabi International Book Fairs

Certifications

Certified Data Scientist in Tableau

Robotic Process Automation Developer Certification in UIPath

DSS Core Designer of Dataiku Data Science Studio ETL Development