

Khyati Bareja

DATA ENGINEER

✉ khyati.bareja@gmail.com 📞 +1(514)5166403 📍 Montreal, Canada

in <https://www.linkedin.com/in/khyatibareja/>

🌐 <https://sites.google.com/view/khyati-bareja-portfolio/home>

SUMMARY

I am a Data Engineer having 7+ years of working Experience in IT industry, have experience managing team members and end to end delivery by following Project lifecycle with a focus on optimizing data ingestion pipelines and conducting analyses to support business decisions.

Have experience in managing containerized workflows with Kubernetes, building data science and machine learning data pipelines, and creating visualizations. I have recently graduated with Masters in Software Engineering from Concordia University, Montreal, Quebec, Canada.

SKILLS

Google Services
(BigQuery, Composer,
pub/sub, cloud function)



AWS Cloud



Azure Cloud



Talend



Terraform



Docker/Kubernetes



NoSQL- MongoDB



Programming
(Python/Java)



SQL



Others (Jira/Gitlab)



Tableau



EXPERIENCE

Graduate Teaching Assistant

Concordia University (CSSE) Sep 2023 - Apr 2024

- Tutored 40+ students in **Programming and Problem-Solving** course, using slides and conducted weekly Lab Exams.
- Assessed and graded 100+ students on lab assignments, via means of online demo sessions, graded projects and lab exams for **Software Requirements Specifications** course based on clear criteria to ensure fairness in evaluations.

Data Engineer

Dunnhumby Sep 2019 - May 2022

Role: Data Engineer

- Managed workflows by writing DAGs with Airflow and Google Composer.
- Maintained a Data Lake that supported over 100+ data sources and delivered data to more than 80+ ponds.
- Involved in Requirement Gathering for the development activities – such as understanding and onboarding new sources.
- Performed analytics on sales arrangement and sales data to identify trends, offering visualizations predictions and inventory management improvements and shelf arrangements.
- Utilized Google services, including Dataproc, AutoML, and TensorFlow, to construct machine learning data pipelines, preparing data for ML engineers to integrate predictive models.

Sr Software Engineer

Tech Mahindra Aug 2018 - Sep 2019

Role: Big Data Engineer

- Designed and built data pipelines utilizing SQL, Cloud Storage, Cloud Functions to integrate diverse data sources for analytical models and dashboards.
- Developed and productionized multiple Tableau dashboards focusing on key metrics in Yield Rates, Equipment Efficiency, Production Cycle Times, Defect Rates, and Supply Chain Performance, delivering actionable insights.
- Automated essential processes, including GCP and Tableau data table refreshes, as well as push-pull pipelines using Cloud Functions and Cloud Scheduler to improve data update efficiency.

Associate Consultant

Capgemini *Jun 2015 - May 2018*

Role: Data Analyst

- Extracted, transformed, and analyzed data from multiple POS systems and APIs using Hadoop and Hive, to generate business insights through visualizations.
- Designed ETL pipelines with Sqoop to load cleansed and transformed data in Hive for reporting and building visualizations in Talend.
- The source of the data was structured database and PAR Data Central.
- The code development of extraction was done using MSSQL and various components of Talend.
- Generated reports to present sales historic trends, predictions and customer insights.
- Managed a key module from requirements to deployment, integrating sales data to generate trends and forecasts which involved creating a reusable ETL jobs with Hive.

RECENT ACADEMIC PROJECTS

Diabetes Prediction Analysis (Machine Learning Project)

Model designed in phases which involved Data Preprocessing, Exploratory Data Analysis (EDA), Feature Insights (via Correlation Matrix, Model training, Model Evaluation (using confusion matrix, RMSE, F1-score, precision, Feature Importance).

Github: https://github.com/shashank6341/Big-Data-Team-12/blob/master/Team_12.ipynb

Vision Walk- Human Computer Interaction

Vision Walk uses image captioning (Vision Language Pre-training) and real-time translation (Google ML Kit) to provide audio descriptions of the environment by using a combination of auditory and haptic feedback to improve interaction with the environment.

Github: <https://github.com/Khyati-Bareja/Vision-Walk-Human-Computer-Interaction>

OSMDashboard for OpenTracks (OpenSource Project)

OSM Dashboard is a companion application for OpenTracks. It adds the functionality to show tracks on a map using data from OpenStreetMap in real time.

Github: <https://github.com/OpenTracksApp/OSMDashboard>

EDUCATION

Master Of Engineering, Software Engineering

Concordia University *Sep 2022 - Jun 2024*

Montreal, Quebec

Bachelor of Technology, Computer Engineering

Kurukshetra University *Sep 2009 - Aug 2013*

Haryana, India

CERTIFICATIONS & TECHNICAL BLOGS

Hortonworks HDP Certified

Administrator

Credential Url: <https://bcert.me/bc/html/show-badge.html?b=wtmazhf>

• Engineer Data In Google Cloud - Intermediate

GCP Certified

Url: https://www.cloudskillsboost.google/public_profiles/ca1cb18f-10a6-44f5-a804-f2fb31513a49/badges/5998781

Microsoft Certified Azure

Fundamentals

Credential Url: https://www.youracclaim.com/badges/38a9d847-21dd-4761-bf7d-7f43eb436b10/linked_in_profile

Machine Learning – Feature Engineering

Medium Blog:

Url: <https://medium.com/@khyati.bareja/machine-learning-feature-engineering-d7699ee3bdcf>

Deep Learning Using TensorFlow

Medium Blog

Url: <https://medium.com/@khyati.bareja/deep-learning-using-tensorflow-1fe71e6810d0>