Akshay Hazari

akshayhazari@gmail.com   Mob. No: +91-7045272148

Linkedin : https://www.linkedin.com/in/akshay-hazari-04014873/

SO : https://stackoverflow.com/users/2251058/akshay-hazari?tab=profile

Skills	Python, Node.js ,Django, flask , Hive, Hadoop, Big Data, Data mining , Postgresql, Spark, Presto, Redshift, Airflow
LetsTransport (SDE 2)	Created Back End for Client Bookings, Client Dashboard and Client App, For creation and retrieval of vehicle requirements, quotations and handling statuses of bookings. Showing the list of partners available and quotation statuses on client dashboard.**(nodejs,expresso,python,mongoose,ndb)** .
	Created pipeline for handling Client Bookings data (Client Listings) from MongoDb , Google DataStore to Cloud (GCloud) creating custom operators in airflow **(airflow, python, mongodb)**
	Payment automation and validation system for scheduled payments and on demand payments. Modeled and Created Apis for handling Validation handling city budget and api's for finance interface for viewing and correcting validation failures and connected Banks Api's for transactions **(nodejs, mongodb) (Sept 2019 - Present)**
Expedia (Software Engineer)	Part of expedia data platform team (DP2) . (Jan 2018 - Aug 2019)

- **Campaign and Customer Uber ETL** - Modified and Managed the workflow for Response, Visitors , Transactions scripts for multiple sources.(Travelocity, Ebookers, Omniture Response, BEX Brand-Expedia) **(Hive, Presto, Redshift, python)**
- **Integrated Uber ETL** - Built the ETL pipeline for Customer and Campaign merged together. Having the workflow optimized with common processing and staging tables.

  - Worked on new added features in the new pipeline - Travelocity scripts at customer grain , new dimensions and metrics, change of source tables picked from streaming jobs , added refined regexs , notifications for QV dashboard, Removed unseen bugs , added timeouts to all scripts to avoid long running jobs, added many metrics to Daily DQ check script mailed as a Variance Report, updated many Hive settings to make it efficient.

  - Added Channel Dim Refresh query at Customer grain assumed to be costly and achievable reducing skewed joins and other expensive operations and converting joins to map joins by filtering and adding settings and introducing Random distribute by to reduce skewness etc.

  - Analysed the query plans to figure the costly joins and operations making optimizations in joins and settings.

  - Worked on reducing merge joins to Map joins , adding initial filtering of joined tables earlier in workflow to reduce the data sets and the size of the joins. Reduced skewed Reducers and EC issues with distribute by and randomized distribute by, and dropping partitions earlier in the scripts.

  - Optimised the new workflow making it more efficient and reducing it by 40% in execution time and cost **(Hive, python, Presto, Redshift, Kafka)**
- **Loyalty ETL** - Working on Jira's assigned and Managing Loyalty ETL which gives benefits (points) and privileges to customers **(Teradata, Informatica , Java)**

- **Validation Utilitiy for Loyalty 1LP** - Created a utility to check the latest processed data between source and final tables **(python, Postgresql , MSSQL)**
- **Data Quality Framework DQF Backend,Frontend** - Rule based data quality backend framework with apis to check on quality of data processed between source and target showing comparisons on UI Dashboard (Hive , Teradata, PostgreSQL, MSSQL) **(python, Django ,javascript, html)**
- **Spark POC** - Created codes for alternate scripts for Response in Integrated Uber and response merge script with deduplication logic for Customer grain in Uber Work Flow. **(pyspark, scala spark)**
- **Slack Notifications for SLA** - Worked on modifying the notifications to slack on ETL - SLA completion, failure, warning and other instances. **(python)**
- **Hackathon 1** - Created a recommendation engine to schedule a vacation and decide a place based on history of traffic patterns , weather and user behaviour , built backend , setup , algorithm for recommendation **(nmslib , python)**
- **Hackathon 2** - Created backend for hotel waitlist **(python)**

Entropik Technologies (Back End Developer, Learning Front End)	Fast paced environment, prioritizing ,multitasking. (Oct 2016 - Jan 2018)

- **Affectlab (affectlab.io)** - App is used to monitor the Emotion Quotient of a Person as he views a Video , activity on the Web(Instructions for Browsing), Gaming , Natural Activity.

  Stream lined the process of Media Test Creation , Creation of User Testers and allocating tests, Capturing webcam for facial landmarks and detections of facial expressions , storing them on s3, assigning emotions to the tests from EEG and facial expressions.

  Built a dashboard where the App is used along with an eeg headset to provide demographics and user emotions for age groups , gender and other diminsions(Emotions) to be compared with other tests. Solely created the backend .Some parts of the frontend (Dashboard, Test generation and assigning Testers).

  Created a Multi tenant version of the app with middlewares to handle multiple tenants, and modeled and designed a dynamic survey/feedback system for tests Where Admin can add questionnaire while response is recorded. Questionnaire can be Text, options, Numeric ratings, feedback while response would be captured appropriately for different tests.

  Server set ups using fabric to set up nginx and gunicorn on multiple servers, managing staging and production environments and rest of aws setups for rds (databases) etc. on aws **(python, django ,postgresql ,mongodb ,pandas, Java ,javascript)**

- **Facial Coding (facial.affectlab.io)** - Analyzing facial expressions from facial landmarks to identify human emotions as he watches a video, capturing his facial landmarks then deriving emotion through facial expressions.

  Also worked on cross compilation to translate c++ code to javascript (DLIB to Javascript to be used in frontend). Built backend,frontend pipeline, devops. **(python,webrtc,postgresql,javascript,emscripten)**

- **Vision, Semantic, Brand Safety Pipelines** - Feature Extraction from Videos, Image and Video Classification. Used tensorflow to build a simple deep learning

application classifying Images into Nudity, Violence, Drugs, Alcohol.

Created a text and image crawler for websites, determining the safety for a brand to place their advertisement, then classifying brand safe pages. Scaled up image and website text Crawler, classifying brand safe pages. **(RabbitMQ,Scrapy)**,**(Scikit-image,Scikit-learn,dlib,Imageio,Tensorflow)**

- **Video analytics** - Feature Extraction from Videos (Scene breaks , scene fading, face detections, HSV rate of change, number of objects, beats in audio, voice and music detection). Analysis of emotion to feature correlation , built the entire pipeline for data extraction, video and audio features and then prediction based on these features and other inputs like gender, age etc.

  Features - Scene breaks, number of objects, hsv, optical flow, motion, color to emotion relations, beats, audio segmentation by speech and music etc. **(Scikit-image,Scikit-learn,dlib,Imageio,many audio processing libraries)**

- **Chromo (chromo.io)** - App which deals with mobile scrolls ,swipes, motion, posture and analyses user emotions . Created api endpoints and restructured the backend with a mongodb generic filter , created a dashboard with backend apis, scaled the app using celery. **(python , django , mongodb)**

- **Fictionite (fictionite.io)** - App which allows users to read unpublished books as they are created and provide feedback and suggestions from users to its authors. Fictionite had its own Monetary system (Gold). There would be seasons and episodes as the author adds new episodes he can monetize them.

  Created Backend **(python , django , postgresql)**
- **goaktive (http://54.169.5.247/)** -Goaktive provides tracking a count of repeated maneuvers performed by individuals , counting physical exercises.

  Created backend to capture webcam stream store it on s3 and use it on algorithms to give a count of exercise while processing in the background. Created Backend, FrontEnd to add, delete, show the counts, Server setup etc. **(python,django,javascript)**

Rediff.com (Software Engineer)	Part of Engineering, Search Team. ML, BigData - Hadoop, Backend (June 2015 - Oct 2016) - **Active Learning Spam Filter** - Semi supervised learning based continuously learning spam filter for Rediff Message Boards. Notification system to analyse and block spammers and a Gui (cherrypy) to test multiple classifiers . **(python , scikit-learn)** - **Queue System for cache warmup** - Continuously running queue to warmup and repopulate cache for urls having dynamically generated delays. Every town in the queue has dynamic delays to refresh news articles and once a certain delay is reached rabbitmq consumer nodes call to repopulate the cache. System with a realtime Healthwatch to monitor. **(python , RabbitMQ)** - **Newsletters Analysis Rediff Edge** - Entropy weighted clustering,apriori, click stream analysis. Power Iteration Clustering - for graph based clustering on distances, clustering users by node to node similarity,GradientBoostRegressor to Predict Click Rate **(python, R , Spark mllib , SolR)** - **Handling Hypertable ETL for Rediff Shopping Mails, Rediff News Emails (views,visits,clicks, read unread, subs, unsubs, referrer etc)**

Akshay Hazari   akshayhazari@gmail.com   Mob. No: +91-7045272148

1015, Maruthi Apartments, 1st Main Rd, Avalappa Layout, K R Garden,

Murgesh Pallya , Bengaluru, Karnataka 560017

- Scripts to parse source data, Loading data to Hyper Table (open source Big Query columnar database) , Redshift , Oracle , from source after ingestion**(python, awk, Hyper Table, Resdhift, Hadoop)**
- **News Guru Graphs, Queries, Analysis - Rediff Edge** - Queries in a realtime News Guru app to find trending , popular and new News Gurus. Initial data preparation,analysis by writing code scripts in python, R and queries in Redshift. **(aws-redshift(postgresql), Oracle ,R , python,SolR)**
- **Other work** : Hadoop Scripts (EMR) , Spark - Java (EMR) scripts for data mining and analysis, Analysis of mail users,rediff news , uncommon users ,sequential page views, frequent mailers ,shopping trends, data preparation , pre-processing etc. using AWS Redshift, Java, Python (Scikit learn, pandas), R .

Few Lectures at (VJTI, PUCSD) from Rediff.com : Took lectures on Hadoop Mapreduce. Taught building mappers and reducers in Python,R,,Java,Shell.

Apigee Technologies (Internship)	Big Data , Data Analysis and POC on various Technologies (June 2014 - November 2014 , 6 Months) • **Custom Report Generation & Custom Rules and Alerts (Big Data Analysis)** - Scalable report generation mechanism built on API Analytics data, using Hadoop map reduce. The system took report generation requests , report definitions from customers on api data to build the report and subsequently store it for customers to view. Multi-tenanted system , scaled for multiple report generation requests for large datasets with live cluster on cloud handling multiple job requests. Scheduler scheduling jobs , mail notification system . – Custom Report Generation generated reports on fixed rules specifying multiple dimensions in records , functions to perform analysis and other constraints and then sending notifications and alerts. – Custom Rules and Alerts involved generation of comparison reports based on custom rules added to perform comparison analysis on records and accordingly creating notifications and alerts. **(Hadoop, EMR, EC2 ,S3, Oozie , NodeJs , Java)** **Other work at Apigee** : POC on Spark, Spark Sql, Hive PrestoDB, Azkaban Scheduler, Sqoop
Post Graduation Projects	• **Implementation of Select Query in Hadoop** : A map reduce engine, with RDBMS queries - Select query with where, groupby, having, avg as an input where it generates an output stored as a file satisfying the query. **(Hadoop,Java)** • **Amazon Project (Transit Time Analyzer and Route Generator)** Route Generator : Route generator generates a road map between a source and destination location within India using a database of locations from geonames.org . The route based on algorithm very closely resembles that on Google maps and transit times closely match. The routes are generated based on well connected locations. **(python , OpenGL)** Transit Time Analyzer : For a delivery system where there are records of actual transit time taken, this extension of Route Generator gives probable reasons for such exceptions . **(python)**
Education	• B.Sc Computer Science Fergusson College, Pune 60.16 April 2012 • M.C.A Computer Science, Pune University CSD (4.28 GPA on 6) May 2015

Certifications   Udacity - Introduction to Computer Science(Building a Web Search Engine in Python)

Other Activities

- Member of Placement Cell - PUCSD. Placement Cell 2013-14.
- 2nd Runner Up in C Coding Competition - Inspiria 2015(C-Scuffle) at PCCOE.
- Single Engine, Multiengine Land,Instrument Rated, Commercial Pilot - DGCA,FAA CPL Holder - 297.5 Flight Hrs (2007 - 2008)
- Basic Astronomy Course - Jyotirvidya Parisanstha - 2012