

Harsh Pandey

Northwood Dr, South San Francisco, CA 94080

3+ years of data engineering at the forefront of Healthcare & Digital marketing & E-Commerce. Passion for designing a drug's end to end journey as a pipeline, from inception to it's reaching the right patient.

Cell: +1 (317) 793 8959.

Email: harshvardhanpandey@hotmail.com

Linkedin: linkedin.com/in/harshvp.

Github: github.com/geekidharsh

Experience

Sr. Data Engineer, Bioinformatics Research
South San Francisco, CA.

Amgen
June 2019 - Present

Bioinformatics software engineer building data pipelines drug discovery and inception.

- Working on python, AWS and Google cloud warehouses, mining, SQL, turning Rest APIs to data pipelines.
- Support highly readable and scalable analytics in Single Cell RNA and Bulk RNA etc.
- Design and implement scalable research models on AWS and Google Cloud using EC2, Tableau, Aurora, BigQuery.
- Lead technical conversations and requests from research analysts, implement solutions and deliver results.

Data Engineer - Digital Marketing
St. Louis, MO.

MilliporeSigma
Dec 2017 - June 2019

Data Engineer cum scientist with a history of generating over \$120 million in revenue with scalable analytical products, 9-10x return on budget.

Notable Projects:

Advanced Keyword Research Database - Developed Analytical model for Data-Driven Keywords optimization. Consolidated dataset from internal platform & external analytics APIs: Google Analytics, Omniture, etc. Designed a final data model that uses attributes from all sources, weights keyword relevancy and generates final recommendations to bid Ads on.

Results: 30% revenue from ad optimization, higher conversions, \$100 million return on a \$12.6 million budget. Utilized Google BigQuery for warehouse, cluster on google cloud pandas, pySpark, SQL, Power Bi, and Data Studio.

Link Equity Flow Project - Mined product names from a massive catalog of Chemical Safety Documents (MSDS) stored in PDF to recommend relevant product to the user. Fuzzy matched product names in MSDS pages with product inventory, resulting in adding direct <ahref> links products. Changed product HTML markup based on the analysis. Automated the entire flow in python, created ETL, matching in SQL and created a reporting model in Data Studio to monitor these improvements.

Results: Better link equity leading to 30% improved product sales from MSDS and 'Add To Cart' rate spike.

Data Pipelines Migration- Led software development migration to bitbucket without affecting ongoing Reporting environments. Rewrote over 10+ existing pipelines from python 2.7 to python 3.x and cross-machine compatibility of Unix to Windows servers using python OS package. Implemented local dockerization and environment control using Anaconda.

Results: Improved SDLC across developers from multiple departments, faster feature improvement & maintenance.

Analyst - E-Commerce
Burlington, MA

MilliporeSigma
April 2016 - Dec 2017

Overall 33% revenue growth in Paid and SEO combined from some of the projects listed.

Notable Projects:

Search Console App - Ideated, designed, conceptualized and built an end to end python based data pipeline that brings Organic traffic datasets from Search Engines into our internal data warehouse. Built Oauth based connectors to external APIs, clean and load data into Google Bigquery, designed database Schemas, tables. Automated the pipeline using cron.

Results: Powerhouse Database supporting digital growth in SEO, PPC. Team's first big data warehouse. Improved existing dataset by over 15x, helped lead the gateway to future programmatic development and make data-driven decisions in Digital Marketing.

HTTP to HTTPS Migration - Helped the Web Development Team in analyzing the impact of migrating a large e-commerce website successfully from HTTP to HTTPS with a minimal negative impact on organic traffic. Used metrics from organic traffic, product sales, etc. to recommend, HTML canonicalization and batch change in the order of priority. Built Tableau Reports monitoring the impact month over month with KPIs.

Results: The site still brandished a 1% increase in organic, impressive for a large website.

Research Assistant, Dept. of Computer Science
Springfield, IL

University of Illinois
Feb 2015 - Dec 2015

Notable Project:

Data Driven Roadside Safety Campaign - Using medical records Utilized SQL and bash automation scripts to manage databases monthly, perform complex analysis, designed ER models and database for Upper Management. Utilized data from annotated medical records and DMV to understand driving pattern and common casualties and conclude recommendations to improve.

Result: Used feedback to lead statewide social media campaigns on roadside safety awareness and improving safe driving.

Backend Web Engineer
Raipur India

Fine Brains Consulting
June 2013 - Jul 2014

Achievements

Best Scalable SEO Strategy, 2017, 2018. Team Member.

Best In House Digital Team, SMX Awards, Team Member.

Top 10% Python on Stackoverflow: ~3M users helped: Profile: <https://stackoverflow.com/users/story/5893695>

Education

MS, Computer Science. University of Illinois at Springfield, IL. May 2016

BE, Computer Science. CSVTU, Bhilai, India. June 2013

Certification

Google Certified Data Engineer - Google Cloud

Google Analytics Certified - Google Analytics

Google Certified in Building Relient Serverless Platforms - Google Cloud

Skills

Languages: Python, SQL, HTML, CSS, Bash.

Tools and Systems: Google Cloud Platform, NoSQL warehousing, BigQuery, Tableau, Data Studio, Power BI.

ETL & Analytical Tools: Knime, Pentaho, Google Analytics, Webmasters, Brightedge, Ahref.

Publications

"Present Scenario Analysis Of Green Computing Approach In The World Of Information Technology", Undergraduate Academic Research Journal (UARJ), ISSN : 2278 1129, Volume-1, Issue-2, H. Pandey.