

Harsh Pandey

3+ years of data engineering at the forefront of Biotechnology & Digital marketing & E-Commerce. Passion for designing a products end to end journey with data, from its inception to reaching the right consumer.

Address: South San Francisco, CA, United States 94080 **Mobile:** +1 (317) 793 8959.

Personal Email: harshvardhanpandey@hotmail.com

Linkedin: linkedin.com/in/harshvp.

Github: github.com/geekidharsh

Stackoverflow: https://stackoverflow.com/users/5893695/geekidhars

Education

MS, Computer Science.

University of Illinois, Springfield.

May 2016

BE, Computer Science & Engineering

CSVТУ, Bhilai, India

June 2013

Skills

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

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

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

Experience

Senior Data Engineer, Amgen

South San Francisco, CA.

June 2019 - Present

- Bioinformatics software engineer building data pipelines from drug discovery and inception.
- Technical lead of Single Cell RNA sequencing pipelines as well as supporting analytics in Single Cell RNA and Bulk RNA etc. Liaison between Genome Scientists, Data Scientists and Product Owners.
- Working on python, AWS and Google cloud warehouses, mining, SQL, turning Rest APIs to data pipelines.
- Design and implement scalable research models on AWS and Google Cloud using EC2, Tableau, Aurora, BigQuery.

Data Engineer, MilliporeSigma

St. Louis, MO.

Dec 2017 - June 2019

- Part of Advanced Digital Marketing team whose solutions that have generated over \$120 million in revenue with scalable analytical products with 9-10x return on budget.
- Designed and launched data pipeline that now serve as the building blocks of MilliporeSigma's marketing data lake, using Google Cloud Bucket, Big Query and Google Data Studio.
- Designed and wrote software that stores, manages and updates web logs from Akamai CDN and funnel Log analysis in Sumo Logic, enabling large scale SEO optimization and troubleshooting analysis using these web logs.
- Wrote data ingestion pipeline in python to support unsupervised ingestion of billions of rows of data end to end, using both cron and Google Cloud DataProc.
- Designed highly sophisticated SQL scripts, querying billions of rows of Google Analytics data in Big Query.
- Delivered results to analytics pipelines used by high level stakeholders and product managers.

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.

Associate Engineer, MilliporeSigma

Burlington, MA

April 2016 - Dec 2017

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

SearchConsoleApp:

- Designed, conceptualized and built an end to end pythonian data pipelines 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. 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 brandished a 1% increase in organic, impressive for a large website.*

Research Assistant, University of Illinois

Springfield, IL

Feb 2015 - Dec 2015

Data Driven Traffic 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 analyze driving patterns and common casualties.
- Used feedback to lead statewide social media campaigns on roadside safety awareness and improving safe driving.

Engineering Head, Fine Brains Consulting

Raipur India

June 2013 - Jul 2014

- Implemented innovative marketing strategies for real estate properties digital and social media channels by analyzing offline vs online channels.. Achieved \$1.2M in sales of agricultural plots through digital marketing.
- Led team of 9 members responsible for development and management of multiple web portals and social media accounts of leading clients in hospitality & fashion.

Achievements

- Search Engine Land Best Scalable SEO Strategy, 2017, 2018.
- Best In House Digital Team, SMX Awards, Team Member.
- Top 10% Python on Stackoverflow

Technical Certifications

- Google Certified Data Engineer - Google Cloud
- Google Analytics Certified - Google Analytics
- Google Certified in Building Relient Serverless Platforms - Google Cloud

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.