

# ASMIT KUMAR

+91-9620151287 | asmitkumar11@gmail.com | [Github](#) | [LinkedIn](#) | [Website](#)

---

## SUMMARY

Results-driven engineer with extensive experience in robust, scalable backend application development. Thrives on collaborative problem-solving and rapid prototyping to deliver high-performance solutions efficiently, while maintaining a focus on delivering version one first and iteratively refining towards optimal functionality and user experience.

---

## EXPERIENCE

### Senior Software Engineer, Hearst Magazines

Dec 2020 - Present, Bengaluru

#### Product data-ingestion pipeline

- **Impact:** Achieved a 20% decrease in time taken to create commerce articles
  - Engineered a solution to manage diverse data feeds via FTP, SFTP and direct APIs, overcoming challenges with unstructured data and file formats.
  - Implemented a user-configurable end-to-end solution to minimize code changes.
  - Processed over 15 million SKUs daily and reducing reliance on crawling.
- **Technologies:** Python, Go, FastAPI, PostgreSQL

#### Product matching service

- **Impact:** Saved \$200K spent annually on a third party solution.
  - Crafted a solution for product unification across retailers, employing universal product identifiers and image/text matching.
- **Technologies:** Python, openAI

#### Product categorisation

- **Impact:** Ad revenue up by 8% due to improved targeting.
  - Re-architected the legacy product categorisation service by implementing a mapping system to associate each product with its nearest category in the Google taxonomy.
- **Technologies:** Python, openAI

#### Integrating fallback data sources

- **Impact:** Improved success rates with product data API calls
  - Designed a solution to have a chain of fallbacks to different data sources or even have a custom one enabled based on feedback.
- **Technologies:** Golang, Redis

#### Infrastructure Migration to Kubernetes

- **Impact:** Migration completed with zero downtime.
  - Ensured seamless transition of 60+ microservices to Kubernetes (K8s) with feature parity and load-bearing capabilities.
  - Developed custom data pipelines for rigorous testing, maintaining data integrity during migration and synchronizing 3TB of PostgreSQL data between AWS RDS instances.
- **Technologies:** K8s, AWS, Apache Airflow, Datadog, Pagerduty

#### Sales Event Optimization

- **Impact:** Increased user engagement and positively impacted revenue by 10-15%.
  - Implemented custom Airflow pipelines for timely product updates and dynamic price adjustments during sales events, enhancing pricing strategies.
- **Technologies:** Python, Apache Airflow, PostgreSQL

### Software Engineer, Semantics3 Technologies

Apr 2016 - Nov 2020, Bengaluru

#### Web crawler

- **Impact:** Scrapes 10 mil pages/day across 300 e-commerce websites.
  - Co-developed and maintained in-house.
  - Created a system where one main service downloads web pages by sending tasks to its 400 follower nodes across different cloud platforms like AWS and GCP, as well as other third-party forward proxies.
  - Implemented features like bot evasion techniques, health checks, efficient ip rotation, automatic routing through better performing proxies.
- **Technologies:** Go, Perl, RabbitMQ, PostgreSQL, Redis

## Content extraction

- **Impact:** Addressed data issues promptly, garnered positive API user feedback, and achieved cost savings by halting crawls for sites with data issues.
  - Led a team of six people and maintained extraction rules for over 1000 e-commerce sites.
  - We even crawled google occasionally for some custom crawl related tasks.
  - Wrote an automated QA system to identify and flag any data extraction/quality issues for any supported e-commerce sites.
- **Technologies:** Perl, Go, Regex

## Misc

- **Impact:** Operational Optimization, Client Engagement, and Cost Efficiency
  - Established a systematic process to regularly assess and decommission sites based on API call quantity.
  - Engaged regularly with customers to add features, address data issues, and ensure overall customer success.
  - Implemented data monitoring pipelines across API, data extraction, and crawling processes to ensure optimal uptime.
  - Implemented custom data pipelines tailored to meet diverse customer needs.
- **Technologies:** Perl, Python, Apache Airflow

## Data Engineer, Semantics3 Technologies

Nov 2013 - Apr 2016, Bengaluru

### Supervised data extraction

- **Impact:** Clean data extraction and normalization to ensure adherence to a predefined schema, facilitating efficient data management and analysis.
  - Added supervised extraction for sites based on requirements.
  - Automated weekly complete site crawls to ensure comprehensive data gathering and analysis.
- **Technologies:** Perl, Regex, MongoDB

### Crawl data validator

- **Impact:** Reduction in customer complaints for data.
  - Incorporated checks for data accuracy during extraction, including schema validation, URL verification, and price validation.
- **Technologies:** Perl, Regex

---

## SKILLS

- Programming Languages: Python, Go, Perl
- Databases: PostgreSQL, Redis, MongoDB, Elastic Search
- Technologies: FastAPI, Apache Airflow
- Working knowledge: Java, Node.js (Typescript)
- Containerization: Docker, Kubernetes
- Monitoring and Logging: DataDog, Grafana
- Large language models: Clip, GPT-2 (Huggingface Transformers)
- Misc: Prompt engineering

---

## EDUCATION

### Bachelor of Technology: Computer Science Engineering

Punjab Technical University

Jun 2012

### Class 12

Kendriya Vidyalaya, Dimapur, Nagaland

Apr 2008

### Class 10

St. Francis De Sales, Dimapur, Nagaland

Apr 2006