

Shreyas Hukkeri

Irvine, CA | shukkeri@pm.me | linkedin.com/in/shreyashukkeri

Summary

- SDE II at Amazon Grocery Search building low-latency, highly available services reaching customers in 9 marketplaces
- Co-designed and implemented an inline recommendations framework driving \$1M+/yr in grocery revenue, later expanded to support teams in other product categories and adopted by the core Search org
- Focused on building right-scale solutions that address customer traffic/experience while balancing development effort, infrastructure costs, and project timelines
- Enjoy challenging engineering problems that tackle complexity and scale, especially projects that require collaboration across engineering, design, and product teams

Experience

Amazon Irvine, CA

June 2021 – Present

Software Dev Engineer II (L5)

Inline Post Add-to-cart Recommendations Slot

- Co-designed and led initial implementation of the Inline Expansion Slot, featuring cart-aware recommendations shown after a customer added an item to cart on the Search page
- Expanded functionality of static server-side rendered (SSR) Search page to accommodate dynamic slot creation and content retrieval using jQuery, Javascript, and DOM manipulation
- Designed a two-step content loading strategy by adding Typescript hooks on page-load, and using AJAX retrieval after ATC event to fetch recommendations content from an existing rendering frameworks for faster experimentation
 - Strategy allowed for no regression in Search page-load metrics like time-to-interactive (TTI) or DOM size
- Ensured consistent rendering and interactivity across mobile web (mWeb) and mobile app (iOS and Android)
- Additionally, dove into browser-side rendering frameworks to accommodate dynamic slot resizing and scrolling behavior to continue displaying more recommendations after users interacted with the initially expanded slot
- Initial project experiment contributed to >\$1MM projected annual increase in grocery revenue
- Designing a new surface for recommendations on the Search page, other teams quickly expressed interest in integrating the slot for items across other product categories (like Everyday Essentials, and Sub-Same Day Shipping) so began converting feature to platform to accommodate external contributors
 - Learned to navigate working across external team stakeholders (engineers, PM's, UX) in developing a framework that could accommodate use-case specific functionality and extensibility beyond initial POC implementation
- After formalizing as a new Search-page feature, handed off to core-Search org for continued support and expansion

Whole Foods Market Search migration and expansion of In-store shopping

- Implemented new data models to bring parity of promotional pricing (ex: 'Buy 2 get 1', 'Buy one, get one 50%', etc.) across what was advertised on Wholefoods.com and Amazon grocery search results
- Engaged with product faceout rendering platform teams in Search org to address gaps in localized promotions visibility, a problem unique to selling groceries online with regional/store-level pricing and promotions
- Supported roll out of Multi-banner Same-building initiative introducing the ability to shop and pickup Amazon retail items, while customers shopped in-stores at Whole Foods Market
 - Implemented a dynamically loaded (AJAX) Typescript bottom sheet explaining how the program works, connected to a 'More info' link on the product's faceout for increasing customer familiarity

Broad Query Reformulation on Search Auto-complete

- Designed and led development of a query refinement widget in Search Autocomplete bar using Native AWS stack (Lambda, DynamoDB, S3) serving Daily Peak of 17,500+ TPS (1 Billion+ requests per day) across 9 countries
- Built offline data processing pipeline to cache recommendations and meet <10 ms avg. latency requirement for service
 - Implemented greedy algorithms to crawl through the product classification ontology and map broad search queries to most related/relevant categories, which would be surfaced in the widget
 - Setup asynchronous Lambdas to process, crawl, and generate the relevant widget entries, kicked off by event listeners connected to a file upload in S3 where product teams could place updated broad query lists to target
 - The widget entries were then cached in DynamoDB for retrieval by the Autocomplete service
- The extensibility of the data pipeline enabled re-use and faster experimentation on future projects in Autocomplete

Multi-grocer Shopping widget in Fresh Search

- Developed a multi-tab shopping widget to help customers shop across different grocery stores with a single search
- Worked efficiently across teams to devise creative solutions and deliver the widget 2 months ahead of schedule
- Kickstarted larger platform rewrite from Scala to Java, making it easier to transition existing and future widgets

Additional accomplishments & accolades

- Won three internal hackathons, developing ideas from prototypes to production ready features
- Led the deprecation of a Search widget platform, saving over \$3,500 per month in EC2 infrastructure cost
- Aided in roll-out of Amazon Prime Grocery subscriptions, resulting in double digit percent of sign-ups attributed to Search page promotional ads
- Optimized fleet resources during peak events (Prime Day & BF/CM), saving over \$500 per month and increasing the team's internal fleet credit score by 40% YoY
- Mentored teammates and lead knowledge shares on dev tooling to improve collective understanding and velocity; lunch-and-learns on machine learning, LLM capabilities, and exploring agentic workflows with 'pi' harnesses

Zscaler San Jose, CA

Oct 2020 – Apr 2021

Software Engineer

Terraform Cloud Automation

- Built multi-platform cloud automation tools using Terraform to speed up development and deployment of FreeBSD-based containers/VMs for customers and developers of Zscaler's cybersecurity suite
- Worked directly with customers to support engineering teams in adopting ZScaler in their bespoke infra
- Created on-prem Edge Node deployment scripts for enterprise customers, saving 2 hours of setup per deployment

Skills & Education

Technologies: [AWS (EC2, Lambda, DDB, S3), Docker, QEMU, Git, Linux, Distributed Systems, CI/CD]

Languages: [Java, Python, TypeScript, C++]

Personal proj.: [Self-host homelab server w/ containerized Nextcloud & Tailscale, and VM's for local LLM deployment]

B.S. Computer Science, *University of California, Irvine*