Google Resume

Maricia Scott, Software Engineer

I joined Google in Sept, 2002. This was last updated in April 2012.

Significant Project Accomplishments at Google

(7/2006 - present) Google News

We organize the world's News. Fast. We haven't yet managed to do it before the news actually happens, but we are working on it. News impacts ~1 Billion users/week, through news.google.com, News search, and News Universal.

I transitioned from the Borg frontend to work on backend infrastructure for Google News in 2006, became Tech Lead of Build Infrastructure in Q1 2009, and Tech Lead/Manager for MTV Infrastructure in Q3 2010.

Over the years I have been a key contributor to our pre-indexing ("build") side pipeline, as well as working on many smaller-scale projects . I have also done a lot of production and process work that has helped keep both the product and the team scale.

- **Technical Leadership (TL 2009-2010, TLM 2010-present)**

  My role here includes:
  - Handle technical supervision, prioritization of tasks, bugs and emergency production issues
  - Work with the team to take care of issues in this large, critical system
  - Do team goal planning and new project generation
  - Make sure the“little things” get done, either by me or by delegating
  - Help bring new team members up to speed
  - Over the years, have advised on the design of most projects within News
  - Handle management tasks related to being TLM for a team of ~9

- **News/Web Pipeline Merge (Q1 2011 - present)**

  Since its beginning, News has run a completely separate index-building pipeline from Web. In 2011, Dave Weissman (from Freshdocs team) and I started investigating how to merge News into the Web pipeline. The goal of this 18 month, 15+ person project is to merge the News indexing pipeline into Web (Freshdocs and Alexandria) to reduce maintenance/prod load, share indices for OneTree and share features. Full list of features to merge is at News Features Breakdown and here is the current Snippet. With Dave, I:
  - Co-lead the project
  - Determined project scope and broke into subprojects
  - Coordinated with many teams (News Quality, Springfield, Webrefs, Authors, Images, Sitemaps/Feeds, History, Video, Goldmine, Alexandria, Finance)
  - Provide technical consultation on the News side for many subprojects

- **News OneTree (Q1 2011 - present)**

  One Tree is the serving equivalent for the build merge; the goal is to integrate the News serving code (mustang ascorer, superroot) into the Web serving stack. The base tree (older documents) for News Universal is live; instant is rolling out now. The News onetree implementation work is being done primarily by Deepak’s and the mustang teams. My involvement included:
  - With Dave Weissman, coordination work to make sure things moved smoothly

- Figuring out how to allow work (and launch) of OneTree to progress in parallel with the indexing pipeline merge
- Making sure that the nothing slipped through the crack between build (project above) and serve
- Eval setup and help

- **Production Responsibilities (2007 - present)**

  I am one of the "go-to" people on Google News for debugging the "hard production problems" both within our code, and outside interactions (bigtable issues, borg issues, gfs, etc). I am one of two people that do most of the production maintenance for our Build datacenters (running 350+ jobs when last I checked).
  - Prior to SRE onboarding, wrote many and refactored most of our borgcfgs for build side (10K+ lines of borgcfg); familiarity with all
  - Work with SRE and News team to make sure the system has 99.99% uptime, realtime indexing (1-2 min crawl->index), and reduces pages and outages
  - Have held the pager for 2-3 times/quarter since Q1 2007.
  - Found and initiated current SRE relationship.

- **News Build Pipeline (implementation 2006-2007. production, maintenance, improvements and advice through 2012)**

  The Article Repository Bigtable was a refactoring of the News pipeline; previously all News data was written out to logs by the crawl. The refactored design makes it easier to add new features, modify old content, and modularize processing. Over the last 6 years, the team has scaled the build pipeline to 7 different bigtables containing all of our content (articles, images, hubs, clustering), ~100 scanlets, 5 mapreduces, and several one-off servers. The entire pipeline is running 350+ jobs per Build DC, with an end-to-end processing time of ~1-2 minutes. See Life of an Article for more details.
  - Designed and implemented the initial repository bigtable, including helper functions that make it relatively easy to add new columns quickly.
  - Wrote the first populator, which takes FetchReply's from the crawl log and inserts them into the repository.
  - Hooked the original clustering workflow up to the new repository, making it possible to use in production.
  - Continued maintenance on this critical component of code - it has held up well for 6 years
  - Infrastructure and documentation for how to write and test processes that work on the repo.

- **Multihoming Build Pipeline and Recrawl (implementation 2008-2009. Design advice for build components ongoing)**

  I led the team (of Marisa Bauer, Martin Law and I) working on adding multihoming and recrawl to the build pipeline. Our production build pipeline had been singly homed -- this is fine for many products that can tolerate a 10 hour PCR in their preprocessing stages; but News can't be 10 hours old. We needed a way to make the entire Build pipeline (from Crawl to Index) resilient to a datacenter outage, without going stale. We also needed a way to be able to make more disruptive changes (such as backfilling years of data) without disturbing the running system. At the same time, Marisa was working on recrawl, which required rethinking the way we stored articles, so we wanted an architecture that could address all three problems. We now have a system that stores all copies of articles, then picks a "live" version to store in the primary article repository, and can failover from one of our build datacenters to the other in 3 hours without going stale.
  - Led design process and was the go-to person for multihoming-friendly architecture of new components
  - Modified article repo populators to handle multihoming, including writing extractor scanlet
  - Many code reviews for other parts of the project
  - Productionized the new setup (bringing up two datacenters, borgcfgs, monitoring, etc)

- **Process implementation and scaling (2008 - present)**

  As the News team scaled, the code quantity scaled with it, but many of our ad-hoc procedures did not. Starting in 2008, I worked on tackling some of these issues, primarily in getting a weekly Release Procedure ("Release Train") up and running, which has allowed our team to maintain a fast push cycle while increasing testing and stability of the production pipeline. Creating this process involved a huge amount of coordinating, cajoling, borgcfg refactoring, and new-binary debugging.
  - Implemented weekly release train procedure, coordinating swes in 4 offices, releng, testing
  - Brought up and maintained a "staging" pipeline, where we run final testing and smooth out push procedure for one-off pushes (this is now maintained by release coordinators).
  - With Martin Law, Andy Golding and Larry Sun, automated much of the release train process and verification.
  - Average weekly push pushes changes to 10-20 binaries (including flag flips)
  - Was Build Release Coordinator 25 times from 5/2008 through 9/2009, did my last rotation in Q4 2010.

- **Other projects (ongoing)**

  - Editor's Picks (2010)
    Editor's picks was a project to showcase publisher content on the news front page (see "Editor's Picks" gadget in RHS of Google News). I wrote the prototype experiment for this project and consulted with design; Yogita Mehta took over and did the proper implementation and all the bells and whistles.
  - Hosted News on Google News (2007 - 2010)
    This project adds the hosted news feed to news.google.com (see: Hosted Query). I took over project when another team member went on leave, and am now the primary maintainer (including bug fixes, small new features)
  - Mobile Pipeline Integration (2009)
    With Haavard Kvaalen, Egon Pasztor, Nilesh Agrawal, this project integrates mobile News articles into the regular Google News pipeline, so that we can turn down the separate Mobile News pipeline, which is running on old infrastructure.
  - Testbed fixes (2008 - ongoing)
    As more people joined the team, the cumbersome testbed process did not scale. I helped make it easier to bring up testbeds and debugged countless issues with testbeds.

(12/2009 - 12/2010) New One Pass

In late 2009, helping publishers easily put content behind a paywall became a priority request from Eric Schmidt; the News team carved out a sub-team to work on it. I lead that team, eventually handing off the project to Omi Chandiramani. My role on the project involved:

- Technical leadership, design input, cat herder
- "Acting" PM when PM-less (the project went through 3 PMs)
- Main eng contact working with Checkout team, finance, legal, sales, product, support
- Most backend code reviews (access server, report generation, cancellation tool)
- Coordinated successful Dogfood of product

(7/2004 - 7/2006) Borg Status Console

I worked on the UI/frontend for Borg. This is the main window engineer's have on Borg, and has a huge amount of data that needs to be made visible in a coherent way. My major accomplishments on this project include:

- With UI designer Chad Thornton, redesigned the layout of the Borg UI. I was then responsible for the implementation of the redesign.
- Designed and implemented many features beyond the basic layout, including:
  - Machine page search box
  - User page regular-expression searching for names, users, cmdlines
  - Cookie preferences to allow users to customize the UI
  - Help links for all the major functionality
  - Many smaller features (user limits, usermaps, recent termination log, overrides, pending tasks) that needed to be exposed on the UI when added to the borgmaster, often on a short turnaround.
- Machine utilization accounting. Since these numbers needed to be displayed on the UI, I had to get accurate measurements for all resource usage.
  - Debugged and kept up with resource accounting
  - Improved upon and added to the borgmon resource rules, adding user and cell graphs to the UI.
  - Helped initial efforts by SpaceJam to get resource accounting information

(5/2005 - 9/2006) TV Showtimes (20% project)

This project came out of a grassroots 20% brainstorming session of women engineers. 4 of us began working on the idea of one-boxing TV search results; eventually the project was taken over by a fulltime team.

- Co-tech lead of the project, with Marisa Bauer
- Within the 4-person team, worked on initial concept and design.
- Worked on mustang build/scoring, based on work by VideoGoogle.
- Helped bring the fulltime team up to speed (majority of whom were completely new to Google)

(1/2004 - 7/2004) Blog Search (20% project)

I joined the initial blog search team as a 20% member, with plans to transition full time from Enterprise until the whole project was put on hold for 6 months. While working on it, I:

- Worked on the frontend, including implementing the blogger look-and-feel (designed by Ellen Beldner), added special search operators for blog title/author/etc, and added an rss output format.
- Implemented the initial prototype for blogger profile pages, by using Googler data available in feed format (as is used by Moma).

(10/2002 - 7/2004) Enterprise (Google Search Appliance)

Enterprise was my first project at Google; working in the Testing Technology team I was responsible for testing-related infrastructure. On that project, my major accomplishments include:

- (2/2004-7/2004) MAD (Massive Amounts of Documents). In order to test Enterprise crawling on a large data set, we needed to create or obtain 20 million documents: 30% html, 20% pdf, 50% other(.doc, .ppt, .xls, .txt, .rtf), a 20X increase in the number of documents we previously had. This was accomplished using miniweb (see below) to generate most document forms, and transferring 4 million pdfs from the new science-search team to our servers.
- (4/2003 - 7/2004) Enterprise Automated Testing Framework. Wrote a generic framework for running regression tests on Enterprise machines. Allows someone to specify a config with sets of binaries to run (such as build installation, administrator setup, loadtest, etc), and generates a set of status pages indicating current status, failures, etc. Meant for nightly regression tests.
- (4/2003 - 9/2003) Miniweb. Proxy that simulates a webserver. Creates deterministic, randomly generated webpages with links to other miniweb pages, for use in simulating a intranet for the

Enterprise crawl.
- (10/2002 - 7/2004) Freshmaker The Enterprise Installation tool. Took over the tool written by an intern, revamped it and expanded it to handle 15+ possible different installations of 2 different types of Enterprise systems (deployed and new real-time versions) over both single machines and clusters, and made it push-button so the testers could re-install with minimal manual intervention.
- (1/2003-7/2004) Adminrunner regressiontest, Loadtest, Monitoring test Test infrastructure that can be used within Enterprise regressiontests. Tests the "adminrunner" (SysAdmin UI for the GSA), serving loadtest, monitoring uptime/crashing of backend binaries.

Significant non-Project Accomplishments at Google

(2007 - 2010) Career Mentoring Intergroup

In 2007, I started a GWE-based group to work on Career Mentoring, in response to an obvious need on the GWE mailing list. In 2008, we expanded the program to all of eng worldwide, and in 2009 having been working on a new program MentorsOncall, in order to scale.

- Initiated group - I wanted a mentor, many other people seemed to want one too! Let's do it.
- We are a "go-to" group to talk to when new career development resources are starting up.
- 6-month alpha program with 30 (?) pairs in gwe community in Q3 2007
- 6-month beta program with 60 pairs worldwide over eng in 2008. China-Hr team found program and replicated it for that office.
- MentorsOnCall, a "short-term" mentoring program rolled out in 2009. Piloted in NY/CAM/PIT. Now worldwide, with major groups in US-West, US-East, EMEA, ramping up JAPAC.
- More than 100 mentees paired in MentorsOnCall
- Initial survey results show positive impact for mentees
- Was a mentee in alpha program, and mentor in beta and MentorsOncall

(2009 - present) GoogleGeist Career Development Subgroup

With Andrew Wegley, Nadav Eiron, Yoah Bar-David, Andy Warner, working on an eng-wide award to recognize hard workers on non-launch projects.

(2006 - 2010) Various GWE initiatives

- (2006, '07, '08, '09) US Anita Borg Scholarship committee. Participated in the scholarship retreats in Mountain View
- (2006, 2007) Australia Anita Borg Scholarship comittee. Attended retreat in 2006 in Sydney.
- (Spring 2007) With Marisa Bauer, Lucy Zhang and Alice Tull, started a series of 20% brainstorming sessions for female engineers.
    - 3 teams of female engineers began investigating 20% projects such as gmail features and video closed-captioning.
- (Winter 2006) Organization committee member for National Engineering Week events for middle and high school girls
- (Winter 2006) Organization committee member for Take Your Child to Work day
- (Winter 2005) Femeng diversity engineering interviews

(2003 - present) Interviewing and Recruiting

- (2003 - 2010) Interviews; particularly college hires
- (2007 - 2011) New college grad Hiring Committee
- (2005) Stanford University oncampus interviews

(2/2007) EDGE Leadership Training

(2011) Mars Leadership Training

(Fall 2006): Head First Design Patterns reading group

Co-organized (with Will Robinson) and participated in a reading group for the Head First Design Patterns book.

Awards

- Peer Bonuses (12 Total): 2004 (1), 2007 (1), 2008 (3), 2009 (1), 2010 (3), 2011 (2), 2012 (1)
- Spot Bonuses (8 Total): 2003 (1), 2004 (1), 2008 (1), 2009 (1), 2010 (2), 2011 (2)
- Patents submitted (3 Total): 2007 (1), 2011 (2)