

James Harley Gorrell

46 Hampstead Rd NW Calgary
403-999-1519 (cell) | harley@panix.com

Profile

Experienced computer programmer seeks to advance the understanding of biological systems. Enthusiastic with good professional communications skills and takes delight in a challenge.

Skills

- Programming: Python, C++, Java, Perl, elisp, and others...
- SwEng: K8s, docker, CI/CD with Jenkins & github, Sphinx, ...
- Databases: Postgres, Sqlite, Mysql...
- OSes: Linux, Darwin, OpenBSD, and others...
- Sys Admin: Linux clusters and servers, Ansible, NFS, Networking...

Employment

Consultant, HBA Spectro, 2019 -- Present

- Setup and administer linux hosts and postgres databases used for traffic modeling and GIS visualizations. (~10 hosts; ~700GiB memory; ~120TB disk)
- Generated documentation with Sphinx.
- Setup monitoring with Prometheus and Grafana.
- Administer office network and backups.
- Wrote web applications with Python, Flask and Javascript.
- Dockerized applications.

Software Engineer, Healthtell, 2016 -- 2019

- Used Python/Flask to write an application for tracking lab samples used in the development of clinical tests.
- Wrote `Ctrl-F` for auditing desktops for PHI data, using the RE2 library for quicker searches.
- Ported a system for the processing of MALDI traces from Java to Python/R. Also added a feature to watch for new data and process it as data is added. Reduced its runtime from 8hrs to 2.5h & minimal user interaction.
- Moved internal applications to Docker.
- CI/CD with jenkins, along with Sphinx docs.
- Mentored junior SW developers.

Software Engineer, Rackspace, 2016

- Worked on server-side javascript for the "Cloud Monitoring" product at Rackspace.
- Debugged and updated the version of Thrift used for internal API calls.
- Updated internal diagrams & documentation for the system.

Software Engineer, Invitae 2013 -- 2015

- Developer and maintainer of Geneticus an in-house application for the curation of genetic conditions and disease variants.
- Wrote and oversee operation of the data archiver of primary sequencing data and analysis results. Migrating data to S3 and keeping an index in Postgres.
- Support and development of other in house applications and users.

Software Engineer, Affymetrix, 2001 -- 2013

- Used Django to build the Sequence Selection Tool website: A framework for supporting generation of custom reports of Imputation and R2 data for Marketing. Data is visualized in the GenomeBrowser and IGV.
- A lead developer of Affymetrix Power Tools (APT). APT is a C++ framework and programs for the analysis of Affymetrix Gene Chips. APT is able to process thousands of chips at a time in less time than the prior analysis. The APT core is used as the foundation of commercial Affymetrix and third party software.
- Principal developer of the TsvFile and File5 libraries of APT. TsvFile is a text based format; File5 is an easier to use layer built upon HDF5 to simplify common APT usage. Most of APT data is in these two formats.
- Wrote the configure and make system for APT.
- Converted 10 years of CVS history to SVN.
- Administer the Bamboo, Fisheye and JIRA services used by the software development team.
- Wrote the C++/Perl bindings for ExACT, the precursor to APT.
- Wrote Unix and windows code to memory map several formats of legacy Affymetrix data files.
- Programmer for the sequence selection pipeline used in the design of the HG-U133 chip. Written in Perl, the pipeline analyzed probe selection regions from the outputs of RepeatMasker, Stackpack and BLAST searches of NCBI databases (Refseq, Unigene) and matches to the human genome (UCSC hg7).
- Built the Emeryville Compute Cluster, consisting of 200 Linux nodes. Starting with the requirements, assembled the cluster from 1U servers, installed and customized open source software. (Torque) Nodes boot via PXE to install the OS. Continue to administer the cluster and its services. (DNS, NIS, NFS)
- Wrote custom Postgres functions for the analysis of expression data within a database. These functions present external data files as tables which can be selected from in user SQL. These 'set returning functions' allow for ad-hoc exploration of expression data.
- Postgres database administration -- Administer 4 Postgres databases (~4TB). Wrote programs to perform backups via point-in-time snapshots or normal dumps. The backups are written to a tape library.
- Software package installer - compile custom kernels, bioinformatics and many other packages for use at the Emeryville site. Investigate and solve user problems with these packages.
- System administrator of ~30 Unix hosts in conjunction with the Emeryville IT staff.
- Programming Answer Guy - answer questions on a wide range of computing topics.

Sr. Scientific Programmer, Human Genome Sequencing Center Baylor College of Medicine, 1995 -- 2001

- Principal programmer for the 'Sequencing Pipeline' at the HGSC, a system for the collection, analysis and assembly of Human genomic data. This system has processed ~5M sequence traces and assembled and submitted ~3,500 BAC clones. The pipeline is written in Perl and integrates the use of many other programs supplied by other centers and the NCBI. Reports were written with Perl/CGI and available on the internal web site.
- Trained junior programmers in programming (principally Perl/CGI) and the HGSC computing environment.
- Acted as backup/senior system administrator for the computing environment of the HGSC. The facilities have grown from a Sparc 20 at my start to ~30 Unix machines (Solaris, Irix, Linux) with 1.4TB of disk and 14TB of tape.
- Selected and installed PBS (Portable Batch System) for our computing cluster.
- Installed and administered an Oracle 7.3 database of lab production and sequencing reads.
- Supported and trained others on the use of specialized software applications. (Phred, Phrap, Consed, Staden, GCG, others...)
- Participant in the HGSC's Whole Genome Shotgun Assembly project. Contributed a prototype to BANDIT, a project to explore genomic analysis.
- Acted as the HGSC's 'Chief Answer Guy', answering questions over a wide spectrum of computing (theory, history and practice) posed by fellow programmers.

Programmer, Project Management Systems, Brown & Root Inc., 1992 -- 1995

- Aladdin -- Author of a system, written in Progress under AIX, to estimate the time and cost to fabricate custom pressure vessels. Tracks customer information and proposal status. Exports a planned budget into the shop floor control system, SYMIX. Imports from DISASU, a vessel design program.
- Management Graphics -- Generates graphs of project status for project and company management. Data is collected from Brown & Root's Integrated Project Management System, written in Oracle, and graphs are generated using Access and Microsoft Graph.
- Equipment & Specification Database -- Designed and wrote a Paradox database to track tagged equipment used in the Exxon SYU refinery and the specifications under which the equipment was purchased.
- Warehouse Barcoding -- Programmed hand held barcode readers using IRL to manage inventory at Brown & Root's central warehouse.

Education

BSc. Computer Science, Texas A&M University, Dec 1991

Publications

- Lander ES, Linton LM, Birren B, Nusbaum C, et. al. *Initial sequencing and analysis of the human genome*. Nature. 2001 Feb 15;409(6822):860-921
- Mark D. Adams, Susan E. Celniker, et. al. *The genome sequence of Drosophila melanogaster*. Science. 2000 Mar 24;287(5461):2185-95. PMID: 10731132; UI: 20196006

- Muzny DM, Metzker ML, Bouck J, Gorrell JH, Ding Y, Maxim E, Gibbs RA. *Using BODIPY dye-primer chemistry in large-scale sequencing*. IEEE Eng Med Biol Mag. 1998 Nov-Dec;17(6):88-93. PMID: 9824768; UI: 99042209
- Bouck J, Miller W, Gorrell JH, Muzny D, Gibbs RA *Analysis of the quality and utility of random shotgun sequencing at low redundancies*. Genome Res. 1998 Oct;8(10):1074-84. PMID: 9799794; UI: 99018119
- Ansari-Lari MA, Oeltjen JC, Schwartz S, Zhang Z, Muzny DM, Lu J, Gorrell JH, Chinault AC, Belmont JW, Miller W, Gibbs RA. *Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6*. Genome Res. 1998 Jan;8(1):29-40. PMID: 9445485; UI: 98112780

Patents

- US 8,855,939 B2 *System, Method, and Computer Product For Exon Array Analysis*

Presentations

- *Informatics for Automation at the BCM-HGSC*, Automation in Mapping and Sequencing, Heidelberg, Germany Feb 1997
- *Large Scale Sequencing Project Management*, International Genome Sequencing and Analysis Conference, Hilton Head SC Oct 1996

Other computing

- Have written a number of GPLed emacs packages, such as [dna-mode](#), and [ssh config mode](#).
- Automated personal environment utilities such as `repo-cmd` to make working with git repos quicker.

Memberships

AAAS & ACM