# Mark McGrew Mims

Physicist turned Software Developer, Data Scientist, then Devops Engineer.
In love with data driven decision making and data engineering... at any scale.

## About

I'm a Data Plumber. I spend most days designing/tuning/fixing/banging-on data pipelines and the infrastructure behind data-intensive applications of all sorts.

Originally trained as a Physicist (Ph.D. researching how quantum computers can learn to tolerate noise and errors), I've since had an incredible career out in the world where Data Science meets DevOps and infrastructure engineering.

I've been lucky enough to work for companies like Canonical, the folks behind the Ubuntu operating system, Infochimps, and now Silicon Valley Data Science. Places where talented teams of people are innovating and leading open-source communities and the tech industry as a whole.

I've been a consultant for a lot of my career, which gives me great insight into a vast range of data-related problems people have as well as how to help solve these problems across industries and technology stacks.

## Current Passions

- training data scientists. I'm building a class that's essentially "Just enough data engineering" for data scientists in Berkeley's MIDS program and I can't wait to start teaching it Fall 2017
- adopting test-driven methodologies into building data pipelines... keeping pipelines strongly connected to the actual problems that need to be solved
- and then pretty much anything to do with AI in infrastructure automation :-)

## Experience

2017-Present   **Lecturer** - Master of Information and Data Science (MIDS)
*UC Berkeley School of Information* (Berkeley)

Design, develop, and teach the introductory *Just Enough Data Engineering* course for the MIDS program.

2014-Present   **Principal Engineer**
*Silicon Valley Data Science* (Mountain View)

Building data science pipelines for various enterprise clients.

Silicon Valley Data Science is a boutique consultancy specializing in solving your toughest Data problems. We provide Data Science, Data Engineering, Architectural Advisory, and Data Strategy services to a wide range of clients across industries.

As a Principal Engineer with SVDS, I typically set technical direction, provide technical guidance/mentoring, and write code to help deliver client and internal projects:

- *Data Engineering* - multiple engagements in Retail and Entertainment - Providing the technical underpinnings of client initiatives to better understand customer activity and take action in an appropriate and timely manner. Solutions vary by client, but were primarily using technologies such as Terraform/Ansible to automate data pipelines on AWS public cloud infrastructure and integrate with on-premise client datacenters, retail stores, etc.
- *Data Strategy* - multiple engagements in Retail and Entertainment
- *Architectural Advisory* - for client engagements across industries
- *Speaking* - conference talks and workshops (Strata, Spark Summit, DataDay Austin/Seattle, Hadoop-With-The-Best, Enterprise Data World)
- *Internal R&D Projects* - spearheaded projects for data-platform and hybrid devops for data pipelines and build/maintain some of our internal tooling (docker images and ansible roles for cdh and cm-api development)

2013-2014   **Principal Data Architect**
*Infochimps* (Austin)
now *Computer Sciences Corporation (CSC)* (Falls-Church)

Building managed data science pipelines for the enterprise.

Infochimps, rebranded as CSC's Big Data and Analytics (BD&A) Group, provides a wide range of fully managed data science and analytics services for the enterprise. My primary accomplishment here was driving work to adapt the Infochimps cloud-based products over

to "dedicated rack" openstack-based solutions to meet the data pipeline needs of CSC's enterprise customers.

As the Principal Data Architect, my contributions included:

- *Product development* - Worked with a team of architects and product leads to define the actual product offerings, then communicate these offerings to sales and marketing teams. Helped build tools to simplify pricing and hardware configuration
- *Solution architecture* - Brought in to pinch-hit infrastructure and architecture with customers. Built service deployments for large insurance and financial services organizations... including quite popular telemetry based data pipelines within the insurance industry
- *Infochimps platform architecture* - Worked alongside an incredibly talented development team to refactor and adapt the Infochimps architecture to integrate with other CSC acquisitions and meet the data pipeline needs CSC's customer base
- *Development planning* - Worked with the development team to help capture and translate requirements into actionable projects/tasks and help schedule and prioritize these for development
- *Ops process development* - Helped the infochimps ops team through the training, process development, and tooling needed to adapt to ops challenges of hybrid public/private cloud managed services
- *Devops* - Helped develop tools around RH-OSP Foreman-based Openstack deployments and integrated what were primarily chef-based infochimps platform component workloads

Ultimately responsible for all tooling and process to support production rollout and lifecycle for dedicated-rack BD&A product installs.

| | |
|---|---|
| 2011-2013 | **Software Engineer** - Ubuntu Server Team<br>*Canonical Ltd.* (London) |

Building DevOps tools for Ubuntu Server.

Part of the team working to build Juju, a new suite of DevOps tools for Ubuntu Server. Developing juju charms to orchestrate various services throughout the enterprise. Designing / developing APIs and tools surrounding the juju DevOps stack. Integrate / test deployments on LXC, bare metal-as-a-service (MaaS) as well as EC2 and OpenStack cloud infrastructures.

Implement data-intensive service stacks and using Juju to capture, test, and model data science pipelines.

| | |
|---|---|
| 2012-2013 | **Visiting Scholar** - Department of Physics<br>*Utah State University* (Logan) |

Research in data management and data modeling.

Interested in data plumbing and the toolchains surrounding data science pipelines and how they effect subsequent results at various scales. Working to apply Test-Driven and Behavior-Driven Software Development techniques to data science pipelines as "Sanity-Driven Data Science."

| | |
|---|---|
| 2010-2011 | **Cloud Architect**<br>*Archethought* (Austin / Boulder) |

Building private and hybrid clouds and cloud applications for universities.

Archethought is a consulting firm specializing in designing and building *private and hybrid clouds* for colleges and universities around the world. This helps universities take advantage of more efficient Virtualization technologies, provide infrastructure as a service within the university, and safely explore various emerging Digital Library technologies.

Design and Deliver all software, networking, configuration and monitoring needed to set up and support a Cloud Computing System, Storage Systems, and cloud-based High Performance Computing Systems. Help integrate the Cloud Computing System with existing systems and applications throughout the university environment. This includes a web-based (Rails) Cloud Management Console with account, instance, image, and storage management.

Technology Used: Eucalyptus, AWS/EC2 API, RightAWS, Ruby, Rails, Chef.

| | |
|---|---|
| 2005-2011 | **Founder / Chief Scientist**<br>*Agile Dynamics* (Austin) |

Data Science Consulting.

Designed and built a *data-driven decision support system* for the Jamaican Ministry of Education. This USAID-sponsored project serves primary and secondary educational institutions throughout the nation and gives the Ministry previously unknown visibility into the state of education on the Island. This application was built using Rails and scales dynamically in EC2 using Chef.

Designed the next generation of application for a company offering Fleet/Inventory Tracking services. This company provides web services and whitelabel web portals to track assets

using their proprietary GPS tracking hardware. The new design helped prepare for *integration into Machine to Machine (M2M) data market*.

Provide *ad-hoc data-munging* services to a variety of businesses. Designed/built web-based bulk data importers for textbook distributors to manage inventory from publishers with various proprietary and standard (ONIX) formats. Designed/built web-based bulk data importers for a game company wanting a portal for customers to manage game content.

Provided quantitative marketing tools and services for the visualization and *modeling of social networks*. Allowed for trend identification and analysis, growth rate predictions, and what-if scenarios for various network and Web-2.0 businesses. This was developed using Ruby/MySQL with Rails/GraphViz visualization.

Provided environmental simulation and modeling solutions to track pollutants in the Florida Everglades. Created numerical *hydrodynamic mass balance models* that are used to calculate tax incentives/penalties for surrounding commercial land. This was developed using Java/SWT/WebStart and interfaced with a variety of legacy apps and databases.

Technology Used: Ruby, Rails, Java, SWT/JFace.

| | |
|---|---|
| 2009-2011 | **Board of Directors** |

*LoneStarRuby Foundation* (Austin)

LSRF organizes an annual LoneStarRuby Conference

| | |
|---|---|
| 2002-2005 | **Chief Technology Officer** |

*Rational Systems* (Houston)

Building Operations/Optimization systems for the Energy Industry.

Directed delivery of two complete product lines, Rational Pipe(TM) and Rational Catalyst(TM), from conception.

Rational Pipe is software designed to *manage the commercial activities of interstate natural gas pipelines*, including contracts, CRM, tariffs, capacity release, nominations, allocations and invoicing. It was the result of a 140+ man-year, joint development project between Rational Systems and a major US interstate natural gas pipeline, utilizing Rational's Rights-Based engine.

Chief Architect for this $30M project delivered on time and on budget. Provided Technical leadership for a team of approximately thirty developers and twenty testers. Directly developed components across the system, including: gas flow, physical pipe, scheduling, and the JMX-based system management console.

Rational Catalyst is a business simulation and analysis framework used in energy production, exploration, and gathering. It is software that enables *collaborative business modeling* by integrating small disparate models of various aspects of the business together making model data available across the enterprise. Catalyst packages data mining, revision control for both data and models, and various visualization tools including configurable executive dashboards into one complete package for business analysis.

Chief Architect for the Rational Catalyst team of four developers and two testers. Directly developed add-in interface components for Microsoft Excel 2000 using MFC/ATL/COM plugins in C++.

Technology Used: Java, C#, C++, J2EE Design/Development, .NET, Business modeling, MFC, ATL, COM, Tibco, SQLServer 2000 with Analysis Services, Enterprise Hardware (Compaq/HP) running Windows 2000 Server, Windows 2003 Server, Red Hat 9 and Fedora Core 2-3, Microsoft SharePoint, Linux.

| | |
|---|---|
| 2000-2002 | **Lead Software Architect** |

*The AEgis Technologies Group, Inc.* (Austin / Huntsville)

Building IDEs for Simulation Engineers.

Principal architect of AEgis' AcslXtreme(TM) product line, a suite of *commercial simulation tools* based on the industry standard ACSL(TM) (Advanced Continuous Simulation Language). Leader of a development team responsible for refactoring and modernizing the ACSL language as well as developing a complete modern development environment for simulation engineers. Responsible for coordinating all technical activities and artifacts throughout the lifecycle of the project.

Directly developed software components across the product line: for ACSL language translation, compilation, interpretation, symbolic mathematical manipulation, numerical integration and analysis, numerical optimization, build management, simulation execution management, communications infrastructure (using both distributing and componenting technologies), and developing user interface component APIs.

Technology Used: C/C++, C#, Java, .NET, MATLAB, VB, FORTRAN, UNIX and Win32 systems programming, MFC, COM/DCOM, CORBA, SOAP, HLA, ANTLR, lex/yacc, UML, RUP, GoF design patterns, object-oriented design, component-based design, Windows .NET, Various flavors of UNIX/Linux (some components native, UI(MFC) components ported using Bristol porting tools).

| | |
|---|---|
| 1996-1998 | **Software Developer** |

*Wesson International, Inc.* (Austin)
now *Adacel Technologies, Ltd.* (Calgary)

Building Air Traffic Control simulators.

Responsible for creating and maintaining realistic aircraft movement and intelligent pilot behavior in a multi-platform, scaleable *air traffic control (ATC) simulator*.

Integrated tower ATC, radar ATC, and flight simulators in order to simultaneously train tower controllers, radar controllers, and pilots. Distributed the system using CORBA and the US Defense Department's High Level Architecture (HLA). Spearheaded the simulator port to C++ on a POSIX-compliant kernel.

In addition to movement and pilot intelligence in a soft real-time environment, responsibilities included on-site customization for systems installed in Alaska and Hong Kong, graphics programming using SGI's IRIS Performer toolkit, and developing networking tools to assist in distributing the simulators.

Technology Used: C/C++, Tcl/Tk, UNIX and Win32 systems programming, (soft) real-time process scheduling/event management, resource conflict resolution/management, network programming using TCP/IP and NetBIOS, Silicon Graphics O_2, Onyx Reality Engine, and Onyx2 Infinite Reality high-end graphics systems running IRIX(UNIX), i386 hardware running Linux, Win95, NT-4.0, and an in-house real-time OS over DOS/4GW.

| | |
|---|---|
| 1994-2000 | **Instructor** - Department of Physics<br>*The University of Texas* (Austin) |
| | Physical Science I: Mechanics (AI, Instructor of Record) |
| 1995-1996 | **Lecturer** - Department of Physics<br>*Austin Community College* (Austin) |
| | Intro to General Physics I |
| | Engineering Physics I |

## Education

| | |
|---|---|
| 1992-2000 | **Ph.D. in Physics**<br>*The University of Texas at Austin* |
| | Dissertation: "Dynamical Stability of Quantum Algorithms." Supervisor: E.C.G. Sudarshan Created a numerical model to characterize noise in Grover's quantum search algorithm. This model was then used to determine the maximum amount of noise that the bare algorithm can tolerate before failing. This is useful in determining exactly which emerging technologies will prove to be viable for implementing quantum computers. Technology Used: C++, Perl, BASH script, LaTeX, numerical solutions to ODEs, randomization, various matrix calculations (using blitz++, TNT, and LAPACK). |
| 1988-1992 | **B.S. in Physics**<br>*The University of Texas at Austin* |
| 1988-1992 | **B.S. in Mathematics**<br>*The University of Texas at Austin* |
| | Thesis: "Path Integration on Multiply Connected Configuration Spaces." Supervisor: Ce'cile DeWitt-Morette |

mark.mims@gmail.com • +1(512)981-6467 • markmims.com
San Francisco, CA - USA