RICHARD PETTI

richard.petti@gmail.com ⋄ https://www.richardpetti.com ⋄ https://www.linkedin.com/in/richpetti

TECHNICAL SKILLS

Languages	Python, C++, Cypher, SQL, CSH/BASH, R, LaTeX
Data analysis	Pandas, Numpy, Scikit-Learn, CERN-ROOT, Excel
Platforms and Tools	Azure, AWS, ElasticSearch, Nifi, Spark, Neo4j, postgres, Git

WORK EXPERIENCE

Knoema - Head of Data Ingest NY, Lead Data Engineer Feb. 2020 - Current

Team lead for the NY branch of the data ingest team for the tech startup Knoema offering a suite of products including data warehousing and centralization of publicly available data sets along with proprietary data through partnerships with data providers. Promoted to the head of the team after acquisition of Adaptive Management. Main responsibilities include:

- Building and maintaining ETLs from over 40 different data vendors.
- Maintain client engagement and delivery.
- Responsible for quality assurance on ETLs built and deployed.

Adaptive Management - Data Engineer (acquired by Knoema) Jan. 2019 - Feb. 2020

Data Engineer on the ingest team for a tech startup focusing on data integration of alternative data.

- Responsible for building and maintaining ETL pipelines, feeding data from over 40 data vendors, standardizing the format of the data for consumption for the system, using tools such as Nifi, SparkSQL, Zeppelin and Python.
- Worked directly with clients for Data Engineering as a Service contracts, building out custom data feeds to meet the client needs.
- Worked to build out a Python based extraction layer to compliment the existing Nifi extraction layer.
- Developed an entity mapping service to identify common companies and link that information to the internal system identifiers, utilizing a tech stack comprised of Python, Neo4j and Elastic Search.
- Developed and maintained code base supporting the refresh and syncing of information between the main product platform and the knowledge graph supporting the entity mapping service.

CA Technologies, Inc. - Principal Data Engineer Aug. 2017 - Nov. 2018

Data Engineer on the business platform team.

- Work within the engineering team to provide design and implementation of backend API services for business applications in the cloud (Python, Flask, Azure, Neo4J, Redis, Nginx, Git).
- Technical lead for development of multiple products and services for the business, coordinating work with contractors and offshore developers.
- Improved existing application speed by a factor of six, while making the backend more scalable.
- Develop and deliver from within an Agile, cross-functional team (UI developers, product managers, QA team, software developers).
- Perform POC studies for possible technology solutions or product features (such as a clustering algorithm for recommendation of a cross-sale product to an customer based on existing market sales).
- Work with the team to improve development/deployment tools (such as Docker) and procedures, aiming for continuous improvement.
- Debug, troubleshoot and maintain services in production.
- Migrated/upgraded data storage system and related code moving from Community edition Neo4j to Enterprise edition Neo4j, improving overall hardening of the system.

Rodale - Data Scientist Mar. 2017 - Aug. 2017

Data Scientist on Analytics and Growth team at a publishing company.

- Worked with ad ops, editorial, product, and marketing teams to provide analytics tools and support.
- Automated reports providing critical analytics at the fingertips to the non-technical teams with output to daily email attachments, Google Sheets updates, and online dashboards (Python, Django, Postgres, AWS EB).
- Built automated ETL process to push data from many sources into the BigQuery data warehouse.
- Built a dashboard to pull together disparate sources of revenue and metrics from the data warehouse.
- Built an extension to the Google Sheets API to make creating tables and charts more user friendly.
- Topic analysis using Natural Language Processing techniques to provide more detailed context on content within a content channel based on article text (R and Python sklearn).

Brookhaven National Laboratory - Research Associate Aug. 2014 - Mar. 2017

Researcher in experimental science (high-energy nuclear/particle physics).

- Supported the lab mission for the development of a future $1B experimental accelerator facility (eRHIC) through simulation studies.
- Lead researcher for development of four different detector components through simulations.
- Generated pseudo-data utilizing the 50,000 processing core computing farm at Brookhaven Lab and the Condor batch job scheduler.
- Setup an analysis pipeline to transform initially generated pseudo-data for ingestion to different stages of the simulation to final output and visual representation of the results, deployed to the farm.
- Improved the simulation package (C++, CERN-ROOT, https://wiki.bnl.gov/eic/index.php/Eicroot) for public use by adding new features as needed (such as new data interface).
- Provided code documentation (Doxygen, updated nightly) to support users and developers of the simulation package.
- Develop Monte Carlo based software to simulate physics data and develop analysis methods and tools.

Stony Brook University - Research Associate/Research Assistant May 2007 - Aug. 2014
This role includes the period of my dissertation research followed by a short post-doc.

- Analyzed a TB sized data set collected by the PHENIX experiment consisting of 3 billion physics events with up to hundreds of particle tracks in each event (C++, ROOT, Condor, RACF).
- Built an analysis package designed to be deployed onto the computing farm at Brookhaven to process data and create visualizations used to draw inference about the system of study.
- Developed a novel method of particle identification allowing measurement in a new energy range.

EDUCATION

The Data Incubator Jan. 2016 - Feb. 2016
Graduate of Fellowship program
Highly selective and competitive Data Science Bootcamp training academics for industry.

Stony Brook University Sep. 2005 - Dec. 2013
Ph.D in Physics
Thesis: Low Momentum Direct Photons as a Probe of Heavy Ion Collisions.

SUNY Brockport Sep. 2001 - May 2005
B.S. in Physics and Mathematics
Honors & Awards: Summa Cum Laude, Physics Department Scholar.

PROJECTS OUTSIDE OF WORK

Personal Website

- Used modern front end libraries to build a personal branding website (React, CSS, Javascript)
- https://www.richardpetti.com

Lead Management System

- Full stack work, building a web application using the Django framework (Python) for overall project structure, with front end components written using React, Javascript, CSS.
- Designed to be a lead management tool, allowing users to store lead information, view lead information and perform basic analytics on the leads available.
- MVP demo: https://lead-management-system.herokuapp.com
- https://github.com/RichInCode/LMS

CitiBike Helper App: Predicting and Understanding Rider Behavior

- https://citibikehelper.herokuapp.com
- Developed a website with a user interface (HTML, CSS).
- Analyzed 3GB of historical data from the Citi Bike Sharing Program (NYC) (Python, Pandas, Sci-kit lean, Flask).
- Created visualizations (plots and maps) to display trends in rider behavior, such as indicating popular routes and time of ride start.
- Combined bike data with local weather data to train a Random Forest Regressor model to predict the number of riders expected in a zone in a given the day of the week and weather related information through user input parameters.
- Connect to the live data API to help users choose nearby stations with free bikes or parking spots.
- Developed an app to display and predict user trends in the Citi Bike Sharing program data https://citibikehelper.herokuapp.com