

Ben Bubnick

DATA SCIENCE · DATA CURATION · DATA INTEGRATION

Cleveland Heights, Ohio, 44118 | (216) 556-1612

✉ ben.bubnick@gmail.com | 📱 benbubnick

Summary

Database Design	ETL/EDW design with SSIS/SQL/T-SQL/PostgreSQL in CI/CD pipeline
SME/Utility Roles	AI & Machine Learning, Data Curation/Wrangling & Integration
Big Data Platform	Python, Pig, Ruby, kafka, and Hive/Impala on Hadoop & Azure
Health Care Analytics	HEDIS Measures & Tailored Analytics from EMR, HL7, 837, CMS, etc.
Demonstrated Success	Lab Services Award, Eminence and Excellence Award, Honey Badger

Experience

Merative

Jul. 2022 - Present

DATA SCIENTIST

Python, Azure, Postman

Technical lead on the delivery of large & complex AI and Cognitive led transformation programs and works with Merative & Partners to optimize tools and delivery methods.

- Led HIPAA patient deidentification project from 200+ million individual patients for content analytics
- Created testing framework for ETL & implementation processes on new data platform

IBM

Jun. 2015 - Jul. 2022

ADVISORY DATA SCIENTIST

Nov. 2021 - Jul. 2022
Python, GoLang, Azure, Docker, SQL/PostgreSQL, Tableau

Technical lead on delivery of Cognitive & AI led programs using diverse data sources, such as electronic health record data, insurance claims data, scheduling and financial data, social determinants of health data.

- Created machine learning Covid-19 tool that reduced need for hospitalization identification
- Led delivery on tailored analytics and reporting project for client administrative data
- Developed data migration validation analysis on 150+ billion patient records
- Developed surgical smart case card using combined CNN/RNN model

Dec. 2020 - Nov. 2021

Promotion SENIOR DATA CURATION ARCHITECT

Gradle, Python, Hive/Impala, Spark, kafka, Kibana, SQL

Senior member of Data Onboarding & Governance for integration of healthcare data and enrichment of metadata. Configuration of data models, analytics, reports, and visualizations.

- Developed clinical data integration of 100+ billion structured & semi-structured records
- Developed topic modelling project on nearly 6 million semi-structured provider notes data
- Curated 25 pages of AI workflow documentation and mentored new hires in AI tuning process
- Created AI Surgical Scheduling Optimization tool currently evolving into new offering

Nov. 2017 - Dec. 2020

DATA CURATION ARCHITECT

Ruby, Pig, SQL/SSIS/Oracle, Hive/Impala, kafka, Kibana

Lead member of Data Onboarding & Governance, member of the Watson Foundation for Health Solutions Architecture team.

- Managed multi-org enterprise integration with over 70 interconnected data sources
- Developed new data validation and review process that reduced total rework by 25%
- Reduced mapping rework by 45% with through new knowledge-driven initiatives

Sep. 2016 - Nov. 2017

SOFTWARE DEVELOPER *Ruby, Pig, Perl, Jenkins, SQL/mysql/Oracle, DB2, PL/SQL*
 Technical delivery lead for client integrations which has included the ordering, installation, and configuration of required hardware, including VPN devices to support client integrations.

- Reduced time spent on knowledge transfer by ~80% by reorganizing project documentation
- Eliminated need for a \$30,000 annual contract by developing data curation tool
- Reduced cost-per-hire by \$5,210 by re-developing education program

Nov. 2015 - Sep. 2016

DATA SCIENTIST *Ruby, Pig, Jenkins, Hadoop/HDFS/CDH, SQL/mysql/Oracle*
 Data Ingestions, pre-analytic enrichments, analytic augmentation, data governance, and implementation of assigned client integrations.

- Managed team of 12 remote data science contractors for 36 integration projects
- Created over 125 new knowledge base articles, making up 30% of the team's online content
- Developed automated data integration tool to cut out 28% unnecessary development time

Jun. 2015 - Nov. 2015

DATA SCIENTIST CONSULTANT *Ruby, Pig, React, NodeJS, JavaScript, SQL, Oracle, Jenkins*
 Identified clinical, financial, and operational data elements within different health provider systems to develop data transformations and providing meaning to the data.

AmTrust Financial Services Mar. 2014 - Jun. 2015
SOFTWARE ENGINEER *VB.NET, ASP.NET, C#, CSS, Javascript, HTML5, SQL, PL/SQL*
 Created multi-tier MVC components and web forms for commercial insurance policy generating software.

PPG Industries Apr. 2012 - Jun. 2013
COLOR SCIENTIST *VB.NET, C#, SQL, C++, FORTRAN*
 Developed radiative transfer models, spectroscopic profile analysis, and cross-instrument correlation software.

University of Cincinnati Sep. 2006 - Sep. 2010
RESEARCH ASSISTANT *FORTRAN, IRAF, IDL, JMP, LabVIEW, Maple, Mathematica, PDL*
 Gathered data in the short-wavelength, crossdispersed mode SXD on a 0.8–5.4 μm spectrograph at NASA IRTF.

Awards

2018 - 2022	Gold Champion , For 1,477 classroom hours and earning 43 digital credentials	IBM
2018	Lab Services Award , For coordinating internal teams to deliver on time	IBM
2017	Eminence and Excellence Award , For developing a new data curation tool	IBM
2012	Color Lab Automation Award , For expanding data collection and analysis reports	PPG Industries
2010	Mary J. Hanna Fellowship , For STEM students pursuing research-based Masters	UC

Education

2022	Certification , Data Science	IBM
2015	Certification , Data Science	Johns Hopkins University
2010	Master of Science , Astrophysics	University of Cincinnati
2004	Bachelor of Fine Arts , Printmaking	Ohio University