# Ben Bubnick

DATA SCIENCE · DATA CURATION · DATA INTEGRATION

*3419 Superior Park Drive, Cleveland Heights, Ohio, 44118, U.S.*

(+1) 216-556-1612 | ✉ ben.bubnick@gmail.com | ⌨ bubnicbf | in 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 LP**                                                                  *Clinical Operations*
DATA SCIENTIST                                                                  *Jul. 2022 - Present*

Tools: Python, Azure, Postman

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

**IBM**                                                                  *Healthcare Data Analytics*
ADVISORY DATA SCIENTIST                                                                  *Jun. 2015 - Jul. 2022*
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, and social determinants of health data.

### Advisory Data Scientist                                                                  *Nov. 2021 - Jul. 2022*

Tools: Python, GoLang, Azure, Docker, K8s, C#, SQL/PostgreSQL, Tableau

- 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

*Promotion* **Senior Data Curation Architect**                                                                  *Dec. 2020 - Nov. 2021*

Tools: Gradle, Python/R, Hadoop/HDFS, Hive/Impala, Spark, kafka, Docker, K8s, Elasticsearch, 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

### Data Curation Architect                                                                  *Nov. 2017 - Nov. 2020*

Tools: Ruby, Pig, Java, Gradle, Perl, Hadoop/HDFS, SQL/SSIS/Oracle, NoSQL, Hive/Impala, kafka, Elasticsearch, 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

### Software Developer
*Sep. 2016 - Sep. 2017*

Tools: Ruby, Pig, Java, Perl, Jenkins, Hadoop/HDFS/CDH, SQL/mySQL/Oracle, DB2, IMS, 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

### Data Scientist
*Nov. 2015 - Sep. 2016*

Tools: Ruby, Pig, Hadoop/HDFS/CDH, Shell, SQL/mySQL/Oracle, Jenkins, Java

Data Ingestions, pre-analytic enrichments, analytic augmentation, data governance, and implementation of assigned client integrations.
- Mnagaed 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

### Data Scientist Consultant
*Jun. 2015 - Nov. 2015*

Tools: Ruby, Pig, Hadoop/HDFS, React, NodeJS, JavaScript, Shell, 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.
- Developed cron-based widget to define blackout periods for client data pulls
- Performed troubleshooting on production issues across multiple environments
- Principal integration specialist for 10+ client database warehouses

### Software Engineer
*AmTrust Financial Services*

COMMERCIAL PACKAGE POLICIES
*Mar. 2014 - Jun. 2015*

Tools: 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.

### Color Scientist
*PPG*

AUTOMOTIVE REFINISH
*Apr. 2012 - Jun. 2013*

Tools: VB.NET, C#, SQL, C++, FORTRAN

Developed radiative transfer models, spectroscopic profile analysis, and cross-instrument correlation software.

### Research Assistant
*University of Cincinnati*

DEPARTMENT OF PHYSICS
*Sep. 2006 - Sep. 2010*

Tools: FORTRAN, Dataplot, IRAF, IDL, JMP, LabVIEW, Maple, Mathematica, MATLAB, PDL, SAS/GRAPH

Gathered data in the short-wavelength, crossdispersed mode SXD on a 0.8–5.4 μm spectrograph at NASA IRTF.

# Education

### Certification in Data Science
*IBM*

CAPSTONE: SURGICAL SCHEDULING OPTIMIZATION AND PROCEDURE DURATION PREDICTION
*2022*

### Certification in Data Science
*Johns Hopkins University*

CAPSTONE: AI PAINT FORMULA PREDICTION FROM SPECTROSCOPIC MEASUREMENTS
*2015*

### Master of Science in Astrophysics
*University of Cincinnati*

THESIS: MASSIVE STELLAR CLUSTERS IN THE DISK OF THE MILKY WAY GALAXY
*2010*

### Bachelor of Fine Arts in Printmaking
*Ohio University*

THESIS: A NOVEL METHOD FOR LOW-TOXIC PHOTO-ETCH PRINTING
*2004*