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 | 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

Data Scientist, Data Curation Architect

Nov. 2021 - Jul. 2022

Advisory Data Scientist

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, and any other relevant sources.

- 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/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

Nov. 2017 - Dec. 2020

Data Curation Architect

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

Sep. 2016 - Nov. 2017

Software Developer

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

Nov. 2015 - Sep. 2016

Data Scientist

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.

- Managaed 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, 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

AmTrust Financial Services

Mar. 2014 - Jun. 2015

Software Engineer

VB.NET, ASP.NET, C#, CSS, Javascript, HTML5, SQL, PL/SQL

Developed web forms for commercial insurance using ISO standards and created multi-tier MVC components, Web APIs, and UIs.

-
-
-

PPG Industries

Apr. 2012 - Jun. 2013

Color Scientist

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

Developed software and radiative transfer models, spectroscopic profile analysis on color match samples, and cross-instrument correlation software.

-
-
-

University of Cincinnati

Sep. 2006 - Sep. 2010

Research Assistant

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

Gathered data using SpeX, a 0.8–5.4 µm spectrograph, at the NASA Infrared Telescope Facility. Observations in the short-wavelength, crossdispersed mode SXD.

-
-
-

Volunteer

CSforAll

Oct. 2018 - May. 2019

Instructor

Our goal was to provide computer science instruction in all of the District's high schools, with a special emphasis on students who do not have access to a computer or a stable Internet connection at home.

HIT in the CLE

Mar. 2017 - Mar. 2019

Coach

Teams of five students work together to solve a challenge using a large data collection. The purpose is to educate students understand the potential of a career in health information technology and to assist IT companies in Northeast Ohio in developing a talent pipeline.

Code for America

Jan. 2015 - Jun. 2015

ETL Guild

As IT technology evolves, the number and quality of data available to planners and managers changes. The initiative was brought about to apply data science tools to improve and expand public record access.

T.C.P. World Academy

Sep. 2008 - May. 2009

Tutor

Our goal was to teach students how to understand the value of their learning experience and how to engage in a global society. Students at T.C.P. World Academy become academically involved and independent learners by participating in society simulation activities for higher learning.

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*

Talks

AI Hackathon 2020: Surgery Scheduling Optimization

Dec. 2020

A unique algorithm was designed to accurately anticipate surgery duration using a three-stage process that first uses previous utilization data and current waiting list information to manage case mix distribution.

IBM

REACH Initiative: IBM Watson Care Insights

2019

Analysis of longitudinal patient records and real-time data points through HL7 feeds to provide insights into the patient's data, and address challenges by integrating within the physician's workflow in the EHR system.

IBM

Content & Data Analytics Learning Exchange

2018

Introduction of implementation of machine learning algorithms using real-world health data analytics.

IBM

2018 IRI Fall Networks Conference

Sep. 2018

Cognitive computing using NLP to extract insights from the unstructured data that holds most of the relevant information and the necessary transformation at the earliest stages of the data integration pipeline.

INNOVATION RESEARCH INTERCHANGE

Cross Team Training Series

2016

Seven part series in data ingestion approach using the Hadoop software framework. Data can be aggregated much more quickly and cost-effectively using a data lake framework than in traditional data warehouses.

EXPLORYS

Publications

Surgical Scheduling Optimization and Procedure Duration Prediction [a]

Dec. 2020

Improvement on traditional fixed ratio methods for total procedure time estimation, using linear regression models based on relevant variables.

IBM

Process and Requirements for CCD Data Integration

Oct. 2017

Technical white paper on processes and requirements, overcoming the many difficulties with flat file data curation of CCDs.

IBM

AI Paint Formula Prediction from Spectroscopic Measurements [b]

May. 2015

Naive bayesian color and effects matching algorithm for the formula of complex automotive paint formula.

SHINYAPPS.IO

Characterization of Surface Coatings using RT Models for Color Matching

Dec. 2012

Approximation of Chandrasekhar radiative transfer model for atmospheric scattering is applied to color matching.

USPTO

Massive Stellar Clusters in the Disk of the Milky Way Galaxy [c]

Dec. 2010

Using spectroscopic and photometric analysis of two open clusters to determine the structure of the plane of the Milky Way Galaxy.

OHIOLINK ETD

Near-Infrared Spectroscopy of Candidates Members of the Galactic Cluster [BDS2003] 107

Feb. 2008

New near-infrared classification spectra for nine candidate members and comparative 2MASS photometry for several cluster members.

PASP

A Novel Method for Low-Toxic Photo-Etch Printing [d]

May. 2003

Aquatint etching techniques, which have long been employed in intaglio printmaking, may be applied to a novel method for photo-etch procedures.

MAPC