Paul Duncanson
Email: pbduncanson@gmail.com

## SUMMARY OF QUALIFICATIONS

Lead Data/Software/ML Engineer/Architect with extensive knowledge of Application/ Framework/ ETL Development, Machine Learning, GPT, NLP, API's, Object Oriented, Service Oriented and Functional methodologies, Algorithms and Design Patterns.

## TECHNICAL SKILLS

Languages: Scala/Java, C++/C, Go, Python, TypeScript/Javascript, Node.JS/ReactJS/ Angular, Objective-C/Swift

Operating Systems: Linux, Mac OS, iOS, Windows

Streaming Platforms/Databases: Kafka/Kinesis, S3, Redshift, DynamoDB, Cassandra, Flink, Microsoft SQL Server, Oracle, MySQL MongoDB/Mongoose, PostgreSQL, Pinecone

Packages/Frameworks: AKKA, Scikit-Learn, TensorFlow, PyTorch, OpenCV, Pandas, spaCy/ NLTK, NumPy, Apache Spark, Airflow, Spring Boot, SNS/RabbitMQ, Docker, ZIO/Cats, Tesseract

Software: Jupyter/Colab, Sagemaker, MatLab, IntelliJ/Eclipse, Xcode, Visual Studio Code

Hosting Providers: AWS, Azure, GCP

- C/C++ – 10 years
- Java/Scala– 8 years
- Go - 6 years
- Python - 8 years
- Spark/Airflow – 8 years
- AWS/GCP/Azure – 8 years
- APIs / Microservices – 9 years
- Restful Web Services – 10 years
- Javascript/TypeScrypt/ReactJS - 8 years
- CoreNLP/SpaCy/Tesseract, Matlab, Machine Learning/TensorFlow/SciKit/PyTorch – 8 years

## PROESSIONAL EXPERIENCE

Pyramid Consulting — Date: 07/2022 - current

Role: Principal Software Engineer/OCR

- Developed a Knowledge Graph with a California Probate Law Firm to distinguish precedence rulings in response to a query dialog against trained Mixtral8x7B LLM
- Built NLP pipeline utilizing several ML classification techniques
- Conducted meetings with Discover to understand their specific needs, leading to the development of a custom Document Query Application and an OCR Check Reader
- Document Query Application utilizes Python/Scala/ZIO/Spark/Pytorch/Kafka/Flink/ Sagemaker/Glue/Airflow, Pinecone, S3 and Snowflake
- OCR Check Reader employed Gaussian Blur image processing techniques utilizing OpenCV to capture and clean MICR data to achieve 92% accuracy

Microsoft Research Date: 01/2022 – 07/2022
Role: Lead Software Engineer/Architect

- Played a key role in understanding and translating client requirements into the development of a Machine Learning Framework that dramatically increases productivity with an intuitive design that supports the entire ML Pipeline
- Development life cycle Technologies/languages supported also included Python, PyTorch, Scala Spark Batch/Streaming, F-Score Measurement, Tableau, Azure Blob Storage and Data Lake Store
- Refinement was iteratively achieved through frequent interactions with user community that influenced how the framework can target local development to seamlessly transition from local to hosted deployment for shared collaborative development and maintenance by making use of containerization using Anaconda, Docker and Kubernetes
- An NLP trained model was also included through on-going feedback from consumers as a proof of concept that provides text classification and summarization

Core Logic Date: 08/2021 – 01/2022
Role: Lead Architect

- Led the effort in refining and implementing customer requirements into the development of a GCP Vertex AI architecture that includes a migration plan
- Trained Core Logic employees on how to utilize the GCP Vertex AI architecture by conducting recorded video sessions to support the development effort of 6 core forecasting applications

Conifer Health Date: 01/2019 – 08/2021
Role: Sr. Software Engineer/Machine Learning/NLP

- Continuously refined accuracy and completeness of a Knowledge Graph with SME's
- Created NER models to represent diseases, symptoms, medications and procedures Developed NLP pipeline that ingests diagnostic data from multiple sources by utilizing Spark, and Kafka
- Upgraded ingestion platform to distinguish a given patient across multiple providers Utilized TensorFlow/Keras/PyTorch trained models to classify treatments with specific patient diagnosis to alleviate uncertainty with clinical decisions.

Synchrony Date: 01/2018 – 12/2018
Role: Sr. Software Engineer/Machine Learning

- Developed and trained Classification/Regression models utilizing multiple credit card features to distinguish fraud behavior
- Built ETL pipeline that automates ingestion with optimized performance across Spark distributed platform
- Analytic pattern recognition characteristics can be applied through a parameterized pivoting technique
- Technologies utilized: UML/Scala/Spark/Streaming/Parquet/Hadoop/MLib/GraphX/PyTorch

Shell Corporation Date: 01/2017 - 12/2017
Role: Lead Software Developer/Architect

- Actively reported project progress and sought approval from Shell's Board of Directors, emphasizing customer-centric development approaches
- Led the gathering of specifications and business goals from clients, ensuring the architectural solution aligns with their needs
- Designed and developed the AWS IoT analytics engine and data collection framework that captures GPS data from a customized Linux-based hardware device installed on each chassis going in and out of the Port of Long Beach initially and then all ports throughout the U.S.
- Utilized Athena for query across multiple data sources for real time billing and trend analysis
- Gathered specifications, business goals and feedback of first iteration effort from developed hardware to and from the overall architectural solution
- Developed AWS Firehose, AWS Lambda, C++/C code application on IoT device, Javascript/ Angular developed to support tenant login dashboard with Docker
- Developed Spark Streaming platform to generate routing of trailers for highest logistical availability with lowest mileage expense

Ephesoft Corporation Date: 01/2015 – 01/2017
Role: Principle Engineer

- Hands-on technical lead involved with training a forming team while developing features utilizing Maven, IntelliJ, ReactJS/Node.JS/Scala, Kafka, Flink, Akka, Docker, AngularJS and the Selenium Automation Testing Framework
- Developed document router tool that makes use of Tesseract to perform Optical Character Recognition (OCR) on batches of document images that are routed through a visually designed data-driven work flow utilizing Airflow DAG components
- Applied TF-IDF vectorization method and K-Means algorithm to determine document grouping
- Applied Naive Bayes algorithm to classify documents in relation with a given batch and in relation to all collected documents
- Work flow of captured email attachments and files from a designated subdirectory are fed into Flink/Kafka Topics to optimize throughput with parallelized consumers
- Built a set of Micro Services hosted on a. Cluster of Azure virtual machines to provide OCR document processing through Tesseract and Nuance
- Employed a B-Cubed algorithm that scores CoreNLP's Coreference output to optimize noun phrase correlations
- Developed Machine Learning algorithms utilizing TensorFlow/PyTorch that categorize documents with a multi-classifier that distinguishes document types (i.e. invoices, receipts, offer letter, credit report, etc.)

Invigorate Software Date: 01/2014 - 01/2015
Role: Principle/Lead Developer

- Developed a cloud-based image and story sharing platform, PicSavour, utilizing AWS/ Apache Spark Streaming/Flink/Kafka/Cassandra

- Developed load balancing driver on EC2 to provide dynamically scaled session support Designed and Developed Scala Machine Learning algorithms on Apache Spark platform to capture relationships between shared images
- Utilized Amazon Recognition library and Open CV for facial recognition. Custom code developed utilizing TensorFlow convolutions on image data to automatically tag detected faces with sufficiently similar facial features
- Microservices written in Javascript that distributes work across a set of AWS GPU's to provide real-time tagging feedback on the mobile device as photos are taken within the mobile app.
- Utilized Kafka cluster that captures image meta data records into log entries within partitions and topics associated with login credentials and geographical location Designed a denormalized set of tables on DynamoDB for fast geographical, logical and/or chronological query

Emotiva Corporation/Momentum Data Systems. Date: 01/2013 - 01/2014
Role: Lead Engineer

- Developed two high-end home theater sound systems, Emotiva XMC-1, Fusion C++/ Objective C/OOP/Python code developed on corresponding platforms Linux/OS X/ Window's
- AWS IoT utilized to transfer room correction coefficient data between Dirac and Emotiva devices across MQTT
- Captured video/audio selections with Apache Kafka cluster for query with Cassandra Sink providing video selection trends and best alternative choices
- Developed a cloud-based recommender engine that populates ad banner space on mobile device
- Led architecture redesign and development of device driver services to provide a reusable set of object-oriented components utilizing latest C++ standards (I.e. C++17) to reduce spin-up time for new board design from one year to 3 months

Intel Corporation Date: 01/2012 - 01/2013
Role: Lead Developer

- Utilized TermIO/FastBoot/C++/Objective-C to provide the firmware download application for all Intel IoT devices
- Developed Intel's Smart Watch iPhone-based Software Development Kit for Intel in Swift
- Developed Bluetooth LE communications stack between iPhone and SmartWatch
- Developed a cloud-based recommender engine that populates ad banner space on mobile device
- Utilized Jenkins/Jira/Git for continuous integration, task management and version control

EDUCATION

B. S., Information Systems Management, University of San Francisco, 3.8 Honor Roll

Continuing Education: Stanford University (Machine Learning, NLP, Linear Algebra) - 2022

Currently enrolled in M.S., Data Science, University of California, Berkeley - 2025