

ABHISHEK JANGALWA

Senior Data Scientist at Oracle, with 7+ years of experience, specializing in scalable big data applications | San Francisco, California
+1 213-210-1939 | jangalwa@usc.edu | <http://linkedin.com/in/jangalwa> | <http://jangalwa.me> | <https://github.com/jangalwa>

TECHNICAL SKILLS

Programming Languages: Scala, Python, Java, PHP, SQL, JavaScript, C, C++, HTML, CSS, Shell

Databases: Postgres, Oracle, MongoDB, MySQL, Firebase, DynamoDB, Hive, Elasticsearch, Jenna (Triplestore)

Libraries & Frameworks: Spark, Hadoop, jQuery, Ajax, Kafka, Scikit Learn, Flask, Matplotlib, Tensorflow, Keras, RDF, Luigi, Langchain

Systems & Tools: Kubernetes, AWS, OCI, Unix, Git, Docker, MapReduce, Six Sigma, IntelliJ, Grafana, Federated Machine Learning

PROFESSIONAL EXPERIENCE

Senior Software Developer AI/ML, Oracle Health, Denver, CO

July 2024 – Present

- Solving clinical data challenges like summarizing doctor's notes, entity extraction & linking using **Large Language Models**.
- Creating a **Semantic Knowledge Graph** to enhance and evaluate accuracy of LLM responses for condition-drug mapping.
- Developing **Python APIs** to fetch and serve LLM responses in real time over Oracle Cloud and Microsoft Azure.

Senior Data Scientist, Identity Graph Core Team at Oracle Advertising, Denver, CO

July 2021 – June 2024

- Designed and developed **entity resolution** systems, creating anonymized profiles for targeted advertising using terabyte-scale data and cloud technologies, coordinating with multiple stakeholders, internal teams and external clients.
- Created a pipeline for Connected TV devices, processing 5B+ records in minutes with Scala & Spark on OCI Dataflow and AWS, **boosting graph reach by 20%** while preserving **Differential Privacy with Federated Learning**.
- **Reduced operating cost by 30%** by implementing testing frameworks for **Python and Scala** libraries.
- Improved legacy reporting and ingest systems, enhancing performance and **reducing failures by 70%**.

Data Scientist, Audience Modeling Team at Oracle Advertising, Denver, CO

Aug 2019 – July 2021

- Built **scalable applications** for targeted ads using machine learning and graph processing on **AWS and OCI**.
- **Productionized a research prototype generating \$20M** annually by optimizing Python & **Spark** scripts.
- Devised **Orchestration** (Workflow management) and metadata storage projects with AKKA, Elasticsearch, and Postgres.

Research Engineer Intern at Evidation Health Inc., Santa Barbara, CA

Jan 2019 - May 2019

- **Developed a data lake and ETL pipeline** for storing and serving US pollen data using Python and AWS.
- Programmed continuous **scrapping** of various open-source data sources to fetch relevant data.
- Created a **real-time Flask-based REST API** for data access, ensuring efficient querying.
- Conducted research to optimize data storage solutions for the Pollen-data-lake project.

Machine Learning Pipeline Developer at USC Information Sciences Institute, Los Angeles, CA

Nov 2017 - Jan 2019

- **Developed ML pipelines for health sensor data**, enabling quick feature extraction and model creation.
- **Improved data processing speed by 8000%** using PySpark, handling up to 20 TB of time-series data.
- Co-authored a paper on human movement pattern analysis.

Full Stack Developer, Founder Director at Paritosh Software, Mandsaur, India

Jan 2010 - Apr 2017

- Founded and grew the company, focusing on product development and client outreach.
- Developed responsive websites and **LAMP**-based software for **50+ clients**, including government agencies.
- Enhanced public policy awareness and market rate information dissemination in rural areas.
- Trained 100+ undergraduate engineering students in software and web development.

EDUCATION

University of Southern California, Los Angeles, CA

Aug 2017 - May 2019

Master of Science, Data Informatics

Courses: Machine learning, Data Mining, Data Management, Knowledge Graph, Information Retrieval

Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India

Aug 2006 - June 2010

Bachelor of Engineering, Computer Science

Courses: Analysis & design of algorithms, Data structures, Object oriented programming, Database, Artificial Intelligence

PUBLICATIONS

- Toward Robust Interpretable Human Movement Pattern Analysis in a Workplace Setting [IEEE Xplore]