Sungwook Yoon

Data Developer / Engineer

Contact

🌐 sungwookyoon

✉ sungwook.yoon@gmail.com

📞 (480) 277-3458

🔗 /in/sungwookyoon

📍 Philadelphia, PA

Technical Skills

- Python
- Scala
- Bash
- SQL
- C

Frameworks & Tools

- AWS
- GCP
- BigQuery
- Databricks
- Spark
- Hadoop
- Claude Code
- EMR
- Lambda
- Cloud Run
- DBT
- Github Actions
- Terraform
- Elastic Search
- Kubernetes
- Kafka

Strengths

- System Design
- Technical Leadership
- Agile/Scrum
- Code Review
- Technical Writing
- Mentorship

Education

B.E. Electrical Engineering
Seoul National University

Ph.D. Computer Engineering
Purdue University
Thesis on Artifical Intelligence
Applying ML to Decision Making
Won International Competitions
Won Runner Up JAIR paper award
Won Influenctial Paper Award

Professional Summary

I have journeyed through modern revolution in data space. Started my professional career in backend of Ruby on Rails Postgres. I then moved on to Hadoop, then AWS, then Databricks and now I am working with Bigquery and GCP. Through the Journey, I learned a lot on modern data architecture and scaling factors. I can describe and build modern pipelines on most of Cloud platforms. I am a self learner. I have learned AWS, Databricks or GCP without training. I learned through self research, trial and errors. I have built / coded up many data pipelines from raw ingestion via API or Kafka stream to BI dashboards. Having used and learned AWS extensively for long time, transitioning to GCP has been pretty straighforward, since these Cloud services offer similar products with slightly different nuances. Now I am leveraging Claude Code to serve customized Dashboards for end data customers to see the status and quality. Well, even this Resume is being built using Claude (Not content side but format and etc).

In terms of business domain knowledge, I have worked on Adevertisement industry for long tiem. My exposure to AdTech was with Comcast. I have built data solutions that analyzes viewing behaviors in traditional and OTT videos. At current company, I am building my domain knowledge on Influencer Marketing. It is an amazing experience how influencer marketing is currently done and the rapid change happening in AdTech world.

With avalanche of new LLM based technologies pouring in, I expect particular domain experience is becoming less important and what remains as competency is the mastery of overall knowledge on the system you are directing the LLM tools. I frequently compare current LLM revolution with airplanes. Pilots may not control everything on plane and leave it to automation. But, Pilots know every control to control the automation. Same for LLM tools. One must know all the details of what LLM is doing to efficiently drive it. I am exicded to see new changes ahd enjoy riding the changes.

Work Experience

📅 Jul 2025 – Present 📍 Remote
Data Engineer
Later
- Developed Data Ingestion layer. Get the data from various Affiliates APIs then ingested into GCS. Leveraged Cloud Run.
- Developed Data Ingestion from AWS S3 to Bigquery, leveraging Cloud Run and Bigquery.
- Developed Data Infrastructure. Provided guidelines and principles on BigQuery and GCP resources.
- Developed Custom WebApp and Dashboards for Internal Customers using React/Typescript/NodeJs, leveraging Claude Code
- Developed Custom Data status / Quality framework.
- Provided guidelines on how to label and analyze the costs on GCP platforms.
- Used Linear, Notion, MCP of them to automate Claude generated Notion pages and Linear tasks.

📅 Feb 2018 – Apr 2025 📍 Remote
Data Engineer
Comcast
- Implemented data ingestion and ETL operation on various platforms (AWS, Hadoop and Databricks)
- Built fault tolerant ETL operations on AWS, mostly via Spark scala or Pyspark jobs on EMR.
- Built data monitoring and data quality operations on various platforms (AWS, Hadoop and Databricks)
- Migrated data operations from OnPrem to Hadoop, Hadoop to AWS, AWS to Databricks
- Scaled data operations to 100Billion+ size
- Data consumption from Kafka via Spark Streaming
- Data consumption from Kinesis via Firehose on AWS
- Operationalize Machine Learning codes into AWS via Step, EMR and Lambda
- Massive parallelization through Lambda on AWS and Ray on Databricks
- API implementation on APIGateway and Appsync
- Event driven data processing through SNS/SQS/Lambda on AWS
- Enable events driven data processing between AWS and Databricks
- Developed KPI for Data team
- Enabled Data Scientists work into Production, Helping them to scale and operation.

📅 Mar 2014 – Feb 2018 📍 San Jose, CA
Professional Services
MapR
- Close to 100% Professional Services Deliveries
- Engaged on customer specific projects to help the customer's goals in data operation
- Leveraged various Hadoop echo systems, Hive, Spark, Oozie, HBase and many other Hadoop components
- Presented Hadoop Experiences in various Meetups