

Sungwook Yoon

<http://sungwookyoon.com>
sungwook.yoon@gmail.com | 480.277.3458

EDUCATION

PURDUE UNIVERSITY

PH.D IN COMPUTER ENGINEERING
 Machine Learning | Planning
 Classic Artificial Intelligence

SEOUL NATIONAL UNIVERSITY

BSEE IN ELECTRICAL ENGINEERING
 Control
 Instrumentation

LINKS

LinkedIn:// [sungwookyoon](#)

PROJECTS

DATA MIGRATION

On Prem SQL to Hadoop
 On Prem Hadoop to AWS
 AWS to Databricks

ETL

Data Ingestion and Operationalization on Hadoop
 Data Ingestion and Operationalization on AWS
 Data Ingestion and Operationalization on Databricks
 Implemented Alert and Data Quality Checks via Cloudwatch and Montecarlo

SKILLS

PROGRAMMING

Language
 Scala • SQL • Python
 Shell Script • Java • C
 Spark
 Pyspark • Spark-streaming • Kafka
 AWS
 Lambda • EMR • EC2 • ECR • Redshift
 Step • Cloudwatch • Appsync
 Platform
 Terraform • Boto3
 Databricks
 Unity Catalog • Workflow • Ray

EXPERIENCE

COMCAST | PRINCIPAL ENGINEER

2016 – Current | Philadelphia, PA

- Deep experience in Advertisement data
- Along with Data Operation, I provide RCA on Data issues
- 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
- Migrated data operations from Hadoop to AWS
- Migrated data operations from AWS to Databricks
- Scaled data operations to 100Billion+ size
- Data consumption from Kafka via Spark Streaming
- Data consumption from Kinesis via Firehose on AWS
- Worked with Kafka Schema Registry
- Built fault tolerant scalable data solutions to handle data faults and system faults.
- 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
- Lambda functions via Python and Container

MAPR | PRINCIPAL DATA SCIENTIST

2014 – 2018 | San Jose, CA

- Close to 100% Professional Services Deliveries
- Delivered Hadoop System Installation
- Engaged on customer specific projects to help the customer's goals in data operations
- Leveraged various Hadoop ecosystem systems, Hive, Spark, Oozie, HBase and many other Hadoop components
- Presented Hadoop Experiences in various Meetups

RESEARCH

PALO ALTO RESEARCH CENTER | RESEARCH STAFF

2008 – 2012 | Palo Alto, CA

- Published Research Papers in various conferences and Journals in Artificial Intelligence
- Worked on various DARPA projects for research deliveries

AWARDS

2021	Influential Paper Award	ICAPS
2011	Best Paper Runnerup	Journal of Artificial Intelligence
2009	Best Machine Learning	International Learning and Planning Competition
2004	Winner	International Planning Competition