Chengzhi Zhao

New York, New York

Sr. Data Platform Engineer

Summary

Professional data platform/infrastructure engineering experience in programming, distributed system, batch & real-time data pipeline, and data visualization. Successful history of effectively developing large scale data pipelines and data warehousing solutions. Experience in developing data-driven applications using open source libraries. Committer of Apache Airflow.

- Scala, Java, Python
- Hadoop, Hive, Spark core, Spark Streaming, Flink, Storm, Sqoop
- Docker, Kubernetes
- Airflow, SSIS, Pentaho, Talend
- AWS (S3, EMR, ECS, Kinesis, Lambda, DynamoDB, Aurora) and GCP (GCS, Pubsub, Dataproc, BigQuery)
- SSRS, SSAS, Kibana, Tableau, BIRT
- SQL, Redshift, MySql, SQL Server, Teradata, DB2, ElasticScearch
- Redis, DynamoDB, MongoDB
- HTML, CSS, JavaScript, JQuery, d3.js, XML, JSON
- Github, Bitbucket, Jira, Team Foundation Server
- Travis
- Shell scripting

Experience

Sr. Data Platform Engineer at Meetup
February 2018 - Present (1 year 8 months)

- Reduced data pipeline latency from 24 hours to minutes to improve the data analysis and machine learning modeling quality with near real-time dataset using MySQL(binlog), Kinesis and Flink.
- Designed and implemented ETL pipeline with up to 500 distributed concurrent tasks using Apache Airflow.
- Reduced ETL operation cost by 30% on ETL pipeline by processing delta dataset only in Spark (Scala).
- Improve data pipeline stability and scalability by utilizing Kubernetes and auto scaling.
- Built CI/CD pipeline for private python packages on Nexus 3 server.

Data Engineer at Meetup

April 2017 - February 2018 (11 months)

Developer II at Walmart
December 2014 - April 2017 (2 years 5 months)
- Created revenue reconciliation matching rules engine to improve wireless payment to receivable match rate to 98% and reduced unapplied cash by 5% monthly.
- Developed pharmacy revenue reconciliation engine.
- Built data warehouse in dimensional modeling with star/snowflake schema and built ETL process using SSIS, Teradata, Hive.
- Built near real-time ETL and dashboard to process sales metrics data.
- Led agile team to develop monthly wireless revenue reconciliation and automate journal entry process for accounting department.

Developer at Meridian Behavioral Healthcare, Inc.
June 2013 - December 2014 (1 year 7 months)
- Developed an incident management system for incident reports that track records.
- Developed a data-driven Survey Scheduler to automate the process of generating and delivering surveys for each client using Java, BIRD reporting engine and QUARTZ
- Built an ETL process to extraction 5,000 records/day, validated and loaded data to data warehouse using SSIS, SSAS

Software Engineer Internship at Infinite Energy Inc
January 2013 - May 2013 (5 months)

Programmer at University of Florida
May 2012 - May 2013 (1 year 1 month)

Software Engineer Internship at HARN Museum of Art
January 2012 - April 2012 (4 months)

Marketing & Management Internship at PING AN INSURANCE (GROUP) COMPANY OF CHINA ,LTD
June 2009 - August 2009 (3 months)
- Assisted in organizing products, customer, and salesperson conferences.
- Arranged social activities and led group discussions, delivered core value to the customer and gathered feedback.
- Interviewed "Best Sales Person of the Week" and generated motivating reports that shared experiences with the local company, with each report reviewed more than 300 times and received good comments.

Education
University of Florida - Warrington College Of Business

Master of Science, Information System & Operation Management, 2011 - 2013

Activities and Societies: Association of Information System

Wuhan University of Technology

Bachelor of Management, Marketing, 2007 - 2011

Activities and Societies: Student Union, Association of Marketing, Association of Future Management

Chengzhi Zhao

New York, New York

Sr. Data Platform Engineer