

# Pat Thomas

## Big Data Engineer

Big Data Engineer with six years of experience in Hadoop, Spark, and Kafka to process petabyte-scale datasets. Delivers production-grade pipelines for analytics and ML.

### CONTACT



youremail@example.com



LinkedIn | Portfolio



Houston, TX

### KEY SKILLS

- Hadoop Ecosystem (Spark, Hive, HDFS)
- Apache Kafka & Airflow
- Cluster tuning & administration
- Ansible automation

### EDUCATION

M.S. in Data Engineering  
University of Houston, TX, May 2016

B.S. in Computer Engineering  
Rice University, Houston, TX, May 2014

### PROFESSIONAL EXPERIENCE

**BIG DATA ENGINEER | DATAMINDS CORP., HOUSTON, TX**  
MAY 2018 – PRESENT

- Designed Spark jobs processing 10 TB of log data daily, improving pipeline throughput by 70%
- Built Kafka clusters for real-time event streaming, reducing data latency to <1 second
- Tuned HDFS and YARN configurations, increasing cluster utilization by 35%
- Created Airflow workflows to coordinate Spark and Kafka jobs

**DATA ENGINEER | TECHSCALE SOLUTIONS, HOUSTON, TX**  
JANUARY 2016 – APRIL 2018

- Developed Hive tables and optimized MapReduce jobs, cutting ETL time by 40%
- Implemented Parquet file formats to reduce storage costs by 50%
- Automated cluster provisioning with Ansible, reducing setup time from days to hours
- Provided 24/7 on-call support for critical data pipelines

### CERTIFICATIONS

- Cloudera Certified Professional: Data Engineer, March 2019
- Senior Big Data Engineer, Data Science Council of America, May 2019