Akshay Jain's resume

email-me | **LinkedIn profile** | **Github profile**

Work experience of all my professional and personal projects

AdColony (acquired by Digital Turbine)

Senior Software Engineer (Apr 2018 - Present)

Bothell, WA

- Tech lead and mentor for the Bigdata team
- Migrated Databricks jobs to Spark over Kubernetes. The jobs aggregate data from S3/GCS using Spark and use Airflow to ingest into Druid.
- Built a generic data pipeline using Apache Beam, GCP Cloud Dataflow for migrating AWS DynamoDB data into GCP BigTable.
- Migrated Databricks jobs to Spark over Kubernetes. The jobs aggregate data from S3/GCS using Spark and use Airflow to ingest into Druid.
- Architected and implemented AWS to GCP infrastructure business unit migration using Terraform
- Created generic AWS Data pipeline using AWS EMR to clean up non-TTL records from AWS DyanmoDB – leading to a `$1M annualized savings` in storage costs.
- Led personal information obfuscation project for entire business org to meet GDPR legal requirements.
- Built automation scripts to aggregate TBs event data and compact them on Redshift/Google BigQuery in hourly/daily/monthly time windows.
- Built monitoring & alerting tools for all the Kafka based streaming applications using Lag exporter, Prometheus, Grafana hosted on Kubernetes cluster as Helm releases.
- Memory & application instrumentation that led to 20% month-over-month cost savings for the BigData team.
- Designed & implemented Monix task based HTTP service (with rate-limiting) to populate data onto GCS and GBQ.
- Built event triggers using Google cloud function to trigger app processing whenever a Google BigQuery job dumped data onto Google cloud storage. The trigger was an HTTP ping to an existing app to notify the data was ready for processing.
- Created scalable scripts to process tons of data (in TBs) from Google cloud and pump it the data processing pipeline for replaying HTTP requests.
- Architected designs for end-to-end request tracing in micro-services based big-data pipeline. This helped analyze each request that enters in the big-data system, through the various interactions/transformations it goes through, before ending up in the destination data sources.
- Created load-testing frameworks for HTTP services using Gatling that helped analyze and improve the performance by reducing the response time by 50%.
- Analyzed big-data components for cost-savings effort. Used Prometheus + Grafana and AWS's internal tools like Cloudwatch, EC2 to monitor the current usage of services. Recommended & implemented optimum resource usage (EC2 types and counts) that helped reduce the monthly cost by 20%.

Software Engineer (Mar 2017 - Mar 2018)

Bellevue, WA

- Created a highly scalable RESTful web-service using Scala + Finch to support adding new categories of devices (segments) for AdColony's web portal for ad-targeting.
- Created a real-time streaming application to consume events from AWS Kinesis streams triggered by Lambda to populate Google BigQuery for data analysis.
- Built a highly scalable & highly available application that syncs events from AWS S3 to Google Cloud storage and loads them to BigQuery – with automatic retries, alerting (using Slack) and monitoring (Prometheus + Grafana).
- Built many customizable Bash scripts to help automate tasks for the team. For example:
  - Loading data from Google Cloud storage into BigQuery by the data's correct partition date.
  - Sanitizing really huge compressed data files (in GBs) for data processing.
- Designed a multi-threaded highly scalable application to read and process 15k+ requests per second from Google BigQuery and enrich the data on AWS DynamoDB.
- Instrumental in setting up a testing environment to address end-to-end request tracking and data availability.

Bank of America

Software Engineer, Vice president (Jan 2017 - Mar 2017)

Seattle, WA

- Built an ETL process using Cassandra, Spark for data pre-processing for machine learning models.
- Designed & built a highly scalable multi-threaded cross-datacenter distributed system using Akka framework.
  - The cluster application did message transformation and replication from & to Kafka.
  - Also implemented the JMX-plugin of akka-cluster to monitor the application.

Software engineer, Asst. Vice-President (Feb 2015 - Jan 2017)

Seattle, WA

- Data migration - Developed a multi-threaded, highly configurable J2EE application to migrate data from Mainframe to Oracle with the following features:
  - Check pointing - ability to start/stop/resume processing input file at any time.
  - Locking mechanism - so that only 1 instance of the application runs at any time.
  - Dynamic loading of properties - number of threads, timeouts, number of records to process etc.
  - Shell script to trigger the conversion process.
- Big data analytics and pre-processing framework setup
  - Used Spark, Cassandra, SOLR and Kafka to create a big data analytics project. (Details below)
- Built many automation scripts using Bash and Expect:
  - Passwordless SSH into new VMs

Intuit

Software developer co-op (Jul 2014 - Dec 2014)

San Diego, CA

- Enhanced the TurboTax Mac app:
  - Added new features for the TY14 tax season.
  - Enabled PDF attachments for NY users - compliance feature.
  - Added new XML screens for ACA (affordable care act) post filing experience.
- Replaced server based tax calculations, with embedded Javascript, for TaxCaster app on iOS (using Javascriptcore) and Android (using Rhino and Gson), resulting in increase in speed by 50%.
  - Improved the efficiency by totally eliminating the network latency
  - Also allowed the app to be used in offline mode (as opposed to being network based).

Rochester Institute of Technology

Graduate Research Assistant (Aug 2013 - May 2014)

Rochester, NY

- Python based web-server developer and manager for co-robotics project.
  - Developed a web based application to communicate with all the corobots in the system.
  - The PHP and JQuery based website, backed with a multi-threaded Python server process and MySQL database, was hosted on an Ubuntu VM, running an Apache 2 web server (LAMP stack).
  - The web application has the following features:
    - Remote monitoring of corobots (using AJAX and JSON). - Current X-Y co-ordinates, status, destination of the corobots.
  - Remote deployment to corobots. - Upload code to be executed on corobots. - Choose a destination to navigate an IDLE corobot to. - Currently, we provide Python and Java API. Hence, the application supports, extension based remote deployment of code.
  - Individual workspaces for students. - Upload files, view/deploy uploaded files, download user specific logs.
  - Integrated with RIT password service over LDAPS and HTTPS.

Motorola Solutions

Summer Intern (Jun 2013 - Aug 2013)

Holtsville, NY

- Developed a web application using Python, for automatic daily collection of data.
  - Support for MySQL and ORACLE databases with the web application, to generate user interested data.
  - Output spreadsheet and graphs generated were used for tracking the progress and analysis of products.
  - Basic configuration and initial setup of ISS web server for hosting the web application.
  - Application enhanced the process of metric collection for the products under test cycle.

Infosys

Systems Engineer (Jul 2010 - Jun 2012)

Pune, India

- Handled the Mainframe based applications, their routine scheduling and their feed to other applications in the pipeline.
- Responsible for analysis and quick resolution of incidents, pertaining to the application.
- Actively involved in bug fixing and addition of new features.
- Wrote unit and system test routines as part of the enhancement of the project.

Polygon

Engineering Intern (May 2009 - May 2010)

Mumbai, India

- Developed a prototype for an RF based remote control for HVAC systems.
- Added basic functionality of the HVAC remote system, along with an option, to allow control for multiple HVACs, by changing the program feature.
- Also added functionality for automatic scheduling and timing.
- The prototype was designed using Proteus 7.1. The coding was done in Keil-C.

Education

Rochester institute of technology

Masters of Science (M.S.) Computer science (2012-2014)

CGPA = 3.51

SVKM’s Narsee Monjee Institute of Management Studies (NMIMS)

B.Tech (Electronics) (2006-2010)

CGPA = 3.43

Nath valley school

SSC & HSC (2000-2006)

Licenses & certifications

CKA: Certified Kubernetes Administrator

The Linux Foundation Foundation

Issued Jun 2022 · Expires Jun 2025

Credential ID LF-tcmewjc3tu

DataStax Apache Cassandra™ Professional Certification

DataStax

Issued Sep 2016

Technologies & skills:

`Airflow • Apache Druid • Apache Spark • Databricks • Docker • Google BigQuery • Helm • Kafka • Terraform`

Coding languages: `Bash • Java • Python • Scala`

Databases: `MySQL • Postgres`

Platforms: `AWS • GCP • Kubernetes`