

# ABHINAV PUNDIR

San Jose, CA 95134 | (669) 224-0552 | pundirabhinav@yahoo.com

## Work History

---

### Principal Software Developer, Mail Intelligence

01/2015 to Current

Yahoo! Inc. – Sunnyvale, CA

#### ML Email Extraction System

- Spearheaded built of end-to-end scalable email extraction system on long tail domains using Neural Networks by working cross functionally with data scientist, product and editors to generate ~\$7M in revenue QoQ.
- Built training pipeline that scale to millions of emails, wrote ML extractions libraries and lead engineers to design sampling, evaluation and model playground systems.
- Tech Stack: PySpark, TensorFlow on Spark, NER, Java.

#### ML Real Time Inference Stack

- Designed and implemented ML based real time micro services powering 10M auto suggestions and 20M smart views daily in Yahoo mail app achieved through containerization, incorporating scalability and security, serializing models and simplifying deployment process for On-Premises K8s.
- Published work at Yahoo's annual conference that lead stack to be de facto standards for Real Time ML Inference and is currently leveraged by various teams at Yahoo.
- Tech Stack : Kubernetes, TF Serving/Saved Model, Docker, Envoy Proxy, mTLS, Splunk, Grafana.

#### ML Batch Processing Inference Stack

- Developed language/runtime independent scalable system to run ML models inside Hadoop containers powering all ML batch systems for Mail Intelligence including pipelines that extract data from 20 million emails daily.
- Collaborated with Hadoop, Security and BigML teams to prototype, resolve security issues and perf benchmarking that resulted in generic (Linear, DNN and Sequence models), scalable (tens of thousands of yarn containers in parallel) and efficient (35% reduction in inference latency) stack.
- Tech Stack : Hadoop, Pig, Spark, GRPC, Unix Domain Socket, TF Serving, Protobuf.

#### Email Clustering

- Lead various clustering initiatives based on structural similarity that is used by Mail systems to identify email for sampling, extraction and logging.
- Designed clustering that scale to 4B emails (40TB) daily, invented graph based clustering algorithms for better performance (5X less clusters) and built micro services that handles read/write requests for millions of clusters daily.
- Tech Stack: Java, Hive, HBase, Html xpath, Hashing, Graph algorithms, Sharding, Jetty, Cache.

#### Email Extraction

- Designed and developed scalable batch processing email extraction pipeline running hourly and retroactively; processing 4B emails daily, extracting data from 800M emails and responsible for generating \$100M in revenue YOY.
- Tech Stack: Oozie, Pig udf, Pig distributed cache, Druid.

### Senior Software Engineer

07/2013 to 01/2015

Yahoo Software Development Centre – Bangalore, India

- Content Asset Tracking: Perf benchmarking of HBase to find right persistent storage tracking millions of events/day.
- URL Metadata Service: Developed spouts and bolts of Storm based service that extracts metadata of URL scaling up to 1K QPS with 100 ms average latency.

## Education

---

**M Tech:** Computer Science

07/2013

**IIIT** - Bangalore

- Gold medalist (3.91/4)