



Data Pipelines Can Be Simple

SUMMARY

We partnered with our agricultural seed customer to create sales rep & grower facing products including a mobile application that wirelessly syncs planting, harvest and application data from the field back to a central location. The harvest yield data is used to measure the efficacy of both grower operations and seed varieties planted worldwide. After we created a mobile product to wirelessly collect harvest data below budget and ahead of schedule, we began to have conversations with our customer about their pain points with processing the potentially petabytes of harvest data in a reliable and efficient manner.

CHALLENGE

The established pipeline from harvest data ingest to end-user data visualization had taken years to assemble, involved highly complex transforms & hand-offs and required multiple large teams to maintain. With a week of budget remaining after finishing the mobile app, they green-lighted an attempt by us to produce a worldwide heat map from an entire season of raw harvest data.

SOLUTION

A data pipeline was quickly stood up. Large binary files were landed in S3, events fired SNS->SQS, ECS cluster workers tore through the data, writing it to Elasticsearch. In a day, a yield heat map of an entire year's harvest was visible in a Kibana world overlay map. Our skeptical customer zoomed in on specific fields, and declared: "This is spot on!"

BENEFITS

Solving Big Problems With Simple Architectures

In 1 week, 2 devs built a pipeline that processed an entire year of nationwide harvest data in 1 day.

Easily Build Scalable Data Pipelines

An auto-scaling ECS cluster let us tailor our queue workers and match ingest load.

Modern AWS Product Selections Let You Tune It Further

Use of Kinesis, Kinesis Firehose and Amazon Elasticsearch Service can make it faster.