ASTRONOMER

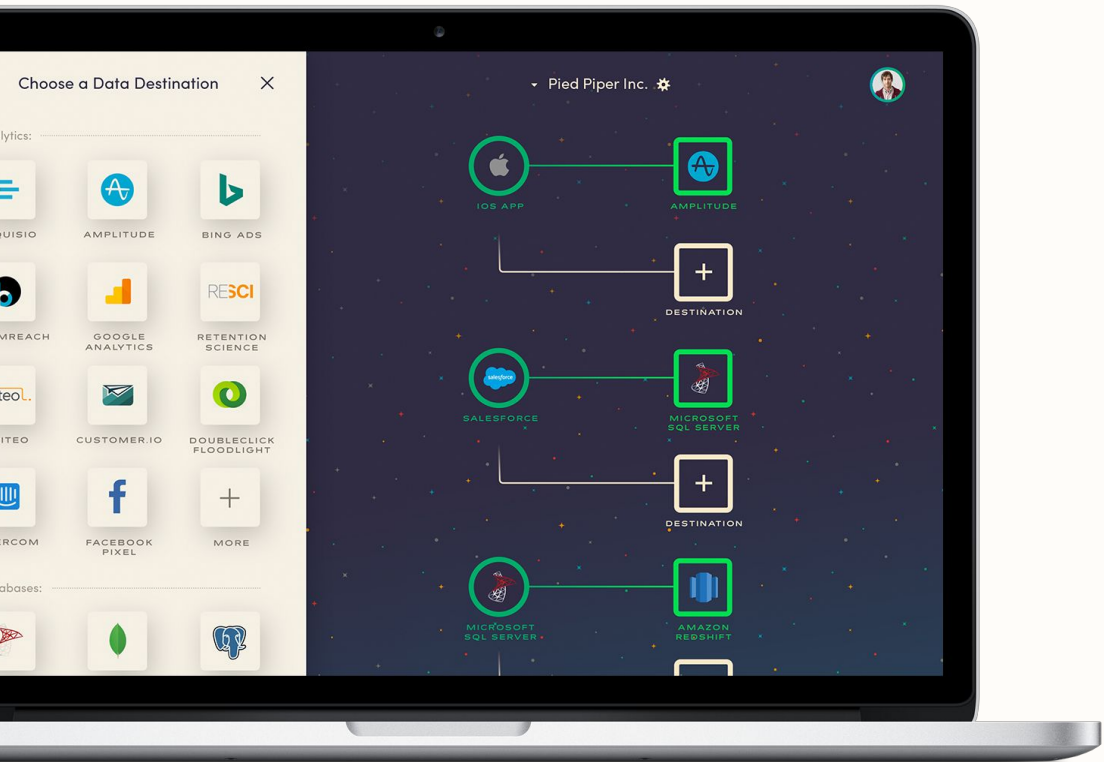## Greg Neiheisel

CTO

Astronomer

Data Engineering Platform

Streaming data
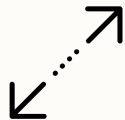Data pipelines
Code first ETL

Early Priorities

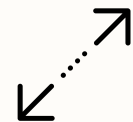Quick prototyping

Get data in motion

Ease of scale

Astronomer V1

Lambda + API Gateway

Cloudwatch for Monitoring

Kinesis + Elastic Beanstalk

Trouble in paradise

Strategic Obstacles

Companies view Amazon as direct competition

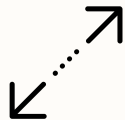Acquisition talks

Open source philosophy
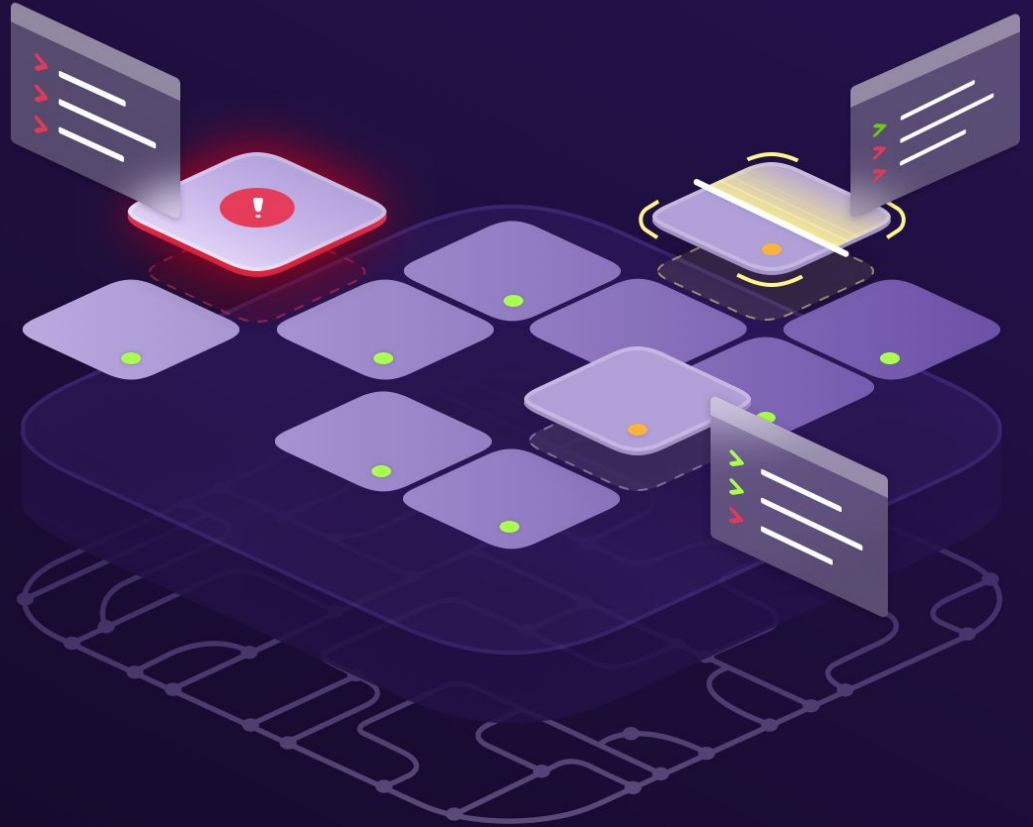
Engineering Obstacles

Access to customer data

Need a better tool for ETL

Deeply ingrained in the AWS ecosystem
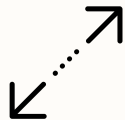
DC/OS

Single Unified Platform

DC/OS at Astronomer

Apache Airflow & Spark on Mesos

Marathon (Kubernetes?) replaces
Elastic Beanstalk

Foundation for open source DE platform

Apache Airflow
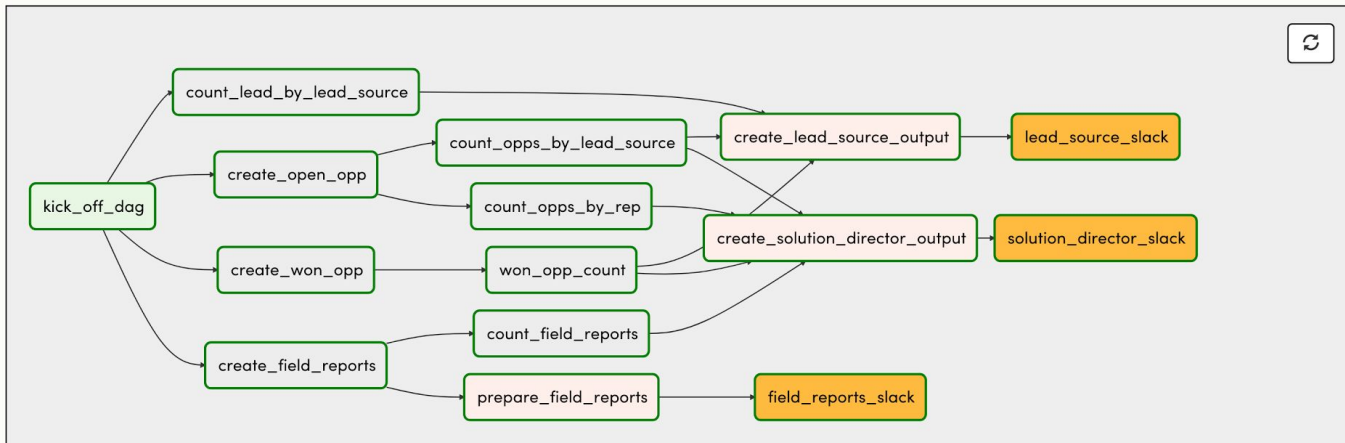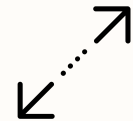
Airflow on Mesos

Leverage community-contributed Mesos executor

Up and running quickly

Scales to millions of tasks daily

Airflow at Astronomer

Behind the scenes to Managed service

Intelligent Redshift loading

Dependency driven tasks

Not all AWS tools are created equal
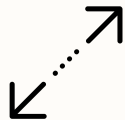
Kinesis to Kafka

Issues with Kinesis

Buggy Kinesis Client Library

Not available everywhere

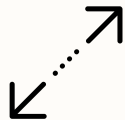Unable to tap into the Kafka ecosystem

The road to Kafka

Rewriting API and processors in Go

Improve provisioning, monitoring and testing

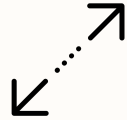Run systems in parallel

Kong and the inevitable end of API Gateway

Kong

Replaces API Gateway

Auth, rate limiting, lambda invocations for APIs

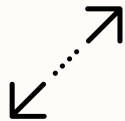Backed by Cassandra

CloudFormation + Ansible to Terraform

Terraform

Infrastructure as code

100% repeatable installs

Ease of scale
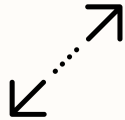
Rebuilding CloudWatch

Prometheus

All nodes monitored out of the box
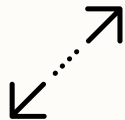
Write our own exporters

Ease of scale

ELK

Centralized logging
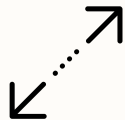
Aggregated queries across instances

KairosDB

Time series events collected via REST

Extremely durable, backed by Cassandra
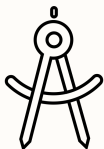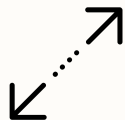
Rollups must be handled externally

R&D

Kafka Connect sources/sinks

Ceph or Minio

Druid

Istio, Weave, Kubernetes

Astronomer.io

Greg Neiheisel
Twitter: @schniebot
LinkedIn: greg-neiheisel