Live Bart Ridership Architecture EC2 Instance - tasks managed vis airflow Bart Live Arrival Data for every station (updated every Static s3 Webpage 10 minutes) 1) Daily predicted ridership per station Amazon Kinesis Ridership 2) Actual daily capacity per station Firehose prediction 3) Percent full calculate updated S3 Data Lake Live Weather Data (updated daily every 10 minutes) Weather data forecast for ía boto next 14 days (once per day) **Normalized Data** Ridership prediction (via boto) (managed via airflow) **Data normalization Static Historic Data** to 3NF Historic Bart hourly ridership Spark EMR data from 2016 Cluster EMR cluster tasks managed MongoDB via airflow on EC2 instance Historic Weather data from 2016 GBoost model to predict ridership