**Lookout for Equipment**

**Author: Ivan Miskulin and Jimmy Vila**

Amazon Lookout for Equipment Analyzes Sensor Data to Help Detect Equipment Failure

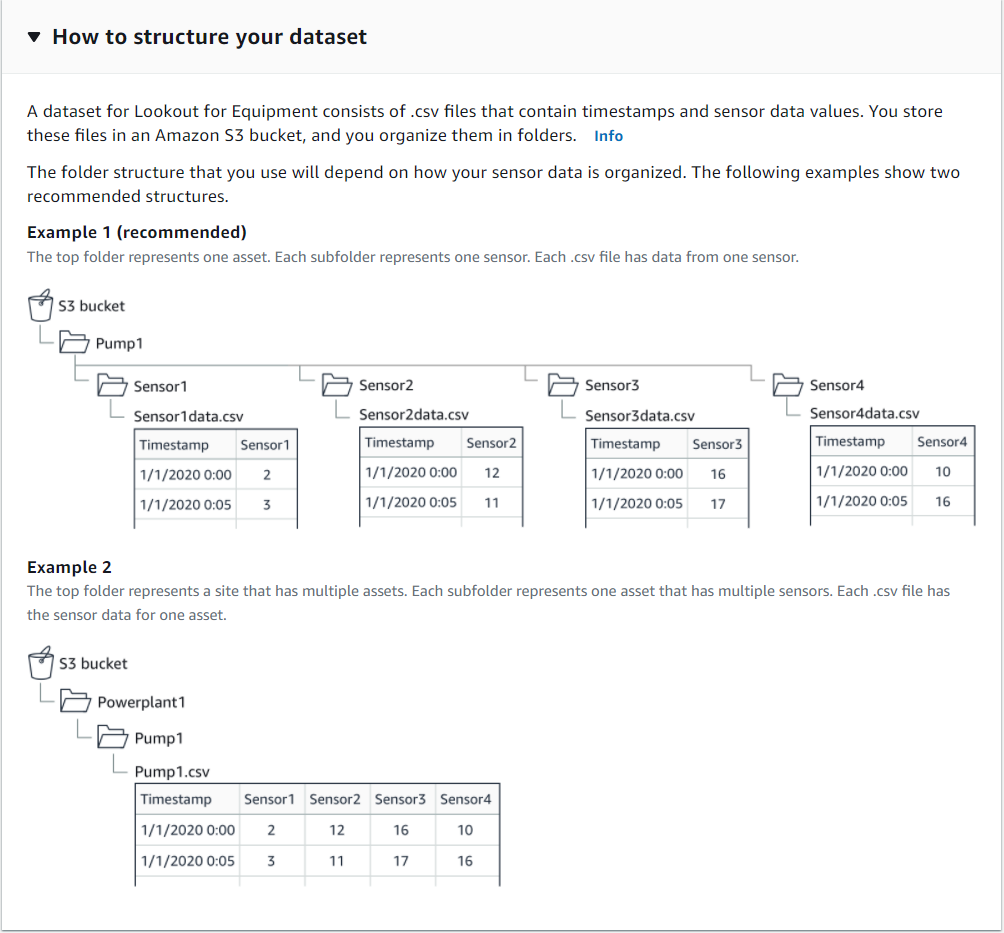
<https://aws.amazon.com/blogs/aws/new-amazon-lookout-for-equipment-analyzes-sensor-data-to-help-detect-equipment-failure/>

L4E can be set up in 5 steps:

1. Upload data to S3 buckets
2. Create datasets
3. Ingest data
4. Create a model
5. Schedule inference (if you need real-time analysis)

**1. Upload data to S3 bucket**

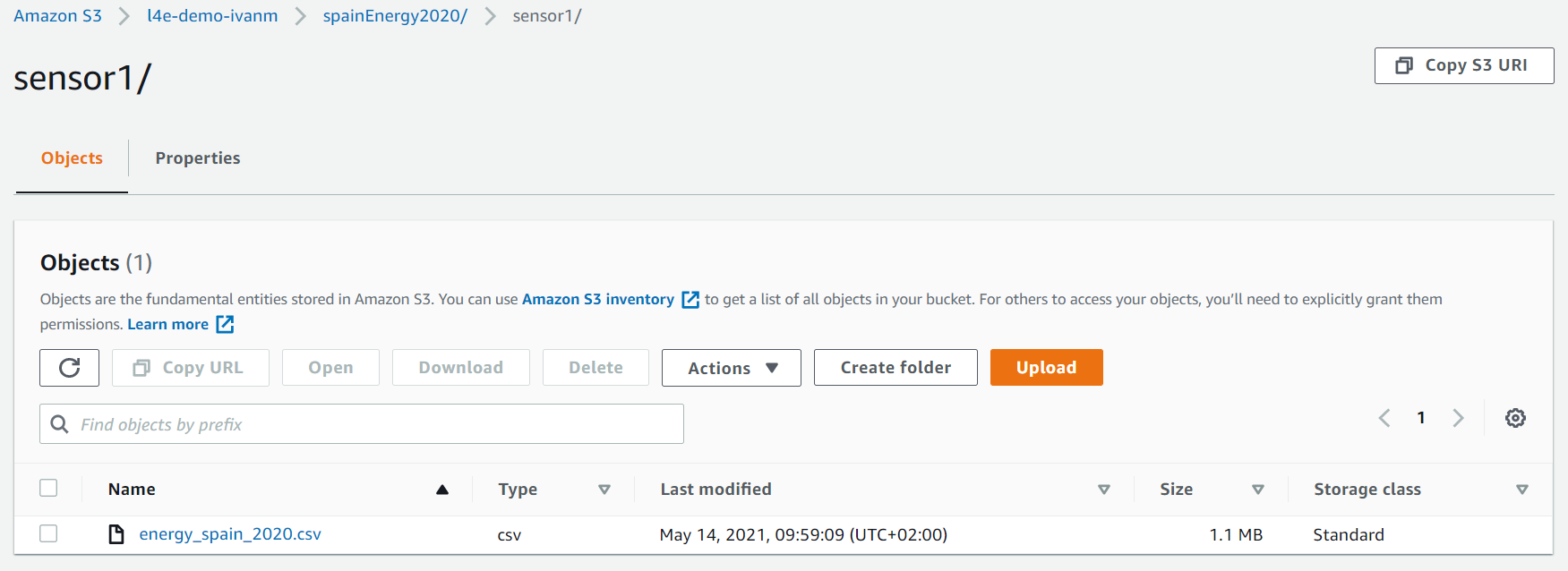
A dataset for Lookout for Equipment consists of .csv files that contain timestamps and sensor data values. You store these files in an Amazon S3 bucket, and you organize them in folders (Example 1). If you have only one sensor, use the structure of Example 2.



Save .csv file in any S3 bucket. Folder name and the folder structure have an impact to the Data Schema.

Note that there is only one .csv file in each folder.

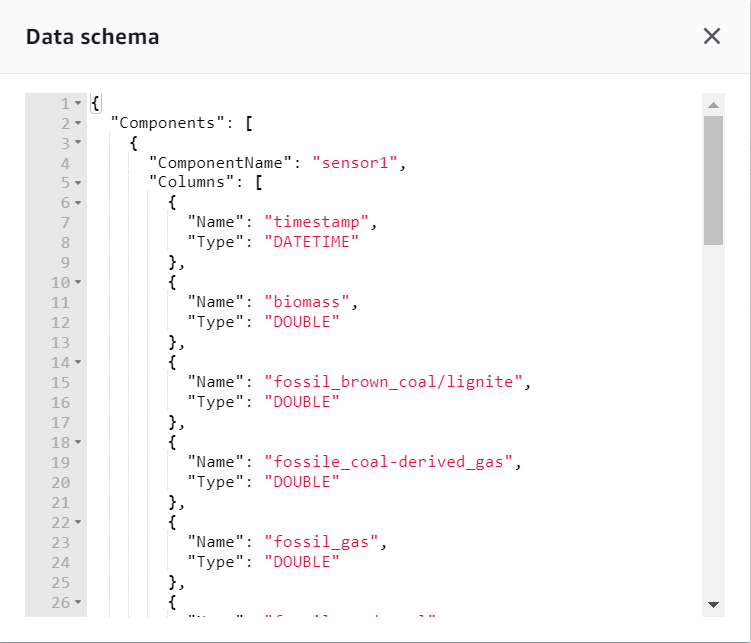
In the example below all data was saved to one .csv file which was saved in the folder /sensor1.

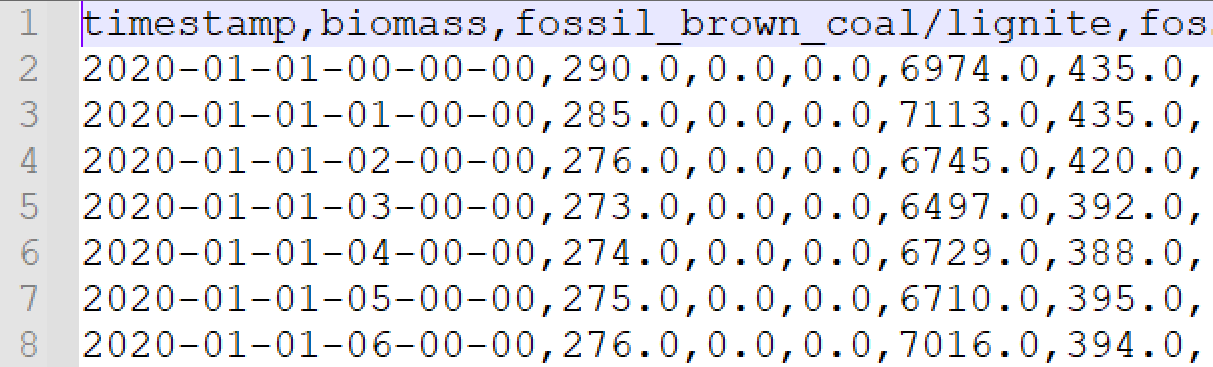


**2. Create datasets**

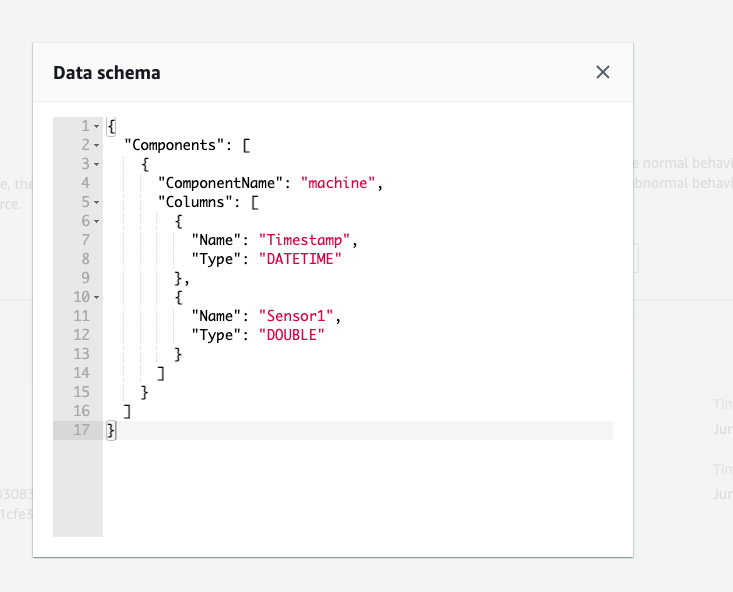
The “ComponentName” in the Data schema must match the name of the folder where the .csv file is saved.

Column names and types must match .csv file.





Schema for one sensor



**3. Ingest data**

To Ingest data you must specify the s3 path where the folder containing the data is located.

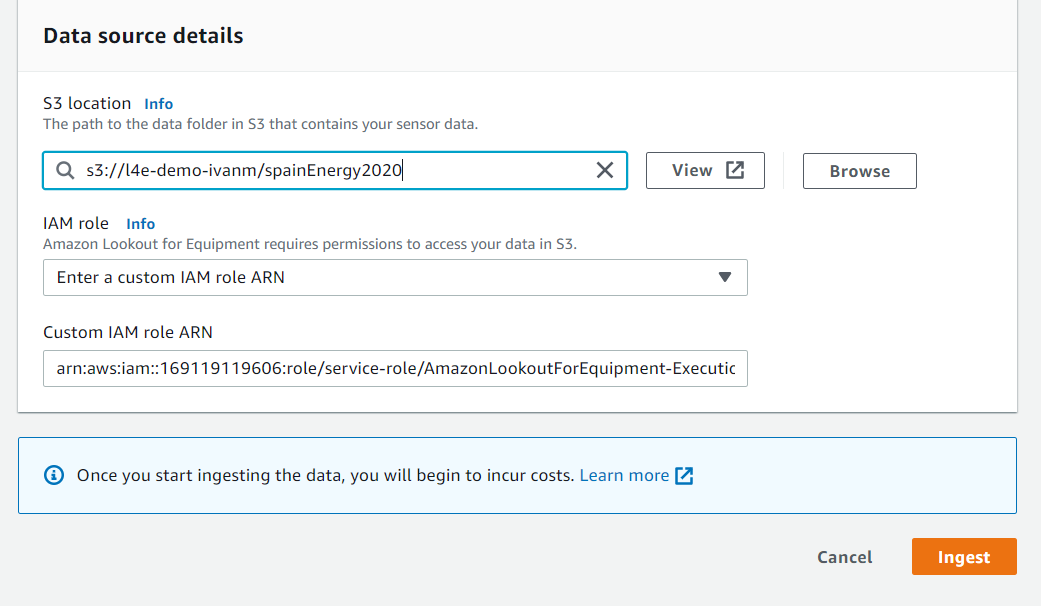
Example:

S3 location for the data is:

s3://l4e-demo-ivanm/spainEnergy2020/sensor1/energy\_spain\_2020.csv

In the Data schema it was specified that the first component is sensor1 so it will look for folder named sensor1 and load data from that folder. In this example path will be:  
s3://l4e-demo-ivanm/spainEnergy2020

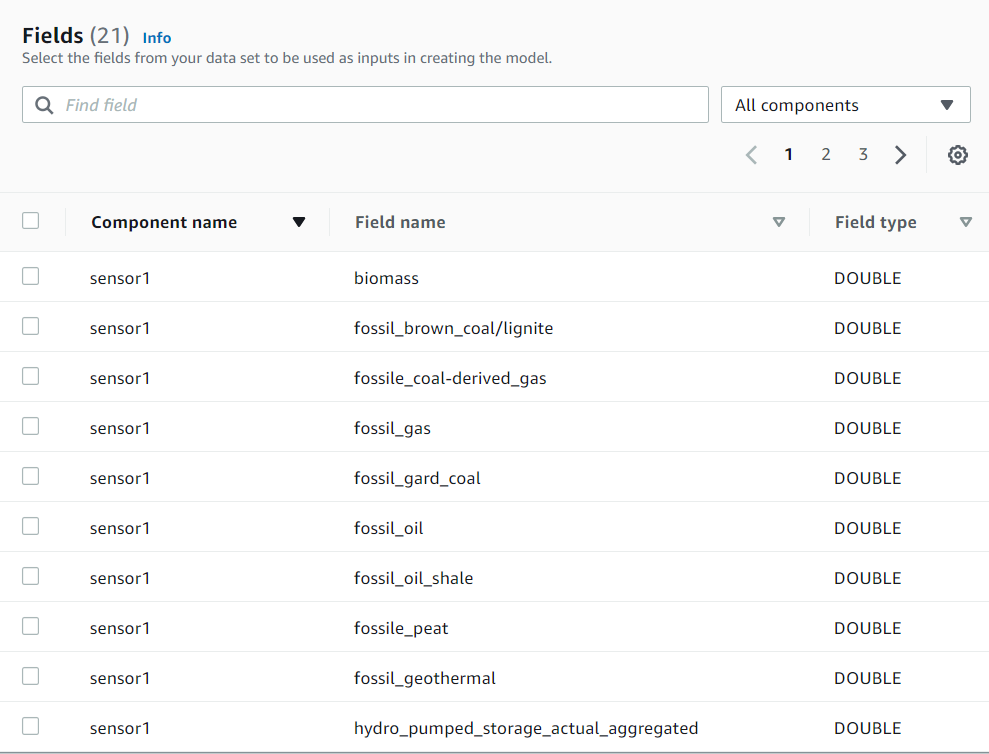
(In section 1. Upload data to S3 bucket was noted that there can only be one .csv file in each folder)



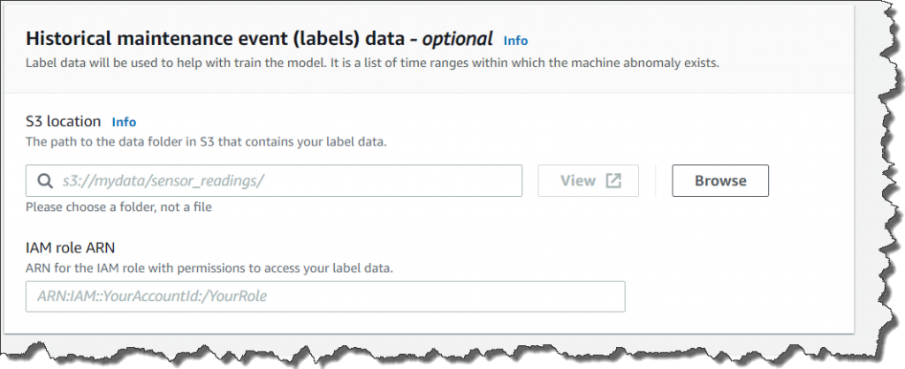
**4. Create a model**

After successfully ingesting data you can create a model. Fields shows us the list of fields in the ingested data. You can select fields you want L4E to learn. L4E automatically finds and trains correlations from multiple specified fields and creates a model.

Checking the checkbox of the Component name checks fields only for current page, e.g. page 1.

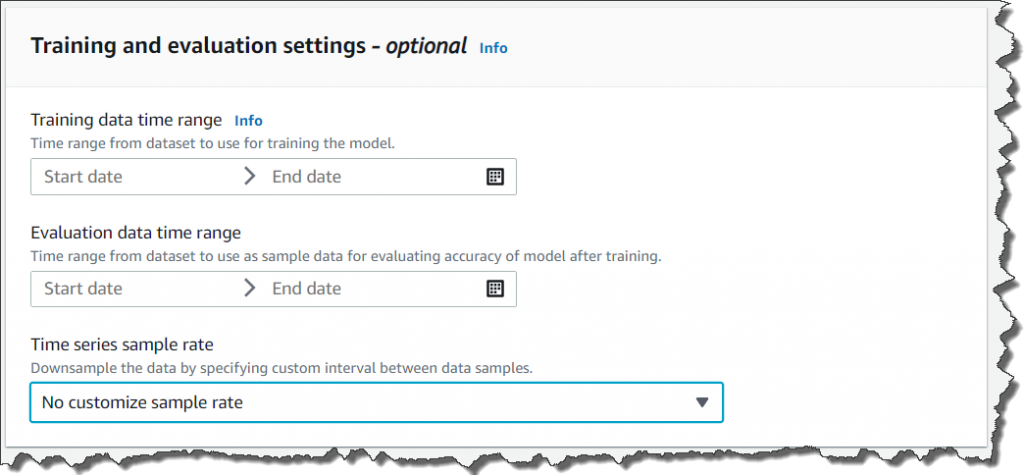


If you are sure that your data has some unusual data included, you can optionally set the windows to exclude that data.



Optionally, you can divide ingested data for training and then for evaluation. The data specified during the evaluation period is checked compared to the trained model.

**Note**: ***the upload frequency must be compatible with the sampling rate selected at training time.*** *For example, if a model was trained with 30 minutes resampling, asking for 5 minutes won't work.*



Once you select **Create**, Lookout for Equipment starts to train your model. This process takes minutes to hours depending on your data volume. After training is finished, you can evaluate your model with the evaluation period data.

**5. Schedule Inference**

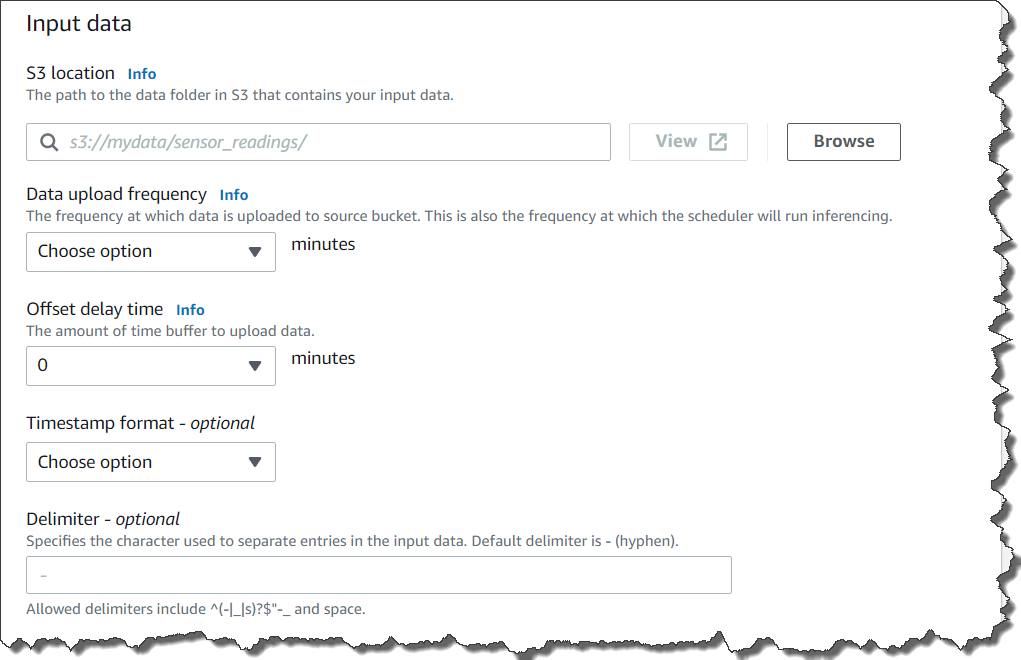
Now it is time to analyze your real time data. Select Schedule Inference and set up your S3 buckets for input. Again, you need to be careful as to which folder you specify as your input location for live data.

Here you need to define input folder and an output folder.

Example:

Input: S3://<BucketName>/realtime\_sensor\_input/

Output: S3://<BucketName>/realtime\_sensor\_output/



Note again: ***the upload frequency must be compatible with the sampling rate selected at training time***

**Sensor input data format (Very Important)**

Input: S3://<BucketName>/realtime\_sensor\_input/

If the schema has the follow format bellow, the correct name for each input will be this: **ComponetName\_Timestamp.csv**

