



Weather Data Project Brief

We love talking about weather in the UK, come wind, rain or sun we never tire of discussing it! It seems though that no matter how well prepared we are we never have our cagoule packed when needed.

Michael Fish famously said on the night of 15th October 1987 *"Earlier on today, apparently, a woman rung the BBC and said she heard there was a hurricane on the way... well, if you're watching, don't worry, there isn't!"*. He couldn't have been more wrong – that evening large parts of Britain were hit by the worst storm in over 300 hundred years leading to the tragic loss of 19 lives.

<https://www.youtube.com/watch?v=eKPQLI5rupg>

We have been tasked with a project to ingest, blend and analyse forecasted weather data from the MetOffice DataPoint API and Twitter Streaming API to compare the accuracy of MetOffice forecasting to a quantified record of live weather retrieved from the Twitter firehose.

We need to work with certain enforced constraints including:

- Twitter Streaming API limits: <https://support.twitter.com/articles/15364>
- Connection Parameter Filters: <https://dev.twitter.com/streaming/overview/request-parameters>
- Met Office DataPoint documentation:
 - <http://www.metoffice.gov.uk/datapoint/support/documentation/code-definitions>
 - Availability of last 24 hours only and forecasted data

We would like you to utilise all the big data skills gained so far. Initially this may involved storing your json data as an nvarchar(max) field within an SQL Server Database but as we progress through the course we will want to utilise better processes as they lend themselves to the technology i.e:

- Agile Sprint planning for the task allocation
- GitHub repo for project work
- NoSQL Mongo for initial storage of the json documents
- Hadoop HDFS for persisted storage of data files
- Sqoop for ingestion to/from SQL
- Hadoop Oozie scheduling
- Python Plot.ly for presentation of Analysis

We need you to work in small Agile teams for this project and remember to please keep your agile board up to date! Your work will need to be well presented and eventually be communicated to stakeholder level.

Final tip: don't get yourself rate limited by Twitter!