

Live Streaming weather data and crunching big weather numbers

Semester: Spring 2016

Instructor: Prof. Dino Konstantopoulos

Author: Anantha Shankar Arun Kumar

Siddhartha Singh

Table of Contents

1. Motivation
2. Description of Dataset and API
3. Problems faced
4. The Process
5. Conclusion
6. References

Motivation

Weather plays an important role in everyday life and it affects everyone in one-way or the other. Our main emphasis was to develop an intelligent system, which can process historic as well as real time weather data and predict the weather of a specific location.

The applications of this project can be used in many ways. Intelligent cars and intelligent homes can suggest a driver to modify or automatically modify the behavior and provide greater comfort. This can help a person to avoid alerts and also help a person plan vacation.

Description of dataset and API

Dataset

The data summaries provided here are based on data exchanged under the World Meteorological Organization (WMO) World Weather Watch Program. The following is a description of the global surface summary of day product produced by the National Climatic Data Center (NCDC).

The data are available via:

- 1) WWW -- <http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html>
- 2) FTP -- <ftp://ftp.ncdc.noaa.gov/pub/data/gsod>

Global summary of day data for 18 surface meteorological elements are derived from the synoptic/hourly observations contained in USAF DATSAV3 Surface data and Federal Climate Complex Integrated Surface Data (ISD).

In deriving the summary of day data, a minimum of 4 observations for the day must be present (allows for stations which report 4 synoptic observations/day).

Forecast API

The Forecast API allows you to look up the weather anywhere on the globe, returning:

- Current conditions
- Minute-by-minute forecasts out to 1 hour
- Hour-by-hour forecasts out to 48 hours
- Day-by-day forecasts out to 7 days

An account was created on <https://developer.forecast.io/> which facilitates 1000 free API calls per day. The API is very well documented and very well supported.

Problems Faced

The integration of all the technologies used by us to achieve what we wanted was a big challenge. The technologies like Apache Kafka, Apache Storm, Apache Hive, Apache HBase, Apache Hadoop etc were used to provide a scalable and fault tolerant application but their implementation was a tough challenge.

The output of Apache Storm was stored in HBase in a structured manner, which proved a tough challenge.

The data, which was linearly regressed and stored in HBase, was also a tough job.

The Process

- We used **Apache Kafka** to produce events by calling **Forecast.io API**, the data is then ingested and processed through **Apache Storm** and then stored in multiple sources like **HBase** and **Hive**. This data is finally produced through **Tableau** to show different projections and connection between the data.
- We ran **multiple map reduce** on the GSOD data of 27 years and then ran **linear regression** which, predict the forecast for a day in the future.
- We compared the output of both API data and Linear regression from historical data to find how accurate our model is.
- We also derived various relations between humidity and precipitation, cloud cover and humidity, Max and Min temp of different cities for a particular day.
- We pulled historical data for 20 biggest cities in the world and data from Forecast.io API.

Initial Data -

984290-99999-1990 UNREGISTERED

984290-99999-1990 x

1 018989429999991990011010004+14517+1210085Y-MT+0015RPMW V0209991C00001220001C0N03000IN1+02401+02851101351ADDA106000091AG14000AY141061AY241061GA1011+0075 00895GF61910911810108001801KA1240N+02101MA181291999999M1212021+99999M1051MW261REMSY017333 20210 56999MTE015THICK HZ ALQDS?

2 0071584299999919900110108001801KA1240N+02101MA18129199999M1212021+99999M1051MW261REMSY017333 20210 56999MTE015THICK HZ ALQDS?

3 0071584299999919900110108001801KA1240N+02101MA18129199999M1212021+99999M1051MW261REMSY017333 20210 56999MTE015THICK HZ ALQDS?

4 0071584299999919900110108001801KA1240N+02101MA18129199999M1212021+99999M1051MW261REMSY017333 20210 56999MTE015THICK HZ ALQDS?

5 01133984299999919900118108304+14517+1210085Y-MT+0015RPMW V0203608N001220001N0080008IN1+02781+02121101241ADDAA106000091AG14000AY141031AY241031GA1021+0075 00895GF1910911810108001801MA1101911999999M1210111+99999M1051MW261REMET004IAQ?

6 007898429999991990011010404+14517+1210085Y-MT+0015RPMW V0202880N0012220001CN1120081N+0291+02101999999ADAG129996GA1021+0075096898MA11010811999999REMET089H Z ALQDS?EQD001+000000PRSMW

7 0066998429999991990011010504+14517+1210085Y-MT+0015RPMW V0202707N0012220001CN1120081N+0291+02101999999ADAG129996GA1021+0075096898MA11010811999999M1210111+99999M1051MW261REMSY011333 56999MTE013LG HZ ALQDS?

8 0159984299999919900110106504+14517+1210085Y-MT+0015RPMW V0202707N0012220001N01080001IN1+03501+02151108941ADDAA106000091AG14000AY141061AY241061GA1021+0075 00895GF1029101182108001801MA1108911999999M1210111+99999M1051MW261REMSY011333 56999MTE013LG HZ ALQDS?

9 0078984299999919900110107514+14517+1210085Y-MT+0015RPMW V0202880N0012220001CN1120081N+02781+02101999999ADAG129996GA1011+0075096898MA1100911999999REMET060NSIG?EQD001+000000PRSMW

10 0078984299999919900110108004+14517+1210085Y-MT+0015RPMW V0202880N0012220001CN1120081N+03101+02101999999ADAG129996GA1011+0075096898MA1100911999999REMET060NSIG?EQD001+000000PRSMW

11 014998429999991990011010904+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN1120081N+0291+02115108931ADDAA106000091AG14000AY121031AY221031GF101991011 011080001010001MA1108911999999M1210110+99999M1051MW261REMSY011333 56999MTE006NSIG?EQD001+000000PRSMW

12 0066998429999991990011011004+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN00800001IN1+0291+02101999999ADAG129996GA1031+007809029MA1100911999999M12061REMET012LHQ AFew CS?

13 0061984299999919900110110804+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN00800001IN1+0291+02101999999ADAG129996GA1031+007809029MA1101081201999999M12061REMET005LHA0?

14 01599842999999199001101110404+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN00800001IN1+0291+02101999999ADAG129996GA1031+007809029MA1101081201999999M12061REMET005LHA0?

15 00619842999999199001101110804+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN00800001IN1+0291+02101999999ADAG129996GA1031+007809029MA1101081201999999M12061REMET005LHA0?

16 00789842999999199001101110804+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN00800001IN1+0291+02101999999ADAG129996GA1031+007809029MA1101081201999999M12061REMET005LHA0?

17 01549842999999199001101110504+14517+1210085Y-MT+0015RPMW V0200801N0012220001N00800001IN1+02451+02171101151ADDAA106000091AG14000AY141031AY241031GF102999199 011080001010001MA110891199999M12061REMSY011333 56999MTE005LHA0?EQD001+00000025COTLC

18 00549842999999199001101110604+14517+1210085Y-MT+0015RPMW V0201601N00621999999N00080001IN1+02401+02101999999ADAG129996MA1101081999999M12061REMET005LHA0?

19 00459842999999199001101110704+14517+1210085Y-MT+0015RPMW V0209991C0001800000800001IN1+02401+02081999999ADAG129996MA1101081999999M12061REMET005LHA0?

20 01829842999999199001101110804+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN00800001IN1+02401+02101999999ADAG129996GA1031+007809029MA1101081201999999M12061REMET005LHA0?

21 00519842999999199001101110804+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN00800001IN1+02401+02101999999ADAG129996GA1031+007809029MA1101081201999999M12061REMET005LHA0?

22 00789842999999199001101110804+14517+1210085Y-MT+0015RPMW V0200901N0012220001CN00800001IN1+02401+02101999999ADAG129996GA1031+007809029MA1101081201999999M12061REMET005LHA0?

23 00628842999999199001101120804+14517+1210085Y-MT+0015RPMW V0209991C00012220001CN00800001IN1+02401+02101999999ADAG129996GA1021+0075096898MA1101081201999999M12061REMET006NSIG?EQD001+00000023COTLC

24 00789842999999199001101120804+14517+1210085Y-MT+0015RPMW V0209991C00012220001CN00800001IN1+02401+02101999999ADAG129996GA1021+0075096898MA1101081201999999M12061REMET006NSIG?EQD001+00000023COTLC

25 00628842999999199001101120804+14517+1210085Y-MT+0015RPMW V0209991C00012220001CN00800001IN1+02401+02101999999ADAG129996GA1021+0075096898MA1101081201999999M12061REMET006NSIG?EQD001+00000023COTLC

26 00789842999999199001101120804+14517+1210085Y-MT+0015RPMW V0209991C00012220001CN00800001IN1+02401+02101999999ADAG129996GA1021+0075096898MA1101081201999999M12061REMET006NSIG?EQD001+00000023COTLC

27 00628842999999199001101120804+14517+1210085Y-MT+0015RPMW V0209991C00012220001CN00800001IN1+02401+02101999999ADAG129996GA1021+0075096898MA1101081201999999M12061REMET006NSIG?EQD001+00000023COTLC

28 00789842999999199001101120804+14517+1210085Y-MT+0015RPMW V0209991C00012220001CN00800001IN1+02401+02101999999ADAG129996GA1021+0075096898MA1101081201999999M12061REMET006NSIG?EQD001+00000023COTLC

29 00628842999999199001101120804+14517+1210085Y-MT+0015RPMW V0209991C00012220001CN00800001IN1+02401+02101999999ADAG129996GA1021+0075096898MA1101081201999999M12061REMET006NSIG?EQD001+00000023COTLC

The initial data was like this and we extracted max temperature, min temperature and standard pressure from all the files.

After 1st Map reduce job, we got a file which was much more readable like this

	part-00000	*
1	1990_01_01_Beijing	Beijing,01-01,1,7,-6,9,1032,3
2	1990_01_01_Boston	Boston,01-01,11,1,11,1,1004,7
3	1990_01_01_Cairo	Cairo,01-01,19,4,8,5,1018,2
4	1990_01_01_Chicago	Chicago,01-01,0,0,-3,3,1024,5
5	1990_01_01_Karachi	Karachi,01-01,25,0,12,4,1018,0
6	1990_01_01_Las_Vegas	Las Vegas,01-01,13,9,-4,4,1024,2
7	1990_01_01_London	London,01-01,6,6,5,5,1018,3
8	1990_01_01_Los_Angeles	Los Angeles,01-01,16,1,13,3,1021,8
9	1990_01_01_Madrid	Madrid,01-01,8,6,7,8,1021,6
10	1990_01_01_Manila	Manila,01-01,30,5,21,7,1013,5
11	1990_01_01_New_Delhi	New Delhi,01-01,14,4,8,9,1019,1
12	1990_01_01_New_York_City	New York City,01-01,10,6,10,0,1012,6
13	1990_01_01_Paris	Paris,01-01,0,0,-0,8,1019,7
14	1990_01_01_Sydney	Sydney,01-01,31,8,20,5,1018,5
15	1990_01_01_Toronto	Toronto,01-01,1,5,-0,5,1020,2
16	1990_01_01_Washington_DC	Washington DC,01-01,7,8,7,8,1021,0
17	1990_01_02_Beijing	Beijing,01-02,0,0,-14,0,1033,8
18	1990_01_02_Boston	Boston,01-02,4,4,-5,6,1024,4
19	1990_01_02_Cairo	Cairo,01-02,25,2,10,8,1013,2
20	1990_01_02_Chicago	Chicago,01-02,3,9,-1,7,1025,6
21	1990_01_02_Karachi	Karachi,01-02,23,0,16,0,1017,2
22	1990_01_02_Las_Vegas	Las Vegas,01-02,11,7,0,0,1012,7
23	1990_01_02_London	London,01-02,9,0,6,5,1018,5
24	1990_01_02_Los_Angeles	Los Angeles,01-02,16,1,10,0,1016,7
25	1990_01_02_Madrid	Madrid,01-02,7,0,5,4,1024,4
26	1990_01_02_Manila	Manila,01-02,30,4,20,3,1013,7
27	1990_01_02_New_Delhi	New Delhi,01-02,13,6,10,1,1020,4
28	1990_01_02_New_York_City	New York City,01-02,5,6,-6,1,1028,2
29	1990_01_02_Paris	Paris,01-02,2,6,2,6,1022,9
30	1990_01_02_Sydney	Sydney,01-02,23,4,17,8,1017,8
31	1990_01_02_Toronto	Toronto,01-02,1,0,-4,0,1022,2
32	1990_01_02_Washington_DC	Washington DC,01-02,7,8,-3,2,1032,2
33	1990_01_03_Beijing	Beijing,01-03,0,0,-8,2,1031,7
34	1990_01_03_Boston	Boston,01-03,8,9,0,0,1025,4
35	1990_01_03_Cairo	Cairo,01-03,19,9,11,2,1012,6
36	1990_01_03_Chicago	Chicago,01-03,5,6,3,3,1020,1
37	1990_01_03_Karachi	Karachi,01-03,18,7,18,2,1017,6
38	1990_01_03_Las_Vegas	Las Vegas,01-03,8,9,-0,6,1026,0
39	1990_01_03_London	London,01-03,5,8,3,9,1018,7
40	1990_01_03_Los_Angeles	Los Angeles,01-03,16,7,6,1,1024,5
41	1990_01_03_Madrid	Madrid,01-03,8,8,7,8,1014,8
42	1990_01_03_Manila	Manila,01-03,29,0,23,7,1014,7

The data was then processed in Apache storm and then liner regression was performed on the data.

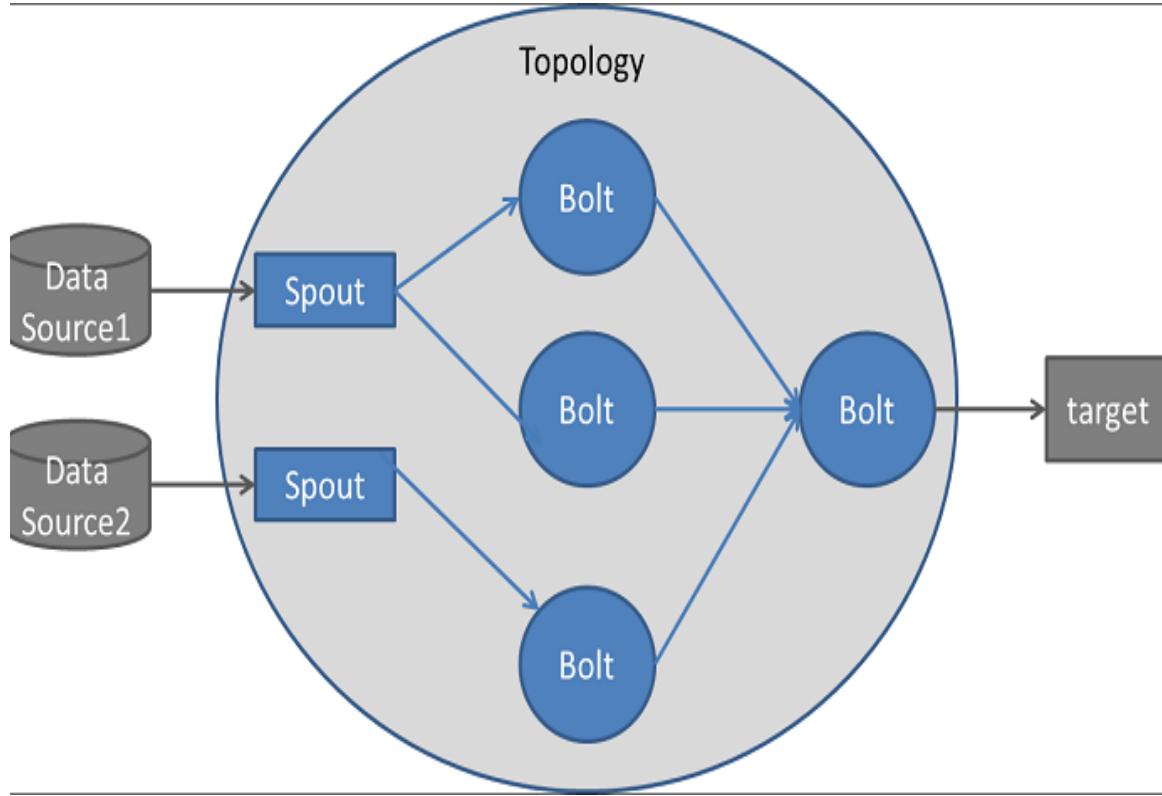
After linear regression the data looked like this

	equation_params	*
1	Las Vegas,26.982905227380826,7.208079337396301,974.5517209580154,0.004344894822405119,5.523614101975358E-4,0.19791039750393455	
2	Mexico City,27.727751480378284,10.952659374431644,973.6093494644077,0.0068999282676638326,-0.0013493768692220134,0.1974720781513116	
3	Toronto,25.641256304942374,12.822600509255958,975.3497973820238,0.007482257168143633,-5.157975445153076E-4,0.198301333425599	
4	New York City,25.42137932491476,15.00059464060682,976.2954422453495,0.006788495590612947,-0.0013812511955755553,0.19856183377893066	
5	Karachi,32.34575915201243,28.435391641985156,975.1392764549958,0.0037173706548719538,0.004368119846065627,0.1981488114456936	
6	New Delhi,37.15178266209754,35.298722096503006,974.150500118816,0.0046413759499345925,-2.1242890001554176E-5,0.1979620117302483	
7	Manila,37.33889294047263,35.6313575850622,973.7490541384255,0.006862235535126813,-1.2463532551911562E-4,0.1978137379434835	
8	Sydney,35.45655174502969,33.58322772699227,974.4548629155919,0.0063069596308965095,3.195468234576694E-4,0.19801695940116937	
9	Seoul,34.007324087047,32.02191169072674,975.0039176538512,0.005819171254125485,5.399238686634765E-4,0.19816664231689127	
10	Tokyo,32.84835526013797,30.758330155729233,975.4437992479609,0.0054278311652021915,7.759116960675007E-4,0.1982927438966976	
11	Cairo,33.685661009708184,31.66191185393958,975.41664864642335,0.005310614529433604,1.764704489276688E-4,0.19825681083417945	
12	Los Angeles,32.881425304784244,31.1190722956849,975.482098418865,0.004804997428809511,-3.660392605139331E-4,0.19824695797836378	
13	Rio de Janeiro,32.32434691984106,30.786933877403634,975.528447606812,0.004242967538283077,-0.0012130488776117168,0.19828948234455535	
14	Washington DC,31.675609265768948,29.754252722759002,975.8379322150537,0.004289040408121022,-5.861803305899225E-4,0.1983758832510665	
15	Beijing,30.99062342058239,30.7647671363771,976.0448245506667,0.004259448073747038,-0.0010378146840476041,0.19836949188566502	
16	Chicago,30.16014225691921,29.534043221955937,976.2268321677808,0.003849982300902414,-0.00112572136748218,0.19848350752890617	
17	London,29.34128871954179,28.58392746306205,976.3203496635124,0.003802609596478001,-7.076906487275904E-4,0.19852713271415504	
18	Paris,29.410417537107904,27.93262994499335,976.4180216621644,0.004788986034961728,-5.819006335650459E-4,0.19872552560801313	
19	Madrid,29.040177047539256,27.335155204879396,976.5824095464495,0.004579273865928295,-6.676873218575881E-4,0.1988493827600851	
20	Boston,28.808276945610796,27.01758185580858,976.6937107566837,0.006144960529750921,0.0013563663988749393,0.19887354613548291	
21		

For the real time weather forecasting we used Apache storm,

Apache Storm

- Apache Storm is a free and open source distributed real-time computation system.
 - Storm makes it easy to reliably process unbounded streams of data.
 - Storm can be used with any programming language.
-



Various components of Apache Storm :-

- **Tuple** - Main data structure in Storm. It is a list of ordered elements.
- **Stream** - Stream is an unordered sequence of tuples.
- **Spouts** - Source of stream. Storm accepts input data from data sources like Apache Kafka queue, Twitter Streaming API etc.
- **Bolts** – Bolts are logical processing units. Spouts pass data to bolts and bolts process and produce a new output stream.

Storm vs Hadoop

Advantages of using Apache Storm

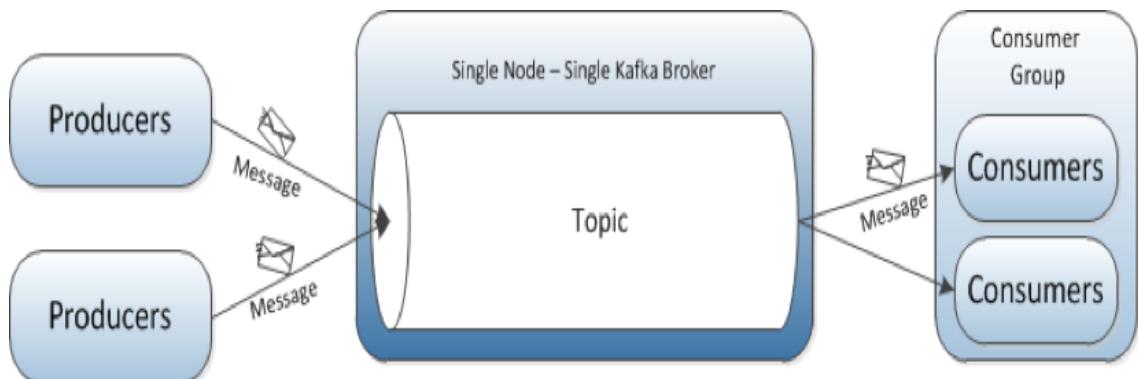
- Allows real time processing
- Storm is unbelievably fast because it has enormous power of processing the data.
- Storm performs data refresh and end to end delivery in seconds . It has very low latency.
- Storm provides guaranteed data processing even if any of the connected nodes in the cluster die or messages are lost.

Distributed Messaging System

- Is based on the concept of reliable message queuing.
- Messages are queued asynchronously between client applications and messaging systems.
- Most of the messaging patterns follow the publish-subscribe model where senders of the messages are called publishers and those who want to receive the messages are called subscribers.

Apache Kafka

- Kafka is a distributed, partitioned, replicated commit log service.
- Kafka maintains feeds of messages in categories called topics.
- Producers - Processes that publish messages to a Kafka topic.
- Consumers – Processes that subscribe to topic and process the feed of published messages.
- Broker – Kafka is run as a cluster comprised of one or more servers, each of which is called a broker.



Apache Hbase

- Provides real-time, random read and write access to tables storing billions of rows and millions of columns.
- In this case, once we store this rapidly and continuously growing dataset, we will be able to perform a swift lookup for analytics regardless of the data size.

Apache Hive

- Facilitates querying and managing large datasets residing in distributed storage.
- Hive provides mechanism to project structure onto this data and query the data using a SQL-like language called HiveQL.

Statistics and Results

Storm UI Statistics:

Below two screen shots statistic shows the how many stream messages have been emitted, transferred through storm which confirms the processing and streaming of storm

Storm UI

Topology summary

Name	Id	Owner	Status	Uptime	Num workers	Num executors	Num tasks	Replication count	Scheduler Info
weather-event-processor	weather-event-processor-3-1461927402		ACTIVE	6h 0m 0s	1	6	6	1	

Topology actions

Activate Deactivate Rebalance Kill Change Log Level

Topology stats

Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed
10m Os	5580	5620	1526.511	940	0
3h 0m Os	7700	6740	1542.679	1120	0
1d 0h 0m Os	23840	22000	4990.967	3620	0
All time	23840	22000	4990.967	3620	0

Spouts (All time)

Search: <input type="text"/>										
Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Error Host	Error Port	Last error
kafkaSpout	1	1	7580	10800	4990.967	3620	0			

Showing 1 to 1 of 1 entries

Input stats (All time)

Search: <input type="text"/>						
Component	Stream	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
kafkaSpout	default	53.698	3980	5.111	3980	0

Showing 1 to 1 of 1 entries

Output stats (All time)

Search: <input type="text"/>						
Stream	Emitted	Transferred				
__ack_ack	3980	3980				
__metrics	720	0				
__system	20	0				

Showing 1 to 3 of 3 entries

Executors

Search: <input type="text"/>											
Id	Uptime	Host	Port	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
[4-4]	6h 0m 24s	sandbox.hortonworks.com	6700	2360	2000	0.000	1.310	2000	8.430	2000	0
[5-5]	6h 0m 24s	sandbox.hortonworks.com	6700	2360	1980	0.209	106.616	1980	1.758	1980	0

Showing 1 to 2 of 2 entries

Below screen shot shows the statistics of stream processing of bolts of storm

Bolt stats

Window	Emitted	Transferred	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
10m Os	920	880	139.467	900	0.283	920	0
3h 0m Os	1820	1480	141.479	1460	0.541	1480	0
1d 0h 0m Os	4720	3980	53.698	3980	5.111	3980	0
All time	4720	3980	53.698	3980	5.111	3980	0

Showing 1 to 4 of 4 entries

Input stats (All time)

Component	Stream	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
kafkaSpout	default	53.698	3980	5.111	3980	0

Showing 1 to 1 of 1 entries

Output stats (All time)

Stream	Emitted	Transferred
_ack_ack	3980	3980
_metrics	720	0
_system	20	0

Showing 1 to 3 of 3 entries

Topology actions

Activate Deactivate Rebalance Kill Change Log Level

Topology stats

Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed
10m Os	5580	5620	1526.511	940	0
3h 0m Os	7700	6740	1542.679	1120	0
1d 0h 0m Os	23840	22000	4990.967	3620	0
All time	23840	22000	4990.967	3620	0

Spouts (All time)

ID	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Error Host	Error Port	Last error
kafkaSpout	1	1	7580	10800	4990.967	3620	0			

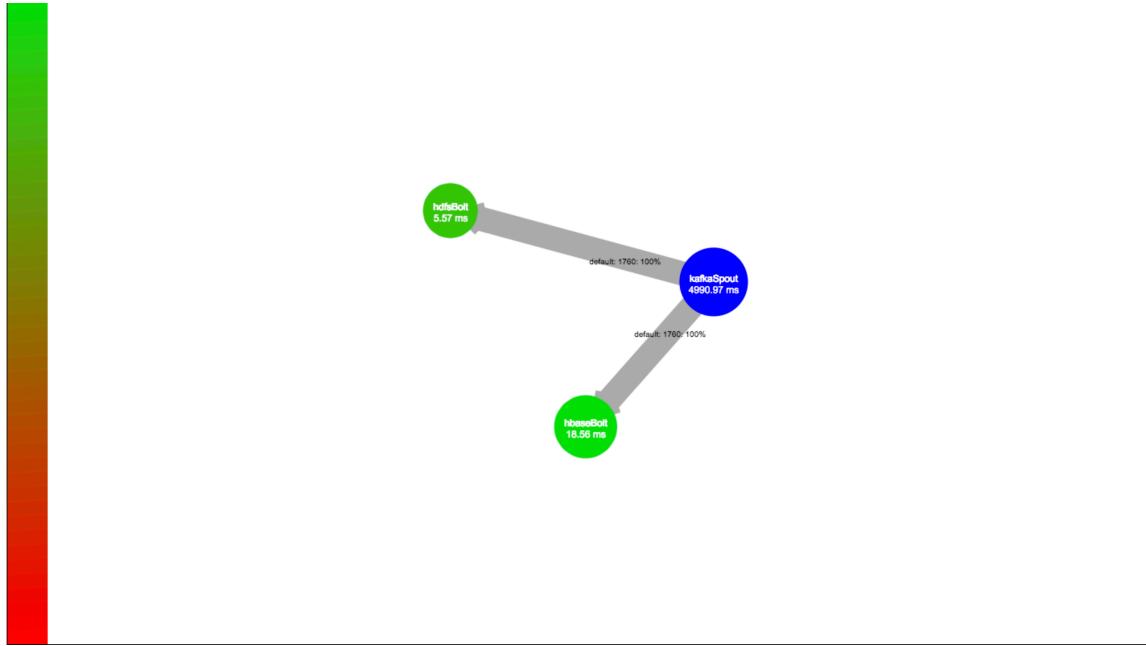
Showing 1 to 1 of 1 entries

Bolts (All time)

ID	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed	Error Host	Error Port	Last error
_acker	1	1	3960	3600	0.002	0.506	10840	0.297	10840	0			
hbaseBolt	2	2	7960	3980	0.017	29.189	3600	18.558	3620	0			
hdfsBolt	2	2	4340	3620	0.135	24.403	3620	5.575	3620	0			

Showing 1 to 3 of 3 entries

Below screen shot shows the percentage of stream processed and transferred to each bolt of storm i.e., hdfs and hbase received from kafka spout



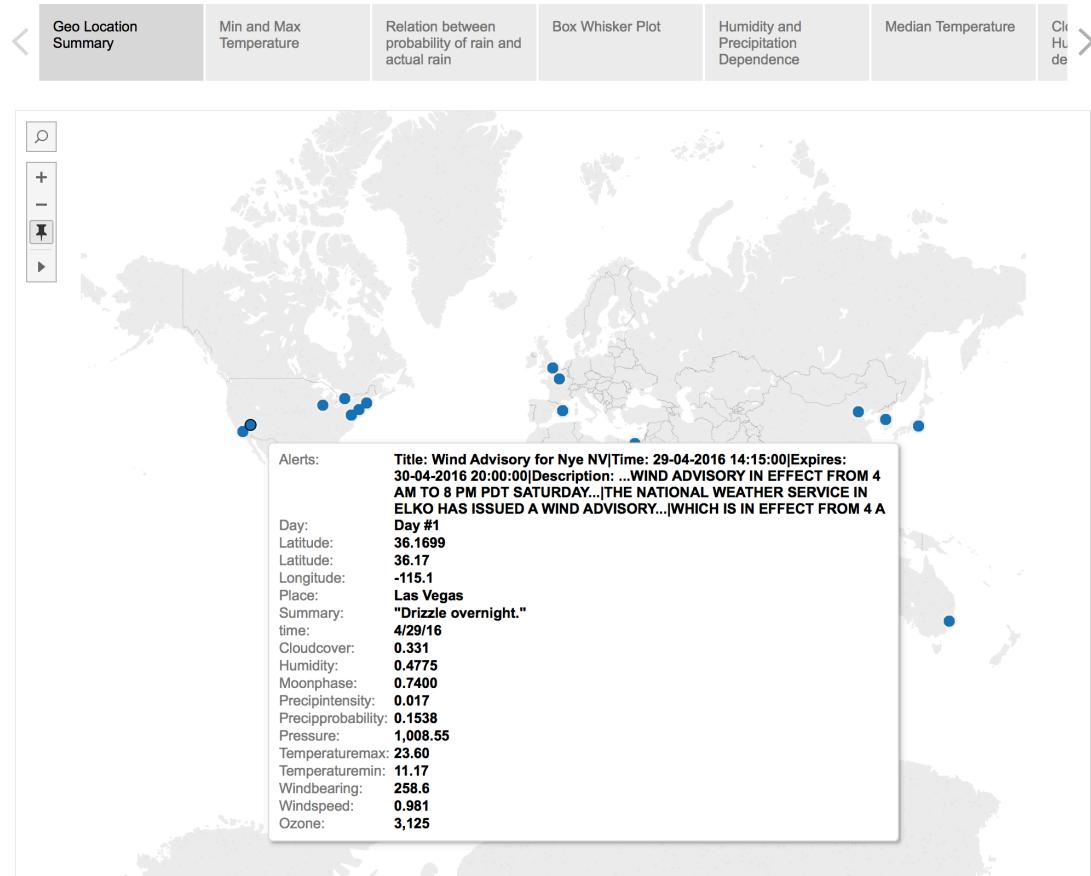
Below screen shot is the statistics of kafka spout stream processing

kafkaSpout		weather-event-processor		1	1																														
Spout stats																																			
<table border="1"> <thead> <tr> <th>Window</th><th>Emitted</th><th>Transferred</th><th>Complete latency (ms)</th><th>Acked</th><th>Failed</th></tr> </thead> <tbody> <tr> <td>10m 0s</td><td>2720</td><td>1840</td><td>1526.511</td><td>940</td><td>0</td></tr> <tr> <td>3h 0m 0s</td><td>3260</td><td>2380</td><td>1542.679</td><td>1120</td><td>0</td></tr> <tr> <td>1d 0h 0m 0s</td><td>10800</td><td>7580</td><td>4990.967</td><td>3620</td><td>0</td></tr> <tr> <td>All time</td><td>10800</td><td>7580</td><td>4990.967</td><td>3620</td><td>0</td></tr> </tbody> </table>						Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed	10m 0s	2720	1840	1526.511	940	0	3h 0m 0s	3260	2380	1542.679	1120	0	1d 0h 0m 0s	10800	7580	4990.967	3620	0	All time	10800	7580	4990.967	3620	0
Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed																														
10m 0s	2720	1840	1526.511	940	0																														
3h 0m 0s	3260	2380	1542.679	1120	0																														
1d 0h 0m 0s	10800	7580	4990.967	3620	0																														
All time	10800	7580	4990.967	3620	0																														
Output stats (All time)																																			
<table border="1"> <thead> <tr> <th>Stream</th><th>Emitted</th><th>Transferred</th><th>Complete latency (ms)</th><th>Acked</th><th>Failed</th></tr> </thead> <tbody> <tr> <td>_ack_init</td><td>3680</td><td>3680</td><td>0</td><td>0</td><td>0</td></tr> <tr> <td>_metrics</td><td>340</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr> <td>default</td><td>3560</td><td>7120</td><td>4990.967</td><td>3620</td><td>0</td></tr> </tbody> </table>						Stream	Emitted	Transferred	Complete latency (ms)	Acked	Failed	_ack_init	3680	3680	0	0	0	_metrics	340	0	0	0	0	default	3560	7120	4990.967	3620	0						
Stream	Emitted	Transferred	Complete latency (ms)	Acked	Failed																														
_ack_init	3680	3680	0	0	0																														
_metrics	340	0	0	0	0																														
default	3560	7120	4990.967	3620	0																														
Showing 1 to 3 of 3 entries																																			
Executors (All time)																																			
<table border="1"> <thead> <tr> <th>ID</th><th>Uptime</th><th>Host</th><th>Port</th><th>Emitted</th><th>Transferred</th><th>Complete latency (ms)</th><th>Acked</th><th>Failed</th></tr> </thead> <tbody> <tr> <td>[6-6]</td><td>5h 56m 23s</td><td>sandbox.hortonworks.com</td><td>6700</td><td>7580</td><td>10800</td><td>4990.967</td><td>3620</td><td>0</td></tr> </tbody> </table>						ID	Uptime	Host	Port	Emitted	Transferred	Complete latency (ms)	Acked	Failed	[6-6]	5h 56m 23s	sandbox.hortonworks.com	6700	7580	10800	4990.967	3620	0												
ID	Uptime	Host	Port	Emitted	Transferred	Complete latency (ms)	Acked	Failed																											
[6-6]	5h 56m 23s	sandbox.hortonworks.com	6700	7580	10800	4990.967	3620	0																											
Showing 1 to 1 of 1 entries																																			

Tableau Analysis and Statistics

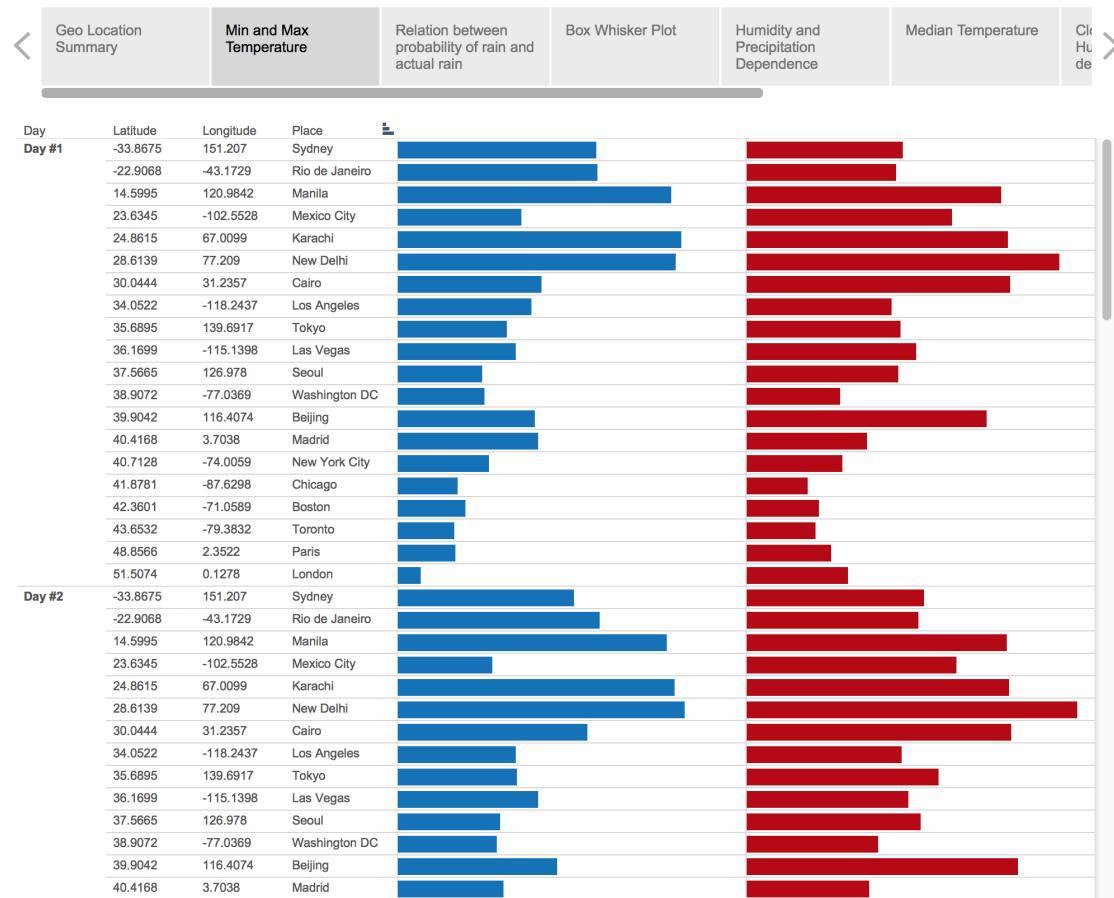
Weather Forecast information plotted on map for each city

Weather Forecast

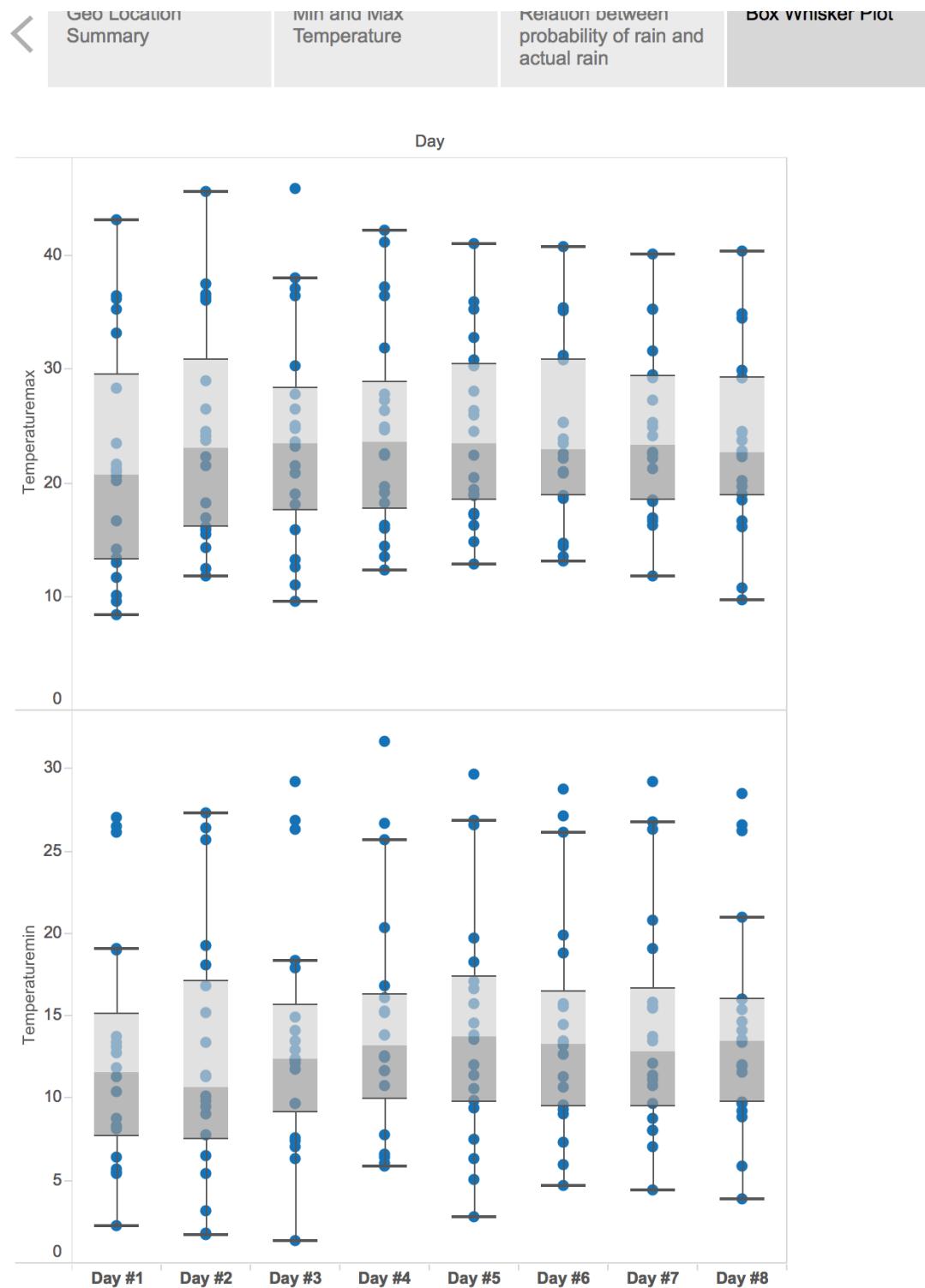


Min and Max temperature statistics for each place based on each day

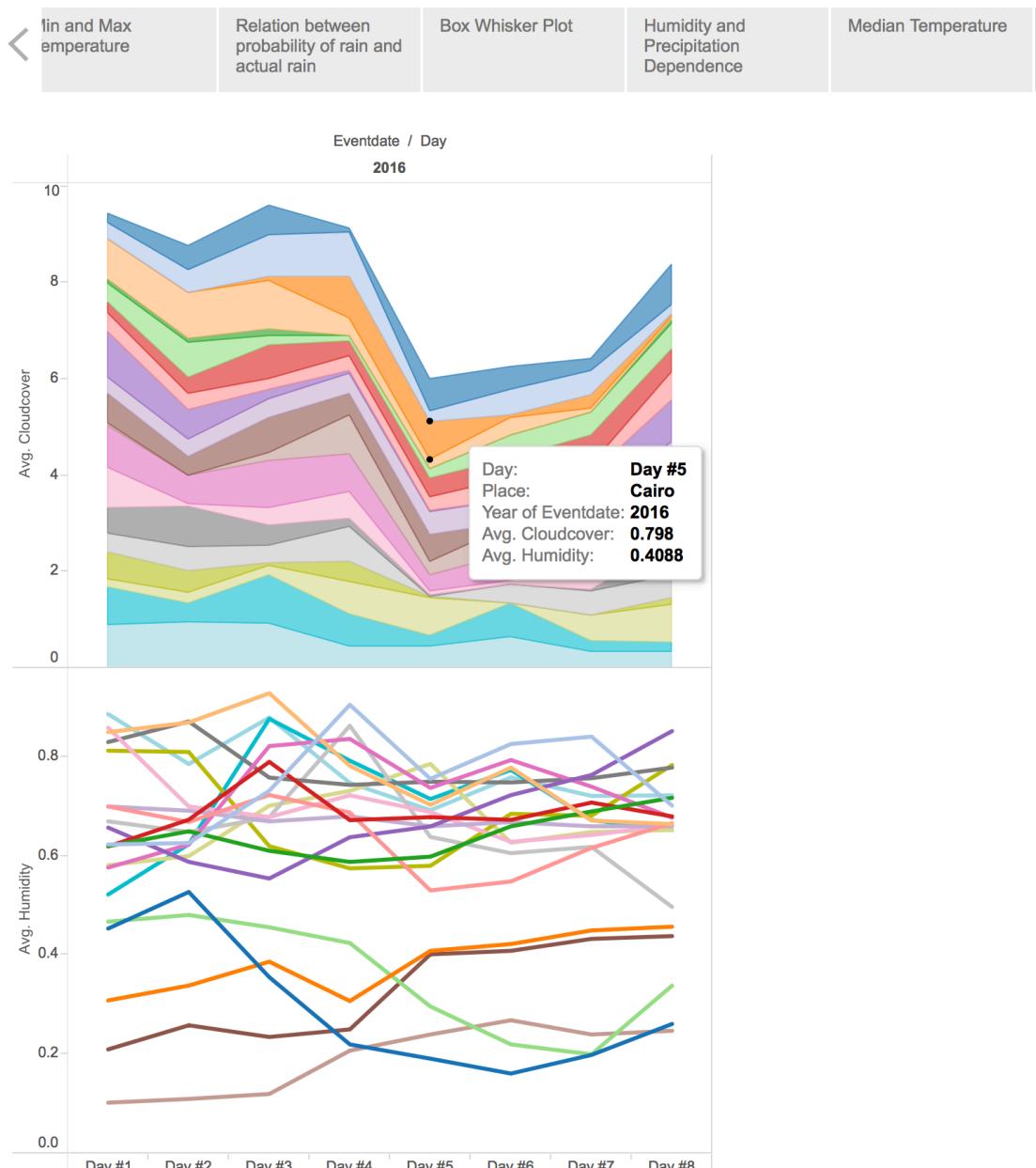
Weather Forecast



Box-Whisker plot shows the which city has max temperature and min temperature for a day sorted in descending



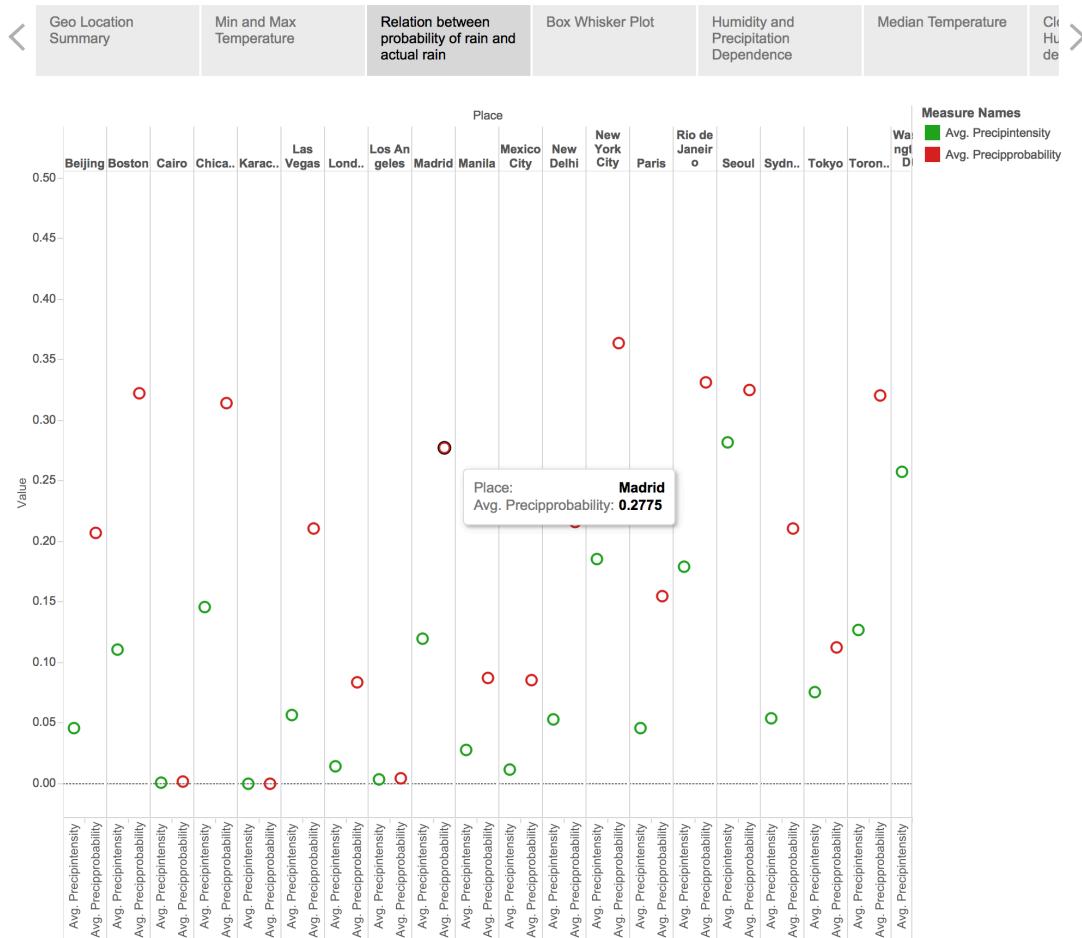
Cloud coverage v/s humidity analysis



Humidity v/s Precipitation plot



Statistics on probability of rain and actual rain



Plot on tableau forecast on Maximum Temperature and Minimum Temperature



Plot Simple Linear Regression Forecast analysis on Maximum Temperature, Minimum Temperature and Pressure



Conclusion

- We saw how to store streaming data into multiple sources for persistence and showcased it using a business intelligence tools.
- We compared the forecast value we are getting from the API with the value we are getting by linear regression of historical data.