Multimedia Big Data Computing for In-depth Event Analysis

Ruben Tous, Jordi Torres and Eduard Ayguadé
Barcelona Supercomputing Center (BSC)
Universitat Politècnica de Catalunya - BarcelonaTech (UPC)
Barcelona, Spain
rtous@ac.upc.edu, torres@ac.upc.edu, eduard.ayguade@bsc.es

***Abstract*—While the most part of ”big data” systems target text-based analytics, multimedia data, which makes up about 2/3 of internet traffic, provide unprecedented opportunities for understanding and responding to real world situations and challenges. Multimedia Big Data Computing is the new topic that focus on all aspects of distributed computing systems that enable massive scale image and video analytics. During the course of this paper we describe BPEM (Big Picture Event Monitor), a Multimedia Big Data Computing framework that operates over streams of digital photos generated by online communities, and enables monitoring the relationship between real world events and social media user reaction in real-time. As a case example, the paper examines publicly available social media data that relate to the Mobile World Congress 2014 that has been harvested and analyzed using the described system.**

***Keywords*-big data, multimedia, spark, movile world congress, barcelona, multimodal, image, analysis**

I. Introduction

While the most part of efforts are focusing on text-based big data analytics, multimedia data, which makes up about 2/3 of internet traffic [1], provide unprecedented opportunities for understanding and responding to real world situations and challenges. Most current big data systems are very restrictive as then cannot handle data types other than text or numbers. The challenge with multimedia big data, is that images and audio analysis require much more sophisticated algorithms and much more computing resources than any other kind of structured or unstructured data.

This paper presents BPEM (Big Picture Event Monitor), a Multimedia Big Data Computing platform that operates over freely available online images from sources such as Instagram or Twitter. The platform includes tools for the continuous analysis of the harvested images, generating a range of social and behavioral metrics in real-time. The system takes profit of the changing nature of the data and their visual content, thus allowing to infer knowledge beyond the capabilities of batch-based systems and text-only analytics. The system is able to process an incoming bandwidth of gigabit/s scale, as it works over a continuous stream of hundreds of digital photos per second. Meeting this goal requires the execution of high performance image analysis and pattern matching algorithms over a distributed and scalable stream processing framework.

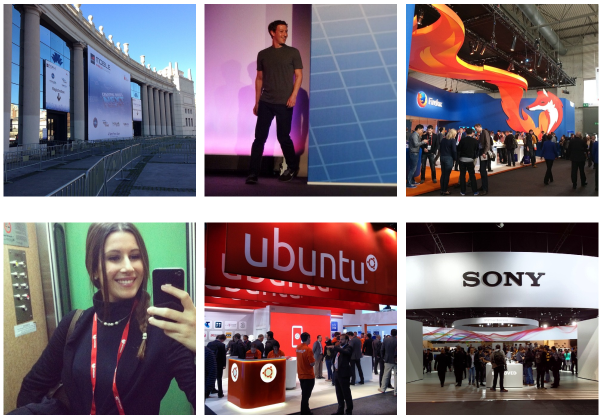

Figure 1. Example Instagram photos from 2014 Mobile World Congress in Barcelona

As a case example, the paper describes how the system was used to analyze photos uploaded to Instagram during the Mobile World Congress 2014, the world’s largest exhibition for the mobile industry that took place from February 24-27 in Barcelona, Spain. Figure 1 shows some example photos from the event.

II. A workflow methodology for understanding real world events through image analysis

The system described in this paper implements the conceptual scheme depicted in Figure 2. The workflow of the system comprises 4 main stages: *data acquisition*, *stream processing*, *storage* and *analytics/visualization*. The following subsections describe these stages in detail.

A. Data acquisition

The *data acquisition* stage consists on capturing (the descriptors of) new images related to an event as they are published on the underlying sources (e.g. Instagram, Twitter). Descriptors of the images (including the URL pointing to the image content) are acquired using the APIs provided by these underlying sources. These APIs impose limits over the amount of images that can be obtained during a certain period of time. So, processing the entire stream of images produced by a given API is not possible. APIs

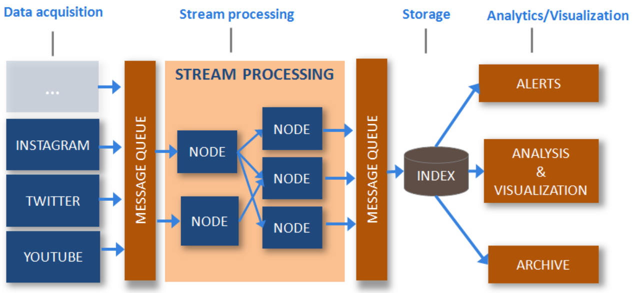

Figure 2. System conceptual scheme

provide the possibility to subscribe to certain filters, such as tags or geolocation bounding boxes. These filters produce partial streams that may be overlapped. In order to capture images related to an event, our system first needs information about the time intervals and geographical areas of the event, plus a list of tags that the event attendees will presumably use. These data is used by the system to program a set of subscriptions to the underlying sources. At the proper time, each subscription will produce a continuous stream of images that we call ”channel”. Because of nature of the events the BPEM platform targets, the throughput of the channels may be extremely volatile, requiring a proper scalability strategy. Besides the images taken at the event, we are also interested in (1) the images that the event attendees take at locations near the event venue (e.g. host city landmarks), to infer interests and behavioral patterns related to these locations, and (2) a sample of arbitrary images authored by the event attendees (to infer user demographic profiles and other latent attributes). We start intercepting these added data once we detect a new attendee.

In order to minimize data traffic among nodes, and to locate computation near the needed data, the images are not downloaded duting this stage. Instead, only the metadata is obtained, and passed to the next stage. In order to pass the metadata record to the following processing stages in a scalable and reliable manner, we employ a distributed and persisted messaging queue. We are currently implementing this functionality with Apache Kafka [2]. Kafka is an open-source message broker that provides a platform for handling real-time data feeds. Data streams are partitioned and spread over a cluster of machines to allow data streams larger than the capability of any single machine. Messages are persisted on disk and replicated within the cluster to prevent data loss.

B. Image stream processing

A multimedia big data computing framework should be able to process a continuous stream of near one thousand digital images per second, so it must process each image in soft real-time; otherwise the content should be dropped. This requirement can be satisfied through a stream processing framework with a distributed and scalable architecture. As we wanted an independent and open-source component for our software stack, we selected Apache Spark for this layer. Spark is an all-in-one distributed computation framework that enables both stream and batch processing. In addition, Spark provides integrated libraries for machine learning (MLlib) and graphs computation (GraphX), both designed for scalable data parallelism. As stated in the previous subsection, in order to decouple the stream processor from the input and output channels we employ a distributed and persisted message queue. Image descriptors are consumed from the queue during the *stream processing* stage, in which each descriptor is pulled by one node of the cluster. Is at this point when the image content is downloaded from the underlying service. Once downloaded in one node of the cluster, the same node will proceed to the multi-modal data enrichment process.

The differential aspect of a Multimedia Big Data Computing platform is the involvement of all data kinds (structured, unstructured text, visual and audio content) during the processing and analysis stages. A characteristic functionality of such systems is the automatic and transparent annotation of the incoming images with semantic information inferred from their audiovisual contents. This process, that we call *multi-modal data enrichment*, takes place during the *stream processing* stage, and it is analogous to text-based data enrichment processes such as sentiment analysis or other natural language processing (NLP) related techniques.

C. Indexing and storage

The descriptors, containing both high-level and low-level metadata generated during the processing stage, need to be temporally stored. On the one hand, these descriptors are later necessary for certain tasks such as duplicate detection. On the other hand, they enable the system to respond also to non-continuous queries performed by users. Thus, the target system has to perform asynchronous writes to a storage layer. In our case, this layer is composed of two different components: (1) A locality-sensitive hashing (LSH) index for replica-detection and (2), a Couchbase Server [3] for metadata indexing. On the one hand, the LSH index, which uses a hashing mechanism such that neighboring data samples have similar hash codes, provides us an efficient mechanism for approximate nearest neighbor search. The LSH index is used to detect near-replicas of images, which may be useful, e.g., for detecting trending images or images infringing copyright. On the other hand, Couchbase Server is an open-source distributed (shared-nothing architecture) NoSQL document-oriented database. We employ Couchbase Server to store the JSON metadata descriptors with administrative and other high-level data about the images.

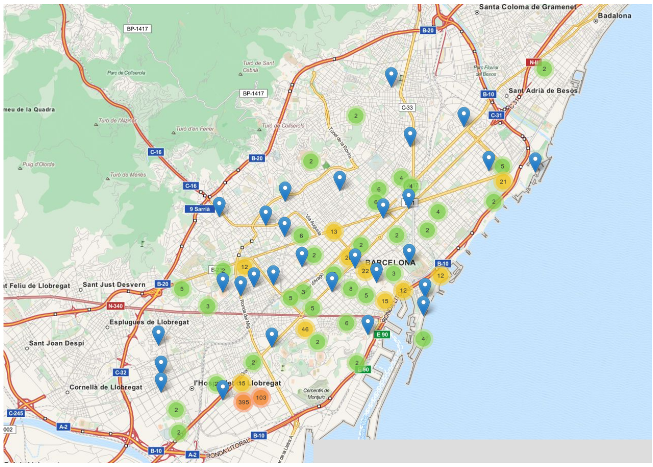

Figure 3. Geolocated Instagram photos in Barcelona during the MWC 2014

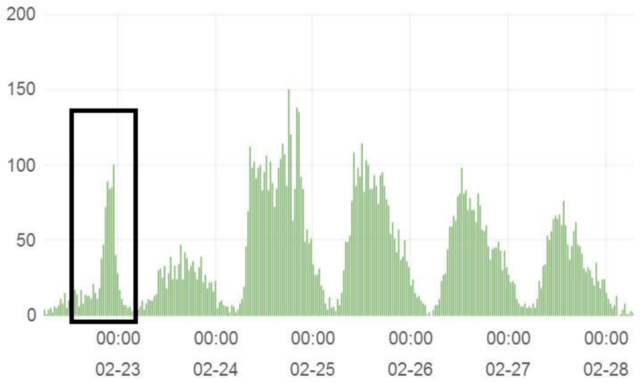

Figure 4. Histograms of images uploaded to Instagram during the 2014 Mobile World Congress (including the tag ”zuckerberg” as a capturing filter)

D. Analytics and visualization

To provide a powerful analytics and visualization interface BPEM relies on the third party dashboard Kibana [4]. Kibana is a web-based interface that enables real time querying and visualization of a JSON-based dataset. As kibana depends on the Elasticsearch [4] search server, we currently are indexing the JSON descriptors also there. BPEM enables querying the latest photos from a certain event using the Lucene query language [5]. Query conditions can address metadata fields inferred during the *multi-modal data enrichment* phase, thus allowing, for instance, to search for photos showing a certain logo or a certain number of faces. Query results can be visualized with histograms, pie charts and geomaps (see Figure 3).

III. A CASE EXAMPLE: 2014 MOBILE WORLD CONGRESS

The GSMA Mobile World Congress is the world's largest exhibition for the mobile industry. The 2014 edition took place from February 24-27 at the Fira Gran Via venue in Barcelona, Spain and was attended by more than 85,000 visitors from 201 countries. 1,800 exhibiting companies showcased their products and services across 98,000 net square meters of exhibition and hospitality space. Economic analysis has indicated that the 2014 Mobile World Congress contributed more than 356 million euros and 7,220 part-time jobs to the local economy [6].

A. Basic statistics based on metadata

We captured and analyzed 22,311 Instagram images during the 2014 Mobile World Congress. This amount, that is certainly low, was not due the contractual limitations imposed by the sources but by the low proportion of geotagged content and the disordered usage of tags. 15,621 of the captured photos (70%) were geotagged. 12,149 photos (54.45%) were tagged with one of the filtering tags, from which 5,459 were also geotagged. So, 10,162 photos (45.55%) were captured only because of the geotagging, 6,690 photos (29.98%) were captured only because of the tags and 5,459 photos (24.47%) were captured through both kind of filters. We don't know the total number of photos taken during the congress, not even the ones uploaded to Instagram, as many people do not use the tags recommended by the organizers or simply do not use any tag. Besides, many people do not has geotagging enabled. According to [7] only 5% of photos are uploaded to Instagram (7% according to [8]), and less than 30% of Instagram photos are geotagged. Performing a Fermi estimate, this means that we may have captured about the 40% of the Instagram photos and about the 2% of all the photos taken during the congress. Figure 4 shows the temporal distribution of the images. For the four days in which the event took place the histogram reflects the congress agenda. Every day activity started at 09:00, with one keynote. February 24 shows a particular pattern not only because it was the first day but also because Mark Zuckerberg, one of the co-founders of the popular social networking site Facebook, gave a keynote speech at 18:00 (it was the only day with a keynote in the afternoon).

The data analysis stage may be hampered by a wrong definition of the data acquisition filters, and so the acquisition strategy has to be carefully designed. Figure 4 shows an unexpected peak of image uploads during the afternoon of the 22th of February. This was unexpected because the Mobile World Congress took place between the 24th and 27th. The problem was that we included the tag ”zuckerberg” among the ones to capture at the data acquisition stage. As Zuckerberg was one of the keynote speakers at the congress we wanted to capture photos tagged with his name. However, on the afternoon of February 22, the messaging app Whatsapp experienced and outage of more than three hours, only a few days after Facebook officially announced the acquisition of the company. That situation sparked criticism on social networks and Instagram was flooded with critical and parodic images.

B. *Multi-modal analysis*

Besides some basic statistics about the event, the multi-modal analysis of the images can provide insights beyond possibilities of text-only analytics. As we not only capture photos taken during the event, but also photos related to the host city and a sample of arbitrary images authored by the attendees, we are able to infer and visualize knowledge in different scopes:

1) Knowledge about the event
2) Knowledge about interests and behavioral patterns related to the host city
3) Knowledge about the attendees

The first scope, the knowledge about the event, is the one that we can infer from the photos acquired by geolocation or by the event's tags. These photos tell us information about the attendees behavioral patterns and opinions related to different aspects of the event. On the one hand, with simple queries over the basic metadata we can infer the spatio-temporal habits of the attendees and we can also detect spatio-temporal trends in real-time. Additionally, we perform certain machine vision tasks such as the detection of the logos of the exhibiting companies (see Figure 1 for some examples).

The second scope, the knowledge about the host city, is the one that we can infer from the photos not related to the event but authored by individuals identified as event's attendees during the event duration. This knowledge may be used by various stakeholders in the field of tourism, e.g. local government officials, to analyze the impact of the event and to support policy and strategic decisions.

The third scope, the knowledge about the attendees, is probably the most relevant one. We have a sample of arbitrary photos taken by an attendee (taken outside the spatiotemporal context of the event). These data can reveal user interests, preferences and opinions, as well as trends and activity patterns.

IV. Conclusion

In this paper we have contributed a workflow methodology for understanding real world events by analyzing streams of digital photos generated by online communities. In addition to the problems related to the traditional streaming data workloads (e.g. tweets and logs), such as volume and burstiness, processing social photo streams introduce novel difficulties. On the one hand, each photo has to be analyzed with computer vision algorithms in order to enrich its metadata. These algorithms are highly demanding in terms of computational resources. In order to avoid load imbalance between cluster nodes a proper pipeline parallelism strategy is required. On the other hand, algorithms for search/indexing and data analysis, which involve multiple photos simultaneously (e.g. clustering), work over data records with a number of dimensions that exceed those of traditional analytics by several orders of magnitude, thus challenging current search/analysis technologies providing scalable data parallelism (e.g. Mahout and Spark's MLlib). As a proof of concept we have developed BPEM (Big Picture Event Monitor), a Multimedia Big Data Computing framework for event analysis. The system has been tested with Instagram photos related to the Mobile World Congress 2014.

Acknowledgment

This work is partially supported by the Spanish Ministry of Economy and Competitivity under contract TIN2012-34557, by the BSC-CNS Severo Ochoa program (SEV-2011-00067), and by the SGR programme (2014-SGR-1051) of the Catalan Government.

References

[1] J. R. Smith, "Riding the multimedia big data wave," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '13. New York, NY, USA: ACM, 2013, pp. 1–2. [Online]. Available: http://doi.acm.org/10.1145/2484028.2494492

[2] "Kafka - a high throughput message broker," http://kafka.apache.org/ (Accessed October 6, 2014).

[3] M. C. Brown, *Getting Started with Couchbase Server - Extreme Scalability at Your Fingertips.* O'Reilly, 2012.

[4] "Elasticsearch official website," http://www.elasticsearch.org/ (Accessed October 6, 2014).

[5] "Apache lucene official website," http://lucene.apache.org/ (Accessed October 6, 2014).

[6] "Gsma mobile world congress 2014 shatters previous records," http://www.gsma.com/newsroom/gsma-mobile-world-congress-2014-shatters-previous-records/ (Accessed October 6, 2014).

[7] "Charles arthur. instagram pictures reveal belfast as the uk's happiest city. the guardian. monday 13 january 2014." http://www.theguardian.com/technology/2014/jan/13/instagram-pictures-belfast-uk-happiest-city-jetpac/ (Accessed October 6, 2014).

[8] C. Smith, "Here's why instagram's demographics are so attractive to brands," http://www.businessinsider.com/instagram-demographics-2013-12/ (Accessed October 6, 2014).