

## Bohr: Similarity Aware Geo-distributed Data Analytic

**SPEAKER** Mr Hangyu Li

PhD Student  
Department of Computer Science  
City University of Hong Kong  
Hong Kong

**DATE** 31 May 2017 (Wednesday)

**TIME** 10:00 am - 10:30 am

**VENUE** G7315, 7th Floor  
Green Zone, Academic 1  
City University of Hong Kong  
83 Tat Chee Avenue  
Kowloon Tong

### ABSTRACT

In this talk we propose Bohr, a similarity aware geo-distributed data analytics system that minimizes query completion time. The key idea is to exploit similarity between data in different data centers (DCs), and transfer *similar* data from the bottleneck DC to other sites with more WAN bandwidth. Though these sites have more input data to process, these data are more similar and can be more efficiently aggregated by the combiner to reduce the intermediate data that needs to be shuffled across the WAN. Thus our similarity aware approach reduces the shuffle time and in turn the query completion time (QCT).

We design and implement Bohr based on OLAP data cubes to perform efficient similarity checking among datasets in different sites. Evaluation across ten sites of AWS EC2 shows that Bohr decreases the QCT by 30% compared to state-of-the-art solutions.

This paper will be presented at the 9th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud '17), July 10-11, Santa Clara, USA

Supervisors: Dr Sarana Nutanong and Dr Henry Xu

Research interests: WAN networks; big data analytics

**All are welcome!**



In case of questions, please contact Dr NUTANONG Sarana / Dr XU Hong Henry at Tel: 3442 7762 / 3442 4840, E-mail: [snutanon@cityu.edu.hk](mailto:snutanon@cityu.edu.hk) / [henry.xu@cityu.edu.hk](mailto:henry.xu@cityu.edu.hk), or visit the CS Departmental Seminar Web at <http://www.cs.cityu.edu.hk/news/seminars/seminars.html>.

