

Really Big Data: Analytics on Graphs with Trillions of Edges

Willy Zwaenepoel

*School of Computer and Communication Sciences
EPFL*

Amphi 25

4, place Jussieu
75005 Paris
Metro Jussieu

**22 Mars 2016
à 18h00**

Big graphs occur naturally in many applications, most obviously in social networks, but also in many other areas such as biology and forensics. Current approaches to processing large graphs use either supercomputers or very large clusters. In both cases the entire graph must reside in memory before it can be processed. We are pursuing an alternative approach, processing graphs from secondary storage. While this comes with a performance penalty, it makes analytics on very large graphs feasible on a small number of commodity machines. I will describe the challenges involved in doing so, and the Chaos system that we have built to meet these challenges. With Chaos we have been able to process an 8-trillion-edge graph on 32 machines, a new milestone for graph size on a small cluster.

This is joint work with Laurent Bindschaedler (EPFL), Jasmina Malicevic (EPFL) and Amitabha Roy (Intel Labs).

Willy Zwaenepoel received his BS/MS from the University of Gent, Belgium, and his PhD from Stanford University. He is currently a Professor of Computer Science at EPFL. His interests are in operating systems and distributed systems. He is a Fellow of the ACM and the IEEE, and has received several other awards for his research and teaching. He has also been involved in a number of startups.

