

## Cover Letter

A wise man once said: "Data is a precious thing and will last longer than the systems themselves". That man was Tim Berners-Lee.

Indeed, we are living in the world where Data is all that matters: how to acquire, store, transfer, transform and generate informative knowledge out of it. For this reason, and for many others, my career objective is to act as an expert in data-driven research and development. I am really excited to apply for this position in Data61 as my goals align perfectly with the vision of the company which is to create a positive data-driven future and I would like to be a part of it. After reviewing the job description and reading the company's science vision, I believe Data61 is a perfect fit for my aspirations and my skill sets.

My name is Rohit Kumar and I am a final year Ph.D. candidate doing a joint Ph.D. in temporal graph mining and distributed processing from ULB, Belgium and UPC, Spain. I will be finishing my Ph.D. in June 2018. My research focuses on designing efficient one pass algorithms to analysis large interaction networks such as social media network. I have also worked on Influence maximization and Information propagation problems using Interaction network. While visiting UPC Barcelona, I worked on distributed graph processing and large graph data management, where I focused on using cost model based approach for adaptive graph partitioning on Apache GraphX. Apart from my research, I also work part-time as big data consultant working on different big data platforms, and conduct online courses for Big data systems like Hadoop ecosystem, Spark, Machine Learning using Python and Deep learning for an Indian company.

I also have 6+ years of industry experience in leading a team and owning the design and development of a suite of products using Java/J2EE. I was a technical team lead and product owner and my responsibilities involved collaborating with multiple teams such as UI team for new UI designs and integration with the existing system, QA team for testing and bug fixes, Infrastructure team for timely deployment and daily build and DevOps related issues.

I also have a huge interest in Machine Learning and looking to work towards fields where my background in large data management and graph processing could help in designing and developing efficient machine learning models for real-world problems. I feel working in the Data61 lab will provide me the opportunity to work on real-world application problems where I can apply my knowledge and skill sets for the benefit of my ultimate goal, which is doing what it takes to improve myself and my surrounding.



# RohitKumar

## Big Data Scientist

### Address

Avenue General Medicine  
Derache 28  
Brussels 1050  
Belgium

### Phone

+32 465 931 350

### Mail

[rohit1308k@gmail.com](mailto:rohit1308k@gmail.com)

### Web

[rohit13k.github.io](https://rohit13k.github.io)

### Programming

Java	★★★★★
Scala	★★★★☆
C++	★★★★☆
Python	★★★★☆

### Databases

MySQL	★★★★★
MS SQL	★★★★☆
Oracle	★★★☆☆
Coherence	★★★★☆

### Big Data Systems

Spark	★★★★★
Hadoop	★★★★☆
Flink	★★★☆☆
Kafka	★★★★☆
NoSQL	★★★★☆

## Experience

Aug. 14 - Now

### PhD Researcher

[Université Libre de Bruxelles \(ULB\), Belgium](#)

Expected, a Ph.D. in "Graph Data stream mining and distributed processing" in June 2018 from ULB (Brussels) and UPC (Barcelona). The research focuses on designing efficient one pass algorithms to analysis large interaction networks such as social media network.

Currently, working on distributed graph processing on Apache Spark GraphX. Worked on other Distributed processing systems like Hadoop, MongoDB, Kafka, Hive and Cassandra. Worked on large graph querying and processing using Giraph, Apache GraphX and Neo4j.

Teaching/Lab assistant for data-warehouse course for one semester for a class of 20 master students.

Sep. 17 - Now

### Online Trainer (part time)

[Digital Vidya, India](#)

Designed and delivered online course on Big Data Engineering course (Hadoop, Spark, Hive, Pig, Cassandra, NoSQL) for Digital Vidya for a batch of 10 students.

Delivered online course on Data science with Python (NumPy, Pandas, scikit-learn) and Scala (Spark Mllib) for Digital Vidya online courses.

Sep. 16 - Oct. 16

### Data Engineer (part time)

[Disperse.io, UK](#)

Worked on analysing and discovering trends in Wi-Fi data for London metro stations. Worked as freelance data engineer in data cleaning and data analysis using Spark-SQL

Jun. 14 - Jul. 14

### Software Designer

[Royal bank of Scotland, Delhi, India](#)

Developed data pipelines with Java for the banking system using Oracle Coherence.

Apr. 11 - May. 14

### Assistant Consultant

[Tata Consultancy Services, Mumbai, India](#)

Technical team lead and Scrum Master for the entire online assessment platform, an online cloud based system for conducting online examinations and certifications.

Handling pre sales and new product offerings as a technical expert for the education domain. Developed and deployed the complete Eco-system of the online assessment platform in multi-tenant architecture for software as a service model.

Technologies used: J2EE, Java, MySQL, SVN, Struts 2.

Mar. 11 - Sep. 11

### Researcher

[Tata Research Development and Design Center, Pune, India](#)

Worked on data privacy and data security. Worked on data masking techniques for privacy preserving data mining.

Technologies used: C++, CVS

Nov. 07 - Mar. 11

### Assistant System Engineer

[Tata Consultancy Services, Chennai, India](#)

Lead architect to develop and deploy the online assessment platform for TCS hiring throughout India. Performance and Security Lead for the online assessment platform. Taught Algorithms and Java to new employees as a part of learning and development program.

Technologies used: Java, CVS, Haskel, Java Security library, MS SQL.

## Languages

English ★★★★★  
Hindi ★★★★★  
Spanish ★☆☆☆☆

## Personal Skills



Jun. 05 - Jun. 07 **Summer Research Intern** [JNC SAR, IISC, Bangalore, India](#)  
Did two summer(2 months) and one winter(1 month) research internship in Nanoscience. Worked with P.hD students on analyzing nano material properties under electron microscopes.

## Education

Aug. 14 - Now **Ph.D Candidate** [Université Libre de Bruxelles & Universitat Politècnica de Catalunya](#)  
Doing a joint PhD from ULB and UPC under joint supervision from Toon Calders and Alberto Abello, Funded by FNRS scholarship.

Aug. 08 - Apr. 11 **M.Sc in Computer Science** [Chennai Mathematical Institute, India](#)  
The Master course was done part time while working in TCS. Main subjects: Discrete Mathematics, Database Systems, Operating systems, Data mining, Cryptography, Statistics, Design and Analysis of Algorithm, Theory of Computation, Programming language Concepts. Grade: **8.44%**  
Thesis: Security aspects of online Assessment platform-Touchstone.

Jul. 04 - Jul. 07 **B.Sc (Hons) in Physics** [Hindu College, Delhi University](#)  
Did summer research internship at IISC Bangalore every summer break. Co-Organized Physics Department fest on 2006. Percentage: **76.1%**

## Awards and Achievements

Apr. 13 **Best Technical Lead Award in ACTION2013 in SMB iON** [TCS](#)  
Awarded best technical lead in the entire department of 500 developers and 25 teams.

Sep. 11 **Innovation of the year award southern region** [Tata Innovista](#)  
Tata group yearly award for innovations across the Tata groups companies.

## Publications

2017 **Activity-Driven Influence Maximization in Social Networks** [ECML/PKDD](#)  
**Rohit Kumar**, Muhammad Aamir Saleem, Toon Calders, Xike Xie and Torben Bach Pedersen

2017 **Cost Model for Pregel on GraphX** [ADBI](#)  
**Rohit Kumar**, Alberto Abello, and Toon Calders

2017 **Information Propagation in Interaction Networks.** [EDBT](#)  
**Rohit Kumar** and Toon Calders

2017 **Location Influence in Location-based Social Networks.** [WSDM](#)  
Muhammad Aamir Saleem, **Rohit Kumar**, Toon Calders, Xike Xie and Torben Bach Pedersen

2017 **Finding simple temporal cycles in an interaction network.** [TD-LSG](#)  
**Rohit Kumar** and Toon Calders

- 2017 **IMaxer: A Unified System for evaluating Influence Maximization Mechanisms in Location-based Social Networks** [CIKM](#)  
Muhammad Aamir Saleem, **Rohit Kumar**, Toon Calders, Xike Xie and Torben Bach Pedersen
- 2017 **Cost Model Based Approach for Graph Partitioning in Spark GraphX** [DBDD](#)  
**Rohit Kumar**, Alberto Abello, and Toon Calders
- 2015 **Maintaining sliding-window neighborhood profiles in interaction networks.** [ECML/PKDD](#)  
**Rohit Kumar**, Toon Calders, Aristides Gionis, and Nikolaj Tatti.

## Patents

- 2009 **System and method for automated competency assessment** [US8915744](#)  
Raman Srinivasan, Priyadharshini Sridhar, Swarna Srinivasan, **Rohit Kumar**, Radhika Jayapaul, Radhika Ganesan, Amit Nath, Vikash Agarwal
- 2013 **Customized question paper generation** [US20130084554A1](#)  
Viral Prakash SHAH, Nawaz Shaikh, **Rohit Kumar**
- 2014 **Secured Computer Based Assessment** [US20140186812A1](#)  
Viral Prakash SHAH, Nawaz Shaikh, **Rohit Kumar**

*Rohit Kumar*  
*February 27, 2018*