

# ROHIT KUMAR

Languages: English(native fluency), Hindi(native), Spanish(A1)

Personal website: <http://rohit13k.github.io/>

## RESEARCH INTEREST

---

Big Data platforms, Data Stream Mining, Graph Stream Mining and Distributed Processing, Data management, Data mining and Machine Learning, Data Security and Privacy.

## SUMMARY

---

Dr Kumar has a diverse background with graduation in Physics and research training in Nanoscience followed by masters in Computer Science and a PhD in Big Data. After doing a successful startup derived from his master thesis and working in Indian IT company for 6.5 years, he moved to Europe to pursue a joint Ph.D. in “Temporal Graph Mining and Distributed Processing” from ULB (Brussels) and UPC (Barcelona). He finished his PhD in June 2018 and was granted excellent cum Lauda grade for the same. Currently, he is working as a Big data architect and engineer for multiple H2020 and Spanish government projects in Eurecat.

## TEACHING EXPERIENCE

---

### Barcelona Technical School

2019 - now

*Masters in Big Data*

*Barcelona, Spain*

- Teaching a course on Big data infrastructure to a batch of 15 students in the Big Data masters course.

### University of Barcelona

2019 - now

*Masters in Big Data*

*Barcelona, Spain*

- Teaching a course on Spark and Big data technology to a batch of Data Science students in a Data Science masters course.

### Digital Vidya, India

2017 - 2018

*Online- Course Instructor*

*India*

- Teaching a course online to a batch of 15 students in Data analytics with Python. Already completed 2 batches.
- Designing a new course curriculum and course material for Big data analytics with Hadoop and Spark.

### ULB, Brussels

2015 - 2016

*Teaching Assistant*

*Brussels, Belgium*

- Teaching Assistant for Data warehouse labs for IT4BI masters course for a batch of 30 students.

### TCS - iGnite

Feb 2008 - July 2010

*Teaching Assistant*

*Chennai, India*

- Teaching Assistant for Java and Fundamental of Programming in TCS iGnite Training Center for three batch of 500 trainees.
- Exam co-coordinator for online Mooshak exams.
- Guided two batch of 24 students for a period of 4-6 months.

## INDUSTRY WORK EXPERIENCE

---

**Eurecat***Researcher*

July 2018 - now

*Barcelona, Spain*

Working on multiple Big data projects as Big data architect and Big data engineer.

- Big Data Architect: Designed the big data platform for a supermarket chain in Spain to help them migrate from their traditional in house data ware system to an on cloud data lake for big data analytical use cases.
- Big Data Architect: Designed and helped in the development of a new platform for an advertising company in Madrid for storing user web clickstream data and providing recommendation and predictions using Spark.
- Big Data Engineer: Developed a plugin in ElasticSearch for data clustering to be used in an EU search platform.
- Big Data Engineer: Developed and helped redesign the Data visualization platform BCNNow as part of DECODE project.
- Researcher: Working on Data privacy tool for an EU Project SMOOTH to help Micro and Small enterprises with GDPR compliance.
- Research Engineer: Worked on developing a large social graph mining engine Kalium using OrientDB and Spark.
- Big Data Engineer: Maintain and support the inhouse cloud for Eurecat using Red Hat Openstack Platform.

**Upwork***Freelance Big Data Engineer*

Aug 2014 - June 2018

*Brussels, Belgium*

- Spark NLP: Core developer to develop and support the Spark NLP library.
- Disperse.io: Worked on analyzing and discovering trends in Wi-Fi data for London metro stations. Worked as a data engineer in data cleaning and data analysis using Spark-SQL.
- Discovery: Worked on smart meter applications. Developed a machine learning based platform on top of Spark streaming to collect, store and analysis data from smart meters.
- L3Networks: Designed and implemented a processing system that could process web logs arriving at the rate of 46GB/min in real-time. Designed the architecture and the data processing pipeline using Apache Kafka, Apache Spark and Apache Hadoop using Java and Scala Languages. The work included deployment of these technologies as well.
- DigitalVidya: Designed and delivered online course on Big Data Engineering course (Hadoop, Spark, Hive, Pig, Cassandra, NoSQL) for DigitalVidya for 3 batches of 10 students. Delivered two courses on Data Science with Python (NumPy, Pandas, scikit-learn) and R.

**Royal bank of Scotland***Software Designer*

June 2014 - July 2014

*New Delhi, India*

- Developing high availability query systems on Oracle Coherence.

**Tata Consultancy Services***Assistant Consultant*

April 2013 - May 2014

*Mumbai, India*

- Technical team lead for the development of TCS-iON Online Assessment platform. The platform is used in many high stack all India exams like JEE, Banking recruitment and government recruitment exams.
- Managed a team of 30 software developers.

**Tata Consultancy Services***IT Analyst*

April 2011 - March 2013

*Mumbai, India*

- Developed and deployed the complete Eco-system of the online assessment platform in multi-tenant architecture for software as a service model.

**Tata Consultancy Services**

Nov 2007 - March 2011

*Assistant System Engineer**Chennai, India*

- Lead architect to develop and deploy the massive online assessment platform for TCS hiring throughout India.
- Performance and Security Lead for the online assessment platform.

**TECHNICAL SKILLS**

---

<b>Programming</b>	Java, Scala, C++, R-basics, Python
<b>Databases</b>	MySQL, MS SQL, MongoDB, Oracle Coherence
<b>Tools</b>	SVN, CVS, github, Eclipse, RStudio, Octave, Matlab, gnuplot
<b>Others</b>	Hadoop, Apache Spark, Apache Storm, Kafka, Struts 2, JDBC

**EDUCATION**

---

**Université Libre de Bruxelles (ULB), Belgium & Universitat Politècnica de Catalunya (UPC), Spain** August 2014 - June 2018

- Joint PhD in Computer Science

**Chennai Mathematical Institute, India**

April 2008 - March 2011

- M.Sc in Computer Science

**Hindu College Delhi University, India**

July 2004 - July 2007

- B.Sc(Hons) in Physics

**RESEARCH EXPERIENCE**

---

**ULB and UPC**

Aug 2014 - June 2018

*PhD**Brussels, Belgium and Barcelona, Spain*

- Ph.D. in “Temporal Graph mining and distributed processing” in June 2018 from ULB (Brussels) and UPC (Barcelona).
- Worked on social network analysis and viral marketing.
- Worked on large scale graph processing and data management.

**TCS - Tata Research Development and Design Center**

April 2011 - September 2011

*Researcher**Pune, India*

- Worked on data privacy, data security, and on developing data masking Tools in C++.

**JNCSAR, IISC, Bangalore**

2004 - 2007

*Summer Research Intern**Bangalore, India*

- Did two summer and one winter training in Nanoscience and Nanomaterials. Worked on synthesis and characterization of Carbon and Zinc Oxide nanotubes.

**PUBLICATIONS**

---

- 2SCENT: An Efficient Algorithm to Enumerate All Simple Temporal Cycles. **Rohit Kumar** and Toon Calders. **VLDB 2018**.
- Towards Location Influence in Location-Based Social Networks. Muhammad Aamir Saleem, **Rohit Kumar**, Toon Calders, Xike Xie and Torben Bach Pedersen. **KAIS 2018**
- Cost Model Based Approach for Graph Partitioning in Spark GraphX. **Rohit Kumar**, Alberto Abello, and Toon Calders. **DBDBD 2017**.

- IMaxer: A Unified System for evaluating Influence Maximization Mechanisms in Location-based Social Networks. Muhammad Aamir Saleem, **Rohit Kumar**, Toon Calders, Xike Xie and Torben Bach Pedersen. **CIKM 2017** (DEMO Paper)
- Finding simple temporal cycles in an interaction network. **Rohit Kumar** and Toon Calders. **ECML/PKDD 2017** (Workshop paper)
- Activity-Driven Influence Maximization in Social Networks. **Rohit Kumar**, Muhammad Aamir Saleem, Toon Calders, Xike Xie and Torben Bach Pedersen. **ECML/PKDD 2017**
- Cost Model for Pregel on GraphX. **Rohit Kumar**, Alberto Abello, and Toon Calders. **ADBIS 2017**
- Information Propagation in Interaction Networks. **Rohit Kumar** and Toon Calders. **EDBT 2017**
- Location Influence in Location-based Social Networks. Muhammad Aamir Saleem, **Rohit Kumar**, Toon Calders, Xike Xie and Torben Bach Pedersen. **WSDM 2017**
- Maintaining Sliding-Window Neighborhood Profiles in Interaction Networks. **Rohit Kumar**, Toon Calders, Aristides Gionis, and Nikolaj Tatti. **ECML/PKDD 2015**

## PATENTS

---

- System and method for automated competency assessment (US8915744) - **Granted**
- Secured Computer Based Assessment (US20140186812A1) - Applied
- Customized question paper generation (US20130084554) - Applied

## AWARDS AND ACHIEVEMENTS

---

- Big Data Talent Award 2018 given in Big Data Coe Barcelona for the Doctoral Thesis.
- FNRS-Belgian National Funds for Scientific research, Scholarship for Doctoral Research Study.
- Best Technical Lead Award in ACTion2013 in SMB iON, TCS.
- Special initiative Award by TCS in May 2013.
- Technical Excellence Award by TCS in March 2012.
- Innovation of the year award southern region in TATA Innovista 2011.
- Senior Diploma in Arts and Painting from Ankan Kala Bibhag, Ravindra Bharti University in 2006.