人大-华师大-InfoSys数据科学联合实验室招聘实习生

由中国人民大学、华东师范大学和印度软件业巨头InfoSys公司联合成立的数据科学联合实验室（简称InfoSys联合实验室）旨在从事大数据相关领域的研究。现面向信息学院二三年级本科生、硕士生和博士生公开招聘实习生若干名，于今年暑期去印度InfoSys总部班加罗尔实习2-3个月，实习内容主要是参与联合实验室的大数据系统研发项目。要求实习学生具有较好的编程基础（熟练使用C和Java等）、较好的英语交流能力，并且能够成为InfoSys联合实验室成员、长期参与联合实验室的大数据系统研发项目。实习期间InfoSys将提供相关的往返旅费和生活费，实习过后参与联合实验室项目也有补助提供。请感兴趣的同学于本月28日前填写附件的表格，并发送到陈跃国老师的电子邮箱（chenyueguo@ruc.edu.cn）。根据申请情况，联合实验室将于5月4日前后（具体时间地点另行通知）对符合基本要求的同学组织面试，确定候选名单，5月10日前推荐给InfoSys公司，并由InfoSys公司确定最终的实习人员名单。具体实习时间预计为6月底至9月初。以下是联合实验室今年设立的两个项目信息：

项目一：日志大数据实时分析系统

（负责人：陈跃国副教授，chenyueguo@ruc.edu.cn）

用户建模对于电信、电商等企业非常重要。它一般是在用户日志等大数据基础上分析得到的用户在众多维度上的属性。这些大规模用户的属性需要被很多应用频繁的更新和访问，以发挥出大数据最精华的价值。这样的应用在现有解决方案下被分为两类独立的大数据系统：数据服务系统与数据分析系统，由于系统的分离难以做到数据的实时分析和利用。在这项研究中，我们试图在电信、电商等大数据实时分析处理的需求背景下，将具有数据服务和数据分析功能的两类开源系统有机的融合起来。系统以HDFS上的列存储技术为基础组织数据，同时融合先进的MPP数据库集群技术，构建高性能、高可扩展性、低成本的日志大数据实时分析处理系统。

项目二：Complex query processing and optimization on big data

（面向大数据的复杂查询和优化技术研究，负责人：陆嘉恒教授，jiahenglu@ruc.edu.cn）

Spark and Spark SQL are emerging as a leading open source portfolio to handle Big Data analytics. However, Spark from the open source community holds a high learning curve to IT professionals, especially on system performance management to better utilize the system resources. The parameter configuration in Spark requires the understanding of the characteristics of the job, data and system resources, which is beyond the knowledge of traditional enterprise IT people. Another interesting scenario about Spark job optimization comes from analytic services (e.g. Elastic Spark) on the cloud. The users, such as data scientists, do not know how to correctly choose the Spark parameters to accelerate the job execution. Therefore, motivated by above scenarios, this project addresses the challenge to job optimization for complex query processing bases on Spark and Spark SQL.

InfoSys数据科学联合实验室实习生报名表格

（请于4月28日前发至chenyueguo@ruc.edu.cn）

|  |  |
| --- | --- |
| 姓名及学号 |  |
| 学生类型及年级 |  |
| 身体健康状况 |  |
| 英语能力说明（例如：4、6级、托福、雅思等成绩） |  |
| 学分绩（本科所有课程，本科非本校的学生请注明学校） |  |
| 程序设计、数据结构、算法类课程成绩 |  |
| 感兴趣的项目 | [ ]项目一 [ ] 项目二 |
| 是否愿意长期参与InfoSys联合实验室大数据系统研发项目 |  |
| 项目开发经历 |  |