Faculty of Science and Engineering

Profile report: Computational Infrastructure for Big Data Processing / Distributed Systems

- Discipline: Computer Science
- Level: Tenure-track Assistant professor with an education profile
- Fte: Full time (0.8-1.0 FTE)

1. Scientific discipline

The domain of Distributed Systems covers various aspects of coordinating computer network-based applications that aim to achieve a common goal. The complexity of distributed systems comes from different sources: failure of components, lack of global trust, overhead of network communication, amount of data transferred between individual components, heterogeneity of the environment, etc. Since the emergence of the Internet and the Web, the majority of applications have inherent requirements of being distributed in order to be able to continuously provide their services to the users. New research subfields are forming due to the continuously growing amount of data that can only be stored and processed in a distributed fashion.

2. Vacancy

This position is opened by the Board of the Faculty of Science and Engineering (ref: PT/gl/22/00181) and will be embedded in the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, basic unit Distributed Systems. The criteria and conditions pertaining to the position are described in the document 'Assistant professor with an education profile'.

3. Selection committee (BAC)

Prof. dr. A. Lazovik (chair)	Programme and Education Director for Computing Science
Prof. dr. N. Taatgen	Professor of Artificial Intelligence, Bernoulli Institute Research Director
Prof. dr. D. Karastoyanova	Professor of Information Systems
Dr. F. Cnossen	Education Director for Artificial Intelligence
Dr. A. Rastogi	Assistant Professor in Software Engineering
Floris Westerman	Student member
Prof. dr. M. Aiello	External member, Professor of Distributed Systems

HR advisor:
M. Laning, MSc

Advisor:
A.G. Gringhuis, MSc   Policy Officer Bernoulli Institute

4. Area of expertise

Recent prominent achievements in the area of data science would not be possible without our improved understanding on how to build distributed supporting infrastructures that let us process large amounts of data. Without knowing how to collect, store and process the data in a distributed fashion, many new algorithms that inherently require large amounts of data would not be possible to realize. This new field is often referred to as *"Big Data engineering"*. However, our knowledge on how to build such systems is still limited: the majority of such systems (while being distributed) are still controlled in a centralized fashion and put strict requirements on the availability of data, the location of processing nodes, the network latencies, etc. In practice, many existing

assumptions taken for granted do not always hold. For example, there might be legacy data spread across several organisations or departments within the same organisation forming independent "silos" that would require increased location-awareness for data processing.

Overall, we expect a focus on both theoretical algorithms and practical applications that would aim at solving problems in domains that would require novel distributed infrastructures involving processing of large amounts of data. Domains of particular interest are: Internet-of-Things and pervasive computing; blockchain, smart contracts and their applications; Big Data processing; SOA and cloud computing; energy and water management, etc.

This education tenure track position aims at coordinating and synchronizing the educational activities in the fields of Big Data, data engineering, high performance and high-throughput systems. The Distributed Systems group will provide a natural research embedding for a teaching-oriented tenure track position with a Computational Infrastructure for Big Data Processing profile.

5. Embedding: institute (and base unit)

The position will be embedded in the research group Distributed Systems. The group performs fundamental research, delivers education at the frontier of the state of the art in dynamic complex distributed systems using formal engineering tools, and seeks applications with societal impact. The main research interests of the group are in the areas of AI planning, discrete optimization, large-scale distributed data processing in highly distributed environments, having as main application domains Internet-of-Things, building automation, large-scale data analytics, business processes and energy distributed infrastructures. The research results have been field-tested in collaboration with industry. One of such applications eventually led to founding the Sustainable Buildings company that applies the optimization algorithms in practice.

The Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence is part of the Faculty of Science and Engineering (FSE). The profile of the institute centers around modelling, computation, and cognition with a focus on science and technology, keeping a balanced mix of fundamental and applied aspects. The Bernoulli Institute comprises five mathematics research groups, seven computer science groups, and four groups in the field of artificial intelligence. The constituting research groups participate in various national research schools and most of the PhD students are enrolled in an educational programme and take part in other activities offered by these schools. The Bernoulli Institute aims to strengthen the current research portfolio in Mathematics, Computer Science and Artificial Intelligence by expanding both in fundamental areas that have a prominent role in education as well as in directions that are essential for new technological and societal developments.

The Bernoulli Institute has a leading role in the recently established cross-disciplinary research theme on Data Science and Systems Complexity (DSSC) within the Faculty of Mathematics and Natural Sciences. This concerns a research cluster of 60+ researchers in a number of basic disciplines (mathematics, computer science, artificial intelligence, systems & control, engineering, astronomy) and various scientific application domains. The ambition is to understand and solve Big Data problems by exploiting the joint perspectives from both data science and complexity science. The institute is also heavily involved in the Groningen Cognitive Systems and Materials Center (CogniGron), which is a joint venture between the Bernoulli Institute and the Zernike Institute for Advanced Materials. It comprises researchers from materials science, physics, chemistry, mathematics, computer science and artificial intelligence. The center provides structure, coherence, and visibility for a joint research program in the direction of cognitive systems and materials. Healthy aging and energy are two other important university-wide research topics that the Bernoulli Institute is involved in.

6. Local and (inter)national position

Nationally, most universities have strong research efforts in distributed systems. In particular, there are strong groups at the VU University, TU Eindhoven, and University of Twente. The Distributed Systems group in Groningen is among the leaders in the field of Smart IoT and the AI methods supporting automated reasoning in distributed smart environments, e.g., planning and scheduling, discrete optimization, machine learning.

At the national level the group participates in the Dutch Research School in Programming and Algorithmics (IPA), the Advanced School for Computing and Imaging (ASCI), and the School for Information and Knowledge Systems (SIKS). The research group has established close collaboration and participated in joint projects with national research institutes (TNO, CWI, and Astron), and various Dutch companies (Philips, DNV GL, Enexis, Alliander, Vitens, Anchormen, Nerdalize). The group leader, Lazovik, was a co-founder of the Sustainable Building start-up company, whose business proposition was based on the research of the Distributed Systems group. At the international level the research group has been involved in several EU-funded research projects (e.g., Smart Homes, GreenerBuildings, Virtual Factories), has established collaborations with major international companies (IBM, Shell, Cognizant) and has cooperation and exchange programmes with many universities (e.g., Rome, Stuttgart, Vienna, Modena, Bournemouth).

In Computer Science, the Bernoulli Institute has a strong (inter)national position (as evidenced by participation in NWO and EU projects, publications in renowned journals and conferences, memberships of editorial boards and program committees, conference chairing, etc.). Its expertise lies in intelligent systems (biologically inspired computational modelling, machine learning, morphological image processing); pervasive middleware and energy distribution infrastructures; architecting of software-intensive systems and object-oriented software design; adaptive information systems, middleware, enterprise, services and cloud computing and autonomous process performance improvement; computer architecture, microarchitecture and reconfigurable computing; data and information visualization, and visual analytics.

7. Expected contributions to teaching

The successful candidate is expected to contribute to the teaching programmes of the bachelor and master programmes of Computer Science in the Undergraduate and Graduate Schools of Science of the FSE. He/she is expected to take the roles as courses coordinator, lecturer, and examiner. He/she will contribute both to teaching existing courses, and to the development of new courses in the area of distributed systems. This includes the supervision of bachelor and master theses. The teaching tasks amount to at most 60% of the total time for a tenure-track assistant professor with an education profile. To foster educational developments, the candidate will apply for grants that support teaching innovation.

In order to coordinate and synchronize the educational activities in the fields of Big Data, data engineering, high performance and high-throughput systems, the main task of the successful candidate is to play a leading role in contributing to education in the areas of Big Data and supporting computational infrastructures for large-scale data processing. The candidate will mainly contribute to the Computing Science programmes, but education activities should also accommodate students from other disciplines, e.g., students with an AI background, potentially extending to service education in the various topics of Big Data processing for the rest of the faculty.

In order to facilitate the connection between high school students and the Computing Science programmes, the candidate is expected to organise activities to raise the awareness of prospective students from high schools, e.g., by engaging in dedicated educational activities at the high school level and the organisation of introductory events for high school students.

8. Expected contributions to research

The candidate is expected to contribute to the ongoing research activities of the Distributed Systems research group with focus on computational infrastructure for Big Data processing. There is room for various research directions, and the specific direction will depend on the profile of the candidate. The co-supervision of PhD students will be a concrete mechanism for contributing to the group's research efforts. Additionally, the candidate will carry out research in the field of education, and develop and adopt novel didactic methods used in teaching of Big Data processing.

9. Expected contributions to the organisation

The candidate is expected to have an active interest and to provide a positive contribution to the management and organisational tasks of the institute. At the level of FSE, the candidate will contribute to the organisation of the faculty, for example by participating in working groups and committees in the area of education. The candidate will participate in relevant national and international organisations.