

### **Project 1: IT Operations Analytics**

#### **Project Description:**

IT Operation analytics aims to mine the actionable insights with Big Data and advanced analytics techniques from the huge volume machine generated data including logs, events, configurations, metrics, system changes and etc. within highly distributed environments (e.g. Cloud, internet/enterprise applications) in order to facilitate the faster troubleshooting and proactive anomaly detection towards an intelligent IT operations. More specifically, we expect an intern to:

- Develop innovative model and algorithms for anomaly detection & root cause analysis by correlating the multi-type of operational data
- Generate the anomaly patterns and remediation actions with machine learning techniques

#### **Desired Skill:**

Ph.D. or Master Students majored in computer science or related areas with the following skills or experience:

- Distributed systems and cloud computing
- Big data analytics with Hadoop, Spark, etc
- Data mining and machine learning
- Hands-on experience on building software systems
- Creative, independent and Self-motivated
- Team collaboration

### **Project2: Transformation between Scripts and Domain Specific Language**

#### **Project Description:**

In enterprise's journey to embrace cloud, some modern frameworks such as Chef and Puppet are introduced to gain efficiency and reduce cost. However, there are many legacy scripts which have been developed to operate and control the existing infrastructure, middleware, applications, so how to leverage, reuse and manage the legacy scripts in the modern infrastructure framework is very important. More specifically, we expect an intern to:

- Design and develop an innovative methods and algorithms to transform a legacy bash scripts to some Chef recipe or Puppet module while keeping the idempotency of the recipes and modules.
- Abstracting the general methods and algorithms to do transformation between different domain specific languages (DSL).

#### **Desired Skill:**

Ph.D. or Master Students majored in computer science or related areas with the following skills or experience:

- Cloud computing
- Compiling Principle and Model Transformation
- Hands-on experience on bash scripts, Linux.
- Creative, independent and Self-motivated
- Team collaboration

### **Project 3: Cognitive Internet of Things(IoT) Analytics Platform Services**

The Internet of Things (IoT) research of IBM Research-China (CRL) is leading to develop world class technology and cognitive solution research innovations enabling the enterprises

across industries to accelerate the transformation. In this project, we will work together to develop Cloud based cognitive IoT analytics platform services processing and analyzing the physical sensors data, especially unstructured data such as acoustic, video, text, leveraging advanced analytics technologies such as machine learning, graph theory to discover business insights enabling industry solution innovations, especially unstructured data such as acoustic, video etc.

**Desired Skill:**

Ph.D. or Master Students majored in Computer Science, Electronic Engineering, Industrial Automation or related areas with the following skills or experience:

- ▣ Acoustic Signal Processing, Voice Recognition, Feature Engineering, Machine Learning or Data Mining
- ▣ Graph Theory and Applications, Graph Signal Processing, Sensor Data Fusion or Machine Learning
- ▣ Hands-on experience on building software systems and algorithm optimization
- ▣ Experience on real world use case is preferred
- ▣ Creative, independent and self-motivated

**Project 4: Cognitive Internet of Things (IoT) Industry Solutions**

**Project description:**

The Internet of Things (IoT) research of IBM Research-China (CRL) is leading to develop world class technology and cognitive solution research innovations enabling the enterprises across industries to accelerate the transformation. In this project, we will work together to develop Cloud and Mobile innovations to effectively processing and analyzing the physical sensors data, correlated with contextual information like weather, map, knowledge base etc to develop Cognitive IoT industry solution innovations focusing on connected manufacturing industry 4.0 operations, connected vehicle, connected healthcare and connected machinery innovations etc.

- Design/develop innovate machine learning algorithm, data analysis model to discover insight from multiple source IoT sensor data.
- Design/develop contextual knowledge graph model and knowledge learning algorithm leveraging Cognitive analytics (Machine Learning, Deep Learning, Natural Language Processing, etc.)
- Publish and author high quality research papers/patents, Design/integrate with mobile application/solution.

**Desired skill:**

- Ph.D. or Master Students majored in computer science, electronic engineering, industrial automation, civil engineering or related areas with the following skills or experience:
- Data mining and machine learning skills.
- Hands-on experience on building software systems (e.g.: Mobile, Web etc).
- Creative, independent and Self-motivated, Team collaboration.
- Experience in industry oriented research and development is a plus

**Project 5: BlueSCAN: Cognitive Compliance Advisor**

**Project Description:**

Financial Services Industry today is more complex than ever witnessing high losses related to compliance issues, needs for Cognitive Compliance. Huge amounts of legal documents (e.g. regulations, contracts and policies) written in natural language requiring professionals interpretation and judgment, and thus the compliance checking process is high-cost, low-efficient and inconsistent due to the manual review required. BlueSCAN tries to provide a

automatic regulation tracking and semantic compliance checking capability by leveraging natural language processing and deep semantic cognitive analysis. It can significantly enhance contract professional productivity with significantly cost reduction.

More specifically, we expect intern to develop develop and deliver innovative technologies for BlueSCAN leveraging a variety of Natural Language Processing (NLP) and Machine Learning related capabilities. These technologies include (but not limited to)

- *N Natural Language Understanding (NLU) and Dialogue Management,*
- *Ontology and Knowledge Graph,*
- *Knowledge Representation and Reasoning,*
- *Supervised learning and Unsupervised learning,*

#### **Desired Skill:**

We are looking for candidates with Ph.D in Linguistics, Information Science, Computer Science, Applied Mathematics, or related areas, with the following qualifications to strengthen our team:

- *Outstanding researcher with a proven track-record on both conceptual and empirical aspects in the areas of NLP, or a related area;*
- *Good software development experience and programming skills to implement algorithms and conduct experiments.*
- *Strong analytical skills;*
- *Creativity, enthusiasm and self-motivation.*
- *Team-work and good Chinese/English communication skill*

#### **Project 6: Cognitive Healthcare**

To conduct research on developing big data analytics and cognitive computing methods for healthcare data, building data mining and machine learning models on real world clinical/behavior/genomic data, as well as evaluate and improve the model performance.

We are looking for Ph.D or Master students majoring in Medical Informatics, Information Science, Computer Science, Statistics, Applied Mathematics, or related areas, with the following desired skills:

#### **Desired skills**

- Strong in machine learning, data mining and statistics.
- Strong in software development, proficiency in at least one advanced programming language, such as Python and Java.
- Familiar with data analysis tools and libraries, such as SPSS, Weka, R and/or scikit-learn, etc.
- Knowledge of clinical/genomic/behavior data is a plus.
- Familiar with big data platforms and tools, such as HDFS, Hadoop, Mahout and Spark is a plus.

#### **Project 7: Industry Based Data Analytics & Optimization**

##### **Job Description**

The Industries and Solutions department aims to establish the industry thought leadership and innovative solutions through the synergy of advanced technologies and deep industry

insights. It focuses on advanced analytics, mathematical optimization and data mining. The focused industries include but are not limited to:

- Green Horizon initiative for environmental protection, to cooperate with China Government to address severe air pollution issues and help the government solve this extremely challenging problem (See Media Report in English:

- <http://uk.reuters.com/article/2014/07/06/uk-china-pollution-data-idUKKBN0FB0QC20140706>,

- Media Report in Chinese: <http://world.huanqiu.com/exclusive/2014-07/5051361.html>

)

- Energy & Utility (e.g., intelligent outage scheduling, high-precision weather forecasting, oil pipeline leak detection)

- Logistic & supply chain management (e.g. logistics network optimization, transportation optimization and vehicle routing, inventory optimization, demand forecasting, production planning and scheduling, sales and operations planning, supply chain risk, etc.)

- Smart commerce (e.g., customer analytics, precision marketing/recommendation, pricing, search, social analytics, etc.)

### **Job Requirements:**

We are looking for candidates with PhD or Master in Operation Research, Statistics, Artificial Intelligence, Machine Learning, Computer Science, Meteorology, Atmospheric or related areas with the following qualifications:

- Outstanding track-record in the areas of data analytic, optimization, simulation, high-performance computing or a related area.

- Excellent programming and system development skills.

- Self-motivated, responsible, good team-work and communication skill.

- Industry experience is a plus.

- Experience in data analytics tool, big data analytics or cloud computing is a plus