<table>
<thead>
<tr>
<th>No.</th>
<th>Project Domain</th>
<th>Project Name</th>
<th>Company</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cloud, networks, hardware</td>
<td>Correlating power consumption in Data Centers with Resource Utilization</td>
<td>Alcatel Lucent</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Cloud, networks</td>
<td>Development of a traffic generator and monitoring tools based on Intel’s DPDK Pktgen</td>
<td>Alcatel Lucent</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>UI designer\Web &amp; DB application, Optimization, AI</td>
<td>Advocacy engineering dashboard</td>
<td>Amdocs</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Web application , DB, UI , HTML5</td>
<td>Refactoring Business DB Viewer Tool</td>
<td>Amdocs</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>Big Data , Analytics , BI , Fraud , Mobile Financial Services, Machine Learning, Predictive Modeling, Rule based system</td>
<td>Discovering Fraud cases in Mobile Financial Services (MFS) world</td>
<td>Amdocs</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>Security, Communication networks</td>
<td>Java web firewall</td>
<td>Amdocs</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>Automation, Generation, Java, Web, Javadoc, Rest</td>
<td>Javadoc tool for REST mapping later</td>
<td>Amdocs</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>Information System, Development process management tool</td>
<td>Operations management tool</td>
<td>Amdocs</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>Automation, Search, Algorithm</td>
<td>Code parsing engine for test-automation</td>
<td>Amdocs</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>Networking, Linux user space programming, Intel DPDK, x86 multi-threading</td>
<td>TCP plug-in for high-speed network Traffic Generator</td>
<td>Cisco System</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>Computer graphics</td>
<td>Real-time volumetric rendering on distributed-shared memory hardware</td>
<td>HP Laboratories</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>Computer graphics</td>
<td>Real-time ray-tracing on distributed-shared memory hardware</td>
<td>HP Laboratories</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>Computer graphics</td>
<td>3D graph data visualization</td>
<td>HP Laboratories</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>Various (mainly web, operating systems, and data bases)</td>
<td>Production server agent for monitoring – various technologies</td>
<td>HP Software</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>Networks (In Containers – focus on Docker)</td>
<td>Micro-service topology discovery</td>
<td>HP Software</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>Information Processing, Internet</td>
<td>Enabling a files-based application to work on the cloud</td>
<td>IBM</td>
<td>16</td>
</tr>
<tr>
<td>17</td>
<td>CSP, Authorization, Database Security</td>
<td>From constraint satisfaction problem to access control management</td>
<td>IBM</td>
<td>16</td>
</tr>
<tr>
<td>18</td>
<td>CANCALED</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Cloud Services, Data Analytics</td>
<td>WANsOn</td>
<td>IBM</td>
<td>18</td>
</tr>
<tr>
<td>20</td>
<td>Computer Vision, Machine Learning</td>
<td>Classification of medical images using semi-supervised learning</td>
<td>IBM</td>
<td>18</td>
</tr>
<tr>
<td>21</td>
<td>Computer Vision \ Image Processing</td>
<td>3D structure reconstruction from Video</td>
<td>IBM</td>
<td>19</td>
</tr>
<tr>
<td>22</td>
<td>Computer Vision \ Image Processing</td>
<td>Automatic calibration of stereoscopic augmented reality glasses</td>
<td>IBM</td>
<td>20</td>
</tr>
<tr>
<td>23</td>
<td>Cloud Services and Machine Learning</td>
<td>Text understanding as a service on the cloud</td>
<td>IBM</td>
<td>20</td>
</tr>
<tr>
<td>24</td>
<td>Cloud, PAAS, Docker, Object store</td>
<td>Video transcoding service using FFmpeg and Docker on IBM BlueMix cloud</td>
<td>IBM</td>
<td>21</td>
</tr>
<tr>
<td>25</td>
<td>Web Application, Docker containers</td>
<td>Web based Docker management application</td>
<td>IBM</td>
<td>22</td>
</tr>
<tr>
<td>26</td>
<td>Cloud, Continuous Integration</td>
<td>Continuous Integration server as a service</td>
<td>IBM</td>
<td>22</td>
</tr>
<tr>
<td>27</td>
<td>Distributed systems</td>
<td>Making sense of CAP theorem</td>
<td>IBM</td>
<td>23</td>
</tr>
<tr>
<td>28</td>
<td>compilation, binary translation and program monitoring</td>
<td>Dynamic binary instrumentation for performance monitoring and performance of tuning of running applications</td>
<td>Intel</td>
<td>24</td>
</tr>
<tr>
<td>29</td>
<td>algorithm, information retrieval, matching</td>
<td>Matching know-how graphs</td>
<td>Intel</td>
<td>25</td>
</tr>
<tr>
<td>30</td>
<td>Security</td>
<td>Man in the cloud</td>
<td>Intel</td>
<td>26</td>
</tr>
<tr>
<td>31</td>
<td>Machine Learning, Artificial Intelligence, Combinatorial Optimization</td>
<td>Statistical relational learning via factorization</td>
<td>LogicBlox</td>
<td>27</td>
</tr>
<tr>
<td>32</td>
<td>Machine Learning</td>
<td>Deep Learning with Factorization Machines</td>
<td>LogicBlox</td>
<td>29</td>
</tr>
<tr>
<td>33</td>
<td>Embedded systems (Marvell high performance SoCs)</td>
<td>Real time performance monitor for embedded systems</td>
<td>Marvell</td>
<td>31</td>
</tr>
<tr>
<td>34</td>
<td>Software Systems</td>
<td>CT- Platinum DMS – platform independent utility for monitoring detection measurement system</td>
<td>Philips Medical Systems</td>
<td>32</td>
</tr>
<tr>
<td>35</td>
<td>Software Systems</td>
<td>CT- REST remote embedded scripting tool</td>
<td>Philips Medical Systems</td>
<td>33</td>
</tr>
<tr>
<td>36</td>
<td>Software Systems</td>
<td>TEvA – Tile Evolution Application</td>
<td>Philips Medical Systems</td>
<td>34</td>
</tr>
</tbody>
</table>
| 37 | Software Systems | TILTAN.CoM – Tile Test Analyzer
Configuration Manager | Philips Medical Systems | 36 |
| 38 | Operating Systems, Internet of Things (IoT), Computer Communications | Internet of Things (IoT) - drone flight controller on Android | Qualcomm | 37 |
| 39 | Image Processing | Synergy of modern video codecs and augmented reality techniques | Qualcomm | 38 |
| 40 | Data Structures, File Systems | Mobile storage synchronization profiling | Qualcomm | 39 |
| 41 | Data Structures, File Systems | Mobile storage profiling | Qualcomm | 41 |
| 42 | Mobile Applications, Image Processing, Algorithms | Camera guided map tracking (Android) | Qualcomm | 42 |
| 43 | Operating Systems, Internet of Things (IoT), Computer Communications | Internet of Things (IoT) - drone flight controller on Linux | Qualcomm | 43 |
| 44 | Operating Systems, Compilers | MicroController debugger IDE | Qualcomm | 44 |
| 45 | Big Data | Cloudera manager service management and monitoring | Akamai | 45 |
| 46 | Big Data & monitor | Yoda cluster monitor | Akamai | 46 |
| 47 | Big Data & monitor | Yoda resource management | Akamai | 46 |
| 48 | Big Data | Hive adaptor for data-catalog | Akamai | 47 |
| 49 | Electronic Design Automation | Signal Routing optimization in a multi-FPGA emulation system | Synopsys | 47 |
| 50 | Image Classification | Automatic storytelling | Shutterfly | 48 |
| 51 | Ranking of images on mobile devices | Mobile Picteligence | Shutterfly | 48 |
Company: Alcatel Lucent

Project Domain: cloud, networks, hardware

Project: Correlating Power Consumption in Data Centers with Resource Utilization

Supervisors: Udi Margolin and Sharon Mendel-Brin

Project Description: Cloud Nodes (data centers for clouds) typically include many compute, storage and networking equipment that provide the physical resources for creation of a cloud. Such physical resources use electrical power to run and need cooling (air conditioning) to maintain working temperature as needed for proper functioning.

One of the largest challenges for a data center and cloud owner is to keep energy consumption to a minimum while giving his customers the SLA he committed in terms of resilience, performance and capacity.

Cloud nodes which are designed for Telco applications (VNFS - Virtual Network Functions) are influenced by the load on the application(s) they are running and thus may have times in which a large portion of the resources are idle.

The purpose of the project is to learn the behavior of the energy consumption in a Telco Data Center and how it correlates with the load on the resources.

In this project you will have the opportunity to:

1) Part 1:
   a. Learn how a data center is built and the APIs required to collect information
   b. Create a monitoring, storage and visualization tool for collected parameters

2) Part 2:
   a. Experience tuning of Openstack placement configuration parameters
   Experience usage of VM migration with Openstack

Programming Languages and Development Platform: Programming: Scripting (language to be decided), Visualization (HTML5)
Environment: RH Linux, KVM Virtualization

Required Background: Students from the 5th semester and above should be qualified to handle the project.
PROJECT 2

Company: Alcatel Lucent

Project Domain: cloud, networks

Project: Development of a traffic generator and monitoring tools based on Intel’s DPDK Pktgen

Supervisor: Shachar Beiser

Project Description: Emerging servers within data centers and particularly in NFV (Network Function Virtualization) cloud are expected to support high data rates. For example, servers with multiple 40Gbps or even 100 Gbps NICs can handle traffic of around 400Gbps. Benchmarking those servers and evaluating the supported data rate is a known challenge. To that end, Intel developed its DPDK pktgen utility, based on its highly debated DPDK (Data Plane Development Kit). However, this utility is quite simple and provides only the basic functionalities. In this project the students would build a packet generator and a measurement tool based on the Intel naive utility. It is expected that the students will get familiar with this important development kit (DPDK) as well with measurement and benchmark methodology. Finally, a friendly user interface should be developed to support the tools functionalities and to allow an easy usage of the tool.

Programming Languages and Development Platforms: C, JAVA/Python.

Courses:

234118 Computer Organization and Programming
236334 Introduction to Computer Networks
236341 Internet Networking
234122 Introduction to Systems Programming

PROJECT 3

Company: Amdocs

Project Domain: UI designer\Web & DB application, Optimization, AI

Project: Advocacy Engineering Dashboard

Supervisor: Guy Sinter

Project Description: Project requires a development of a WEB Dashboard UI and associated backend DB to have an ability to view and track Advocacy Engineering (AE) team task status – e.g. CR (Customer Change Request) status, CR SLA Service Level Agreement (this is a predefine time commitment to the customer for AE maintenance task definition), Assigned developer.
Background details:
AE team responsible to solve issues (Customer change Request CR’s) arriving from both internal and external customers.
AE development lifecycle is by issuing Patch Bundle (PB) per version with heart beat of 2 weeks, Within the PB lifecycle there is the CR life cycle which is composed from these basic steps:
1. Accept the CR
2. Assign to developer
3. Root cause analysis
4. Solution discussion
5. Development
   a. CR fixing
   b. Adding missing automation test cases
6. Review
7. Test
8. Submit to PB Branch

The AE Dashboard should visualize for each PB currently in development the PB Automation status, the CR’s planned for this PB etc.
Also the dashboard should visualize CR’s which are currently not part of PB and their SLA and have the ability to plan in which PB’s they should be released.

Advising algorithm for AE best task assignment for PB release:
The algorithm will take into account all Q accepted tasks and calculate the best assignment optimization, taking into account number of AE employee, employee availability and constrains, knowledge seniority, CR complexity, calculated value of CR time to solve, number of PB involved, PB due date, CR SLA etc.

Features:
• High level status View:
• number of CR’s in queue
• Status indicator – on track/slight delay/major delay
• Patch Bundle view:
• PB Release Stream Automation status
• PB Temp stream Automation status
• CR’s associated to this PB
• PB state – planned, in dev, CCF (this means no more submissions to PB stream), released
• CR (defect) view:
• CR Status and state
• Assigned developer
• SLA
• Associated PB
• Actual work
• Estimated work
• Root cause
• Solution description

The user should have the ability to update the UI both manually and through REST API’s.
**Recommended Background:**
Must: WEB Development, DB applications
Nice to have knowledge: HTML 5, REST.

**PROJECT 4**

**Company:** Amdocs

**Project Domain:** Web application, DB, UI, HTML5

**Project:** Refactoring Business DB Viewer Tool

**Supervisor:** Zahi Gil

**Project Description:** DB Viewer is an existing tool which can create a dynamic view on RDBMS. This tool is currently written in Java Swing technology on top of Spring platform. We would like to refactor this Amdocs Tool from Java Swing technology to Web Application (HTML 5, JQuery, etc.). We would like to have full functionality of the existing tool capabilities (Please see below tool main functionality).

The steps are:
1. Get familiar with the current architecture and design of the tool (architecture reverses engineering, run and learn, code review and more).
2. Design the target architecture
3. Build a work plan
4. Execute with mile stones and roadshow
5. Productization (SDK and Documentation)

**Current Tool Screenshot:**

![Current Tool Screenshot](image)

**Tool main functionality:**
- Login/security mechanism
- Designer mode and user mode (permissions and role based)
- Define Data Entities
- Define query per data entity
- Define the data to display per data entity
- Connection to multiple data source
- Changes on the fly (during run time)
- Repository for the SQL queries
• Context to search and display the data entities
• Drill down per data entity (tree hierarchy)
• Conditional formatting configurable (icons, background, and more)

**Recommended Background:** Java, Swing Spring, HTML 5, JQuery, SQL, ORACLE, Java web server knowledge and Eclipse

---

**PROJECT 5**

**Company:** Amdocs

**Project Domain:** Big Data, Analytics, BI, Fraud, Mobile Financial Services, Machine Learning, Predictive Modeling, Rule based system

**Project:** Discovering Fraud cases in Mobile Financial Services (MFS) world

**Supervisor:** Leon Malalel

**Project Description:** The goal is to auto-discover, using Machine Learning algorithms and techniques or rule based system (or both), cases of Fraud in MFS transactions and alert them in real-time.

**Processing:**
The solution needs to scan every MFS transaction and calculate a score which is the propensity of the transaction to be legal or fraudulent, simulation data structure will be provided by the mentor. The score calculation mechanism will be built based on past cases of legal / fraudulent cases. High score means high propensity to be Fraud case. Transaction which have high score (= Fraudulent) will be alerted to a decision maker, so he can decided to take action, in real-time, while the transaction is still “alive”.

“*A significant problem of MFS is that companies that provide mobile money services are actually deploying small branches of banks in all kind of places, that usually don’t have the security and infrastructure of a bank and it creates potential for fraud*”
Output:
1. List of all the transactions analyzed with their propensity to be fraud score. This will later be used for further analysis and discovery of new patterns and trends, for model tuning purposes. The data should be saved in a standard data base and presented on UI
2. Alerts, in real-time, for this transactions above a given threshold. The solution should include a dedicated UI for the alerts.

Recommended Background: Analytics and Machine learning, statistics.

PROJECT 6

Company: Amdocs

Project Domain: Security, Communication networks

Project: Java Web Firewall

Supervisor: Dani Livne

Project Description:

General background: Amdocs products use a client server architecture, where the client is using Java Web Start and JNLP technologies. The client is a UI application, and using Java software it is able to send serialized Java objects over to the Amdocs server. The problem with Java serialization is that it is hard to monitor the traffic (unlike more modern UI approach like HTML 5) and alert on security threats like SQL Injections and XSS.

Project Objective: The students will need to combine the following technologies in order to build a tool that will protect Amdocs servers:

1. Use a flexible open source proxy (sniffer) tool in order to interject the traffic from the client to the server. We recommend Burp proxy tool since it has a convenient extension for Java deserialization classes.

2. Use an extension to the proxy tool in order to deserialize Java objects from the network traffic. We need to make sure that the extension is using Amdocs JARs for deserialization.

3. Send deserialized objects to an open source security scan software. We recommend MOD security for that.

4. Analyze the results from the security scan software, and report back in case of a problem (using the UI and an alert based on EMAIL or SNMP).

The students will have to learn the Burp extension and check if there are other better alternatives. The same with MOD security.

Responsibilities: Develop the Java Web Security module. The module should be able to intercept and alert on known SQL injections and XSS vulnerability.
**Programming Languages and Development Platform:** Programming skills in Java, application security background is advantage.

---

**PROJECT 7**

**Company:** Amdocs

**Project Domain:** Automation, Generation, Java, Web, Javadoc, Rest

**Project:** **Javadoc tool for REST mapping later**

**Supervisor:** Michael Mashian

**Project Description:** Multi-Channel Self Service (MCSS) is a client application built from client-side layer running on the end-user’s device (Browser, Tablet, Mobile), and a server side layer which exposes REST services that expose operations of underlying backend systems while encapsulating the complexity of their data model. In order to do so, MCSS introduces its own data model, which is designed to be simple, small in size, and UI oriented in terms of what-you-need-is-what-you-get.

In order to perform the data conversion to/from data model of various backend systems, a set of Java based mapper objects has been created. A Mapper is created per each Source & Target type duo (where Source & Target are from different data models), and its inner mapping logic is responsible to map between the (first level) attributes of the Source object to the mapped attributes of the Target object. Mappers can be reused by other mappers in case the Source & Target attribute duo is mapped by another mapper. Different mappers for similar Source & Target type duo can be distinguished by context.

The goals of the project are to create a tool that documents the underlying mapping relation of a given set of mappers, as follows:

- The tool generated a Javadoc structured document
- The document is categorized by packages and mappers (classes)
- Each mapper document describes
  - The Source & Target it represents
  - The mapping relation between the set of Source & Target attributes (primitives)
  - In case of Objects, a link to another mapper will be embedded
- The document Chained mapper hierarchy should be reflected
  - Including link to mappers within the hierarchy

**Recommended background:** Java programming knowledge; Familiarity with the Javadoc tool; Ability to work independently, following a given HLD; Open-minded; Innovative.

---

**PROJECT 8**

**Company:** Amdocs

**Project Domain:** Information System, Development process management tool
Project: Operations Management Tool

Supervisor: Mariam Khalil

Project Description: There is a need to develop management web application:

Requirements:
- Task management – add, edit, assign, remove task
- Define task as entity and keep metadata about this entity (Name, Description, Start time, End time, Average Duration ...)
- Define a request entity and keep metadata about this entity
- Prioritize between different activities.
- Email notifications and reminder - assigned task, postponed, completed
- Keep a repository (Data Base) in which all tasks will be stored
- Provide easy framework to extract reports on the repository and optimize the load and ensure balance between the different working hours by moving activities execution time
- The application should be a web application with a central repository
- At the beginning the application should support up to 100 users, 1000 predefine activities and same number for ad-hoc requests

Project Deliverables:
- Front end Tool that will provide end 2 end solution to the above requirements

Recommended Background:
- Basic understanding in Service Oriented Architecture and SQL
- Client server web application using: HTML5, PHP or Java script

PROJECT 9

Company: Amdocs

Project Domain: Automation, Search, Algorithm

Project: Code parsing engine for test-automation

Supervisor: Gilli Shama

Project Description: Amdocs and its more than 22,000 employees serve customers in over 80 countries, with revenue of $3.6 billion in fiscal 2014. Amdocs testing is the world leading provider of testing in the telecommunication industry.

Amdocs Testing has an in-house automation console that records and enables to edit test automation activities, and run them in series. The automation console holds a repository of test automation activities for re-use.

The goals of the project are to create a smart search engine in the automation repository. The user will record a full long business flow. The search engine will find a series of activities from the repository that will best to cover the recorded business flow in terms of testing coverage.

The developers team role is to define what is the best testing coverage from all the found activities for the defined business flow, review several possible algorithms for this problem,
define the best algorithm to aligned this testing coverage problem, and program the algorithm (no need in front end development), present verification to the coverage problem.

Business flow example and testing activities from the repository will be provided.

**Recommended Background:** Strong algorithm capabilities; Java programming knowledge; Innovative.

---

**PROJECT 10**

**Company:** Cisco System

**Project Domain:** Networking, Linux user space programming, Intel DPDK, x86 multi-threading

**Project:** TCP plug-in for High-Speed Network Traffic Generator

**Supervisor:** Hanoch Haim

**Project Description:** TRex is a low cost, high speed, stateful network traffic generator recently open-sourced by Cisco System. Stateful traffic generators are critical tools for testing network switches a router with realistic speed load scenario. More info can be found [here](#). TRex provides a single tool solution for network L2-3 and L4-7 traffic generation and analysis.

The project goal is to add a high speed multi-instance lock-free TCP stack plug-in to TRex software.

The work includes:
1. Developing tools for offline TCP traffic analysis
2. Understanding of DPDK/RSS NIC configuration
3. Integration of High speed TCP stack (based on mTCP project) with TRex as plugin

**Programming Languages and Development Platforms:** Linux, DPDK, C/C++/Python.

**Courses:** 236330, 234247, 236703

---

**PROJECT 11**

**Company:** HP Laboratories

**Project Domain:** Computer graphics

**Project:** Real-time volumetric rendering on distributed-shared memory hardware

**Supervisor:** Sagi Schein

**Project Description:** HP is developing a novel hardware architecture called The Machine ([http://www.hpl.hp.com/research/systems-research/themachine/](http://www.hpl.hp.com/research/systems-research/themachine/)). This new memory-centric computing architecture replaces the familiar disk with a large shared non-volatile memory space. The Machine hardware is constructed from a set of System on Chip (SoC) each with its own independent OS which are communicating over a fast communication fabric.
In this project we explore the merits of using this new system for real-time volumetric rendering. Basic approach for volumetric rendering simulates the physical process of shoot rays of light through semi-opaque material. As rays are reflected and refracted their contributions to the final image are accumulated on a frame-buffer to form high quality renderings. The processes is time consuming and relatively slow. In this project we want to effectively utilize the large amount of fast persistent memory and the fast interconnect of the Machine to achieve high quality real-time volumetric rendering. The end product will enable the user to interactively adapt aspects of the volume rendering pipeline (e.g. change material properties, camera).

In this project students will gain experience in system programming for future hardware architecture. The students will further gain experience in medical imaging and rendering techniques.

**Programming Languages and Development Platforms:** The project is implemented in C++ on a Linux “The Machine” functional simulator.

**Courses:** A course on computer graphics (234325 or compatible) and a course on distributed computing (236370 or compatible).

---

**PROJECT 12**

**Company:** HP Laboratories

**Project Domain:** Computer graphics

**Project:** Real-time ray-tracing on distributed-shared memory hardware

**Supervisor:** Sagi Schein

**Project Description:** HP is developing a novel hardware architecture called The Machine ([http://www.hpl.hp.com/research/systems-research/themachine/](http://www.hpl.hp.com/research/systems-research/themachine/)). This new memory-centric computing architecture replaces the familiar disk with a large shared non-volatile memory space. The Machine hardware is constructed from a set of System on Chip (SoC) each with its own independent OS which are communicating over a fast communication fabric.

In this project we explore the merits of using this new system for real-time ray tracing of large scene. Ray-Tracing simulates the physical process of shooting rays of light through an artificial scene. The contributions of reflected and refracted rays are accumulated into a frame-buffer to form the final image. The process is time consuming and view dependent. In this project attempt to utilize the large amount of fast persistent memory and the fast interconnect of the Machine to store large spatial index that could accelerate the production time and allow real-time responses. In addition smart caching of rays could be used to further accelerate the production of rays.

In this project students will gain experience in system programming for a future hardware architecture. The students will further gain experience in ray-tracing and image synthesis methods.
Programming Languages and Development Platforms: The project is implemented in C++ on a Linux “The Machine” functional simulator.

Courses: A course on computer graphics (234325 or compatible) and a course on distributed computing (236370 or compatible).

PROJECT 13

Company: HP Laboratories  
Project Domain: Computer graphics  
Project: 3D graph data visualization  
Supervisor: Sagi Schein  

Project Description: In recent years graphs are becoming increasingly important in the field of data analytics. The graph structure captures interactions between entities, different graph properties such as the valance of vertices, or the diameter of the graph may help practitioners in analyzing their data.

In this project we focus a special type of graph, time varying graph, which encodes for each edge its insertion time. The goal of this project is to build a visualization tool that could load a time varying graph and visualize it in 3D. The tool will need to compute a layout for the graph and render it in 3D. The tool will then enable the user to intuitively control the camera, and focus on vertices, edges and times interactively.

The students will gain knowledge in graph layout tools, 3D rendering and UI/UX design.

Programming Languages and Development Platforms: There are multiple alternatives for implementing this project. Either using Javascript or a free game engine is an option for this project.

Courses: A course on computer graphics (234325 or compatible).

PROJECT 14

Company: HP Software  
Project Domain: Various (mainly web, operating systems and data bases)  
Project: Production Server Agent for Monitoring – Various Technologies  
Supervisor: Amichai Nitsan  

Project Description: Production servers are servers that run software that is constantly being used by users (think about your bank online application). These servers should be available all the time without interruptions.
Monitoring agents are software components that are deployed into these servers for monitoring them. I.e. track server execution, identify when the software runs too slow, or maybe even not functioning at all, and alert the operator about it. Also it should collect the relevant data that will help to fix the problems. The real challenge with coding such agents is to do all monitoring and data collection without hurting the server performance and availability.

This project is about experimenting with writing production server monitoring agents for new technologies and modern environments.

**Programming Languages and Development Platforms:** This will be based on the specific environment the students will choose. But it is safe to assume development will be one of: Java / JavaScript / C# / Ruby / PHP (again, dependent on the choice of the students).

**Courses:**
- 234123 Operating Systems
- 234319 Programming Languages

---

**PROJECT 15**

**Company:** HP Software

**Project Domain:** Networks (In Containers – focus on Docker)

**Project:** *Micro-Service Topology Discovery*

**Supervisor:** Avi Saidian

**Project Description:** Docker is rapidly gaining a place as the next virtualization technology. Being lightweight, fast, developer friendly and highly portable is promising a large adaptation of container technology.

In this project we will explore Docker using a monitoring perspective to discover containers and their dependency in a certain environment.

We will create a tool which discover start/stop of containers and sniff the network calls to discover topology map and dependency between many micro-services.

Our tool should take into consideration the Software Defined Networking capability of Docker.

**Programming Languages and Development Platforms:** Java/Go, Intellij/Eclipse.

**Courses:**
- 236334 Introduction to Computer Networks
- 236341 Internet Networking
- 236700 Software design
- 236370 Concurrent and Distributed Programming
PROJECT 16

Company: IBM

Project Domain: Information processing, Internet

Project: Enabling a files-based application to work on the cloud

Supervisor: Uri Shani

Project Description:

We want to run software that was developed for single computer systems with the standard file system and have it work on the cloud.

We choose to use a simple data model and a small number of use cases. In the file system, files are written only once and do not go through updates. We have a simple naming scheme that makes them easier to manage. In order to work on the cloud, we choose an efficient and well-established management system with REST API - CLOUDANT.

Our goal is to convert the use cases of the software to use the same data model using the API of CLOUDANT.

Steps of the project:

1. Getting familiar with IBM BlueMix. Assistance in obtaining accounts and interfaces will be provided by IBM.
2. Getting familiar with CLOUDANT API.
3. Getting familiar with the use cases and obtaining a sample of the database to be converted and displayed in several scenarios.
4. Preparing and checking the development documents and the trials.
5. Implementation of the interface.
6. Testing the trials that were designed and agreed upon by IBM, and checking against the database provided.
7. Final report.

Results:

1. Development and testing documents. Summary of the project also includes a final report and test results.
2. Implementation in Java.
3. Results of the execution against the database provided.

Programming Languages and Development Platforms: Java, Eclipse, IBM BlueMix.


PROJECT 17

Company: IBM

Project Domain: CSP, Authorization, Database Security

Project: From Constraint Satisfaction Problem to Access Control Management

Supervisor: Boris Rozenberg
Project Description: Role Based Access Control (RBAC) is the widely used access control model. To use an RBAC system, one needs to first identify a complete set of roles, including permission role assignments and role user assignments. This process, known as role engineering. Since many organizations already have some form of user permission assignments defined, it is reasonable to identify roles from this existing information. This process is known as role mining. Recently, a method that transforms the role mining problem into a constraint satisfaction problem has been published [1]. The transformation allows to discover the optimal RBAC state based on customized optimization metrics.
What is requested in this project is to extend the proposed method in several directions and to implement the overall approach, using open source CSP solvers, such as [2].

Programming Languages and Development Platforms: Java.

Pre-requisites: The understanding of CSP fundamentals is needed.

References:
PROJECT 19

Company: IBM Research – Haifa

Project Domain: Cloud Services, Data Analytics

Project: WhatsOn

Supervisors: Simona Rabinovici-Cohen and Meni Amran

Project Description: What’s on the material you should learn?
The project aims to build a tool for medical students to help them learn for exams. The tool will accept the material on which a student is tested and a set of keywords, and will automatically generate questions (and possibly answers) based on this material. Med school students, learning for exams, will use the tool to get generated questions that s/he can use to practice for the exam based on the relevant text book and keywords. Example keywords can be “treatment”, “most common cause of”, “symptoms”, etc. The tool will be useful for other users as well, such as university professors.

The solution will include a simple front-end, leaving the main efforts for the back-end. The back-end of the solution will analyze medical text books and use this information to smartly build different variations of questions and their corresponding answers. The solution involves identifying the relevant text in the material using information retrieval (IR) and possibly natural language processing (NLP) techniques, converting it into questions and further retrieval and phrasing of optional answers. It will leverage medical knowledge to enhance its strength. The back-end will use IBM Cloud services such as AlchemyAPI, Watson Text Analytics, and possibly other services including image analytics services. It will also include self-developed analytics modules running on Apache Spark. The back-end will be developed in the Java programming language using the Eclipse IDE and it will run on Windows and Linux.

Programming Languages and Development Platforms: Java, Eclipse, IBM Cloud Services, SPARK.

PROJECT 20

Company: IBM Research – Haifa

Project Domain: Computer Vision, Machine Learning

Project: Classification of medical images using semi-supervised learning

Supervisors: Rami Ben-Ari and Guy Amit

Project Description: Computerized analysis of medical images often utilizes machine learning techniques, in order to derive complex relationships between image features and their semantic meanings. Supervised learning approaches require costly and laborious effort of image annotation by medical experts, which is often a bottleneck in medical imaging research. In this project, we will explore a semi-supervised approach for classification of medical images. This approach fuses a small amount of labeled data with a larger set of unlabeled data to yield a reliable classifier. The application domain of this project will be automatic lesion classification in
X-ray mammograms and magnetic resonance images of the breast. The project will combine advanced computer vision techniques for content-based retrieval of unannotated images, with state-of-the-art machine learning algorithms. The students will have a chance to gain experience in the field of medical imaging analytics.

Programming Languages and Development Platforms: Matlab (image processing and statistics toolbox).

Courses:
- Experience in Matlab programming
- Background in machine learning
- Background in computer vision / image processing

PROJECT 21

Company: IBM
Project Domain: Computer Vision / Image Processing

Project: 3D structure reconstruction from video

Supervisor: Yochay Tzur

Project Description: Reconstruction of 3d shapes and environments is an important task in many applications, such as visual-based navigation, 3d objects recognition, tracking and augmented reality etc. Therefore, various reconstruction approaches were proposed over time, based on everything from a single image, a pair of stereo images, a video from a moving camera, or special sensors like the Kinect.

In this project the students will implement a video-based reconstruction method, which will enable a user to “3d scan” objects (and even a whole room), using a simple mobile phone camera. The implemented method will be evaluated in various real-world scenarios, and domain-specific improvements might be applied if needed.

Programming Language and Development Platforms: C++.

Courses: 236327 Signal and Image Processing by Computer (or any other image processing / computer vision course).
PROJECT 22

Company: IBM

Project Domain: Computer Vision / Image Processing

Project: Automatic Calibration of Stereoscopic Augmented Reality Glasses

Supervisor: Yochay Tzur

Project Description: Augmented Reality Glasses (aka “see-through” glasses) are emerging devices, which deliver a new, fascinating experience of Human-Computer Interaction. The glasses use two semi-transparent displays positioned in front of the user eyes, to display information associated with real objects directly on the objects. An integrated camera enables the glasses to “see” the world from the user viewpoint, and respond accordingly. In order to have the graphics accurately displayed on the “reality”, and get a true 3-dimensional experience, one should calibrate the glasses screens specifically for each user. In this project, the students will develop and implement a method for automatic calibration of the AR glasses. This involves working with the video coming from the glasses’ camera, rendering to the stereo displays, and interacting with the device using finger gestures.

Programming Language and Development Platforms: Java, Eclipse, C++ (optional).

Courses: 236327 Signal and Image Processing by Computer (or any other image processing / computer vision course)

PROJECT 23

Company: IBM

Project Domain: Cloud Services and Machine Learning

Project: Text understanding as a service on the cloud

Supervisor: Einat Kermany, Moran Gavish

Project Description: IBM Bluemix is a powerful cloud platform for building, running, and managing applications. Deep learning is a set of machine learning algorithms that are inspired by a class of theories of brain development. Recently published papers present exciting results on numerous machine learning tasks in various application domains.
The aim of this project is to build a prototype implementation of deep learning algorithms for text representation that shall be available as a service on the cloud. Students in this project will implement a neural network based Auto encoder for representing textual sentences and will deploy it as a service on IBM Bluemix. The service will allow automatic detection of semantical similar sentences.

**Programming Language and Development Platforms:** Java or Python.

**Required prerequisites:**
- machine learning

**PROJECT 24**

**Company:** IBM

**Project Domain:** Cloud, PAAS, Docker, Object store

**Project:** Video transcoding service using FFmpeg and Docker on IBM BlueMix cloud

**Supervisors:** Ophir Azulai and Udi Barzelay

**Project Description:** Bluemix is an implementation of IBM's Open Cloud Architecture based on Cloud Foundry, an open source Platform as a Service (PaaS). Bluemix delivers enterprise-level services that can easily integrate with your cloud applications without you needing to know how to install or configure them. Bluemix provides access to a wide variety of services that can be incorporated into an application.

Docker containers wrap up a piece of software in a complete filesystem that contains everything it needs to run: code, runtime, system tools, and system libraries – anything you can install on a server. This guarantees that it will always run the same, regardless of the environment it is running in.

FFmpeg is a free software project that produces libraries and programs for handling multimedia data.

In this project we will develop a scalable video transcoding service on BlueMix using Docker and FFmpeg. Users can upload videos to the BlueMix's object store, and then use the transcoding service to transcode the file to another video format. The output file will be stored back in the object store.

**Programming Language and Development Platforms:** C++, Java, Linux.

**Courses:** Software Engineering Methods (234321) or Object Oriented Programming (236703).
**PROJECT 25**

**Company:** IBM  
**Project Domain:**  
**Project:** Web Based Docker management application  
**Supervisor:** Alexey Roytman, Artem Barger  

**Project Description:** With the recent proliferation of micro services, container technology has emerged as a natural fit to support such architectures. Docker is a well-known container platform which has nurtured a powerful ecosystem of utilities for the management and rapid development of micro-service applications. Kitematic is an application developed for Apple OS X to help manage Docker container images, to import, create and run them on local machines through a powerful yet user-friendly UI.

In this project you will have an opportunity to become intimately familiar with Docker and its agent API and to develop a web based application which similarly to Kitematic will allow remote management of Docker container images, to allow distributed execution and monitoring of containers across nodes.

**The project has two phases, which could be split into two different teams:**
1. Design the application backend  
2. Develop and implement powerful Web UI.

**Useful links:**
1. [https://www.docker.com/](https://www.docker.com/)  
2. [https://www.docker.com/docker-swarm](https://www.docker.com/docker-swarm)  
3. [https://docs.docker.com/machine/](https://docs.docker.com/machine/)  
4. [https://docs.docker.com/compose/](https://docs.docker.com/compose/)  
5. [https://kitematic.com/docs/](https://kitematic.com/docs/)

**Programming languages:** Any.

**PROJECT 26**

**Company:** IBM  
**Project Domain:**  
**Project:** Continuous Integration server as a service  
**Supervisor:** Artem Barger  

**Project Description:** IBM Bluemix is an open-standarts, cloud platform for building, running and managing applications. Built on Cloud Foundry open source technology, Bluemix makes
application development easier with Platform as a service (PaaS) the main goal is to simplify implementation and application delivery providing services that are ready to use.

In this project you will leverage IBM Bluemix infrastructure to create and build Continuous Integration server as a service, by integrating well know open source CI solution - Jenkins.

**Project goals are following:**

1. Integrate with Bluemix Virtual Machines infrastructure to create a deployment of Jenkins server.
2. Expose Jenkins management via Bluemix service dashboard panel.

**Possible project extensions:**

1. Allow to manage more than one Jenkins deployment on single machine (multi tenancy).

**Useful links:**

1. [http://www.martinfowler.com/articles/continuousIntegration.html](http://www.martinfowler.com/articles/continuousIntegration.html)
3. [https://jenkins-ci.org/](https://jenkins-ci.org/)

**Programming languages:** Any.

---

**PROJECT 27**

**Company:** IBM

**Project Domain:**

**Project:** Making sense of CAP theorem

**Supervisor:** Artem Barger, Bruno Wassermann

**Project Description:** The CAP Theorem represents one of the fundamental results of distributed systems. It states that any asynchronous distributed system can satisfy only two out of three of the following desirable properties:

1. Consistency - maintain strong consistency of replicated data;
2. Availability - general service availability;
3. Partitioning - tolerate network partitions.

Nowadays, developers of distributed systems have many off-the-shelf components available to them to implement the handling of their system's data and communication that promise to achieve various combinations of the above system properties to varying degrees, such as MongoDB, Redis, Cassandra, Solr, RabbitMQ, and so on. Developers are confronted with a large design space that they must navigate in order to configure and use these systems in ways that can achieve adequate trade-offs between these properties.

**Project goals is:**
We want to establish how these components actually perform in practice under different configuration profiles.

1. Pick three of the most widely used distributed storage solutions.
2. Understand the space of configuration settings for each of them and identify what their impact on Quality of Service is in terms of the CAP properties according to the documentation.
3. Build a workload generator using the Jepsen framework to test different configurations of these components under network partitions.
4. Run experiments and analyse the results to demonstrate what the actual impact of different configurations on the fulfilment of the CAP properties is in practice.

2. https://aphyr.com/tags/jepsen

**Programming languages:** Any, while to use Jepsen framework students will have to learn Clojure programming language.

---

**PROJECT 28**

**Company:** Intel, IDC

**Project Domain:** compilation, binary translation and program monitoring

**Project:** Dynamic binary instrumentation for performance monitoring and performance of tuning of running applications

**Supervisor:** Gadi Haber

**Project Description:** The proposed project uses an existing Intel tool called Pin. Pin is a tool for the dynamic instrumentation of programs. It provides the ability to write arbitrary code (in C or C++) which are then injected at designated places in the executable. Pin provides a rich API that abstracts away the underlying instruction set idiosyncrasies and allows context information such as register contents to be passed to the injected code as parameters.

The project purpose is to use the Pin SDK in order to form an efficient performance monitoring software tool for long running applications. The tool must have a negligible performance overhead on the running application.

The collected statistics are then used to apply automatic tuning of calls to various library functions and at the same time provide relevant indications to the program developers and system administrators about the application performance efficiency.

**Courses:** 236360 Theory of Compilation.
Company: Intel

Project Domain: algorithm, information retrieval, matching

Project: Matching know-how graphs

Supervisor: Daniel Gross

Project Description: Know-how is an essential ingredient in scientific work and in engineering. We refer to know-how as technical knowledge that helps to achieve stated objectives, while offering most acceptable tradeoffs. Researchers explore relevant know-how to identify outstanding research problem and to justify their scientific contributions. Engineers search and explore solutions that may address their problems, effectively and/or efficiently. However, identifying, analyzing and applying know-how is challenging. Know-how is often buried in the minds of people, or in numerous scientific and technical papers, not readily accessible to support the identification of relevant solutions to problem at hands, and/or to identifying opportunities for advancing the state of the art.

This project, in the broader area of knowledge representation, addresses the know-how management/organization problem by adopting a know-how mapping approach. We develop ontology (meta-model) for mapping out know-how, and a web application to manually capture and store know-how maps of engineering domains in know-how graphs for further querying and reasoning.

We propose a Know-how graph matching/searching project. The goal of the project is to support merging know-how maps produced by one or more authors, representing engineering know-how of closely related engineering problems, to create a composite know-how map. For example, suppose one know-how map represents know-how of one advancement in a security algorithm, and another an additional advancement, each was extracted from a different scientific paper. The purpose of the project is to develop an algorithm that can combine both know-how maps into a combined map that represents the combined know-how of the scientific contributions of both papers in one composite map. More specifically, the objective is to develop an algorithm to search for semantic match of graph fragments (representing portions of a know-how map) within existing know-how graphs.

At the conclusion of this project the students would gain experience in the development of KR related algorithms while working with terminology and meaning in a problem solving context rather than by strings searches only. The students would work with examples that are relevant to industry and industrial problem solving. Finally, the students will gain a deeper understanding of how to systematically evaluate alternative engineering solutions in relation to their key advantages and liabilities, and in relation to relevant to solutions elements at hand; and what it means to make a research contribution in engineering domains.

The students are required to:
- Look for existing solutions and to map them out using the approach we developed.
- Refine/custom/adjust existing solutions to achieve the project goal or to devise a completely new implementation.
- Evaluate and refine their proposed solution.

For more details look here: [https://goo.gl/w6q344](https://goo.gl/w6q344)

**Programming Language and Development Platforms:** Java/ JavaScript, Mongo DB.

**Courses:**
- 234122 Introduction to Systems Programming
- 234218 Data Structures
- 234247 Algorithms 1
- 236359 Algorithms 2 (optional but preferable)
- 236299 Intro. to Natural Language Processing (optional but preferable)

---

**PROJECT 30**

**Company:** Intel

**Project Domain:** Security

**Project:** Man in the cloud

**Supervisor:** Lenz Oron, Ilya Berdichevsky

**Project Description:** Last month, Imperva published a paper called ‘man in the cloud’, in which they describe in details, how a simple malware can practically manage to steal all the files put in a shared folder that is synchronized by a sync service to the cloud. Imperva tested this on all the major cloud services today, including dropbox, box, google drive and OneDrive. The attack is using the sync service to send a token that is generated by the cloud service in order to replace a user/password pair for authentication to the attacker machine, where it is used to impersonate the client and manage to get a copy of his entire folder. The proposed project is to create a safe client using Intel Software Guard Extension (SGX) technology. The client will replace the existing sync service on the client machine, using the standard APIs exposed by the cloud service. That sync service will provide the same capabilities w/o the vulnerability. To do this, the improved service will use SGX to protect the tokens and the SGX version of TLS for safe communication.

**Courses:** 234123 - Operating systems (this should cover other pre-requisites).

**Programming languages and development platforms:** C/C++ on SDV with SGX SDK (if possible - SKL machine with SGX).
PROJECT 31

Company: LogicBlox

Project Domain: Machine Learning, Artificial Intelligence, Combinatorial Optimization

Project: Statistical Relational Learning via Factorization

Supervisor: Prof. Benny Kimelfeld

Project Description:

Overview
Statistical Relational Learning (SRL) is a branch of Machine Learning and Artificial Intelligence that focuses on models that combine relational structures, logic, and uncertainty. The challenge with SRL is the high computational complexity that SRL tasks (learning and inference) entail. The growing public appetite for Big Data analysis intensifies the need for scaling up SRL implementations. In this proposal, the student will implement a novel approximation approach that is based on tensor factorization. With this project, the student will deepen in, and gain experience with, the important concept of SRL, and may have a substantial impact on the research of facilitating practical SRL.

Background
Statistical Relational Learning (SRL) is a branch of Machine Learning and Artificial Intelligence that focuses on models that combine relational structures, logic, and uncertainty. SRL has attracted a lot of attention because of its generality, ease of development, and the applicability to multi-modal data. SRL has been the foundation for building knowledge bases (KB). Markov Logic Networks (MLN) and Probabilistic Soft Logic (PSL) are two state of the art approaches. For MLN based KBs, deepdive is a state of the art tool.

Despite its success in small scale data, there are some challenges in scaling it for real world problems. Generally, the techniques taken to computationally solve SRL tasks (training and inference) are of two approaches. In the first, the relational information is mapped down to a probabilistic graphical model. In the second approach, the relational information is modeled as multidimensional arrays, and efficient solutions are established through various kinds of factorizations. Factorization models give us an embedding of facts and relationships into a latent domain in the same way that Google’s word2vec deep net represents words in a k-dimensional latent space. Factorization models tend to scale better than graphical models. On the other hand, factorization models lack expressivity, since they do not provide any natural embedding for logical rules.

The goal of this project is to increase the expressivity and the complexity (model capacity) of the factorization models, by using more sophisticated methods, such as Factorization Machines along with objective functions borrowed by Probabilistic Soft Logic. One advantage of this approach is that it requires much simpler algorithms for training and inference than probabilistic graphical models, and it reduces the memory footprint. If the project is successful, then the result will provide a framework for easy-to-code solutions for SRL problems on real Big Data domains, and often on one’s personal laptop!

Probabilistic Soft Logic
For an introduction to PSL see video here. For applications of PSL, look at this video.

PSL uses the Lukasiewicz t-norm (Lnorm) to express the degree a rule is satisfied. For example, if we have predicates of the form $R(x) \land L(y)$ and $I(R(x))$ is the degree of satisfaction of $R(x)$ in the range $[0, 1]$, then the degree of satisfaction of the rule is given by

$$\max(I(R(x)) + I(L(y)) - 1, 0).$$

If the predicate is $R(x) \lor L(y)$, then the degree of its satisfaction is given by

$$\min(I(R(x)) + I(L(y)), 1).$$

The *distance to satisfaction* of a formula is given by

$$\max\{0, I(r_{body}) - I(r_{head})\}.$$

Where $r_{body}$ is the premise (or body) of the rule, and $r_{head}$ is the conclusion (head) of the rule. A possible world is obtained by assigning a degree of satisfaction to each ground fact, and its probability is proportional to an exponent in the sum of satisfactions (where each rule has an associated weight coefficient).

**Project Proposal**

In this project the student will implement an SRL solution in a manner that is different from existing solutions, as we describe below. We conjecture that this implementation will be significantly more efficient than existing ones, with possibly more accurate results. The input will include rules and entity domains. In training, the goal will be to determine rule weights by fitting to examples. In inference, the goal will be to assign degrees of truth to facts by maximizing a goal function (that aggregates the degree of truth over all rule groundings).

The technique will be based on Logistic Tensor Factorization (LTF). In LTF, the relational data are modeled as a tensor, according to the following illustration.

The above method factors the relationship tensor $R$ as $X = AR AT$. The loss functions are given below:

$$\arg\min_{X, R} loss(X; A, R) + \lambda_A ||A||^2 + \sum_k \lambda_R ||R_k||^2$$

$$loss(X; A, R) = \sum_k ||X_k - AR_k A^T||^2$$
The problem of the above method is that it can only model binary predicates of the form \( R(x, y) \). We would like to be able to express first-order logical rules of the form:

\[
R_1(x, y, l) \land R_2(z, w) \rightarrow R_3(t)
\]

So, we need to face two challenges.
1. Extension to k-ary predicates with \( k > 2 \).
2. Support of logical rules.

For the first challenge, we propose to use Factorization Machines that can model higher order relationships. For the second problem, we propose to use the objective function that PSL is using.

**Evaluation**
The success of the project will be determined by benchmarks evaluated on standard SRL datasets, and will be compared against PSL.

---

**PROJECT 32**

**Company**: LogicBlox

**Project Domain**: Machine Learning

**Project**: Deep Learning with Factorization Machines

**Supervisor**: Prof. Benny Kimelfeld

**Project Description:**

**Introduction**
Deep Learning has been a highly successful approach in problems like image recognition, speech recognition and text mining. In all of these problems, data points are correlated either temporally or spatially. In the multibillion world of retail business, demand forecasting is a critical problem. A few percentage of accuracy improvement in prediction can save hundreds of millions of dollars in wastage. For many years, the retail community was limited by manually tuned linear-regression models that required careful feature engineering. Very recently, there has been a movement towards a new algorithmic approach - Factorization Machines (FM). The FM algorithm allows to automate a substantial portion of the tuning process. Both linear regression and FM have been preferred due of their natural interpretation. In this project, we intend to extend FM to accommodate a deep architecture while maintaining interpretability.

**Background**
The Factorization Machine (FM) algorithm has been successful in modeling multimodal information in recommender systems. It has also served very well as a regression solution that
can model higher-order interaction terms between variables. For example, if the temperature and the product name are two features in a linear regression problem, then their combination (temperature, product) is another feature that specifies the effect of the combination. Such correlations are often extremely useful; for instance, in high temperatures coke demand goes up, while blanket demand goes down.

The second order FM is:

$$y = f(x) = w_0 + \sum_{i=0}^{d-1} w_i x_i + \sum_{i=0}^{d-1} \sum_{j<i}^{} v_{ij} x_i x_j$$

or

$$y = f(x) = w_0 + \sum_{i=0}^{d-1} w_i x_i + \sum_{i=0}^{d-1} \sum_{j=i+1}^{d-1} v_{ij} x_i x_j$$

The third order FM would be:

$$y = f(x) = w_0 + \sum_{i=0}^{d-1} w_i x_i + \sum_{i=0}^{d-1} \sum_{j=i+1}^{d-1} \sum_{k=1}^{d-1} v_{ijk} x_i x_j x_k$$

where $v \in \mathbb{R}^k$. Although FM has much more capacity than linear regression and it can model more complicated scenarios, it still has some limitations.

Increasing Model capacity with nonlinearities
It has been shown in the deep-learning community that introducing nonlinearities can significantly increase the model capacity. Nonlinearities also promote sparsity, which is important for generating interpretable data representations. This project will focus on demand forecasting where the dependent variable $y$ is continuous. There, the Rectifier Linear Unit (RLU) function is appropriate:

$$r(x) = \max(0, x)$$

We propose the following modification to the original model:

$$y = f(x) = w_0 + \sum_{i=0}^{d-1} w_i x_i + \sum_{i=0}^{d-1} \sum_{j=i+1}^{d-1} v_{ij} r(x_i x_j)$$

Now the FM model can be written as:

$$y = f(x) = w_0 + \sum_{i=0}^{d-1} w_i x_i + \sum_{i=0}^{k-1} r_i(x)$$

This is a single-layer neural network. We can think of adding layers in two different ways. The first option is as follows:

$$v = l(x) = w_0 + \sum_{i=0}^{d-1} w_i x_i + \sum_{i=0}^{k-1} \sum_{j=i+1}^{k-1} v_{ij} x_i x_j + \sum_{i=0}^{k-1} \sum_{j=i+1}^{k-1} r_i(x)$$

The other option is to work on higher-order FM and introduce the nonlinearity on each order of interaction level:

$$y = f(x) = w_0 + \sum_{i=0}^{d-1} w_i x_i + \sum_{i=0}^{k-1} \sum_{j=i+1}^{k-1} v_{ij} x_i x_j + \sum_{i=0}^{k-1} \sum_{j=i+1}^{k-1} r_i(x)$$

Increasing Model Capacity with Recursive Data Partitioning
When data is multimodal, a mixture model is typically more appropriate. In other words, increasing the capacity of your model by adding layers is not as effective as partitioning the data.
and training FMs or nonlinear FMs as described above. Random partitioning is the best you can do in random forests (e.g., when you do not have any prior information on the data). FM and nonlinear FM both map every training point \((x_i, y_i)\) to a \(k\)-dimensional geometric representation. For example, in the nonlinear case we have:

\[
(x_i, y_i) \rightarrow ([r_0(x_i), \ldots, r_{d-1}(x_i)], y_i)
\]

Each dimension of the \(r\)-vector represents a mode. The sparsity or the variance of each dimension is an indication of complexity. Splitting the dataset on the dimension with the highest complexity gives two new datasets that can be retrained with a new FM. This process can be recursively applied to every partition of the data. The result is a binary tree where every node has an associated FM. It is very similar to a decision tree or a kd-tree. Tree based models have been highly successful in different domains. Moreover, such models often offer natural interpretations on the results.

**Evaluation**

All methods for increasing the model capacity of FM will be tested on a real production system for demand forecasting in big US retailers. Any improvement in the accuracy can measure the immediate impacts on savings for the retailers.

---

**PROJECT 33**

**Company:** Marvell

**Project Domain:** Embedded systems (Marvell high performance SoCs)

**Project:** Real time performance monitor for embedded systems

**Supervisors:** Gil Tal and Ami Sytbon

**Project Description:** Marvell SoCs includes high performance ARM CA9 and CA72 (64 bit) processors along with many highly advanced units such as packet processor, high performance DDR controller, FLC (Final level cache) and much more. Such a high performance SoCs needs a real time non-intrusive performance analysis to improve overall system performance and understanding bottlenecks. Purpose of this performance monitor application is to offer a graphical application or plugin to run under windows. The application will connect with a Marvell board and will let the user choose from various performance monitor available by the board (the protocol will be defined by the students and us), reading the performance monitor will be done through HW interfaces, the board will control the color of graphs in real time to enable the user to identify changes is SW states while running. This project will give the students the chance to gather vast knowledge on the latest ARM V8 processors and many interfaces used by high end embedded systems, as well as the ability to experiencing full SW development cycle as we do in Marvell. Ami Sytbon who will be the leader of this project is considered one of the best at Marvell with vast knowledge in object oriented. The project will be developed using C# under the latest Microsoft WPF environment.

**Programming Languages and Development Platforms:** C++.
**Company:** Philips Medical Systems

**Project Domain:** Software Systems

**Project:** **CT- Platinum DMS – Platform Independent Utility for Monitoring Detection Measurement System**

**Supervisor:** Yan Tsitrin

**Project Description:** The Philips Company offers its customers various CT-scanners keeping them at the forefront of CT imaging (see [http://www.healthcare.philips.com/main/products/ct/products/scanners/](http://www.healthcare.philips.com/main/products/ct/products/scanners/) for details). Detection Measurement System (DMS) is one of the major components of a CT-Scanner. It consists of detectors, which primary function in the CT system is to produce digital number (frequency) corresponding to the flux of impinging X-Ray radiation on their top surface.

The DMS architectures (mechanical, electronical etc.) of different CT Scanners are different. In Philips different tools for monitoring of DMS of different kinds have been created. These tools are used by CT developers (e.g., SW or HW engineers, Support engineers, production technicians, etc.) in order to check (and sometimes change) various parameters of the DMS system. Each such tool can be used only in the context of its specific scanner and does not makes sense in context of scanners of other types.

![Figure 1: Philips CT Scanners Families and their DMS monitoring tools.](image)

The objective of this project is to develop Platform Independent Utility for monitoring any kind of DMS (Platinum DMS). The utility shall be able to adjust its visual representation and available functionality in the run-time according to the Scanner type; and then can be used by a CT developer in context of the scanner currently operated.

The student shall be provided with the existing monitoring tools and shall develop generic Platinum DMS, initially in the simulation environment, and then on the real Philips CT-Scanners of Brilliance and iCT families. The development shall be done in .NET, C# and shall include all SW
development stages: requirements definition, design, implementation, debugging, testing, verification.

Programming Languages and Development Platforms: \([C\#]\).

Courses: 234122 introduction to Systems Programming.

Advantage: Background in GUI development, Object-Oriented Design.

PROJECT 35

Company: Philips Medical Systems

Project Domain: Software Systems

Project: CT- REST Remote Embedded Scripting Tool

Supervisor: Yan Tsitrin

Project Description: The interaction between Philips CT Scanner components and the HOST computer controlling the Scanner is done by message passing via the CAN communication bus (for more details see

http://www.cs.technion.ac.il/~cs234313/projects_sites/W15/34/site/).

Detection Measurement System (DMS) is one of the major components of a CT-Scanner. It consists of detectors, which primary function in the CT system is to produce digital number (frequency) corresponding to the flux of impinging X-Ray radiation on their top surface.

The DMS is controlled by embedded SW running on a processor residing on the DMS. Often it is necessary to update the embedded environment (e.g. to write to a DMS register) from the HOST computer. Such updates can be done one-by-one, when each write/read request is send via the CAN.

To improve this tedious and boring process we in Philips developed REST (Remote Embedded Scripting Tool), Figure 1.
REST allows to create a script on the HOST side (1), inject the script (via the CAN) into the embedded processor (2). The script is a LUA (see http://www.lua.org/) based set of instructions executed by the embedded processor. The script is executed by the LUA interpreter (3) which is compiled and linked together with the embedded SW; and the execution results are sent to the HOST back (4). However, this tool is a command-line based and its user-experience should be improved.

**Figure 3: REST: HOST – CT scanner script-based communication.**

In this this project the student shall be provided with the existing command based REST tool and shall develop GUI for it. The GUI shall support editing, compiling and getting results which shall be presented in a convenient way. The development shall be done in .NET, C#, initially in the simulation environment, and then on the real Philips CT-Scanners of **Brilliance** and **iCT** families. It shall include all SW development stages: requirements definition, design, implementation, debugging, testing, verification.

**Programming Languages and Development Platforms:** {C#, C++}.

**Courses:** 234122 introduction to Systems Programming.

**Advantage:** Background in Embedded Systems.

---

**PROJECT 36**

**Company:** Philips Medical Systems

**Project Domain:** Software Systems

**Project:** **TEvA – Tile Evolution Application**

**Supervisor:** Yan Tsitrin

**Project Description:** The Tiles are main building blocks of the DMS (Detection Measurement System). Their primary function in the CT system is to produce digital number (frequency) corresponding to the flux of impinging X-Ray radiation on their top surface. The detector is
composed of an array of small size X-Ray detectors pixels (16x16). The detectors substrate is connected to a mixed signal Front End Electronic (FEE) VLSI chip. 4, 6 or 8 Tiles are mounted on a detector module (depending on the system). Each detector module provides the necessary power to its Tiles and a communication channel to/from the Tiles.

Figure 4:

[a] DMS – Detection system with Modules

[b] Module with Tiles

[c] Tile – Matrix Representation

The DMS modules with the Tiles on them serve the CT-scanners for years since they are installed, and we in Philips are interested in ensuring of their reliability. Reliability demonstration testing of units under use conditions is often impractical because test duration is expected to be in the order of several years while consuming hundreds of units in test samples. Therefore, an accelerated life test must be performed under stress conditions above and beyond the expected use conditions. The results of all tests are combined and extrapolated to use conditions, assuming some time to failure distribution and life-stress relationship, collectively known as reliability model. In the testing process the tests are applied to a Tile in groups (transactions) of 5 tests in each one. After each transaction a pixel in the tile is marked as defective (of one of the tests in the applied transaction fails) or survived otherwise.

The objective of this project is to develop TEvA (Tiles Evolution Application) providing visualization of a tile evolution (i.e. changing its matrix representation from the state where all its pixels are survived up to the moment when their major part is defective). The development shall be done in .NET, C# and shall include all SW development stages: requirements definition, design, implementation, debugging, testing, verification.

Figure 2:

A Tile pixels matrix evolution visualization

Programming Languages and Development Platforms: {C#}.

Courses: 234122 Introduction to Systems Programming.

Advantage: Background in Statistics and SW Reliability.
**COMPANY**: Philips Medical Systems

**PROJECT DOMAIN**: Software Systems

**PROJECT**: TILTAN.Com – Tile Test Analyzer Configuration Manager

**SUPERVISOR**: Yan Tsitrin

**PROJECT DESCRIPTION**: The Tiles are main building blocks of the DMS (Detection Measurement System), one of the major components of a CT-Scanner. Their primary function in the CT system is to produce digital number (frequency) corresponding to the flux of impinging X-Ray radiation on their top surface. The detector is composed of an array of small size X-Ray detectors pixels (16x16) that’s connected to a ceramic substrate by conductive glue. 4, 6 or 8 Tiles are mounted on a detector module (depending on the system).

![Figure 5](image)

The DMS modules with the Tiles on them serve the CT-scanners for years since they are installed, and we in Philips are interested in ensuring of their reliability. Reliability demonstration testing of units is performed under stress conditions (like, temperature and humidity) above and beyond the expected use conditions.

While analyzing these experiments results numerous parameters should be set by the test operator, among them:

- Environmental parameters (temperature, humidity etc.),
- Description of tiles under test,
- Path/Fail criteria for different kinds of test and modules types,
- Test equipment,
- Pre-existing bad detectors
- Etc

At the moment these parameters are mainly hard coded in the Tiltant-SW and the tool should be recompiled in order to change a parameter.

*The objective of this project is to develop Tiltan.Com (Tile Test Analyzer Configuration Manager), a SW tool allowing the test operator to define experiments parameters in a*
**convenient way.** After the parameters are set they shall be stored in xml format and may be used in the current and further analyzing sessions.

The development shall be done in .NET, C# and shall include all SW development stages: requirements definition, design, implementation, debugging, testing, verification.

**Programming Languages and Development Platforms:** [C#].

**Courses:** 234122 introduction to Systems Programming.

**Advantage:** XML knowledge, background in GUI development.

---

**PROJECT 38**

**Company:** Qualcomm

**Project Domain:** Operating Systems, Internet of Things (IoT), Computer Communications

**Project:** Internet of Things (IoT) - Drone flight controller on Android

**Supervisors:** Valeria Perelman and Dany Rybnikov

**Project Description:** Internet of Things is a new concept that defines that all human surroundings can be controlled and monitored through Internet Protocol. Few examples of IoT are always connected automobiles, wearable health monitors, autonomous vehicles and drones, autonomous and connected surveillance systems, baby monitors, connected to Internet homes with connected autonomous air conditioning, connected refrigerators and many more.

For such IoT use cases, the industry and academies are developing small low cost computers that can be embodied within such connected devices as their controllers, connectivity mechanisms and computation engines. One such example is DragonBoard 410c, a board with Qualcomm 4-core 410 processor, WiFi, Bluetooth, GPS, USB and other functionalities. DragonBoard 410c board runs Linux Ununtu and Android 5.1 OSes.

In this particular project, the students will develop a connected drone controller SW that runs on DragonBoard 410c with Android.

The requested goals
1) Develop or port an existing open source SW for controlling drones to Android on DB410c.
2) UnitTest the flight Control SW.
3) Develop a communication protocol and smartphone application to monitor and control the drone from smartphone.
4) Demo the flight control mechanism.
5) Contribute the development to the open source community.
Additional (bonus) goals
1) Identify additional needed hardware to show case the development on a real drone.
2) Demo a real drone with DB410c and developed/ported SW.

Available information
1) Open Source flight control SW project ArduPilot (https://github.com/diydrones/ardupilot). Usually runs on Real Time OS NuttX.

Programming Languages and Development Platforms: C/C++, Java. Android on DB410c.

Required background: Operating Systems, Linux User space and kernel programming, Communication Protocols.

PROJECT 39

Company: Qualcomm

Project Domain: Image Processing

Project: Synergy of modern video codecs and augmented reality techniques

Supervisor: Alex Fradkin

Project Description: Anyone can feel rapid progress in displays and mobile technologies in the last decade.
Small and simple feature phone ten years ago had small 1 or 2 inch screen with 320x240 resolution in best case and awful camera. Nowadays Smartphone striking our imagination with 5-6 inch screens with 4K resolution and good multi-megapixel camera. Modern TV, PC and media markets also adapting 4K resolution and have plans to move to 8K displays in near future.

Another phenomenon of last decade is social networking, video hosting and cloud services rising.
It is hard to find person not familiar with Facebook, Whatsapp or Youtube.

So the future is now, all these innovations resulted in that billions of people film and shearing millions of videos every day using their smartphones. All these people are willing to film and to watch videos with highest possible quality.

Two large video codecs standard updates have been made over last decade to satisfy these needs – H.264/AVC and then it successor H.265 (HEVC). But even the latest and modern one H.265 is based on principals introduced 20 years ago in MPEG and MPEG-2.

Motion estimation and compensation is one of the basic techniques and one of main pillars used in all video codecs for last 20 years. Quality of motion estimation is key element for good video compression. H.265 motion estimation algorithm like all its predecessors is based on direct pixel-by-pixel comparison between subsequent pictures. H.265 half the video bitrate and size for
the same video quality from its predecessor H.264, but this improvement requires 10 time more computational power for basic algorithm like motion estimation.

Augmented Reality is another rapidly developing area. Motion estimation and object tracking is one of basic tasks in this area. Feature based techniques were adapted and widely used for motion estimation tasks in augmented reality. Plenty of new feature points detectors and tracking algorithms were developed for Augmented reality usage over last decade.

In this particular project, the students will use Augmented Reality feature based techniques from OpenCV to improve motion estimation and compensation algorithms in modern video codec like X265.

The requested goals
1) Use open source x265 project to learn modern video codec and it motion estimation/compensation algorithms.
2) Use open source OpenCV library to learn modern techniques for feature points detection and tracking.
3) Develop a video motion estimation/compensation techniques based on OpenCV feature points tracking and merge it into x265 codec.
4) Investigate computational requirements, achieved compression and quality for new codec.
5) Demo of video compression and decompression based on new techniques.
6) Contribute the development to the open source community.

Additional (bonus) goals
1) Optimize developed code to have better compression ratio than original x265.
2) Optimize developed code to have better computational requirements than original x265.

Available information
1) Open Source x265 codec [http://x265.org/](http://x265.org/)
2) OpenCV library [http://opencv.org/](http://opencv.org/)

**Programming Languages and Development Platforms:** C/C++. Linux/Windows PC.

**Required background:** Computer Vision / Video codecs / augmented reality / Image processing.

---

**PROJECT 40**

**Company:** Qualcomm

**Project Domain:** Data Structures, File Systems

**Project:** Mobile storage synchronization profiling

**Supervisor:** Tanya Brokhman

**Project Description:**

---

Page 39 of 49
As mobile computing becomes a bigger part of our everyday life, the demands from mobile storage are changing. The new use-cases require higher storage performance, characterized by fast response time and concurrent execution of I/O requests. Fast storage is a key factor for:

- Smooth UI experience
- Multimedia, such as camera multi-shot scenarios
- Efficient concurrency of multiple applications

The software storage stack for block devices in Linux kernel can be abstracted into the following layers:

- User space (Applications)
- File System
- Block layer
- Low Level Device Driver

Storage performance evaluation requires deep analysis of the storage stack access patterns and utilization of enhanced low level driver features. This project is concerned with the profiling of two storage synchronization mechanisms - FLUSH and BARRIER – which differ in implementation, in performance impact and in data loss prevention capability.

The storage activity is captured using the blktrace utility, a block layer I/O tracing mechanism which provides detailed information about request queue operations up to user space.

BTT is a post-processing tool for blktrace output. BTT takes blktrace traces as input, analyzes the events, and produces much of the information needed for profiling storage behavior.

The project gives the student an opportunity to understand the Linux storage SW stack across its different layers - from the application layer, through the file system, into the block layer and device driver, down to the flash memory card itself. The student will get acquainted with Linux-community and open source development process.

The goal of this project is to enhance the open source BTT tool to allow deeper analysis of the flush and barrier access patterns in different storage usecases. This includes:

- Addition of BARRIER traces to blkparse and BTT
- Collection of FLUSH, FUA and BARRIER statistics
- Output of the above statistics in the standard BTT format

Blktrace output file from real traffic will be provided.

The student will need to deeply understand the storage stack access patterns and basic functionality (an overview presentation will be provided) as well as the blktrace events.

Blkparse and BTT code will be contributed back to open source project. It assumes high code quality; ability to facilitate public code review; team work.

This contribution will be recognized by the community and will be used by the block devices developers.

Requested project milestones

- Initial BU of the blkparse and BTT project and addition of BARRIER support
- Addition of FLUSH, FUA and BARRIER statistics and their presentation
- Contribution of the code to the community
**Programming Languages and Development Platforms:** C, Linux PC.

**Required background:** Data structures, algorithms.

## PROJECT 41

**Company:** Qualcomm

**Project Domain:** Data Structures, File Systems

**Project:** Mobile storage profiling

**Supervisor:** Tanya Brokhman

**Project Description:** As mobile computing becomes a bigger part of our everyday life, the demands from mobile storage are changing. The new use-cases require higher storage performance, characterized by fast response time and concurrent execution of I/O requests. Fast storage is a key factor for:
- Smooth UI experience
- Multimedia, such as camera multi-shot scenarios
- Efficient concurrency of multiple applications

The software storage stack for block devices in Linux kernel can be abstracted into the following layers:
- User space (Applications)
- File System
- Block layer
- Low Level Device Driver

Storage performance evaluation requires deep analysis of the storage stack access patterns and utilization of enhanced low level driver features.

The storage activity is captured using the blktrace utility, a block layer I/O tracing mechanism which provides detailed information about request queue operations up to user space. The traced operations are mainly: adding an I/O request to the block layer’s request queue (Q), dispatching the request to the low level driver (D) and completion of the request (C).

BTT is a post-processing tool for blktrace output. BTT takes blktrace traces as input, analyzes the events, and produces much of the information needed for profiling storage behavior.

The project gives the student an opportunity to understand the Linux storage SW stack across its different layers - from the application layer, through the file system, into the block layer and device driver, down to the flash memory card itself. The student will get acquainted with Linux-community and open source development process.

**The goal of this project**

Enhancement of the open source BTT tool to allow deeper analysis of the storage access patterns and load. This includes:
• Integration of the current solution of read/write separation on top of the latest BTT project
• Separation of merge ratio according to the direction of the activity (read/write)
• Add the ability to track individual process activity
• Separate the latencies results according to the bio / request size
• Add support for reinstert events
• Add additional statistics calculation and histograms (such as C to D or D and D)

Raw blktrace output file from real traffic will be provided.

The student will need to deeply understand the storage stack access patterns and basic functionality (an overview presentation will be provided) as well as the blktrace events.

BTT code will be contributed back to open source project. It assumes high code quality; ability to facilitate public code review; team work.

This contribution will be recognized by the community and will be used by the block device driver developers.

Requested project milestones

• Initial “bring-up” of the BTT project and a complete solution for separation of read/write statistics
• Addition of per-process profiling
• Enhancement of latencies statistics
• Contribution of the code to the community

**Programming Languages and Development Platforms:** C, Linux PC.

**Required background:** Data structures, algorithms.

---

**PROJECT 42**

**Company:** Qualcomm

**Project Domain:** Mobile Applications, Image Processing, Algorithms

**Project:** Camera guided map tracking (Android)

**Supervisor:** Dror Baum

**Project Description:** In the current widespread usage of smart mobile devices the usage of the camera is usually limited to its primary use case – take great pictures and videos. In this project, the students will be asked to develop an additional use case for the camera – use it indoors and find where the phone is located on a map within a building.

The goal of the project is to develop an Android application the core of which is an indoor positioning recognition algorithm. The students will be requested to integrate the Application with previously developed indoor map system and show the performance of their algorithm with the map.
The behavioral outline of the requested application:
- Capture image on the mobile phone
- Process the image and feed it into the map system
- Assess the location of the phone

Requested Project Milestones:
1) Students will study previous related work which has already been done
2) Students will propose, design and implement the image processing algorithm to enhance map tracking using camera captures
3) Student will study the week spots of the current algorithms, design and implement improvements

Programming Languages and Development Platforms: Java for Android, C. Linux/Windows PC.


Recommended Background: Android SDK, Databases.

**PROJECT 43**

**Company:** Qualcomm

**Project Domain:** Operating Systems, Internet of Things (IoT), Computer Communications

**Project:** Internet of Things (IoT) - drone flight controller on Linux

**Supervisors:** Liron Kuch and Constantine Elster

**Project Description:** Internet of Things is a new concept that defines that all human surroundings can be controlled and monitored through Internet Protocol. Few examples of IoT are always connected automobiles, wearable health monitors, autonomous vehicles and drones, autonomous and connected surveillance systems, baby monitors, connected to Internet homes with connected autonomous air conditioning, connected refrigerators and many more.

For such IoT use cases, the industry and academies are developing small low cost computers that can be embodied within such connected devices as their controllers, connectivity mechanisms and computation engines. One such example is DragonBoard 410c, a board with Qualcomm 4-core 410 processor, WiFi, Bluetooth, GPS, USB and other functionalities. DragonBoard 410c board runs Linux Ununtu and Android 5.1 OSes.

In this particular project, the students will develop a connected drone controller SW that runs on DragonBoard 410c with Linux Ubuntu.

**The requested goals**
1) Develop or port an existing open source SW for controlling drones to Linux Ubuntu on DB410c.
2) UnitTest the flight Control SW.
3) Develop a communication protocol and smartphone application to monitor and control the drone from smartphone.
4) Demo the flight control mechanism.
5) Contribute the development to the open source community.

Additional (bonus) goals:
1) Identify additional needed hardware to show case the development on a real drone.
2) Demo a real drone with DB410c and developed/ported SW.

Available information:
1) Open Source flight control SW project ArduPilot
   (https://github.com/diydrones/ardupilot). Usually runs on Real Time OS NuttX.
2) DragonBoard 410c - https://www.96boards.org/products/ce/dragonboard410c/,
   https://www.96boards.org/forums/forum/products/dragonboard410c/

Programming Languages and Development Platforms: C/C++, Java. Linux Ubuntu on DB410c.

Required Background: Operating Systems, Linux User space and kernel programming, Communication Protocols.

PROJECT 44

Company: Qualcomm

Project Domain: Operating Systems, Compilers

Project: MicroController debugger IDE

Supervisor: Oren Reiss

Project Description: Many SoCs (system On Chip) in consumer electronics market are equipped with uControllers which are tailored for specific tasks. Such processor usually has proprietary instruction set, tool chain and debug capabilities. In order to develop software effectively for such uController, developer would probably need IDE which allow him to navigate the code, place breakpoints and control the processor execution (stop, step, run...)
In this project, students will create a generic Debugger IDE for uControllers. They will parse the symbols files (produced by uController’s compiler/linker), present the source code in GUI, correlate between code and current PC (program counter) and provide basic debugger capabilities.
Outline of the system:
Requested Project Milestones:
1) Study existing symbol file structure and build effective parser.
2) Define uController interface and implement a stub.
3) Explore alternative for GUI (Plugin for “on shelf” IDE such Eclipse or “from scratch implementation”).
4) Establish basic debugger capabilities such as: placing breakpoints, stepping...
5) Perform integration and conduct system testing.

Programming Languages and Development Platforms: C/C++/Java/C#

Required Background: Operating Systems, Software Design.

PROJECT 45

Company: Akamai

Project Domain: Big Data

Project: Cloudera manager service management and monitoring

Supervisor: Boris Shubaev

Project Description:

Background: Yoda is our query engine which enables us to perform sub second queries and long running analysis on our Big Data cluster. Log-Agent is our incoming data logger, which supports tens of thousands of connections simultaneously. Cloudera Manager is the industry’s first and most sophisticated management application for Apache Hadoop and the enterprise data hub.

Today we use the Cloudera Manager as management and monitoring tool for Hive, HBase and HDFS

Abstract: We would like to add our proprietary services – Yoda & Log-Agent as services into Cloudera Manager. The project includes deep dive into the big data management and monitoring tools world and development in a Big Data environment across multi machine cluster.

What will the student learn: The student will get the chance to learn the inner workings of the leading Big Data platform, and contribute to a solution that will be moved into a real world production environment.

Programming Languages and Development Platforms: Java, Perl, Bash
PROJECT 46

Company: Akamai

Project Domain: Big Data & Monitor

Project: Yoda cluster monitor

Supervisor: Amir Skovronik

Project Description:

Background: Yoda is our query engine which enables us to perform sub second queries and long running analysis on our Big Data cluster. Yoda cluster monitor is our system to monitor the Yoda queries being execute on our cluster.

Abstract: We would like to replace our current implementation of cluster monitor to support more advance UI capabilities and enable analysis & statistics capabilities. The project will include sophisticated UI implementation as well as develop analysis & statistic solutions in Big Data environment across multi machine cluster.

What will the student learn: To build a web based management and monitoring system from scratch with a major impact on the design and technology selection.

Programming Languages and Development Platforms: Java, JSP, Spring MVC, Jscript, AJAX.

PROJECT 47

Company: Akamai

Project Domain: Big Data & Monitor

Project: Yoda resource management

Supervisor: Amir Skovronik

Project Description:

Background: Yoda is our home-brew query engine - very similar to Facebook Presto / Cloudera Impala. Currently it has no resource management, thus one query can: consume all available memory; run an endless loop taking 100% CPU, etc.

Abstract: The project is to add resource management both in terms of memory and CPU. One way to do it is through YARN. We'll define resource pools, and for each pool we'll spawn a Yoda instance limited by memory and CPU. The project includes researching how other query engines are doing that and trying to come up with a working prototype.

What will the student learn: In depth knowledge of Big Data query engines and develop a big data resource management solution.
Programming Languages and Development Platforms: Java

PROJECT 48

Company: Akamai

Project Domain: Big Data

Project: Hive adaptor for data-catalog

Supervisor: Boris Shubaev

Project Description:
Background: The CSI platform uses its home-brew solution named Data Catalog for managing the data it saves in the Big Data platform, be it HBase, HDFS, etc. This library has two parts: Catalog definition and APIs for reading/writing data defined in the Catalog. The catalog is composed of several schemas, which in turn are composed of several tables. Each table has a record structure (much like a RDBS table) and storage properties (HBase and its configuration, HDFS, etc.) Today we have the ability read/write from each table by using: Apache MapReduce (OutputFormat, InputFormat) our very own query engine named Yoda (similar to Impala/Presto).

Abstract: Our security research team required us to supply them with SQL language for querying the data in the Big Data platform, which is the reason we installed Hive for them. Today for we manually create Hive tables, and also resort to exporting much of our Data into CSV format and define manually Hive tables on top of them. The task is to enable querying in Hive any table defined in Data Catalog.

What will the student learn: Learn about HDFS, Hive and the HDFS MetaData repository

Programming Languages and Development Platforms: Java

PROJECT 49

Company: Synopsys

Project Domain: Electronic Design Automation

Project: Signal Routing optimization in a multi-FPGA emulation system

Supervisor: Dmitry Korchemny

Project Description: A very hot domain which is currently growing in chip development is hardware acceleration for simulating chip design (emulation). This domain requires solution of different large-scale optimization problems. Specifically the optimal assignment of logical wires to physical pins is one such problem. A solution for this has been developed on paper by the advisor, and needs to be implemented and tried out. The project scope will be to implement the solution and to benchmark it on several in-house designs. It will be conducted on Linux using C++.
**Programming languages and development platforms:** Linux, C++, Python.

**Courses:**
- Data Structures 1 234218
- Digital Computers Structure 234267
- Introduction to optimization 236330 (recommended)

---

**PROJECT 50**

**Company:** Shutterfly

**Project Domain:** Image Classification

**Project:** Automatic storytelling

**Supervisor:** Roman Sandler

**Project Description:** Today's image classification tools automatically label an image with a set of more or less relevant tags. They can find people, animals, trees, and buildings. They can also deal with more abstract concepts like cruise or hike. In this project we shall use that technology to annotate albums of images with a short description.

The first step of the project will be automatic tagging of thousands of images with one or more automatic classification tools. Then the performance of these tools will be evaluated and compared. The last stage will be summarization of the image tags for an album to describe the image set as a whole.

**Programming Languages and Development Platform:**
- Python and PHP development in Linux at AWS.
- Work with APIs of third party tools.

---

**PROJECT 51**

**Company:** Shutterfly

**Project Domain:** Ranking of images on mobile devices

**Project:** Mobile Picteligence

**Supervisor:** Eran Cohen

**Project Description:** Statistical ranking of images has proven to provide great value for photo based companies. Most of the current known images ranking techniques are based on backend processing, which include computer vision analysis, face detection/recognition and image quality analysis.

In this project the goal is to develop lightweight ranking system on mobile devices, provide a new curated gallery view, which shows only the better photos.
Programming Languages and Development Platform: Mobile platform can be either IOS/Android platforms.

Required experience:
- Experience in IOS/Android is required
- Course in image analysis is an advantage