

China Center of Excellence

Project Guardian

eBay is a global company, projects within eBay normally require efforts and synergies from teams located in different cities of different countries. This is the challenge we are facing. How to help these teams work more efficient? How to improve the quality of the projects while the cost is reduced? How to detect potential risks at early stage?

“Project Guardian” is our solution to the challenge. We decide to build an artificial intelligence to collect/organize/analysis and publish information to all project members. Just like a very smart project manager, based on large scale of data collected from different channels, such as code/bugs/emails/documents/messages, it will be able to tell us is this project healthy? Or it is highly likely to fail. And it will also be able to send out beautiful daily reports on “what happens today related to this project?”

This is unique opportunity for you to build a tool that will behave just like a real person to solve a real problem in our lives.

In “Project Guardian”, you will face great challenge. You need to know how to talk to different systems, such as exchange server/github/Jira/Lync/Wiki to collect data for analysis purpose. You need to work with no-sql database such as MongoDB to store all kinds of data efficiently, and you may need powerful data statistics tool, such as hadoop, to fetch meaningful and organized information, and you need powerful analysis tool(such as R) to present the information to end readers. Besides, in order to understand what people are talking in emails, you need to find out the way to deal with natural language based on machine learning algorithm.

This is hard, but I guarantee that it will be an interesting journey that pushes you to achieve something that you never imagine.

Need 4 Software Engineers

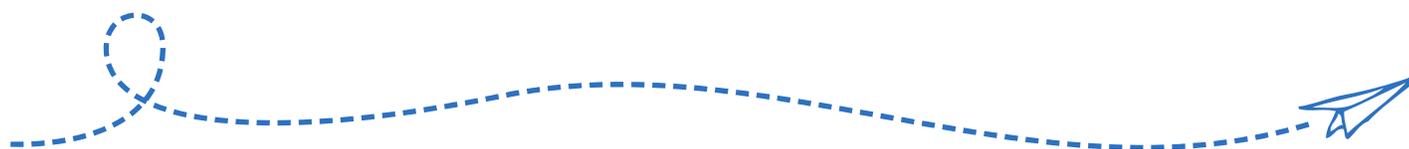
Performance tuning for eBay Web Framework

Raptor is the primary and preferred Java development platform at eBay. It consists of a light-weight framework that is built on top of Spring framework and takes advantage a set of Open-source projects to help develop Web Applications, SOAP & REST Services, Messaging and Batch Applications.

Raptor 2.0 is our latest eBay web framework which includes many open source components. Since those open source products are not naturally designed for eBay use cases, there are spaces to improve and achieve much higher performance.

In this project, you will have opportunities to work with our seasoned developers on solving the items like below:

- Learn low level detail about Tomcat and how is the servlet engine work. Performance tuning on Tomcat/Jasper.



- Investigating into Spring which is a very popular open source framework. Performance tuning for Spring and other open source components.
- Performance tuning for GZIP. Learn GZIP algorithm and the implementation.

Using existing open source projects where appropriate and contributing to existing open source projects is eBay's strategy. Look forward to working with you on exploring those cutting-edge technologies.

Need 2 Software Engineers

Hadoop Eagle active monitoring

Hadoop is good at scalability, but not that good in self-healing. We are facing the challenges of smoothly running Hadoop cluster with thousands of nodes which may run thousands of MapReduce jobs and a hundred million tasks per day. The challenges come from the fact that any single disk, host or any single task could fail, which would probably fail some critical jobs. Without active monitoring, keeping running large Hadoop cluster is mission impossible.

Hadoop Eagle active monitoring project aims at crawling, parsing, storing, searching and analyzing huge amount of MapReduce performance metrics. This project specifically builds the intelligence of determining which node is abnormal by analyzing the raw data in seconds. Once the culprit node is identified, backend remediation process will run and the cluster will recover soon. The intelligence part includes effective modeling on host anomaly and metrics anomaly. Anomaly modeling firstly comes from human study of past failures, but eventually becomes a machine learning process. Timely detection of anomaly is another challenge, which requires anomaly computation and query should be extremely fast, otherwise the issues goes away when we wait the query to return. We use Hbase to store the performance data and use a house-made query language and query framework to search the data.

This project uses cutting-edge technology to analyze Hadoop cluster performance issues. It's a huge challenge but of course interesting. By participating in this project, you can learn Hadoop technology in deep and besides you will learn how to build a large scalable system to store and query huge data.

Need 2 Software Engineers

Dataworks

eBay is a real data-driven company that deals with a massive volume of data everyday. eBay's business is heavily relying on data to make decision and drive innovation. How to make the power of data available to everyone?



Dataworks is a POC (proof of concept) project for eBay's next generation big data analytics platform which makes data works for everyone. Join this project, you will use your software design, analysis and programming skills to make things happen:

- Build a scalable platform for structured and unstructured data integration
- create friendly interfaces to enable people working together on data analytics and visualization

Being a member of this project, you can learn and practice cutting edge big data technologies: cloud computing and NOSQL etc.

Need 3 Software Engineers, 1 Data Engineers

Keyword Web Application

This will be an enterprise intelligent transactional database application powered by terabyte keyword data generated from eBay buyers search behavior, ideally built upon Hadoop + Hbase.

It is expected to be able to:

1. Respond to user queries within seconds.
2. Provide actionable keyword reports in user friendly format

Provide extra mile intelligent assistance telling users "this is something you also want to know."

1. Keywords insights is CRITICAL to eBay market place:

- Keywords represents market demands.
- Hot keywords tell us products that are most wanted by buyers.
- Emerging keywords leads to emerging opportunities eBay want to seize and establish.

2. Keywords insights need to be revealed and delivered to business user in timely manner

3. Keywords insights need to come with business intelligence that goes beyond human brain power

In the project, intern can learn ETL on hadoop, data science on hadoop, utilize HBase to build schema and also can learn building application on top of HBase.

Need 1 Software Engineers, 2 Data Engineers

Search Session Explorer

EBay's search is one of the largest e-commerce search engine in the world. Given the huge traffic of eBay's search, it generates massive amount of user log every day. The search log is extremely crucial for us to understand users' behaviors and continuously improve search experience. Therefore, we want to build a methodology and interactive tool which could easily analyze and visualize our log and provide data insight. For example, for users who query "iphone 5s", by using this search session explorer tool, we could know how many of them just leave after search, how many of them skip all the items in first page and click next page, how many of them click category/aspect refinement and how many of them rewrite the query to other queries, etc. The tool will dramatically help us to understand user behaviors in search session.



In this project, you will have opportunities to learn and use cutting edge techniques such as data mining, Hadoop and R to handle real big data (terabyte per day) in industry, to cooperate with smart engineers/researchers and even world class statistician (technique fellow) in eBay, and to develop an industry product with great business impact with you own hands!

Need 4 Software Engineers

A Real-time and Intelligent Recommendation Engine for Fans

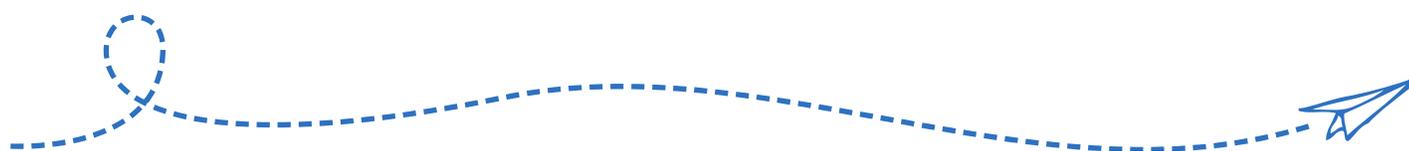
StubHub, the world's largest secondary ticket marketplace, is an eBay company. In US market, in the last month (Feb 2014), there are about 1.7 Million tickets sold in StubHub, more than 16 Million people visiting our site and listing over 3.8 Million tickets.

Our customers include fans and individual investors. To provide better user experience, we plan to build a recommendation engine which can real-time detect the user's preference and recommend appropriate events or tickets to him/her. The calculation is based on our 50+ Terabytes clickstream logs in Hadoop cluster and 10+ Terabytes data warehouse. We want to make the recommendation near to real-time and across multiple devices.

There will be great challenges as well as opportunities for you to learn how the recommendation engine being built from scratch. You have to deal with the TBs data processing in Hadoop platform and think as an architecture to build lightning-fast computing framework. You will join the design of the recommendation algorithm based on our business. We encourage you to creatively solve problems by adopting various cutting-edge technologies. Of course, you won't work alone. Team and our experts will fully support you.

Please join us. Let's make something cool! It's an exciting experience you shouldn't miss.

Need 5 Software Engineers



China Analytics Center

CRM Upsell Segmentation

Ebay new buyers purchase amount (GMB) is declining in 2013, although the number of new buyers is growing. To driven the GMB positively growing, we need to do customer segmentation, and find out the target with high probability to purchase expensive item on ebay.

Intern will work closely with senior modelers and business partners, and by using ebay in-house data associated with other data sources, you will:

Identify and explore data source which could measure user's buying power and purchase intention;
Research the best statistical/mathematical approach that can rank order user's buying power and purchase intention;

Build up model to help BU well target the right customer for marketing campaign.

Need 2 Business Analysts

Ship Cost Impact on ebay MP Equilibrium

Shipping cost is a critical component in Finance and Marketing Plan, and to optimize the financial plan, it is quite important to understand the impact of shipping cost on buyer's behavior.

Intern will work closely with senior shipping analytics, and using the in-house data associated with other data sources, you will:

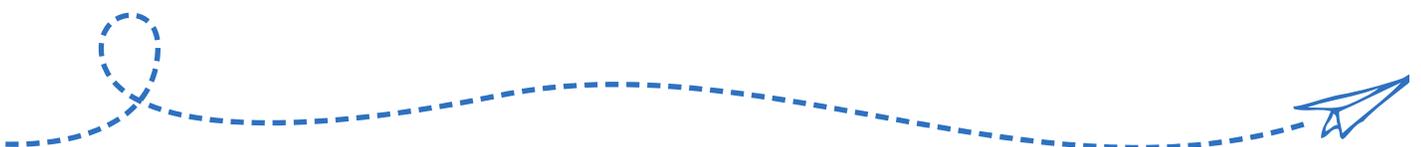
Execute quantitative/qualitative analysis and come up business insight;
Build up model to understand the relation between shipping cost and customer behavior, and identify and evaluate the business opportunity to improve buyer experience and lift site revenue.

Need 2 Business Analysts

Demand and Supply Optimizer

Most sellers on ebay, especially new sellers, are lack of knowledge of buyers demand, and also balance on supply and demand on ebay site will dramatically increase the chance of successful purchase, and provide better buyer experience and driven up site revenue. The purpose of this project is to design an intelligent platform that could identify the supply and demand status at detailed level on ebay and provide valuable info. for sellers to optimize their inventories.

Need 2 Business Analysts



PayPal ATCC

PayPal Seller & Buyer Visual Reporting System

Nowadays, big data is becoming a popular word which people have been discussed frequently. Transaction data is one of most valuable data we need to deal with. As leading company in payment industry, there are 110 million accounts from all over the world are using PayPal to do their payment both online and offline. All those transaction data are stored in PayPal environment. Those could be next treasure not only for company but also for customer who can benefit from.

PayPal Visualization Reporting System is the project which leverages those transaction data. We target to generate all kinds of visual trend reports to our customer to help them know better about their business and expense trend. By providing different dimension (transaction trend, geography related report, buyer age/gender/shopping habit related chart) reports, seller could able to adjust their business accordingly and know their competitor's strategy. From buyer perspective, they could be more interesting on social comparison and self-evaluation. Meanwhile, we also can provide on-demand report in case customer needs some advanced analysis. All those report can be presenting on web-site and mobile.

For this project, we will be building a big data environment on top of Hadoop, extract Paypal existing raw data(near real-time data) into no-sql DB; After that, we will be refining those raw data into intermediate storage(Mysql cluster) with help from hive or map/reduce framework. By prepared data in intermediate storage, we will be using a service to get request from presentation layer and return required data to UI for rendering.

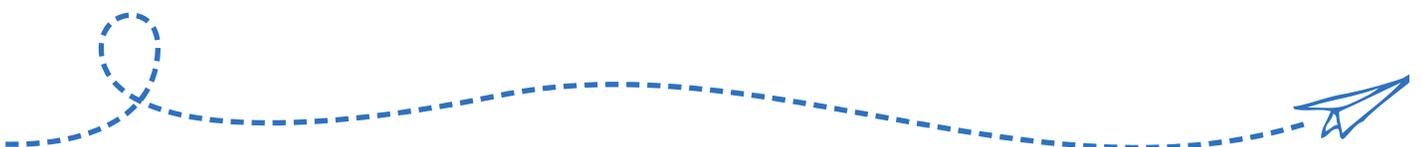
Need 8 Software Engineers

PayPal Risk Management

Online Cop – Fighting with Cybercriminal

Have you ever dream of being a cop to protect the world? Ever heard about hacker and cybercriminal? Ever imagine of fighting with online fraudsters by advanced technology? Now you have the opportunity to make your dream come true. Come and Join us!!

Risk Management is a key enabler of PayPal success and plays a critical role in providing financial security to the company and users. It is a core function responsible to detect and prevent any suspicious activities in PayPal system using cutting edge innovation in science and technology. In "Project Online Cop", you will have the opportunity to be in the front-line of war against cybercriminals and hackers using advanced



detection algorithms, sophisticated policy and working closely with industry top-notch data scientists and fraud specialists in risk management.

Join us, we will bring you business knowledge in online payment industry and risk management domain. You will get exposure to leverage big data platform (Teradata, Hadoop), detect abnormal customer behavior and pattern by advanced data mining technology (modeling/decision tree/algorithms/fashionable analytical skills), optimize risk strategy (SAS/SQL), develop tools/metrics to support anti-fraud models and provide deep dive insight through visualization reporting. What you need to equip yourself is – logical thinking, analytical mindset, intellectual curiosity, last but most importantly – your passion!!

Need 8 Business Analysts

