

## SUMMARY

I am a software engineer with interest in big data and algorithm. I enjoy building tools to simplify everyday job. I enjoy algorithm competitions and practice a lot at spare time. My Topcoder handle is *code6*.

## SKILLS

- Use Python / PHP / SQL in daily development
- Familiar with data warehouse modeling, the design of ETL flow
- Have some experience in data platform components development(scheduler, ETL tool)
- Familiar with Hive, understand of Hadoop ecosystem
- Familiar with common data structure

## EXPERIENCE

**meituan.com**, Beijing, China

*Software Engineer*

**7/2011 – present**

I'm working at **meituan data team**, I focus on data warehouse/platform tools building.

- Upgrade production hadoop version to 2.2.0 and enable hadoop security  
Duration: 11/2013 – 1/2014  
When Hadoop 2 is Generally Available, we start our plan to migrate to yarn and enable hadoop security at the meantime. In this project I'm mainly responsible for hadoop security research and deployment. I spend lots of time testing hadoop security setup process and security integration with client apps, and fixing problems caused by security. The upgrade process is very smooth and we finally start using yarn.
- Upgrade production hive version to hive 0.11  
Duration: 7/2013 – 8/2013  
Hive 0.11 has some nice features and improvements and we decide to move to hive 0.11. I'm in charge of this project. At first I maintain our productional version of hive 0.11 by patching our own commits base on hive 0.11 official release. Then I build a tool which runs automated tests with productional hive sqls in different hive versions and generates diff result. After I fix all the problems I encounter I make final deployment and hive 0.11 runs well later.
- Hive authorization and query service enhancement  
Hive authorization is not completely secure, and hiveserver(hive thrift service) is not stable. I manage to wrapper a thrift service called 'openhive' above hive cli for hive query need of our engineers. Later we open our ETL design service to our engineers, I enhance hive authorization by adding some query hook and fixing authorization bugs.
- Migrate log process from Infobright to hive  
I participate in the Infobright to Hive project and I design and carry out the ETL migration plan. I also build a small tool called diff\_checksum to check the table data consistency between the migration. This project lasts about two months and we transform all our infobright based ETLs(about 80) to hive.
- ETL tool upgrade  
Sqlweaver is our data warehouse ETL tool. At first it only support mysql, I add different database supports by adding a concept called resource handler and add support to hive and shard mysql table. I add another concept called action which enable loose ETL steps construct and code reuse, which works well for complex ETL tasks. Later I build a data synchronization tool base on sqlweaver which supports data synchronization between different type of database(mysql/hive). We use it for our data synchronization tasks.

## AWARDS

<b>Gold Medal</b> - 34th ACM ICPC Asia, Shanghai	<b>10/2009</b>
<b>Honorable Mention</b> - 34th ACM ICPC World Finals	<b>02/2010</b>
<b>Gold Medal</b> - 35th ACM ICPC Asia, Hangzhou	<b>10/2010</b>

## EDUCATION

**Fuzhou University**, Fujian, China

*Software Engineering, B.E.*

**9/2007 – 6/2011**