

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.

- 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 implement hive handler and shard mysql table handler. I add another concept called action which enable loose ETL steps construct and code reuse.
- Data synchronization Tool
I build a data synchronization tool which supports data synchronization between different type of database(mysql/hive). Users only need to fill in the source and target table info and it will do the rest. It use sqoop to transfer between hive(hadoop) and mysql.
- 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 different database. This project lasts about two months and we use hive as our mainly data warehouse infrastructure 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. Later we open our ETL design service to our engineers, I enhance hive authorization by adding some query hook and fixing authorization bugs.
- Query center Service building
Our query center service is the entrance of all queries, including report/api/client. It distributes the query to backend db server(mysql/hive) and caches the query result. I implement the asynchronous query part of the service.

AWARDS

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

EDUCATION

Fuzhou University, Fujian, China

Software Engineering, B.E.

9/2007 – 6/2011