ACTIAN

Student Program

At Actian we provide software solutions to seamlessly manage and connect our customers' operational and analytic data for superior performance, insights, and business outcomes. Our company is headquartered in the United States and has offices worldwide.

We are looking for students to join our ActianX/Vector team. Candidates will work on a cutting-edge high-performance data processing engine, which is the core of the Vector product. This is your chance to become a member of the team behind the fastest analytical database system on the market. Our focus is on high performance implementations in the database kernel and in its distributed version based on the Hadoop environment. We are looking for team players that can work independently in a distributed development team.

We offer

- INTERNSHIPS – An internship with us may last from 3 to 12 months. For each internship, we provide a tailored project to research, design and implement a new functionality into our Vector database.
- MASTER TOPICS – in coordination with your university's examination office and your collaborating professor we will define a master project tailored to your needs and based on our available thesis topics.
- PART-TIME JOBS – Part-time jobs with us provide you with the opportunity to gain your first work experience in a program related field. Your contribution will help improve an already outstanding database product. Working hours and times are flexible and can be discussed when you decide to start a project with us.

Below you will find a list of topics together with a short explanation. These topics are either marked with (I)nternship, (M)aster topic and/or (J)ob.

Vector cloud deployment.

Providing databases as a hybrid on premise and in the cloud is a promising and already growing business. Our goal is to bring an on-premise Vector partially to the cloud and within this project the task is to exploit our cloud storage architecture used in Avalanche and bring it to the Vector product. (I,M)

Load balanced query execution in a clustered environment.

Load-balancing in a cluster is hard, because normally you cannot offload work to another node if that node does not have the data to work on. However, as the HDFS integration of VectorH controls its replication policy, there are opportunities to shift work around to other cluster nodes that already have the data. This requires developing a strategy for data placement, data processing and a work shifting strategy. (I,M)

External tables in Spark

VectorH supports reading data from external data sources such as Spark. The performance of queries accessing such external tables could be improved greatly, e.g., by pushing down selections or even subtrees of the query plan into Spark. The feature could also be extended with support for more sources and data types. (I,M)

Compact hash tables.
Smaller hash tables can be significantly faster, thanks to fewer CPU cache and TLB misses. The goal of this project is to find such compact representations by bit-packing multiple columns and using dictionaries for string data. (M)

PDTs on flash.
The goal of this project is to modify our structure for differential updates (Positional delta Trees - PDTs) to expand to disk. This requires the addition of a layer that resides on disk, most likely a flash disk. The fact that PDTs are expanded to flash would make it possible to store much more updates, hence reduce the checkpoint interval (where PDT updates are merged into the main data storage structures), and lead to the system being able to sustain much higher update workloads. Current research project with TU Ilmenau DBIS(I,M)

Collations.
Understanding the current use of character sets in Ingres including the way these character sets collate data and make these rules available also for Vector. Providing performance in that case is very difficult since some mechanisms heavily rely on expanding characters before processing them. Finding cache efficient algorithms for these cases is also part of the project. As an example, consider the ASCII order for “a”, “b” and “ä”. While ASCII would order these three letters “abä”, the German language typically requires “aäb”. (I,M)

Spatial data type support.
The goal of this project is the integration of geospatial datatype support into Vector. This requires the definition of new Vector datatypes and the integration into all stages of query execution. (I,M)

Tuple layout planning.
In this project, we want to challenge the way data is stored during query processing. In principle, any mix between horizontal and vertical storage (NSM vs. DSM) can be chosen. Some columns may actually be processed in vertical vectors, while other columns are processed in a tuple layout. Horizontal storage of data inside hash tables is already supported but needs to be extended to other operators. (M)

RDF in Vector.
In principle, it should be possible to turn Vector into a highly efficient engine for RDF storage and query evaluation. This entails the storage of quads in a compressed PAX format, and a basic translation of SPARQL to SQL or even direct Vector algebra. (M)

Exploiting co-processors for Vector.
The most powerful piece of hardware in today’s average PC is the GPU, not the CPU. There have been studies how to express database operations of almost every conceivable type in GPUs. However, what is missing is a framework where complex queries consisting of many such operations could work together. (M)

Maintenance of our testing infrastructure.
For our number one scoring TPC-H experiments we need to constantly stay up to date. Test numbers for our own improvements need to be recorded and maintained. In addition, all tests and comparisons need to be kept up-o-date with our competition (Impala, Hawk, SparkSQL and Hive). (J)

Adaptation of conversion functions.

There are many built-in datatype conversion functions that are slow in comparison to an actual optimized implementation. Replacing these functions will directly impact affected queries and lead to noticeable performance improvement. (J)

Actian Germany GmbH
Weimarer Str. 1a; 98693 Ilmenau
Tel: 03677 67 85-0; Fax: 03677 67 85-23
Internet: www.actian.com

Please contact:
Steffen Harre
E-Mail: office_etc@actian.com