BigZetta Systems

Reimagining Big Data

SOLUTION BRIEF

Data Deluge,

End of Moore's Law and Way Forward

With every aspect of our living getting digitized, there is deluge of data getting generated every moment. Enterprises too are creating massive amounts of business data of their own. Data alone itself is not sufficient unless it can be converted into useful insights. Hive has emerged as a popular choice for data analysts due to its extensive support for SQL and distributed query processing over large clusters.

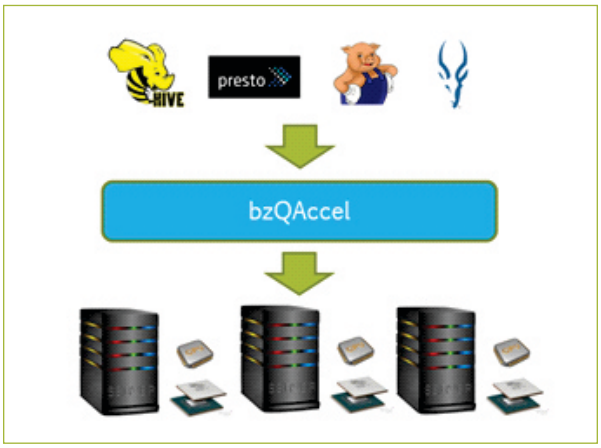
There is insatiable need for speed from analytical queries. At the same time, with Moore's law coming to an end, CPUs have hit a performance wall. Two popular ways of getting more performance out of CPU are: Scale-up and Scale-out. Many studies have clearly established that RAM/disk contention and network I/O severely limit benefits of scaling up/out beyond a point.

FPGAs have emerged as a very popular choice for accelerating variety of algorithms. With ever increasing on-chip/DDR memory sizes, improvement in data transfer speeds and ease of programming, FPGAs are well suited to take on Big Data work-loads.

Realizing the need for speeding up analytical queries, BigZetta has augmented Hive with FPGA based acceleration. With this BigZetta solution (bzQAccel) has demonstrated speed-ups of 4x (and more) over several TPCH benchmark queries.

SOLUTION OVERVIEW

BigZetta's solution, bzQAccel, acts as middleware between query processing and its execution on hardware (CPU+FPGA).



bzQAccel has been architected to provide seamless integration with Hive and is agnostic of Hive version being used. User does not need to make any changes to either query code or cluster setup. In fact, the technology can be deployed on top of a running Hadoop cluster. Only change a user would see would be faster execution of queries.

Accelerating Business Intelligence

KEY BENEFITS OF bzQAccel

- **4x or more** speedup of analytical queries
- **No changes required** to the query or existing software
- Can be deployed on top of an **existing Hive installation** (on-prem or Cloud)
- Deployment **transparent** to end user
- Extendable **to other query processing engines**

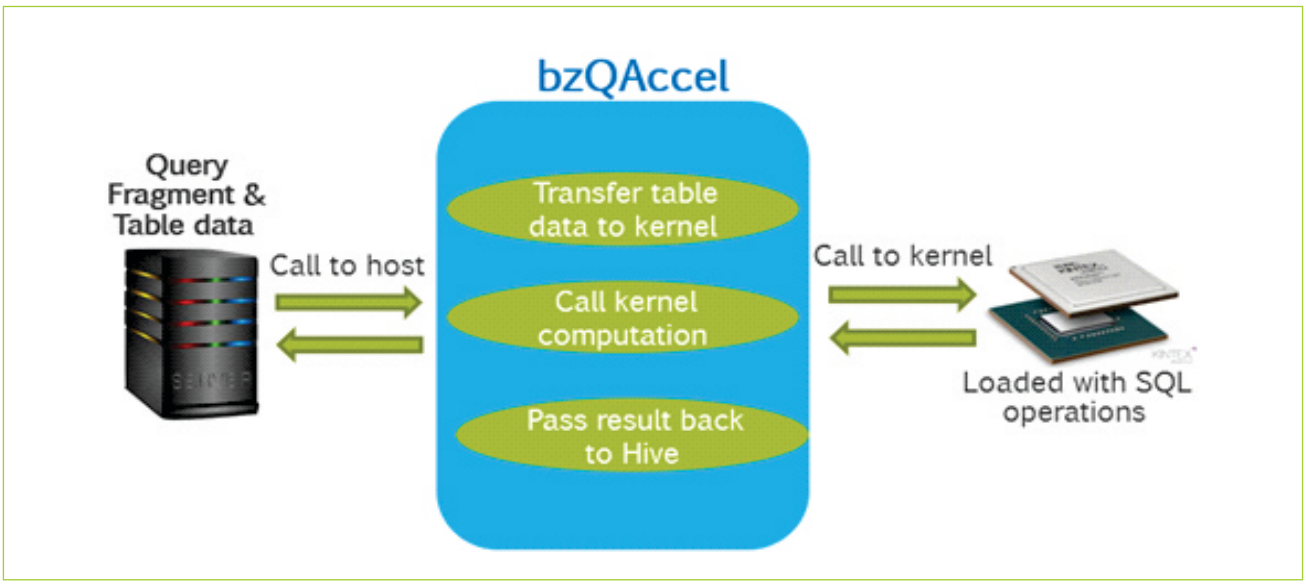
BigZetta Systems

Reimagining Big Data

SOLUTION BRIEF

SOLUTION DETAILS

The key aspect of FPGA based Hive acceleration is the interaction between the CPU (Host) and the FPGA (Device). BigZetta's Query Accelerator (bzQAccel), has been designed to manage the interaction between Hive code running on CPU and offloading of runtime critical operations onto FPGA. Figure below shows the interaction between CPU and FPGA through bzQAccel layer:



KEY FEATURES OF bzQAccel:

- Acts as **middleware** between Hive and underlying hardware
- **Optimizes** query execution plan suited for FPGAs
- Provides **fastest execution** of the plan on FPGAs
- For different queries, **no need to recompile** either the host code or the FPGA kernel
- **Minimal penalty** of data movement (from CPU to FPGA and back)

Results:

On a cluster of machines (CPU+Xilinx Alveo card), several TPCH benchmark queries were run with default Hive and one augmented with bzQAccel. FPGA accelerated Hive consistently outperformed default Hive in runtime by **4x across all the queries**.

Take the Next Step:

- bzQAccel works with Xilinx Alveo accelerators. Learn more at www.xilinx.com/alveo.
- To try bzQAccel, send email to info@bigzetta.com.
- For more information about BigZetta Systems, please visit www.bigzetta.com.