# VITESSE DATA

## FPGA enhanced MPP Data Warehouse

## CPU has peaked; FPGA has room

- 10x speed up of complex hash join
- Enable compute intensive UDF
- Integrated ML libraries

### DEEPGREEN DB on AWS F1

## INTRODUCTION

**Compute: the next barrier after storage**
Recent advances in NVMe SSD has shifted the bottleneck when it comes to data warehouse -- storage is no longer the problem; CPU is the new barrier. In addressing this shift, Vitesse Data compiles OLAP SQL query dynamically using LLVM JIT technology, resulting in improved CPU utilization and 5X improvements to TPCH Q1. Building on this foundation, Deepgreen DB performance has been amplified by utilizing the abundant compute prowess of FPGA.

## PRODUCT OVERVIEW

Vitesse Data Deepgreen DB is a scalable MPP data warehouse solution derived from the open source Greenplum Database Project and maintains 100% compatibility with the open source GPDB project. Deepgreen DB works seamlessly on local appliances or on public clouds.

AWS F1 instances can be configured with up to 8 FPGA instances for a total of 64 vCPU and 976 GB memory. These are tremendously potent nodes for MPP data warehouse provided the software can harness the power of FPGA. Vitesse Data Deepgreen DB Solution is currently available on AWS EC2 F1 instances and the Alveo™ U200 accelerator card.

## SOLUTION OVERVIEW

### Deepgreen DB with FPGA Acceleration

By carefully tuning the database engine to dispatch big chunks of compute intensive operations to Xilinx FPGAs, Deepgreen is able to reduce dependency on CPU and fully utilize the total compute power available on AWS F1. The diagram highlights the speedup of the solution compared to Greenplum DB when running the TPCH benchmark.

## Geospatial Join

UDFs (user defined functions) that require substantial compute power should also exploit FPGA. For example, the query to find all devices serviced by each cell tower is a join of a polygon area and points on a plane. Leveraging Xilinx FPGA acceleration, our Telco clients have seen more than 45X improvements to this critical query:

```
SELECT t.location, count(*)
FROM towers t, devices d
WHERE st_intersection(t.polygon, d.location)
GROUP BY t.location;
```

## In-Database Machine Learning

Using the *transducer* framework provided by Deepgreen DB, snippets of GO or PYTHON code can be interweaved into SQL queries to group and send the records to TensorFlow or other ML libraries for training or inferencing. Deepgreen DB on AWS F1 includes ML libraries that run frictionless on FPGAs. This query predicts house prices:

```
SELECT h.address,
       transducer_col_double(1) as predicted_price,
       transducer($$PHIExec go
       … go code to send data to ML inference library … $$)
FROM house h
WHERE h.state = 'CA';
```

## CONCLUSION

Deepgreen MPP with FPGA enhancements for AWS F1 or on-premise installations is built for petabyte-size data warehouse applications. Taking advantage of the abundant compute resources available on FPGA, many crucial queries can be computed swiftly, increasing the productivity of data scientists.

## TAKE THE NEXT STEP

Learn about Vitesse Data: http://vitessedata.com/
Request a trial: http://vitessedata.com/products/deepgreen-db/
Explore Vitesse Data Deepgreen DB Solution running on Xilinx Alveo[TM]: www.xilinx.com/alveo
For more information: info@vitessedata.com
Sales inquiry: sales@vitessedata.com

## ABOUT VITESSE DATA

Vitesse Data is a venture-funded big-data startup based in the Bay Area. We currently operate out of US, China, and Australia. Our main product is Deepgreen DB, a data warehouse solution that is 100% compatible with Greenplum DB. Compared to Greenplum, Deepgreen is faster, more connected, and more intelligent.

# VITESSE DATA