Trinity: End-to-End In-Database Near-Data Machine Learning Acceleration Platform for Advanced Data Analytics

Ji-Hoon Kim¹, Seunghee Han¹, Kwanghyun Park², Soo-Young Ji³ and Joo-Young Kim¹
In-DBMS Data Analytics

- Three Important yet Independent Technology Trends

**ML-based Advanced Data Analytics**
- Enterprise-level DBMS
- In-DBMS ML support

**Database HW Acceleration**
- GPU-based DBMS
- ASIC/FPGA/GPU

**Near-Data/In-Storage Processing**
- Data-intensive application
- SmartSSD/SmartNIC/HMC
Trinity: In-Database, In-Storage Platform

- Full Stack System for Near-Data-based In-DBMS ML Inference

---

### Trinity: In-Database, In-Storage Platform

<table>
<thead>
<tr>
<th>ML-based Data Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise DBMSs</td>
</tr>
<tr>
<td>MADlib</td>
</tr>
<tr>
<td>Spark ML</td>
</tr>
</tbody>
</table>

### Trinity

- **ML-based Data Analytics**
  - Enterprise DBMSs
  - MADlib
  - Spark ML

### Trinity: In-Database, In-Storage Platform

#### SmartSSD-enabled DBMS (PostgreSQL+)

- Conventional SW Stack
  - CPU Executor
- Extended SW Stack
  - MADlib

#### SmartSSD Executor (XRT Platform)

- Host code (XRT C/C++ API)
- Device code (.sv)
- XRT Linux Kernel Drive

#### SmartSSD

- NAND Flash (3.84TB)
  - PCIe Switch
- FPGA DRAM (4GB)
- FPGA HW Accelerator (i-DPA)
Trinity: In-Database, In-Storage Platform

- Full Stack System for Near-Data-based In-DBMS ML Inference

**Software DBMS**
- Data Format Converting
- Dynamic Offloading Decision

**SW-HW Interface**
- Seamless Integration of SmartSSD

**Hardware Accelerator**
- Direct Page Decoding
- Dynamic Tuple Binding
- Heterogeneous Core Arch. w/ Reconfig. On-chip Interconnect

**SmartSSD-enabled DBMS (PostgreSQL+)**
- Conventional SW Stack
- Extended SW Stack
- CPU Executor
- MADlib

**SmartSSD Executor (XRT Platform)**
- Host code (XRT C/C++ API)
- Device code (.sv)
- XRT Linux Kernel Drive

**SmartSSD**
- NAND Flash (3.84TB)
- PCIe Switch
- FPGA DRAM (4GB)
- HW Accelerator (i-DPA)

Trinity shows up to **57.18x faster** query processing speed than CPU-based DBMS
Computational Storage Device

- **New Hardware Backend: Samsung’s SmartSSD**
  - Xilinx Kintex UltraScale+ FPGA, 4GB DRAM and 3.84TB NAND flash
  - Direct FPGA-to-SSD data access using internal PCIe switch
Software Stack for Trinity

- Seamless Integration of SmartSSD in DBMS
  - Extended SW stack → extended analyzer\(^1\) + optimizer\(^2\)

1) Converting query information to the HW data format
2) Making **runtime offloading decision** to select an optimal HW backend
Extended Query Optimizer

- **Performance Cost Model**
  - Determining optimal hardware backend (SmartSSD vs CPU)
  - Showing **5.3%** & **12.97%** average error + **96% offloading accuracy**

### SmartSSD Cost Model

**Equation-based** performance model

\[ T_{\text{smartssd}} = T_{\text{host}} + T_{\text{transfer}} + T_{\text{kernel}} \]

\[ T_{\text{buffer}} + T_{\text{others}} \]

\[ T_{\text{ssd2fpga}} + T_{\text{fpga2host}} \]

### CPU Cost Model

**Regression-based** performance model

**Latency vs. Query Complexity**

**Fine-tuning**

**Error (%)**

- Hardware
- Cost model

Database Size (GB)

1

10

10000

100000

0.0001

0.001

0.01

0.1

0.5

1

1.5

1.53

2

3.47

15.33

HOTCHIPS 2022

Trinity: End-to-End In-Database Near-Data Machine Learning Acceleration Platform for Advanced Data Analytics
Overall Architecture of i-DPA*

1. **Database Page Decoder**
   - Page & data processing unit
   - Page-level parallelism

2. **Database Tuple Binder**
   - Dynamic tuple binding
   - Tuple-level parallelism

3. **Heterogeneous Core Arch.**
   - Reconfig. on-chip interconnect
   - Task-level parallelism

* i-DPA = in-Database Processing Accelerator
Database Page Decoder

- **Direct Tuple Extraction from Database Page**
  - Removing host interaction ➔ Keep BW benefit of in-storage processing
  - Faster page decoding with **page-level parallelism 😊**

---

**Diagram:**
- SSD (Database) ➔ Page Table ➔ CPU ➔ FPGA ➔ Ctrl, Reg, ALU, RF ➔ Mask info. ➔ Filtering ➔ Data Processing Unit ➔ Page Processing Unit ➔ Page Buffer ➔ Unnecessary Data Copy ➔ PCIe ➔ CH#1 ➔ CH#N

Page-level parallelism

---

**Notes:**
- HOTCHIPS 2022
- Trinity: End-to-End In-Database Near-Data Machine Learning Acceleration Platform for Advanced Data Analytics
### Dynamic Tuple Binding
- Dynamically varying tuple packing density according to the tuple size
- **Tuple-level parallelism ▲ & hardware utilization ▲ 😊**

#### Database Tuple Binder

**CASE 3**
(# of Attr >16)

**CASE 2**
(# of Attr 9-16)

**CASE 1**
(# of Attr 5-8)

**CASE 0**
(# of Attr <5)

- Packing tuple w/ zero-padding
- Increasing parallelism up to 8x
- Set proper aggregation link for final output
Query Processing Core

- Heterogeneous Core Architecture
  - Reconfigurable on-chip interconnect ➔ Enabling flexible data streaming 😊
  - Task-level parallelism across the computing units
FPGA Implementation Result

- System Setup & FPGA Implementation Result

- 2 Intel Xeon Silver 4210 CPUs
  - PostgreSQL v12.6
  - MADlib v1.17.0
- 156GB DIMM
- 3.84TB SmartSSD

Specifications

<table>
<thead>
<tr>
<th></th>
<th>FPGA</th>
<th>Freq.</th>
<th>170MHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface</td>
<td>LUT: 128665</td>
<td>FF: 183077</td>
<td>BRAM: 311.5</td>
</tr>
<tr>
<td>Core0</td>
<td>Core1</td>
<td>Others</td>
<td>Utilization</td>
</tr>
</tbody>
</table>
End-to-End Trinity Evaluation

- Evaluate Against CPU-based DBMS Platform
  - **0.85x – 57.18x** faster query processing than CPU-based DBMS
  - **15.21x** faster than CPU-based DBMS on average
Scaling-up with Multiple SmartSSDs

- **Scale-up the Overall System**
  - SmartSSD: easy to scale-out the number of devices w/ U.2 form factor
  - Deploying 4 SmartSSDs ➔ **200x faster** than CPU-based DBMS

![Diagram of SmartSSDs and CPU with PCIe Sub-system](image)

- With 2 SmartSSD ➔ **1.85x performance gain**
- With 4 SmartSSDs ➔ **3.66x performance gain**

**Linear Performance Improvement**

⇒ **Overall 200x Speedup Achieved**
Conclusion

1. Full Stack System for In-DBMS Advanced Data Analytics
   – 57.18x faster query processing than CPU-based DBMS

2. Software Stack (PostgreSQL+) for SmartSSD Integration
   – Dynamic offloading decision ➔ 96% accuracy

3. Near-Storage based Hardware Accelerator (i-DPA)
   – Direct data page decoding & abundant parallel processing (3-levels)

Trinity: In-Database, Near-Data Machine Learning Acceleration Platform for advanced Data Analytics
Thank You!

- Questions? Feel Free to Contact Me!
  - E-mail: jihoon0708@kaist.ac.kr
  - LinkedIn: https://www.linkedin.com/in/ji-hoon-kim-860429225/
  - Slack: Trinity: In-Database Near-Data ML Acceleration platform