How GPUs Enable XVA Pricing and Risk Calculations for Risk Aggregation

James Mesney | Principal Solutions Engineer | jmesney@kinetica.com
Setting the Scene
What is XVA?

X-Value Adjustment (XVA) refers to “Valuation Adjustments” in relation to derivative instruments held by banks.

The “X” in XVA means “C” for credit, “D” for debt, “F” for funding, “K” for capital...

“Doing” XVA is Risk Modelling.

It’s all about computing potential RISK, now and in the future.

OBJECTIVE: INSULATE THE BANK FROM RISK WHEREEVER POSSIBLE AND ALLOCATE THE RIGHT CAPITAL
Why is XVA Needed?

Pre 2007 - Trades cleared at fair valuations

- Costs for capital and collateral were irrelevant to banks’ investment decisions
- Increasingly complex investment portfolios emerge of variable value

Post 2007 – MAJOR REFORM!! *Counterparty Credit Risk and Basel III Accord*

Elevated Capitalisation + Collateralised Trades = EXPENSE & REDUCED PROFITABILITY!

How can banks operate efficiently AND play by the rules?

**ANSWER=LOTS OF DATA + CLEVER FORECASTING MODELS, LOTS OF COMPUTE**
XVA Challenges

• To Compute Risk
  • Continually and comprehensively measure trading activity, currency movements
  • Consistently and timely – at trading speed... Batch processing is no longer satisfactory
  • 100,000s or millions of trades per day
  • 10,000s of counterparties... in 20+ currencies!

• Calculations – Complex and Time Critical
  • Compute intensive workloads are well suited for the GPU
    • REAL TIME adjustment calculations are very computationally intensive
    • Monte Carlo Simulation
Life After Moore’s Law
40 Years of Microprocessor Trend Data

Transistors (thousands)

1.5X per year

1.1X per year

SpecINT

Single-threaded perf

Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten New plot and data collected for 2010-2015 by K. Rupp
The Rise of GPU Computing

- GPU-Computing perf: 1.5X per year
- SpecINT: 1000X
- By 2025

Single-threaded perf: 1.5X per year

The CPU Bottleneck

With limited processing gains on the horizon, CPUs are further and further behind the growth in data.
GPU Acceleration Overcomes Processing Bottleneck

4,000+ cores per device versus ~16 cores per typical CPU

High performance computing trend to using GPU’s to solve massive processing challenges

Parallel processing is ideal for scanning entire dataset & brute force compute

GPU acceleration brings high performance compute to commodity hardware
Deploying XVA
XVA Architecture

ETL / STREAM PROCESSING

PARALLEL INGEST

ON DEMAND SCALE OUT

Server 'n'

Server 1

In-Database Processing

SQL

Python

Java

C++

CUDA

BIDMach

Caffe

ML Libs

BI / VISUALISATION

Native APIs

SQL

Export

Custom Connectors

BI DASHBOARDS

Tableau

Power BI

CUSTOM APPS

+ DOWNSTREAM APPS

Live Trading data
Counterparties
Options
Currencies
Futures
Market
Foreign Exchange
Bloomberg
Reuters

SAP
Teradata
Oracle
MapR
Hortonworks
Cloudera
IBM

Apache Storm
nifi
ActiveMQ
Informatica
TIBCO

Batch & Streaming Data

XVA Model
Kinetica’s XVA Deployment at Major Multinational Bank

Large financial institution moves counterparty risk analysis from overnight to real-time.

• Data collected and fed to XVA library which computes risk metrics for each trade

• XVA outputs stored in Kinetica database

• Flexible real-time monitoring by traders, auditors and management

• Data retained for historic analyses, machine-learning...
Three formidable forces—a **weak global economy, digitization, and regulation**—threaten to significantly lower profits by as much as $90B for the global banking industry over the next three years.

Financial Services enterprises must reinvent their business by transforming the core—

**Resilience, Reorientation, Renewal**

**Kinetica** for:
- Resilience – Manage Revenue, Costs, Capital, and Risks
- Reorient – Customer-centricity, Digitization, Open Bank
- Renewal - New Markets, Products, Customers

*Source: McKinsey & Co, PwC*
Kinetica: A Distributed, In-Memory, GPU Database

- GPU-accelerated database operations
- Natural language processing based full-text search
- Native GIS and IP-address object support
- Real time data handlers to ingest structured and unstructured data

- No typical tuning, indexing, and tweaking
- Predictable scale out for data ingestion and querying
- Deep integration with open source and commercial frameworks and applications: TensorFlow, Hadoop, Spark, NiFi, Storm, Kafka, Tableau, Kibana and Caravel
- Distributed visualization pipeline built in
Kinetica Enabling Broad Enterprise Solutions

**CROSS INDUSTRY**
- Real-Time Analytics
- Converge AI & BI
- Location-Based Analytics
- IoT Analytics

**RETAIL/CPG**
- Omni-Channel
- Customer Experience
- Supply Chain Optimization
- Targeted Marketing

**FINANCIAL SERVICES**
- Risk Modeling
- Financial Crimes
- Compliance
- Customer Experience

**UTILITIES**
- Smart Meters
- Smart Grid Optimization
- Infrastructure Modernization

**HEALTHCARE**
- Drug Development
- Precision Medicine
- Patient 360

**MEDIA/ENTERTAINMENT**
- Sentiment Analytics
- Recommendation Engines
- Ad Targeting

**COMMUNICATIONS**
- Customer Churn
- Network Optimization
- Content Targeting

**TRAVEL**
- Price Optimization
- Customer Experience
- Equipment Maintenance
Kinetica: Unique Strengths & Capabilities

Taking advantage of the parallel nature of the GPU Kinetica delivers low-latency, high performance analytics on large and steaming data sets.

User-defined functions (UDFs) allow for distributed custom compute directly from within the database.

Simultaneously ingest, explore, analyze, and visualize data within milliseconds to make critical decisions.

Easier to work with large geospatial data sets.

Fast, Distributed Database Engine

In Database Analytics

Supercharge BI

Native Geospatial & Visualization Pipeline
Thank You!

Come get your Kinetica t-shirt and copy of the new O’Reilly book at booth G.01!

James Mesney | Principal Solutions Engineer | jmesney@kinetica.com