FOUNTAINHEAD

Supercomputing with Excel

Making Excel models run faster

Microsoft Offices, New York, 24 January 2011

Microsoft

FOUNTAINHEAD

Outline for the Talk

Part A ~ Making Excel Run Faster

1. Introduction.
2. Excel / VBA models.
3. Excel add-ins (UDFs) and real-time data (RTD).
4. Excel and external tools.

Part B ~ HPC/GPU Supercomputing with Excel

5. Running Excel with GPU.
6. Running Excel in parallel with Windows HPC (by Bill Scheel).
7. Questions & answers.
8. Resources.

FOUNTAINHEAD

Talk Admin

- Introducing Bill Scheel.
- Talk materials, models and code available online & extra resources.
- Vendors in the room: Microsoft (obviously!), Nvidia & others.
- Questions and answers ~ as we go along, and at the end.
- 10 day post-talk discussion forum.
- On-the-spot prize draw for the best question of the session.
- On-the-spot prize draw for handing in the questionnaire.
- Future events:
    - HPC / GPU enthusiast meetup 6pm tonight.
    - 3-day training course, 28-30 March 2011: “Microsoft HPC for Finance”.
    - Additional “brown bag” lunchtime talks.

FOUNTAINHEAD

1. Introduction

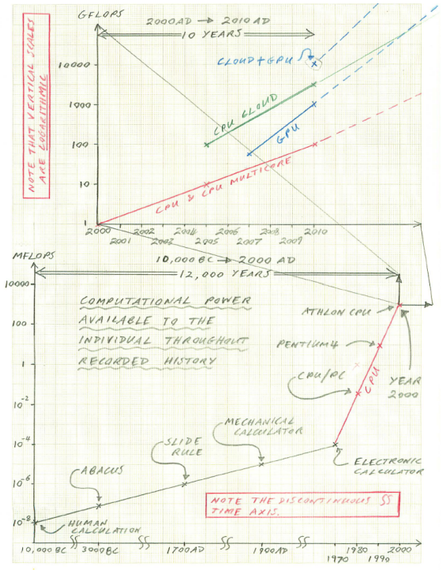
FOUNTAINHEAD

It’s not just about the answers, it’s about the questions too!

- Speed of execution.
- Speed of iteration (equally important).
- Experience (learning by doing).
- Expertise (getting to “expert” more quickly).
- Confidence and the tools to attack more challenging problems.
- … HPC / GPU as the way to get you there … quickly!

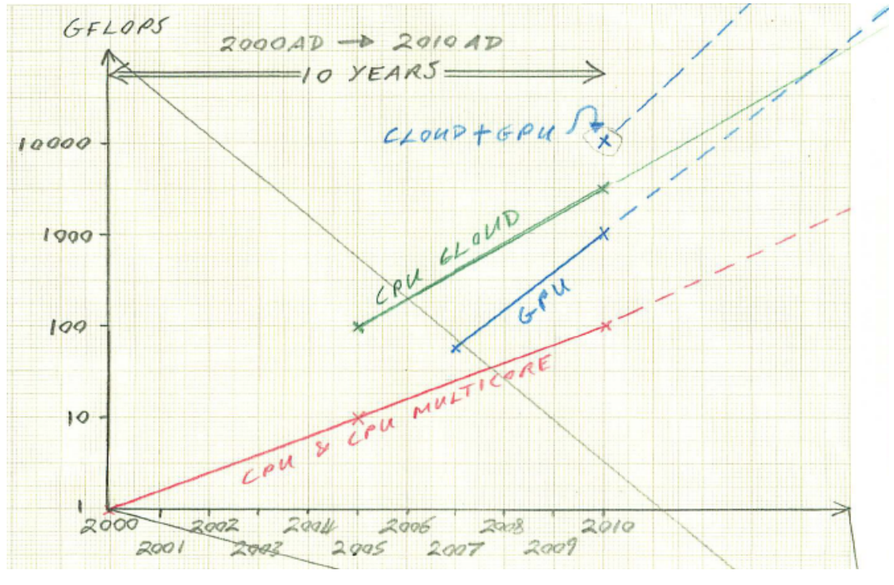
FOUNTAINHEAD

Computational power throughout history



FOUNTAINHEAD

Computational power ~ let's zoom in to the last 10 years



FOUNTAINHEAD

Excel as the front end of choice in finance

• From the trading office to the back office, Excel is the de facto front end of choice for building financial models and applications. And for some very good reasons!
• Ubiquitous.
• Familiar, with no new learning curve on the part of the user.
• Graphics, data analysis tools, and other goodies bundled for free.
• Flexible.
• Extensible.
• Integrated (with Office, Visual Studio, … and 3rd party products).
• High performance, especially with Windows HPC.

FOUNTAINHEAD

2. Excel / VBA Models

FOUNTAINHEAD

Two Kinds of Excel Model

- “F9” (recalculate) spreadsheet models.
- VBA models.

FOUNTAINHEAD

Excel Versions

	Excel 2003	Excel 2007	Excel 2010
32-bits	✓	✓	✓
64-bits	✗	✗	✓
Multithread	✗	✓	✓
Max rows	65,536	1,048,576	1,048,576
Max cols	256	16,384	16,384
Memory	1GB	2GB	2GB / 8TB

FOUNTAINHEAD

Speeding up “F9” models

- By “F9” model I mean a spreadsheet that uses only functions in cells.
- Simply rules of thumb can significantly improve performance:
    - Switch off auto-calculation and use manual-calculation (F9).
    - Use a different function that is faster, but gets the same result.
- Avoid “volatile” functions.
- Forward & backward referencing.
- Avoid circular references (iteration).
- Minimize links between worksheets, and workbooks.
- Minimize used range.
- Speed up lookups.
- … and a whole bunch more available in the talk “resources”.
- And one last tip to help speed up your development of models and general use of Excel - learn Excel’s keyboard short-cuts!

FOUNTAINHEAD

Speeding up VBA models

• Visual Basic for Applications (VBA) is Excel's built in (interpreted) scripting language.
• By coding carefully in VBA, you can speed things up quite a lot. Here are some tips:
    • Switch screen updating off.
    • Turn off the status bar.
    • Use manual calculation.
    • Disable events.
    • Ignore page breaks.
    • Read and write data in big blocks, not small ones.
    • Avoid selecting and activating objects.
    • Use explicit data types (Option Explicit).
    • … and a whole bunch more available in the talk "resources".

FOUNTAINHEAD

3. Excel UDFs and RTDs

FOUNTAINHEAD

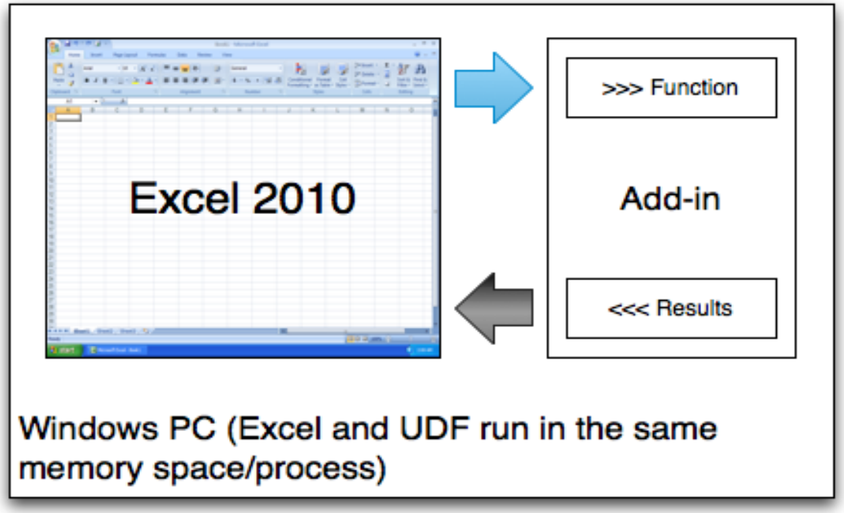
Excel UDFs

- User Defined Functions (UDFs) are custom functions that look and behave in a similar fashion to Excel's own built in functions.
- UDFs are Excel add-ins and are of two types:
  - XLL (really just a DLL with a few pre-defined entry point functions).
  - Managed code (.NET/COM) add-ins.
- All can easily be built seamlessly done with Visual Studio 2010.

FOUNTAINHEAD

Excel Add-ins - Overview

Excel add-ins (XLLs and .NET/COM) provide Excel with UDFs that look, and behave, like Excel's built-in functions.

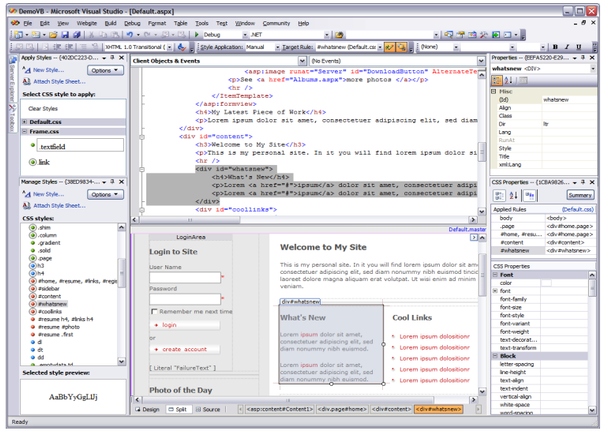


Windows PC (Excel and UDF run in the same memory space/process)

FOUNTAINHEAD

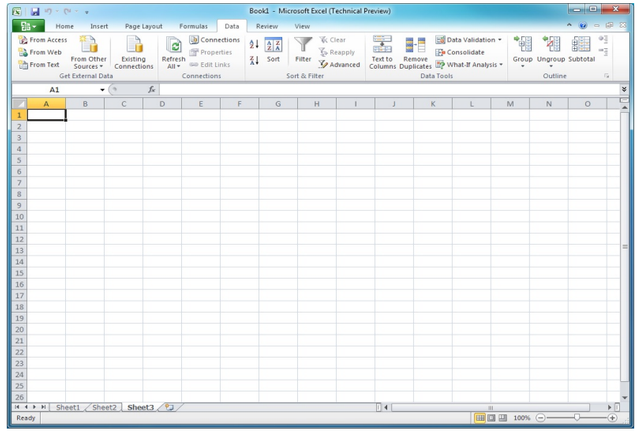
Demo: Building an add-in

Visual Studio



+

Excel



FOUNTAINHEAD

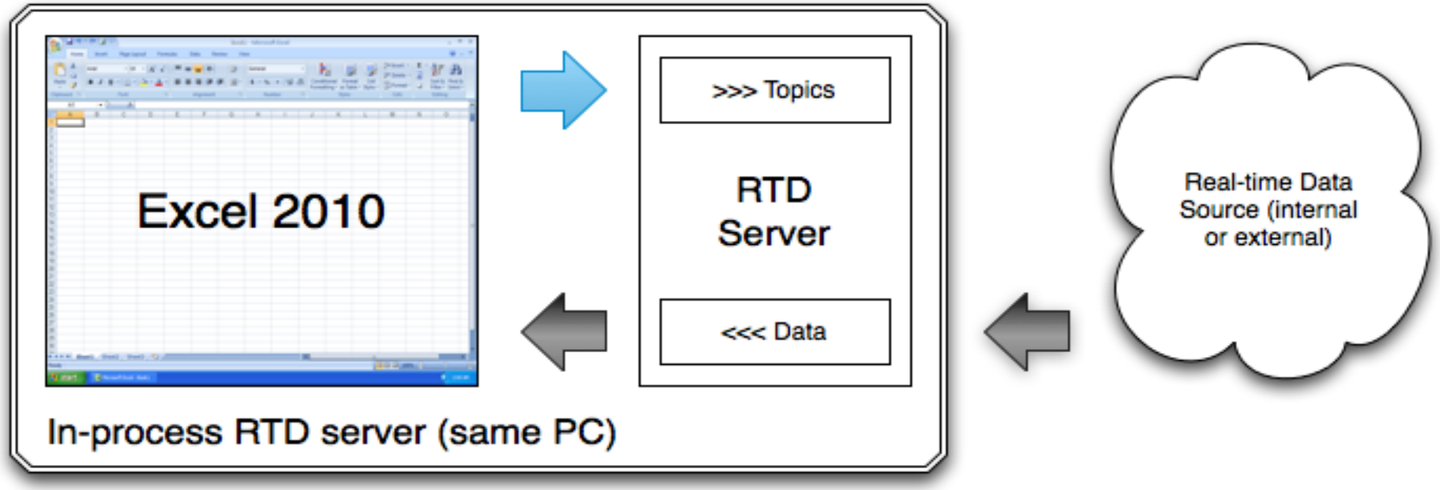
Excel RTD

- Real-time Data (RTD) server (introduced in Excel 2002).
- RTD as in-process (local) and out-of-process (remote) server.
- RTD as an added-value data server (e.g. volatility surface).
- RTD as a compute server (ask, and you shall receive).

FOUNTAINHEAD

Excel RTD - in-process

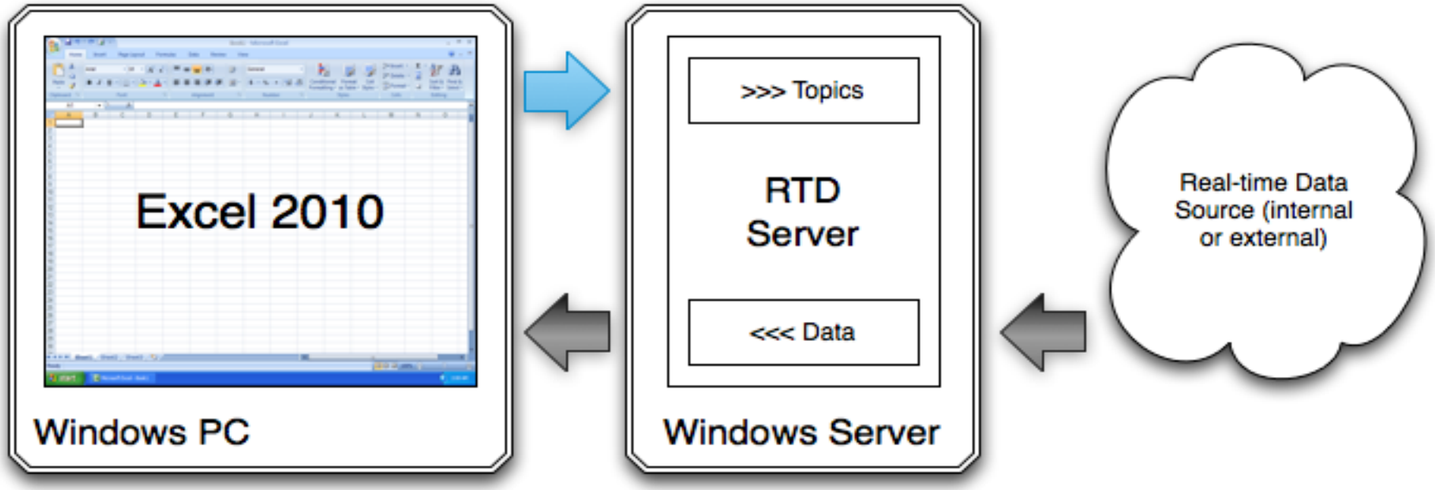
In-process RTD server runs in the same memory space/process as Excel on the same Windows PC:



FOUNTAINHEAD

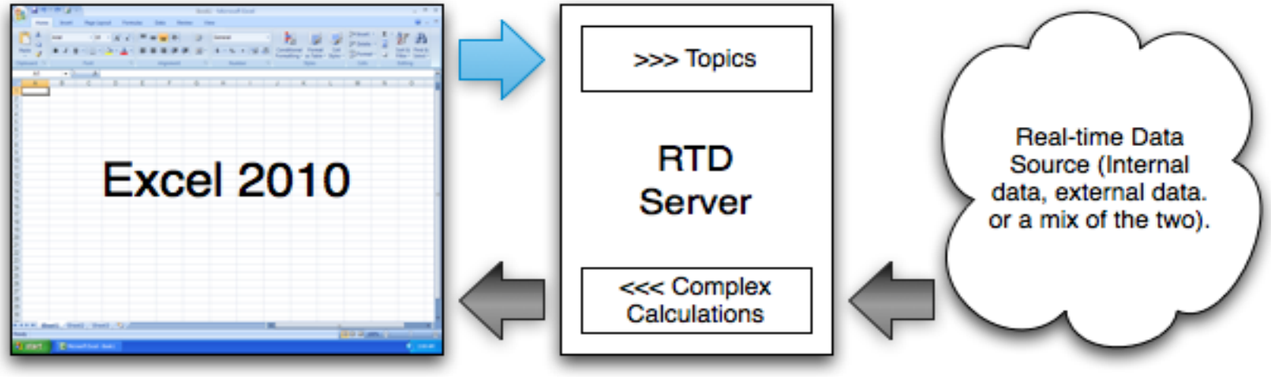
Excel RTD - out-of-process

Out-of-process RTD server runs on a separate Windows server and data is transferred to and fro across the network:



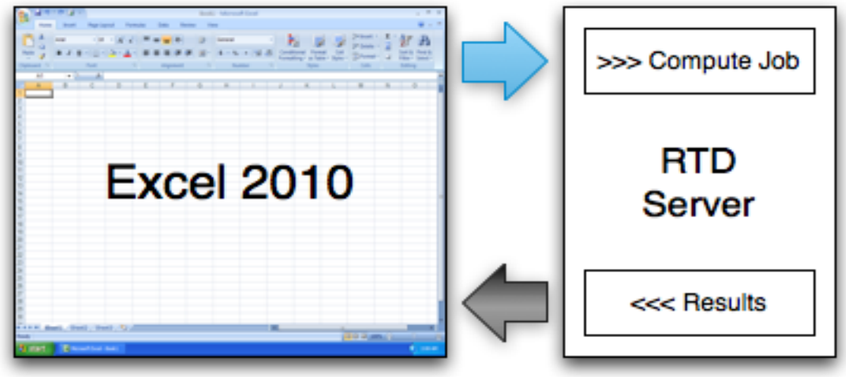
FOUNTAINHEAD

Excel RTD - added-value data



FOUNTAINHEAD

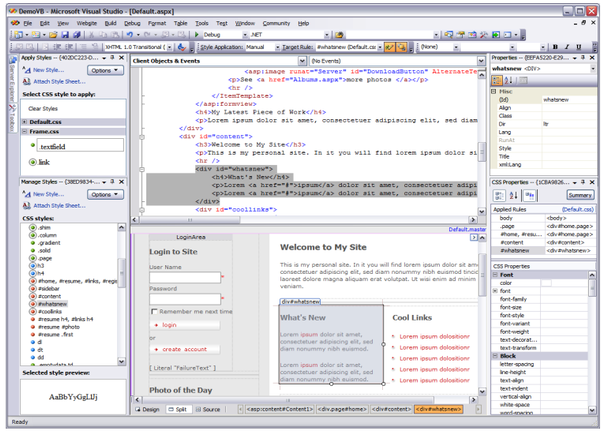
Excel RTD - calculation Engine



FOUNTAINHEAD

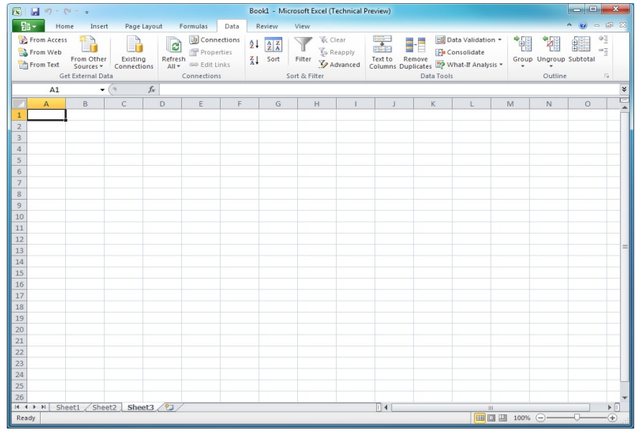
Demo: Building an RTD

Visual Studio



+

Excel



FOUNTAINHEAD

4. Excel and external tools

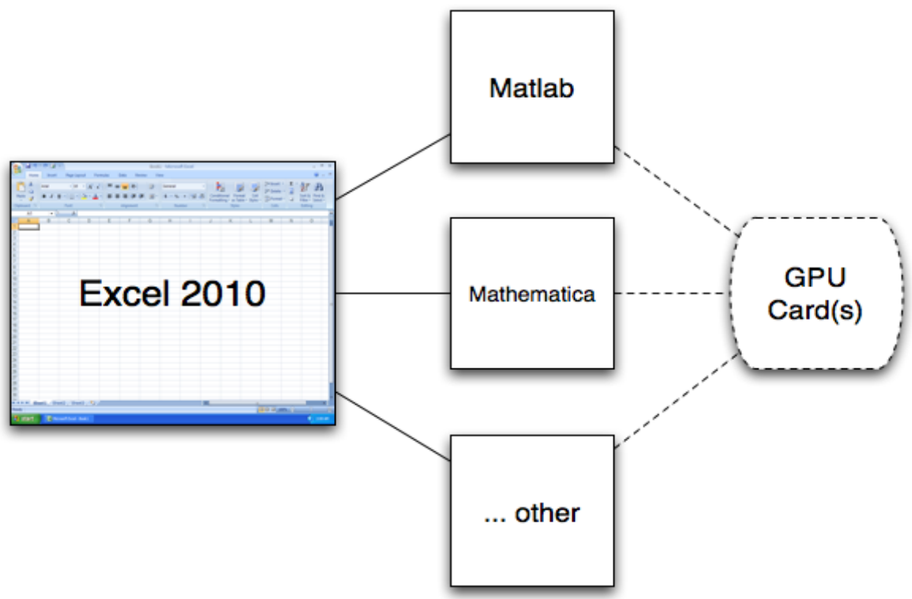
FOUNTAINHEAD

Excel external tools

- Excel can integrate with several 3rd party tools:
  - Matlab (commercial).
  - Mathematica (commercial).
  - R Statistics Package (open source).
  - … and others.
- And some of these tools are now GPU enabled! So that is a very easy way to gain access to GPU supercomputing power.

FOUNTAINHEAD

Excel 3rd party tool integration



FOUNTAINHEAD

Excel and web services

• Excel can make use of web services in the Cloud (for example, on Azure) for both pulling in data (static and real-time) and for offloading computations.
• Accessing web services (via SOAP and REST) from Excel is incredibly easy, both from within a spreadsheet and from VBA code.
• Execution speed can be improved through Excel Calculation Services, which also supports UDFs.

FOUNTAINHEAD

5. Running Excel with GPU

FOUNTAINHEAD

GPU supercomputing

- This slide is for those of you who don't know much about "GPU".
- GPU (Graphics Processing Unit).
- GPGPU (General Purpose GPU programming).
- Massively-parallel.
- Fantastic floating point performance (~1TFlop SP, ~500GF DP).
- For comparison, CPU ~100GFlop.
- GPU memory ~6GB. ECC memory.
- Speedup x10 to x1000.
- CUDA, OpenCL and DirectCompute APIs.

FOUNTAINHEAD

GPU architecture

FOUNTAINHEAD

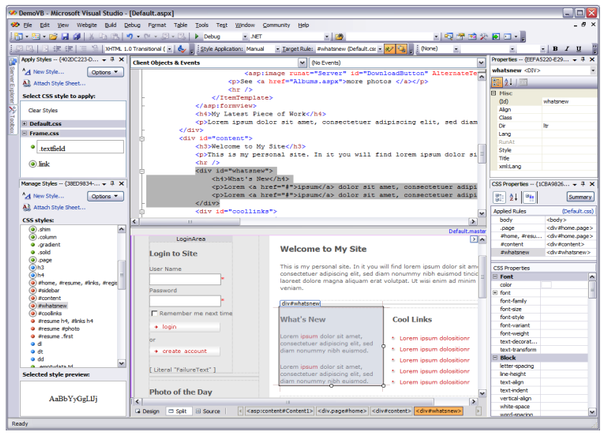
Thinking in parallel

• One of the big challenges of using GPU (and indeed multi-core in general) is “thinking in parallel”.
• By this I mean taking a existing or new application and mapping its architecture and algorithms to best fit GPU. That requires new ways of thinking.
• Some problems are a better fit for parallel programming than others.
• So you first have to figure out whether your problem will benefit from GPU before making the move. But, be creative in how you use GPUs:
    • Use existing libraries of functions.
    • Pre-computation.
    • Lookup tables.
    • Approximations.
    • Example: Fibonnaci series.

FOUNTAINHEAD

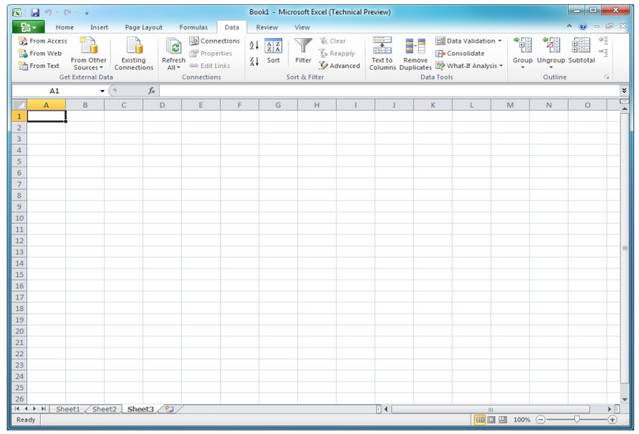
Demo: Calling GPU Functions

Visual Studio



+

Excel



+ GPU/CUDA

FOUNTAINHEAD

6. Running Excel in parallel

FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

‘Dynamo’ Public Access HPC Excel 2010 Dynamic Financial Analysis Model for Property & Casualty Insurance Company

Feature	Description
Probability Dns	750 inverse log normal and beta
Dns for Accounting Vars	64 captured; more calculated
Parallelization Type	F9-er
Conversion Effort	3-5 days
Interesting HPC Consideration	Reverse scenario analysis and replication on a cluster

FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Gaining Modeling Performance Using HPC Excel 2010

Operation (Simulation Partitions)	Microsoft HPC Cluster (approx 225 cores)	Local HPC Cluster (approx 20 Cores)
Standalone 50K Cluster 50K	37.83 mins 3.12 mins	37.45 mins 4.87 mins
50K 75K	3.12 mins 3.77 mins	4.15 mins 9.28 mins
250K 500K 1 mil	8.77 min 15.28 mins 29.6 mins	27.0 mins 1.08 hrs ???

FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Gaining Modeling Precision Using HPC Excel 2010

5th Year Surplus .01 Percentile

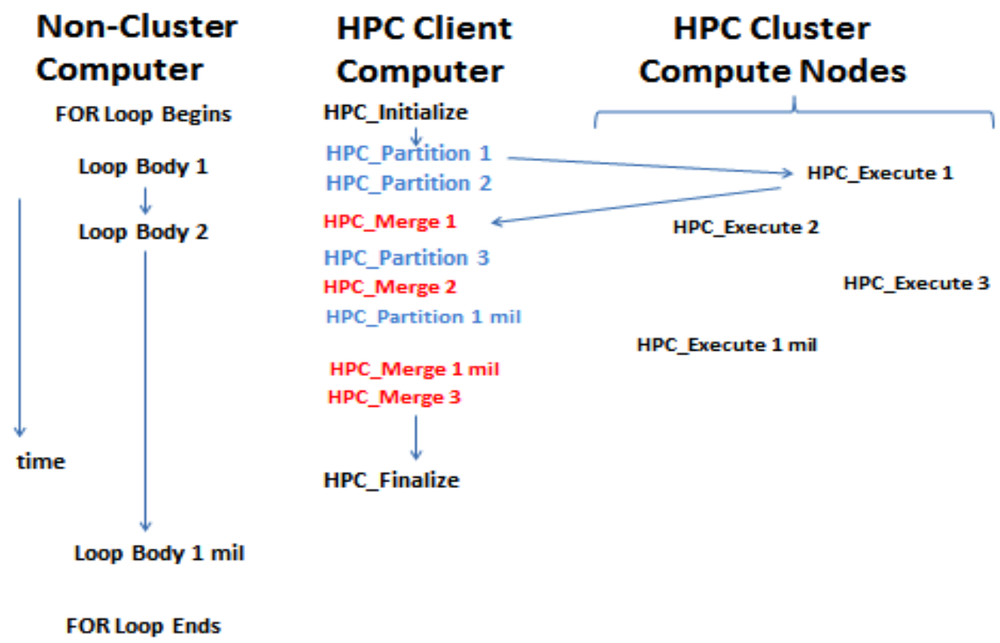
Simulations (Partitions on Compute Nodes)		
5K	25K	100K
14,911	-13,755	9,317
13,727	4,101	8,896
15,905	15,421	12,989
14,189	14,248	7,827
15,065	15,878	3,002

Only when more extreme observations occur during a 100K simulation do we see that we must lower surplus expectations given the business plan.

FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

HPC Excel 2010 Deconstructed VBA Event-Driven For Loop



FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

The HPC Launch

```
Set HPCExcelClient = New ExcelClient
ThisWorkbook.SaveCopyAs sPath
With HPCExcelClient
  .Initialize ThisWorkbook
  .OpenSession sHeadNode, sPath, _
ws.Range(MinResources).Value, _
ws.Range(MaxResources).Value, _
ws.Range(ResourceType).Value
  .Run CalculateOnDesktop
End With
```

FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

(HPC_Initialize)
HPC_Partition

```
Public Function HPC_Partition() As Variant
    If nPartitionCnt + 1 <= numsims
       nPartitionCnt = nPartitionCnt + 1
       HPC_Partition = CreateMsg()
    Else
       HPC_Partition = Null
    End If
    End Function
```

FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

HPC_Execute (compute node Excel instance)

```
Public Function HPC_Execute(data As Variant) As Variant
  'recover in-bound data
  Set ws = ThisWorkbook.Sheets(wsSimulationData)
  Set r = ws.Range(TrialNo)
  nTrial = data(1)

  ReDim data(5)    'data format outbound
  data(1) = ExecuteError.Undefined
  data(3) = nTrial
  data(5) = ComputerName  'compute node

  PartialSimulation nTrial, nEr            ‘basically app.calculate
  If nEr = MyError.NoError Then
    Set r = .Range(StartData).Resize( etc.)
    data(4) = r.Value
  End If
End Function
```

FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

HPC_Merge

```
Public Function HPC_Merge(data As Variant)
  nMergeCnt = nMergeCnt + 1
  ReDim vD(3) As Variant
  vD(1) = nMergeCnt
  vD(2) = Timer
  vD(3) = data
  vMerge(nMergeCnt) = vD
  ‘vMerge global redim’d HPC_Initialize
End Function
```

FOUNTAINHEAD

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

HPC_Finalize

```
Public Function HPC_Finalize()
  'looping over contents of vMerge, 1,2,…,i,….
  vD = vMerge(i)
  nMergeCnt = vD(1)
  dtT = vD(2)
  data = vD(3)
    …
  nTrial = data(3)
  vD = data(4)  'results for a trial
    …
  r(1, 0).Value = nTrial
  r(1, 1).Resize(1, UBound(vD, 2)).Value = vD
End Sub
```

FOUNTAINHEAD

7. Questions & answers

FOUNTAINHEAD

8. Resources

• Contact Andrew Sheppard: **shep@FountainheadTraining.com**
• The presentation materials for the talk, code and additional resources are freely available and can be downloaded from Github.com by searching for the project: **Rocketboost-Kickstart**
• In addition, this talk comes with “10 days of free support”, so to speak, in that you can ask questions relating to the talk and get answers. The support is provided via the “Supercomputing for Finance” LinkedIn discussion group. You can sign up for free at:
http://www.linkedin.com/groups?gid=3673882

FOUNTAINHEAD

Speaker: Andrew Sheppard

Andrew Sheppard is a financial consultant with extensive experience in quantitative financial analysis, trading-desk software development, and technical management. Most recently, from 2006 to 2010, Andrew worked at a New York multi-strategy hedge fund. He also was the manager of an innovative software company based in London that was owned by the hedge fund but run independently. For more than two years, Andrew has been an active developer of GPU (CUDA) massively parallel software in C/C++ for real-time financial trading and risk. Andrew is also the author of the forthcoming book "Programming GPUs”, to be published by O’Reilly (www.oreilly.com).

Speaker: William Scheel, Ph.D.

Bill is a consultant with Capgemini currently working in the area of Solvency II, a regulatory regime promulgating the methodologies and approving the internal models used to measure the adequacy of insurance company surplus. During the last couple of decades, Bill has designed systems for actuarial, accounting, hedge fund and commercial software companies. He served as consultant to the U.S. Department of Labor, the Federal Trade Commission, the U.S. Justice Department and the National Association of Insurance Commissioners. Bill was a Professor of Risk and Insurance for over a decade.