Visual Domain Specific Languages for Actuarial Models: An Industrial Experience Report

Workshop on Domain Specific Languages for Financial Systems

ACM/IEEE 16th International Conference on Model Driven Engineering Languages and Systems (MODELS 2013), Miami, FL

Aon Benfield Securities, Inc.
Annuity Solutions Group (ASG)

October 1, 2013

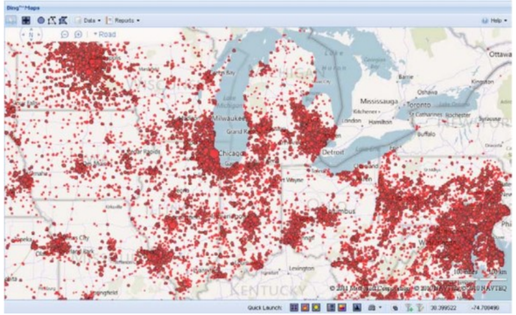
AON BENFIELD

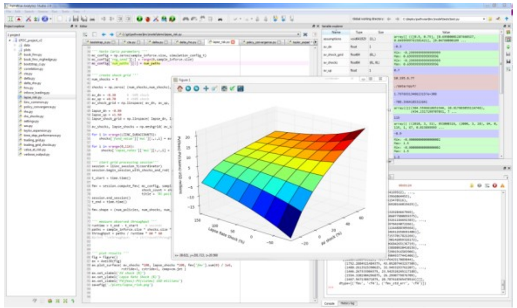
AON BENFIELD

Aon Benfield

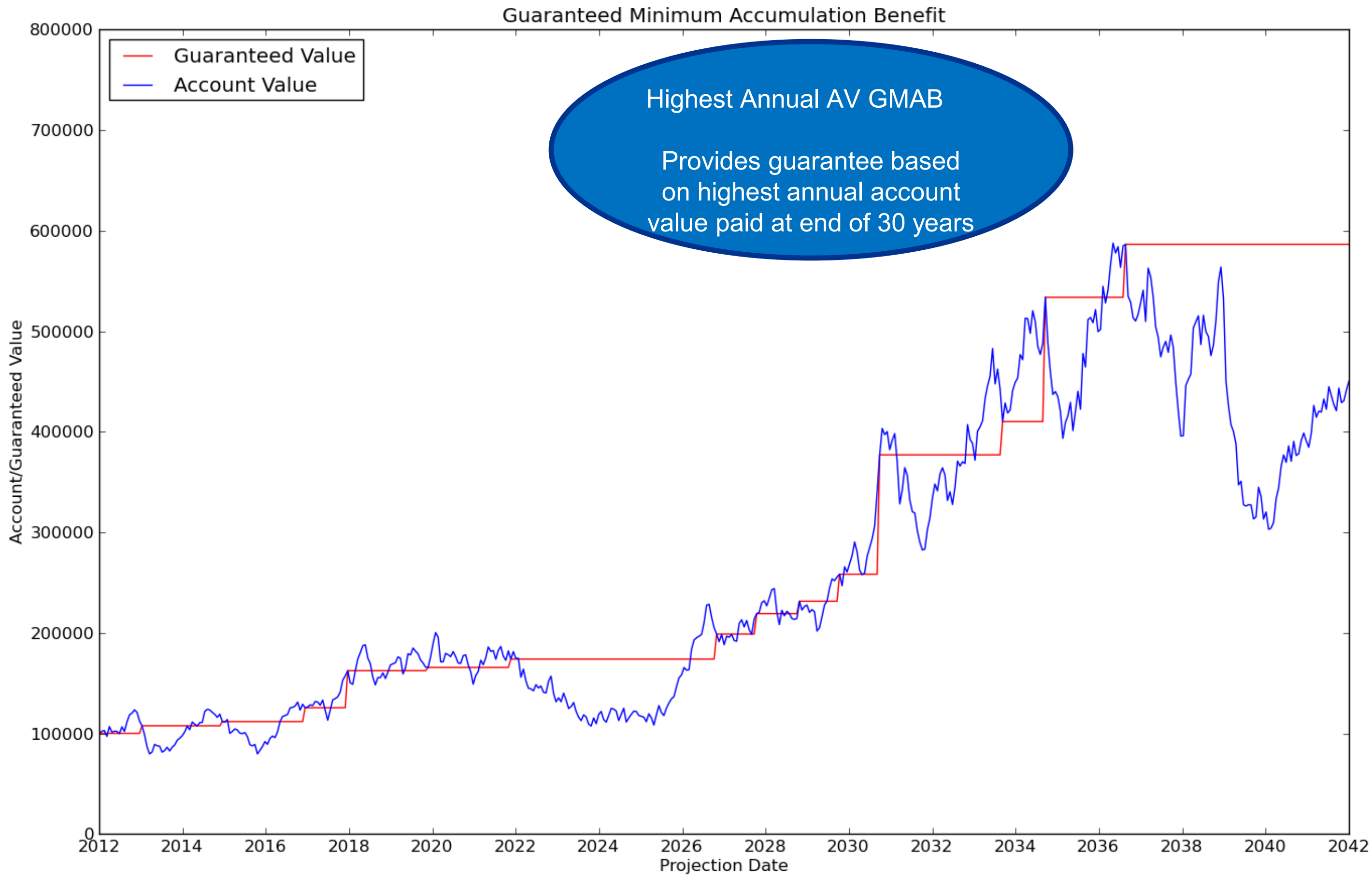
- **Aon Benfield**, a division of Aon plc (NYSE: AON), is the world's leading reinsurance intermediary and full-service capital advisor. We empower our clients to better understand, manage and transfer risk through innovative solutions and personalized access to all forms of global reinsurance capital across treaty, facultative and capital markets. As a trusted advocate, we deliver local reach to the world's markets, and an unparalleled investment in innovative analytics. With more than 80 offices in 50 countries, our worldwide client base has access to the broadest portfolio of integrated capital solutions and services.

- **Aon Benfield Analytics**
  - Aon Benfield Analytics offers clients industry-leading catastrophe management, actuarial, rating agency advisory and risk and capital strategy expertise.
  - Sample risk analytics products
    - **PathWise**
    - ReMetrica
    - ImpactOnDemand
    - CatScore





Industry Overview – Variable Annuities

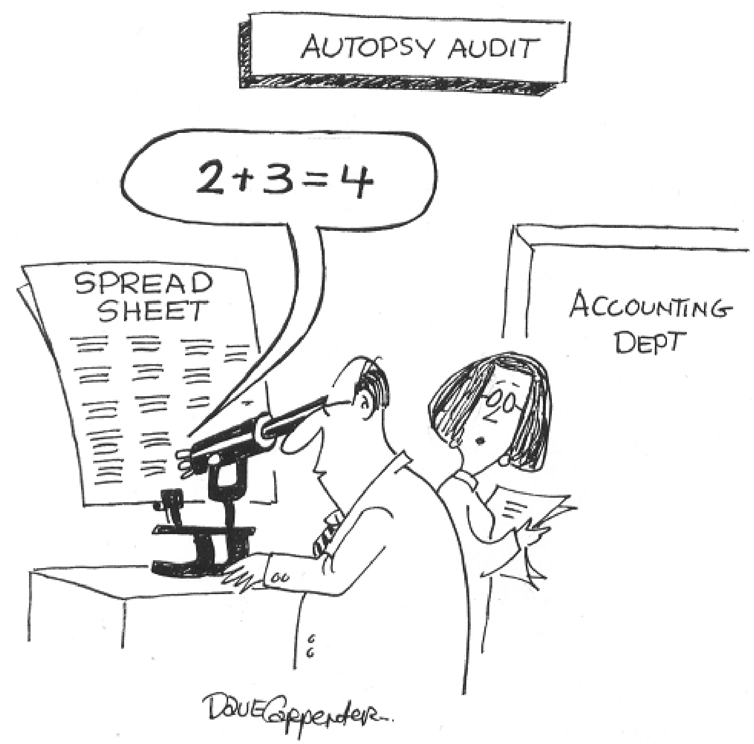


Industry Computational Challenges

- **Business end-users focus**
  - Users are Quantitative Analysts, Actuaries, Traders, Risk Managers, etc
  - The right tools must focus on the end-user requirements
- **Business logic and systems code must be continually adapted to changes**
  - Changing models, financial products, market conditions, and regulatory requirements
  - Changing technologies (Multi-Core, Cell Broadband Engine, GPUs, etc)
- **High Computational Throughput is required**
  - Large-scale real-time Monte Carlo simulations (Support Hedging Programs)
  - Nested simulations (Hedging Back Testing, Capital, Valuation)
  - High end-user productivity (not waiting for huge runs to complete)
- **Mission Critical Operations**
  - The intended use of such systems is mission critical
  - System failures or bugs can be catastrophic for business users
  - Automation and auditability are very important issues

Industry Computational Challenges

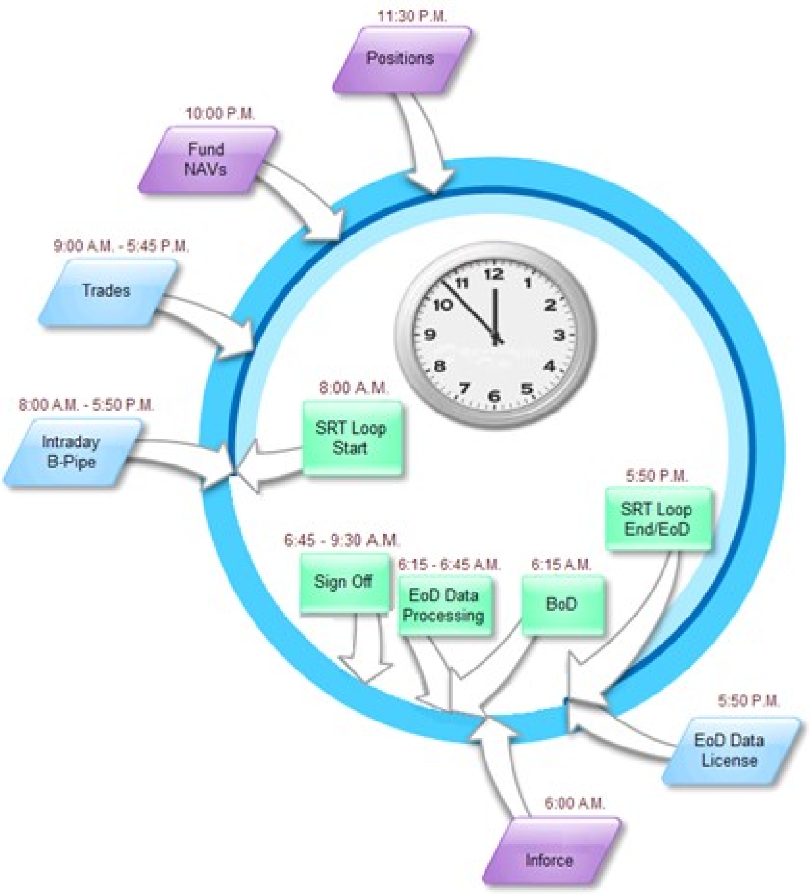
- **Business logic and systems code must be continually adapted to changes**
  - Change is constant
    - Financial modeling innovation
    - Financial products innovation
    - Evolving market conditions
    - Changing regulatory requirements
    - Technological innovation
  - Traditional approaches
    - Enterprise IT systems slow to adapt
    - Shadow IT systems fill the gaps – patchwork of end-user developed, manually operated spreadsheets (potentially thousands of interlinked spreadsheets)
    - Slow, costly, error-prone



"There it is! I've isolated the origin of the firm's demise."

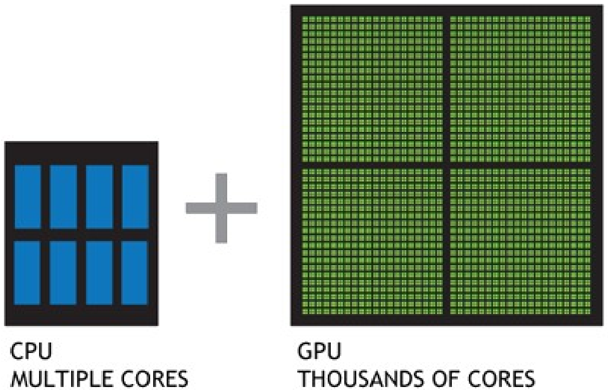
Industry Computational Challenges

- **Mission Critical Operations**
  - **Requirements**
    - High performance, integrated real-time analytics
    - Complex business data-flow management
    - Job scheduling
    - Fault tolerance / failover
    - Operational workflows
    - Reporting presentation layers
    - Audit trails
    - Monitoring and Error Reporting



GPU Computing

- GPU (Graphics Processing Units) are specialized processors that can be used to speed-up parallel computing problems, such as Monte Carlo simulation



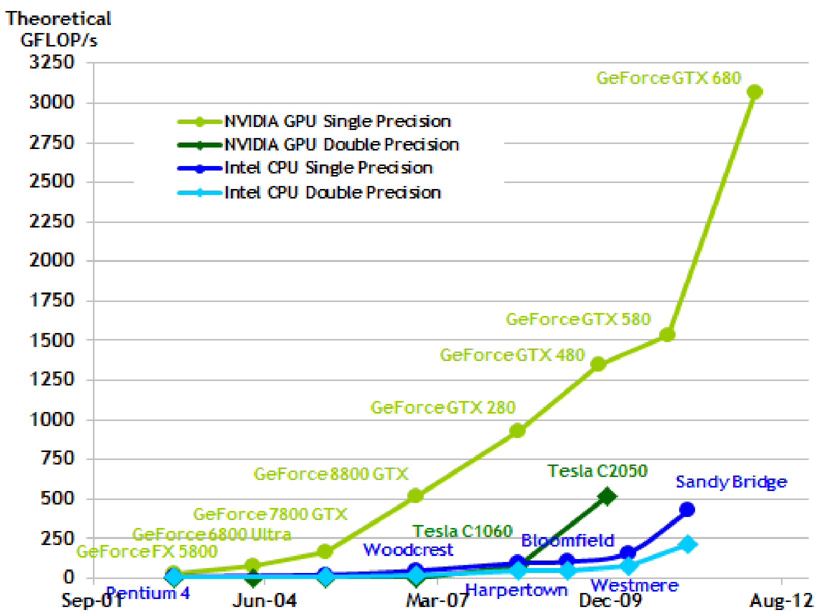
  – **Implications for Variable Annuities Modeling:**
    - **50-500x** speed improvements for stochastic models, when compared to equivalent CPU-based software

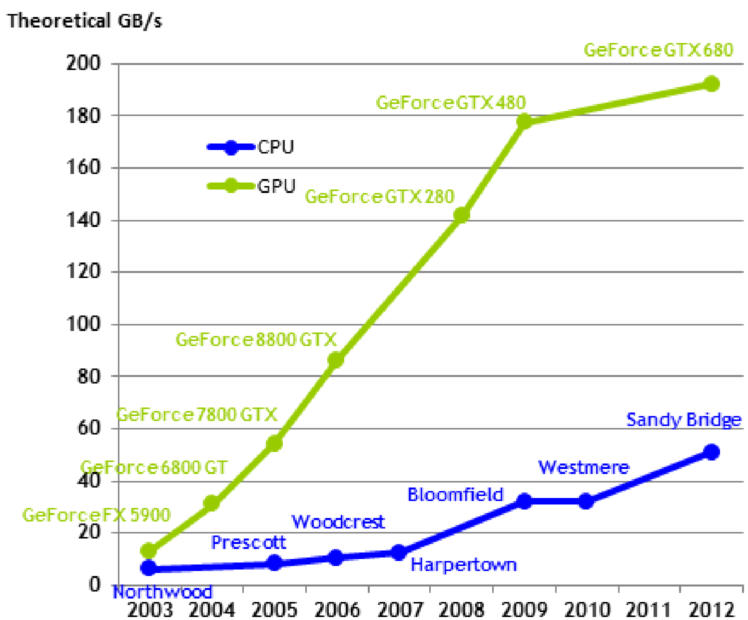
Above: Tesla K10 GPU module, containing 3,072 cores

Source: http://www.nvidia.ca/object/what-is-gpu-computing.html

GPU Computing

- **Performance gap between CPUs and GPUs continues to increase rapidly**





- **Implications for Variable Annuities Modeling:**
  - Cost of GPU-based grids is increasingly lower than cost of equivalent CPU-based grids
  - Complex optimizations for GPUs are increasingly important for VA modeling software (not simply a matter of farming out small sections of legacy code to GPUs)

GPU Computing

- **General Purpose Computing performance on GPUs continues to increase rapidly**

	Tesla M2050	Tesla K10	Improvement
Release Date	May-10	Dec-12	
Cores	448 cores	**3072 cores**	686%
Memory (GDDR5)	3GB	8GB	267%
Memory Bandwidth	148GB/s	320GB/s	216%
Single Precision Peak Performance	1.04 TFLOPS	4.58 TFLOPS	440%
Power Consumption	225W	235W	

  - Tesla M2050 and Tesla K10 have similar hardware and power consumption costs
  - Our VA modeling benchmarks show a 200-300% increase in efficiency (scenarios per second, per GPU or per dollar) when comparing Tesla M2050 GPUs in K10 GPUs

GPU Computing

NVIDIA Kepler GK110 processor

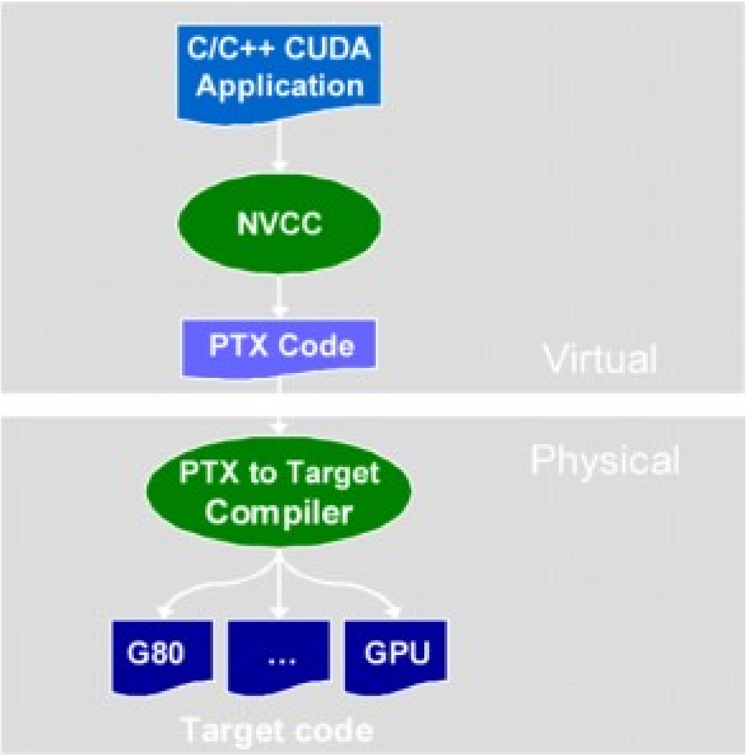
GPU Computing

GK110 processor SMX

GPU Computing

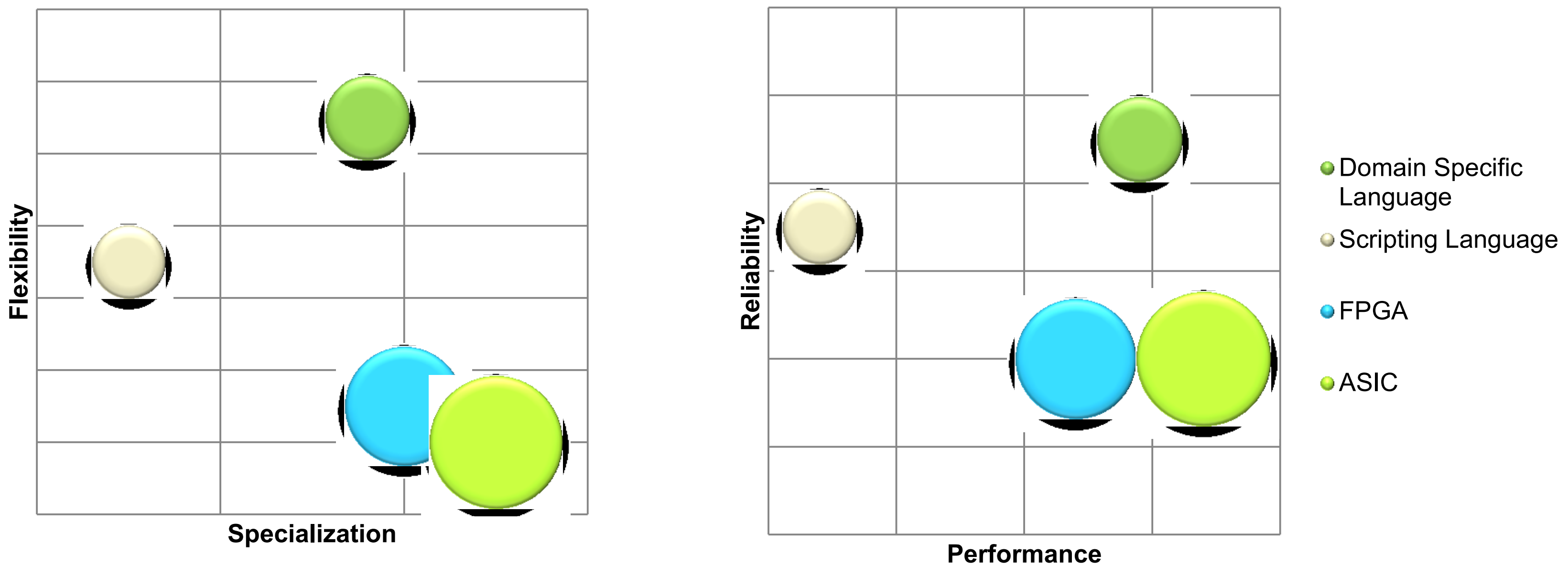
NVIDIA CUDA programming model

```
//Vector size in elements
const int N = 1048576;
//Vector size in bytes
const int dataSize = N * sizeof(float);

//CPU memory allocation
float *h_A = (float *)malloc(dataSize);
float *h_B = (float *)malloc(dataSize);
float *h_C = (float *)malloc(dataSize);

//GPU memory allocation
float *d_A, *d_B, *d_C;
cudaMalloc((void **)&d_A, dataSize));
cudaMalloc((void **)&d_B, dataSize));
cudaMalloc((void **)&d_C, dataSize));

//Initialize h_A[], h_B[]…

//Copy input data to GPU for processing
cudaMemcpy(d_A, h_A, dataSize, cudaMemcpyHostToDevice) );
cudaMemcpy(d_B, h_B, dataSize, cudaMemcpyHostToDevice) );

//Run the core of N / 256 units,256 streams each
//Assuming that N is multiple of 256
vectorAdd<<<N / 256, 256>>>(d_C, d_A, d_B);

//Read GPU results
cudaMemcpy(h_C, d_C, dataSize, cudaMemcpyDeviceToHost) );
```



Domain Specific Languages

Example HPC Solution Trade-Offs

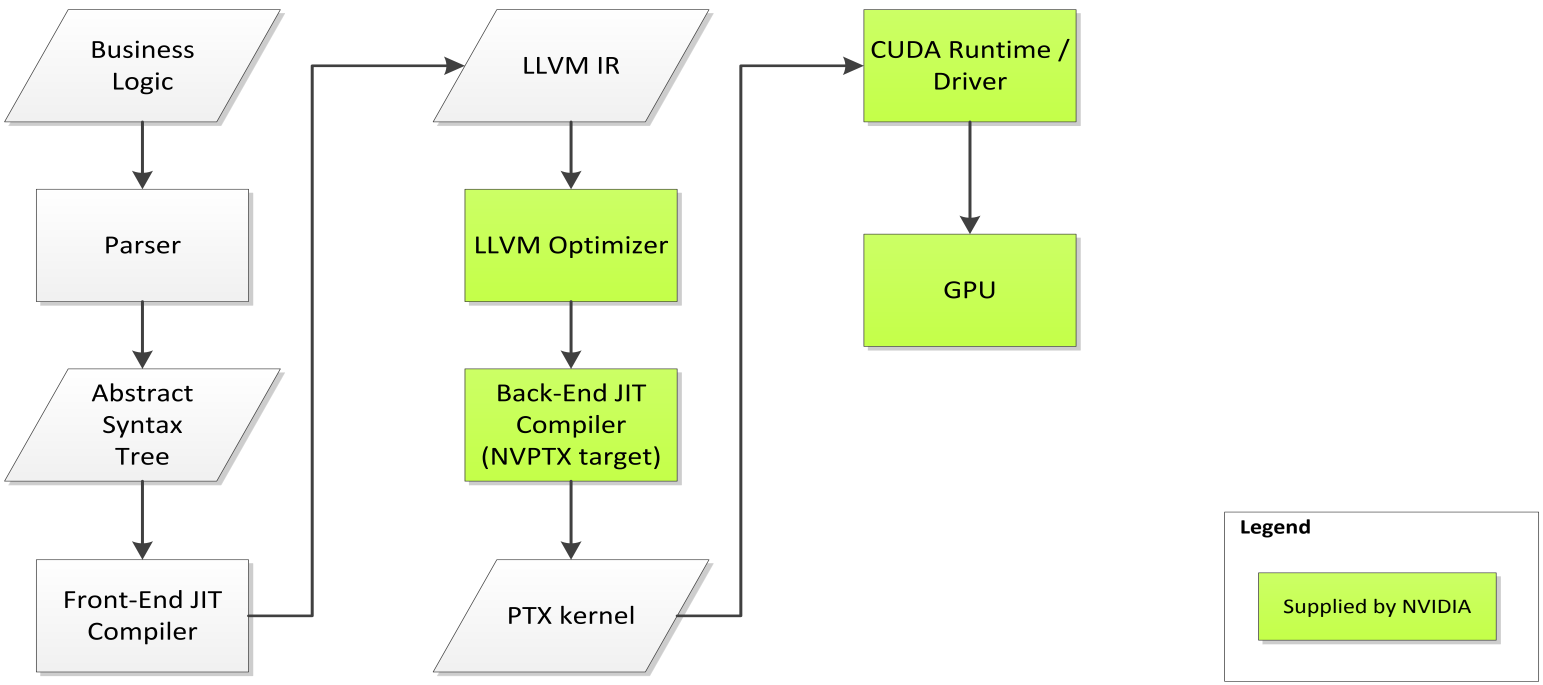


Size of bubble indicates cost (in terms of time and money) of solution

- **Flexibility –** ability to rapidly make changes
- **Specialization –** code specialized to specific hardware
- **Performance –** run-time performance of the solution
- **Reliability –** probable number of bugs in a large system

Domain Specific Languages

GPU DSL compiler architecture



Domain Specific Languages

GPU DSL compiler architecture

DSL

```
foo(x0, x1, x2)
{
   return x0 + x1 * (x2 + 1.0)
}
```

LLVM IR

```
; ModuleID = 'module1'
target triple = "nvptx64"

define double @foo(double %x0, double %x1, double %x2)
   {
entry:
  %x23 = alloca double
  %x12 = alloca double
  %x01 = alloca double
  store double %x0, double* %x01
  store double %x1, double* %x12
  store double %x2, double* %x23
  %x04 = load double* %x01
  %x15 = load double* %x12
  %x26 = load double* %x23
  %faddtmp = fadd double %x26, 1.000000e+00
  %fmultmp = fmul double %x15, %faddtmp
  %faddtmp7 = fadd double %x04, %fmultmp
  ret double %faddtmp7
}
```

Domain Specific Languages

GPU DSL compiler architecture

PTX

```
//
// Generated by LLVM NVPTX Back-End
//

.version 3.1
.target sm_20, texmode_independent
.address_size 64


        // .globl       foo
.entry foo(
        .param .f64 foo_param_0,
        .param .f64 foo_param_1,
        .param .f64 foo_param_2
)
{
        .local .align 8 .b8     __local_depot0[24];
        .reg .b64       %SP;
        .reg .b64       %SPL;
        .reg .pred %p<396>;
        .reg .s16 %rc<396>;


...
```

```
...
        .reg .s16 %rs<396>;
        .reg .s32 %r<396>;
        .reg .s64 %rl<396>;
        .reg .f32 %f<396>;
        .reg .f64 %fl<396>;

        mov.u64         %SPL, __local_depot0;
        cvta.local.u64  %SP, %SPL;
        ld.param.f64    %fl0, [foo_param_0];
        st.f64  [%SP+16], %fl0;
        ld.param.f64    %fl0, [foo_param_1];
        st.f64  [%SP+8], %fl0;
        ld.param.f64    %fl0, [foo_param_2];
        st.f64  [%SP+0], %fl0;
        ld.f64  %fl1, [%SP+16];
        ld.f64  %fl2, [%SP+8];
        add.f64         %fl0, %fl0, 0d3FF0000000000000;
        fma.rn.f64      %fl0, %fl2, %fl0, %fl1;
        st.param.f64    [func_retval0+0], %fl0;
        ret;
}
```

PathWise Platform

AON BENFIELD

PathWise Industry Recognition

- Insurance and Technology Magazine

  "We had ready access to risk information on a regular basis before PathWise, but now the information is refreshed frequently and we're able to make more timely decisions," says Ettles. "**Many calculations that we would have done in hours or days are now done every few minutes** — our information is up-to-date on a real-time basis and we're not taking decisions on information that is stale." Martin Ettles is a senior actuary, finance and risk management, Standard Life

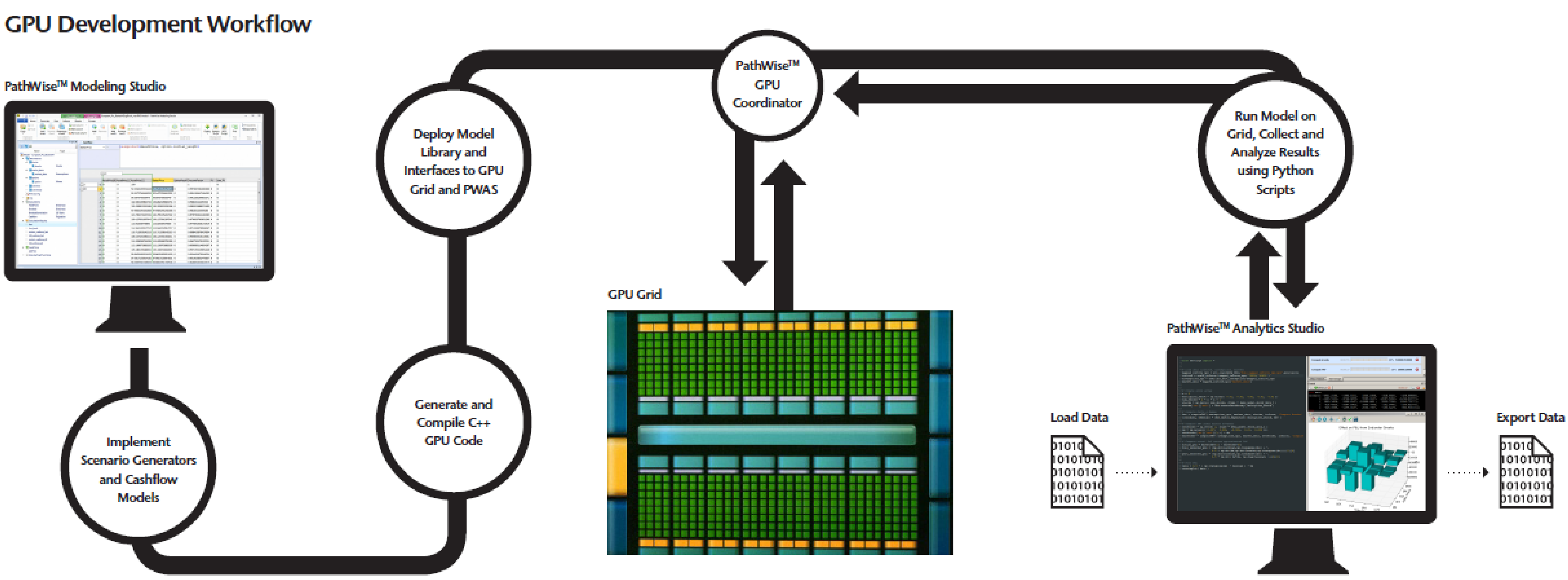
PathWise Industry Recognition

- PathWise won IDC's HPC Innovation Excellence Award in June 2012

  "The new award winners and project leaders announced at ISC'12 are as follows (contact IDC for additional details about the projects):
- **GE Global Research (U.S.).**
- **Department of Defense High Performance Computing Modernization Program (U.S.).**
- **Mary Bird Perkins Cancer Center and Louisiana State University (U.S.).**
- **BGI Shenzhen (China).**
- **Aon Benfield Securities, Inc. (Canada).** Aon has developed the PathWise platform, which uses GPU-based high performance computing to enable quantitative analysts to quickly and easily express financial application kernels such as Monte Carlo simulations using domain-specific interfaces. The computational capabilities offered by the GPU-driven HPC enabled quantitative analysts to accelerate financial computations from days to minutes, with 50-100 times throughput over conventional techniques. The PathWise platform from Aon Benfield achieved an average 90% cost savings both in terms of HPC infrastructure costs and time-to-market, translating to several millions of dollars in savings. Project leader: Peter Phillips, Aamir Mohammad"

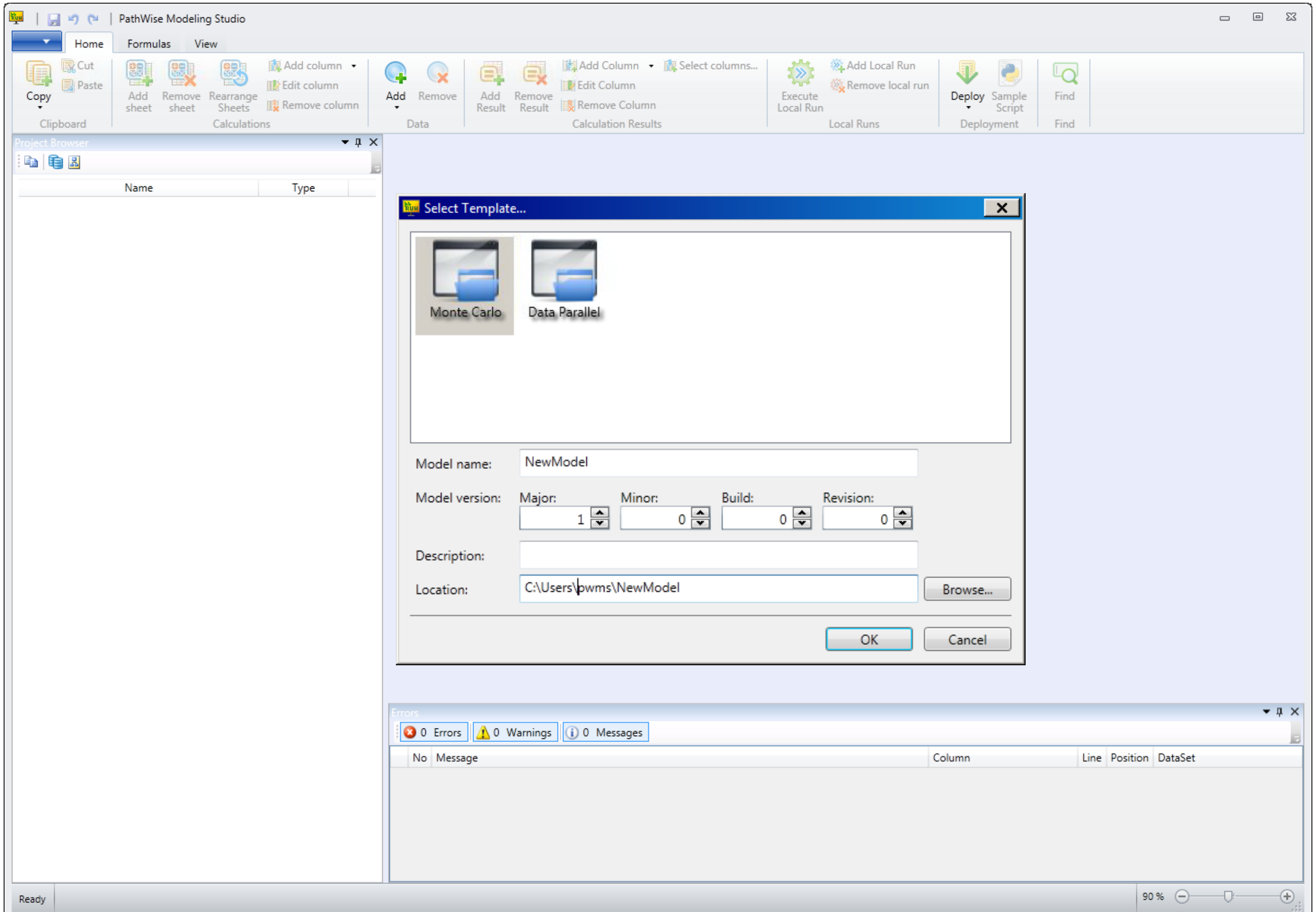
PathWise Modeling Studio

- End-user tools for High Productivity Computing



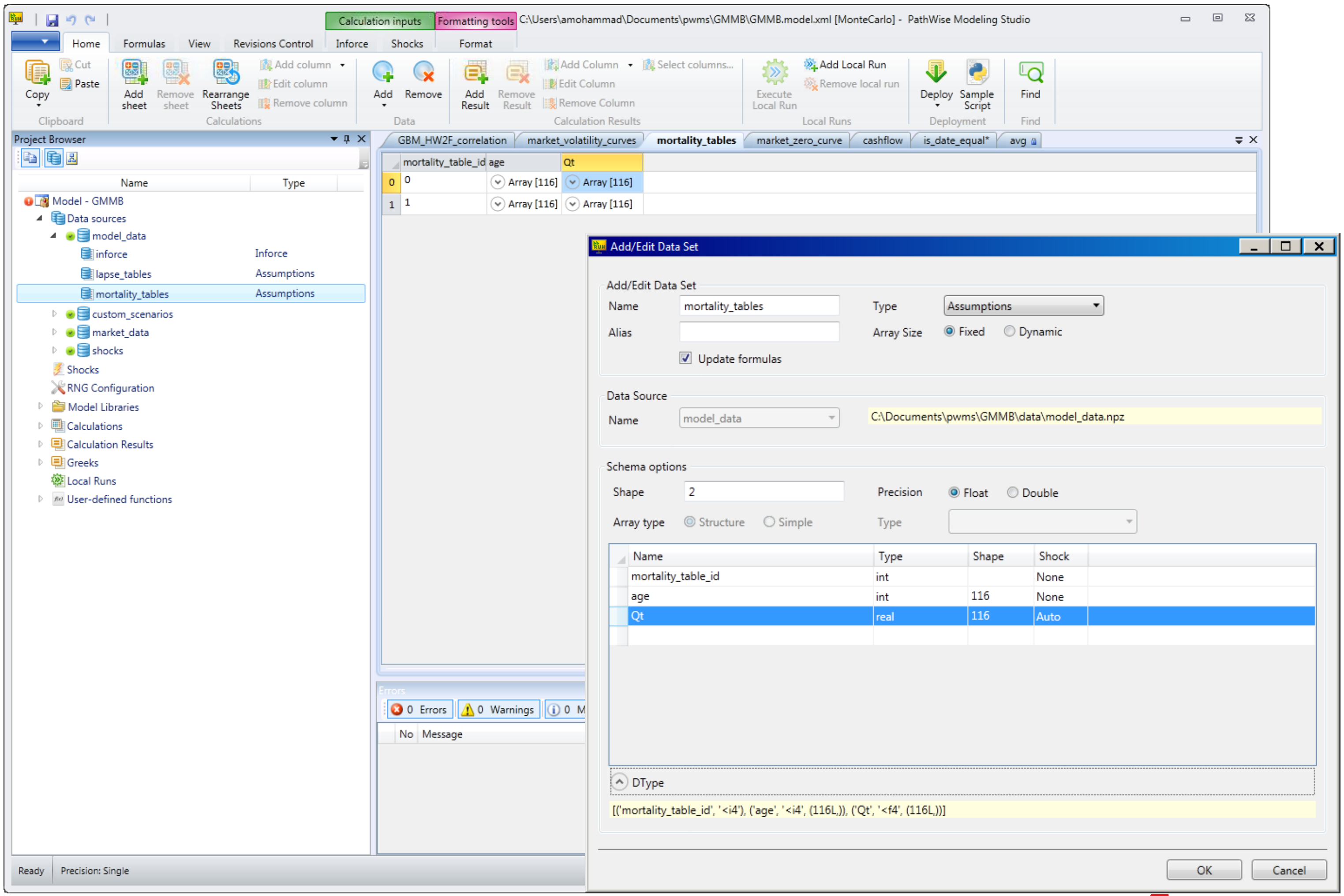
PathWise Modeling Studio

- Create a new model



PathWise Modeling Studio

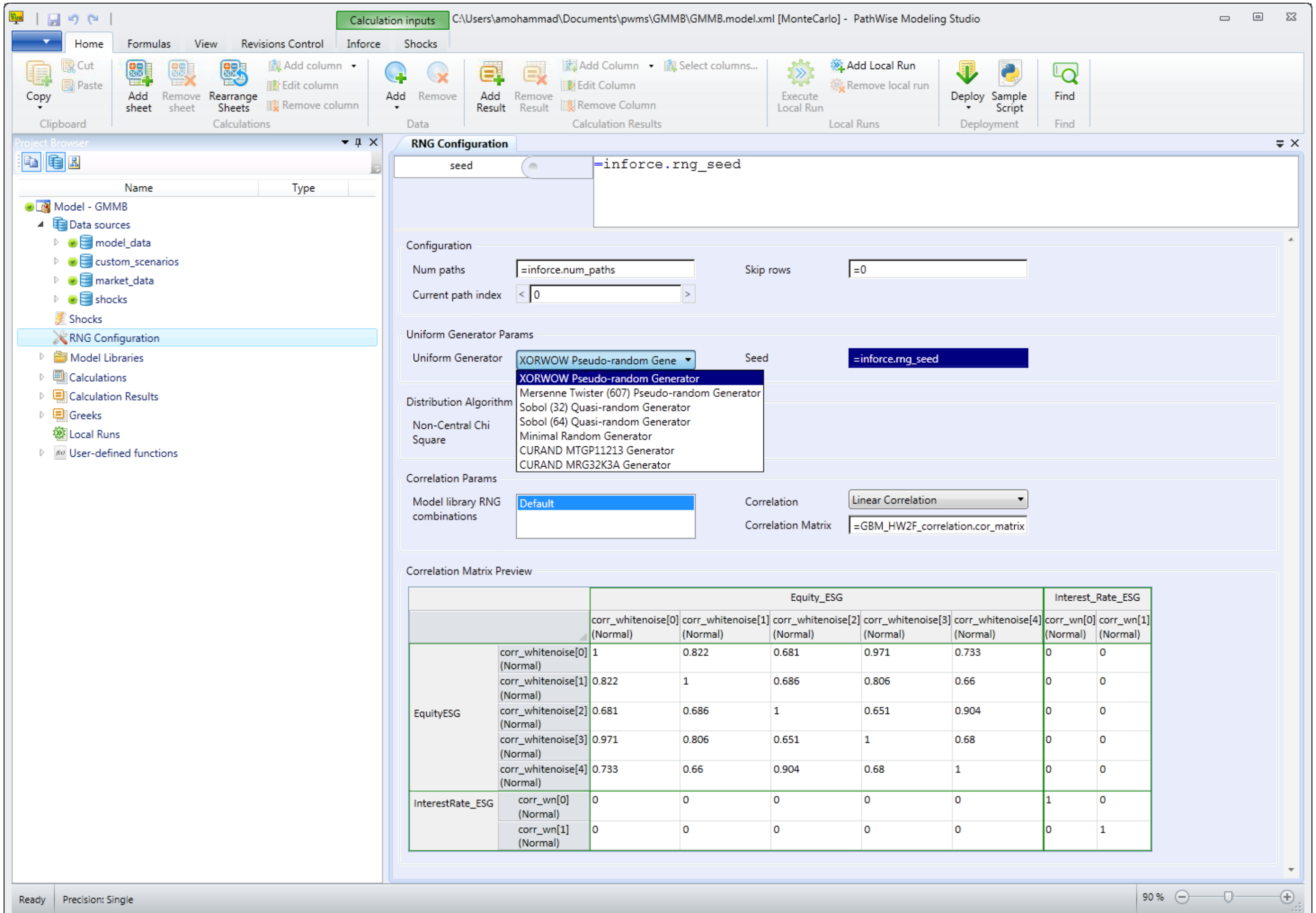
- Define input data structures (customized NumPy data structures)



AON BENFIELD

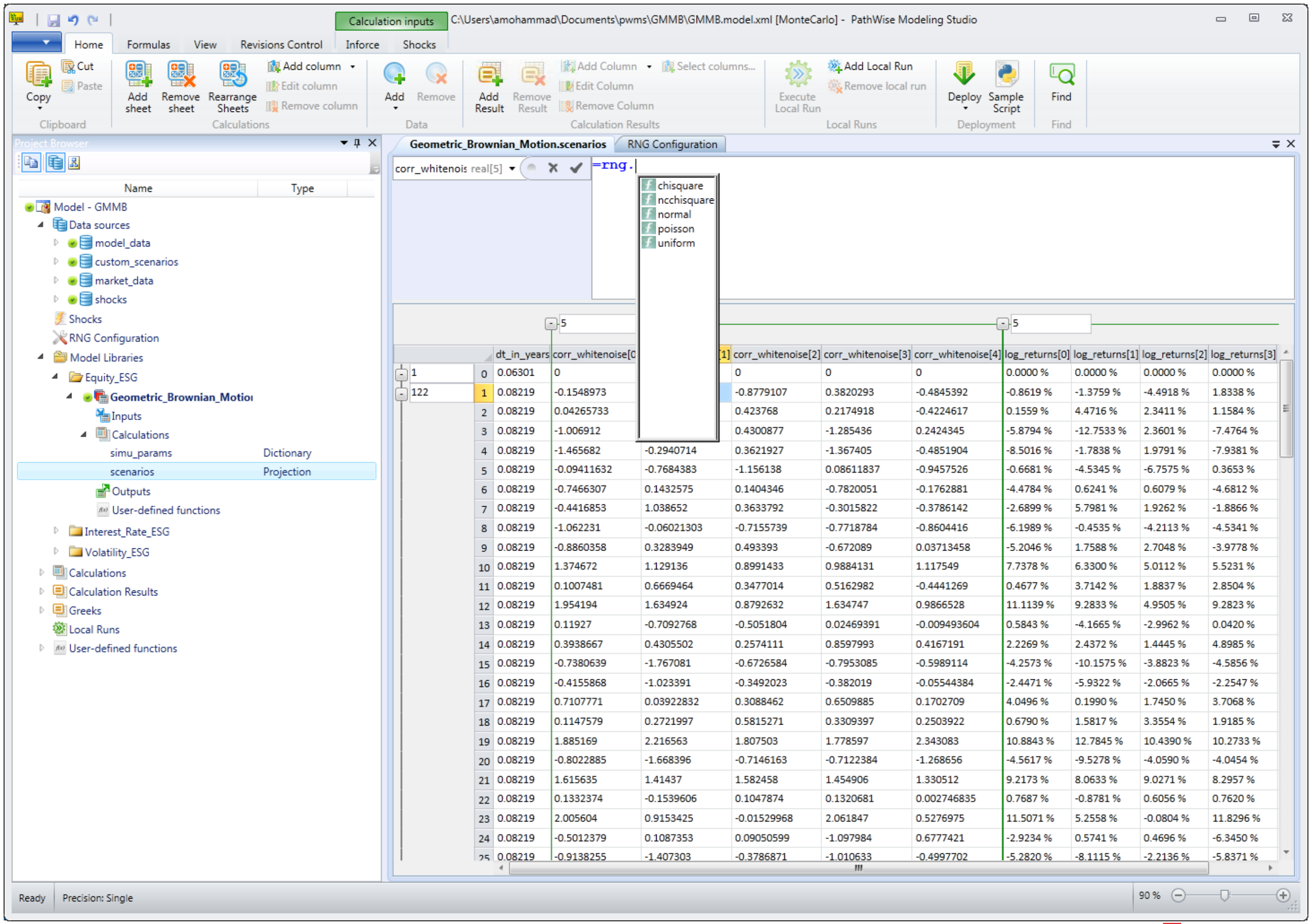
PathWise Modeling Studio

- Setup Random Number Generator options



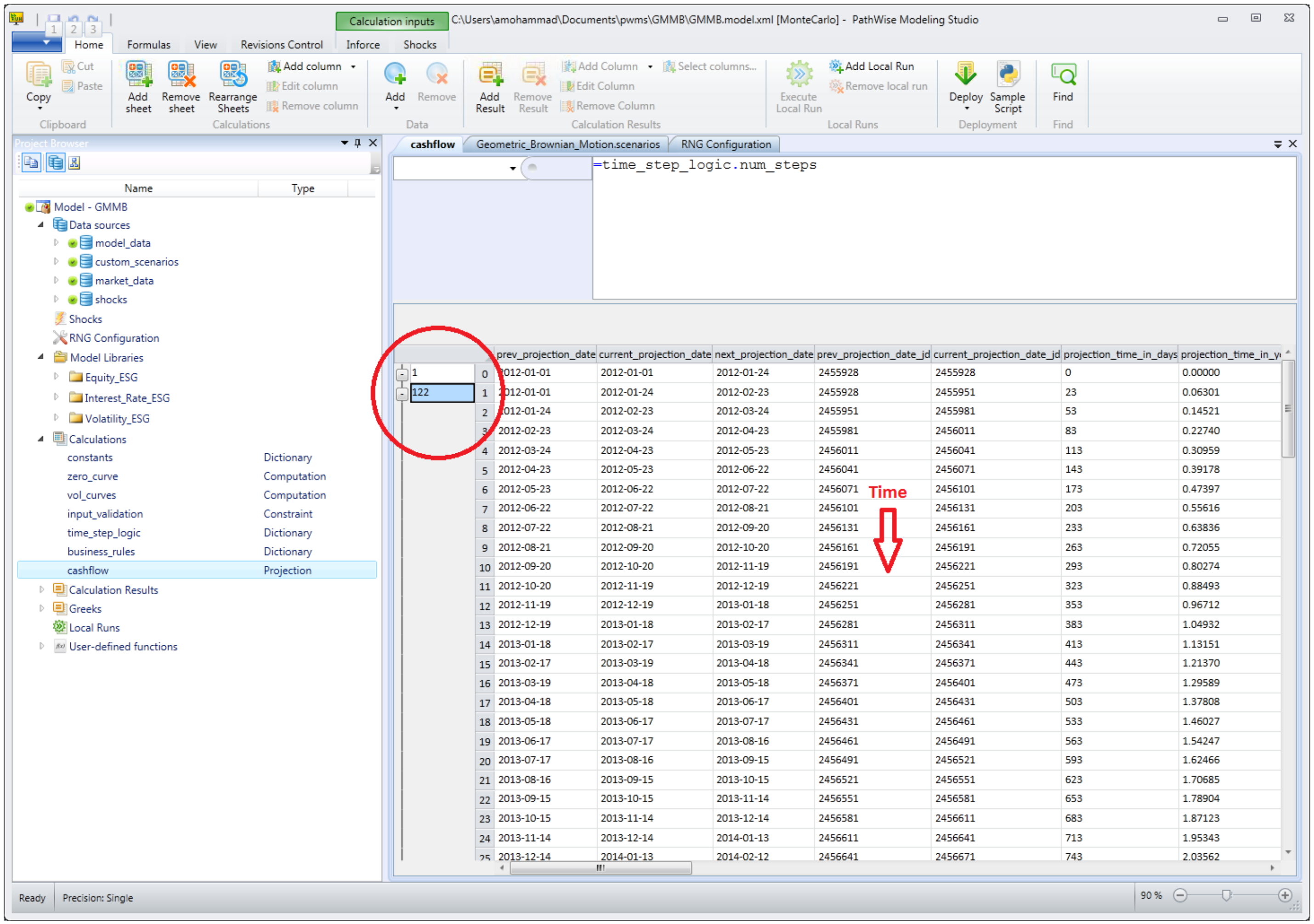
PathWise Modeling Studio

- Import and configure Model Libraries (e.g. pre-built Economic Scenario Generators)



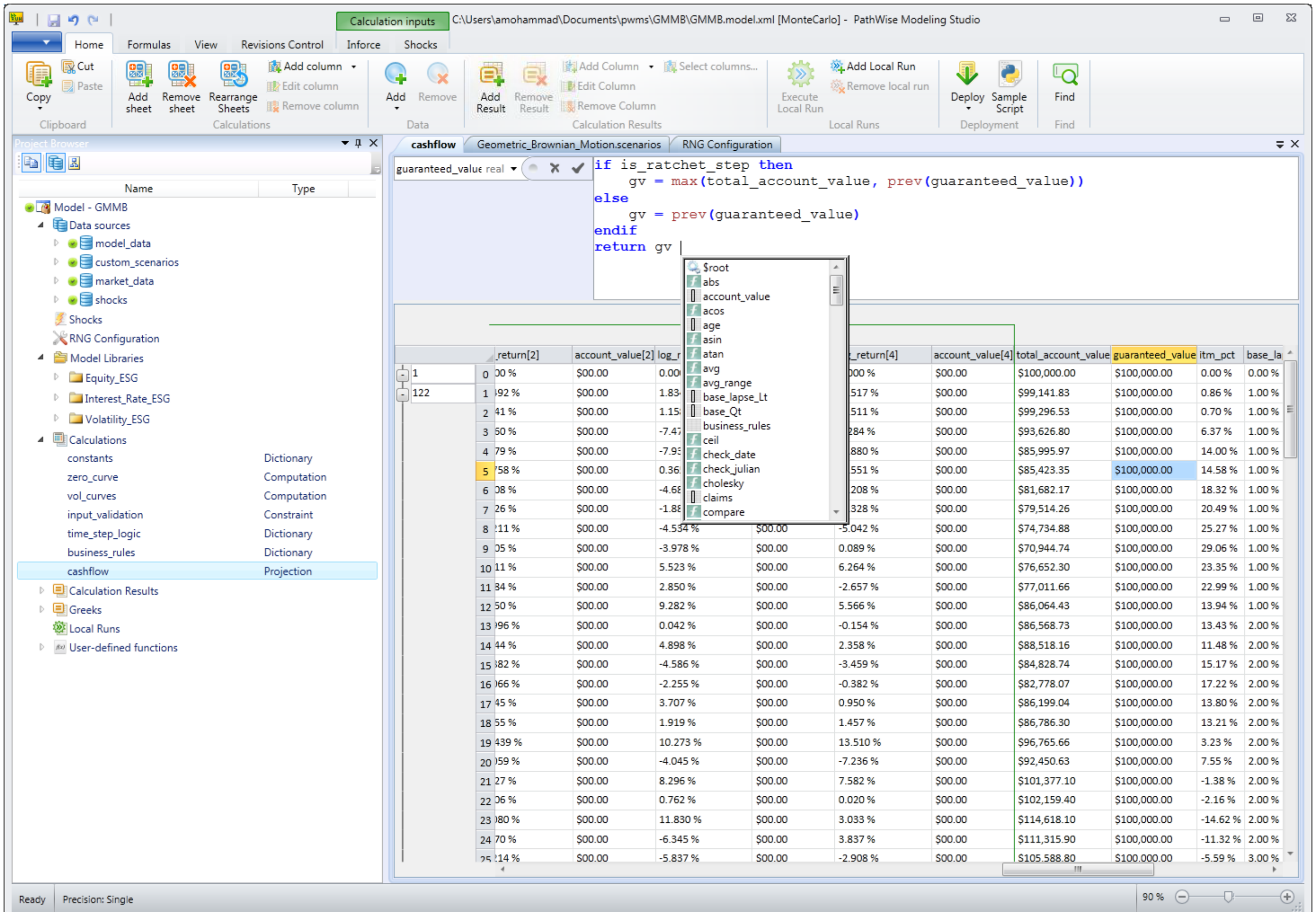
PathWise Modeling Studio

- Calculate number of time-steps to simulate



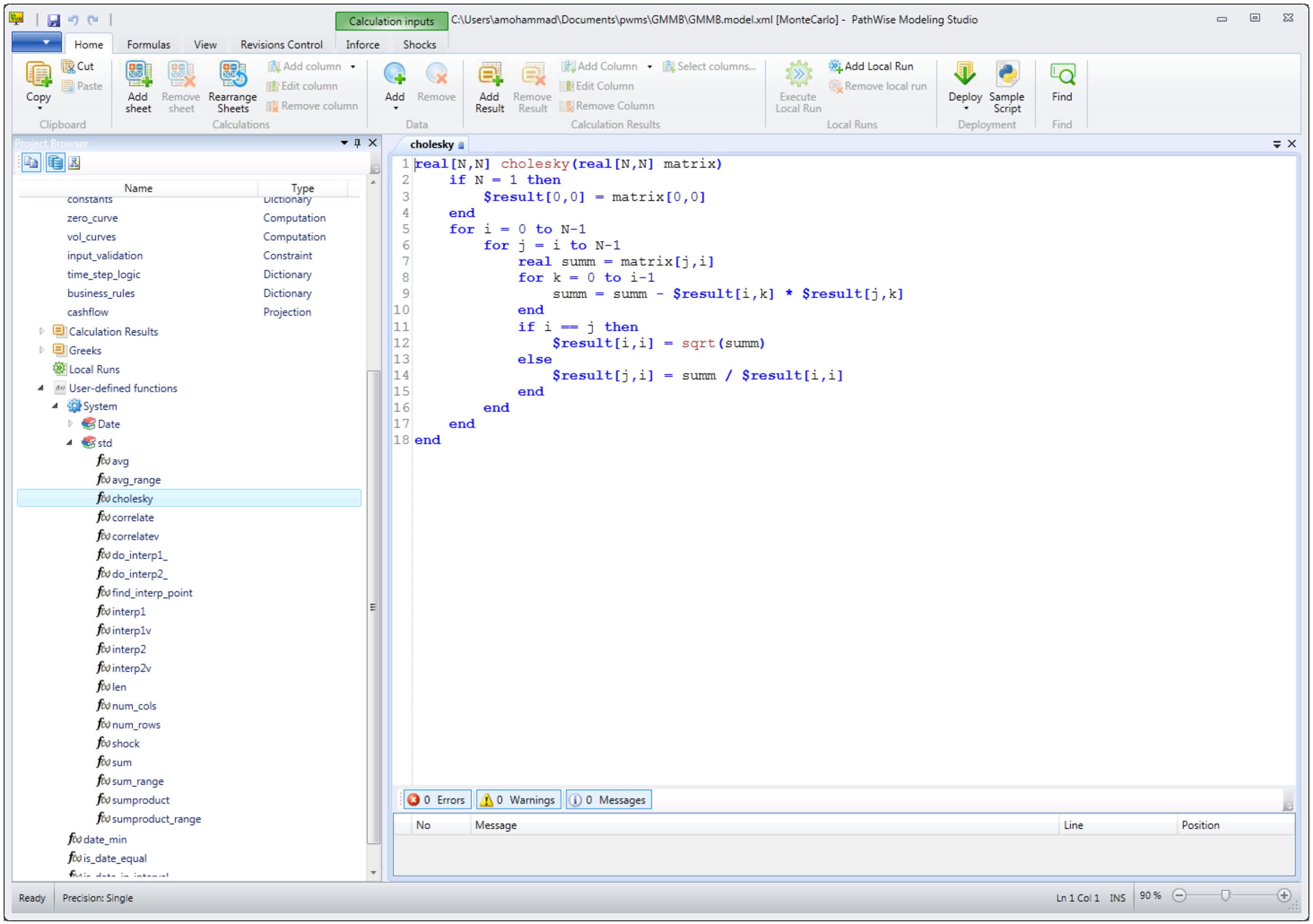
PathWise Modeling Studio

- Define simulation columns and formulas



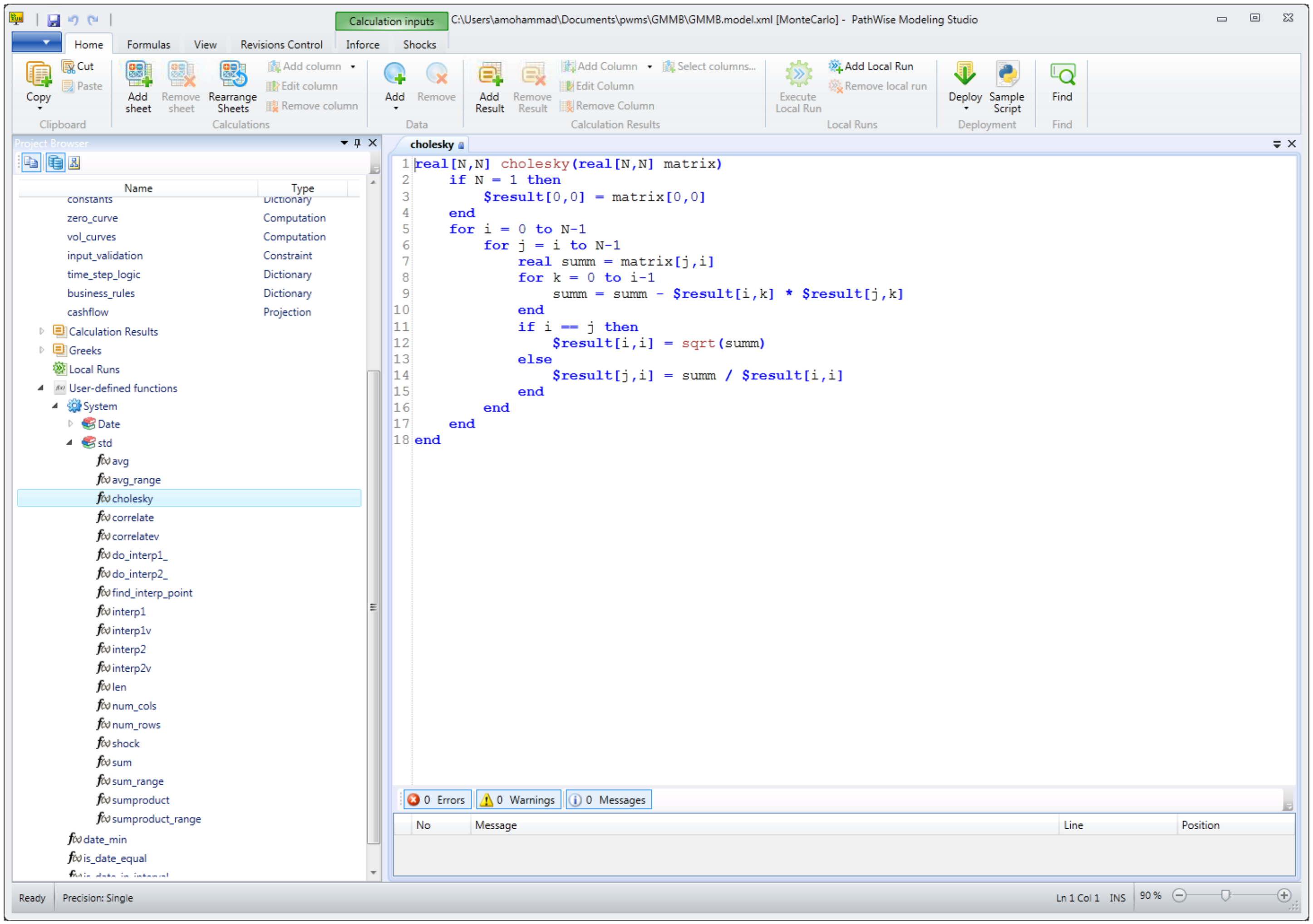
PathWise Modeling Studio

- Encapsulate re-usable logic in UDFs and UDF libraries



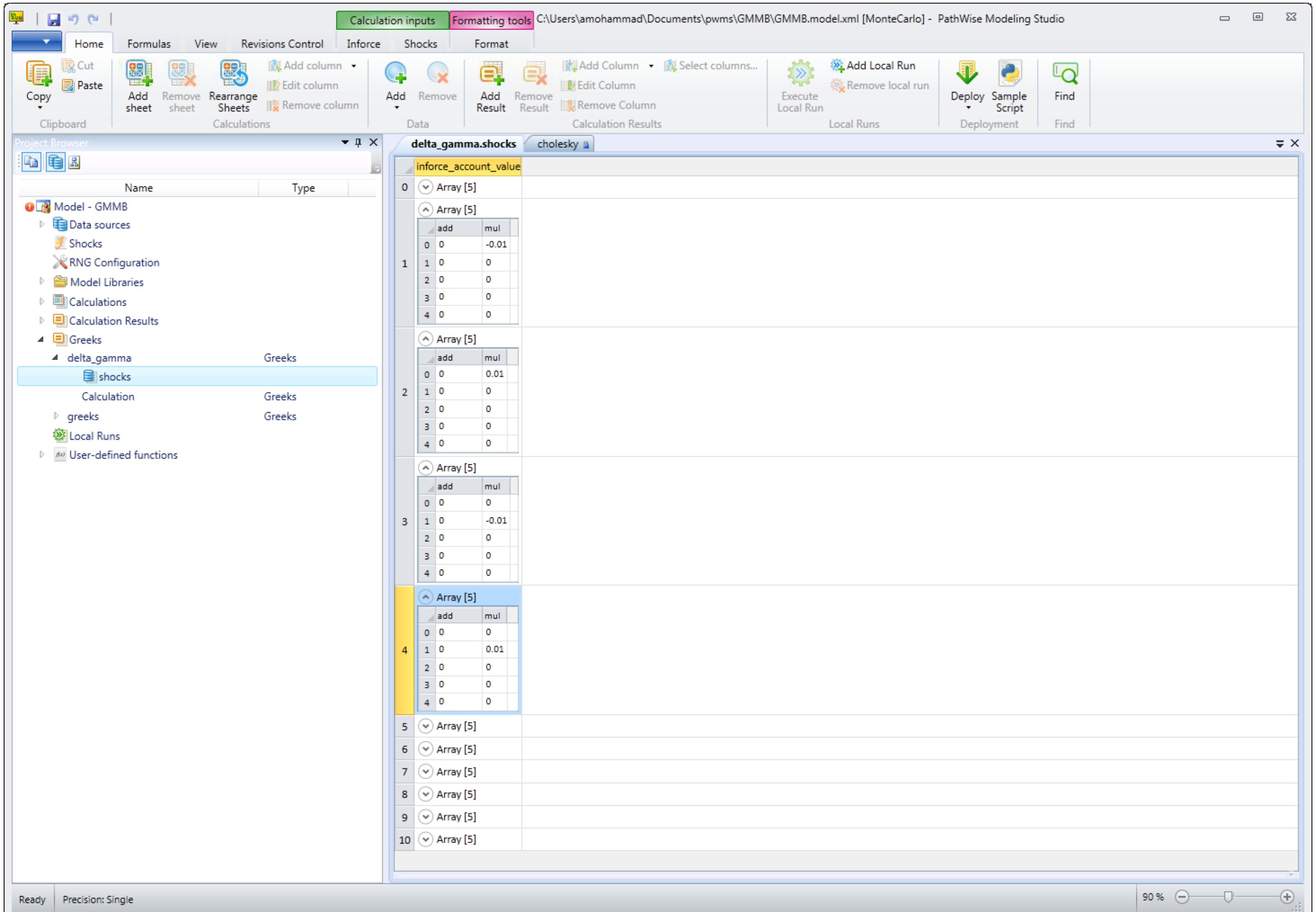
PathWise Modeling Studio

- Encapsulate re-usable logic in UDFs and UDF libraries



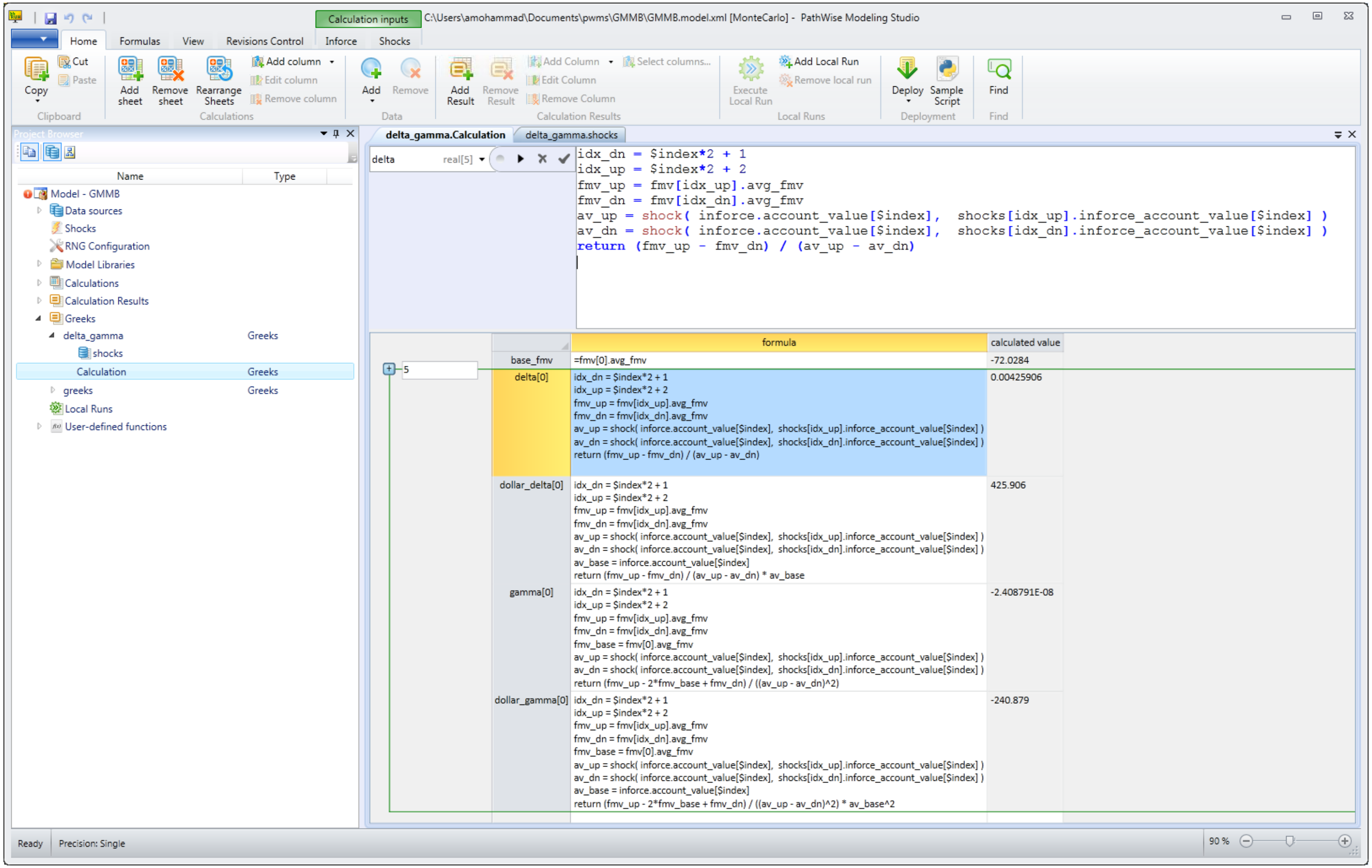
PathWise Modeling Studio

- Define model outputs (e.g. Greeks)



PathWise Modeling Studio

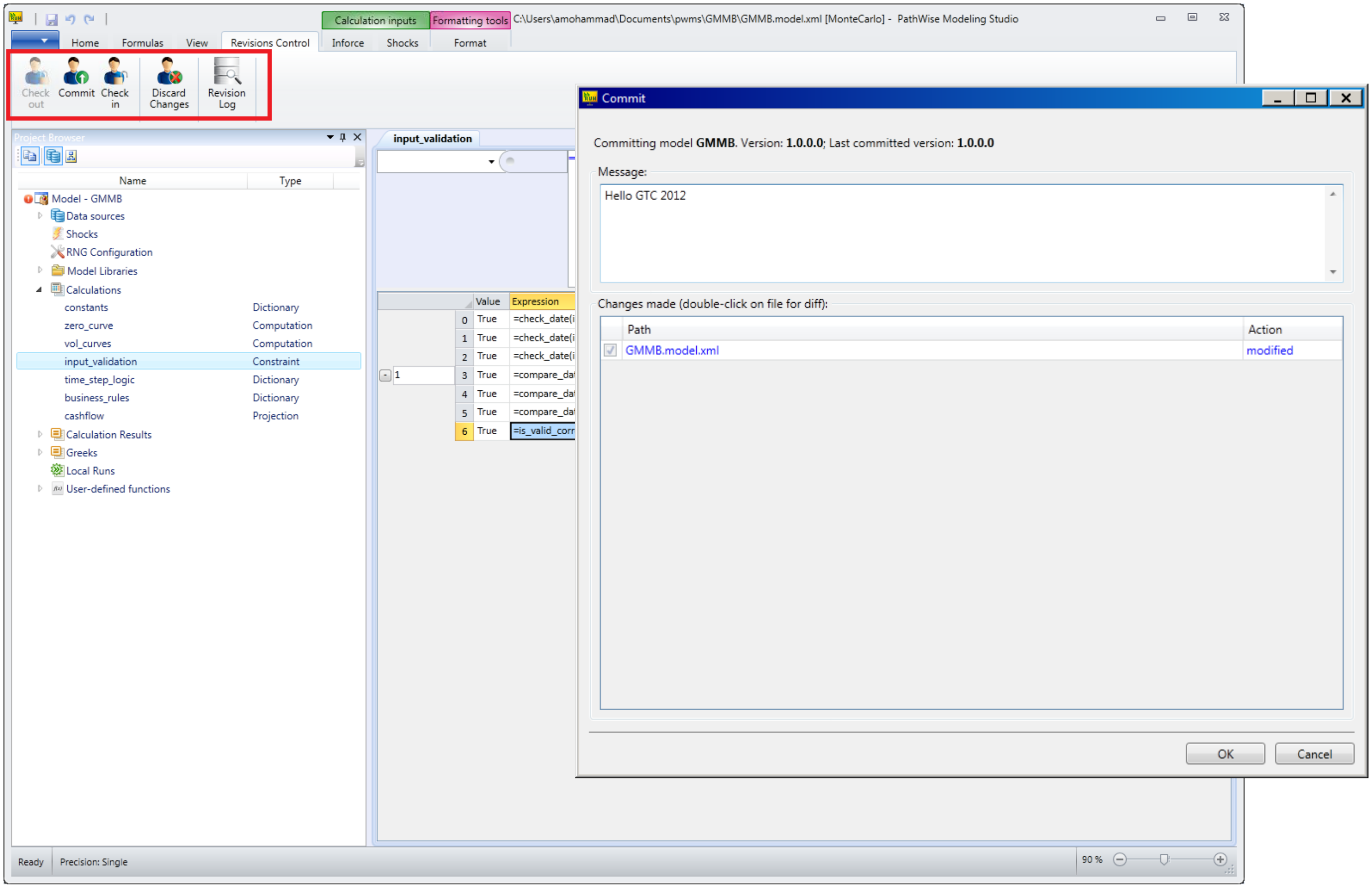
- Define model outputs (e.g. Greeks)



AON BENFIELD

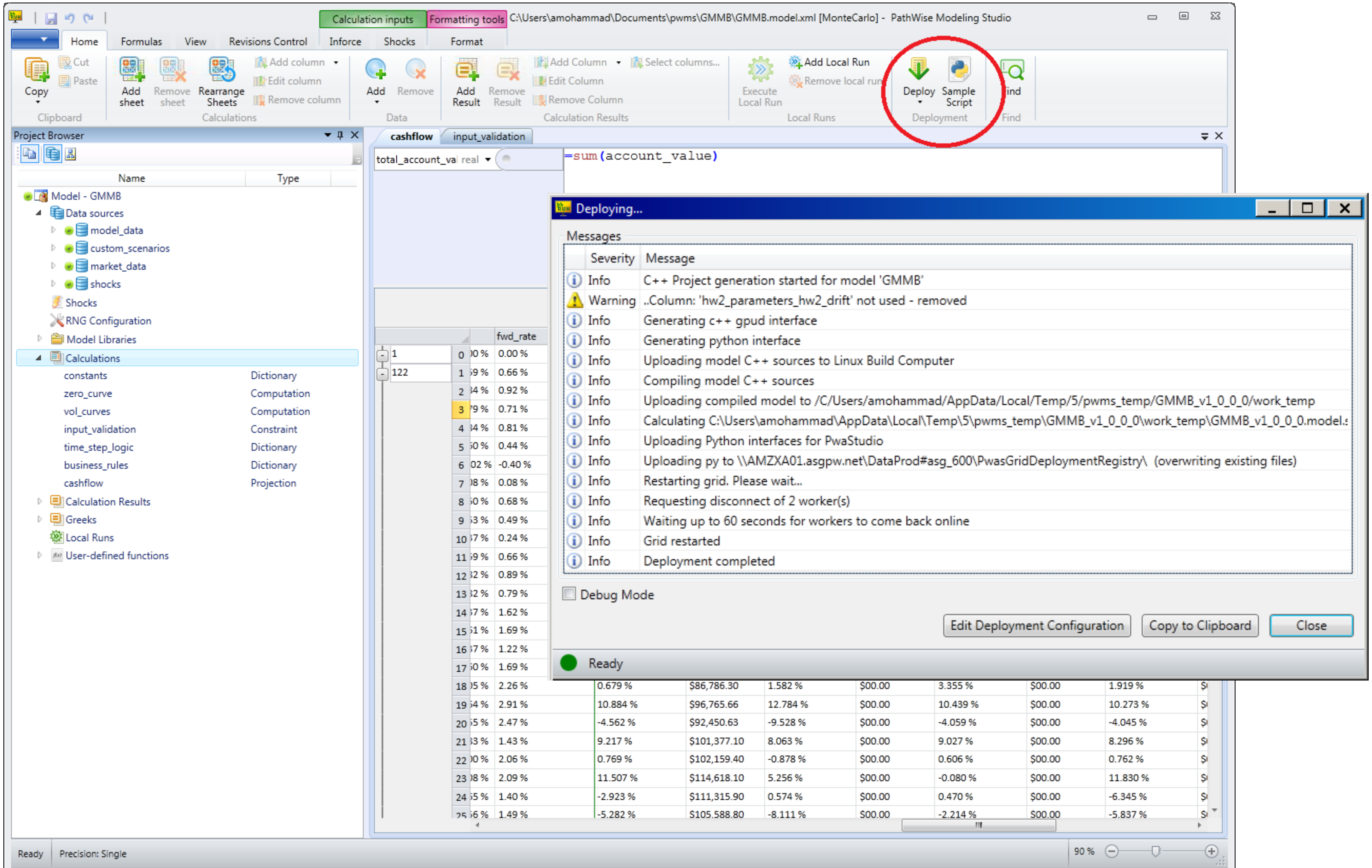
PathWise Modeling Studio

- Commit model to SVN source code repo



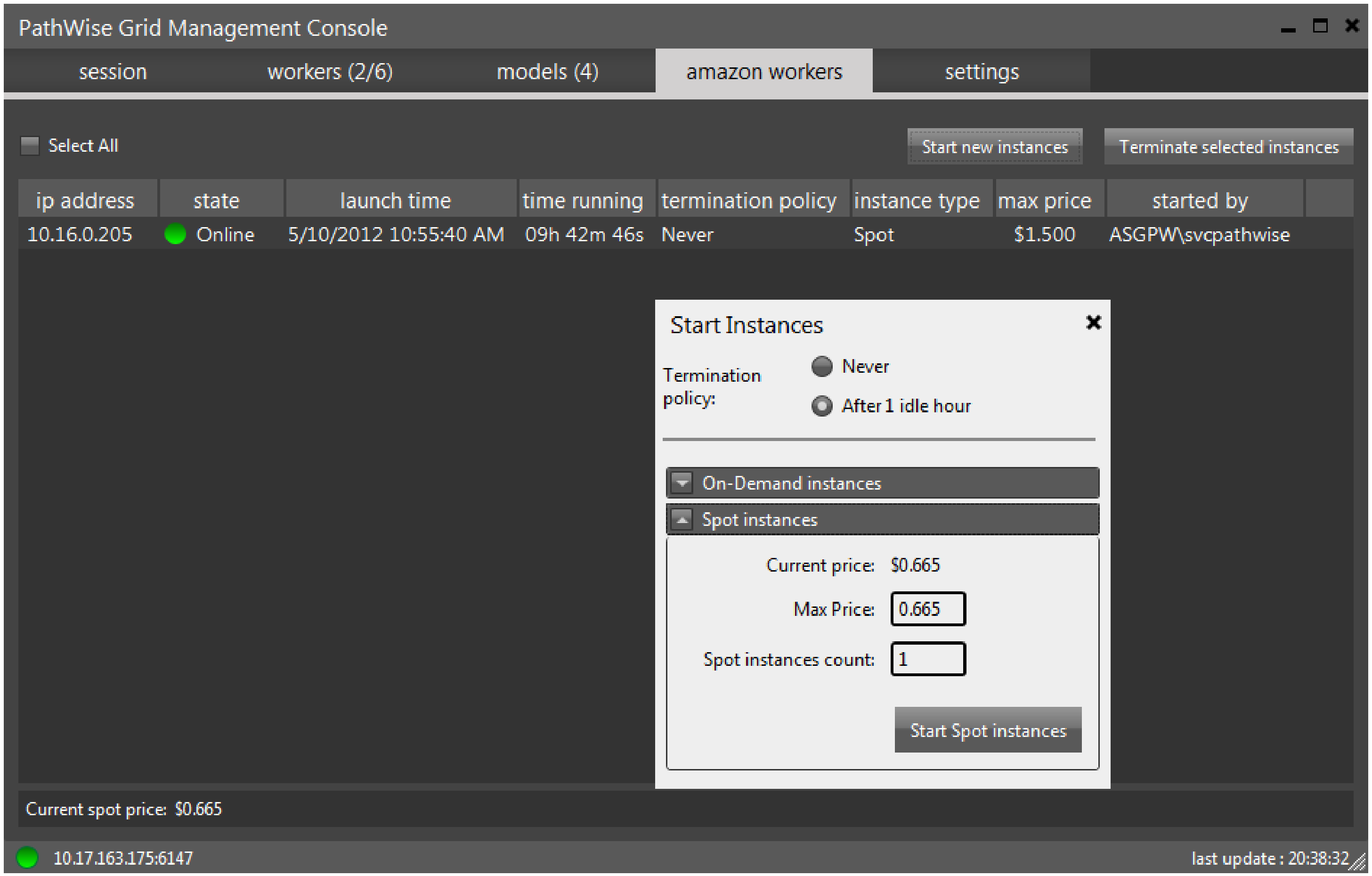
PathWise Modeling Studio

- **Compile and deploy** model to **GPUs**



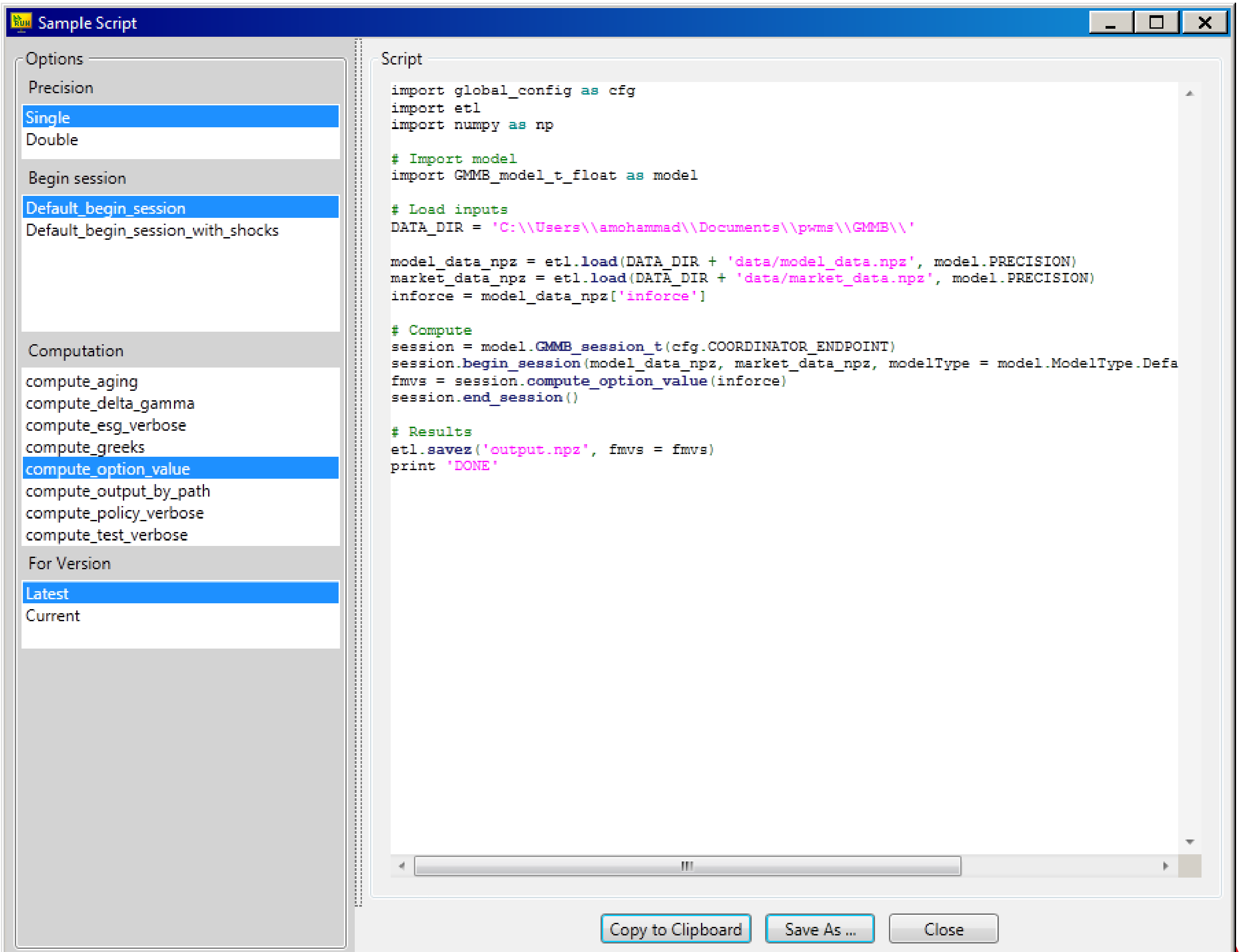
PathWise Modeling Studio

- Add GPU grid workers from the **Cloud**



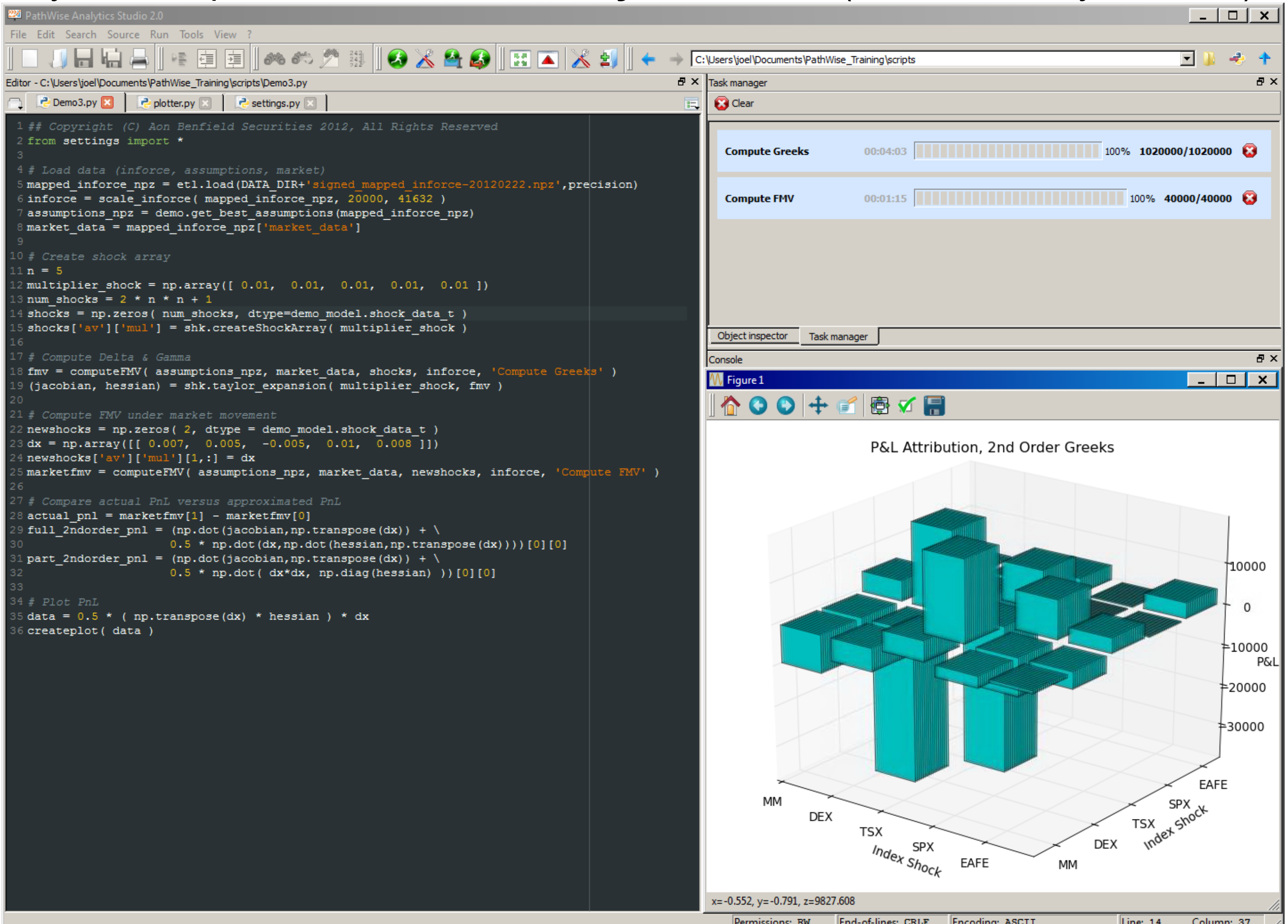
PathWise Modeling Studio

- Generate sample Python script



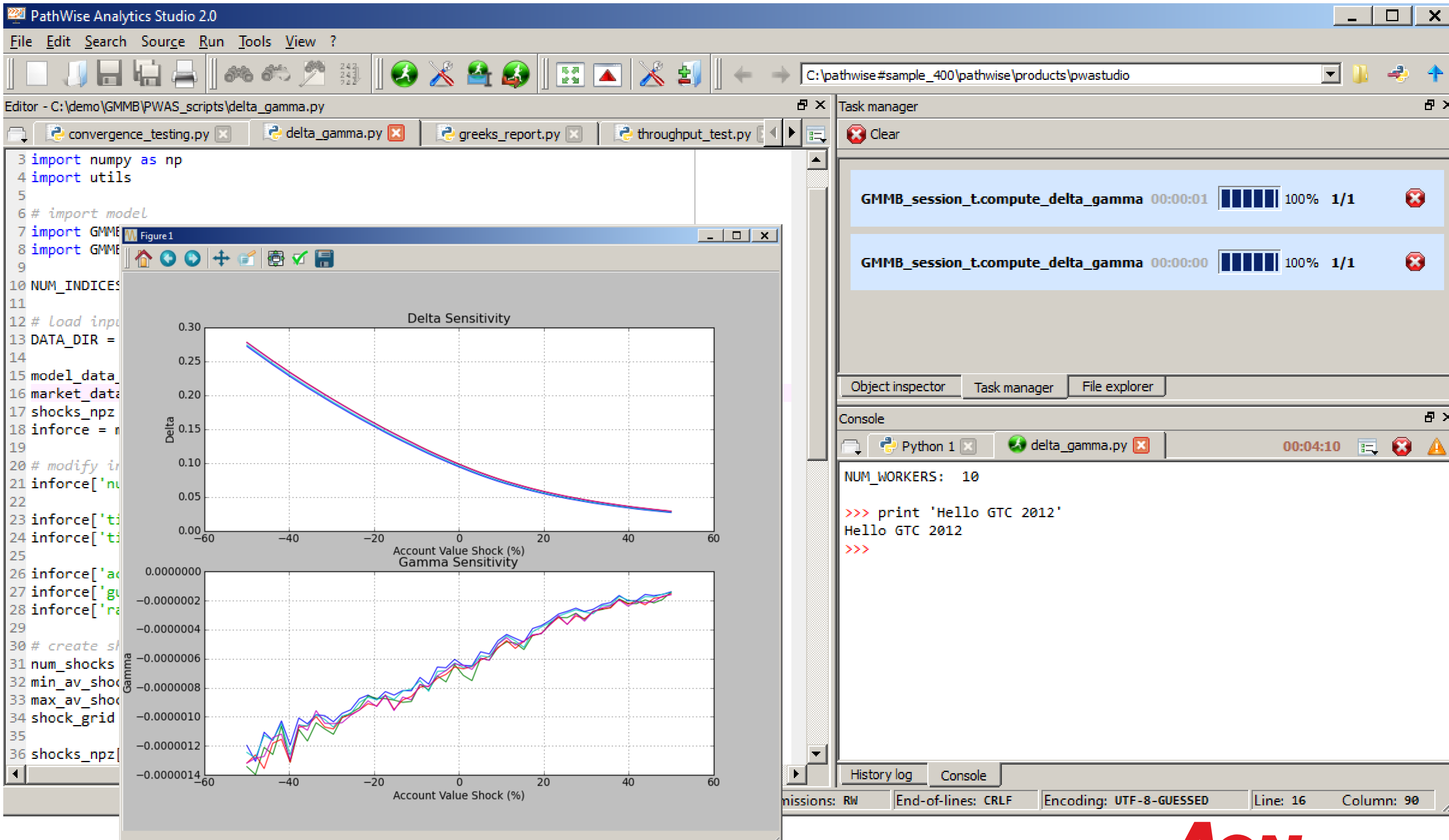
PathWise Analytics Studio

- Run Python scripts from **PathWise Analytics Studio** (customized Python IDE)

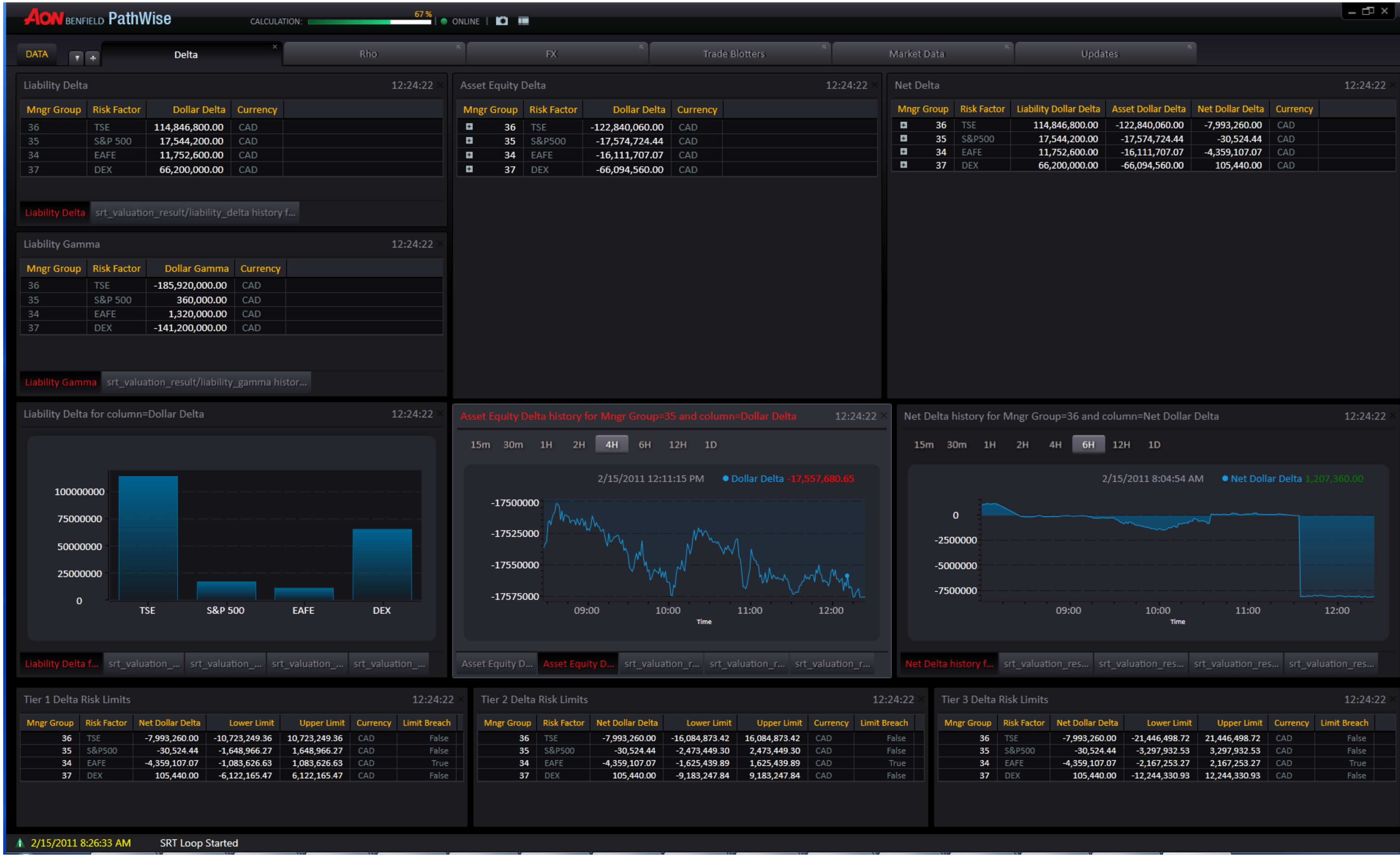


PathWise Analytics Studio

- Run Python scripts from **PathWise Analytics Studio** (customized Python IDE)



PathWise Seriatim Real-Time Risk System



Thank You

AON BENFIELD