Windows HPC Server 2008

High Productivity Computing

Kyril Faenov
General Manager
Windows HPC
Microsoft Corporation

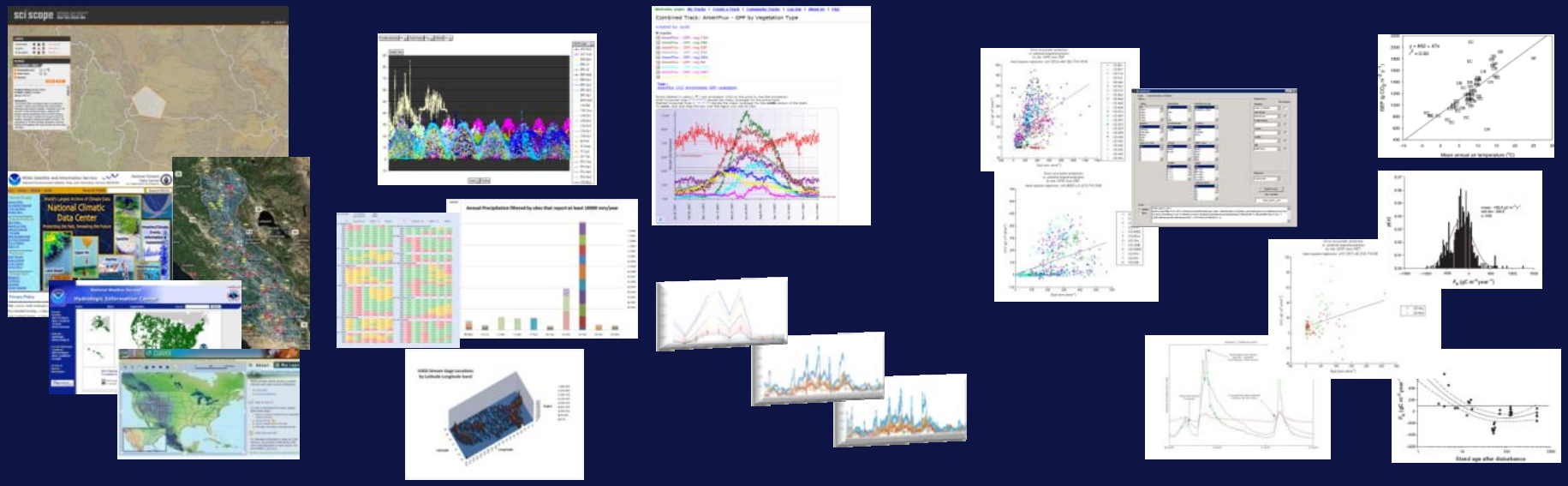
Accelerating Scientific Process

1. Observation **2. Analysis**

4. Validation **3. Simulation**

- Thousand years ago:
  Experimental Science
  - description of natural phenomena
- Last few hundred years:
  Theoretical Science
  - Newton's Laws, Maxwell's Equations …
- Last few decades:
  Computational Science
  - simulation of complex phenomena
- Today:
  'e-Science' or Data-centric Science
  - unify theory, experiment, and simulation

The Data Pipeline



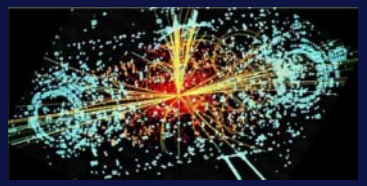
Data Gathering	Discovery and Browsing	Science Exploration	Domain specific analyses	Scientific Output
"Raw" data includes sensor output, data downloaded from agency or collaboration web sites, papers (especially for ancillary data	"Raw" data browsing for discovery (do I have enough data in the right places?), cleaning (does the data look obviously wrong?), and light weight science via browsing	"Science variables" and data summaries for early science exploration and hypothesis testing. Similar to discovery and browsing, but with science variables computed via gap filling, units conversions, or simple equation.	"Science variables" combined with models, other specialized code, or statistics for deep science understanding.	Scientific results via packages such as MatLab or R2. Special rendering package such as ArcGIS. Paper preparation.

Courtesy Catherine VanIngen, MSR

Explosion of Data

Experiments

Simulations

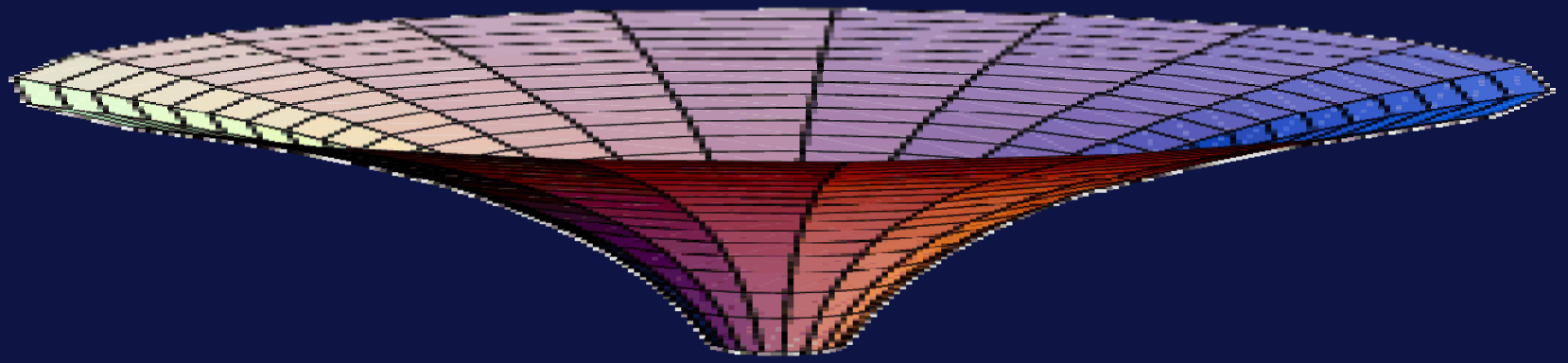


Archives

Literature

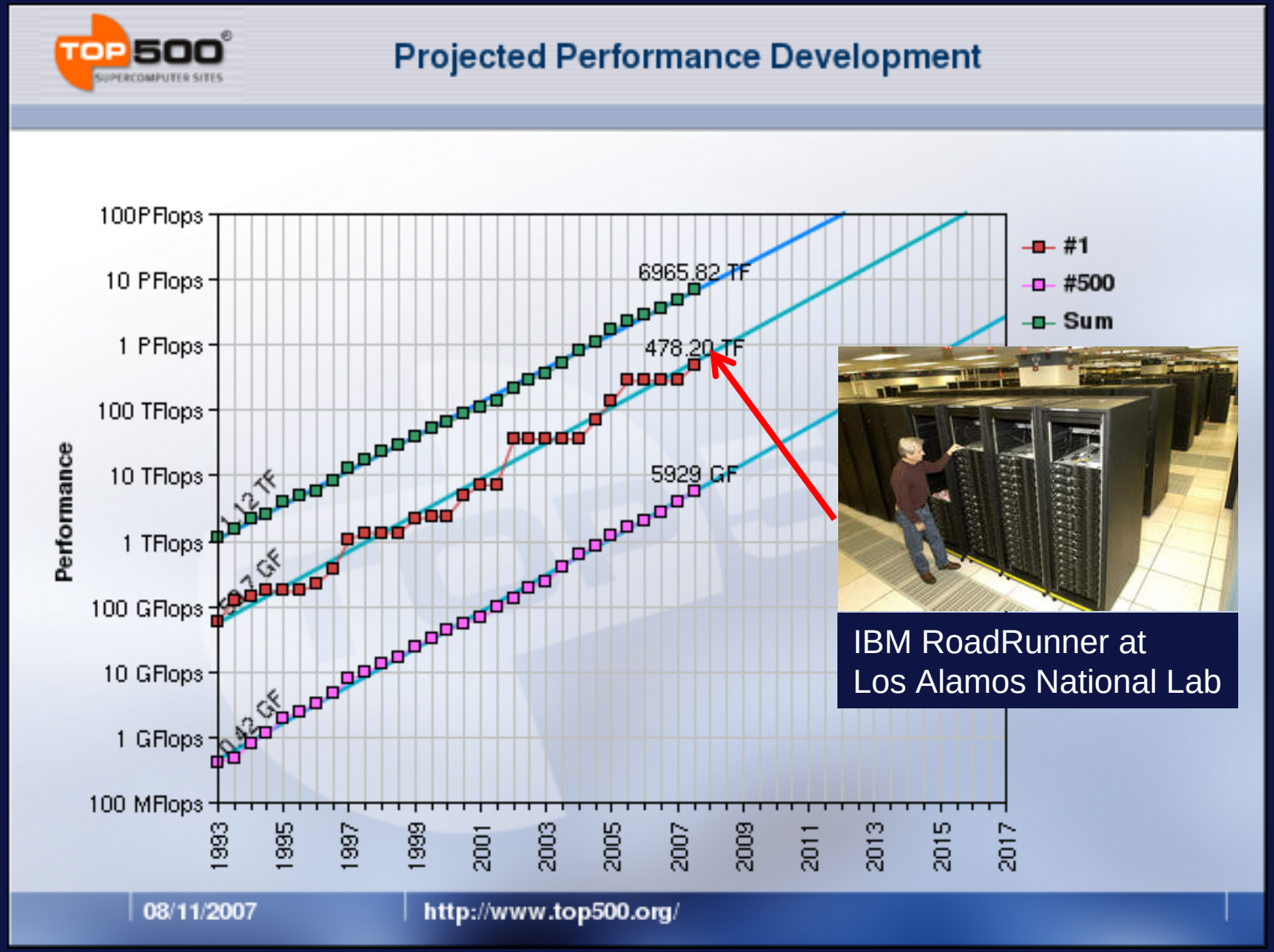
Cornell University Library

arXiv.org



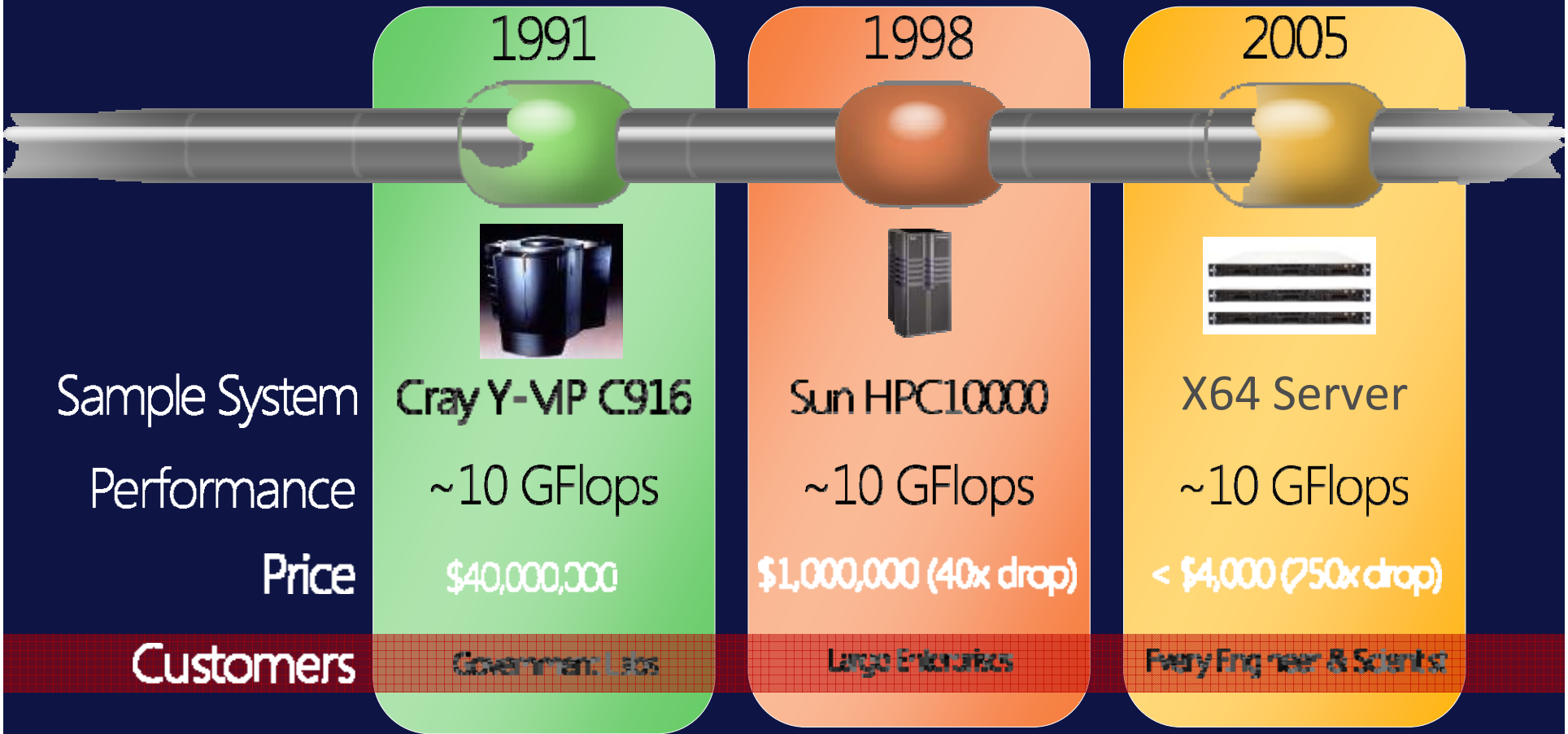
Petabytes
Doubling every
2 years

Supercomputing Reached the Petaflop

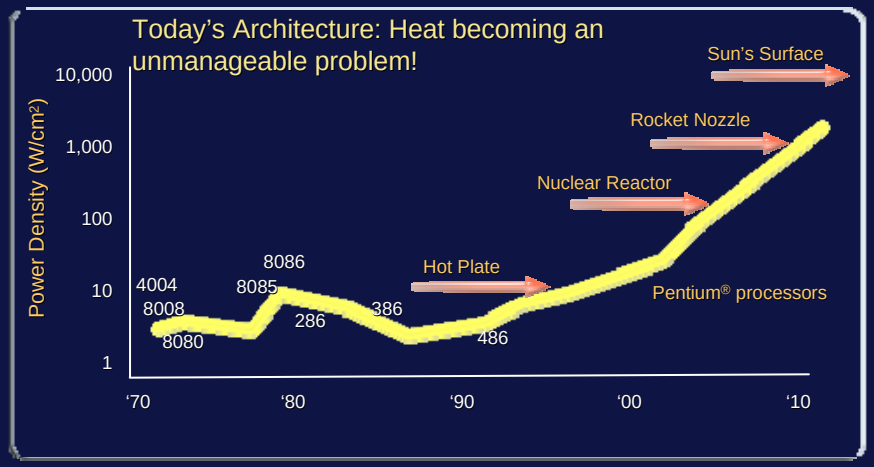


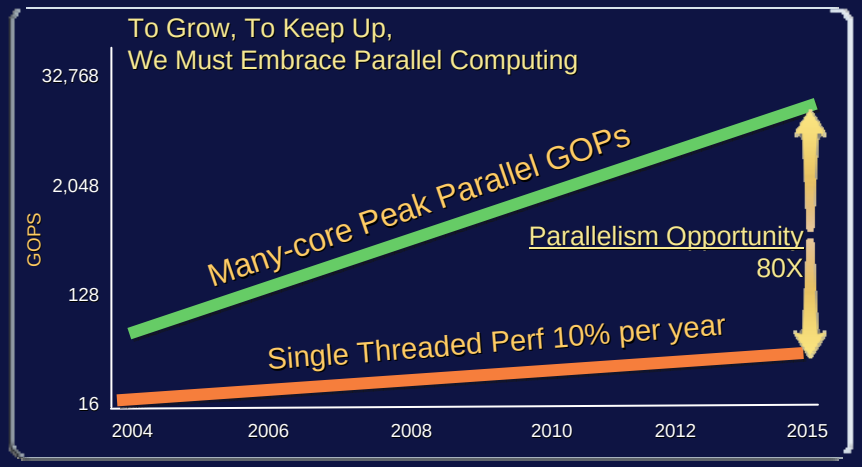
IBM RoadRunner at
Los Alamos National Lab

HPC in Every Lab



Hardware Paradigm Shift





Intel Developer Forum, Spring 2004 - Pat Gelsinger

"… we see a very significant shift in what architectures will look like in the future ... fundamentally the way we've begun to look at doing that is to move from instruction level concurrency to … multiple cores per die. But we're going to continue to go beyond there. And that just won't be in our server lines in the future; this will permeate every architecture that we build. *All* will have *massively* multicore implementations."

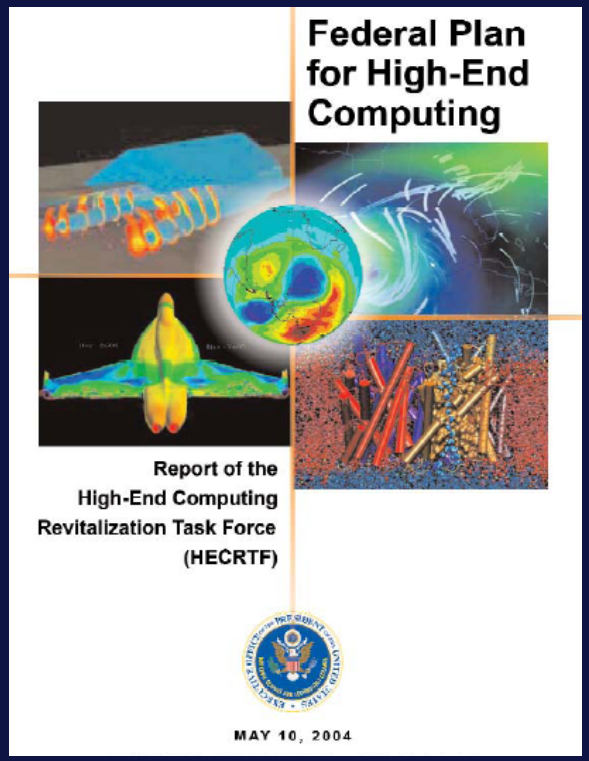
Intel Developer Forum, Spring 2004
Pat Gelsinger
Chief Technology Officer, Senior Vice President
Intel Corporation
February, 19, 2004

Challenge: High Productivity Computing

"Make high-end computing easier and more productive to use.

Emphasis should be placed on time to solution, the major metric of value to high-end computing users…

A common software environment for scientific computation encompassing desktop to high-end systems will enhance productivity gains by promoting ease of use and manageability of systems."



2004 High-End Computing Revitalization Task Force

Office of Science and Technology Policy, Executive Office of the President

Example: Productivity in Cancer Research

FDA Clinical Proteomics Website

1 Regular download of new datasets

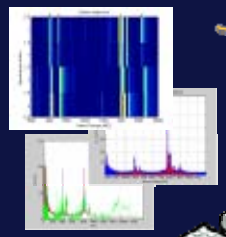
SQL Server Integration Services

2 Signal conditioning and preprocessing

Windows HPC Cluster

SQL 2005

3 Feature extraction



Matlab

4 Classification & training Cancer diagnostics

Personal cluster Windows HPC + Matlab

6 Interoperability with Linux/Grid

Linux Cluster

5 Remote, hosted HPC services

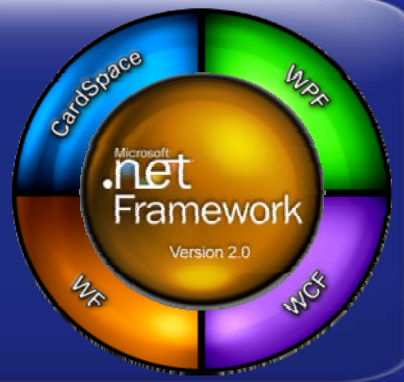
Hosted Windows HPC

Microsoft HPC++ Solution

Application Benefits

The most productive distributed application development environment

Microsoft Office SharePoint Server 2007
Microsoft Office Excel 2007
Microsoft SQL Server 2008
Microsoft Visual Studio 2008



Cluster Benefits

Complete HPC cluster platform integrated with the rest of the enterprise management infrastructure

Windows Server 2008
Active Directory
Microsoft System Center
Windows HPC Server 2008

System Benefits

Cost-effective, reliable and high performance server operating system

Microsoft Hyper-V Server
Windows Server 2008

Analysis

SQL Server 2005

Research Intelligence

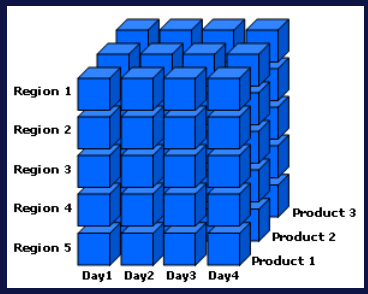
- Data acquisition from source systems and integration
- Data transformation and synthesis

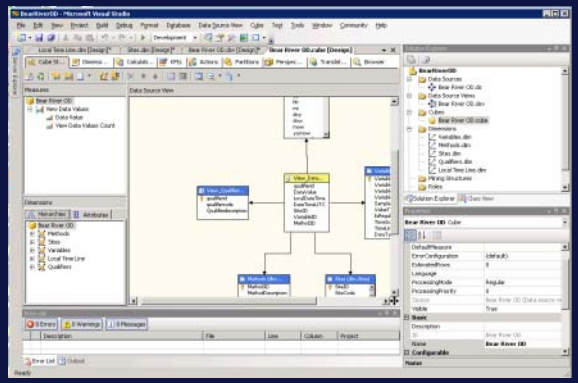
- Data enrichment, with business logic, hierarchical views
- Data discovery via data mining

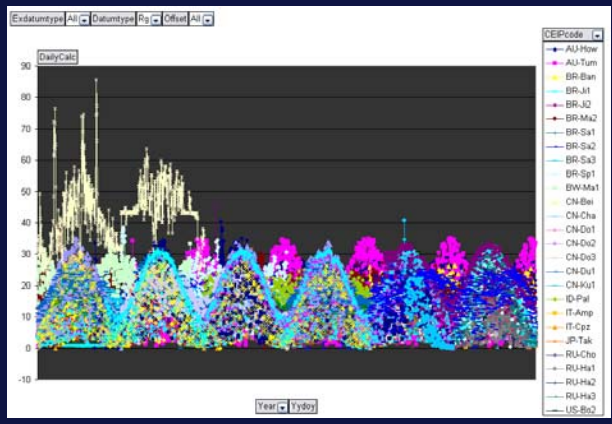
- Data presentation and distribution
- Data access for the masses

Data Cube Basics

- A data cube is a database specifically for data mining (OLAP)
  - Initially developed for commercial needs like tracking sales of Oreos and milk
  - Simple aggregations (sum, min, or max) can be pre-computed for speed
  - Hierarchies for simple filtering with drilldown capability
  - Additional calculations (median) can be computed dynamically or pre-computed
  - All operate along dimensions such as time, site, or datumtype
  - Constructed from a relational database
- Client tool integration is evolving
  - Excel PivotTables allow simple data viewing
  - More powerful analysis and plotting using Matlab and statistics software



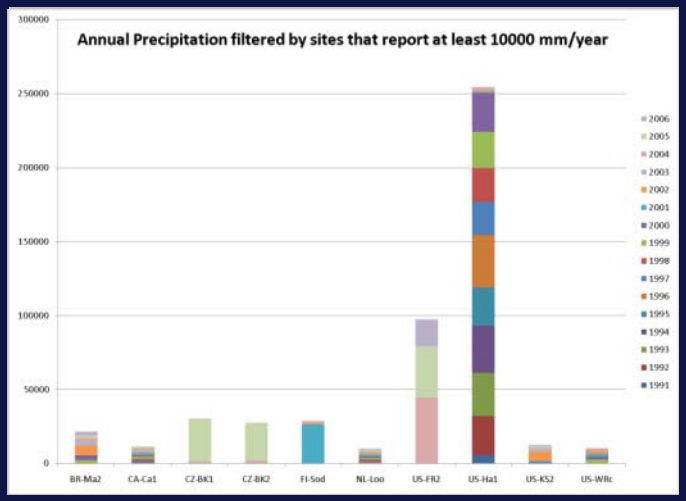


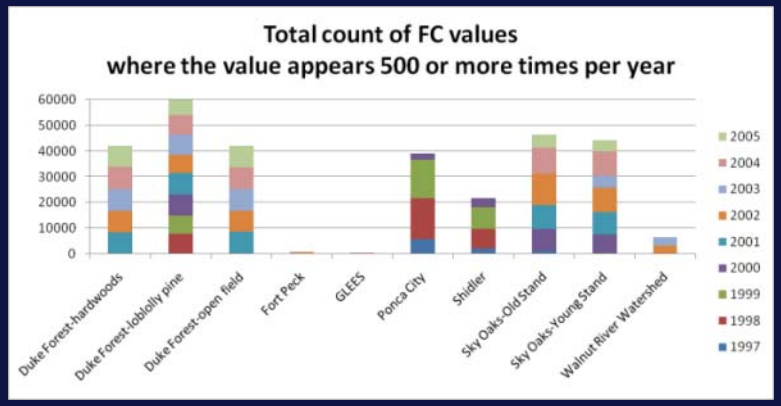


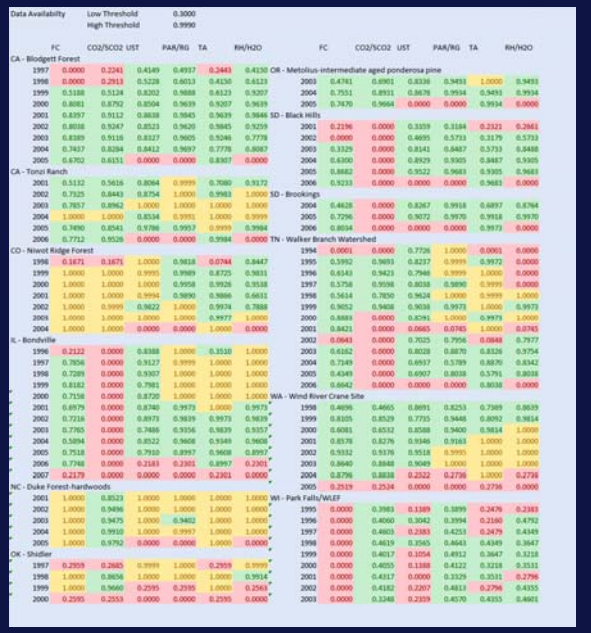
Courtesy Catherine VanIngen, MSR

Browsing for Data Quality

- Data cleaning never ends
  - Running scripts on specific site years often misses the big picture
  - Spike detection and replacement



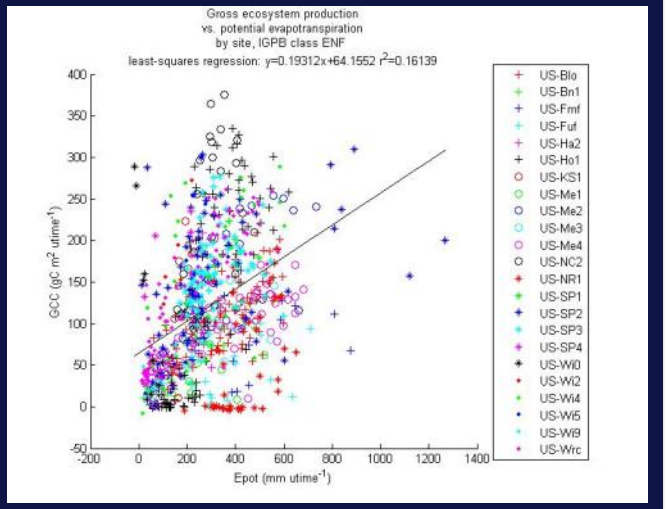


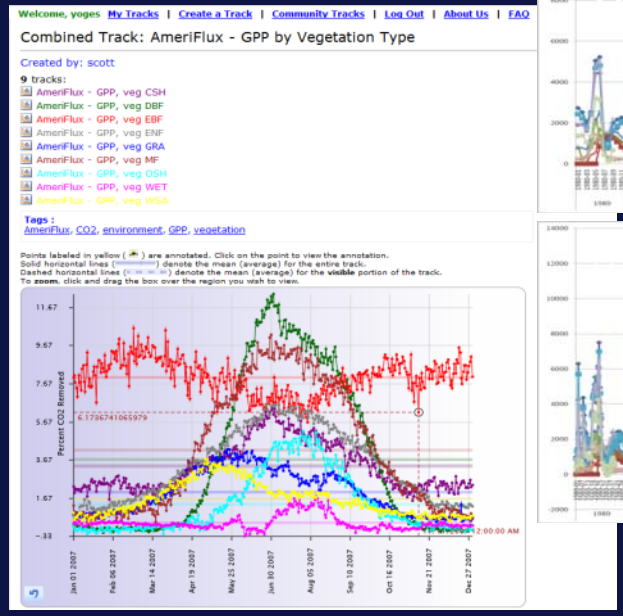


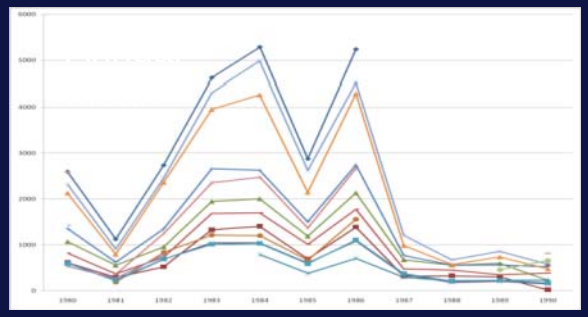
Courtesy Catherine VanIngen, MSR

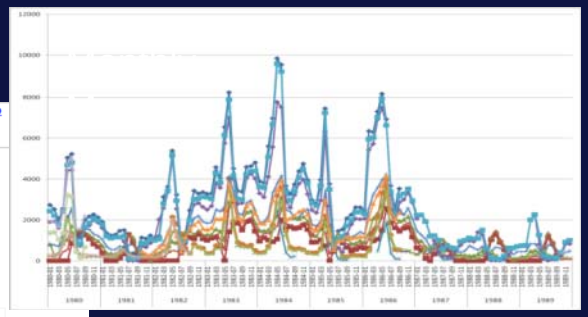
Browsing for Data Analysis

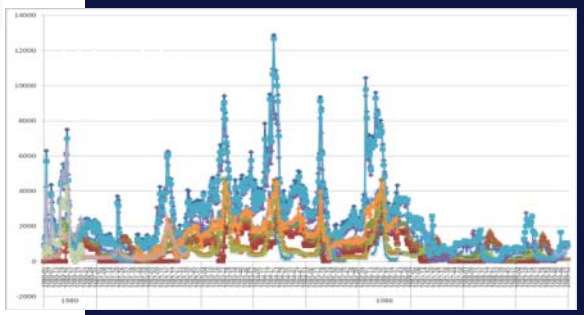
- Plotting is the way of visualizing data
  - Most are discarded so scripting matters





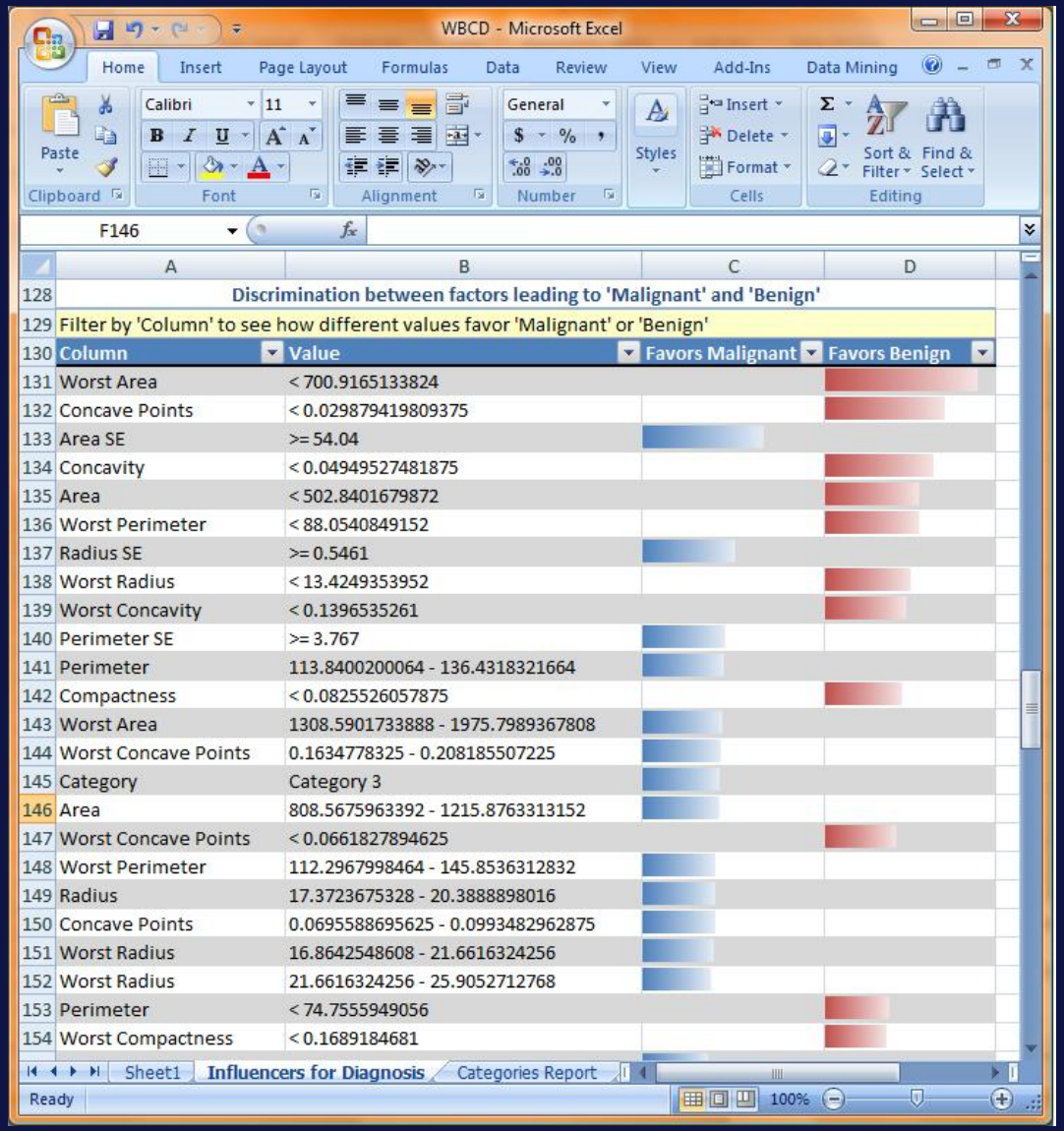






Courtesy Catherine VanIngen, MSR

Datamining with SQL and Excel

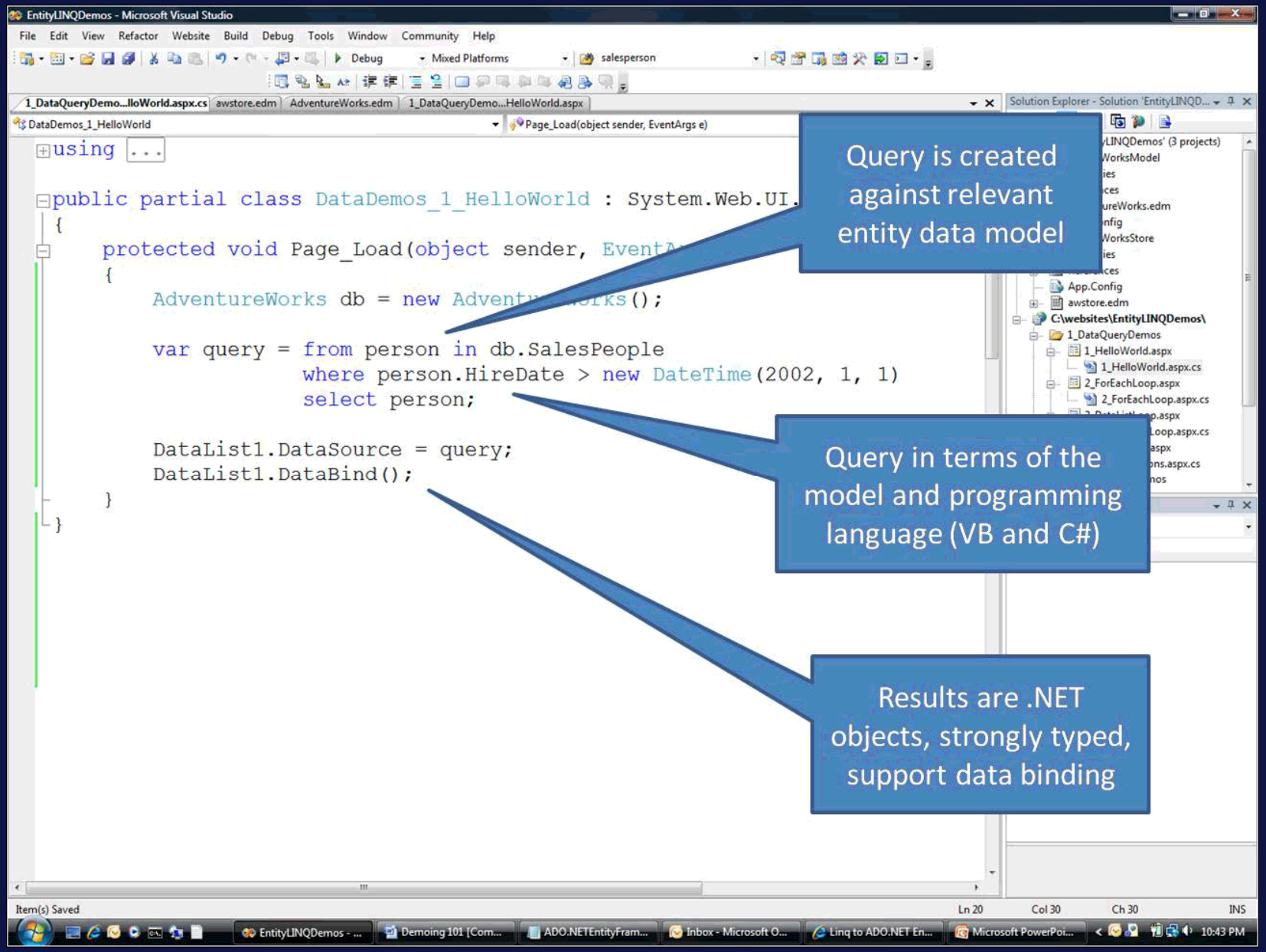


Column	Value	Favors Malignant	Favors Benign
Worst Area	< 700.9165133824		
Concave Points	< 0.029879419809375		
Area SE	>= 54.04		
Concavity	< 0.04949527481875		
Area	< 502.8401679872		
Worst Perimeter	< 88.0540849152		
Radius SE	>= 0.5461		
Worst Radius	< 13.4249353952		
Worst Concavity	< 0.1396535261		
Perimeter SE	>= 3.767		
Perimeter	113.8400200064 - 136.4318321664		
Compactness	< 0.0825526057875		
Worst Area	1308.5901733888 - 1975.7989367808		
Worst Concave Points	0.1634778325 - 0.208185507225		
Category	Category 3		
Area	808.5675963392 - 1215.8763313152		
Worst Concave Points	< 0.0661827894625		
Worst Perimeter	112.2967998464 - 145.8536312832		
Radius	17.3723675328 - 20.3888898016		
Concave Points	0.0695588695625 - 0.0993482962875		
Worst Radius	16.8642548608 - 21.6616324256		
Worst Radius	21.6616324256 - 25.9052712768		
Perimeter	< 74.7555949056		
Worst Compactness	< 0.1689184681		

Integrated algorithms

- Text Mining
- Neural Nets
- Naïve Bayes
- Time Series
- Sequent Clustering
- Decision Trees
- Association Rules

Language Integrated Query



Simulation

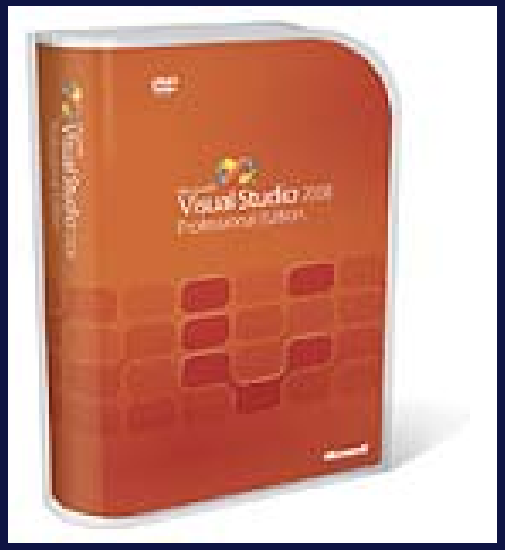
High Productivity Modeling with Visual Studio

Languages/Runtimes

- C++, C#, VB
- F#, Python, Ruby, Jscript
- Fortran (Intel, PGI)
- OpenMP, MPI

Team Development

- Team portal: version control, scheduled build, bug tracking
- Test and stress generation
- Code analysis, Code coverage
- Performance analysis



.Net Framework

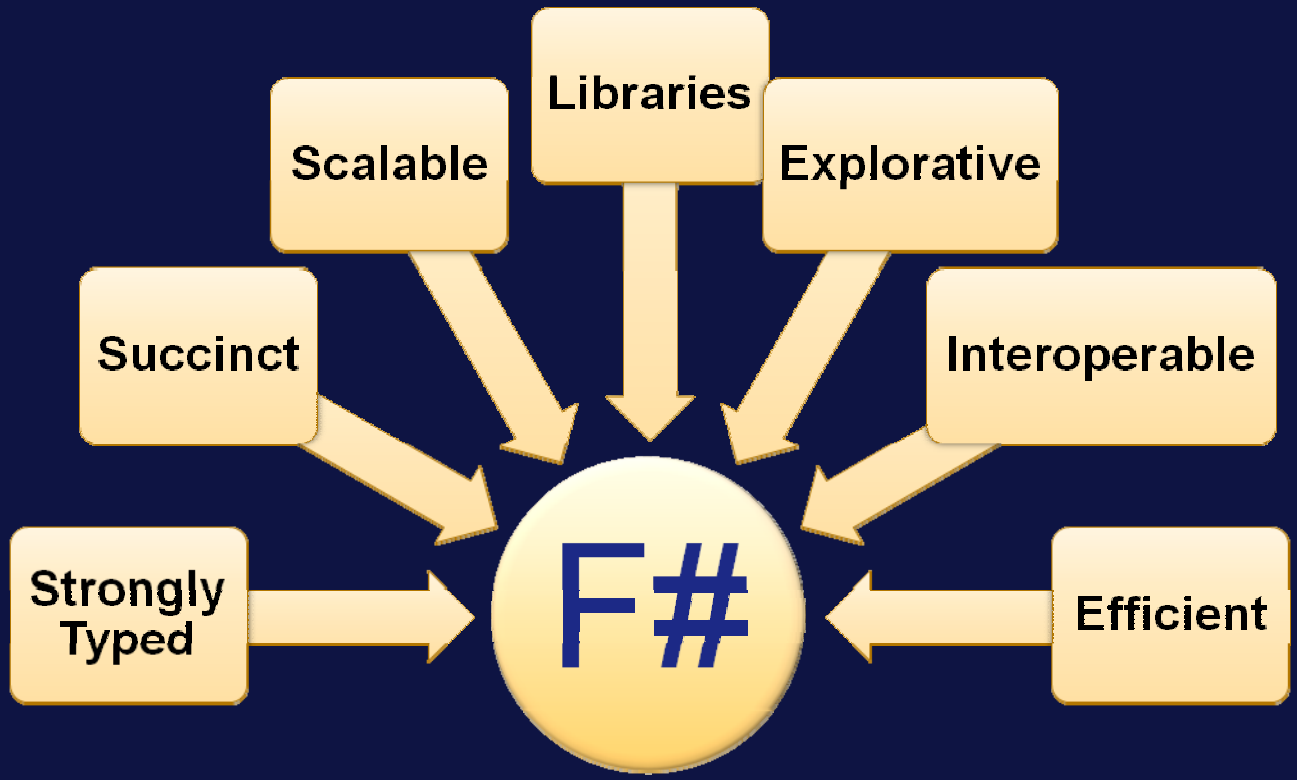
- LINQ: language integrated query
- Dynamic Language Runtime
- Fx/JIT/GC improvements
- Native support for Web Services

IDE

- Rapid application development
- Parallel debugging
- Multiprocessor builds
- Work flow design

F# is...

...a **functional, object-oriented, imperative and explorative** programming language for .NET



Interactive F# Shell

```
C:\fsharpv2>bin\fsi

MSR F# Interactive, (c) Microsoft Corporation, All Rights Reserved
F# Version 1.9.2.9, compiling for .NET Framework Version v2.0.50727

NOTE:
NOTE: See 'fsi --help' for flags
NOTE:
NOTE: Commands: #r <string>;; reference (dynamically load) the given DLL.
NOTE: #I <string>;; add the given search path for referenced DLLs.
NOTE: #use <string>;; accept input from the given file.
NOTE: #load <string> ...<string>;;
NOTE: load the given file(s) as a compilation unit.
NOTE: #time;; toggle timing on/off.
NOTE: #types;; toggle display of types on/off.
NOTE: #quit;; exit.
NOTE:
NOTE: Visit the F# website at http://research.microsoft.com/fsharp.
NOTE: Bug reports to fsbugs@microsoft.com. Enjoy!

> let rec f x = (if x < 2 then x else f (x-1) + f (x-2));;

val f : int -> int

> f 6;;

val it = 8
val it : int
```

Example: Taming Asynchronous I/O

```
using System;
using System.IO;
using System.Threading;

public class BulkImageProcAsync
{
    public const String ImageBaseNam
    public const int numImages = 200
    public const int numPixels = 512

    // ProcessImage has a simple O(N
    // of times you repeat that loop
    // bound or more IO-bound.
    public static int processImageRe

    // Threads must decrement NumIma
    // their access to it through a
    public static int NumImagesToFin
    public static Object[] NumImages
    // WaitObject is signalled when
    public static Object[] WaitObjec
    public class ImageStateObject
    {
        public byte[] pixels;
        public int imageNum;
        public FileStream fs;
    }
```

```
public static void ReadInImageCallback(IAsyncResult as
{
    ImageStateObject state = (ImageStateObject)asyncRe
    Stream stream = state.fs;
    int bytesRead = stream.EndRead(asyncResult);
    if (bytesRead != numPixels)
        throw new Exception(String.Format
            ("In ReadInImageCallback, got the wrong nu
            "bytes from the image: {0}.", bytesRead));
    ProcessImage(state.pixels, state.imageNum);
    stream.Close();

    // Now write out the image.
    // Using asynchronous I/O here appears not to be b
    // It ends up swamping the threadpool, because the
    // threads are blocked on I/O requests that were j
    // the threadpool.
    FileStream fs = new FileStream(ImageBaseName + sta
        ".done", FileMode.Create, FileAccess.Write, Fi
        4096, false);
    fs.Write(state.pixels, 0, numPixels);
    fs.Close();

    // This application model uses too much memory.
    // Releasing memory as soon as possible is a good
    // especially global state.
    state.pixels = null;
    fs = null;
    // Record that an image is finished now.
    lock (NumImagesMutex)
    {
        NumImagesToFinish--;
        if (NumImagesToFinish == 0)
        {
            Monitor.Enter(WaitObject);
            Monitor.Pulse(WaitObject);
            Monitor.Exit(WaitObject);
        }
    }
}
```

```
    public static void ProcessImagesInBulk()
    {
        Console.WriteLine("Processing images...  ");
        long t0 = Environment.TickCount;
        NumImagesToFinish = numImages;
        AsyncCallback readImageCallback = new
            AsyncCallback(ReadInImageCallback);
        for (int i = 0; i < numImages; i++)
        {
            ImageStateObject state = new ImageStateObject();
            state.pixels = new byte[numPixels];
            state.imageNum = i;
            // Very large items are read only once, so you can make the
            // buffer on the FileStream very small to save memory.
            FileStream fs = new FileStream(ImageBaseName + i + ".tmp",
                FileMode.Open, FileAccess.Read, FileShare.Read, 1, true);
            state.fs = fs;
            fs.BeginRead(state.pixels, 0, numPixels, readImageCallback,
                state);
        }

        // Determine whether all images are done being processed.
        // If not, block until all are finished.
        bool mustB      false;
        lock (NumImagesM
        {
            if (NumImagesToFinish > 0)
                mustBlock = true;
        }
        if (mustBlock)
        {
            Console.WriteLine("All worke
                " Blocking until they complete. numLeft: {0}",
                NumImagesToFinish);
            Monitor.Enter(WaitObject);
            Monitor.Wait(WaitObject);
            Monitor.Exit(WaitObject);
        }
        long t1 = Environment.TickCount;
        Console.WriteLine("Total time processing images: {0}ms",
            (t1 - t0));
    }
}
```

Processing 200 images in parallel

Example: Taming Asynchronous I/O

Equivalent F# code (same perf)

Open the file, synchronously

Read from the file, asynchronously

```
let ProcessImageAsync(i) =
    async { let  inStream  = File.OpenRead(sprintf "source%d.jpg" i)
            let! pixels    = inStream.ReadAsync(numPixels)
            let  pixels'   = TransformImage(pixels,i)
            let  outStream = File.OpenWrite(sprintf "result%d.jpg" i)
            do! outStream.WriteAsync(pixels')
            do  Console.WriteLine "done!"  }

let ProcessImagesAsync() =
    Async.Run (Async.Parallel
                  [ for i in 1 .. numImages -> ProcessImageAsync(i) ])
```

This object coordinates

Write the result, asynchronously

"!" = "asynchronous"

Generate the tasks and queue them in parallel

Parallel Extensions to .NET

- Declarative data parallelism (PLINQ)

```
var q = from n in names.AsParallel()
        where n.Name == queryInfo.Name &&
              n.State == queryInfo.State &&
              n.Year >= yearStart &&
              n.Year <= yearEnd
        orderby n.Year ascending
        select n;
```

- Imperative data and task parallelism (TPL)

```
Parallel.For(0, n, i=> {
    result[i] = compute(i);
});
```

- Data structures and coordination constructs

Example: Tree Walk

Sequential

```
static void ProcessNode<T>(Tree<T>
tree, Action<T> action) {
    if (tree == null) return;

    ProcessNode(tree.Left, action);
    ProcessNode(tree.Right, action);
    action(tree.Data);
}
```

Thread Pool

```
static void ProcessNode<T>(Tree<T> tree, Action<T> action) {
    if (tree == null) return;

    Stack<Tree<T>> nodes = new Stack<Tree<T>>();
    Queue<T> data = new Queue<T>();

    nodes.Push(tree);
    while (nodes.Count > 0) {
        Tree<T> node = nodes.Pop();
        data.Enqueue(node.Data);
        if (node.Left != null) nodes.Push(node.Left);
        if (node.Right != null) nodes.Push(node.Right);
    }

    using (ManualResetEvent mre = new ManualResetEvent(false)) {
        int waitCount = Environment.ProcessorCount;

        WaitCallback wc = delegate {
            bool gotItem;
            do {
                T item = default(T);
                lock (data) {
                    if (data.Count > 0) {
                        item = data.Dequeue();
                        gotItem = true;
                    }
                    else gotItem = false;
                }
                if (gotItem) action(item);
            } while (gotItem);

            if (Interlocked.Decrement(ref waitCount) == 0) mre.Set();
        };

        for (int i = 0; i < Environment.ProcessorCount - 1; i++) {
            ThreadPool.QueueUserWorkItem(wc);
        }

        wc(null);
        mre.WaitOne();
    }
}
```

Example: Tree Walk

Parallel Extensions (with Task)

```
static void ProcessNode<T>(Tree<T> tree, Action<T> action) {
    if (tree == null) return;

    Task t = Task.Create(delegate { ProcessNode(tree.Left, action); });
    ProcessNode(tree.Right, action);
    action(tree.Data);
    t.Wait();
}
```

Parallel Extensions (with Parallel)

```
static void ProcessNode<T>(Tree<T> tree, Action<T> action) {
    if (tree == null) return;

    Parallel.Do(
        () => ProcessNode(tree.Left, action),
        () => ProcessNode(tree.Right, action),
        () => action(tree.Data) );
}
```

Parallel Extensions (with PLINQ)

```
static void ProcessNode<T>(Tree<T> tree, Action<T> action) {
    tree.AsParallel().ForAll(action);
}
```

MPI.NET

- Supports all .NET languages (C#, C++, F#, ..., even Visual Basic!)
- Natural expression of MPI in C#

```
if (world.Rank == 0)
  world.Send(“Hello, World!”, 1, 0);
else
  string msg = world.Receive<string>(0, 0);
```

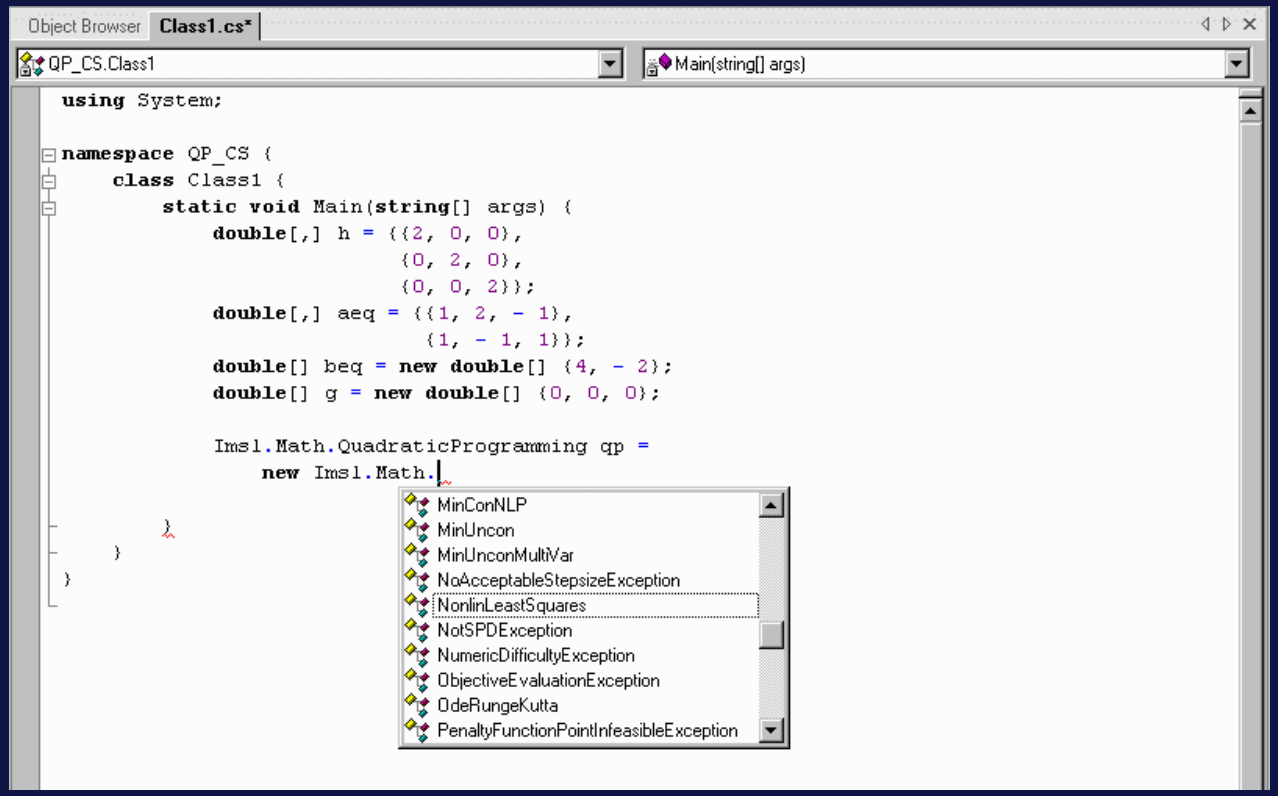
```
string[] hostnames =
comm.Gather(MPI.Environment.ProcessorName, 0);

double pi = 4.0*comm.Reduce(dartsInCircle,(x, y) =>
return x + y, 0) / totalDartsThrown;
```

- Negligible overhead (relative to C) over TCP

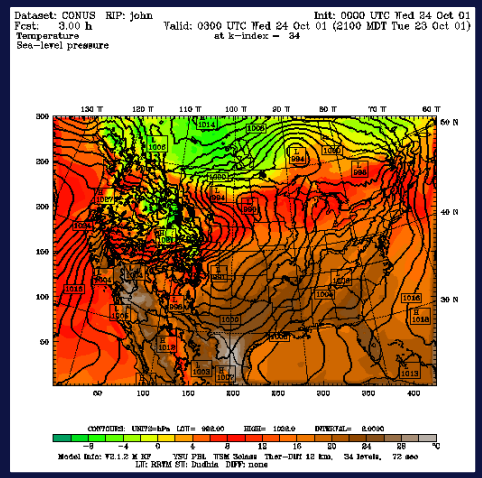
Visual Numerics .NET IMSL Library

- Linear Algebra
- Eigensystems
- Interpolation and Approximation
- Quadrature
- Differential Equations
- Transforms
- Nonlinear Equations
- Optimization
- Basic Statistics
- Nonparametric Tests
- Goodness of Fit
- Regression
- Variances, Covariances and Correlations
- Multivariate Analysis
- Analysis of Variance
- Time Series and Forecasting
- Distribution Functions
- Random Number Generation

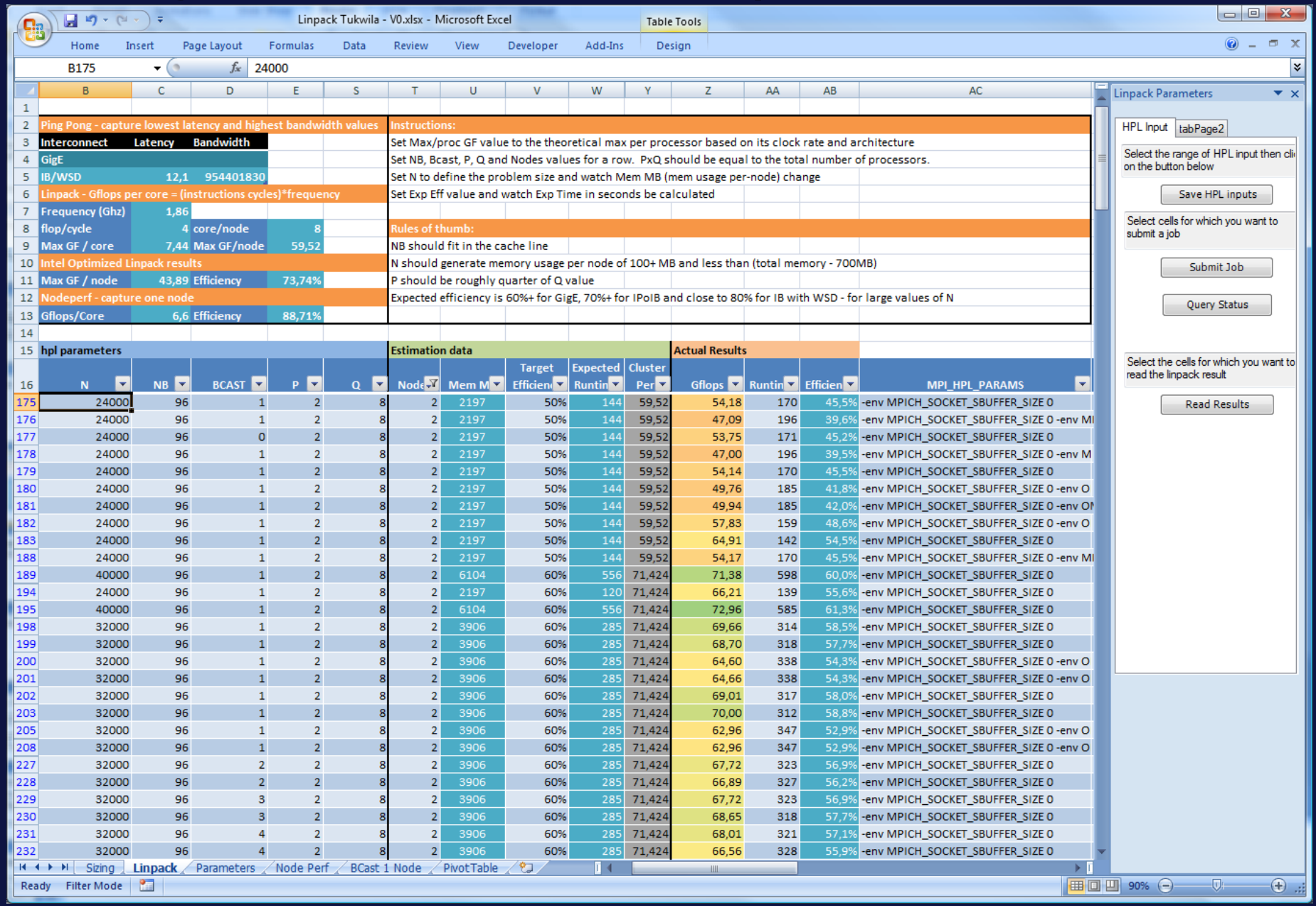


Porting With Subsystem for Unix Applications

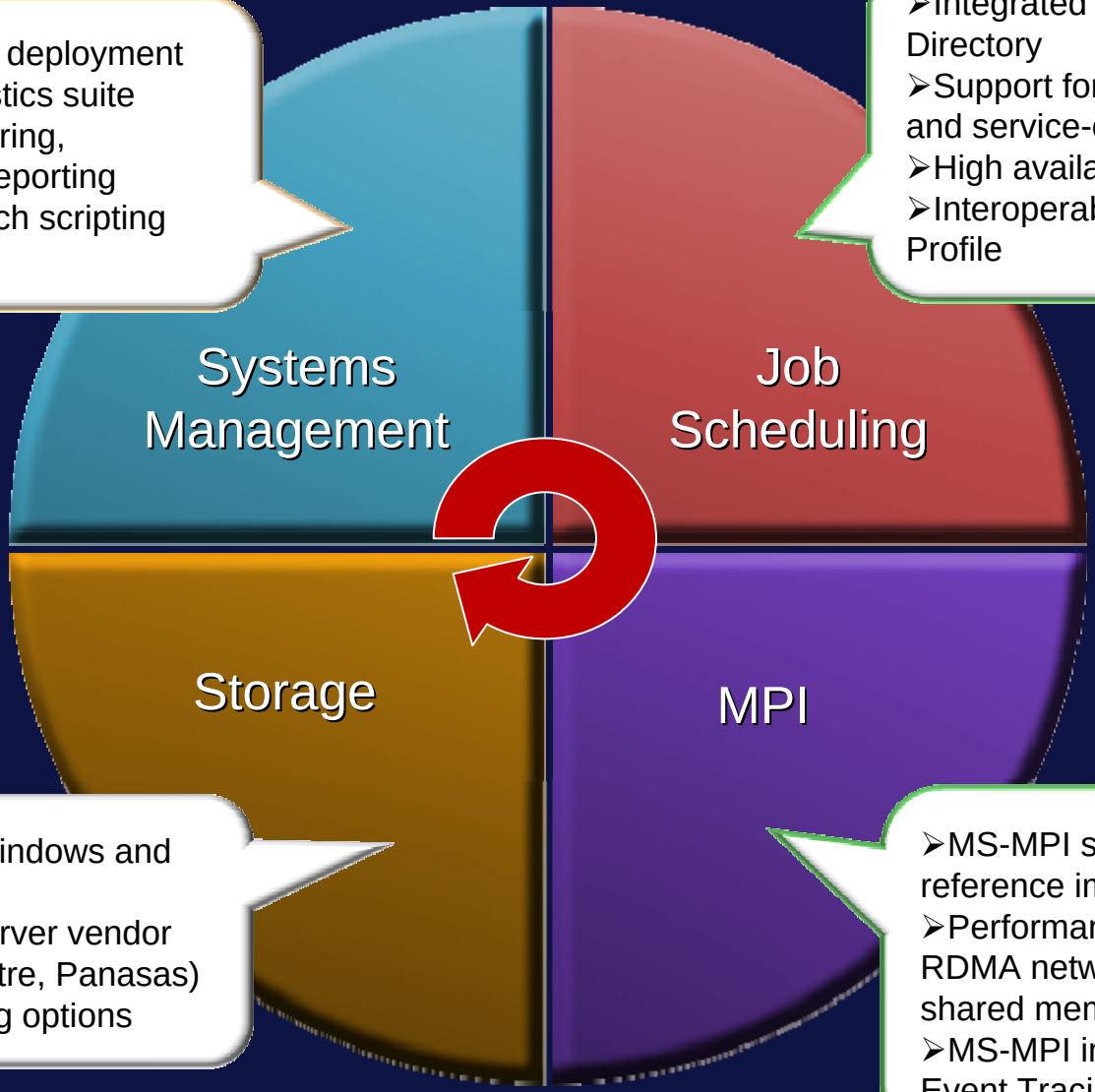
- Subsystem for Unix applications
  - Subsystem for POSIX applications
  - Complete SVR-5 and BSD UNIX environment with 300 commands, utilizes, shell scripts, compilation tools
  - Visual Studio extensions for debugging POSIX applications
  - Support for 32 and 64-bit applications
- Recent port of WRF V2.1.2
  - ~1/3 million lines, Fortran 90 and some C using MPI, OpenMP
  - Traditionally developed for Unix HPC systems
  - Two dynamical cores, full range of physics options
  - Rapid community growth - more than 3,000 registered users
- Porting experience
  - Of 360K lines, fewer than 750 changed to compile and link under Services for Unix
  - Largest number of changes involved the WRF build mechanism (Makefiles, scripts)
  - Level of effort and nature of tasks was not unlike porting to any new version of UNIX
  - Performance on par with the Linux systems



Using Excel to Drive Linpack



Windows HPC Server 2008



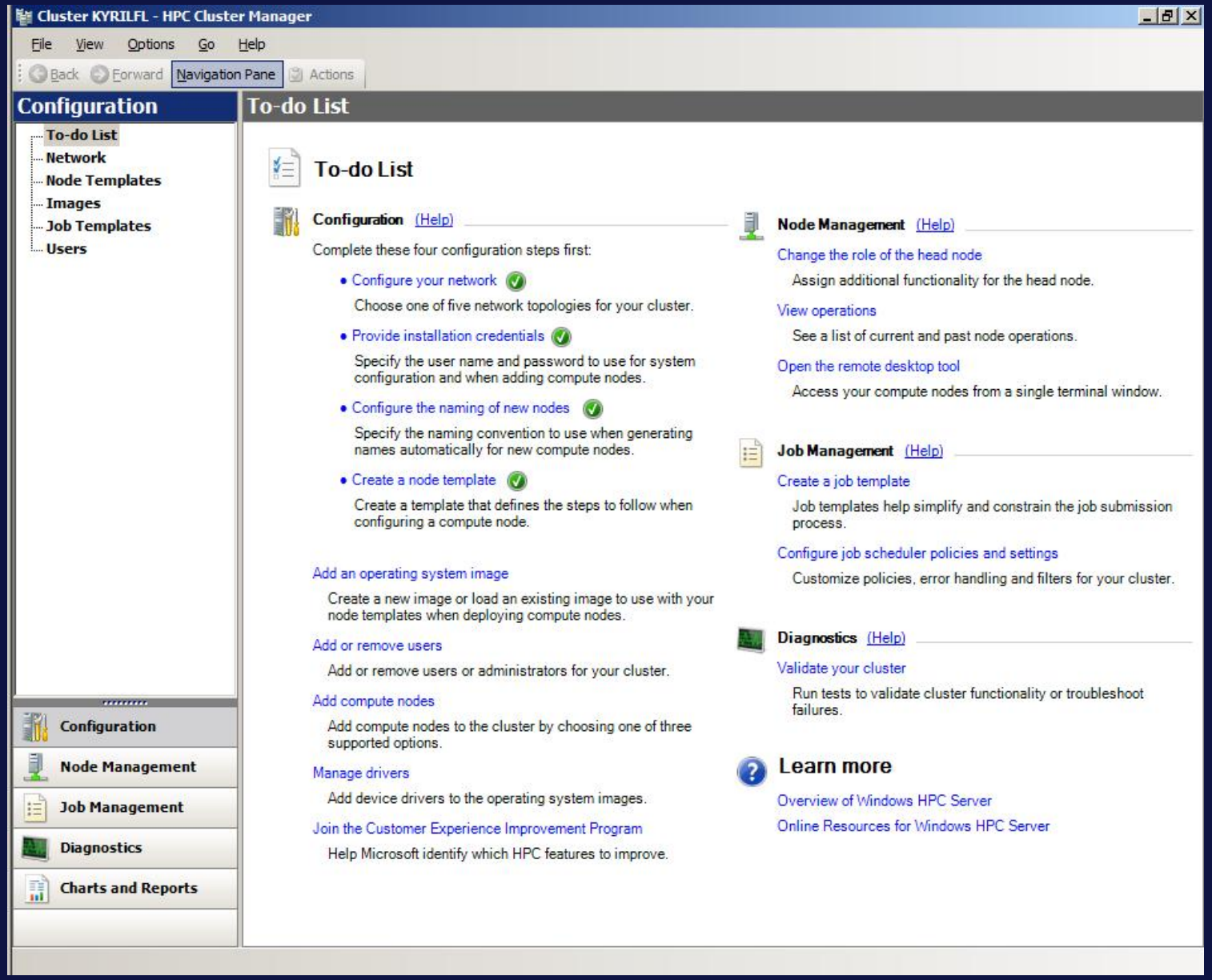
- Rapid large scale deployment and built-in diagnostics suite
- Integrated monitoring, management and reporting
- Familiar UI and rich scripting interface

- Integrated security via Active Directory
- Support for batch, interactive and service-oriented applications
- High availability scheduling
- Interoperability via OGF's HPC Profile

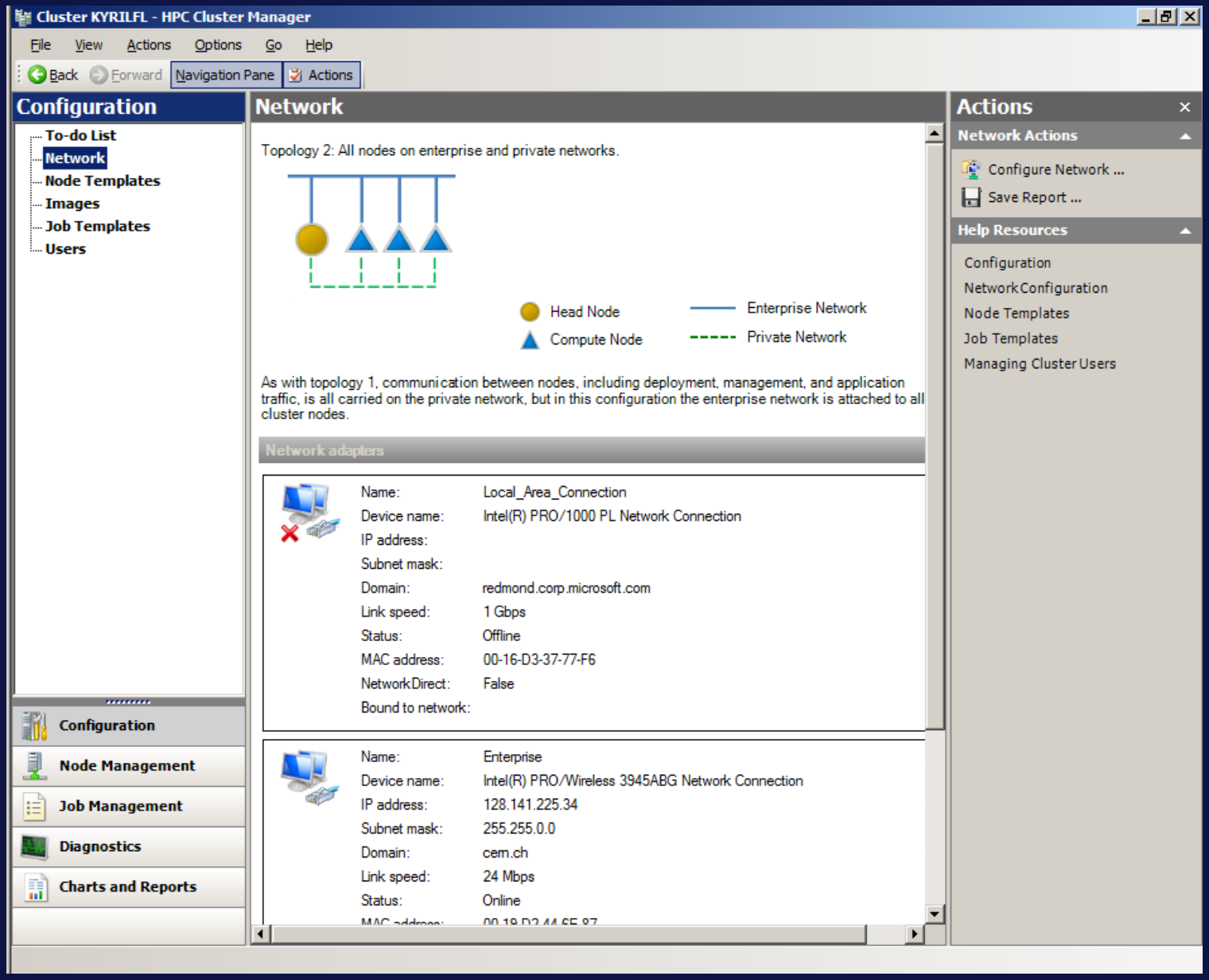
- Access to SQL, Windows and Unix file servers
- Key parallel file server vendor support (GPFS, Lustre, Panasas)
- In-memory caching options

- MS-MPI stack based on MPICH2 reference implementation
- Performance improvements for RDMA networking and multi-core shared memory
- MS-MPI integrated with Windows Event Tracing

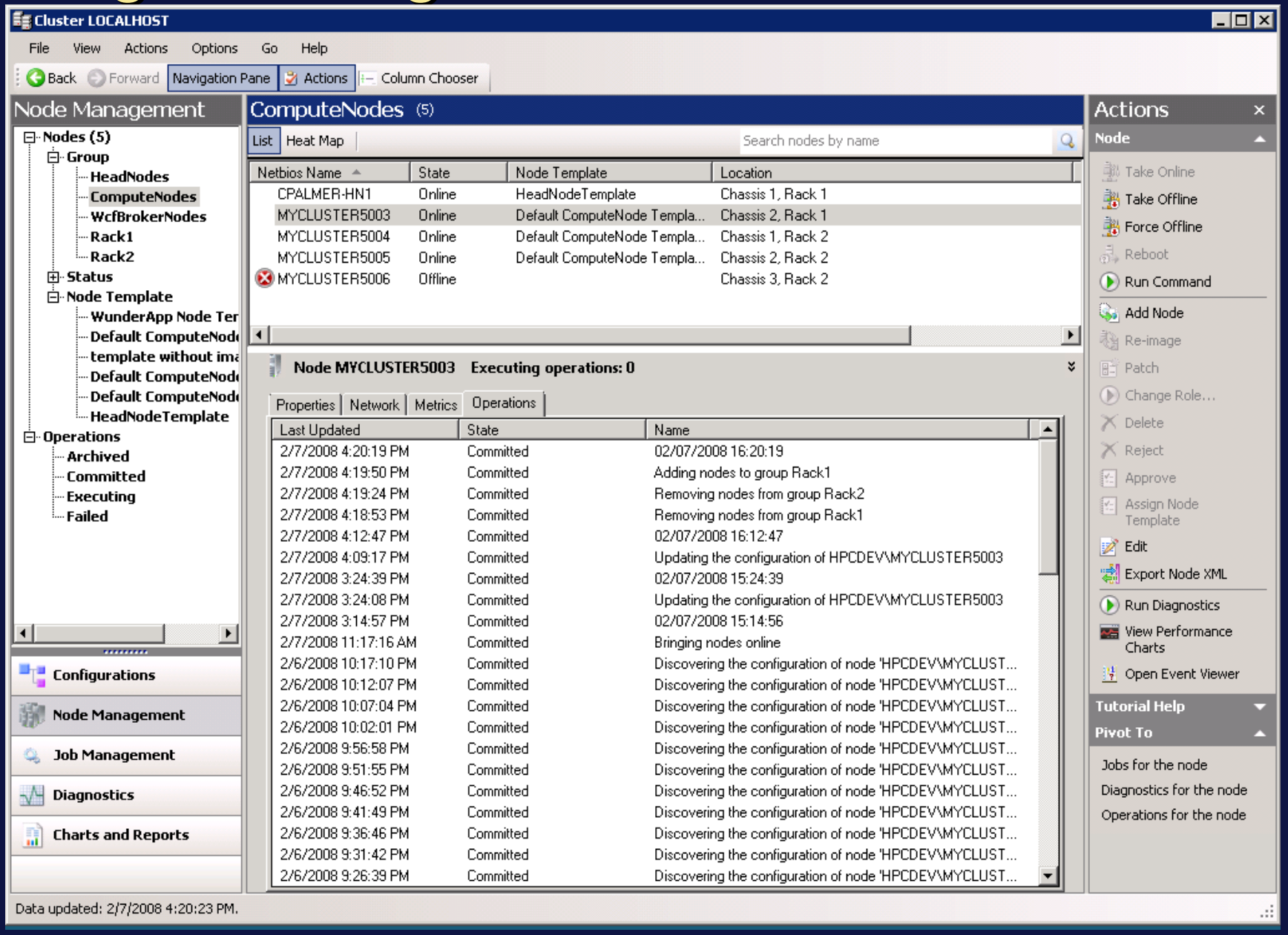
Ease of deployment



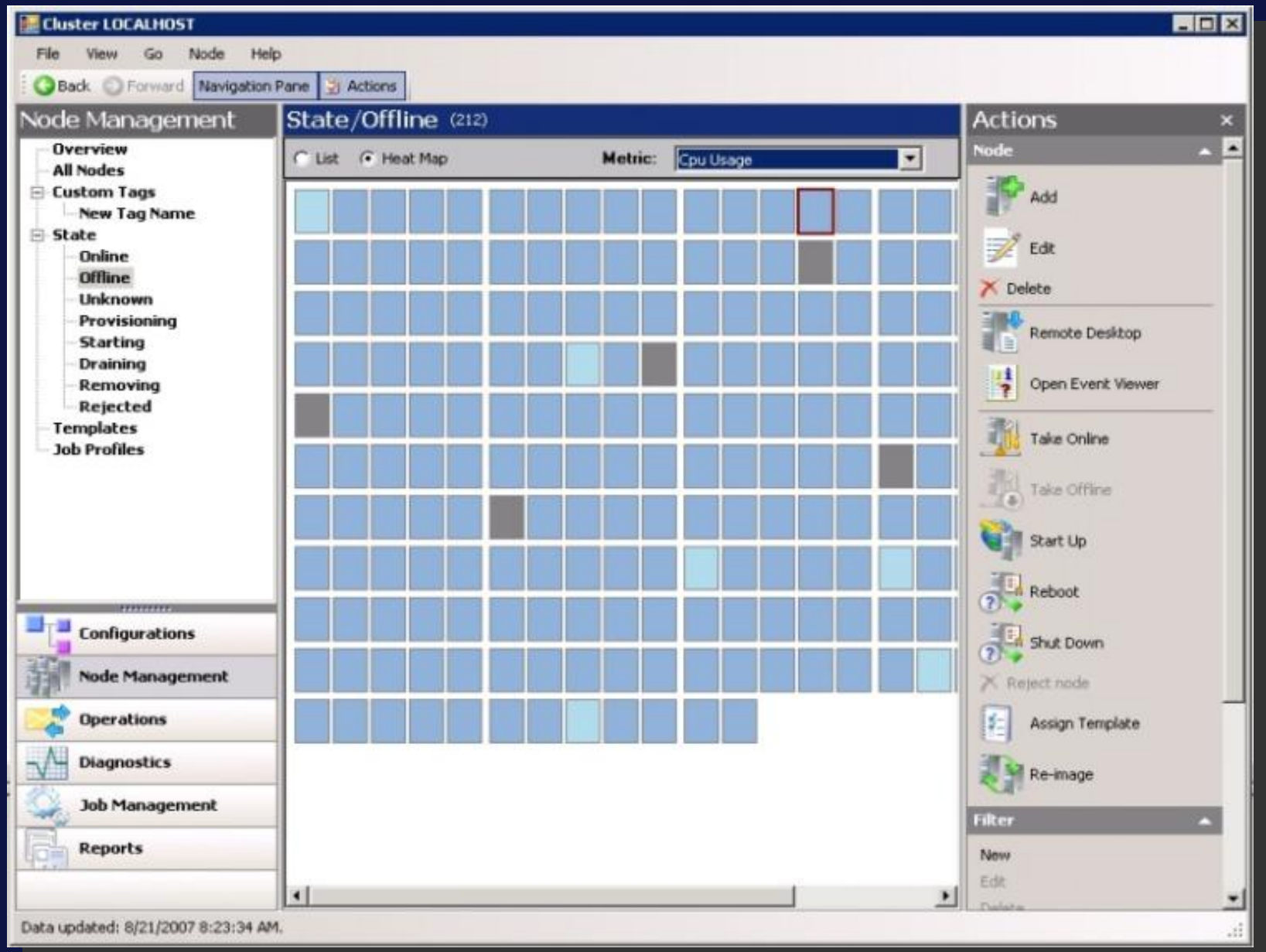
Ease of Deployment



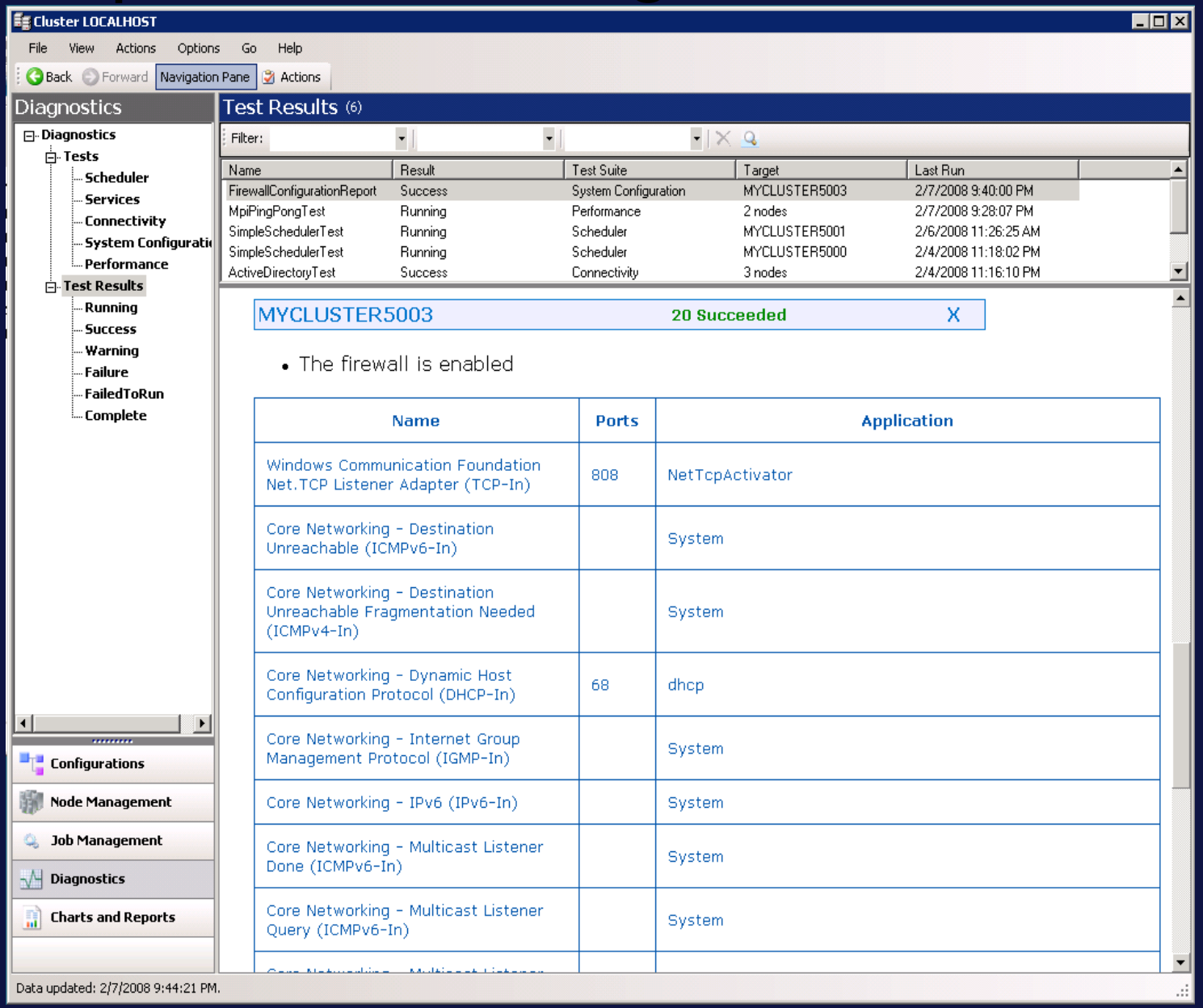
Single Management Console



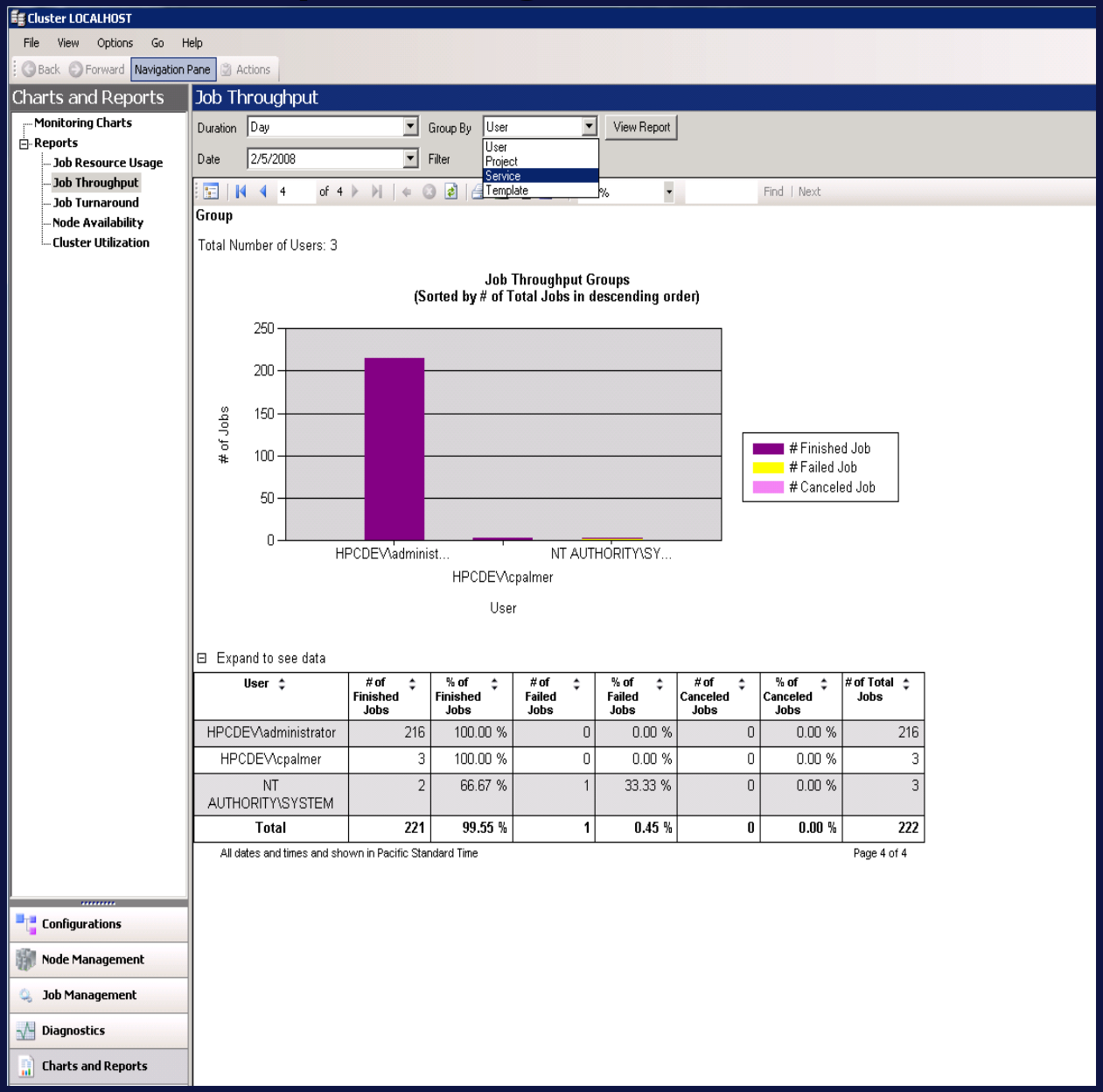
Integrated Monitoring



Comprehensive Diagnostics Suite



Built-in Reporting



User	# of Finished Jobs	% of Finished Jobs	# of Failed Jobs	% of Failed Jobs	# of Canceled Jobs	% of Canceled Jobs	# of Total Jobs
HPCDEV\administrator	216	100.00 %	0	0.00 %	0	0.00 %	216
HPCDEV\cpalmer	3	100.00 %	0	0.00 %	0	0.00 %	3
NT AUTHORITY\SYSTEM	2	66.67 %	1	33.33 %	0	0.00 %	3
Total	**221**	**99.55 %**	**1**	**0.45 %**	**0**	**0.00 %**	**222**

All dates and times and shown in Pacific Standard Time

Page 4 of 4

Spring 2008, NCSA, ~#20
9472 cores, 68.5 TF, 77.7%

Spring 2008, Umea, ~#50
5376 cores, 46 TF, 85.5%

30% efficiency improvement

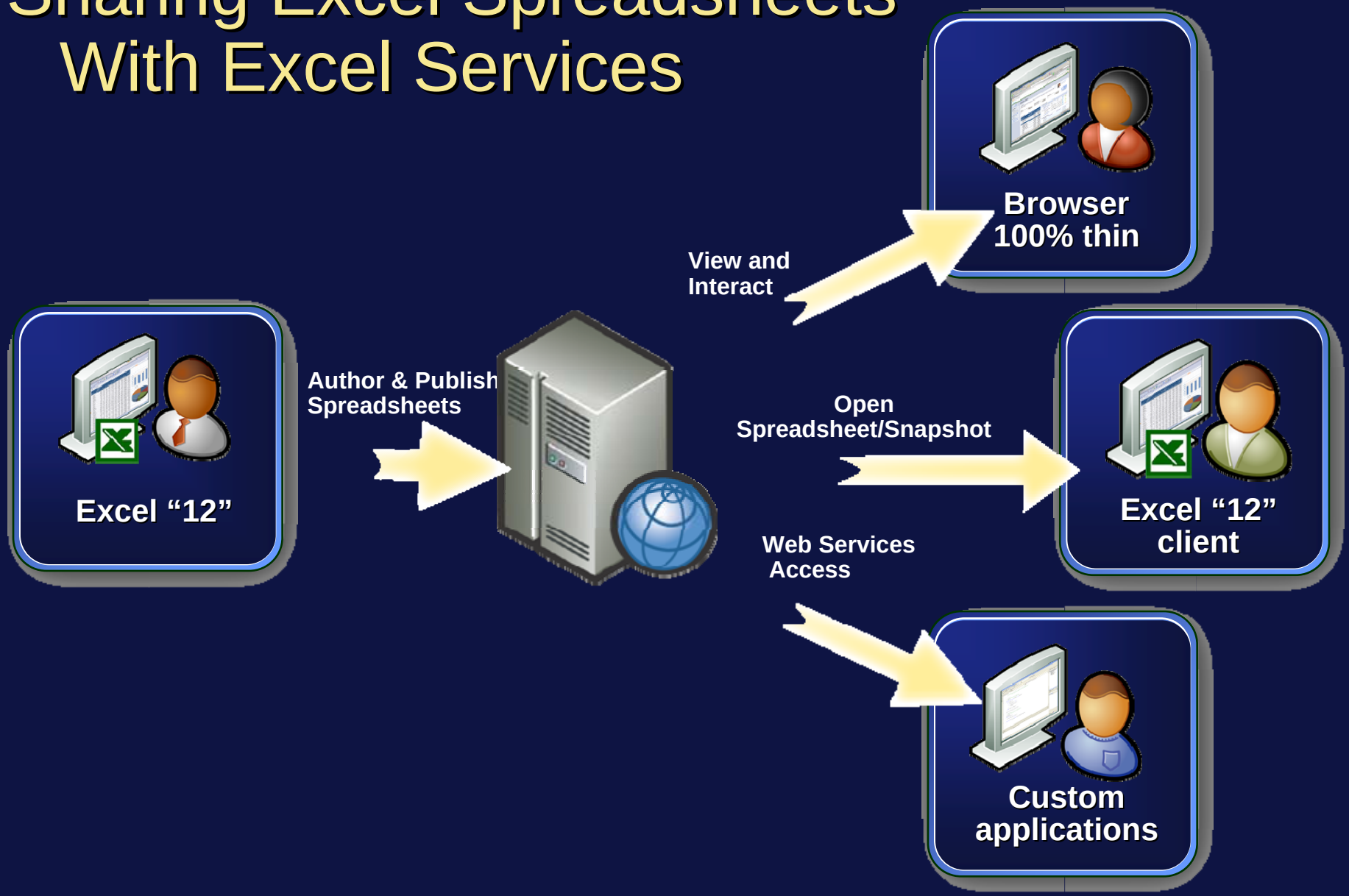
Fall 2007, Microsoft, #116
2048 cores, 11.8 TF, 77.1%

Spring 2007, Microsoft
2048 cores, 9 TF, #106

Spring 2006, NCSA, #130
896 cores, 4.1 TF

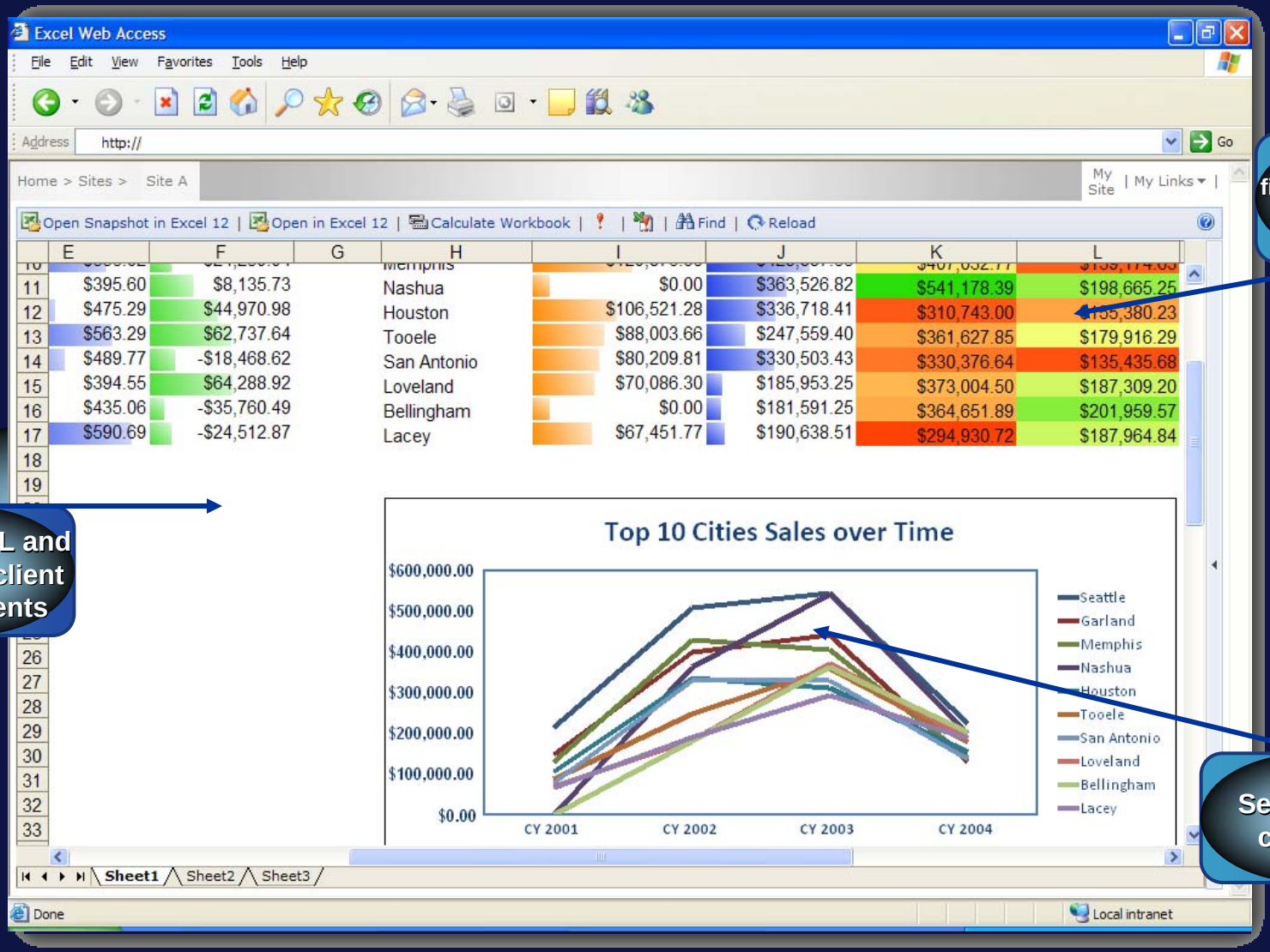
Collaboration

Sharing Excel Spreadsheets With Excel Services



Server-based Spreadsheets

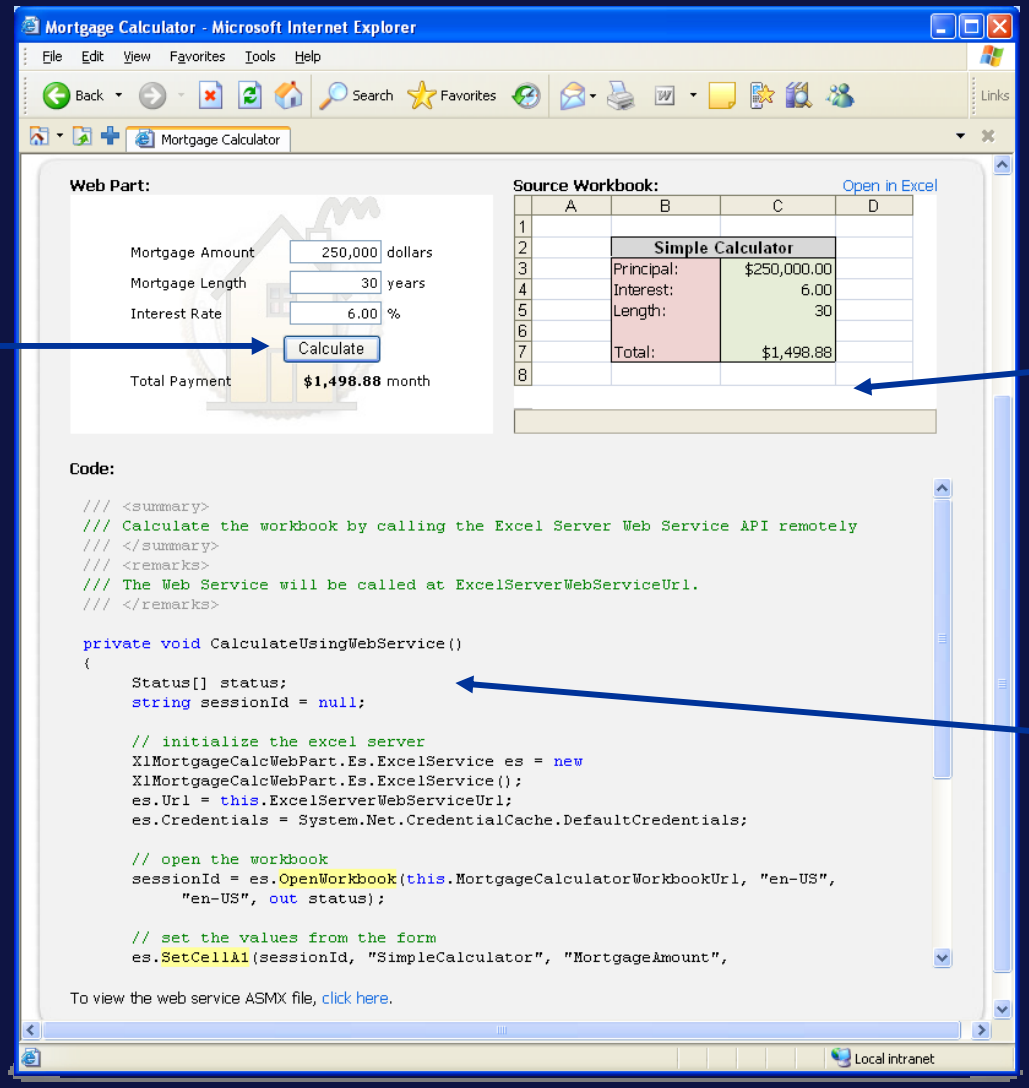
View and interact with spreadsheets in the browser



Reusable Business Logic

Incorporate Excel models in business applications

Calculate spreadsheet and set and get values

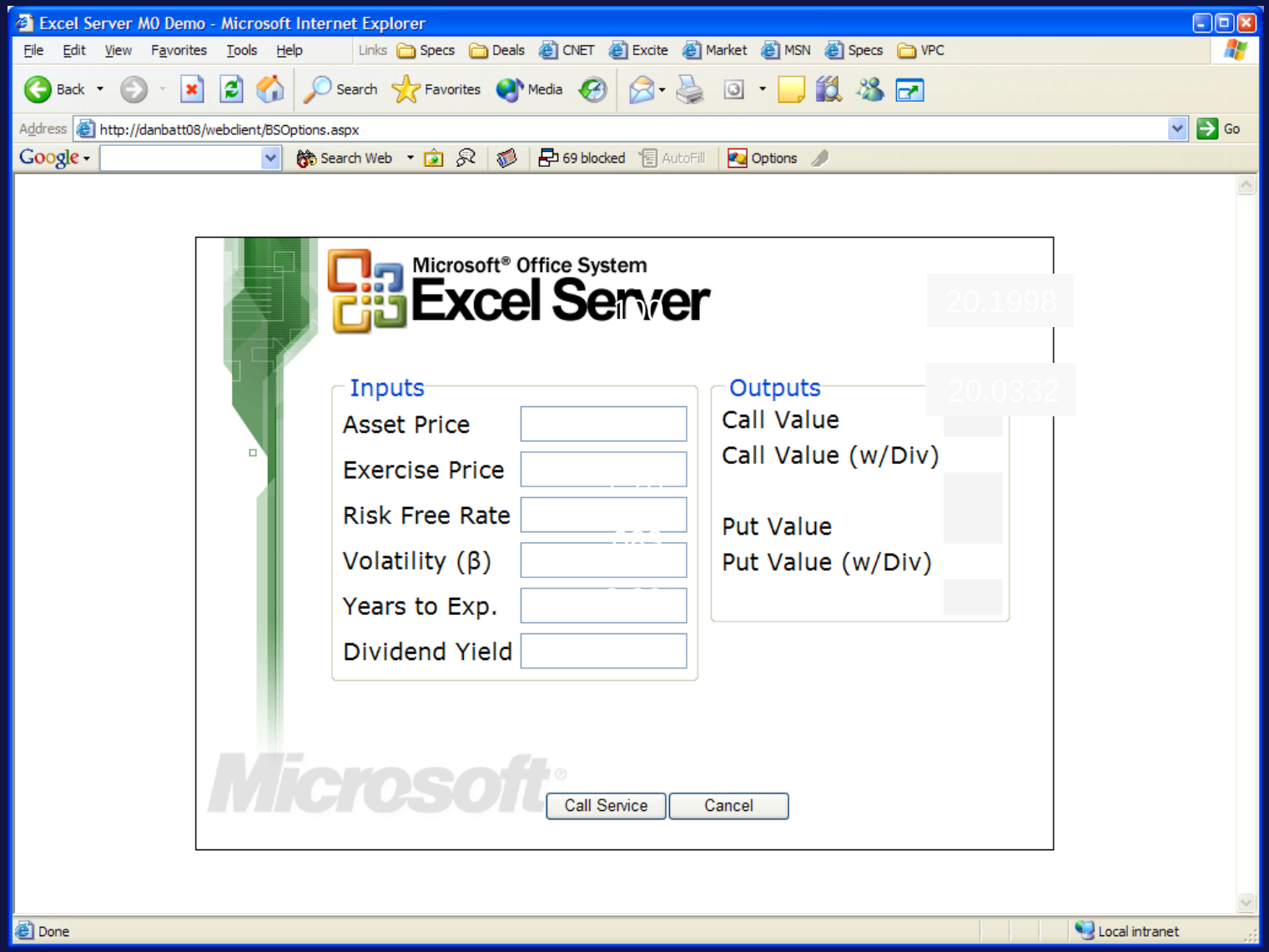


Reusable web-rendered UI

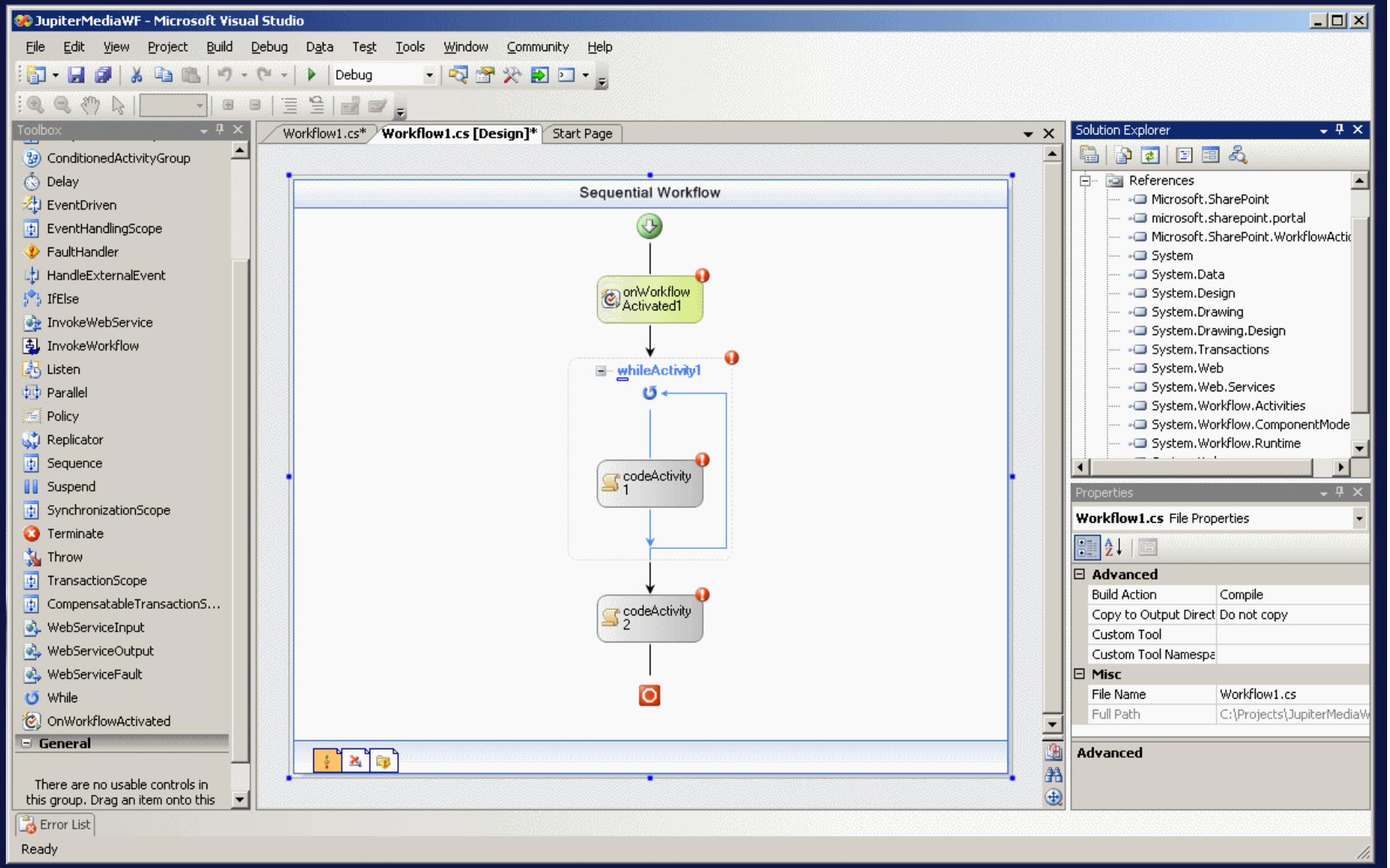
Cross platform web services interface

Reusable Business Logic

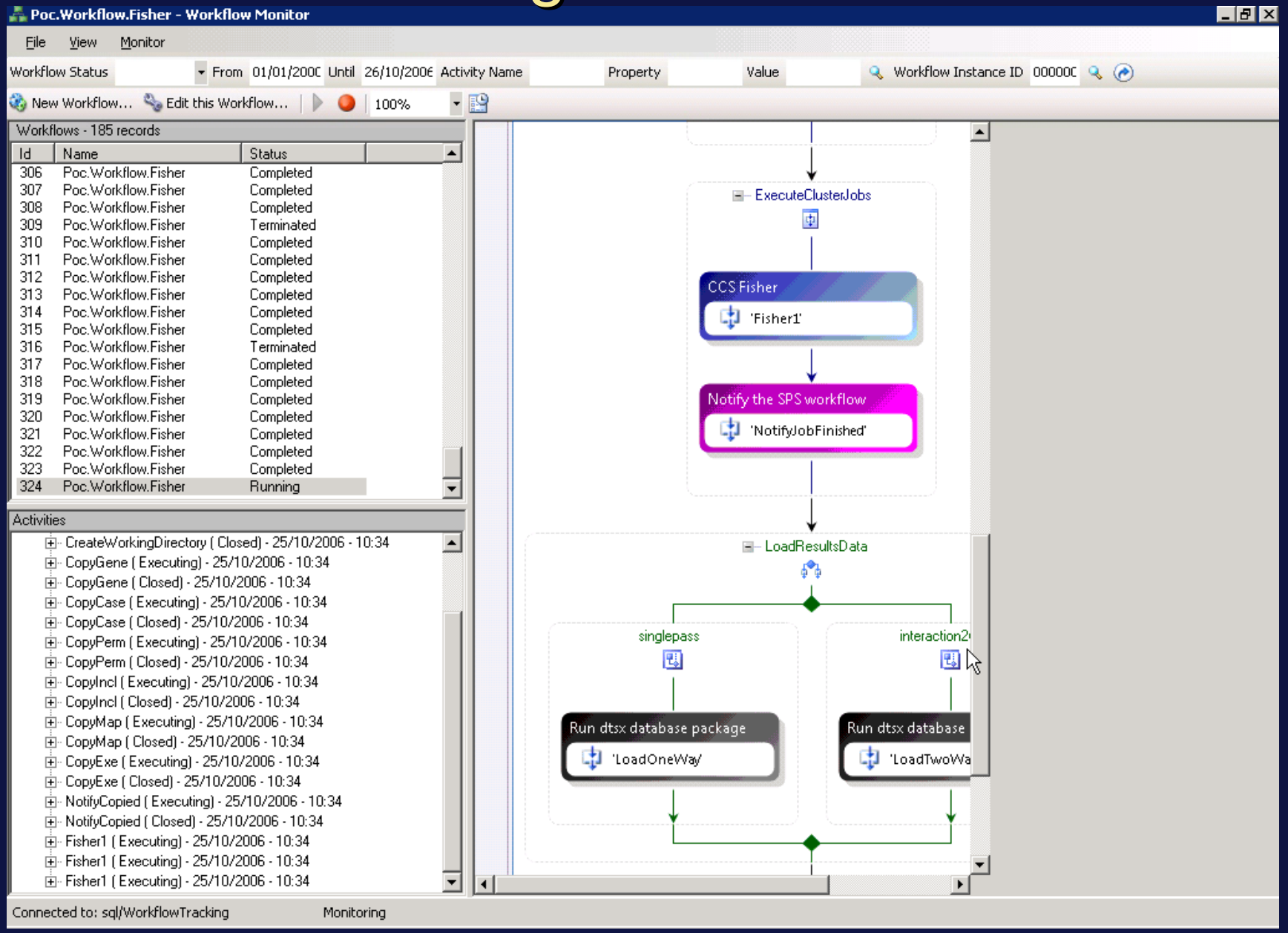
Incorporate Excel models in business applications

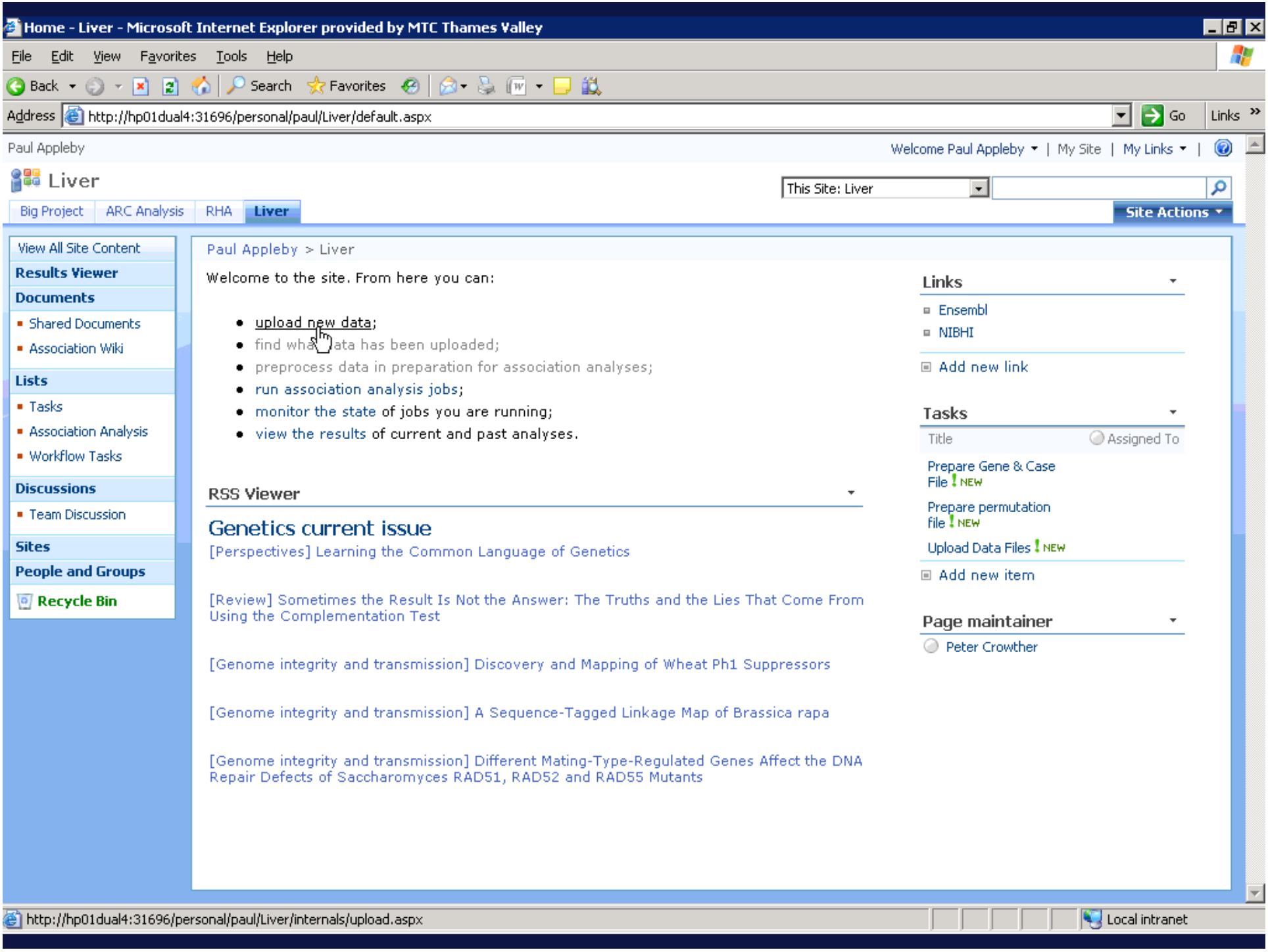


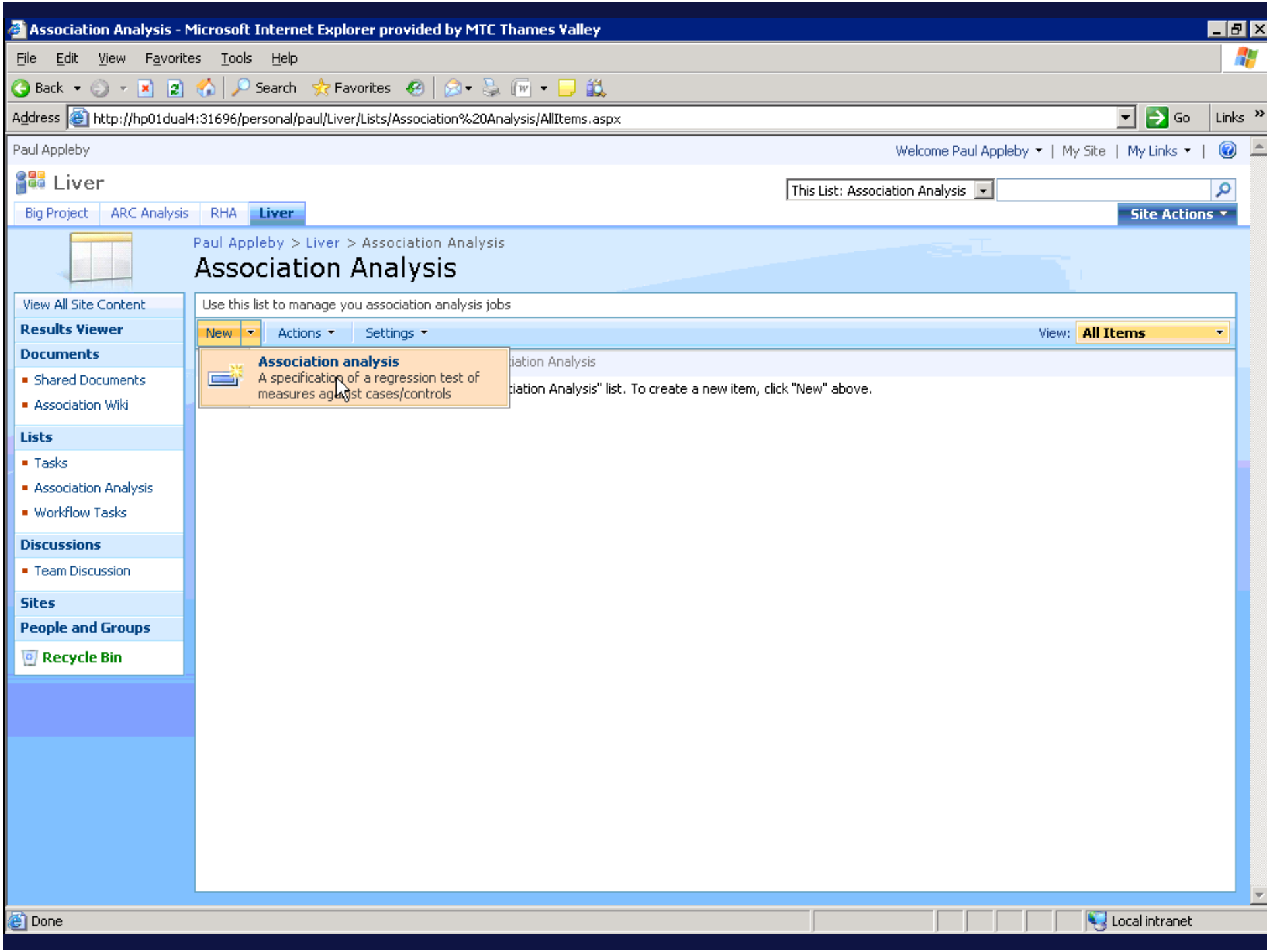
Workflow Design for Sharepoint

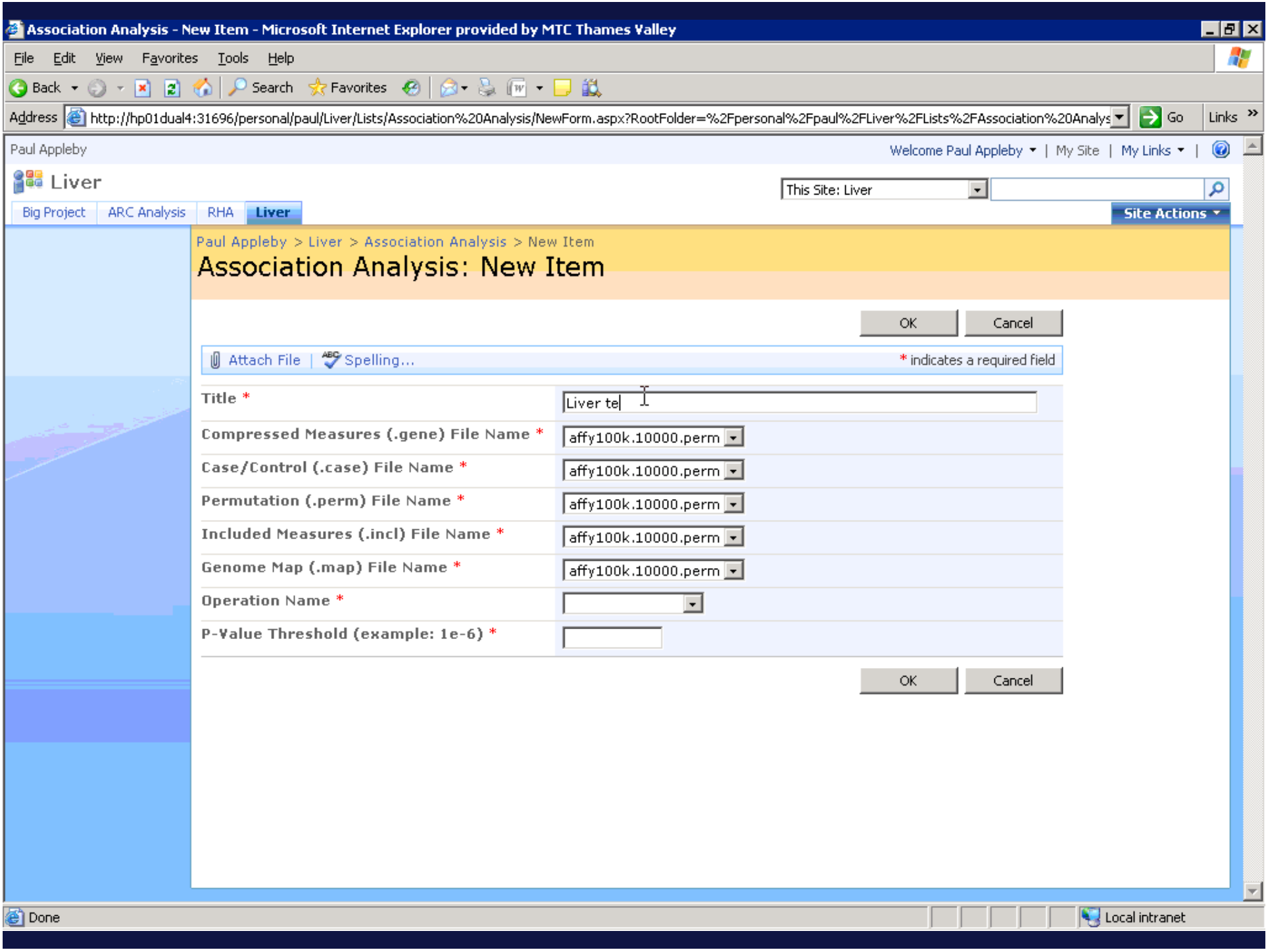


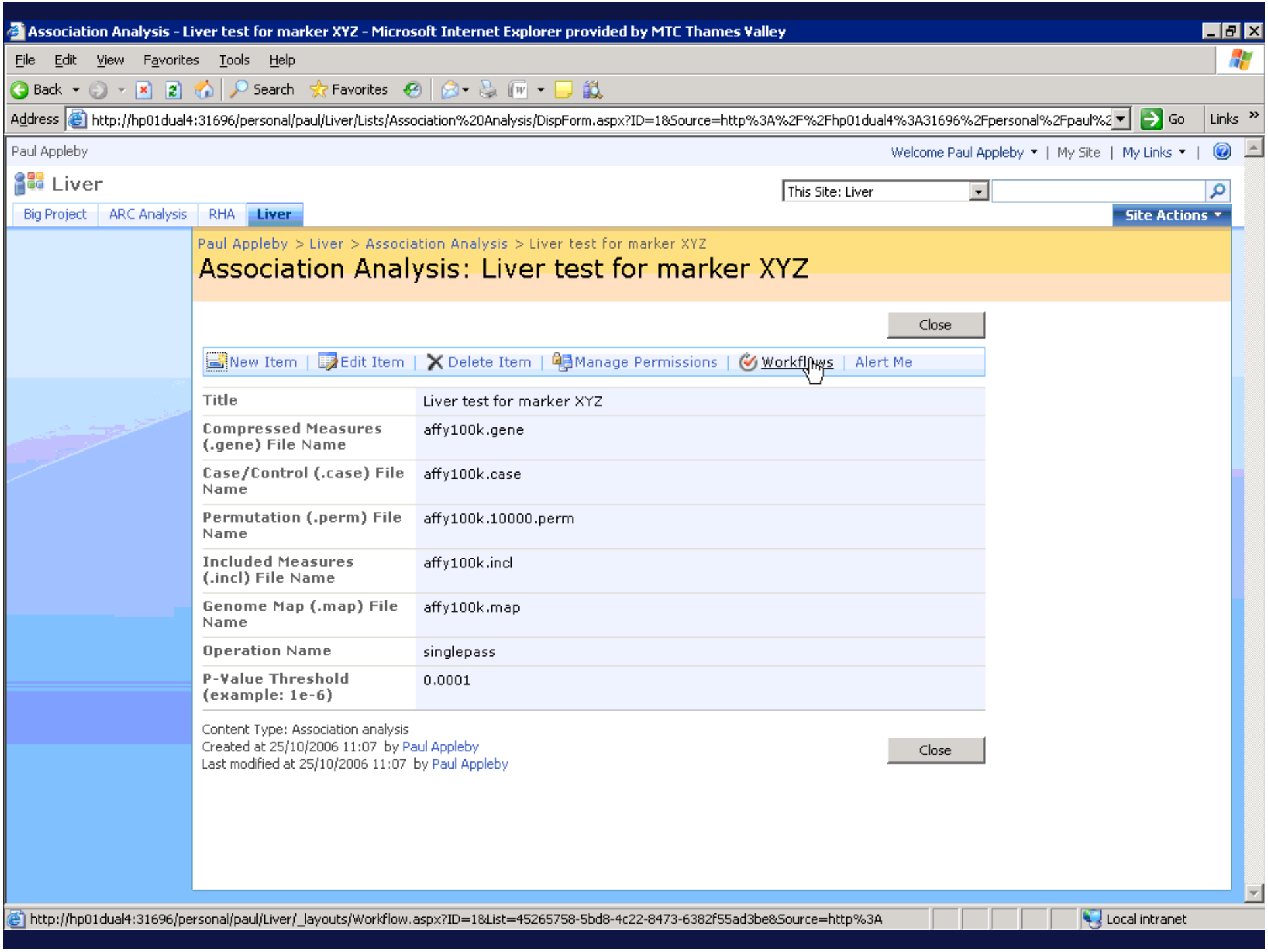
Workflow Tracing

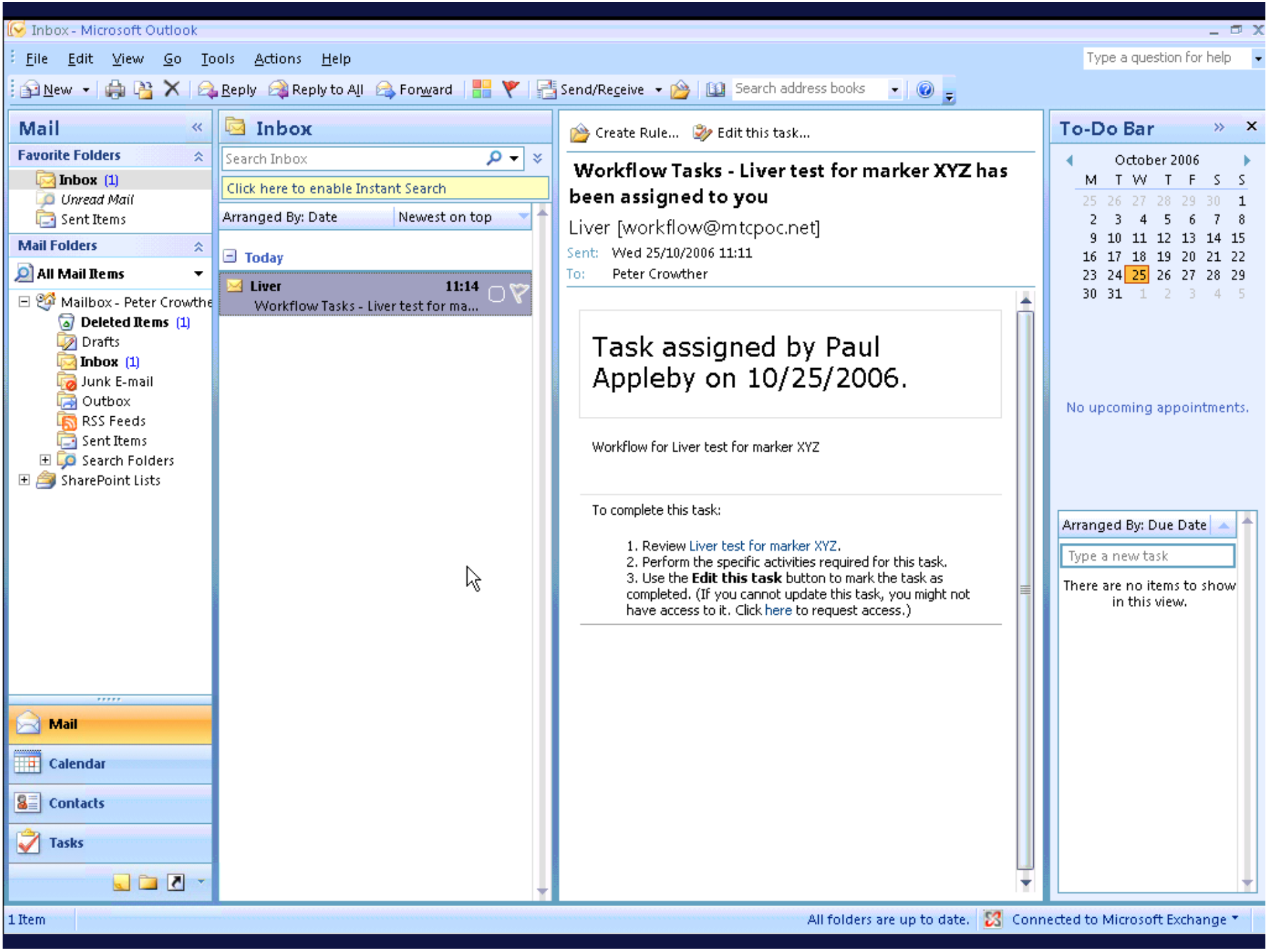


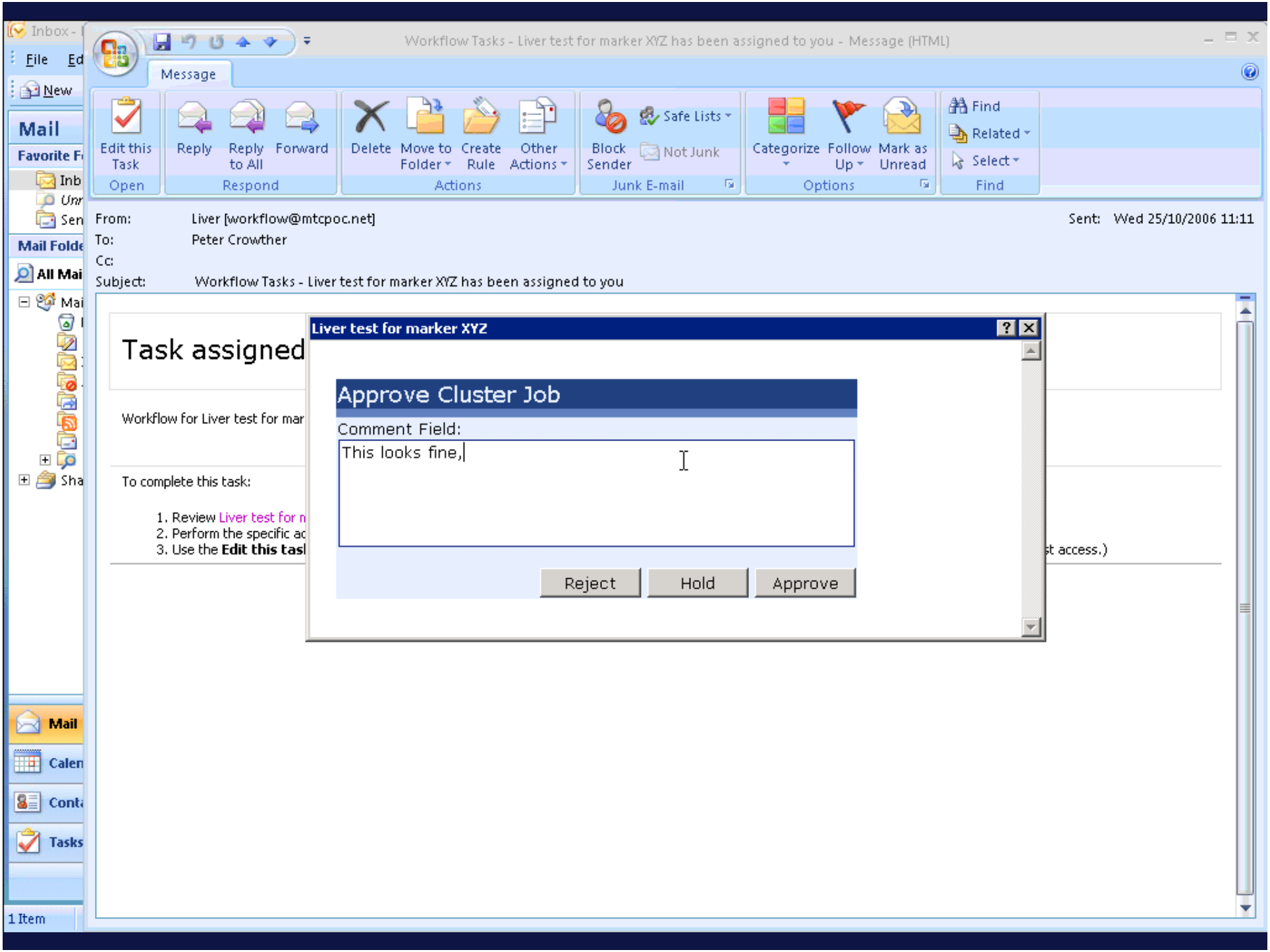

Home - Liver - Microsoft Internet Explorer provided by MTC Thames Valley
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hp01dual4:31696/personal/paul/Liver/default.aspx
Go Links
Paul Appleby
Welcome Paul Appleby | My Site | My Links |
Liver
This Site: Liver
Big Project ARC Analysis RHA Liver
Site Actions
View All Site Content
Results Viewer
Documents
Shared Documents
Association Wiki
Lists
Tasks
Association Analysis
Workflow Tasks
Discussions
Team Discussion
Sites
People and Groups
Recycle Bin
Paul Appleby > Liver
Welcome to the site. From here you can:
upload new data;
find what data has been uploaded;
preprocess data in preparation for association analyses;
run association analysis jobs;
monitor the state of jobs you are running;
view the results of current and past analyses.
RSS Viewer
Genetics current issue
[Perspectives] Learning the Common Language of Genetics
[Review] Sometimes the Result Is Not the Answer: The Truths and the Lies That Come From Using the Complementation Test
[Genome integrity and transmission] Discovery and Mapping of Wheat Ph1 Suppressors
[Genome integrity and transmission] A Sequence-Tagged Linkage Map of Brassica rapa
[Genome integrity and transmission] Different Mating-Type-Regulated Genes Affect the DNA Repair Defects of Saccharomyces RAD51, RAD52 and RAD55 Mutants
Links
Ensembl
NIBHI
Add new link
Tasks
Title
Assigned To
Prepare Gene & Case File ! NEW
Prepare permutation file ! NEW
Upload Data Files ! NEW
Add new item
Page maintainer
Peter Crowther
http://hp01dual4:31696/personal/paul/Liver/internals/upload.aspx
Local intranet

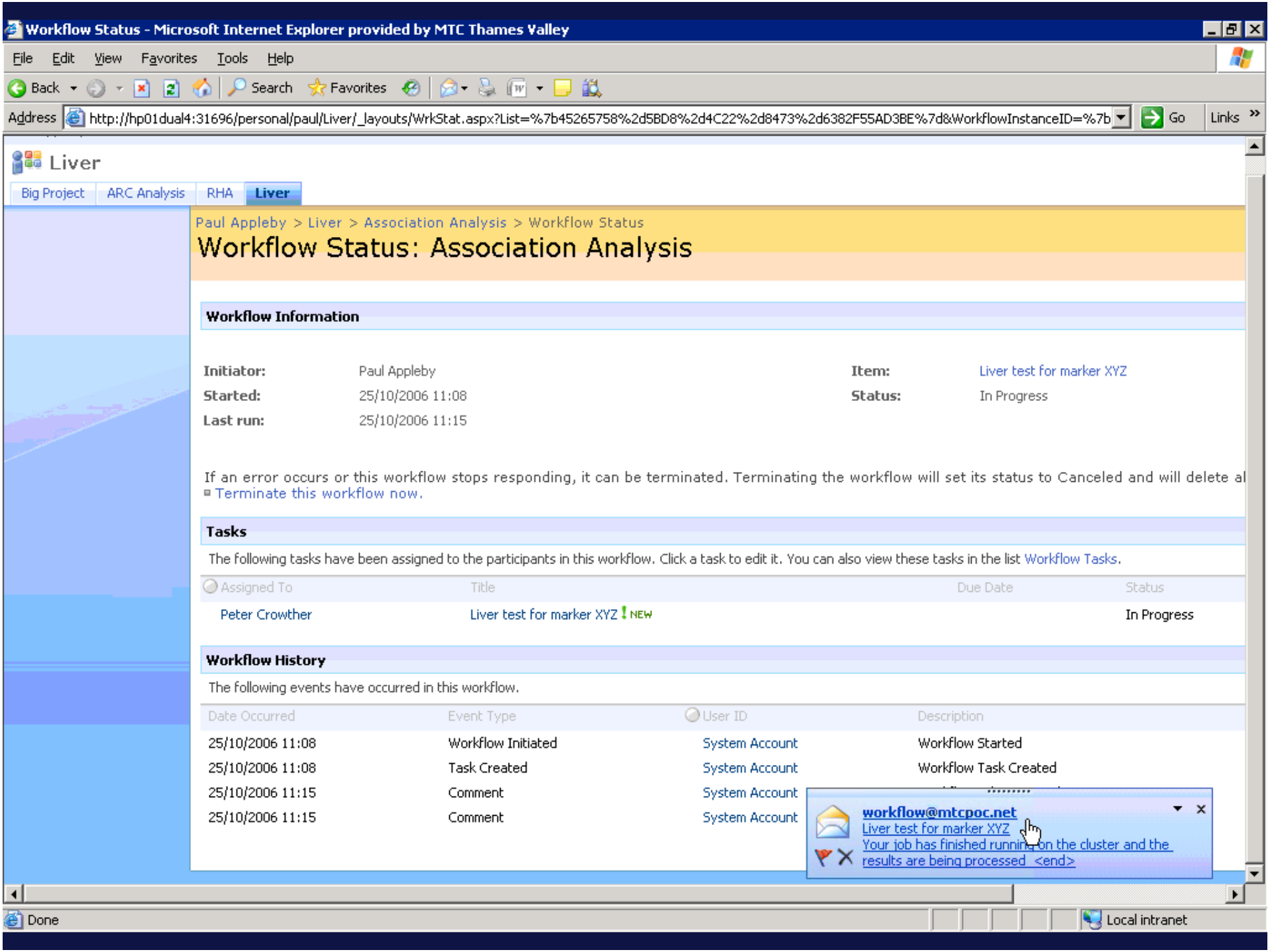

Association Analysis - Microsoft Internet Explorer provided by MTC Thames Valley
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hp01dual4:31696/personal/paul/Liver/Lists/Association%20Analysis/AllItems.aspx
Go Links
Paul Appleby
Welcome Paul Appleby | My Site | My Links |
Liver
This List: Association Analysis
Big Project ARC Analysis RHA Liver
Site Actions
Paul Appleby > Liver > Association Analysis
Association Analysis
View All Site Content
Results Viewer
Documents
Shared Documents
Association Wiki
Lists
Tasks
Association Analysis
Workflow Tasks
Discussions
Team Discussion
Sites
People and Groups
Recycle Bin
Use this list to manage you association analysis jobs
New Actions Settings
View: All Items
Association analysis
A specification of a regression test of measures against cases/controls
iation Analysis
iation Analysis" list. To create a new item, click "New" above.
Done
Local intranet

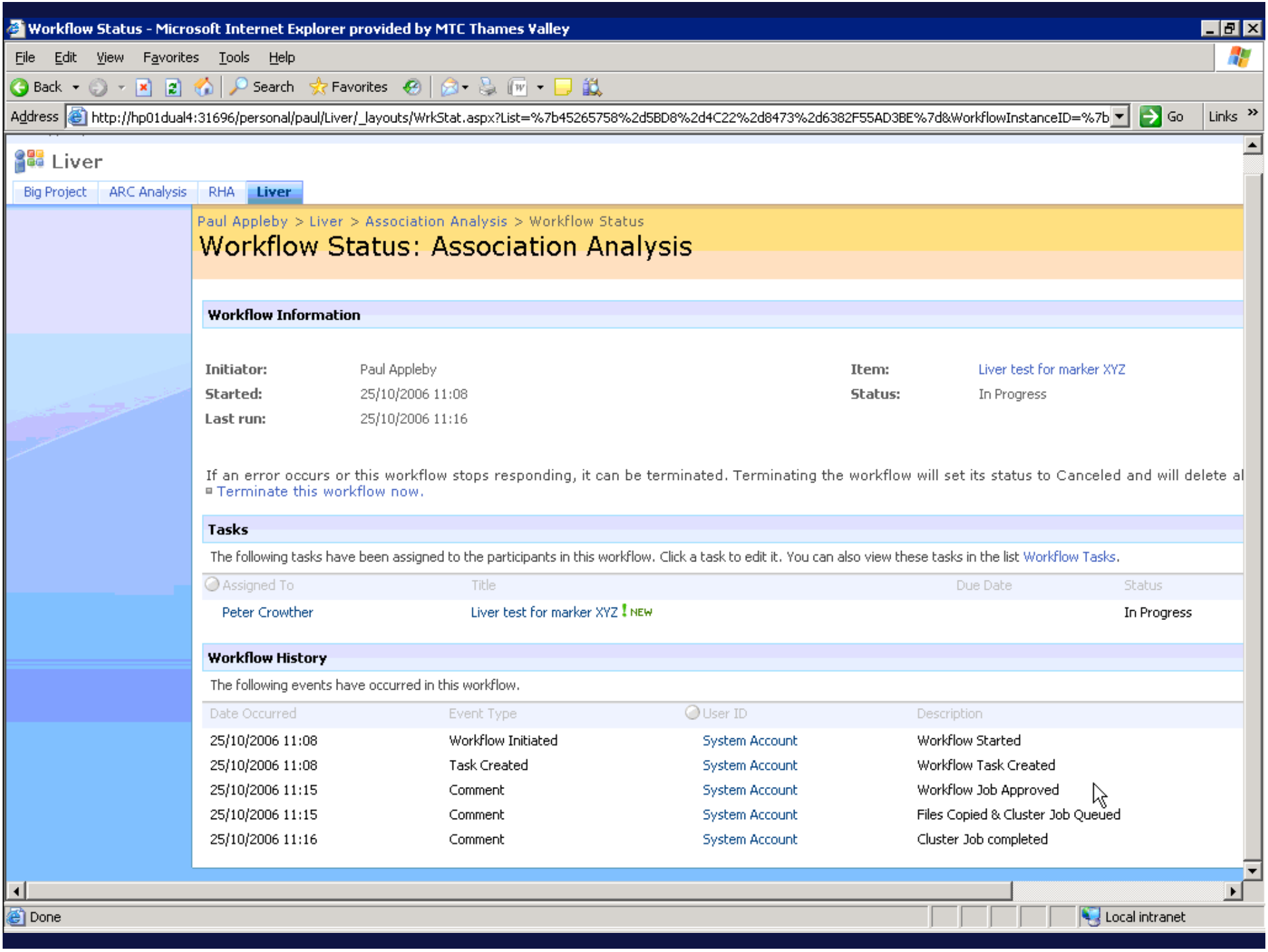

Association Analysis - New Item - Microsoft Internet Explorer provided by MTC Thames Valley
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hp01dual4:31696/personal/paul/Liver/Lists/Association%20Analysis/NewForm.aspx?RootFolder=%2Fpersonal%2Fpaul%2FLiver%2FLists%2FAssociation%20Analys
Go Links
Paul Appleby
Welcome Paul Appleby | My Site | My Links |
Liver
This Site: Liver
Big Project ARC Analysis RHA Liver
Site Actions
Paul Appleby > Liver > Association Analysis > New Item
Association Analysis: New Item
OK Cancel
Attach File | Spelling...
* indicates a required field
Title *
Liver te
Compressed Measures (.gene) File Name *
affy100k.10000.perm
Case/Control (.case) File Name *
affy100k.10000.perm
Permutation (.perm) File Name *
affy100k.10000.perm
Included Measures (.incl) File Name *
affy100k.10000.perm
Genome Map (.map) File Name *
affy100k.10000.perm
Operation Name *
P-Value Threshold (example: 1e-6) *
OK Cancel
Done
Local intranet

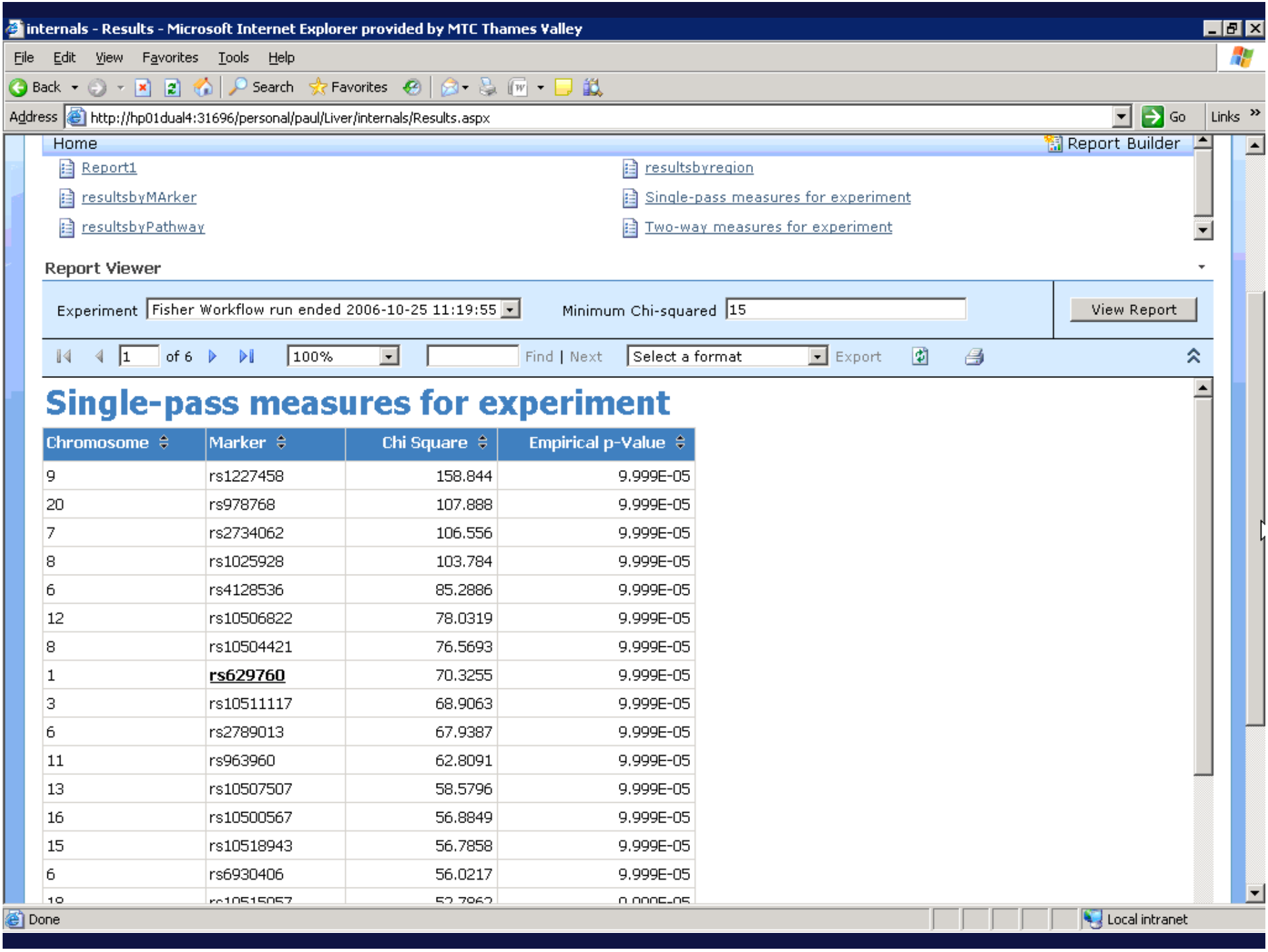

Association Analysis - Liver test for marker XYZ - Microsoft Internet Explorer provided by MTC Thames Valley
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hp01dual4:31696/personal/paul/Liver/Lists/Association%20Analysis/DispForm.aspx?ID=1&Source=http%3A%2F%2Fhp01dual4%3A31696%2Fpersonal%2Fpaul%2
Go Links
Paul Appleby
Welcome Paul Appleby | My Site | My Links |
Liver
This Site: Liver
Big Project ARC Analysis RHA Liver
Site Actions
Paul Appleby > Liver > Association Analysis > Liver test for marker XYZ
Association Analysis: Liver test for marker XYZ
Close
New Item | Edit Item | Delete Item | Manage Permissions | Workflows | Alert Me
Title
Liver test for marker XYZ
Compressed Measures (.gene) File Name
affy100k.gene
Case/Control (.case) File Name
affy100k.case
Permutation (.perm) File Name
affy100k.10000.perm
Included Measures (.incl) File Name
affy100k.incl
Genome Map (.map) File Name
affy100k.map
Operation Name
singlepass
P-Value Threshold (example: 1e-6)
0.0001
Content Type: Association analysis
Created at 25/10/2006 11:07 by Paul Appleby
Last modified at 25/10/2006 11:07 by Paul Appleby
Close
http://hp01dual4:31696/personal/paul/Liver/_layouts/Workflow.aspx?ID=1&List=45265758-5bd8-4c22-8473-6382f55ad3be&Source=http%3A
Local intranet


Inbox - Microsoft Outlook
File Edit View Go Tools Actions Help
Type a question for help
New Reply Reply to All Forward Send/Receive Search address books
Mail
Favorite Folders
Inbox (1)
Unread Mail
Sent Items
Mail Folders
All Mail Items
Mailbox - Peter Crowthe
Deleted Items (1)
Drafts
Inbox (1)
Junk E-mail
Outbox
RSS Feeds
Sent Items
Search Folders
SharePoint Lists
Mail
Calendar
Contacts
Tasks
Inbox
Search Inbox
Click here to enable Instant Search
Arranged By: Date
Newest on top
Today
Liver 11:14
Workflow Tasks - Liver test for ma...
Create Rule...
Edit this task...
Workflow Tasks - Liver test for marker XYZ has been assigned to you
Liver [workflow@mtcpoc.net]
Sent: Wed 25/10/2006 11:11
To: Peter Crowther
Task assigned by Paul Appleby on 10/25/2006.
Workflow for Liver test for marker XYZ
To complete this task:
1. Review Liver test for marker XYZ.
2. Perform the specific activities required for this task.
3. Use the Edit this task button to mark the task as completed. (If you cannot update this task, you might not have access to it. Click here to request access.)
To-Do Bar
October 2006
M T W T F S S
25 26 27 28 29 30 1
2 3 4 5 6 7 8
9 10 11 12 13 14 15
16 17 18 19 20 21 22
23 24 25 26 27 28 29
30 31 1 2 3 4 5
No upcoming appointments.
Arranged By: Due Date
Type a new task
There are no items to show in this view.
1 Item
All folders are up to date.
Connected to Microsoft Exchange


Workflow Tasks - Liver test for marker XYZ has been assigned to you - Message (HTML)
Message
Edit this Task
Reply
Reply to All
Forward
Delete
Move to Folder
Create Rule
Other Actions
Block Sender
Safe Lists
Not Junk
Categorize
Follow Up
Mark as Unread
Find
Related
Select
Open
Respond
Actions
Junk E-mail
Options
Find
From: Liver [workflow@mtcpoc.net]
Sent: Wed 25/10/2006 11:11
To: Peter Crowther
Cc:
Subject: Workflow Tasks - Liver test for marker XYZ has been assigned to you
Task assigned
Workflow for Liver test for mar
To complete this task:
1. Review Liver test for n
2. Perform the specific ac
3. Use the Edit this tas
Liver test for marker XYZ
Approve Cluster Job
Comment Field:
This looks fine,
Reject
Hold
Approve
Mail
Calen
Conta
Tasks
1 Item


Workflow Status - Microsoft Internet Explorer provided by MTC Thames Valley
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hp01dual4:31696/personal/paul/Liver/_layouts/WrkStat.aspx?List=%7b45265758%2d5BD8%2d4C22%2d8473%2d6382F55AD3BE%7d&WorkflowInstanceID=%7b
Go Links
Liver
Big Project ARC Analysis RHA Liver
Paul Appleby > Liver > Association Analysis > Workflow Status
Workflow Status: Association Analysis
Workflow Information
Initiator: Paul Appleby
Started: 25/10/2006 11:08
Last run: 25/10/2006 11:15
Item: Liver test for marker XYZ
Status: In Progress
If an error occurs or this workflow stops responding, it can be terminated. Terminating the workflow will set its status to Canceled and will delete al
Terminate this workflow now.
Tasks
The following tasks have been assigned to the participants in this workflow. Click a task to edit it. You can also view these tasks in the list Workflow Tasks.
Assigned To Title Due Date Status
Peter Crowther Liver test for marker XYZ NEW In Progress
Workflow History
The following events have occurred in this workflow.
Date Occurred Event Type User ID Description
25/10/2006 11:08 Workflow Initiated System Account Workflow Started
25/10/2006 11:08 Task Created System Account Workflow Task Created
25/10/2006 11:15 Comment System Account
25/10/2006 11:15 Comment System Account
workflow@mtcpoc.net
Liver test for marker XYZ
Your job has finished running on the cluster and the results are being processed <end>
Done
Local intranet


Workflow Status - Microsoft Internet Explorer provided by MTC Thames Valley
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hp01dual4:31696/personal/paul/Liver/_layouts/WrkStat.aspx?List=%7b45265758%2d5BD8%2d4C22%2d8473%2d6382F55AD3BE%7d&WorkflowInstanceID=%7b
Go Links
Liver
Big Project ARC Analysis RHA Liver
Paul Appleby > Liver > Association Analysis > Workflow Status
Workflow Status: Association Analysis
Workflow Information
Initiator: Paul Appleby
Started: 25/10/2006 11:08
Last run: 25/10/2006 11:16
Item: Liver test for marker XYZ
Status: In Progress
If an error occurs or this workflow stops responding, it can be terminated. Terminating the workflow will set its status to Canceled and will delete al
Terminate this workflow now.
Tasks
The following tasks have been assigned to the participants in this workflow. Click a task to edit it. You can also view these tasks in the list Workflow Tasks.
Assigned To Title Due Date Status
Peter Crowther Liver test for marker XYZ ! NEW In Progress
Workflow History
The following events have occurred in this workflow.
Date Occurred Event Type User ID Description
25/10/2006 11:08 Workflow Initiated System Account Workflow Started
25/10/2006 11:08 Task Created System Account Workflow Task Created
25/10/2006 11:15 Comment System Account Workflow Job Approved
25/10/2006 11:15 Comment System Account Files Copied & Cluster Job Queued
25/10/2006 11:16 Comment System Account Cluster Job completed
Done
Local intranet


internals - Results - Microsoft Internet Explorer provided by MTC Thames Valley
File Edit View Favorites Tools Help
Back Search Favorites
Address http://hp01dual4:31696/personal/paul/Liver/internals/Results.aspx
Go Links
Home
Report Builder
Report1
resultsbyMArker
resultsbyPathway
resultsbyregion
Single-pass measures for experiment
Two-way measures for experiment
Report Viewer
Experiment Fisher Workflow run ended 2006-10-25 11:19:55
Minimum Chi-squared 15
View Report
1 of 6
100%
Find | Next
Select a format
Export
Single-pass measures for experiment
Chromosome Marker Chi Square Empirical p-Value
9 rs1227458 158.844 9.999E-05
20 rs978768 107.888 9.999E-05
7 rs2734062 106.556 9.999E-05
8 rs1025928 103.784 9.999E-05
6 rs4128536 85.2886 9.999E-05
12 rs10506822 78.0319 9.999E-05
8 rs10504421 76.5693 9.999E-05
1 rs629760 70.3255 9.999E-05
3 rs10511117 68.9063 9.999E-05
6 rs2789013 67.9387 9.999E-05
11 rs963960 62.8091 9.999E-05
13 rs10507507 58.5796 9.999E-05
16 rs10500567 56.8849 9.999E-05
15 rs10518943 56.7858 9.999E-05
6 rs6930406 56.0217 9.999E-05
Done
Local intranet

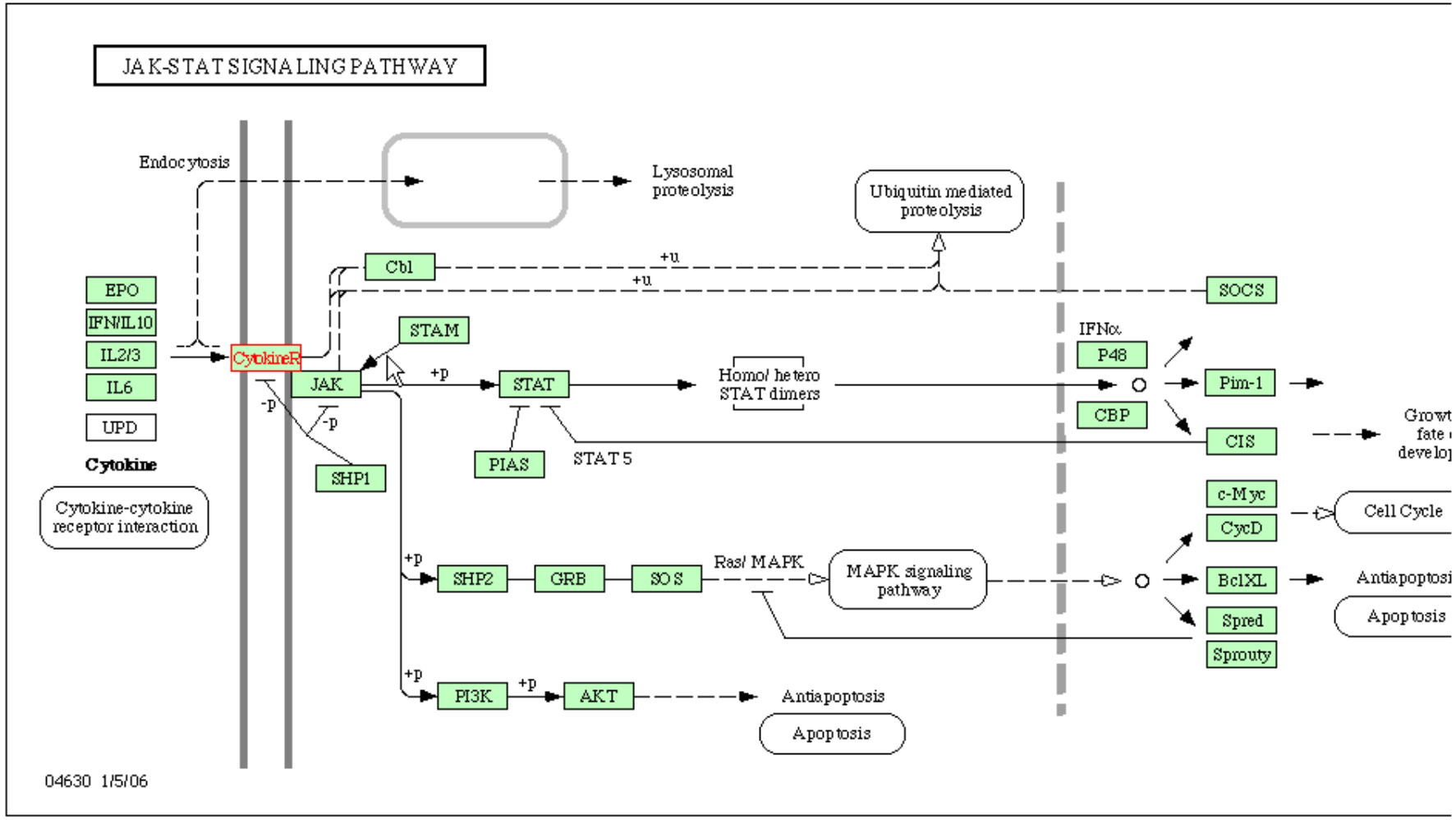
internals - Results - Microsoft Internet Explorer provided by MTC Thames Valley

File Edit View Favorites Tools Help

Back Search Favorites

Address http://hp01dual4:31696/personal/paul/Liver/internals/Results.aspx Go Links

Homo sapiens (human) Go Current selection Select



Done Local intranet

HPC++ Scenario

HPC - Windows Internet Explorer

http://labs.microsofthpc.net/

hpc labs

HPC

Page Tools

Microsoft

Microsoft HPC++ Labs

Microsoft HPC++ CompFin Lab

The Microsoft HPC++ CompFin Lab integrates Microsoft Windows HPC Server, a central market data database and Microsoft productivity products to provide university courses with an online service to publish, execute and manage computational finance models.

Lab Details:

Computing Resources:	■ 256 core/64 node compute cluster ■ 5 TB disk space ■ Low latency interconnect
Rich User Interface:	■ SharePoint 2007 portal to publish, browse and monitor models ■ Excel 2007 as the user interface for models and results ■ Model execution workflow with status email notifications
Central Market Database:	■ 5 year tick data for S&P500 ■ Daily & fundamental data for 10,000 stocks ■ Mortgage Backed Securities pool data
Simplified Modeling:	■ .NET platform with Linq for SQL data access ■ SOA framework for parameter sweep models ■ Local model development and test environment

Who Can Participate?

Microsoft HPC++ Labs

Microsoft HPC++ Labs is an incubation effort within the Microsoft Windows HPC Server product team whose goals include:

- Demonstrate end-to-end integrated HPC solutions using Microsoft Windows HPC Server 2008 with Microsoft productivity and developer products
- Develop expertise in operating end user focused compute and data intensive HPC services
- Incubate Microsoft product extensions required to support HPC++ scenarios "out of the box" on customer premises

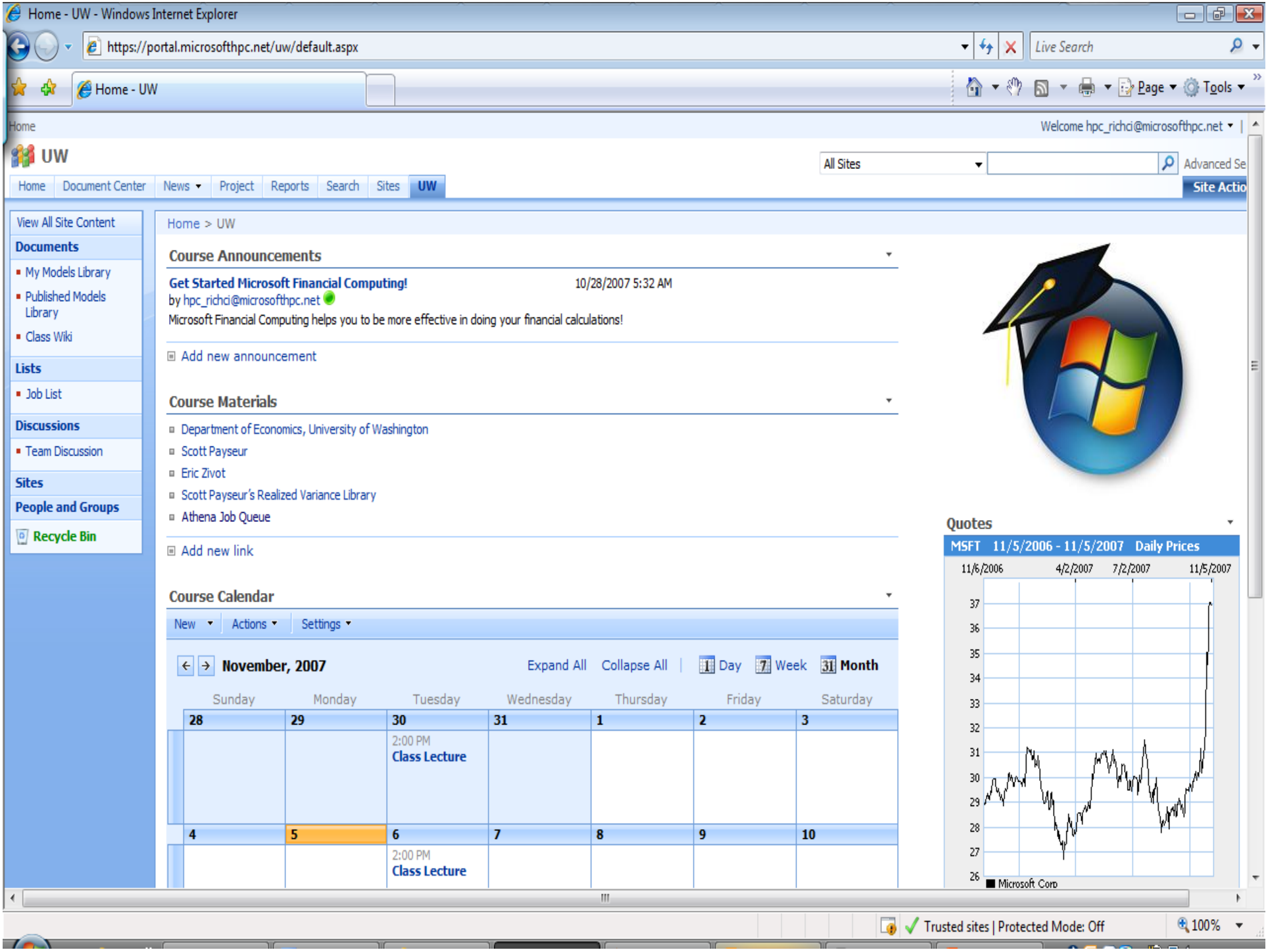
HPC++ Labs is Hiring!

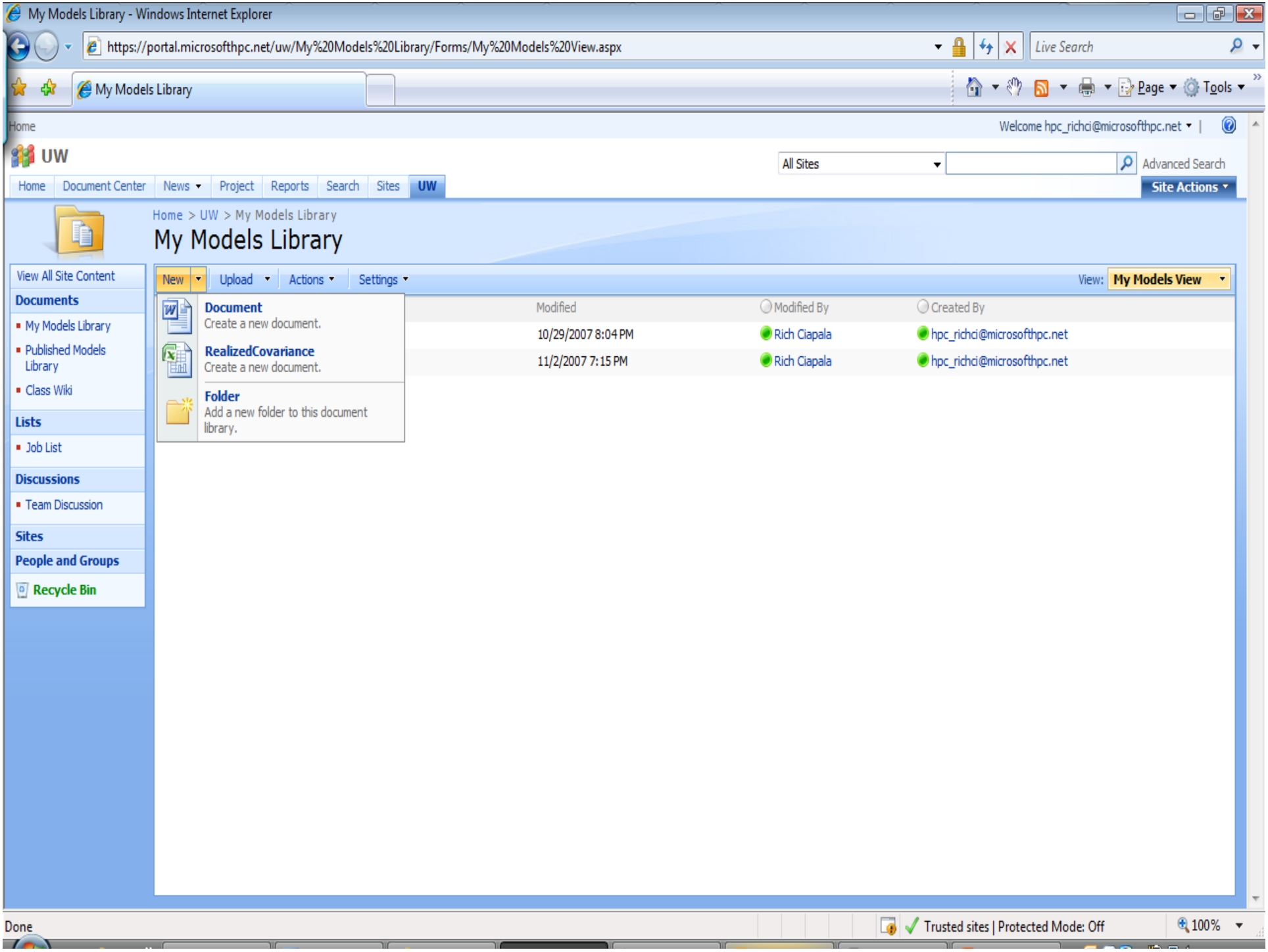
Next Steps

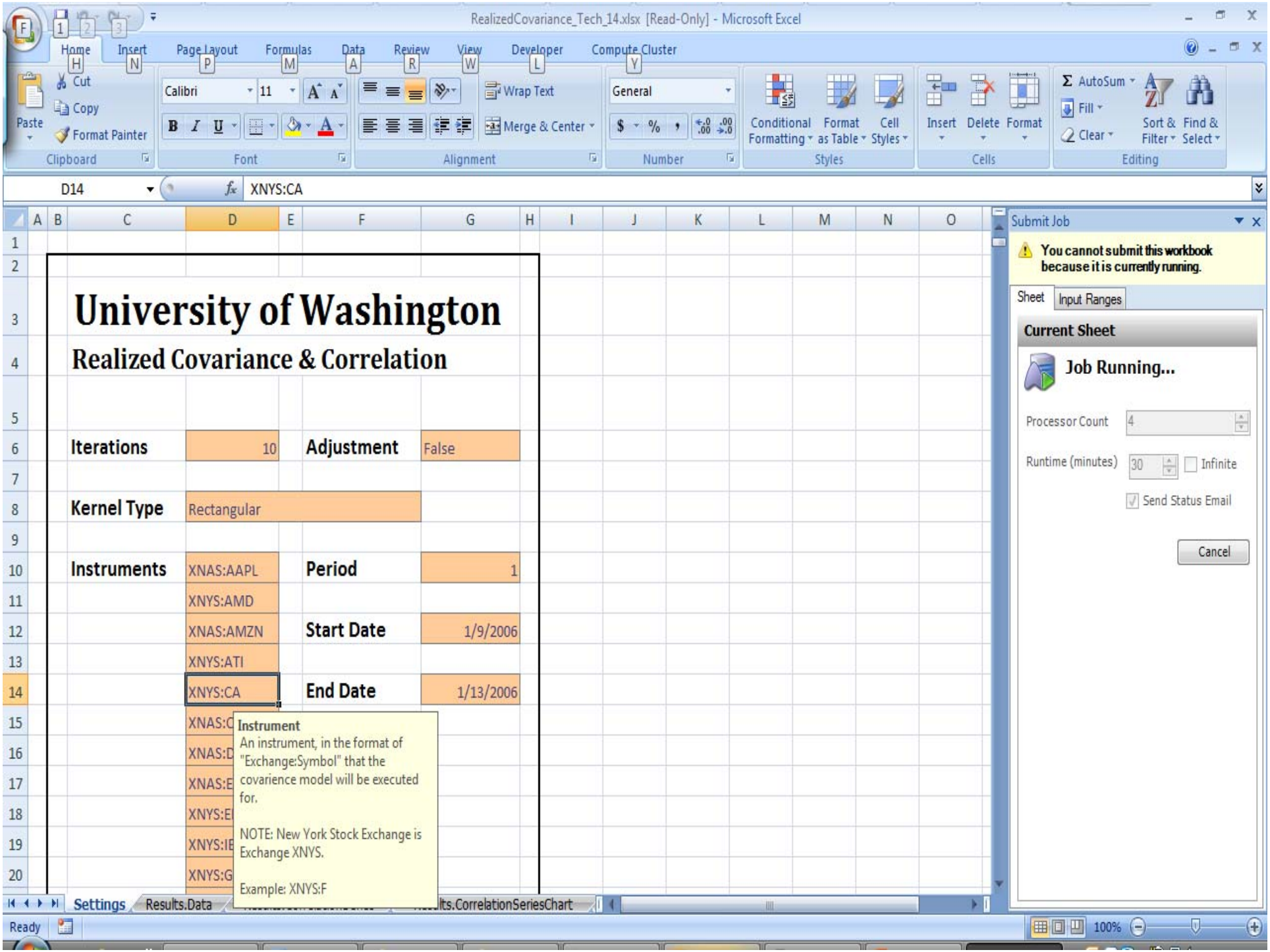
Request Enrollment

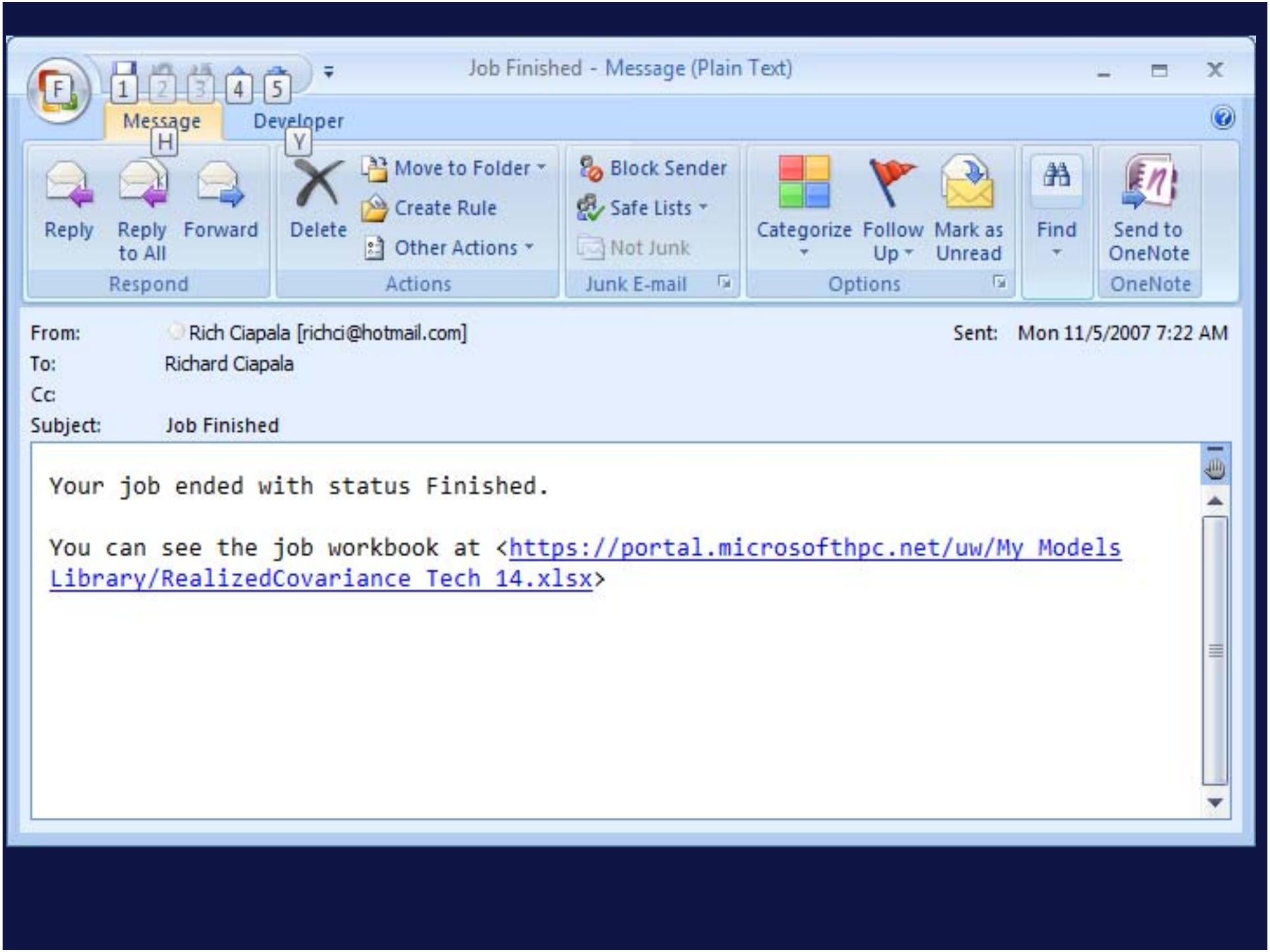
Watch Demonstration

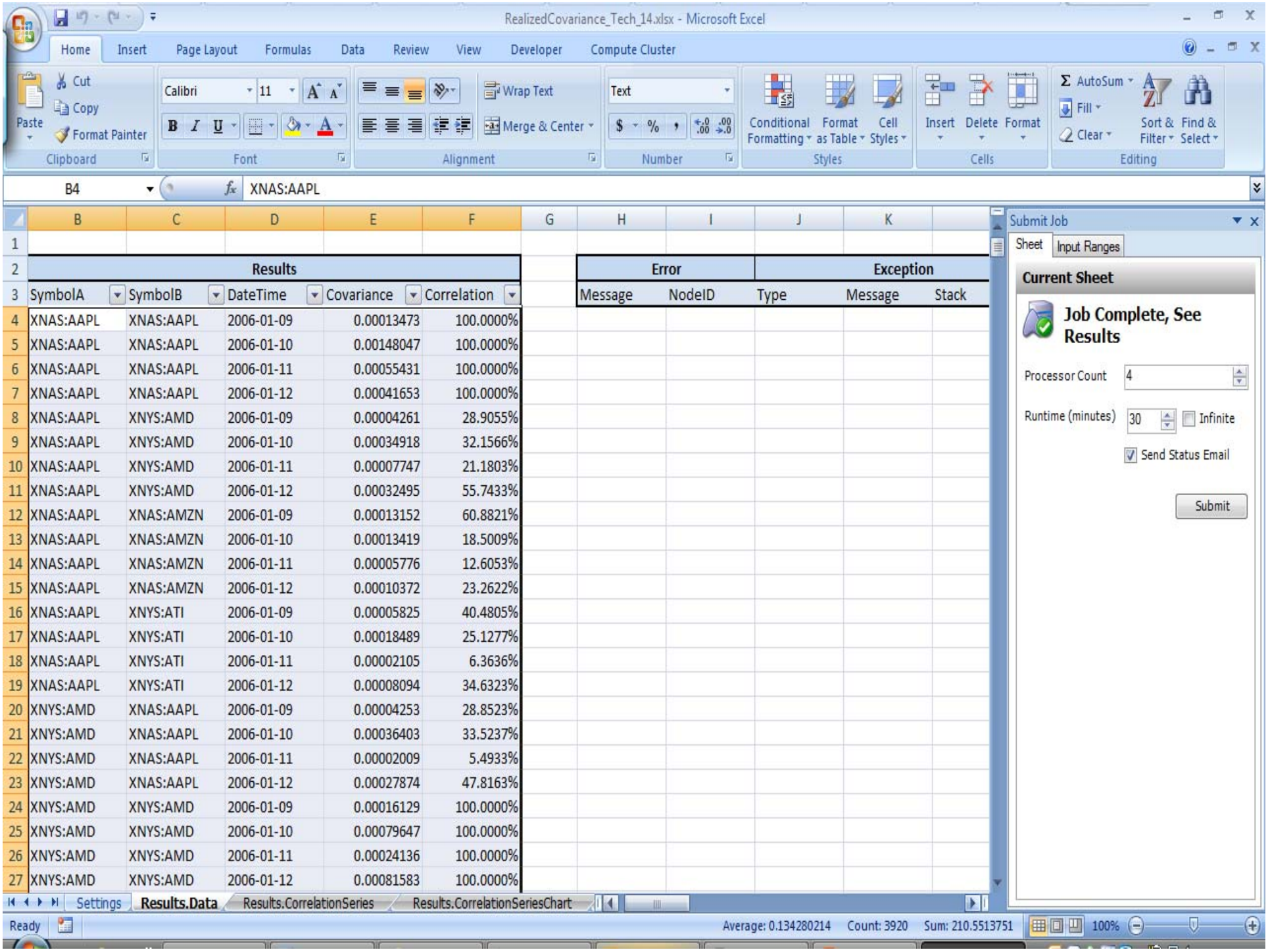
Read Our Blog

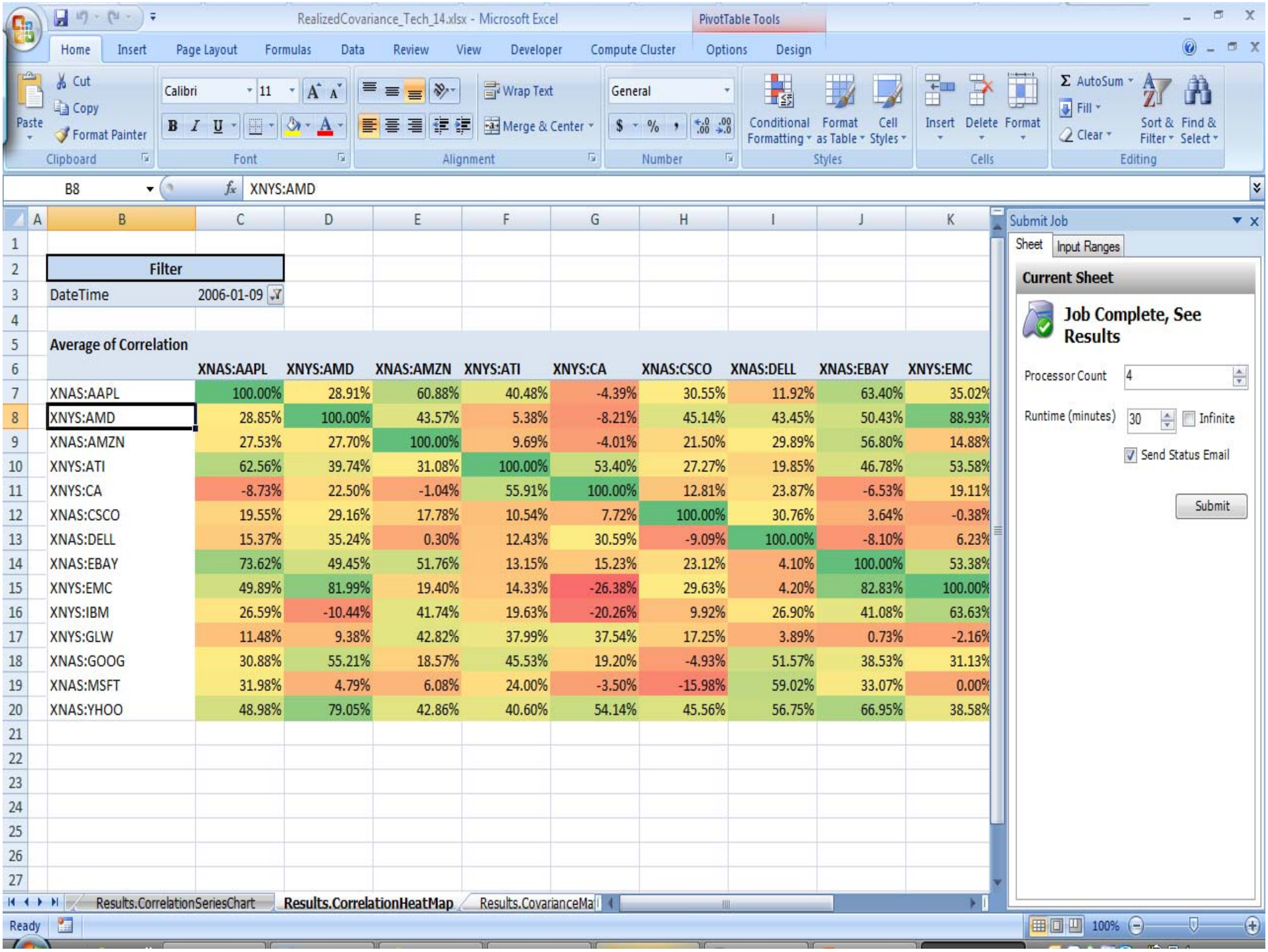

Home - UW - Windows Internet Explorer
https://portal.microsofthpc.net/uw/default.aspx
Live Search
Home - UW
Page
Tools
Home
Welcome hpc_richci@microsofthpc.net
UW
All Sites
Advanced Se
Home
Document Center
News
Project
Reports
Search
Sites
UW
Site Actio
View All Site Content
Documents
My Models Library
Published Models Library
Class Wiki
Lists
Job List
Discussions
Team Discussion
Sites
People and Groups
Recycle Bin
Home > UW
Course Announcements
Get Started Microsoft Financial Computing!
10/28/2007 5:32 AM
by hpc_richci@microsofthpc.net
Microsoft Financial Computing helps you to be more effective in doing your financial calculations!
Add new announcement
Course Materials
Department of Economics, University of Washington
Scott Payseur
Eric Zivot
Scott Payseur's Realized Variance Library
Athena Job Queue
Add new link
Course Calendar
New
Actions
Settings
November, 2007
Expand All
Collapse All
Day
Week
Month
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
28
29
30
31
1
2
3
2:00 PM
Class Lecture
4
5
6
7
8
9
10
2:00 PM
Class Lecture
Quotes
MSFT 11/5/2006 - 11/5/2007 Daily Prices
11/6/2006
4/2/2007
7/2/2007
11/5/2007
37
36
35
34
33
32
31
30
29
28
27
26
Microsoft Corp
Trusted sites | Protected Mode: Off
100%

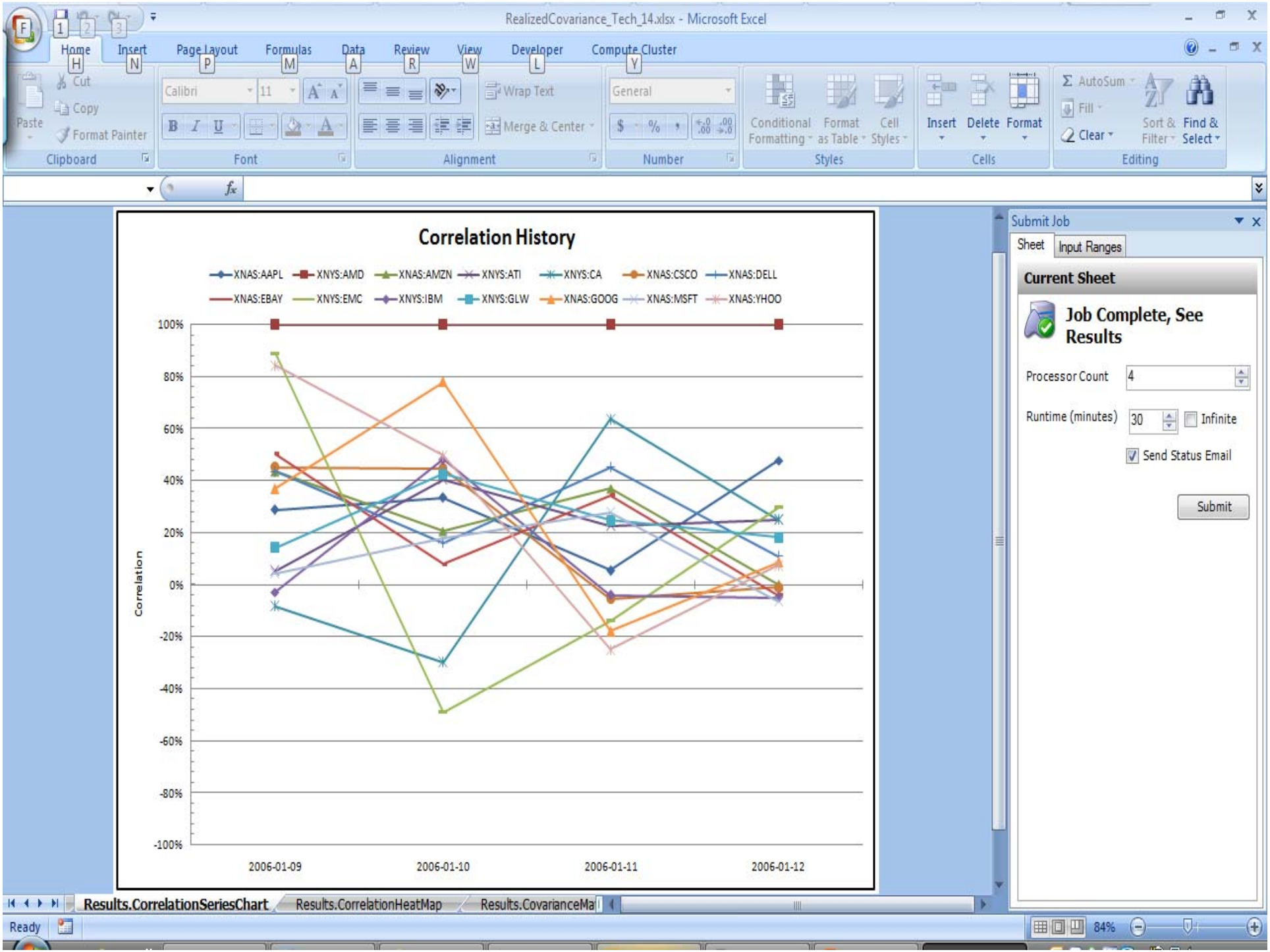

My Models Library - Windows Internet Explorer
https://portal.microsofthpc.net/uw/My%20Models%20Library/Forms/My%20Models%20View.aspx
Live Search
My Models Library
Page
Tools
Home
Welcome hpc_richci@microsofthpc.net
UW
All Sites
Advanced Search
Home
Document Center
News
Project
Reports
Search
Sites
UW
Site Actions
Home > UW > My Models Library
My Models Library
View All Site Content
Documents
My Models Library
Published Models Library
Class Wiki
Lists
Job List
Discussions
Team Discussion
Sites
People and Groups
Recycle Bin
New
Upload
Actions
Settings
View:
My Models View
Document
Create a new document.
RealizedCovariance
Create a new document.
Folder
Add a new folder to this document library.
Modified
Modified By
Created By
10/29/2007 8:04 PM
Rich Ciapala
hpc_richci@microsofthpc.net
11/2/2007 7:15 PM
Rich Ciapala
hpc_richci@microsofthpc.net
Done
Trusted sites | Protected Mode: Off
100%

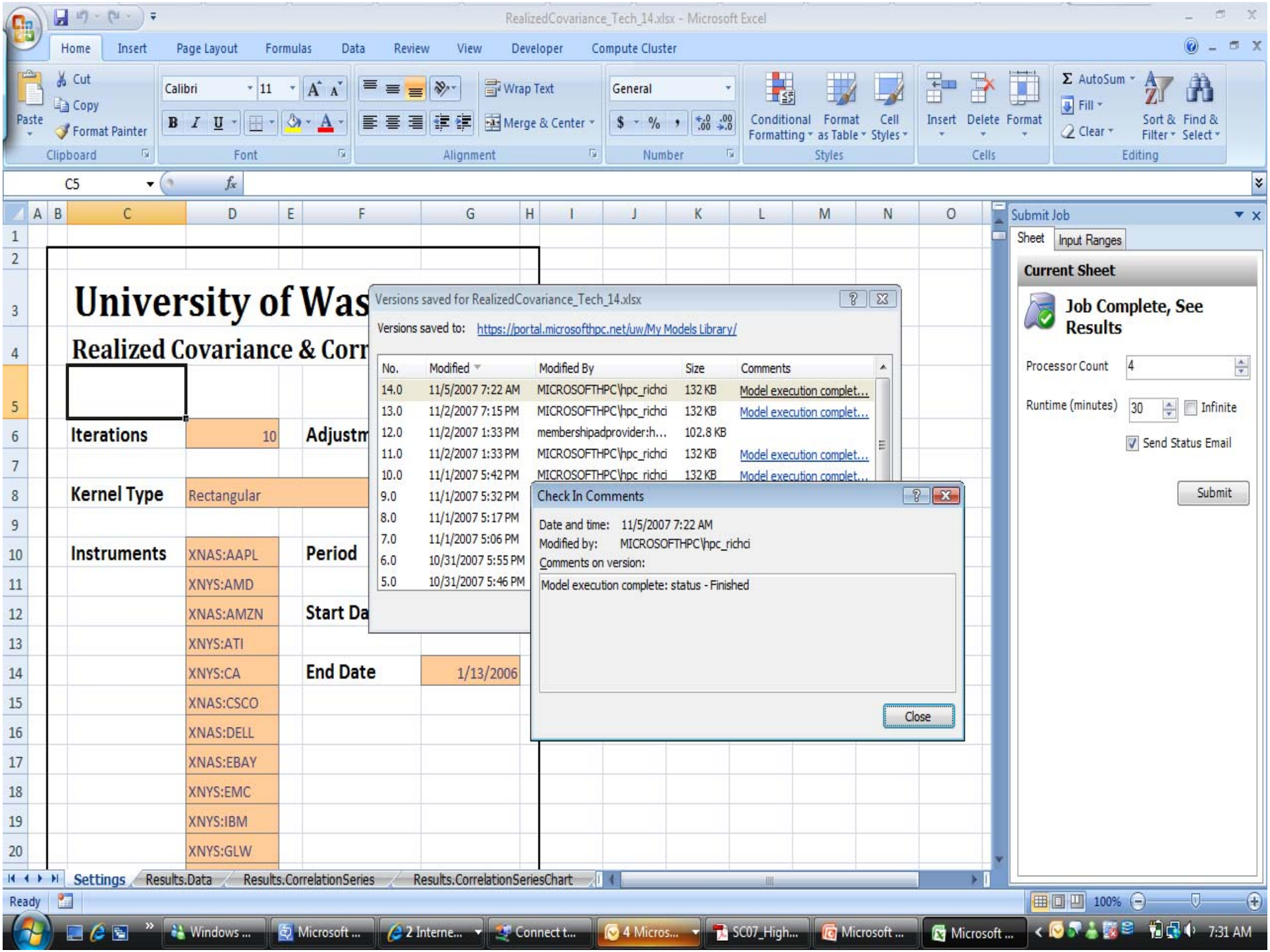

RealizedCovariance_Tech_14.xlsx [Read-Only] - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Compute Cluster
Cut
Copy
Format Painter
Paste
Calibri
11
Wrap Text
Merge & Center
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
D14
XNYS:CA
University of Washington
Realized Covariance & Correlation
Iterations
10
Adjustment
False
Kernel Type
Rectangular
Instruments
XNAS:AAPL
XNYS:AMD
XNAS:AMZN
XNYS:ATI
XNYS:CA
Period
1
Start Date
1/9/2006
End Date
1/13/2006
Instrument
An instrument, in the format of "Exchange:Symbol" that the covarience model will be executed for.
NOTE: New York Stock Exchange is Exchange XNYS.
Example: XNYS:F
Settings
Results.Data
Submit Job
You cannot submit this workbook because it is currently running.
Sheet
Input Ranges
Current Sheet
Job Running...
Processor Count
4
Runtime (minutes)
30
Infinite
Send Status Email
Cancel
Ready
100%


Job Finished - Message (Plain Text)
Message
Developer
Reply
Reply to All
Forward
Delete
Move to Folder
Create Rule
Other Actions
Block Sender
Safe Lists
Not Junk
Categorize
Follow Up
Mark as Unread
Find
Send to OneNote
Respond
Actions
Junk E-mail
Options
OneNote
From: Rich Ciapala [richci@hotmail.com]
Sent: Mon 11/5/2007 7:22 AM
To: Richard Ciapala
Cc:
Subject: Job Finished
Your job ended with status Finished.
You can see the job workbook at <https://portal.microsofthpc.net/uw/My Models Library/RealizedCovariance Tech 14.xlsx>


RealizedCovariance_Tech_14.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Compute Cluster
Paste
Cut
Copy
Format Painter
Clipboard
Calibri
11
Font
Wrap Text
Merge & Center
Alignment
Text
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Editing
B4
XNAS:AAPL
Results
Error
Exception
SymbolA
SymbolB
DateTime
Covariance
Correlation
Message
NodeID
Type
Message
Stack
XNAS:AAPL XNAS:AAPL 2006-01-09 0.00013473 100.0000%
XNAS:AAPL XNAS:AAPL 2006-01-10 0.00148047 100.0000%
XNAS:AAPL XNAS:AAPL 2006-01-11 0.00055431 100.0000%
XNAS:AAPL XNAS:AAPL 2006-01-12 0.00041653 100.0000%
XNAS:AAPL XNYS:AMD 2006-01-09 0.00004261 28.9055%
XNAS:AAPL XNYS:AMD 2006-01-10 0.00034918 32.1566%
XNAS:AAPL XNYS:AMD 2006-01-11 0.00007747 21.1803%
XNAS:AAPL XNYS:AMD 2006-01-12 0.00032495 55.7433%
XNAS:AAPL XNAS:AMZN 2006-01-09 0.00013152 60.8821%
XNAS:AAPL XNAS:AMZN 2006-01-10 0.00013419 18.5009%
XNAS:AAPL XNAS:AMZN 2006-01-11 0.00005776 12.6053%
XNAS:AAPL XNAS:AMZN 2006-01-12 0.00010372 23.2622%
XNAS:AAPL XNYS:ATI 2006-01-09 0.00005825 40.4805%
XNAS:AAPL XNYS:ATI 2006-01-10 0.00018489 25.1277%
XNAS:AAPL XNYS:ATI 2006-01-11 0.00002105 6.3636%
XNAS:AAPL XNYS:ATI 2006-01-12 0.00008094 34.6323%
XNYS:AMD XNAS:AAPL 2006-01-09 0.00004253 28.8523%
XNYS:AMD XNAS:AAPL 2006-01-10 0.00036403 33.5237%
XNYS:AMD XNAS:AAPL 2006-01-11 0.00002009 5.4933%
XNYS:AMD XNAS:AAPL 2006-01-12 0.00027874 47.8163%
XNYS:AMD XNYS:AMD 2006-01-09 0.00016129 100.0000%
XNYS:AMD XNYS:AMD 2006-01-10 0.00079647 100.0000%
XNYS:AMD XNYS:AMD 2006-01-11 0.00024136 100.0000%
XNYS:AMD XNYS:AMD 2006-01-12 0.00081583 100.0000%
Settings
Results.Data
Results.CorrelationSeries
Results.CorrelationSeriesChart
Submit Job
Sheet
Input Ranges
Current Sheet
Job Complete, See Results
Processor Count
4
Runtime (minutes)
30
Infinite
Send Status Email
Submit
Ready
Average: 0.134280214
Count: 3920
Sum: 210.5513751
100%


RealizedCovariance_Tech_14.xlsx - Microsoft Excel
PivotTable Tools
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Compute Cluster
Options
Design
Cut
Copy
Format Painter
Paste
Clipboard
Calibri
11
Font
Wrap Text
Merge & Center
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Editing
B8
XNYS:AMD
Filter
DateTime
2006-01-09
Average of Correlation
XNAS:AAPL
XNYS:AMD
XNAS:AMZN
XNYS:ATI
XNYS:CA
XNAS:CSCO
XNAS:DELL
XNAS:EBAY
XNYS:EMC
XNAS:AAPL 100.00% 28.91% 60.88% 40.48% -4.39% 30.55% 11.92% 63.40% 35.02%
XNYS:AMD 28.85% 100.00% 43.57% 5.38% -8.21% 45.14% 43.45% 50.43% 88.93%
XNAS:AMZN 27.53% 27.70% 100.00% 9.69% -4.01% 21.50% 29.89% 56.80% 14.88%
XNYS:ATI 62.56% 39.74% 31.08% 100.00% 53.40% 27.27% 19.85% 46.78% 53.58%
XNYS:CA -8.73% 22.50% -1.04% 55.91% 100.00% 12.81% 23.87% -6.53% 19.11%
XNAS:CSCO 19.55% 29.16% 17.78% 10.54% 7.72% 100.00% 30.76% 3.64% -0.38%
XNAS:DELL 15.37% 35.24% 0.30% 12.43% 30.59% -9.09% 100.00% -8.10% 6.23%
XNAS:EBAY 73.62% 49.45% 51.76% 13.15% 15.23% 23.12% 4.10% 100.00% 53.38%
XNYS:EMC 49.89% 81.99% 19.40% 14.33% -26.38% 29.63% 4.20% 82.83% 100.00%
XNYS:IBM 26.59% -10.44% 41.74% 19.63% -20.26% 9.92% 26.90% 41.08% 63.63%
XNYS:GLW 11.48% 9.38% 42.82% 37.99% 37.54% 17.25% 3.89% 0.73% -2.16%
XNAS:GOOG 30.88% 55.21% 18.57% 45.53% 19.20% -4.93% 51.57% 38.53% 31.13%
XNAS:MSFT 31.98% 4.79% 6.08% 24.00% -3.50% -15.98% 59.02% 33.07% 0.00%
XNAS:YHOO 48.98% 79.05% 42.86% 40.60% 54.14% 45.56% 56.75% 66.95% 38.58%
Submit Job
Sheet
Input Ranges
Current Sheet
Job Complete, See Results
Processor Count
4
Runtime (minutes)
30
Infinite
Send Status Email
Submit
Results.CorrelationSeriesChart
Results.CorrelationHeatMap
Results.CovarianceMa
Ready
100%


RealizedCovariance_Tech_14.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Compute Cluster
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
Correlation History
XNAS:AAPL
XNYS:AMD
XNAS:AMZN
XNYS:ATI
XNYS:CA
XNAS:CSCO
XNAS:DELL
XNAS:EBAY
XNYS:EMC
XNYS:IBM
XNYS:GLW
XNAS:GOOG
XNAS:MSFT
XNAS:YHOO
100%
80%
60%
40%
20%
0%
-20%
-40%
-60%
-80%
-100%
Correlation
2006-01-09
2006-01-10
2006-01-11
2006-01-12
Submit Job
Sheet
Input Ranges
Current Sheet
Job Complete, See Results
Processor Count
4
Runtime (minutes)
30
Infinite
Send Status Email
Submit
Results.CorrelationSeriesChart
Results.CorrelationHeatMap
Results.CovarianceMa
Ready
84%

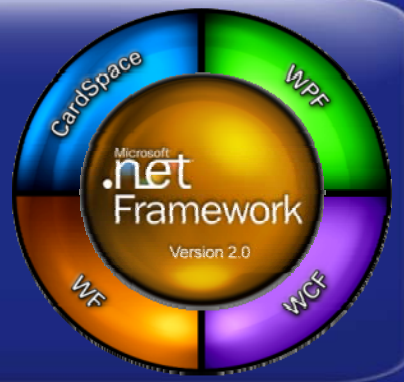

RealizedCovariance_Tech_14.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Compute Cluster
Cut
Copy
Format Painter
Paste
Calibri
11
Wrap Text
Merge & Center
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
C5
University of Was
Realized Covariance & Corr
Iterations
10
Adjustm
Kernel Type
Rectangular
Instruments
XNAS:AAPL
XNYS:AMD
XNAS:AMZN
XNYS:ATI
XNYS:CA
XNAS:CSCO
XNAS:DELL
XNAS:EBAY
XNYS:EMC
XNYS:IBM
XNYS:GLW
Period
Start Da
End Date
1/13/2006
Versions saved for RealizedCovariance_Tech_14.xlsx
Versions saved to: https://portal.microsofthpc.net/uw/My Models Library/
No.
Modified
Modified By
Size
Comments
14.0 11/5/2007 7:22 AM MICROSOFTHPC\hpc_richci 132 KB Model execution complet...
13.0 11/2/2007 7:15 PM MICROSOFTHPC\hpc_richci 132 KB Model execution complet...
12.0 11/2/2007 1:33 PM membershipadprovider:h... 102.8 KB
11.0 11/2/2007 1:33 PM MICROSOFTHPC\hpc_richci 132 KB Model execution complet...
10.0 11/1/2007 5:42 PM MICROSOFTHPC\hpc_richci 132 KB Model execution complet...
9.0 11/1/2007 5:32 PM
8.0 11/1/2007 5:17 PM
7.0 11/1/2007 5:06 PM
6.0 10/31/2007 5:55 PM
5.0 10/31/2007 5:46 PM
Check In Comments
Date and time: 11/5/2007 7:22 AM
Modified by: MICROSOFTHPC\hpc_richci
Comments on version:
Model execution complete: status - Finished
Close
Submit Job
Sheet
Input Ranges
Current Sheet
Job Complete, See Results
Processor Count
4
Runtime (minutes)
30
Infinite
Send Status Email
Submit
Settings
Results.Data
Results.CorrelationSeries
Results.CorrelationSeriesChart
Ready
100%
Windows ...
Microsoft ...
2 Interne...
Connect t...
4 Micros...
SC07_High...
Microsoft ...
Microsoft ...
7:31 AM

Microsoft HPC++ Solution

Application Benefits

The most productive distributed application development environment

Microsoft Office SharePoint Server 2007
Microsoft Office Excel 2007
Microsoft SQL Server 2008
Microsoft Visual Studio 2008



Cluster Benefits

Complete HPC cluster platform integrated with the rest of the enterprise management infrastructure

Windows Server 2008
Active Directory
Microsoft System Center
Windows HPC Server 2008

System Benefits

Cost-effective, reliable and high performance server operating system

Microsoft Hyper-V Server
Windows Server 2008

Resources

- www.microsoft.com/hpc
- www.microsoft.com/science
- www.microsoft.com/servers
- www.microsoft.com/sql
- www.microsoft.com/excel
- research.microsoft.com/fsharp
- www.osl.iu.edu/research/mpi.net
- www.microsoft.com/msdn
- www.microsoft.com/technet

Thank you!