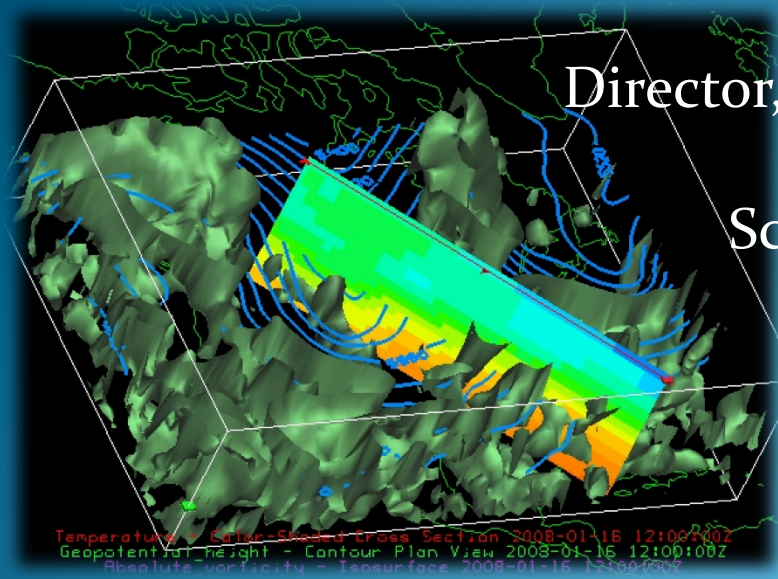




# LEAD II / Trident workflows for timely weather products: the challenge of Vortex2

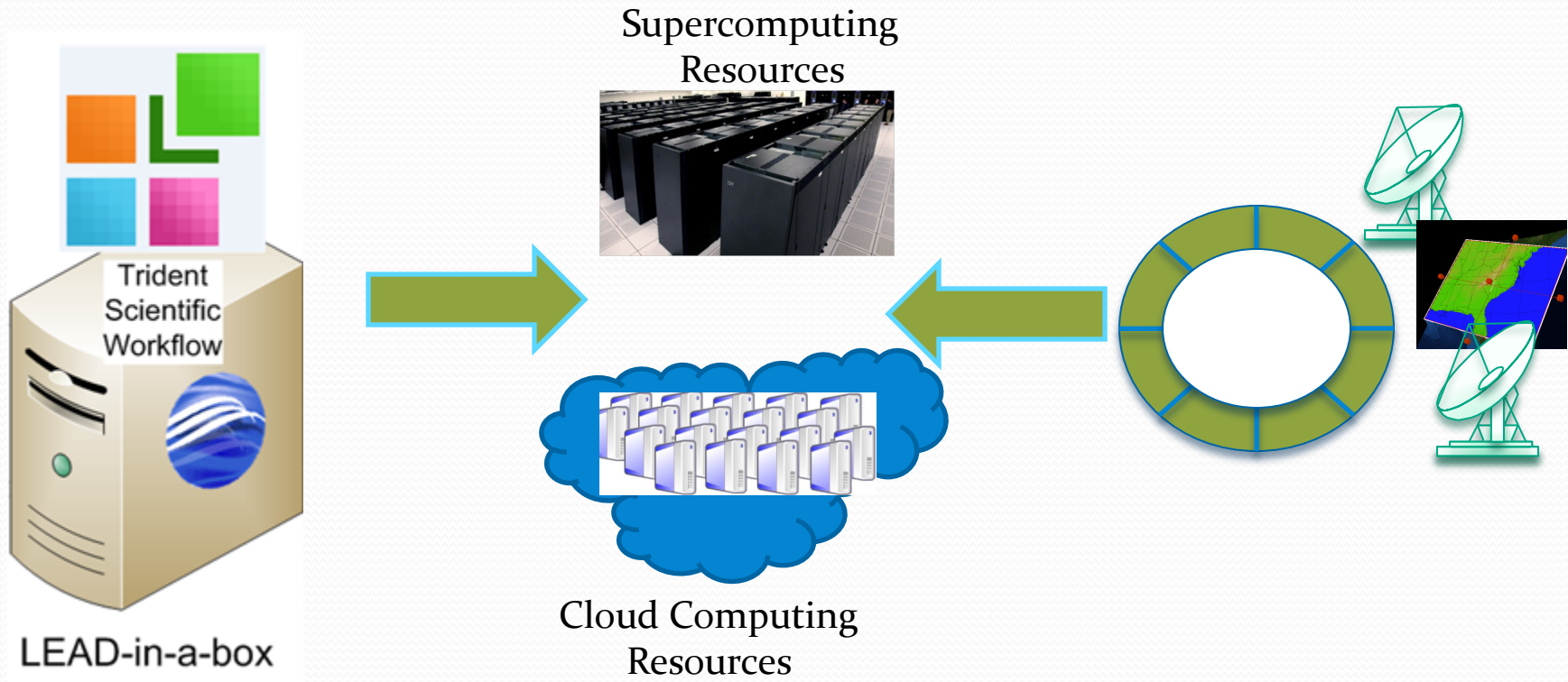


Director, Data To Insight Center of Pervasive  
Technologies Institute  
School of Informatics and Computing  
Indiana University Bloomington

Beth Plale

# Goal: tools for enabling immediate response activity on volumes of recently generated data

- Utilizing Trident Scientific Workflow Workbench



# Current focus: Vortex 2



## Welcome



### • • • What is VORTEX2?

VORTEX2 is by far the largest and most ambitious effort ever made to understand tornadoes. We expect over 100 scientists and over 40 science and support vehicles to participate in this unique, fully nomadic, field program during its second and last field season, May/June 2010. The National Science Foundation (NSF) foundation and the National Oceanic and Atmospheric Administration (NOAA) are contributing over \$10 million towards this effort. Participants will again be drawn from over a dozen

universities, and several government and private organizations. International participants will be drawn from Italy, Netherlands, United Kingdom, Germany, Canada and Australia.

The basic questions driving VORTEX2 are simple to ask, but hard to answer:



The shared Mobile Atmospheric Research and Teaching radar heads toward the Texas Caprock in the panhandle.

### The challenge of observing

Fully nomadic field program studying tornadoes during May/June 2010

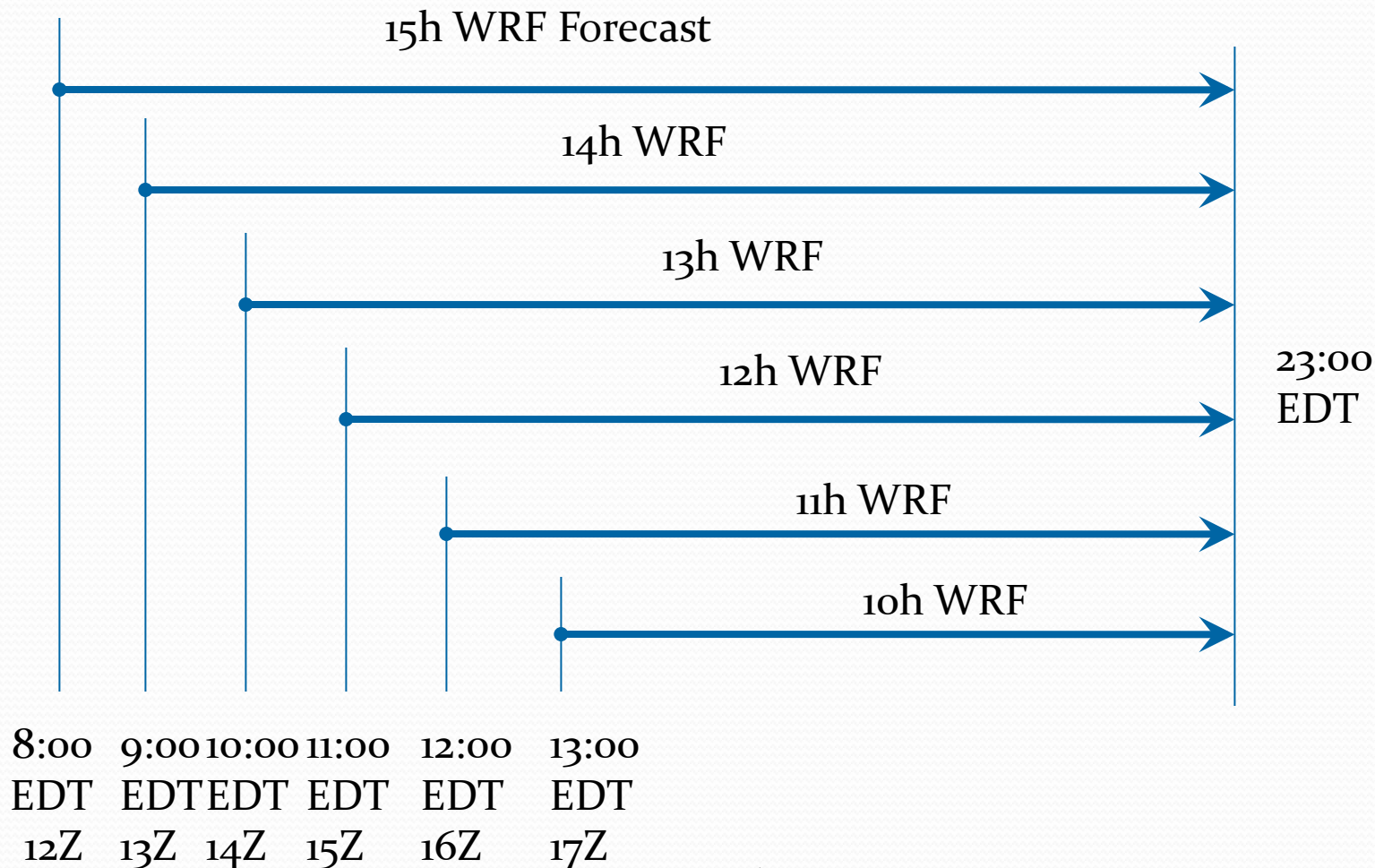
# VORTEX2 2010



- May 1, 2010 to June 15, 2010
- ~6 weeks, 7-days per week
- Some down days when weather is calm
- With Oklahoma University (Keith Brewster) and Univ North Carolina Chapel Hill (Craig Mattocks)



# LEAD II / Vortex 2 run schedule : daily 01 May – 15 June



RUC data:

8:00EDT/12z forecast start kicks off with 10Z RUC,

9:00EDT/13z forecast start kicks off with 11Z RUC, ...



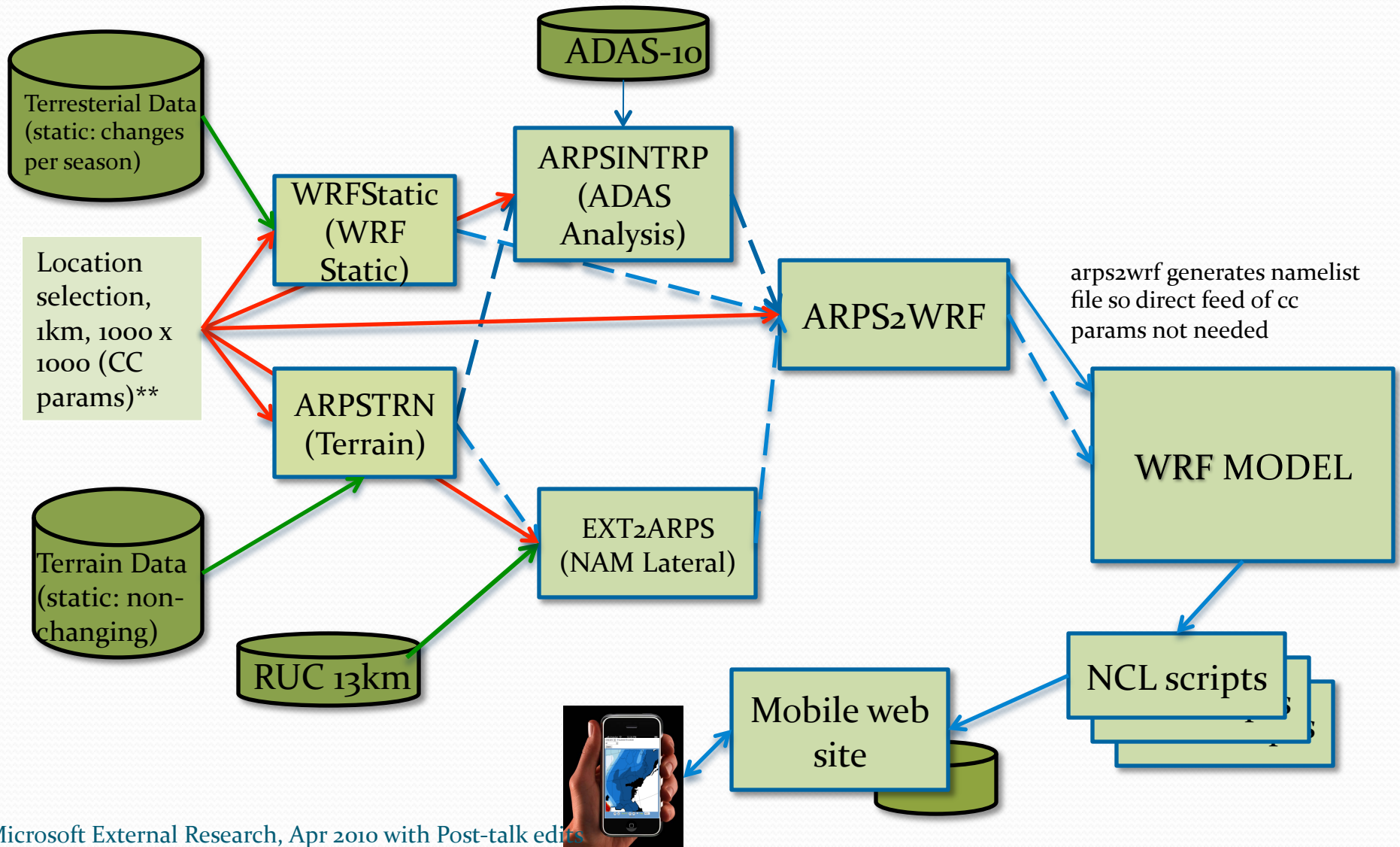
# LEAD II VORTEX2 Workflow

- User-Selected Domain Center
  - Our field guy will advise daily on area of interest
- 1000x1000 4-km WRF Domain
- Initialized with ADAS-10 and Rapid Update Cycle RUC-13km
- Our edge?
  - Dynamic forecasting allows binding to most recent initial conditions data.
  - Oklahoma Univ's assimilation tools gives radar data and better cloud modeling.

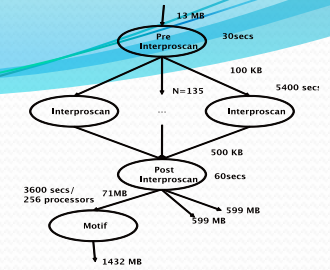
# LEAD II Vortex2 Workflow

ADAS-10 data:

gsiftp://gridftp.bigred.iu.teragrid.org:2812//N/dc/projects/lead/ldm/pub/other/lead/ADAS/10km/ad2010040621.nc



# Workflow Categorization



## I. Size

- Total Number of Tasks
- Number of Parallel Tasks – max number of parallel tasks (width)
- Longest Chain - number of tasks in longest chain
- **LEAD II: not large on any of these dimensions.**

## II. Resource Usage

- Max task processor width – max concurrent processors.
  - **LEAD II: 16 on Windows box, >1000 on Big Red**
- Total Computation time
  - **LEAD II: Time budget is one hour; still working on optimal distribution.**
- Data Sizes – sizes of inputs, outputs and intermediate data products
  - **Ours is < 100 GB.**

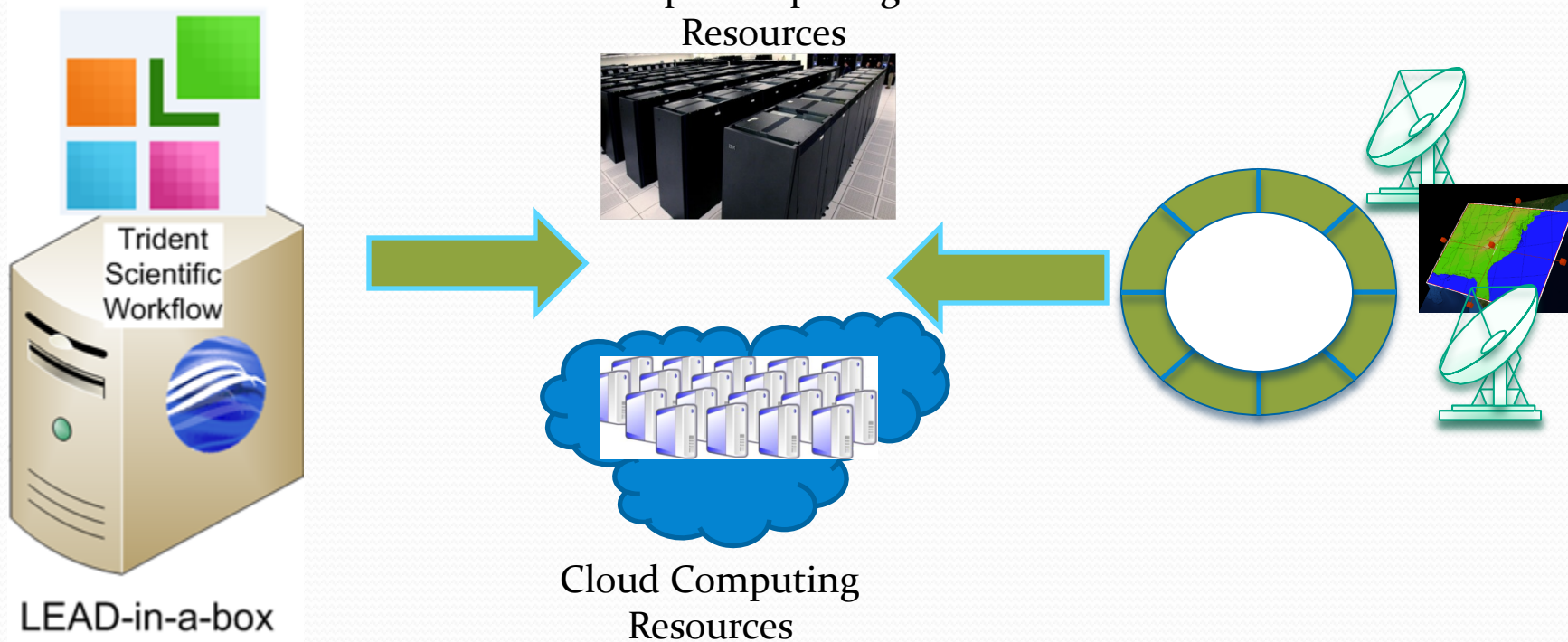
## III. Structural pattern

- Sequential
- Parallel
- Parallel-split - one task's output feeds to multiple tasks.
- Parallel-merge - multiple tasks merge into one task.
- Parallel-merge-split - parallel-merge and parallel-split.
- Mesh - task dependencies are interleaved.
- **LEAD II: Several nodes are parallel but merge within node before control transferred.**

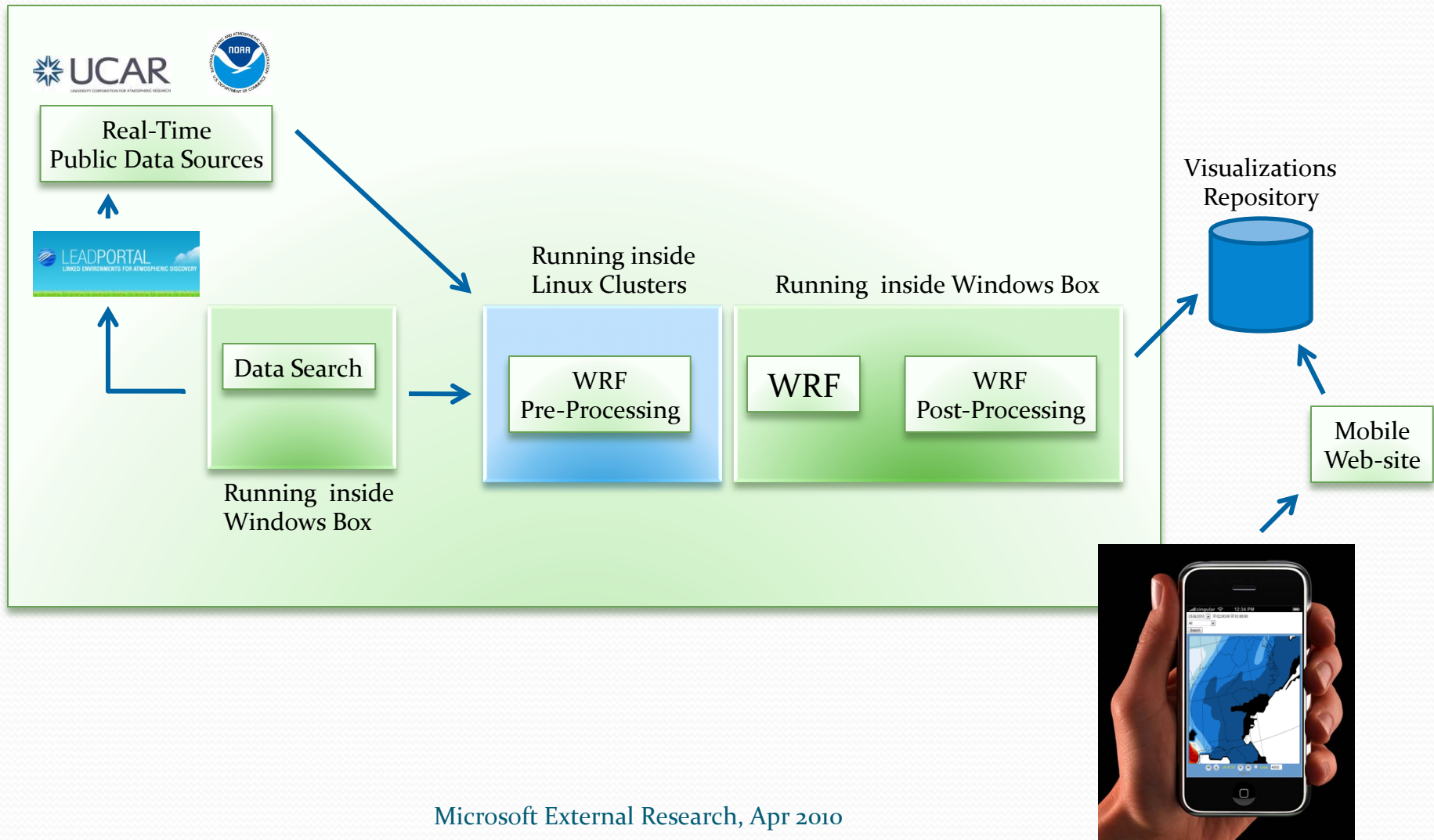


# High Level View

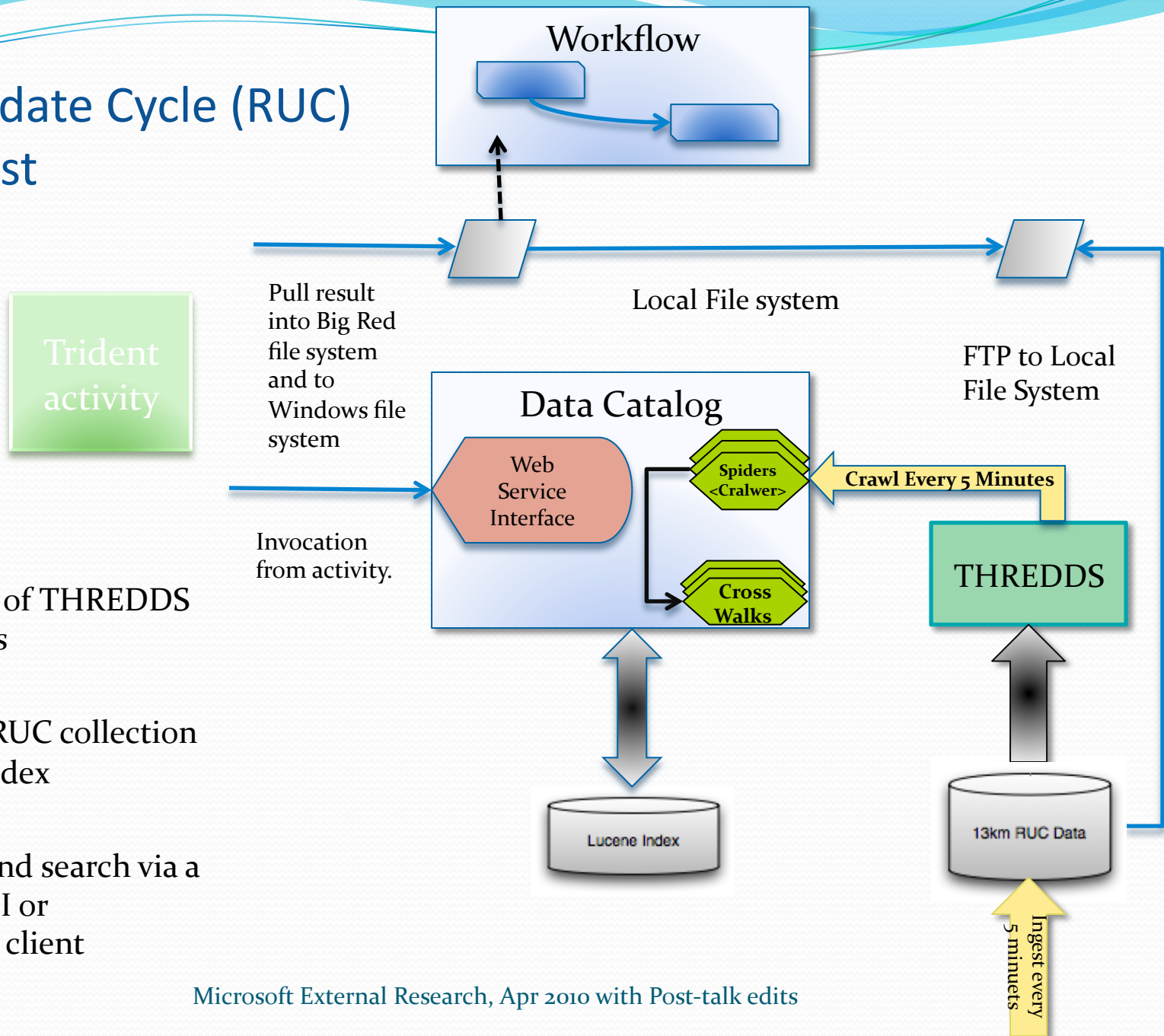
- Trident activities control run of jobs on large-scale computing resources



# Executing workflow from Trident



# Rapid Update Cycle (RUC) data ingest

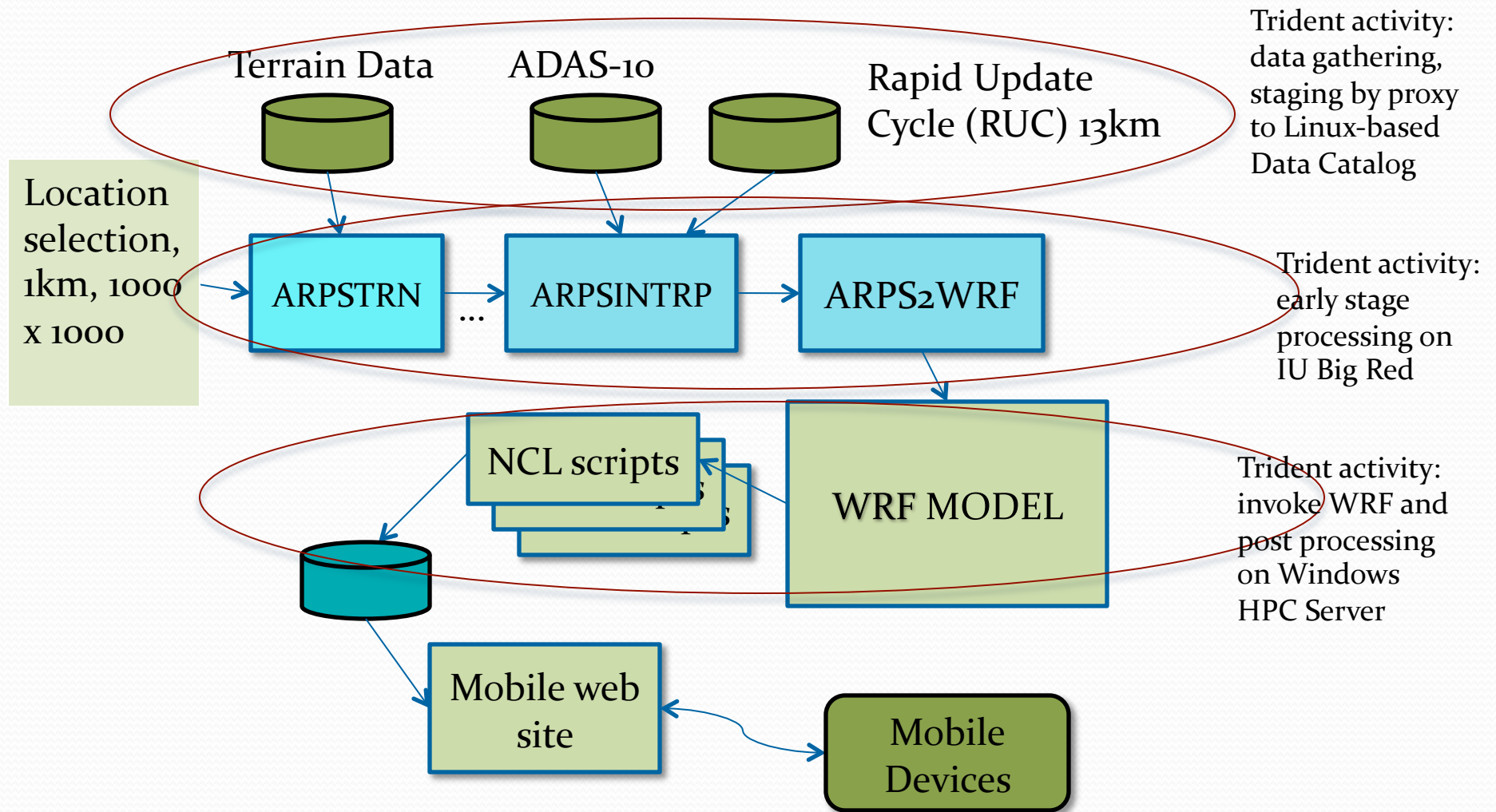


Network crawl of THREDDS (XML) catalogs

Rapid refresh RUC collection registration; index

Support data and search via a web service API or command-line client

# Distribution of Invocation

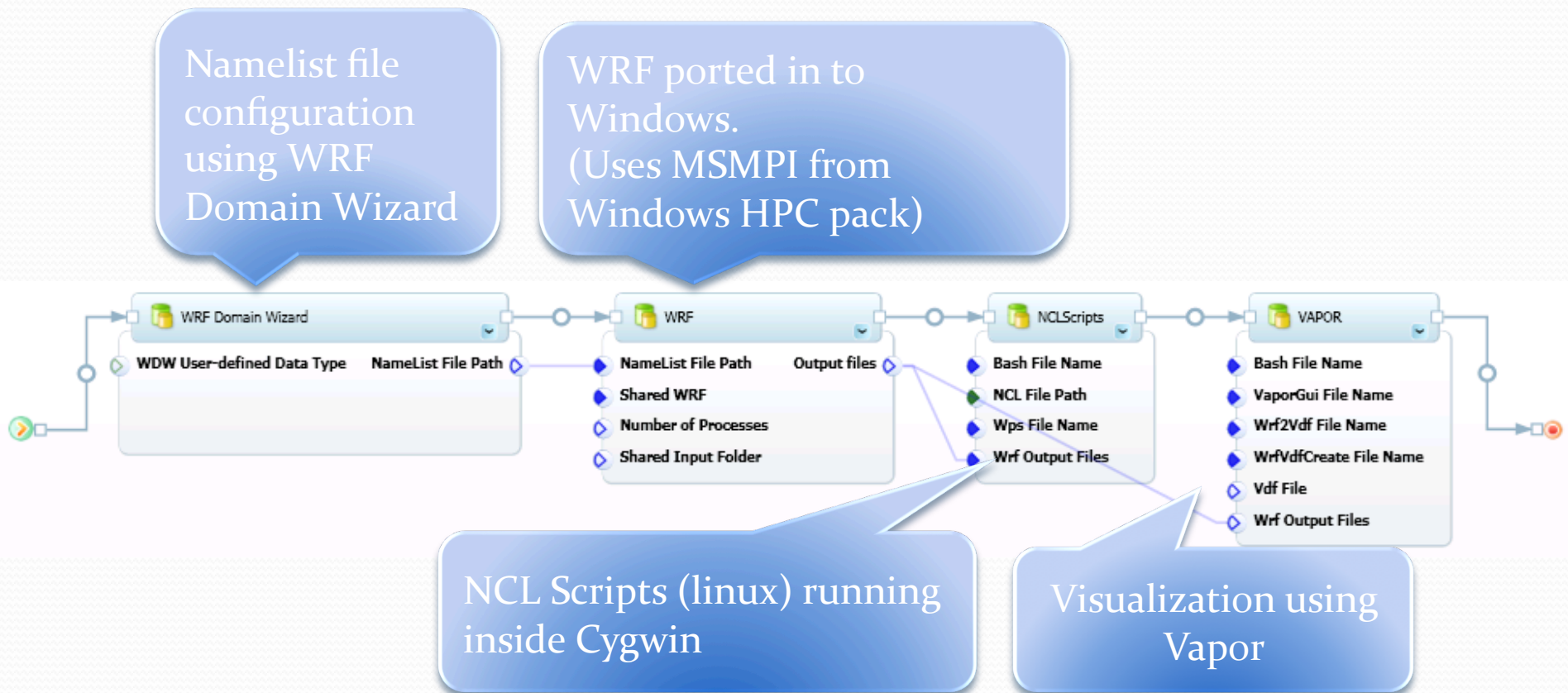




# Invocation of Linux application codes from Windows-hosted Workflow

- Many scientific applications need a Linux environment to execute
- Options to run Linux applications from Windows are:
  - Port the application to Windows
  - Use Linux emulator
  - Proxy activity to Linux services
- Cygwin, a Linux emulator, can run most Linux applications
- *LEAD-in-the-box demonstrated for first time at SC09 Trident orchestrated workflow activities running Linux applications through Cygwin*

# Demonstration workflow



# Workflow Runs Inside Cygwin

The screenshot displays a Windows desktop environment. On the left, the 'SCDemo - Workflow Composer' application is open, showing a workflow diagram with activities: WRF Domain Wizard, WRF, NCLScripts, and VAPOR. The 'Activities' pane on the left lists various workflow components, with 'NCLScripts' highlighted. The main window shows the 'Processing Status' tab with a table of activity events.

Activity	Event	Time	Details
SCDemo	Created	1/14/2010 2:40:53 PM	
SCDemo	Started	1/14/2010 2:40:54 PM	
SequentialWorkflowActivit...	Executing	1/14/2010 2:40:54 PM	
TypeInitializerSetter0	Executing	1/14/2010 2:40:56 PM	
TypeInitializerSetter0	Closed	1/14/2010 2:40:56 PM	
WRF Domain Wizard	Executing	1/14/2010 2:40:56 PM	
WRF Domain Wizard	Closed	1/14/2010 2:41:21 PM	
WRF	Executing	1/14/2010 2:41:21 PM	
WRF	Closed	1/14/2010 2:41:38 PM	
NCLScripts	Executing	1/14/2010 2:41:38 PM	

In the center, a Cygwin terminal window titled 'C:\cygwin\bin\ncl\_scripts' shows the execution of a command: `convert -depth 8 -rotate 0 -density 144 -crop 852.736x700.9+147.949+31 0.54 /tmp/tiffcap3_0002.ps tiff:/tmp/tiffcap3_0002.tiff`. The terminal output shows variable information for 'cmd'.

On the right, the Windows Task Manager 'Processes' tab is open, showing a list of running processes. The 'bash.exe \*32' process is highlighted, indicating the workflow is running in a Cygwin shell.

Image Name	PID	User Name	CPU	Memory (K)	Command
audodg.exe	1216	LOCAL ...	00	1,140 K	
bash.exe *32	3960	Administ...	00	1,560 K	bash -lk
cmd.exe	2436	Administ...	00	1,712 K	cmd /c "
ConsoleApplication...	3700	Administ...	00	6,868 K	"E:\D2IA
csrss.exe	528	SYSTEM	00	1,936 K	C:\Wind
csrss.exe	580	SYSTEM	00	9,504 K	C:\Wind
csrss.exe	1164	SYSTEM	00	2,680 K	C:\Wind
devenv.exe *32	2920	Administ...	00	55,444 K	"C:\Prog
dwm.exe	2220	Administ...	00	1,476 K	"C:\Winc
explorer.exe	2548	Administ...	00	22,080 K	C:\Wind
freesh.exe *32	3056	Administ...	00	2,776 K	"C:\Prog
inetinfo.exe	1488	SYSTEM	00	7,896 K	"C:\Wind
justched.exe	2892	Administ...	00	2,432 K	"C:\Prog
LogonUI.exe	992	SYSTEM	00	6,960 K	"LogonUI
lsass.exe	676	SYSTEM	00	6,084 K	C:\Wind
lsim.exe	684	SYSTEM	00	2,704 K	C:\Wind
msdte.exe	1056	NETWO...	00	3,156 K	C:\Wind
msvsmn.exe	3344	Administ...	00	4,292 K	"C:\Prog
nvSCAPIsvr.exe ...	2024	SYSTEM	00	2,248 K	"C:\Prog
nvsvsc.exe	888	SYSTEM	00	1,288 K	"C:\Wind
nvsvsc.exe	1436	SYSTEM	00	2,504 K	C:\Wind
PresentationFontC...	2428	LOCAL ...	00	5,636 K	C:\Wind
rdpclip.exe	2424	Administ...	00	1,992 K	rdpclip
services.exe	628	SYSTEM	00	3,596 K	C:\Wind
SLsvc.exe	372	NETWO...	00	7,720 K	C:\Wind
smss.exe	460	SYSTEM	00	396 K	\System
SMSvcHost.exe	1700	LOCAL ...	00	7,620 K	"C:\Winc
spoolsv.exe	1412	SYSTEM	00	6,852 K	C:\Wind

Microsoft External Research, Apr 2010 with Post-talk edits

[NCL Home](#) > [Application examples](#) > [Models](#) || [Data files for some examples](#)

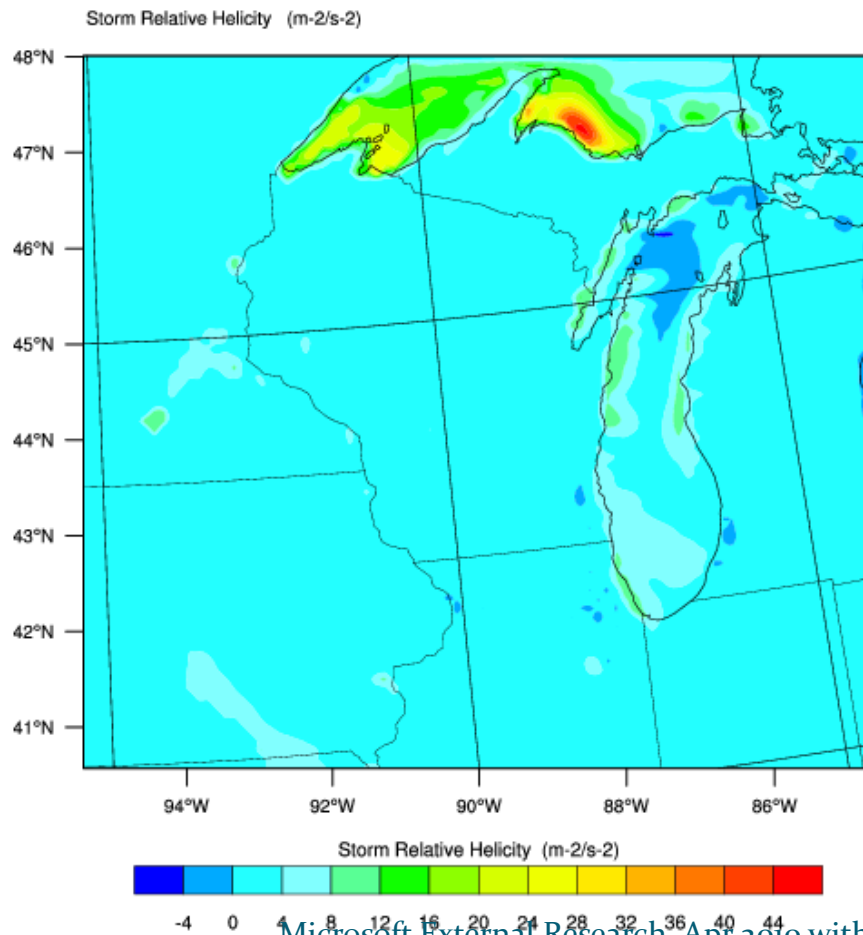
Example pages containing: [tips](#) | [resources](#) | [functions/procedures](#)

## WRF

### WRF Helicity

LEAD REAL-TIME WRF

Init: 2001-06-11\_12:00:00

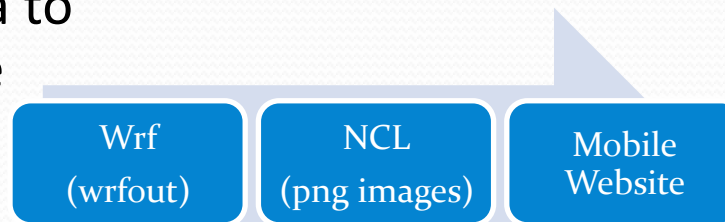
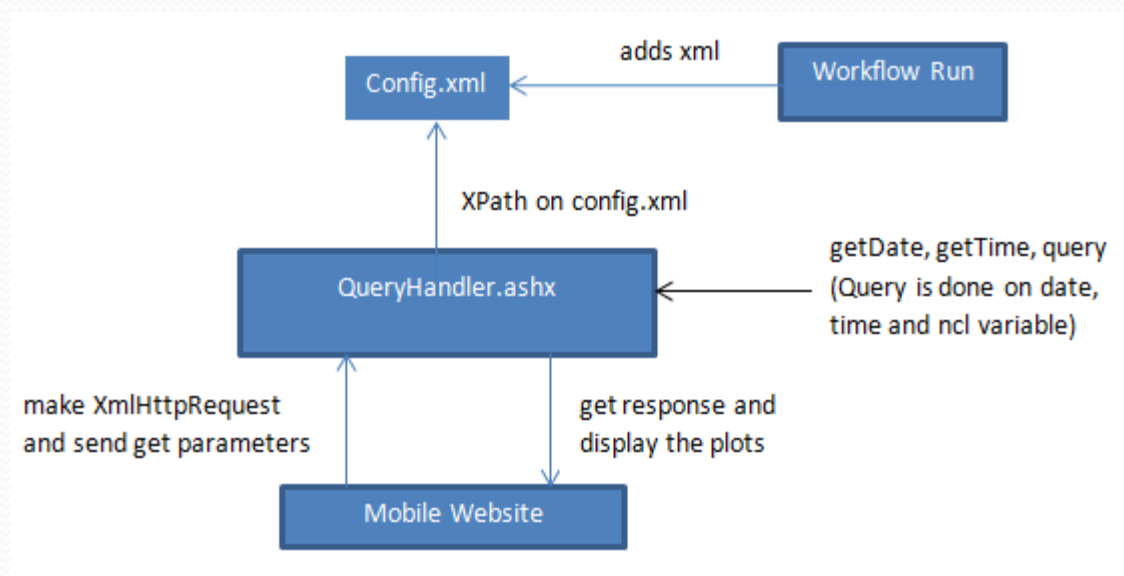


Visualization  
using NCAR  
Command  
Language (NCL)  
scripts for radar,  
helicity



# Mobile Website

- Mobile website
  - Maintains config.xml file
  - Queries config.xml file using QueryHandler
  - Displays plots
- Each workflow run
  - Generates log xml file
  - Adds xml data to config.xml file



# Phone presentation

Select Date  ← Date of the workflow run  
3/4/2010

3/4/2010  02:00:00  01:00:00 ← Time of the workflow run

PressureLevels  ← NCL variable

Search

NCL Plot

Navigation Bar

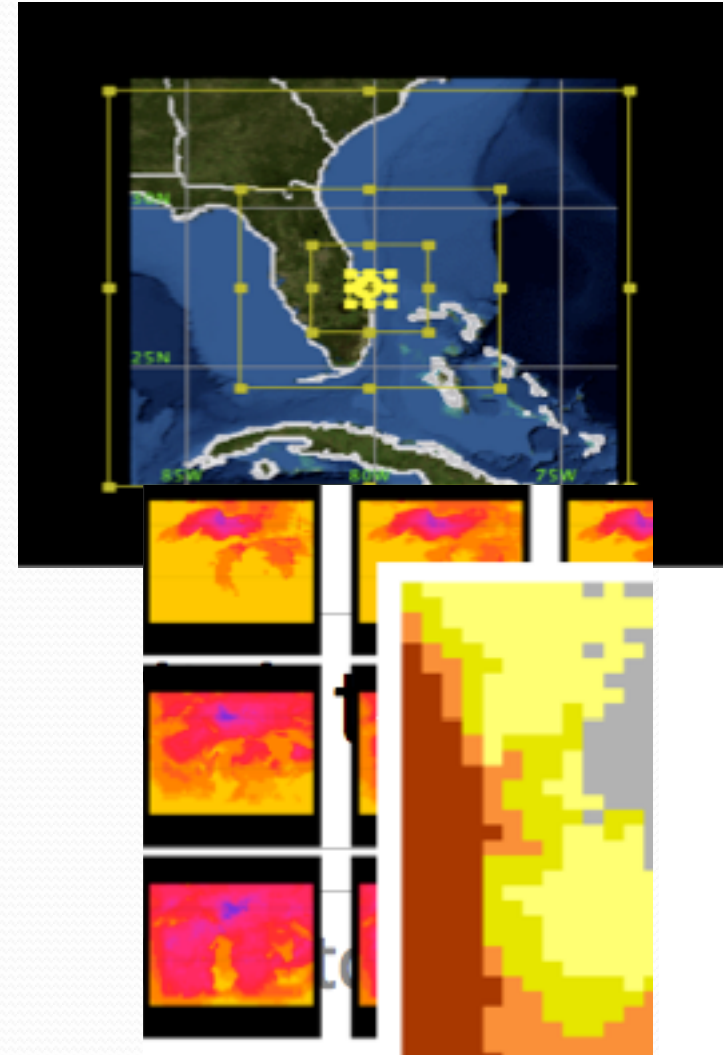
Caption of the plot

17 of 19 Loop 4000  
pressureLevels-1

Microsoft External Research, Apr 2010 with Post-talk edits

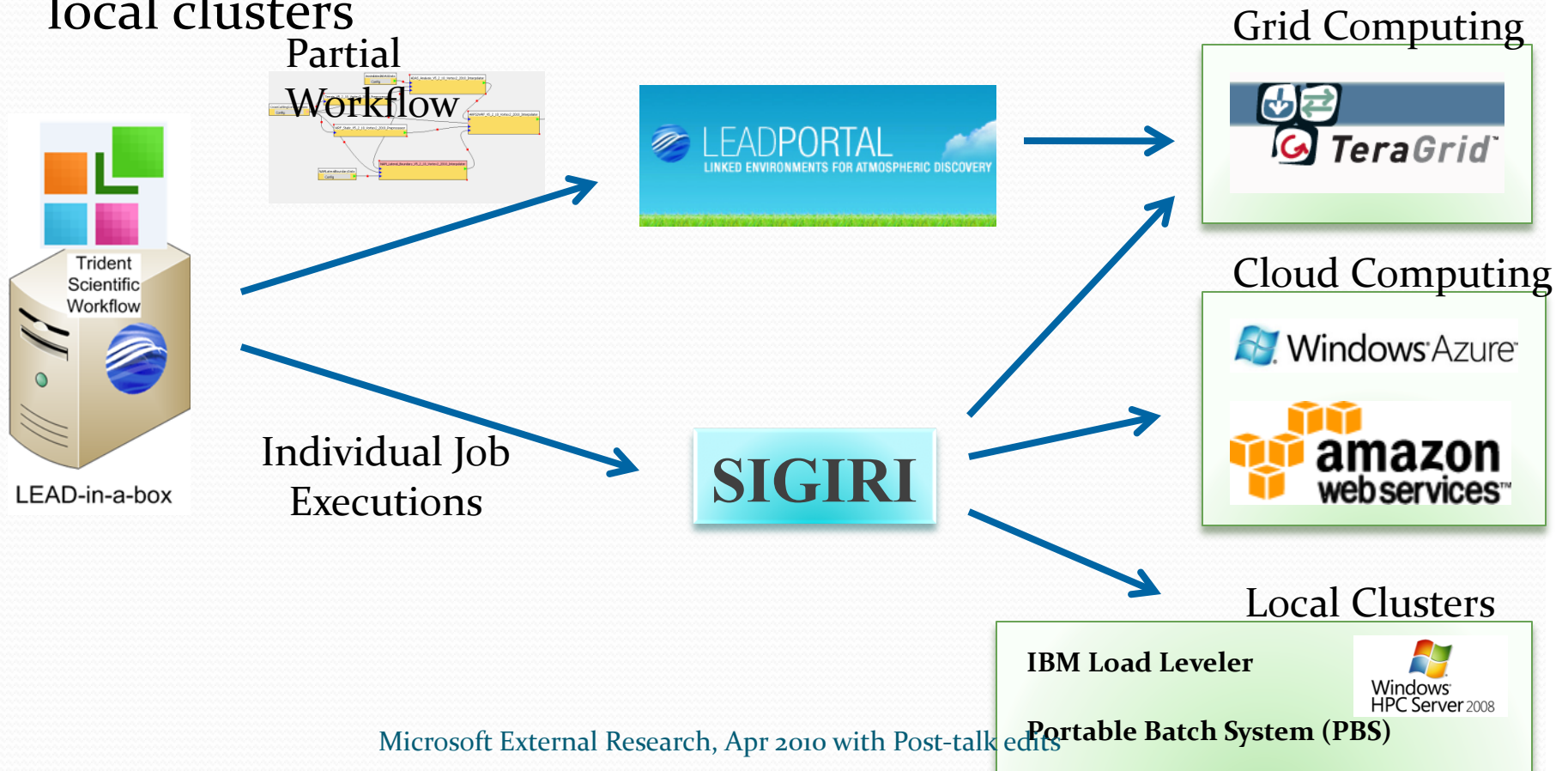
# Overarching goal: Tools that enable real time activity on volumes of real time

- Trident activities that interact with real time constantly refreshing weather data sources.
- Trident accessible repositories capable of programmatic delivery of the precise result that workflow binding requires (e.g., OGC WCS and WFS in FAA NextGen air traffic control).
- Workflow aware scheduling solution across cloud.
- LEAD-in-a-Box (24 core box) that meets local mesoscale needs. E.g., continuous energy output of wind farms (Lidar, WRF); agriculture



# WRF Experiments with Trident – A Future Direction

- Support for job executions in Cloud computing and other local clusters





# Team members



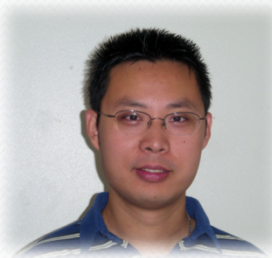
IU



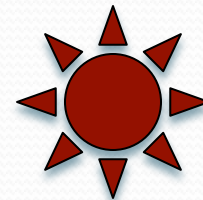
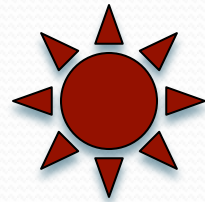
IU



IU



IU



Oklahoma  
Univ  
- Field guy for  
Vortex2



UNC



IU

Microsoft External Research, Apr 2010 with Post-talk edits