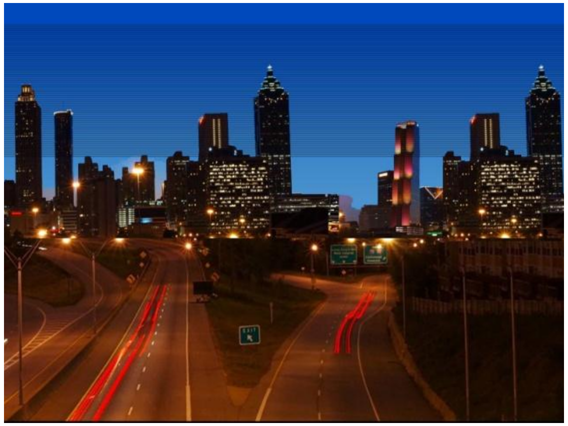
Computational Challenges of Implementing the Atlanta Regional Commission Activity-Based Modeling System



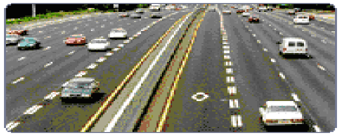
ATLANTA REGION
PLAN
2040

PB 100 YEARS

PBSJ

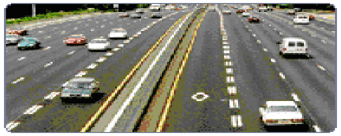
May 11, 2010

ARC
ATLANTA REGIONAL COMMISSION



Presentation Outline

- ARC Activity-Based Model Quick Introduction
- Implementation Design Goals
- Hardware and Software Setup
- Non-CT-RAMP Distribution and Threading
- CT-RAMP Distribution and Threading
- Overall System Setup
- Future Year Runs
- Next Steps
- Conclusions



ARC Activity-Based Modeling System

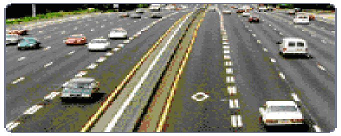
- Based on the CT-RAMP[1] family of ABMs developed in New York, NY, Columbus OH (MORPC) and others
  - Explicit intra-household interactions
  - Continuous temporal dimension (Hourly time periods)
  - Integration of location, time-of-day, and mode choice models
  - Java-based package for AB model implementation

- Implemented with the existing Cube-based networks, GUI and ancillary models (external model, truck model, assignments, etc)

- Households: 1.7 million in 2005, 2.7 million in 2030

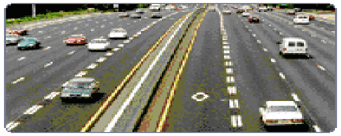
- Model development parallel effort with MTC

[1] Coordinated Travel-Regional Activity-Based Modeling Platform

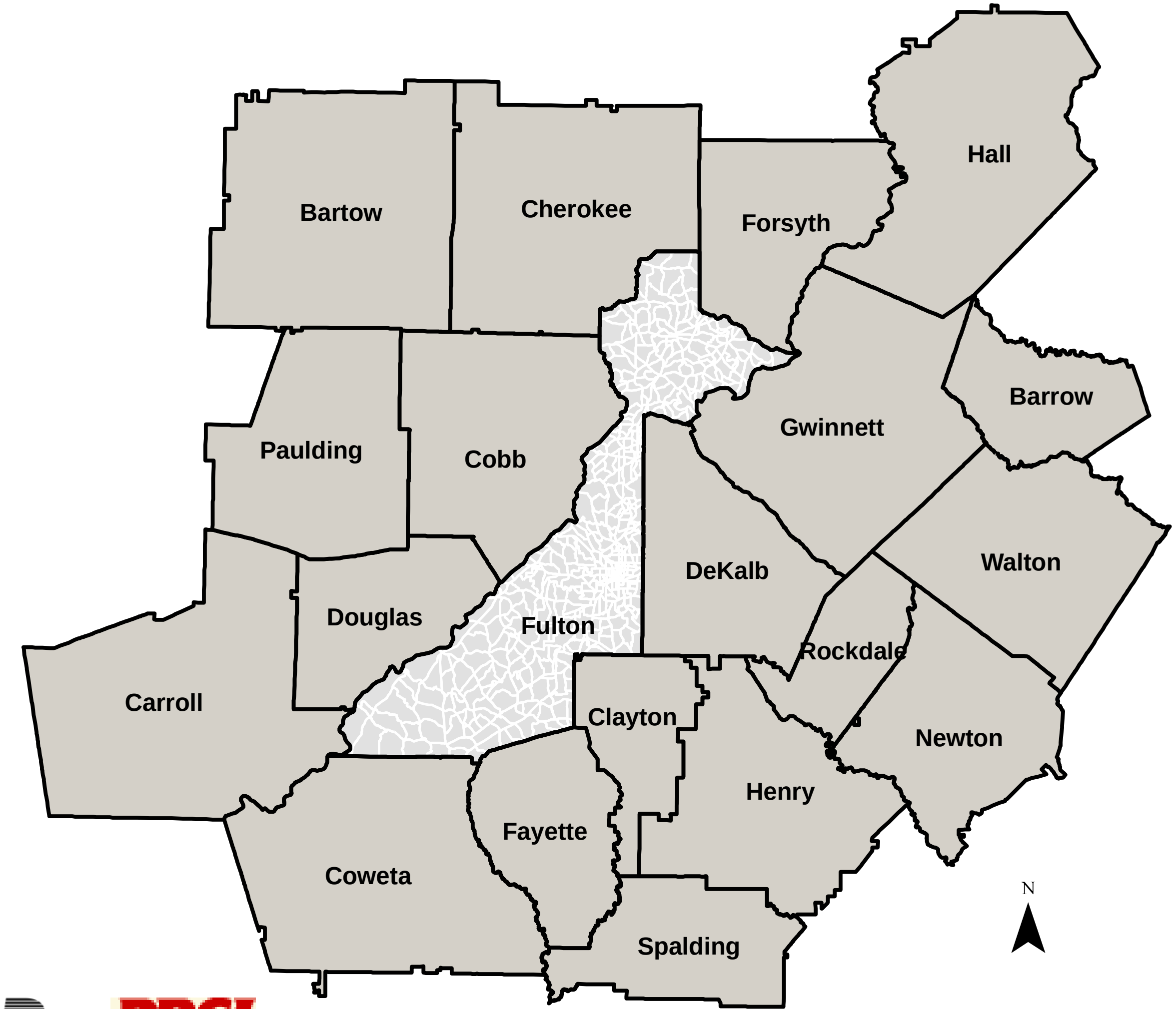


Project History

- 2003 → 2006
  - Models estimated, population synthesizer developed (as presented @ ITM 2006 in Austin TX)
- 2007 → 2008
  - Model implementation, calibration started
- 2009 → April 2010
  - **Calibration/validation completed**, documentation, deployment at ARC, and sensitivity testing
- Remainder 2010
  - Enhanced data reporting and visualization of outputs



Treatment of Space



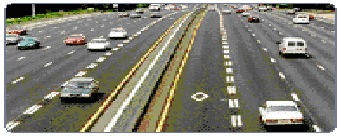
- 2027 TAZs

- TAZs subdivided into transit accessibility:

- Short walk (1/3 mi)
- Long walk (2/3 mi)
- No walk (> 2/3 mi)

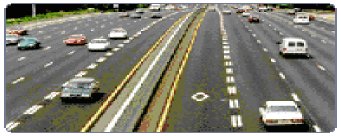
- All origins and destinations identified by TAZ and sub-zone

- 6081 total alternatives in destination choice



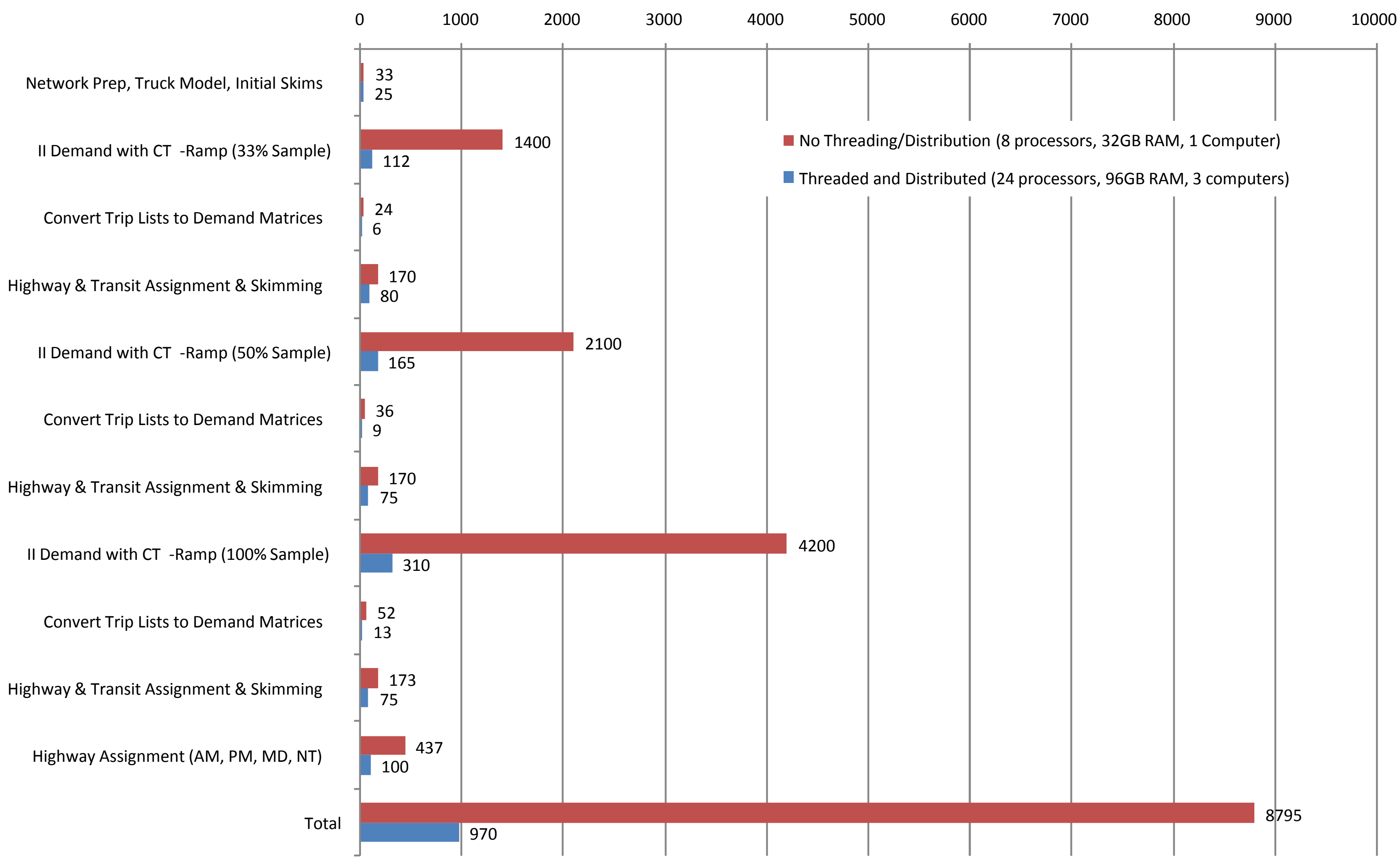
Implementation Design Goals

- Overnight run time → Model Relevance
  - Around 16 hours
  - Requires distribution and threading
- Commodity hardware → Minimize total lifetime cost
  - Hardware available today from common vendors; reasonably priced
- Easy to Setup and Use → Staff acceptance
  - Not too complicated to setup, run, debug, etc



ATLANTA REGION PLAN 2040

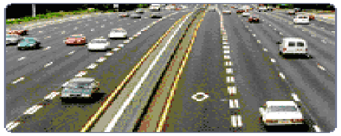
ARC ABM Run Times (min)



PB 100 YEARS

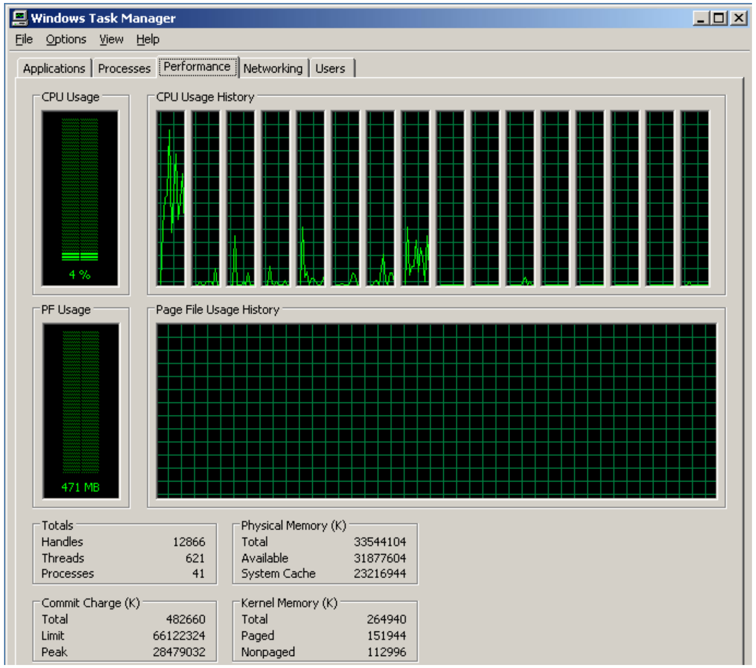
PBSJ

ARC ATLANTA REGIONAL COMMISSION



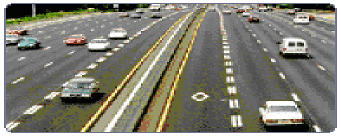
Hardware and Software Setup

- Three Windows Server 2003 64bit Machines:
  - Dual Quad Core Intel Xeon X5570 2.93 GHz with Hyper-Threading → 16 threads
  - 32 GB of RAM
  - Cube Voyager + 8 seat Cube Cluster license



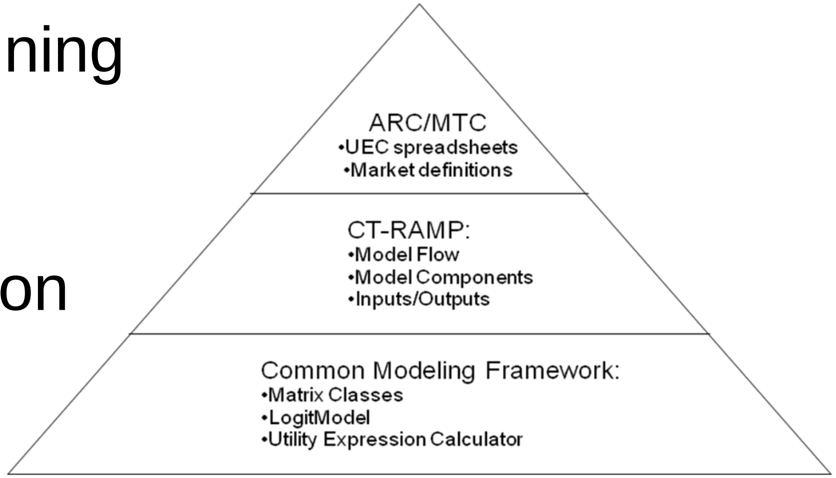
```
Totals                        Physical Memory (K)
Handles          12866        Total            33544104
Threads            621        Available        31877604
Processes           41        System Cache     23216944

Commit Charge (K)             Kernel Memory (K)
Total           482660        Total              264940
Limit         66122324        Paged              151944
Peak          28479032        Nonpaged           112996
```

- Total cost ~ $30,000 in 2009



Hardware and Software Setup

- 64 bit OS for large memory addresses
- 64 bit Java for CT-RAMP
- 32 bit Java to integrate with Cube's native matrix I/O DLL
- Cube Base for the GUI
- Cube Voyager + Cluster for running the model, assignment, etc
- Java CT-RAMP software
- 64 bit R for reporting/visualization

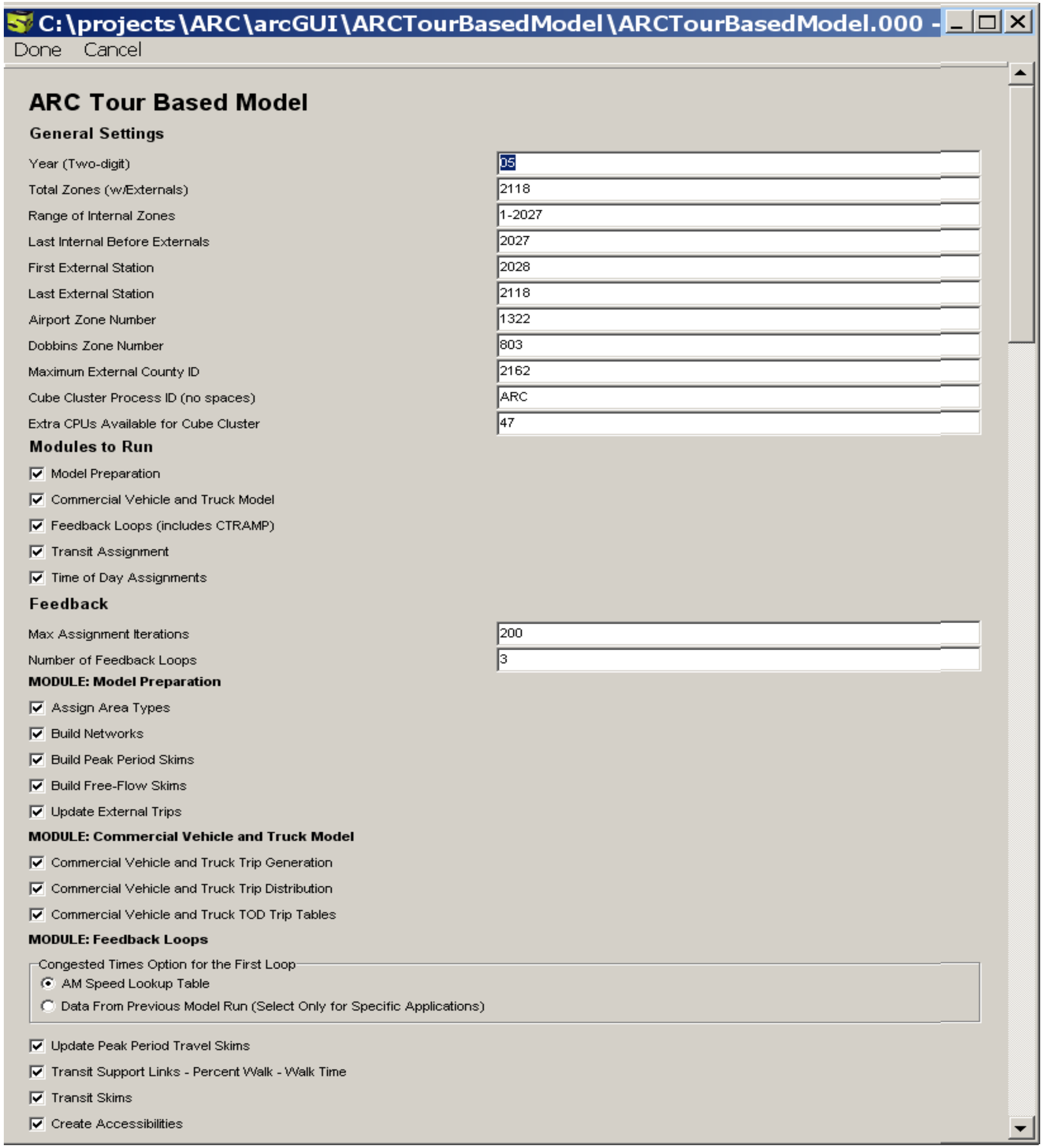


ARC/MTC
•UEC spreadsheets
•Market definitions

CT-RAMP:
•Model Flow
•Model Components
•Inputs/Outputs

Common Modeling Framework:
•Matrix Classes
•LogitModel
•Utility Expression Calculator

User Interface



C:\projects\ARC\arcGUI\ARCTourBasedModel\ARCTourBasedModel.000 -
Done
Cancel

ARC Tour Based Model

General Settings

Year (Two-digit)	05
Total Zones (w/Externals)	2118
Range of Internal Zones	1-2027
Last Internal Before Externals	2027
First External Station	2028
Last External Station	2118
Airport Zone Number	1322
Dobbins Zone Number	803
Maximum External County ID	2162
Cube Cluster Process ID (no spaces)	ARC
Extra CPUs Available for Cube Cluster	47

Modules to Run

☑ Model Preparation
☑ Commercial Vehicle and Truck Model
☑ Feedback Loops (includes CTRAMP)
☑ Transit Assignment
☑ Time of Day Assignments

Feedback

Max Assignment Iterations	200
Number of Feedback Loops	3

MODULE: Model Preparation

☑ Assign Area Types
☑ Build Networks
☑ Build Peak Period Skims
☑ Build Free-Flow Skims
☑ Update External Trips

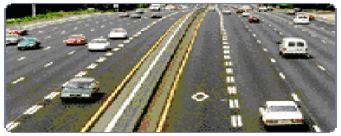
MODULE: Commercial Vehicle and Truck Model

☑ Commercial Vehicle and Truck Trip Generation
☑ Commercial Vehicle and Truck Trip Distribution
☑ Commercial Vehicle and Truck TOD Trip Tables

MODULE: Feedback Loops

Congested Times Option for the First Loop
● AM Speed Lookup Table
○ Data From Previous Model Run (Select Only for Specific Applications)

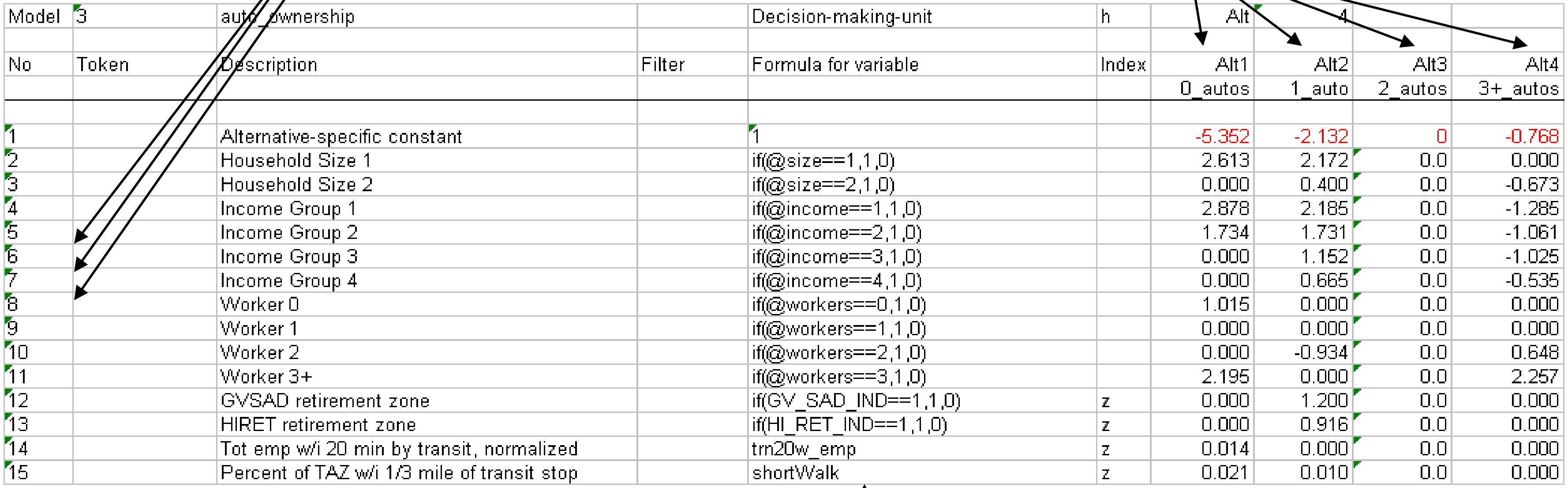
☑ Update Peak Period Travel Skims
☑ Transit Support Links - Percent Walk - Walk Time
☑ Transit Skims
☑ Create Accessibilities



CT-RAMP Model UEC Example

A row for each utility term

A column for each alternative (0, 1, 2, and 3+ autos)

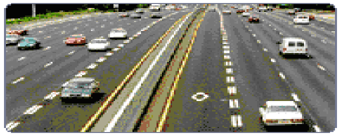


Model	3	auto_ownership		Decision-making-unit	h	Alt	4		
No	Token	Description	Filter	Formula for variable	Index	Alt1	Alt2	Alt3	Alt4
						0_autos	1_auto	2_autos	3+_autos
1		Alternative-specific constant		1		-5.352	-2.132	0	-0.768
2		Household Size 1		if(@size==1,1,0)		2.613	2.172	0.0	0.000
3		Household Size 2		if(@size==2,1,0)		0.000	0.400	0.0	-0.673
4		Income Group 1		if(@income==1,1,0)		2.878	2.185	0.0	-1.285
5		Income Group 2		if(@income==2,1,0)		1.734	1.731	0.0	-1.061
6		Income Group 3		if(@income==3,1,0)		0.000	1.152	0.0	-1.025
7		Income Group 4		if(@income==4,1,0)		0.000	0.665	0.0	-0.535
8		Worker 0		if(@workers==0,1,0)		1.015	0.000	0.0	0.000
9		Worker 1		if(@workers==1,1,0)		0.000	0.000	0.0	0.000
10		Worker 2		if(@workers==2,1,0)		0.000	-0.934	0.0	0.648
11		Worker 3+		if(@workers==3,1,0)		2.195	0.000	0.0	2.257
12		GVSAD retirement zone		if(GV_SAD_IND==1,1,0)	z	0.000	1.200	0.0	0.000
13		HIRET retirement zone		if(HI_RET_IND==1,1,0)	z	0.000	0.916	0.0	0.000
14		Tot emp w/i 20 min by transit, normalized		trn20w_emp	z	0.014	0.000	0.0	0.000
15		Percent of TAZ w/i 1/3 mile of transit stop		shortWalk	z	0.021	0.010	0.0	0.000

A description for the term

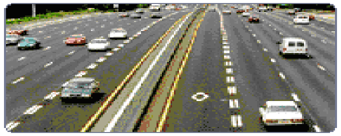
A formula field for computing data items

Coefficients for each term and alternative



Distributing and Threading Non-CT-RAMP

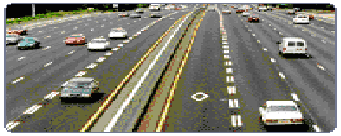
- ARC ABM started with the ARC trip-based model
- Replaced Internal-Internal (II) with CT-RAMP
- Other models in Cube Voyager
  - Network processing
  - Commercial vehicle model
  - Airport model
  - External model
  - Non-II time-of-day model
  - Transit network building
  - Highway and transit assignment
  - Skimming
- Needed to distribute and thread Cube components



Distributing and Threading in Cube Cluster

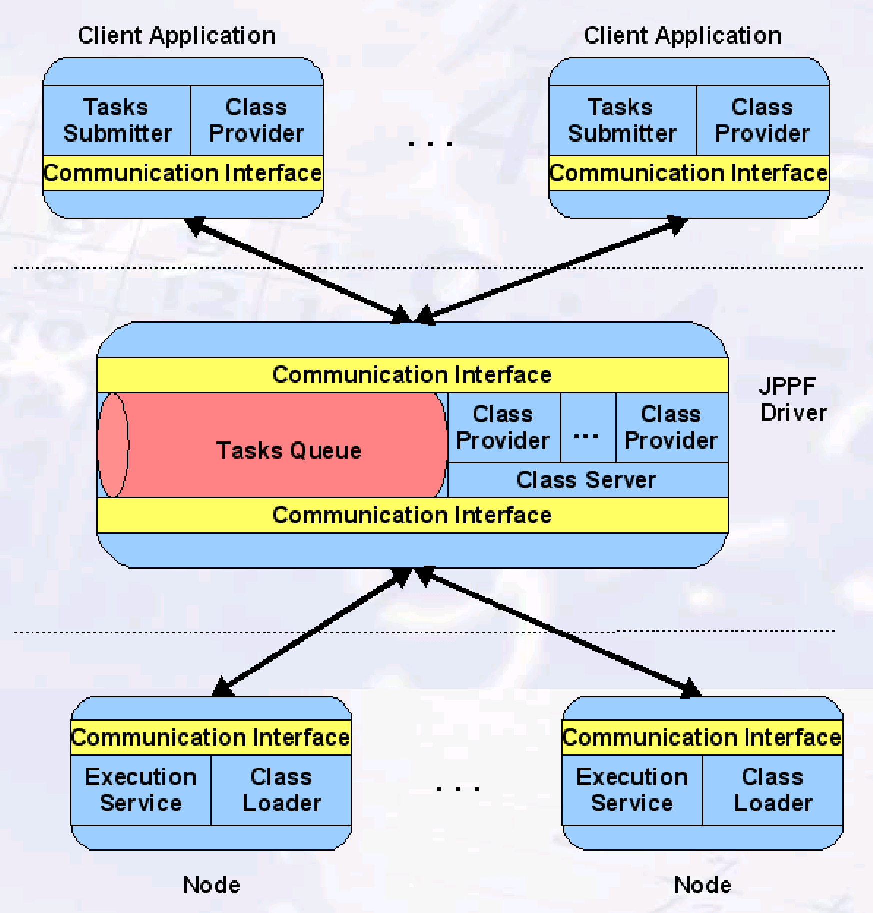
- DistributeINTRAStep
  - Multithreading by origin zone
  - Flexible configuration
  - Highway assignment → AM (x4), MD, PM, NT
  - Matrix processing → Creating time-of-day matrices, etc
  - SPEED UP: ~4X

- DistributeMULTIStep
  - Distribute programs across processors and wait for completion
  - Requires explicit assignment of tasks to processes
  - Highway assignment by Time-Of-Day
  - Creating assignment matrices by Time-Of-Day
  - Transit assignments by Time-Of-Day
  - SPEED UP: ~3X



Distributing and Threading CT-RAMP

- Decompose computations by Households
  - Distribute by groups of households (2000 at a time)
- Implement Java Parallel Processing Framework (JPPF)
  - Open source library to run and manage the distribution of parallel tasks
  - Most computation done on the worker nodes



Client Application

Tasks Submitter | Class Provider

Communication Interface

. . .

Client Application

Tasks Submitter | Class Provider

Communication Interface

Communication Interface

Tasks Queue

Class Provider ... Class Provider

Class Server

Communication Interface

JPPF Driver

Communication Interface

Execution Service | Class Loader

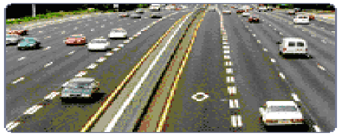
Node

. . .

Communication Interface

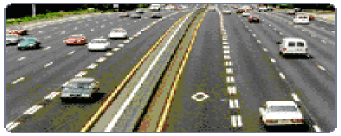
Execution Service | Class Loader

Node



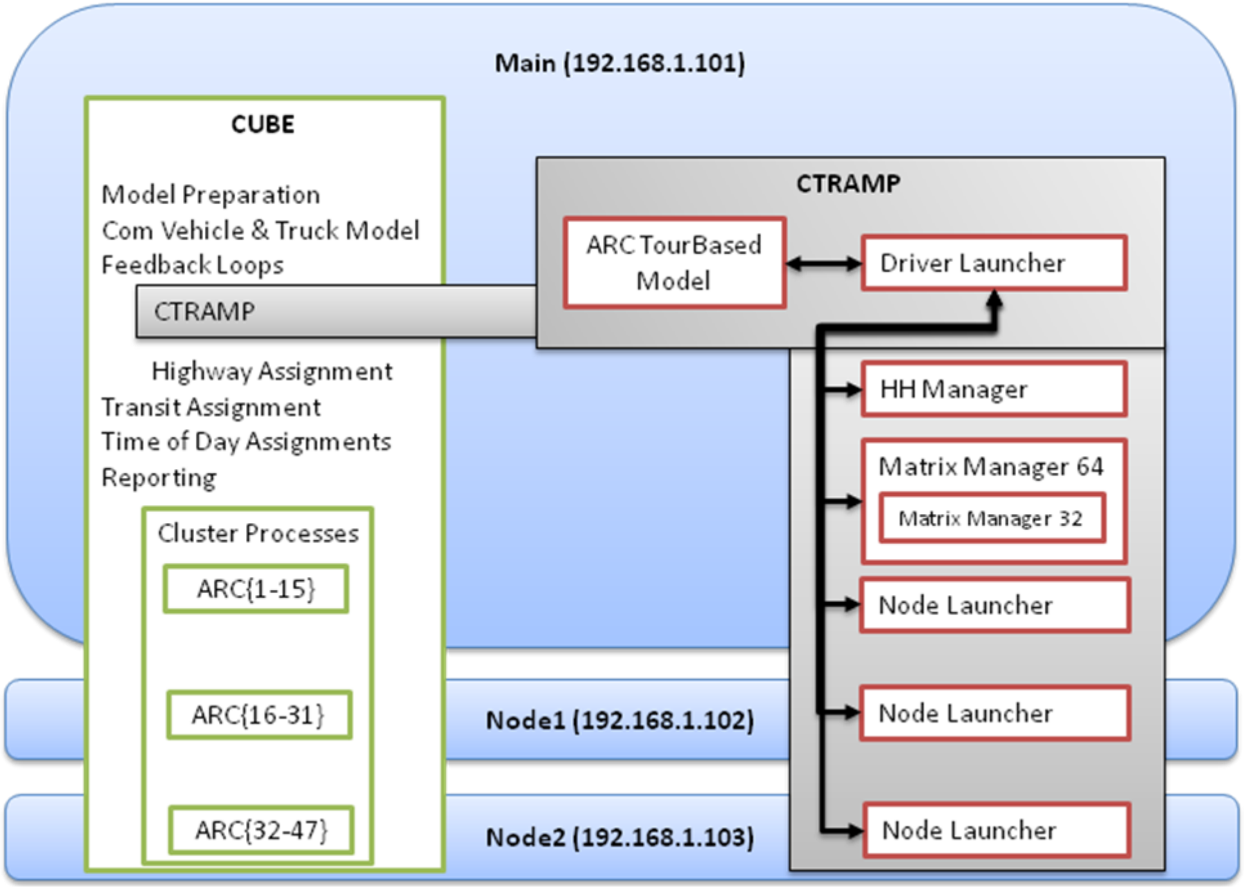
Distributing and Threading CT-RAMP

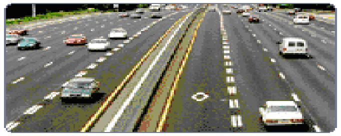
- Main Cube model script calls the JPPF client to start CT-RAMP
- ~1.76 million households split into 880 tasks of 2000 HHs
- CT-RAMP data managed through:
  - Household Manager – manages all HH and person data into RAM for quick I/O
  - Matrix Manager – reads all the matrix data into RAM for quick I/O
- Run a sample of HHs to save time: 33% → 50% → 100%
- HHs and Persons store a random number seed to avoid random number sequence order of processing problems
- SPEED UP: 9X



Overall System Setup

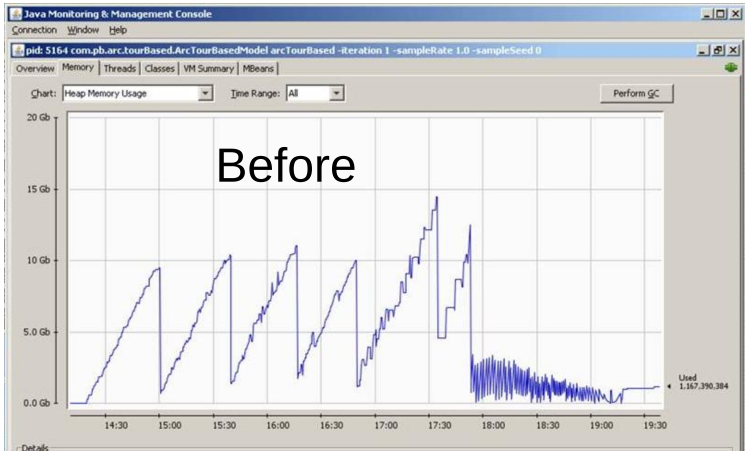
- Cube runs the show and calls all Java processes
- User starts the remote processes on the 2nd and 3rd machine (for now)
- Everything talks to one mapped network folder location





Future Year Runs

- 2.6 million households in 2030
- Significant increase in congestion
- Initial memory “leak” in task → distribution found and fixed that wasn’t a problem in the base year
- Approximately 20 - 21 hour run times depending on the scenario



Java Monitoring & Management Console

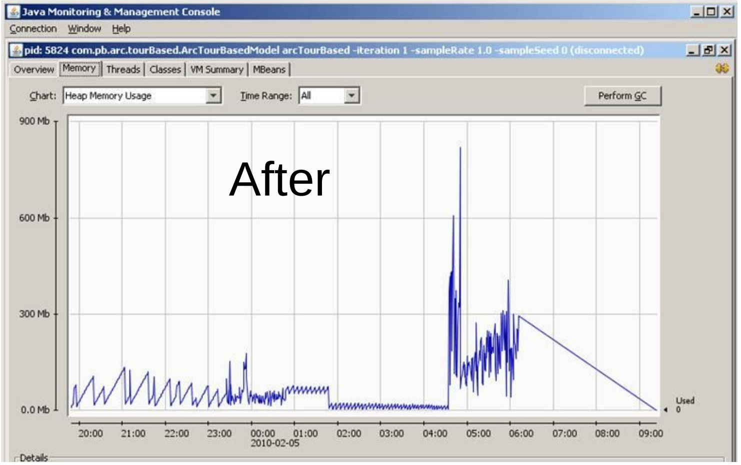
pid: 5164 com.pb.arc.tourBased.ArcTourBasedModel arcTourBased -iteration 1 -sampleRate 1.0 -sampleSeed 0

Chart: Heap Memory Usage

Time Range: All

Used 1,167,390,384

Before



Java Monitoring & Management Console

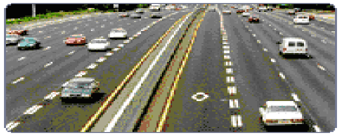
pid: 5824 com.pb.arc.tourBased.ArcTourBasedModel arcTourBased -iteration 1 -sampleRate 1.0 -sampleSeed 0 (disconnected)

Chart: Heap Memory Usage

Time Range: All

Used 0

After

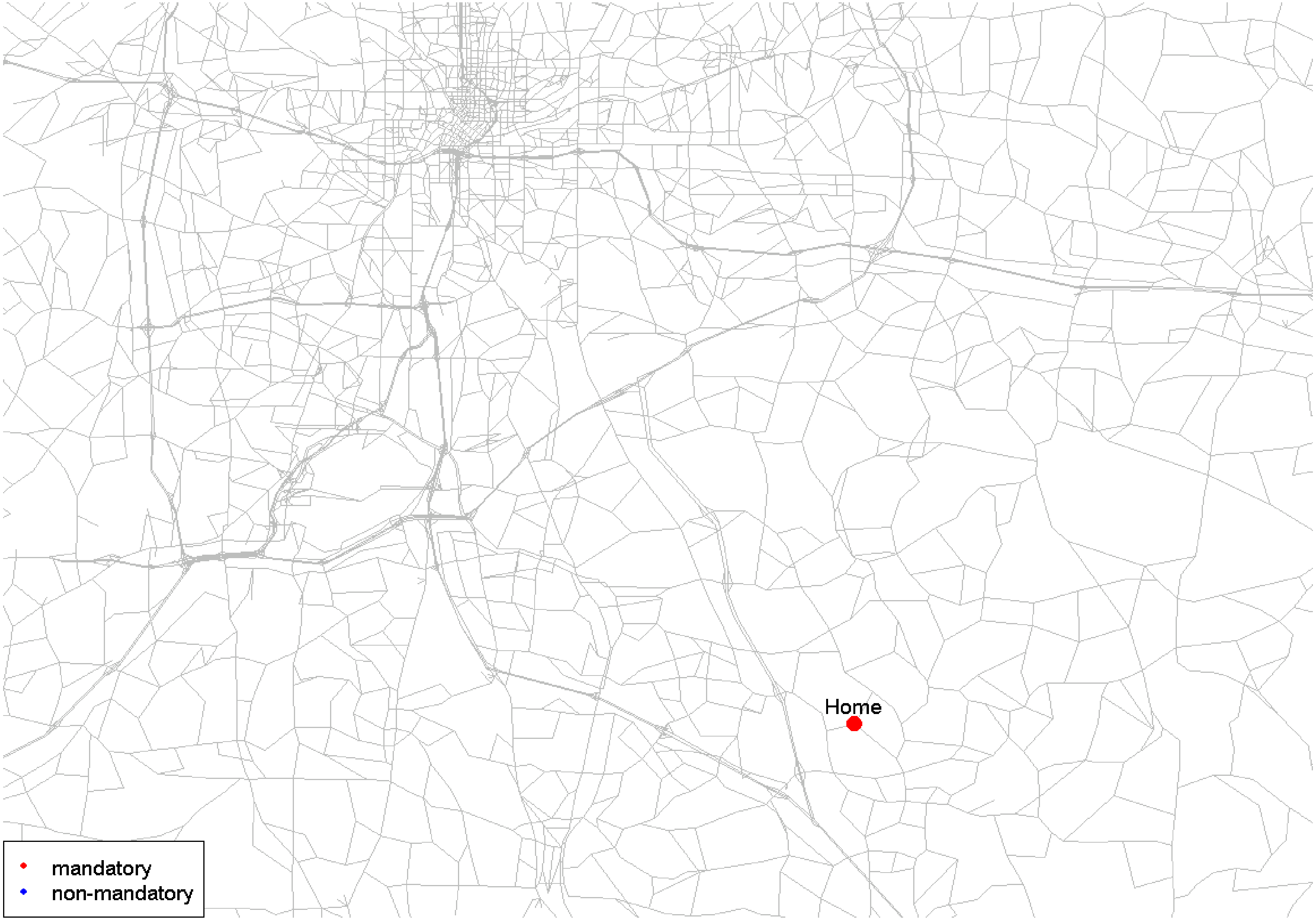


Next Steps

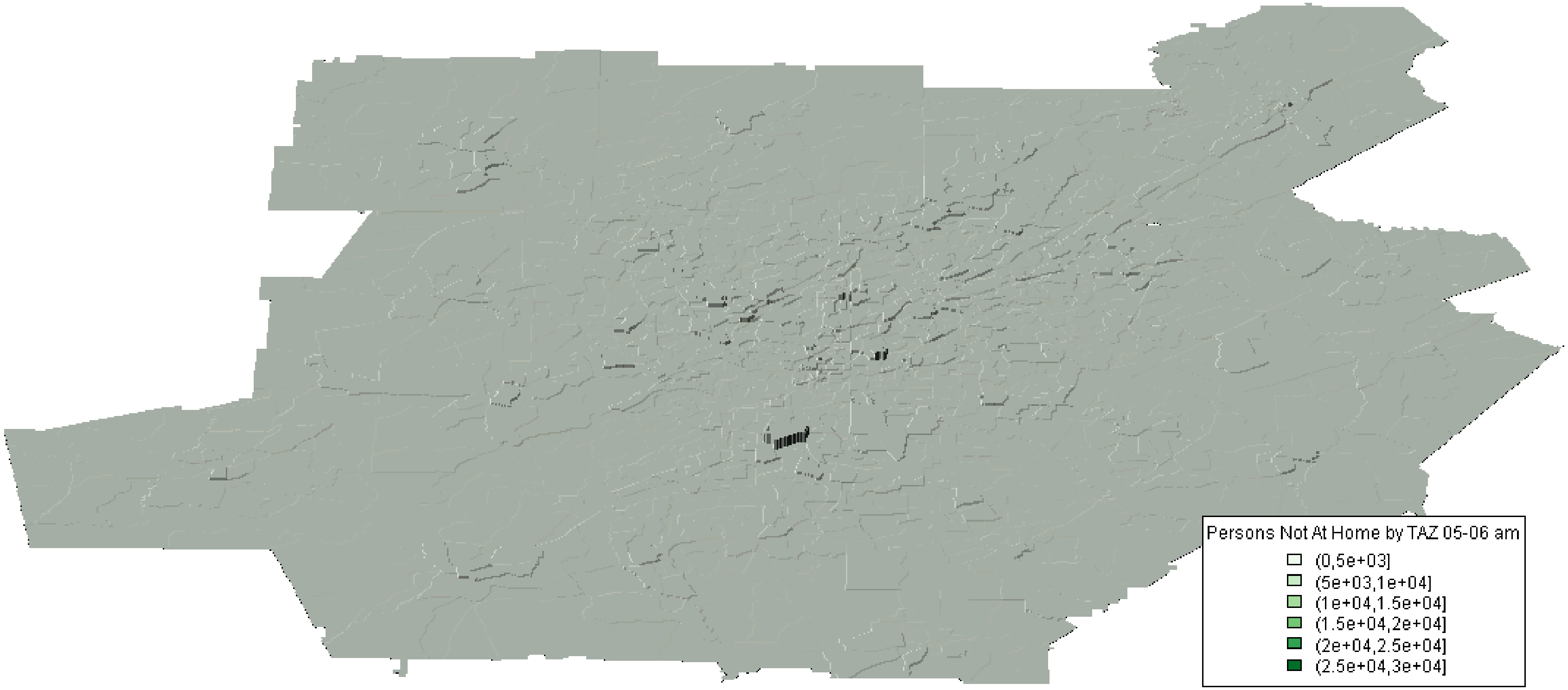
- Improved run times testing
  - Benefit of adding an additional computer
  - Reconfiguring the system setup (moving the matrix manager; adding a 3rd worker node, etc)
  - Additional optimization of CT-RAMP code
- System Setup
  - Automatically start and stop remote processes (with JPPF or WMI)
- Model reporting and visualization tools
  - Some preliminary visualization done
  - Develop dynamic visualization tools (animation, etc) with Adobe Flex/Flash and other tools
- Transit on-board survey analysis and mode choice re-calibration
- Utilize NAVTEQ-conflate highway networks and ARC’s new 2010 20-county 6,000+ TAZ system
- Longer term: Potential integration with PECAS and TRANSIMS

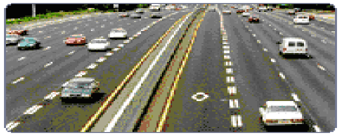
Tracing of Activities/Tours

Person id= 1018897 type= Full-time worker



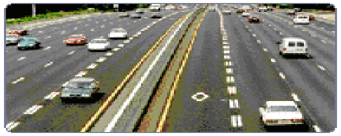
Persons Not At Home By TAZ and Hour





Conclusions

- Significant run time improvements through distribution and threading (16 hours versus 146 hours)
- Leverage commodity hardware and software (and open source software)
- Scalable architecture – can add more hardware and will get even faster!
- Economies of scale for ARC & MTC in model co-development, a +/- 20% cost saving for ARC & MTC
- Model relevance with reasonable run times



ATLANTA REGION PLAN 2040

Questions and Discussion

Ben Stabler, PB, Stabler@pbworld.com

Jim Hicks, PB, HicksJi@pbworld.com

Joel Freedman, PB, Freedman@pbworkld.com

Guy Rousseau, ARC, GRousseau@atlantaregional.com

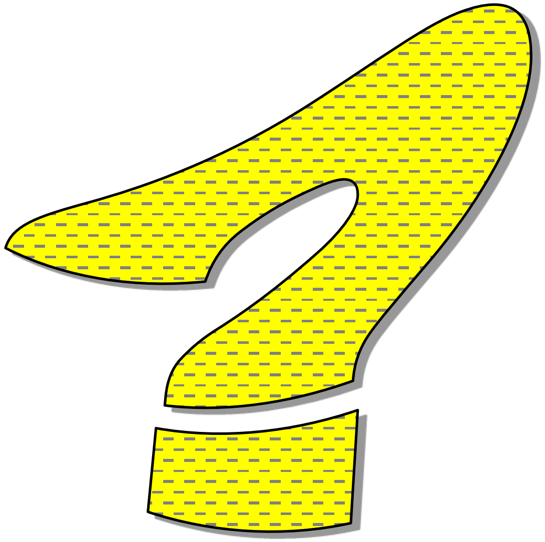
Kyeil Kim, ARC, kykim@atlantaregional.com

Jonathan Nicholson, PBS&J, JLNicholson@pbsj.com

Chris Simons, PBS&J, CNSimons@pbsj.com

Chuck Purvis, MTC, CPurvis@mtc.ca.gov

David Ory, MTC, DOry@mtc.ca.gov



PB 100 YEARS

PBSJ

ARC ATLANTA REGIONAL COMMISSION