



# Joint Experimentation, Data Management and Analysis Enabled by Trans-Continentially Distributed Linux Clusters

**Computational Sciences**



28 June 06

Dan M. Davis

Robert F. Lucas

{ddavis,rflucas}@isi.edu

Review Joint Experimentation

Urban Resolve 2015 Phase 2

Managing the Data



## Larger Scale

Global scale vs. theatre

## Higher Fidelity

Greater complexity in models of entities  
(sensors, people, vehicles, etc.)

## Urban Operations

Millions of civilians

All of the above require dramatic increase in computing power relative to network of workstations previously available.



## Driven by Joint Advanced Warfighting Project (JAWP)

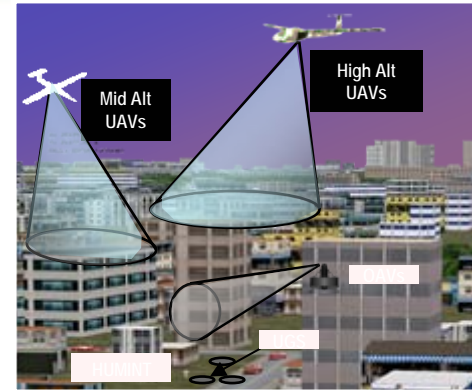
Large urban battle space

100-200K entities (current)

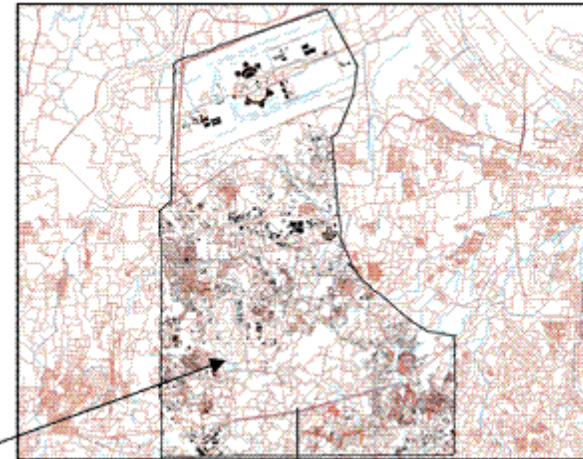
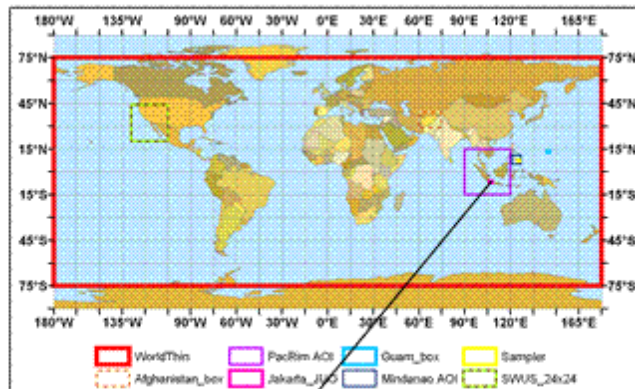
2 million entities (soon)

## Support for Warfighter

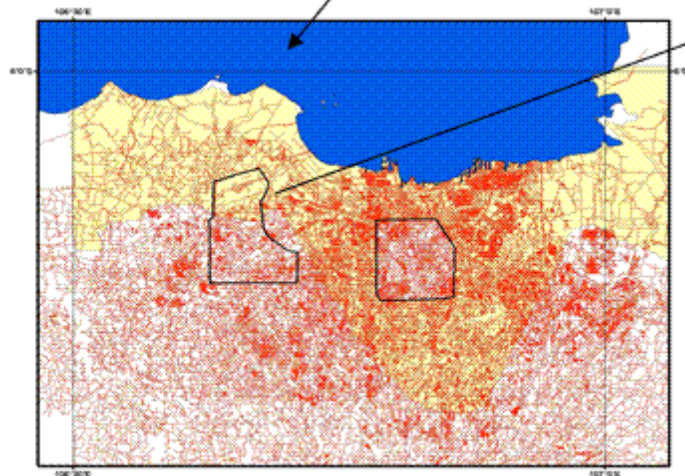
Developing future military tactics  
and doctrine



# Terrain for Typical Experiment



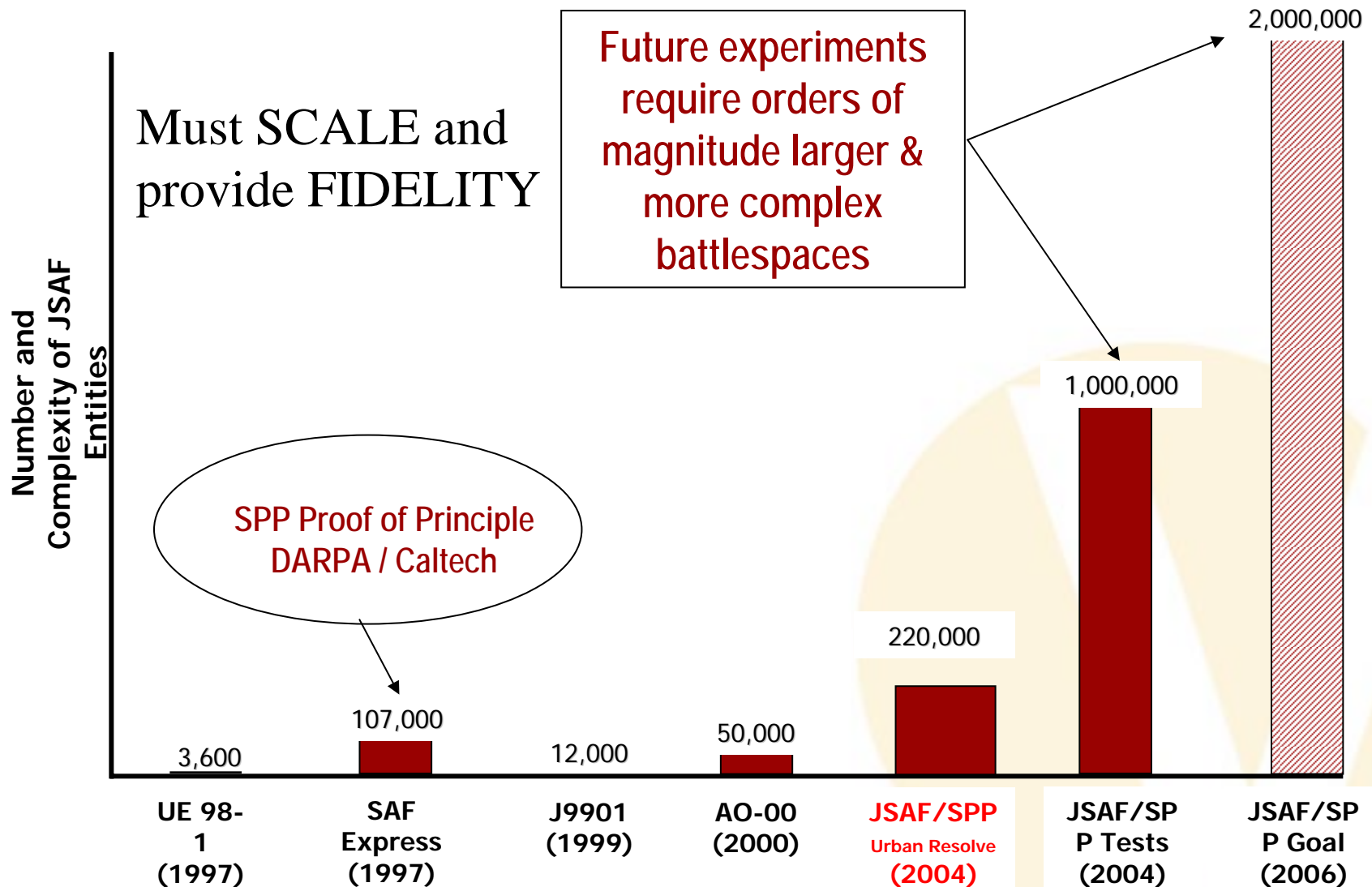
• Geo-Specific Terrain in Western AOI



• 1.8 million Buildings & Structures



• 47,000 Buildings & Structures in AOI



# 3D Stealth View



X: 15858.73 m  
Y: 36892.93 m  
CELL: 22715 (vF0108)

ENTITY COUNT: 1701 (total) / 99 (diguy)

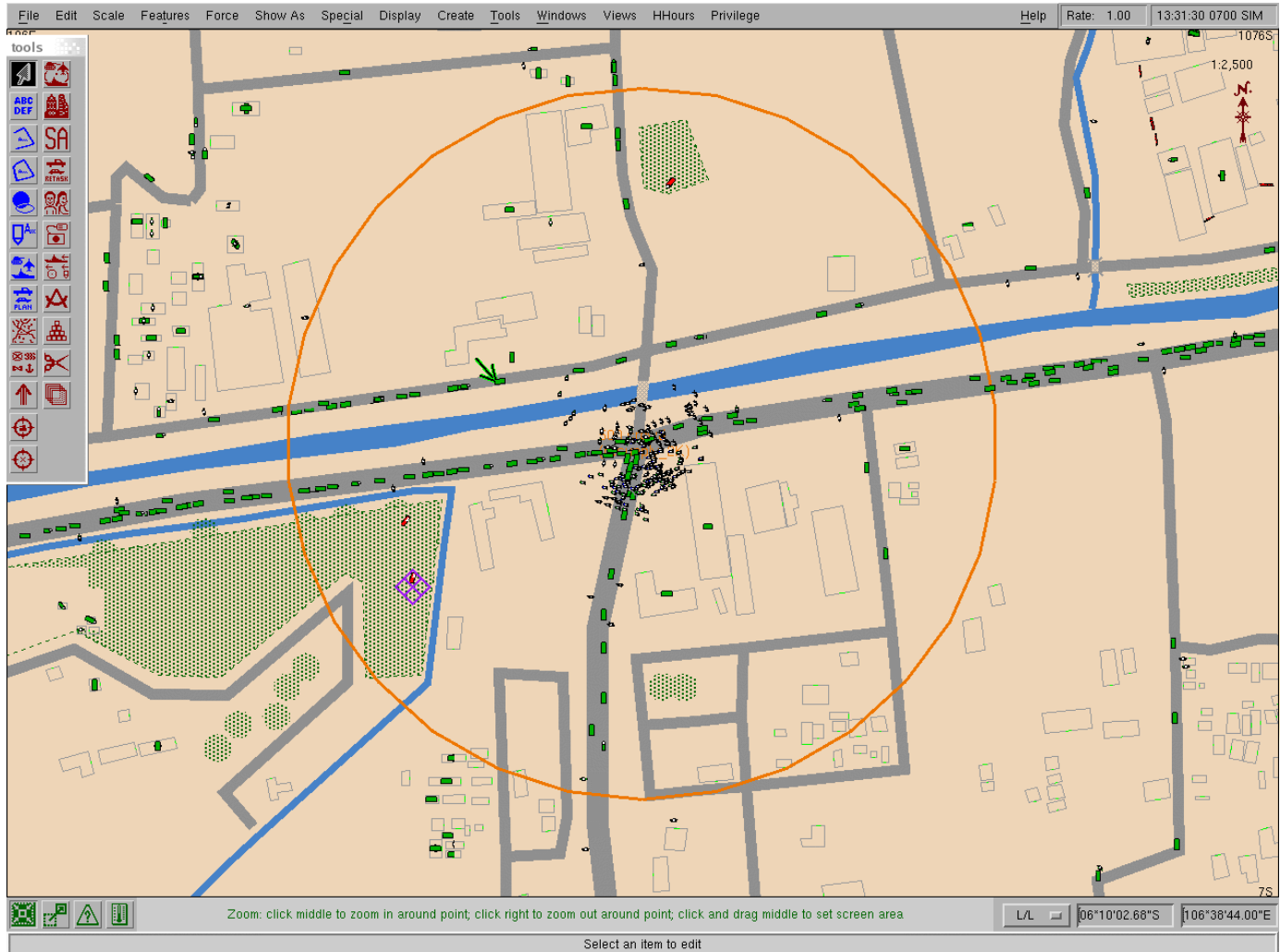
ELEVATION: 22.88 m  
HEADING: 139.30 degs  
MODE: FREE  
FLD VIEW: 40.00 degs  
VIEW RNG: 5000.0 m  
SPEED: 0.00 m/s



The image shows a 3D rendered city street scene. In the foreground, a multi-lane road has several vehicles including blue sedans, a red car, a white pickup truck, and a silver truck. Pedestrians are scattered across the grassy area and sidewalks. The background features a mix of architectural styles, including a prominent yellow building with a curved facade and a tall blue glass skyscraper. The sky is clear and light blue.

Windows taskbar at the bottom shows:  
- Taskbar icons: Start, Internet Explorer, Firefox, VLC, Skype, etc.  
- Taskbar buttons: SI.com - NCAA Football - Miami, OpenScene DPP, ModStealth [USAF / JUO 12.5 / ...], safuser@arachnid: - Shell -  
- System tray: Volume, Network, Safesoft, Date/Time: 07/27/04

# Plan View Display





# Technical Successes



## 1 Million Entities

Clutter and operational

December, 2002

Consistent and stable service scheduled and delivered

Using DREN, both Classified and Unclass

Technology adopted by CENTCOM

Fielded to Baghdad

JSAF Experiments Demonstrated to:

Reps. Ford, Thornberry, and Forbes

Former Speaker Newt Gingrich

Sen Hillary Clinton

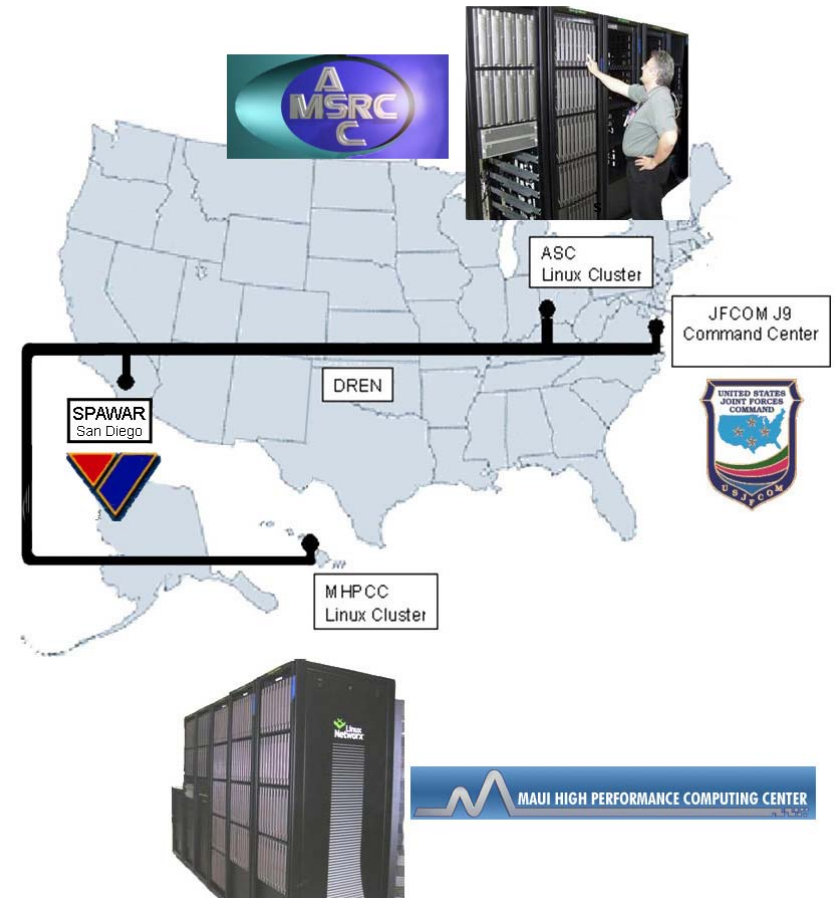
Gen Abizaid

DDR&E

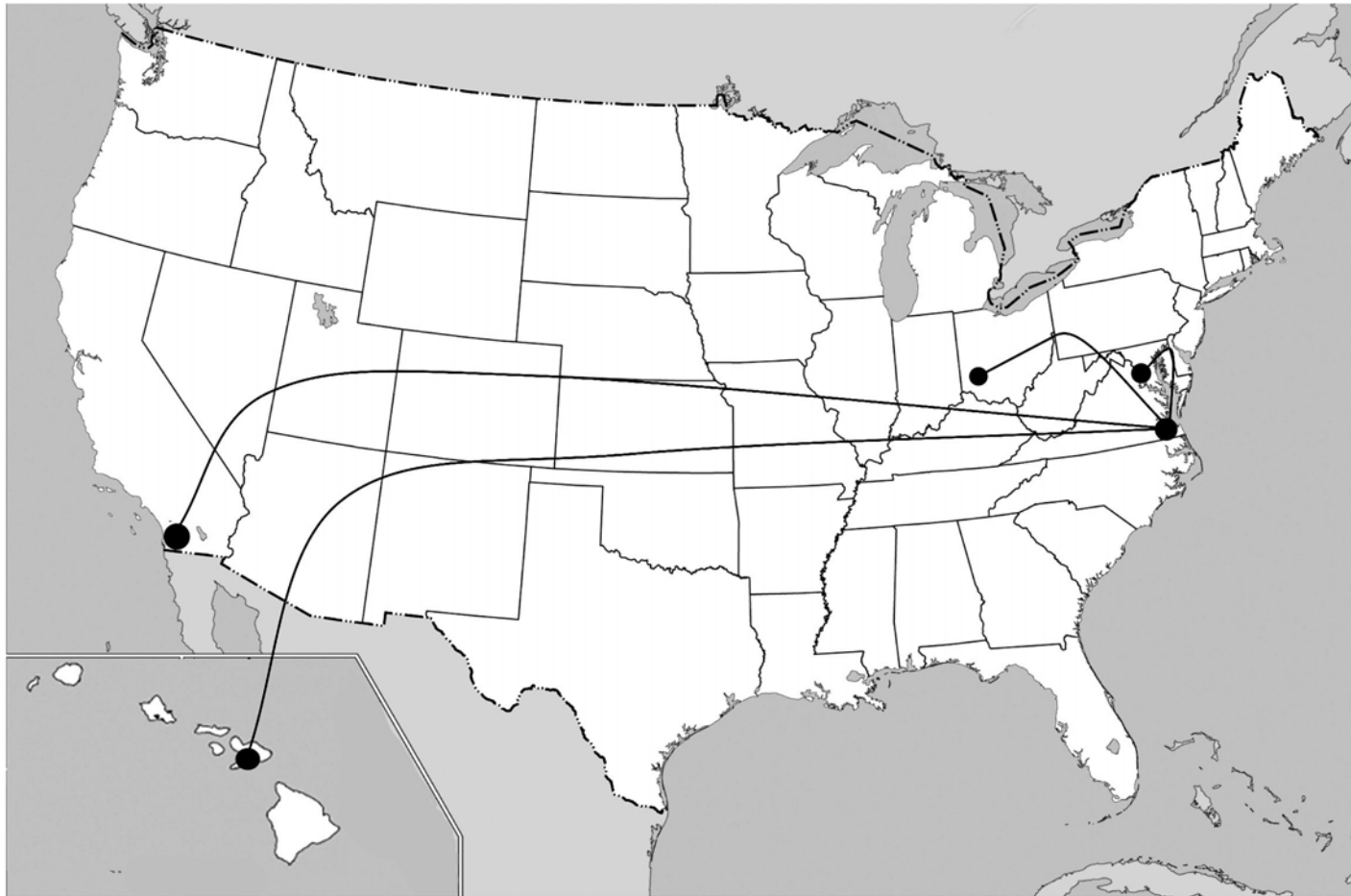
DDS&T



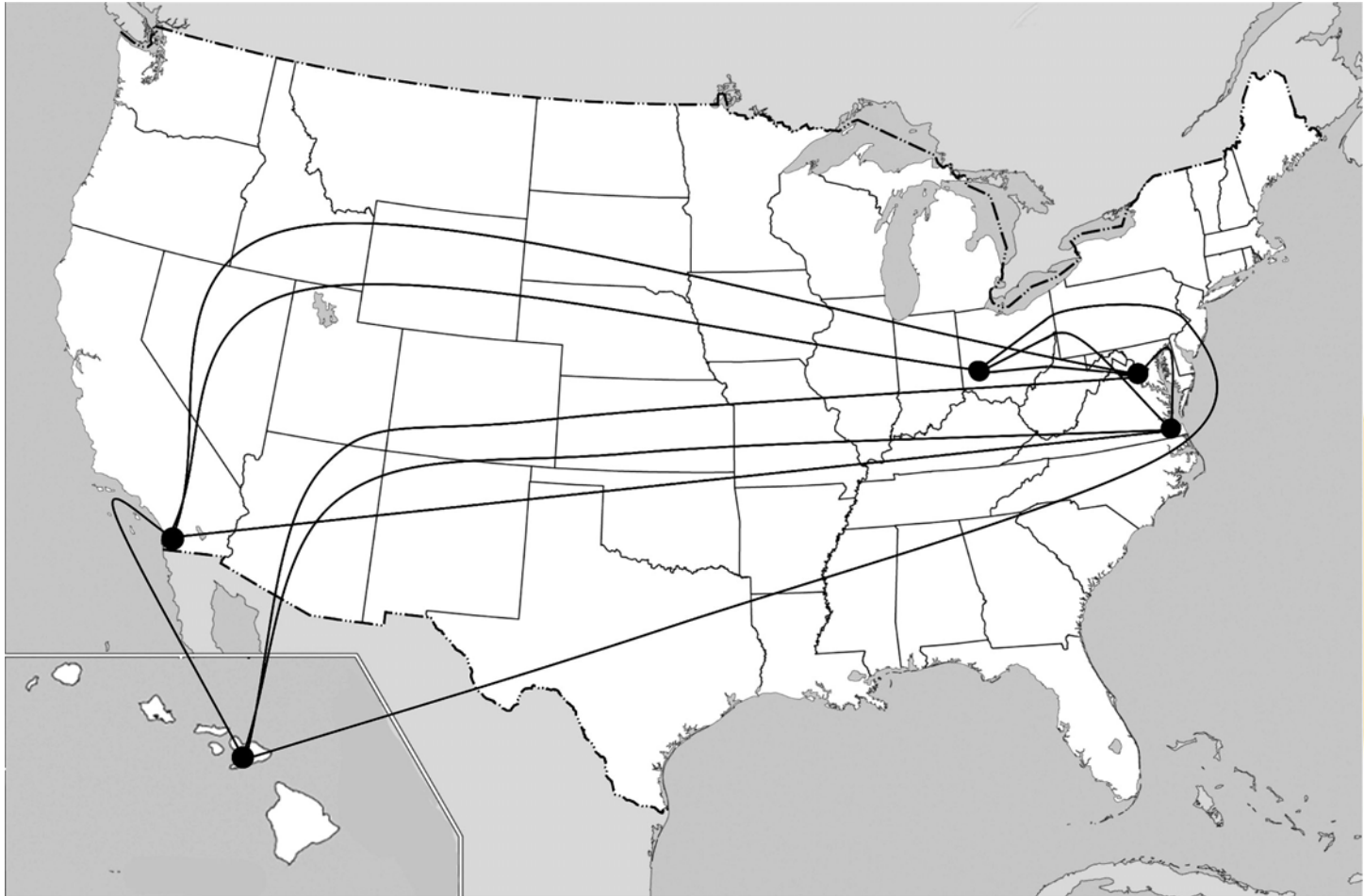
Deployed, spring '04  
MHPCC & ASC-MSRC  
2 Linux Clusters  
24x7 support by HPCMP  
DREN Connectivity  
Users in VA and CA  
Application tolerates  
network latency  
Real-time interactive  
supercomputing



# JFCOM Net Diagram (Notional)



# JFCOM Mesh Diagram (Notional)





## Urban Resolve Participants



JFCOM – Suffolk Virginia (Tidewater Area)

TEC – Ft. Belvoir Virginia

DTRA – Ft. Belvoir Virginia

SPAWAR – San Diego California

US Army – Ft. Knox, Kentucky

US Army – Ft. Leavenworth Kansas

ASC MSRC – WPAFB, Ohio

MHPCC – Maui Hawai'i



## Challenges

**Security Policies**

**Usage model**

**These limit the resources accessible to FMS community**

**Not to say that we don't appreciate ...**

**Stability**

**Support**

**Encouragement**



JFCOM has IMMEDIATE need for more entities (10X)  
Memory on Nodes and in Tertiary Storage very limited

TeraByte a week

Keeping only 20% of current data

Need 10X more entities

Need 10X behavior improvement

Now doing face validity

Need more quantitative, statistical  
approach

Caltech – Dr. Thomas Gottschalk

NPS – Profs Sanchez and Lucas

Data mining efforts now commencing



## Two Key Challenges



Collect the “fire hoses” of data generated by large-scale distributed sensor rich environments

Without interfering with communication

Without interfering with simulator performance

Maximally exploit the collected data efficiently

Without overwhelming users

Without losing critical content

Goal:

Unified distributed logging/analysis infrastructure, helps users and the computing/networking infrastructure managers



# Limitation of the Original System – Does not scale



## Two separate data analysis systems

One for near-real time during the event

Another one for post event processing

### For near-real time

Too much data access over wide-area network

### For post event processing

1-2 weeks to stage data to centralized  
data store

Discards Green entities (80%)

	<i>Compressed Raw Data (Terabyte)</i>	<i>Relational Database (Terabyte)</i>	<i>Relational Database Growth Factor</i>
Urban Reslove Phase I	0.5	0.8	
+ Include green data	0.5	4.0	5X
+ SLAMEM data	1.0	8.0	2X
+ Scale to 1 million entities	10.0	80.0	10X

Data is NOT static during runs, but users need to access

- Logger continuously inserts new data from the simulation

Need distributed query to combine remote data sources

- Distributed logger inserts data into SDG data store at each site

Problems

- Local cache invalid with respect to inserts

- Cannot preposition data to optimize queries

ISI Strategy: explore trade-offs

- Compute on demand for better efficiency

- Compute on insert for faster queries

- Variable fidelity: periodic updates

- Dynamic pre-computation:

  - detect frequent queries



## Handling Distributed Data

Analyze data in place

- Data is generated on distributed nodes

- Leave data where it is generated

Distribute data access so data appears to be at a single site

Take advantage of HPC hardware capabilities

- Large capacity data storage

- High bandwidth network

- Data archival

Exploit JSAF query characteristics

- Limited number of joins

- Counting/aggregation type queries

- Size of data product is several orders

  - of magnitude less than raw data size



## Aggregate(Top) Level

<div style="background-color: red; color: white; padding: 5px;">TARGETS</div> <div style="background-color: blue; color: white; padding: 5px;">AGGREGATES</div>	Civilian School Bus	Civilian SUV	Civilian Medium Car	Civilian Medium Truck	Civilian Small Car	Civilian Large Truck	Civilian Small Truck	Civilian Large Car	Civilian Bus	Civilian Limo	MEL	UAZ469B	BTR80	Total
High Altitude <sup>1</sup>	463	237	234	240	266	389	230	254	266	218	121	4	3	2925
Medium Altitude <sup>1</sup>	12	6	4	7	7	4	8	4	4	5	3	0	0	64
<b>Totals</b>	<b>475</b>	<b>243</b>	<b>238</b>	<b>247</b>	<b>273</b>	<b>393</b>	<b>238</b>	<b>258</b>	<b>270</b>	<b>223</b>	<b>124</b>	<b>4</b>	<b>3</b>	<b>2989</b>

Summarizes sensor contact reports

Positive and negative sensor detections

Displays two dimensional views of the data

Provides three levels of drill-down

Sensor platform type vs. target class

Sensor platforms vs. sensor modes

List of contact reports

## Platform/Mode Level

TARGET TYPE	MEL			Total
<div style="background-color: red; color: white; padding: 5px;">MODE(S)</div> <div style="background-color: blue; color: white; padding: 5px;">PLATFORM</div>	SPOT_IR	STRIP_IR	MTI	
30 Lacrosse*	20	63	0	83
38 Lacrosse*	0	0	27	27
47 Lacrosse*	0	0	11	11
<b>Total</b>	<b>20</b>	<b>63</b>	<b>38</b>	<b>121</b>

Sensor Mode(s): **SPOT\_IR STRIP\_IR MTI**

Raw data has other dimensions of potential interest

Detection status

Time, location

Terrain type, entity density

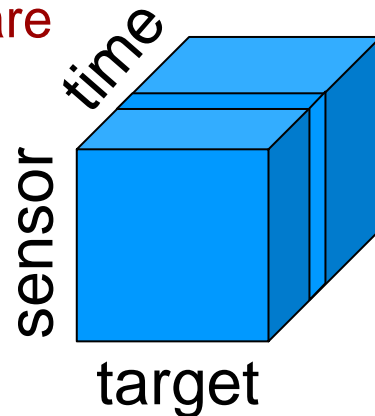
Weather condition

Each dimension can be aggregated at multiple levels

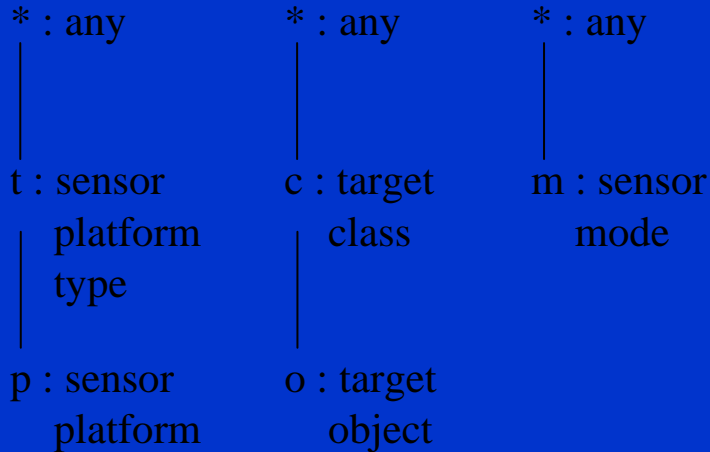
Time: minutes, hours, days

Location: country, grid square

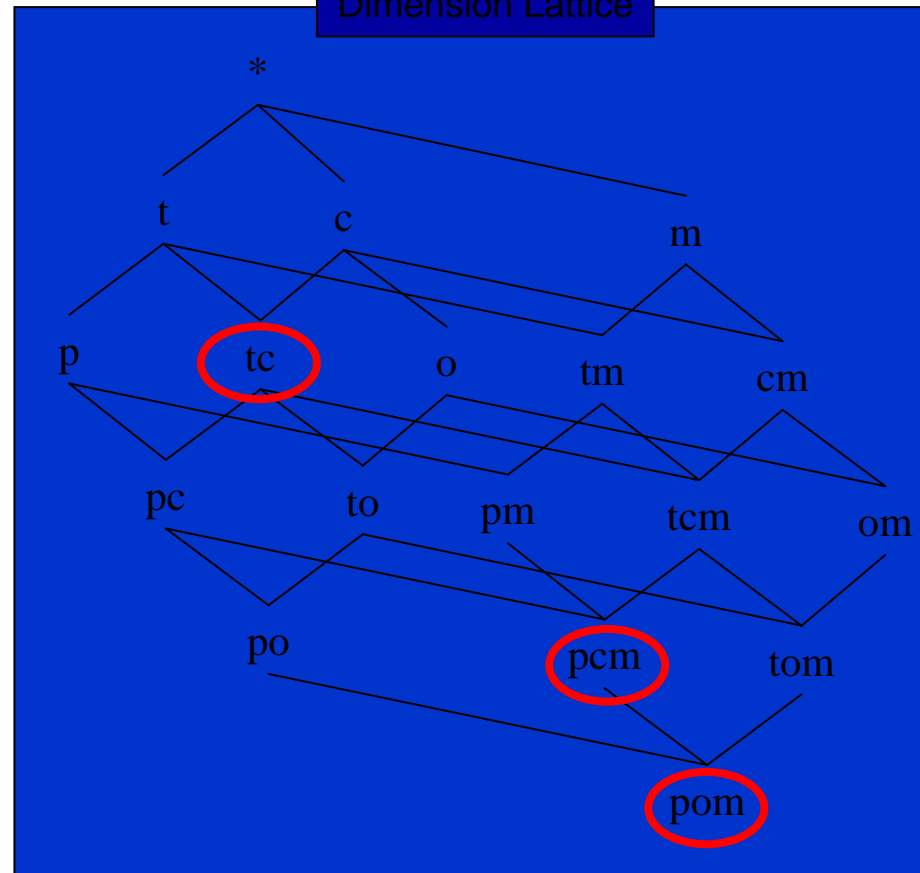
Collapse and expand multiple dimensions for viewing



## Dimensions



## Dimension Lattice



Sensor/Target Scoreboard drill-downs in the context multidimensional analysis

Data classified along 3 dimensions

Drill-down to 3 nodes in the dimensional lattice

# Cube Dimension Editor

**Cube Dimension Editor**

File Tabs View Tools Options

Sensors Targets Detection Status

name	dimension	nodeIndex	valueIndex	lookupString
root	1	312	0	
└─ Munitions	1	313	-454	
└─ Structures	1	313	-458	
└─ Generic	1	314	-457	
└─ Land	1	313	-224	
└─ Indonesia	1	314	-31	
└─ Sweden	1	314	-44	
└─ vehicle_Sweden_CIV_Bus	1	315	37	1-1-205-27-2-0-0
└─ vehicle_Sweden_CIV_Sm_Car	1	315	38	1-1-205-27-3-0-0
└─ vehicle_Sweden_CIV_Bulldozer_Truck	1	315	43	1-1-205-27-1-6-0
└─ vehicle_Sweden_CIV_SUV	1	315	40	1-1-205-27-3-3-0
└─ vehicle_Sweden_CIV_Med_Car	1	315	39	1-1-205-27-3-1-0
└─ vehicle_Sweden_CIV_Liquid_Chemical_C1	1	315	41	1-1-205-27-4-0-0
└─ vehicle_Sweden_CIV_Garbage_Truck	1	315	36	1-1-205-27-1-4-0
└─ vehicle_Sweden_CIV_Concrete_Truck	1	315	42	1-1-205-27-1-5-0
└─ vehicle_Sweden_CIV_Sm_Truck	1	315	34	1-1-205-27-1-0-0
└─ vehicle_Sweden_CIV_Lg_Truck	1	315	35	1-1-205-27-1-2-0
└─ Iran	1	314	-33	

Name: Targets

Description: Types of targets

Column Name: target\_id

Root	Category	Country	Entity
<root=0>	<Undefined=-472>	<Generic=-457>	<vehicle_Launcher=1>
	<Radars=-468>	<US=-453>	<vehicle_SEAsia_107mm_Launcher=3>
	<Ammo Pallets=-465>	<USSR=-378>	<vehicle_SEAsia_57mm_Launcher=4>
	<Structures=-458>	<Iraq=-368>	<vehicle_SEAsia_82mm_Launcher=5>
	<Munitions=-454>	<China=-349>	<vehicle_SEAsia_120mm_Launcher=6>

Scalable Data Grid: a distributed data management application/middleware that effectively:

- Collects and stores high volumes of data at very high data rates from geographically distributed sources

- Accesses, queries and analyzes the distributed data

- Utilizes the distributed computing resources on HPC

- Provides a multidimensional framework for viewing the data

Potential application areas

- Large scale distributed simulations

- Instrumented live training exercises

- High volume instrumented physical experiments

- Virtually any distributed data environment using HPC resources



## New capabilities for JFCOM Joint Experiments

Global scale

Higher fidelity

Orders of magnitude  
more entities

They were the most effective due to:

HPCMP provision of clusters at MHPCC  
and ASC-MSRC

Development under the watchful eye  
of parallel architects

Support by experienced staff at the centers

Assistance of PET personnel (Pratt & Amburn)





**“And now,  
a message from our sponsor.”**



This material is based on research sponsored by the Air Force Research Laboratory under agreement numbers F30602-02-C-0213 and FA8750-05-2-0204. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes, notwithstanding any copyright notation appearing thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.

