Real-Time Regional Gateway Cloud Analytics for Forward Users

[redacted]

BLUF

RTRG:

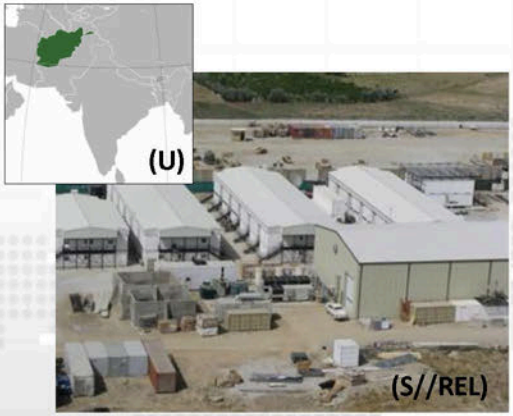
…brings near real-time intelligence to the warfighter

…"grew up" supporting operations in Iraq

…RTRG is now a global architecture

…leveraging the emerging cloud architecture to answer questions we have not been able to do before

RTRG in Afghanistan



(S//REL) Area 82 – Bagram Air Base
Home of RTRG AF1 & gmTote

- 20 days data retention from AFPAK
- 200-600 daily users
- New upgrades include two systems in Kabul

(TS//REL)

Mission Areas: Tracking high-value targets (HVT), Counter-Insurgency (COIN), Counter-IED (CIED)

Organizations Using RTRG:

- CSG* Afghanistan
- U.S. Marine Corps (USMC) 1st and 2nd Radio Battalions
- U.S. Army SIGINT analysts at BCT* level
- U.S. Air Force National Tactical Integration
- Jalalabad Fusion Cell (USMC)
- S2 TOPI
- NSA-G SWAN Counternarcotics Team
- All special operations task forces

(TS//SI//REL)

"RTRG is the most significant SIGINT support to the war fighter in the last decade"
– General David Petraeus

"USSOCOM has enduring and critical needs for the tools and data that RT-RG provides"
- Admiral William McRaven

*CST- Cryptologic Support Team *BCT – Brigade Combat Team *CSG - Cryptologic Support Group

CSG Afghanistan Statistics

- In 2011, RTRG in Afghanistan
  - Played a key role in 90% of all SIGINT developed operations
  - Leading to 2270 capture/kill operations
  - 6534 enemies killed in action
  - 1117 detainees

RTRG in the Gulf and Horn of Africa

Monitoring Iranian Navy (IRIN) in Straits of Hormuz

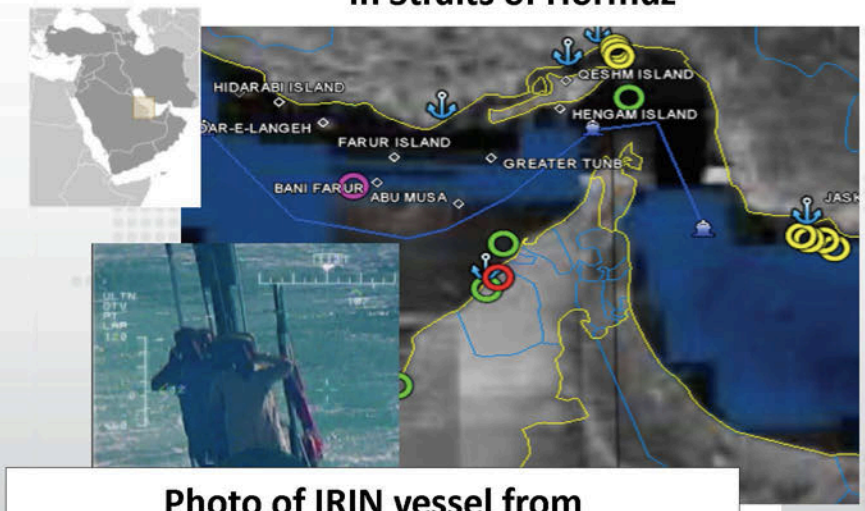


Photo of IRIN vessel from US Navy aircraft, located by RTRG

Supporting CENTCOM Maritime (NAVCENT)

Navy Information Operations Command- Bahrain (NIOC-B)

Missions supported: (TS//SI//REL)

- Iran, Yemen, Persian Gulf
- Recent successes include monitoring of Iranian naval assets

RTRG Afloat on subsurface platform USS Georgia (SSGN-729)

Missions supported: (TS//SI//REL)

- Horn of Africa: In first week of mission, system received 31 million GSM events, leading to 10 high-value target voice ID, and 90 tactical tip-offs

RTRG Global Operations

US-2 – Denver
Global Maritime & ELINT

US-1 Fort Meade
Mission Assurance

IQ2 Iraq*
COIN, CIED

AF-1,5 Bagram & Kabul
COIN,CIED

USS Georgia &
USS Florida

US-8 NSA-Texas
Counternarcotics, Maritime,
Mexico & SOUTHCOM Support

US-3 JFCOM
Joint Forces Command

GE-3 Germany**
AFPAK
Continuity-of-Operations

BH-1 NIOC Bahrain**
AFRICOM, EUCOM,
PACOM

KO-1,2 USFK Pangyo**
PACOM & North Korea

* In draw-down

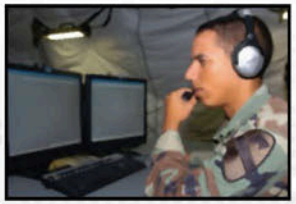
** ECC - NSA European Technical Center USFK – U.S. Forces, Korea NIOC – Navy Information Operations Command

Outline

- RTRG Mission Overview
- RTRG System: Today and Tomorrow
- Target-Centric and Network-Centric Cloud Analytics
- Future Work

Current RTRG Architecture

RTRG System



Supporting Tactical Users

Forward Data Centers

Goldminer — Metadata Search

GeoT — Geospatial Tools

Agent Logic — Alerting Tools

Sharkfinn — Selector Enrichment

SKS — Report & Doc Manager

Panopticon — Target Management

Services Layer: Web services, authentication, auditing

Publish and Subscribe Messaging

Oracle relational database & dimensional data model

Ingest and Enrichment Pipeline:
Flexible, high-speed architecture for parser and data processors

Data Feeds: *DRT* *JUGGERNAUT* *TALIS/MATTERHORN* *KL* *VOICESAIL* *RETURNSPRING*
CIDNE *LOPERS* *AIRHANDLER* *LYCANTHROPE* *.. and a growing list of others**

(DNR, DNI collect, tipping and reporting)

A successful architecture for several years. Demand for more data feeds, longer retention, and data-intensive analytics has driven RTRG to seek new solutions

*Based on Afghanistan RTRG data flow

RTRG Data Challenges

Current Challenges

- Data Storage & Retention
  - "Patterns of Life" analysis needs require 6+ months of data from world-wide collection
  - A typical system has capacity for only 4-6 weeks of regional data (~90% user queries are within seven days of "now")
- Data Use & Computation
  - Analytic processes should make maximum use of all available data to find small signals
  - Relational databases are unsuited to sophisticated analytics such as correlation and matching
- Data & Technology Heterogeneity
  - New types of data must be added to the system continually
  - With traditional databases, schema modifications are difficult
  - Exotic data management solutions are difficult to adopt due to limited expertise

Cloud Architectures for Analytics

Google

hadoop

Apache

Cloudbase
Scalable BigTable Implementation with Security

Emerging NSA Cloud Reference Architecture is well-suited for developing analytics on intelligence data

***Scalable*:** **Distributed file systems and databases are built on clusters of commodity hardware, leveraging open source projects and industrial solutions**

***Computable*:** **The MapReduce programming model simplifies writing efficient parallel computations that operate over large volumes of data**

***Flexible*:** **Cloud technologies enable flexible schema and leverage large open-source efforts**

Scalability & Computation

Data Challenges in AFPAK

Current RTRG (AF1)

- Current database is 27 terabytes (TB)
- Retention is ~30 days

Future Cloud enabled system

- Even a modest cloud system (3 rack) for storage will be at least 125 TB of storage
- 5x increase in available space
- Actual retention improvement depends on how the space resources are allocated

Cloud supports more data feeds & more days of historical data

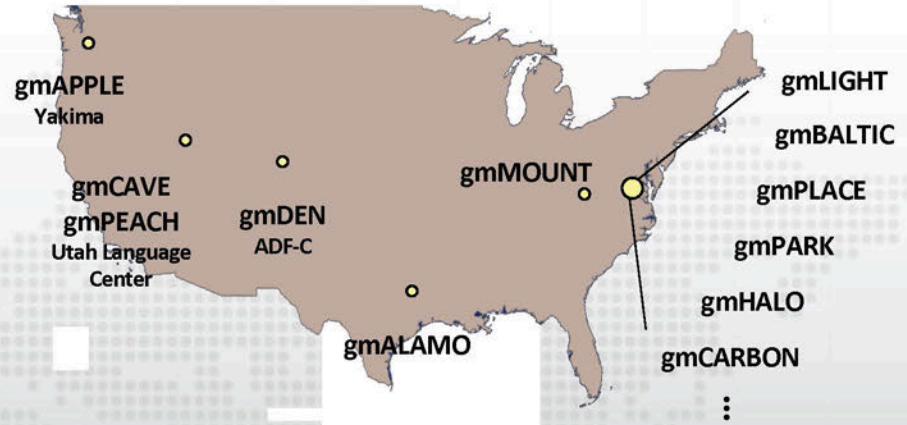
Analytic Challenges from Iraq & AFPAK

- Many analytics used by RTRG are based on R6 SORTINGLEAD event summaries
- Event summaries were originally created on relational databases
- Collection increased dramatically, and a mapreduce implementation was needed
- For new analytics with present day collection volumes, a practical parallel execution model is crucial

Cloud supports large-scale analytics

NSA Cloud Computing Enterprise

- **5-12 racks commodity hardware**
  - 150+ data nodes
  - 16 GB RAM each
- **Apache Hadoop**
  - 100s of terabytes of storage
  - Stores 10s to 100s of billions of events
- **NSA Cloudbase**
  - 100s of nodes serving BigTable implementation
  - Stores 100s of billions of entries

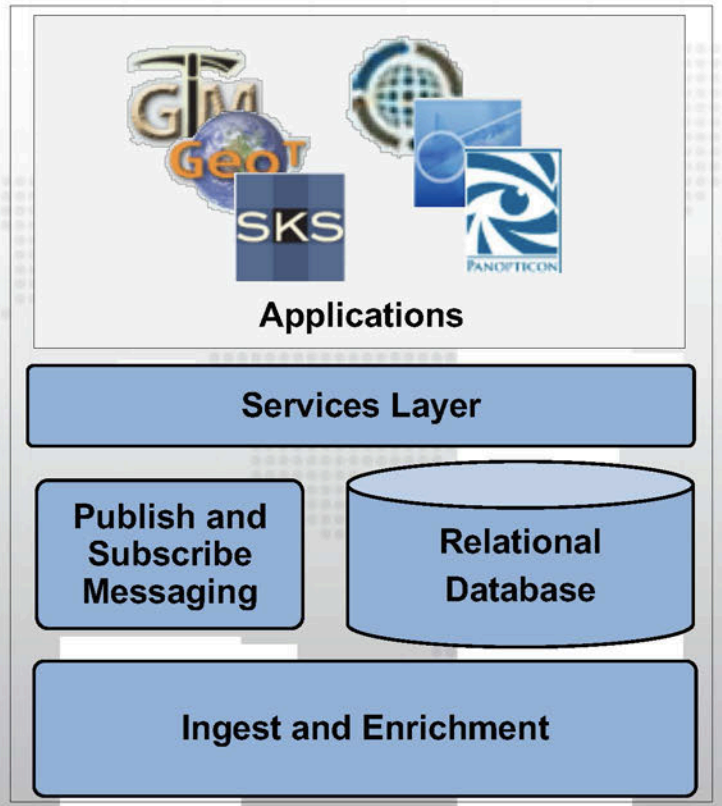


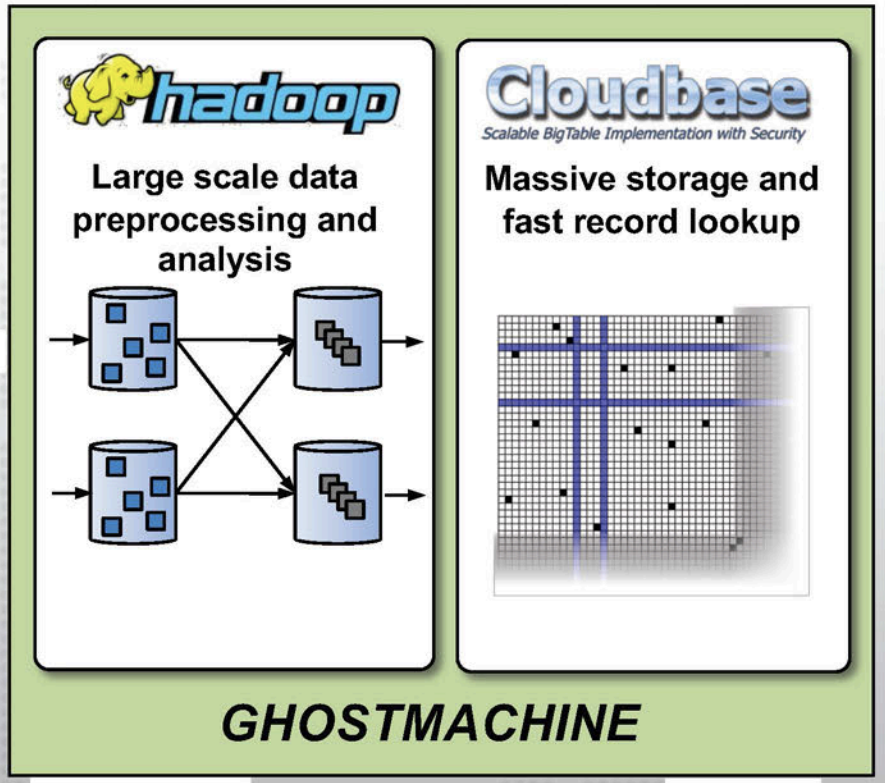
GHOSTMACHINE is a data-intensive cloud system with many fielded instances worldwide

Hadoop clusters have been demonstrated with 10s of petabytes and over 10,000 cores

Map does not include all GHOSTMACHINE/SiteStore systems

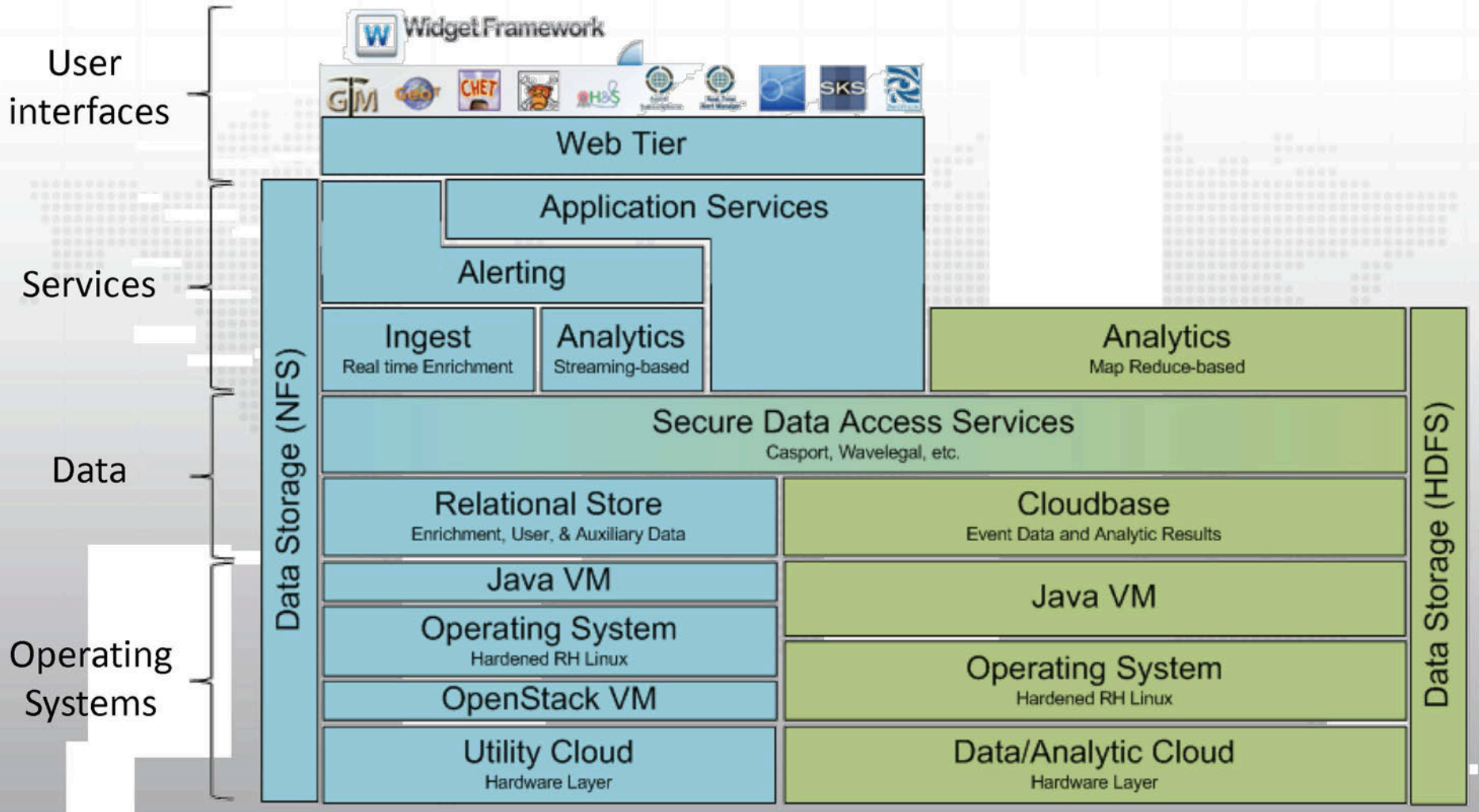
Today's "Hybrid" Architecture





RTRG and GHOSTMACHINE systems are paired with one another: MapReduce analytic results are fed back to RTRG relational database

Tomorrow's Architecture

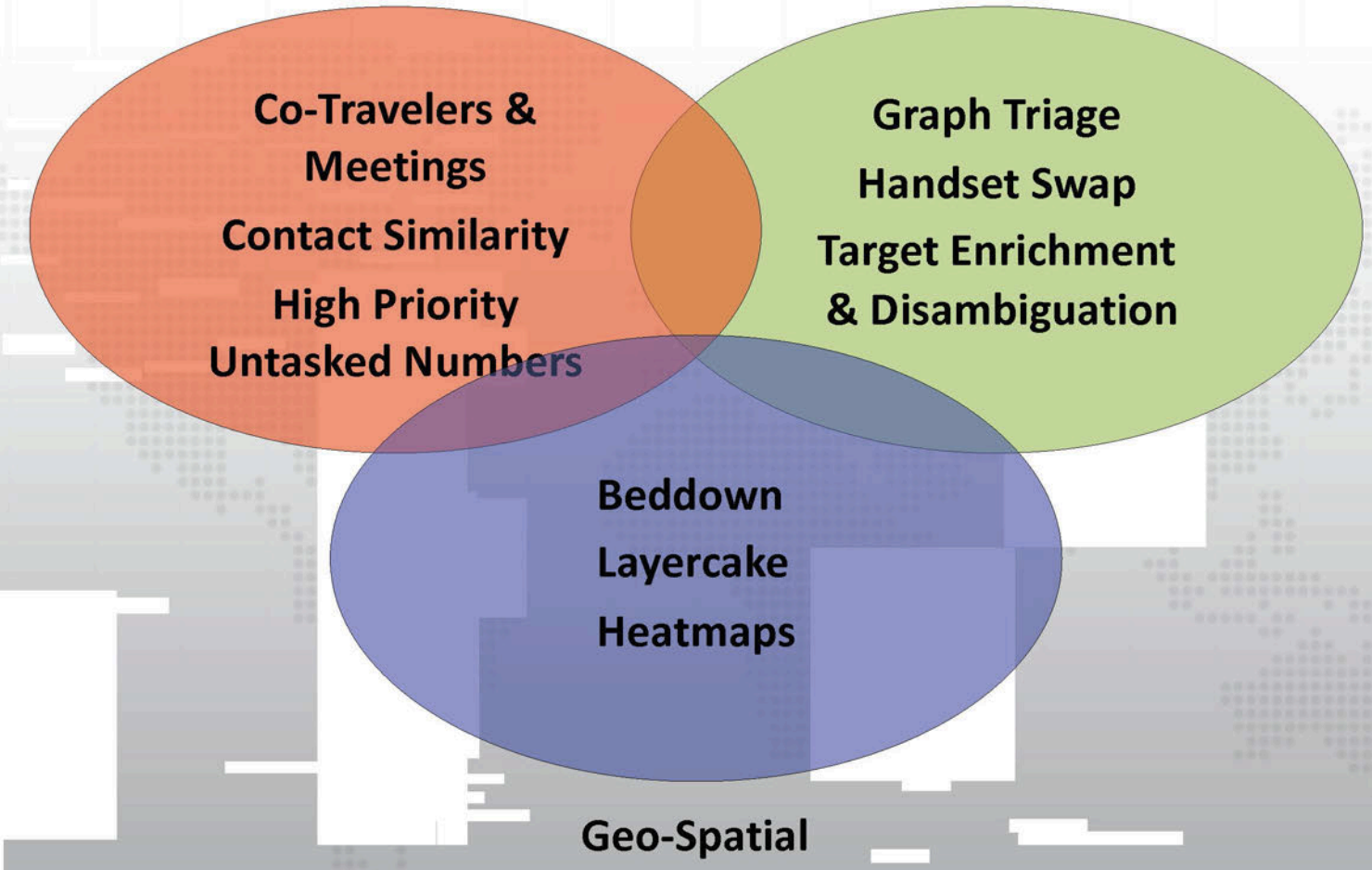


The Cloud will bring new data-intensive capabilities, support existing missions, and align RTRG installations with emerging NSA and IC standards

Outline

- RTRG Mission Overview
- RTRG System: Today and Tomorrow
- Target-Centric and Network-Centric Cloud Analytics
- Future Work

Analytics Overview



The data-intensive computing capabilities in the system enables a set of graph/network analytics and target development analytics

Meeting Target Development Challenges

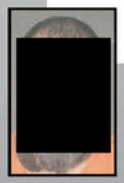
Past Analysts manually queried multiple, independent repositories, aggregating results in Excel, taking hours or work for search and refinement

Now RT-RG provides a streamlined, integrated workflow saving analyst effort

Tips

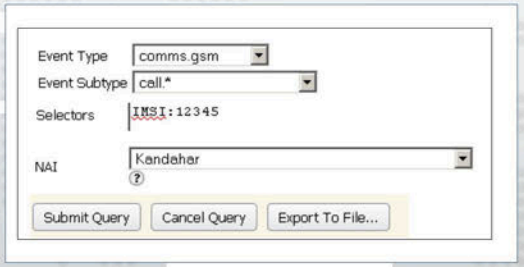
Selectors

GIST Report

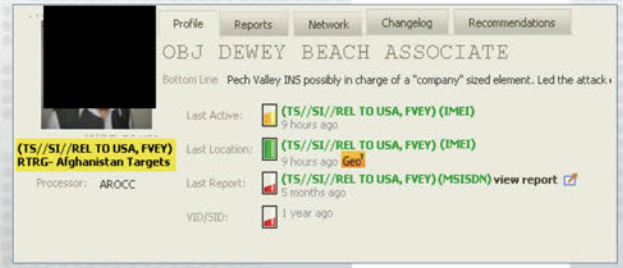


Detainee Reports

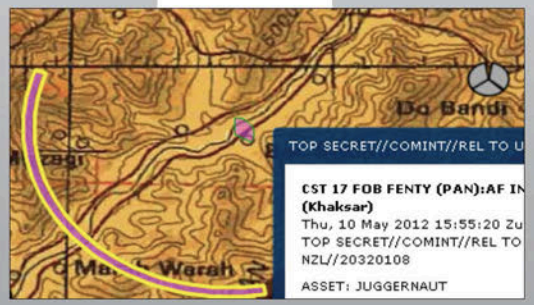
Target Development



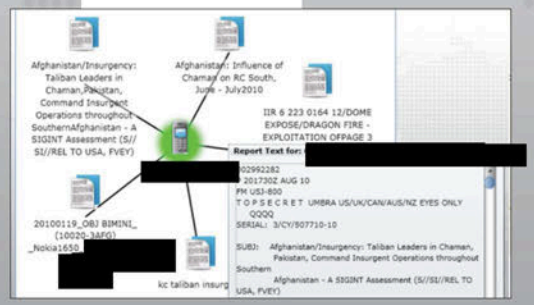
Metadata Search



Collaboration



Geospatial Alerting



Enrichment

Outcomes

Geolocation

Capture

Target Development with Meetings

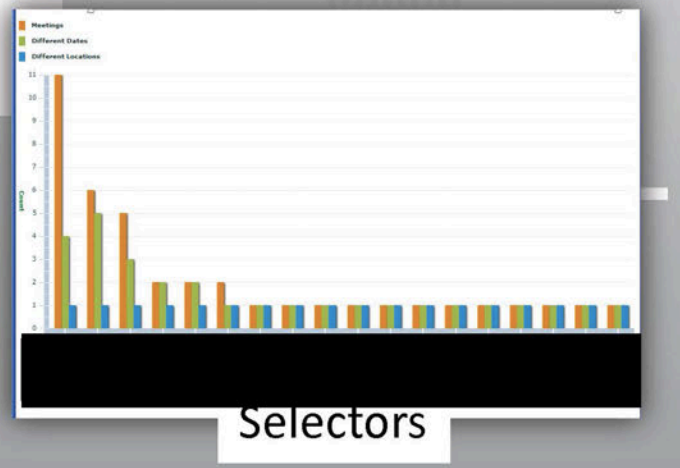
Who is at the same UCELLID at the same time?

Manual Process

- Take your selector and query for every unique location he has been and at what time
- Query for other selectors who have been at the same places at the same times (impossible or painful)
- OR compare to another known set of selectors to find overlap (excel / ArcGIS / JEMA) (limiting to what you know)
- Summary statistics on the matching IMSIs using excel or ArcGIS

Cloud Process

- Pre-calculates all UCELLID overlaps between tasked selectors
- Simply query your selector in cloud-generated QFD and view summary statistics



Target Development with Co-Travelers*

Is there a pair traveling together?

Manual Process

- You could use the same manual process from Meetings, however, this would not find co-travelers on different networks
- Manual comparison of pairs of **known** selectors is possible with ArcGIS or similar spatial tools - You must know the pairs up front

Cloud Process

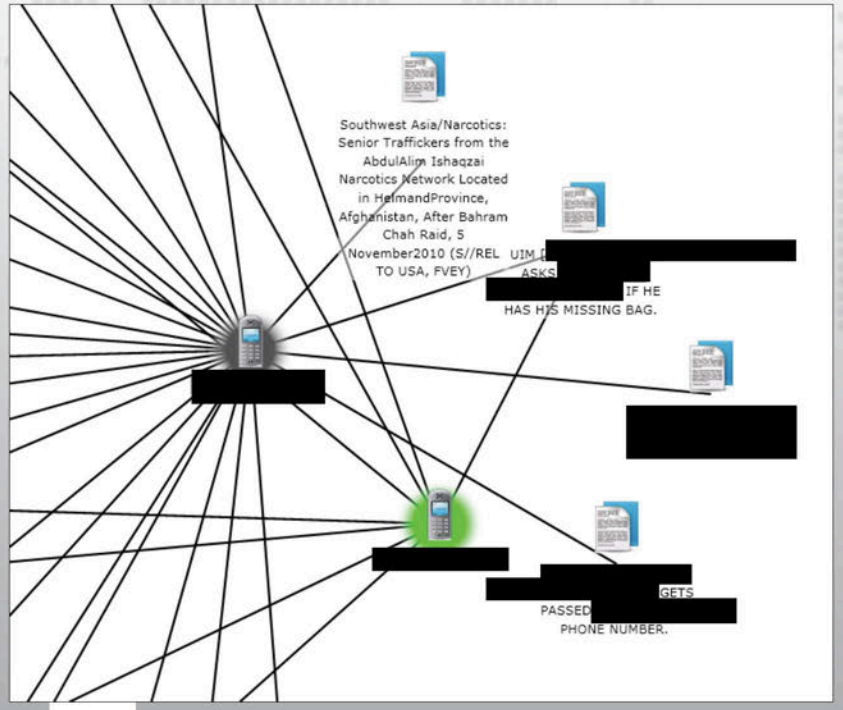
- Measures miles-per-hour (MPH) between tasked selectors as they move around.
- Low average MPH = co-traveling.
- Simply query for your selector to view statistics on average MPH, days calculated, etc.

*Also known as "Sidekicks"

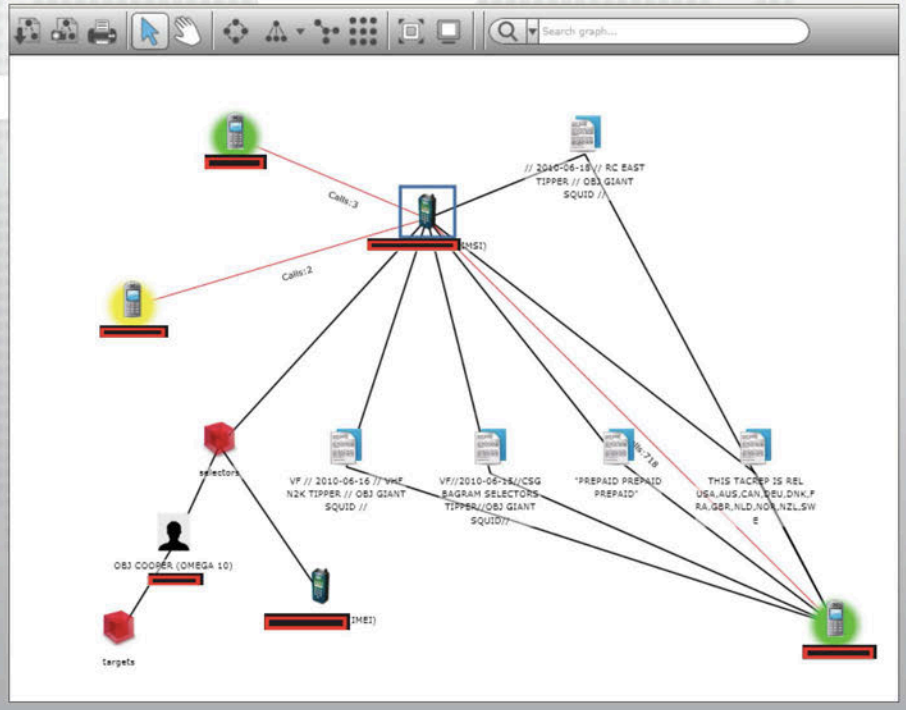
Meeting Network Challenges

Past Manually query multiple repositories and build network with Analyst Notebook (ANB) - amount of labor can be prohibitive

Now RT-RG tools exist for contact chaining for selector-to-selector & selector-to-report graphs, with more analytics and tools to come



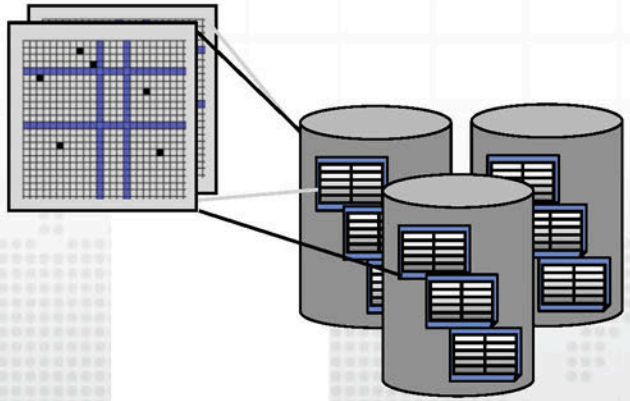
Selector-to-Report Graph from Enrichment



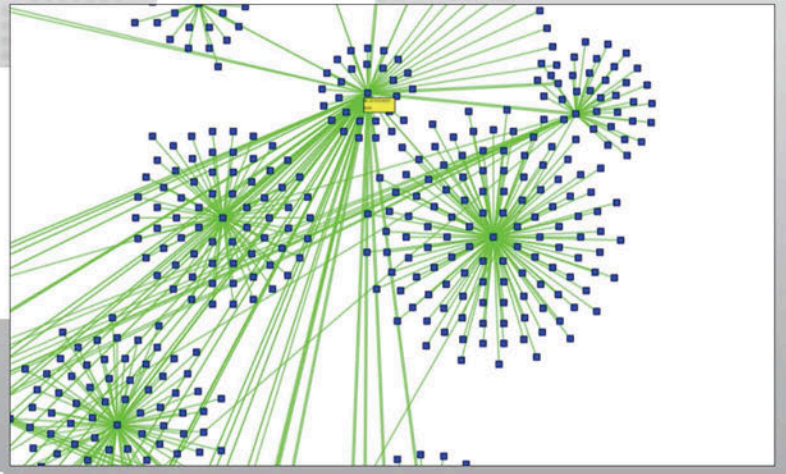
MAINWAY graph in RTRG UI

Graph Analysis with Furious Chainsaw

- **Graph representation is natural for DNR**
- **Result is Furious Chainsaw**
  - **Prototype on Cloudbase**
  - **Now supports contact chains and trends**
  - **Will support other graph analytics in the future**
- **Triage capability for forward users to complement Enterprise databases**
- **Enables chaining *and* other analytics, provides foundation for graph algorithms**



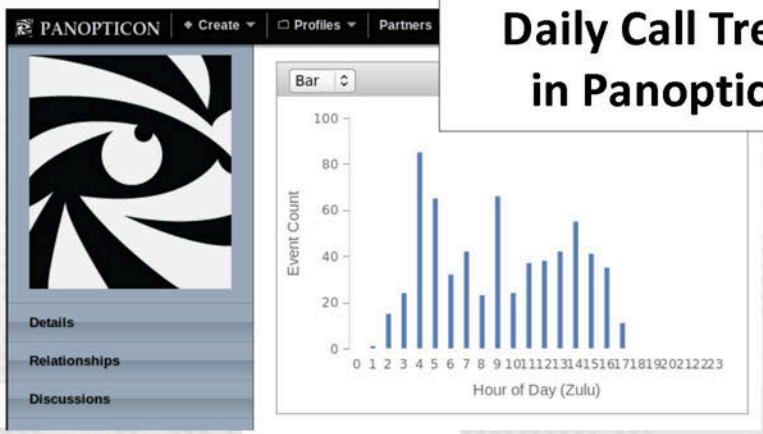
Metadata matrix in Cloudbase supports fast graph traversal



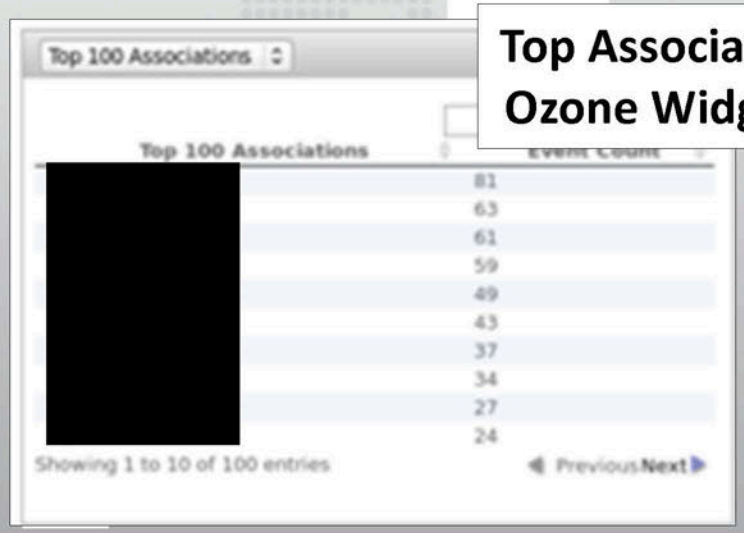
Graph View in Renoir – but many other analytics are possible

Graph Triage: Multiple Views

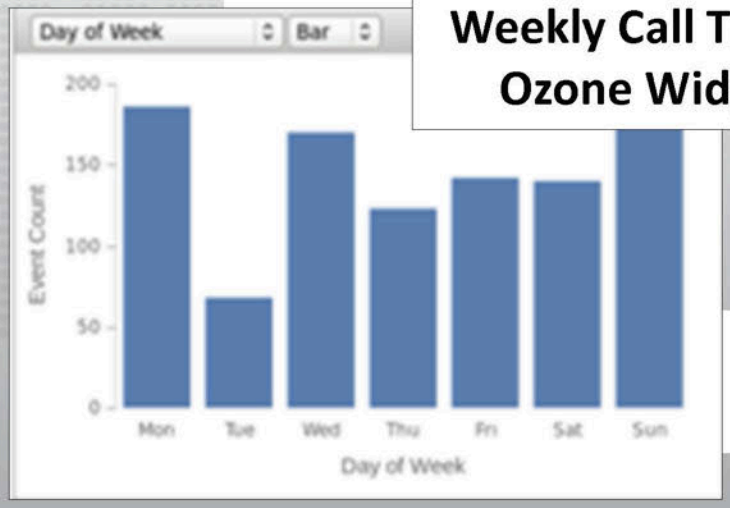
Contact Neighborhood Ozone Widget



Daily Call Trends in Panopticon



Top Associates Ozone Widget



Weekly Call Trends Ozone Widget

DNR graphs in Furious Chainsaw tables in Cloudbase support a wide range of fast queries and analytics

Unstructured Data Exploitation

SHARKFINN

- Selector extraction, normalization, and enrichment
- Flexible free-text query interface
- Graph, text, and spreadsheet output formats

SKS Structured Knowledge Space

- Entity extraction (people, organizations, times, geos)
- Keyword, faceted, and people search
- Document clustering
- Arabic name expansion

sharkfury

- Integrated data ingest using Niagara Files (NiFi)
- SharkQuery: search by selector, entity, location, and keyword
- SharkDocs: query, sharing, and collaboration on user uploaded documents
- Visualization of results in query overview, table, graph, and map
- Cloudbase and HDFS for scalable text analytics platform

Target Development - SharkFinn / SKS

SKS

"OBJ SMITHERS" AND OBJECTIVES:OBJ TOPPER Previous 30 Days

1. Perform keyword search with seed OBJ on report library

OBJECTIVES
OBJ SMITHERS (19)
OBJ TOPPER (19)
OBJ SPRINGFIELD (12)
OBJ APU (6)
OBJ GROUNDSMAN WILLY (6)
OBJ WIGGUM (4)
OBJ MAYFIELD (3)
OBJ NEVIS (1)
OBJ TRAJAN (1)

2. Finding co-occurring objectives

SELECTORS

[redacted]

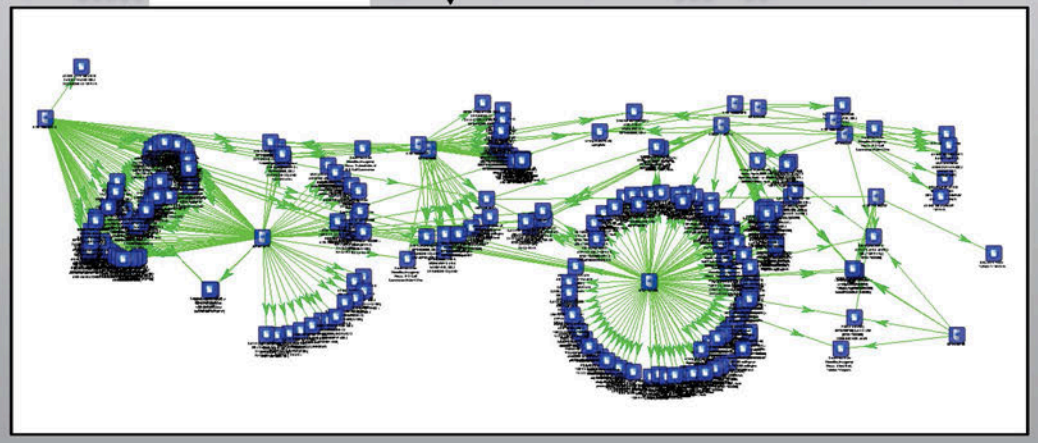
3. Selector co-occurrence

4. Additional Search & Filtering

	SELECTOR
1	[redacted]
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	

5. Correlate found selectors with Octave/UTT and Panopticon

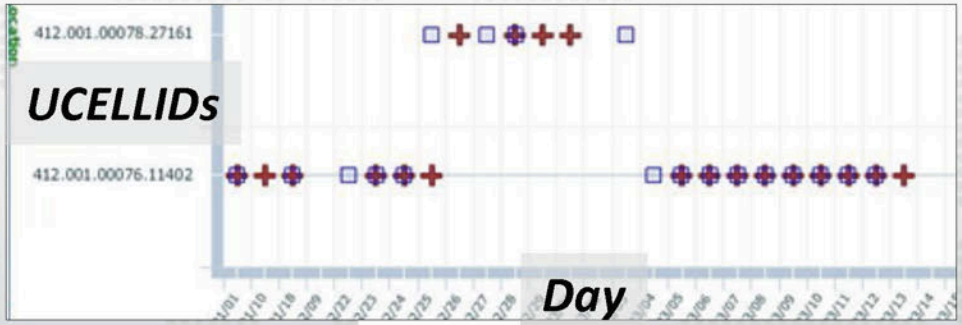
6. Export selector & Report Graph



Meeting Geolocation Challenges

Past Analysts manually correlated locations using map viewers or spreadsheets, aggregating data from multiple sources

Now Analytics and alerts push target information by subscription



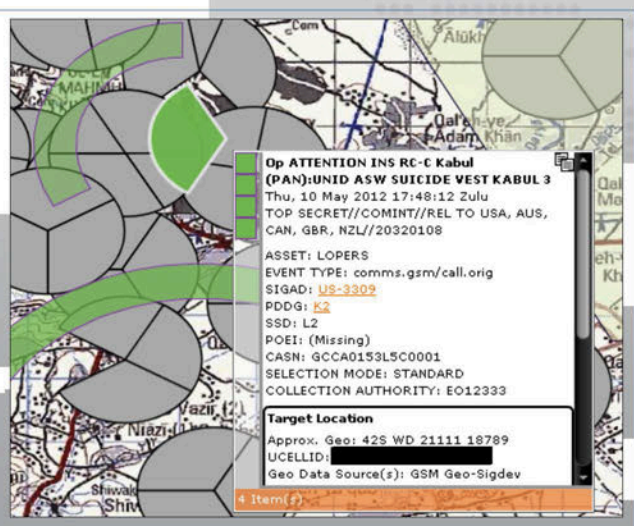
First Event

Last Event

RT-RG pattern of life analytic for target location cues

Analysts are notified by alerts, based on:

- **Geospatial NAI***
- **Tasking status**
- **SMS content**
- **Selectors, callsigns, frequencies**
- **…**



* NAI = Named Area of Interest

Target Geolocation with Bed Down

Find the most consistent location of the day's first/last event

First Events + Last Events = Bed Down

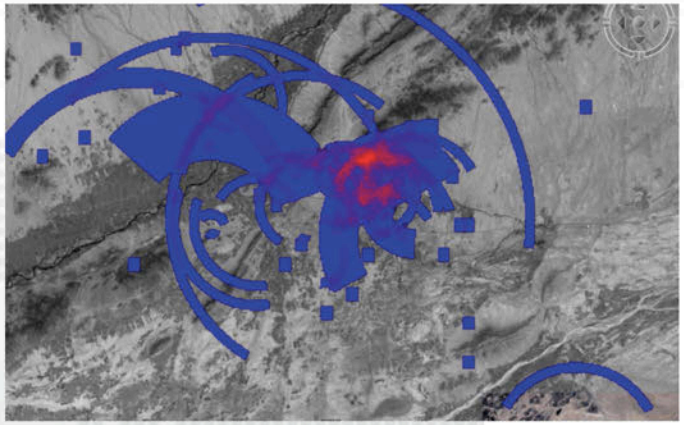
Manual Process – One Selector At a Time

- Query all events for your selector.
- Mark first and last events manually.
- OR
- Enter in a tool like CheekyMonkey to view gaps in activity.
- Slow process to do one selector at a time

Cloud Process- All Tasked Selectors

- Pre-calculates first and last events in local time for ALL selectors.
- Will calculate estimated Bed Down at query time.
- Can query multiple selectors in seconds, find common overlap.

Target Geolocation with LayerCake



LayerCake – find geospatial overlap of a set of targets.

Manual Process

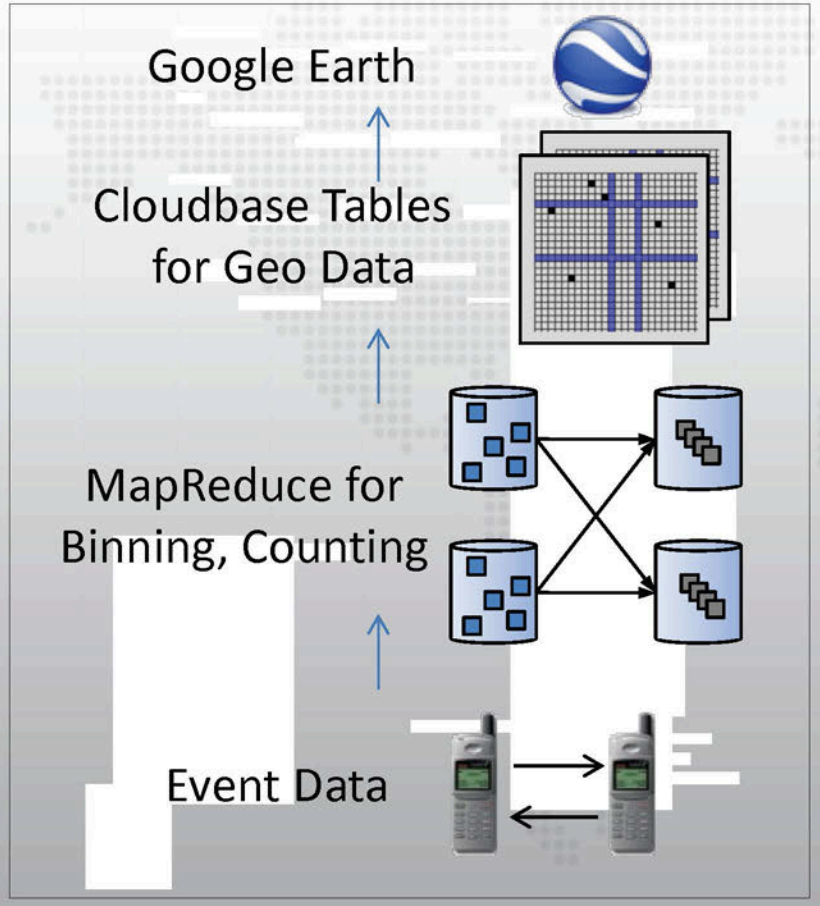
- Query all events for your selectors.
- Display events spatially on mapping software (impossible to view polygon overlaps).
- Rasterize and do "raster math" to determine max overlap. (very complex and expensive task in most GIS tools).

Cloud Process

- Pre-calculates unique locations visited for ALL selectors.
- Raster heat maps drawn at query time.

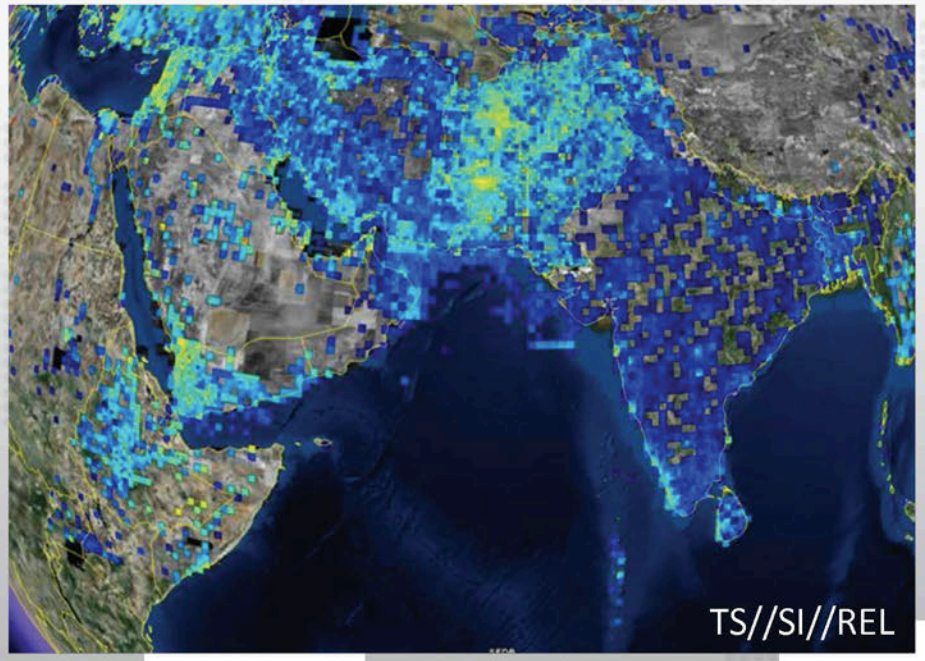
Spatial Visualization of Data

Using MapReduce, analytics can count and aggregate over large number of daily events to give a multi-resolution spatial visualization of the data



Analytic Data Flow

Synoptic View



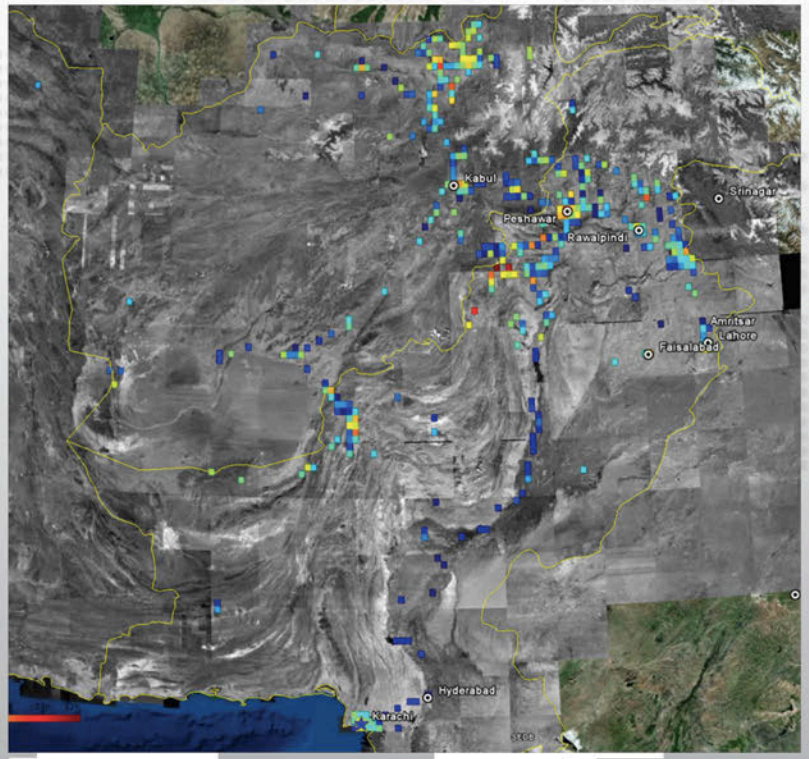
Call volume map of RTRG (AF1) collection
(A day in Dec, 2011)

Fewer DNR events — Fewer events

Spatial Visualization of Data

Cloud enabled analytic allows viewing the spatial data at many levels-of-detail

"Meso-scale" View



Heatmap of activity of a set of targets

Detail View – "Mashup" with heatmap

Using Cybertrans translation service on SMS messages, integrated with heatmaps

Fewer calls — More calls

Outline

- RTRG Mission Overview
- RTRG System: Today and Tomorrow
- Target-Centric and Network-Centric Cloud Analytics
- Future Work

RTRG Tomorrow

- Improved DNI capabilities; focus on convergence
- Integrating active SIGINT capabilities
- Increased CT and expeditionary capabilities
- Better tools for faster analytic development
- Incorporation of content analysis and HLT capabilities
- Improved integration between target and population analytics

RTRG Planned Cloud Instances

gmSeminole
NSA Georgia

gmGulf
NIOC-Bahrain

gmZilla
Kabul

Summary

- **RTRG has been a successful regional data store and exploitation system for COIN, CIED and other missions**
- **Moving to NSA Cloud infrastructure**
  - **More historical data**
  - **Deeper analysis using parallel programs**
  - **Allows for more flexible deployments to IC, DoD service installations**
- **Continuing to support advanced analytics for current and future operations**