Test and Evaluation/Science and Technology Program

Rapid Data Analyzer for NST (RDAN)

2014 Test Technology Workshop

November 5-6, 2014

Mr. Andrew Shaffer (Technical Lead)
Applied Research Laboratory, The Pennsylvania State University

Mr. Gilbert Torres (Government Lead)
Net-Centric Systems Test Executing Agent

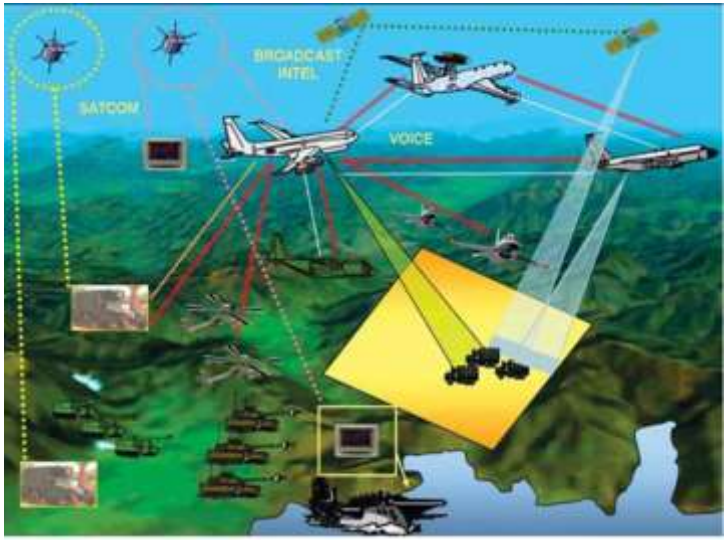
TRMC T&E/S&T Acknowledgement

This project is funded by the Test Resource Management Center (TRMC) Test and Evaluation/Science & Technology (T&E/S&T) Program through the U.S. Army Program Executive Office for Simulation, Training, and Instrumentation (PEO STRI) under Contract No. W900KK-13-C-0015.

Outline

- T&E Need
- System Overview
- Cloud Computing Background
- Functional Diagram
- Query & Analysis Tools
- Key Science and Technology Innovations
- Potential Use Cases
- Summary and Future Work

RDAN T&E Need
Rapid Analysis of High Volume Data ("Data to Decision")



Modern distributed T&E events generate large amounts of unstructured data that are hard to analyze

Slow response time from Data to Decision

Diverse data formats collected using separate T&E systems

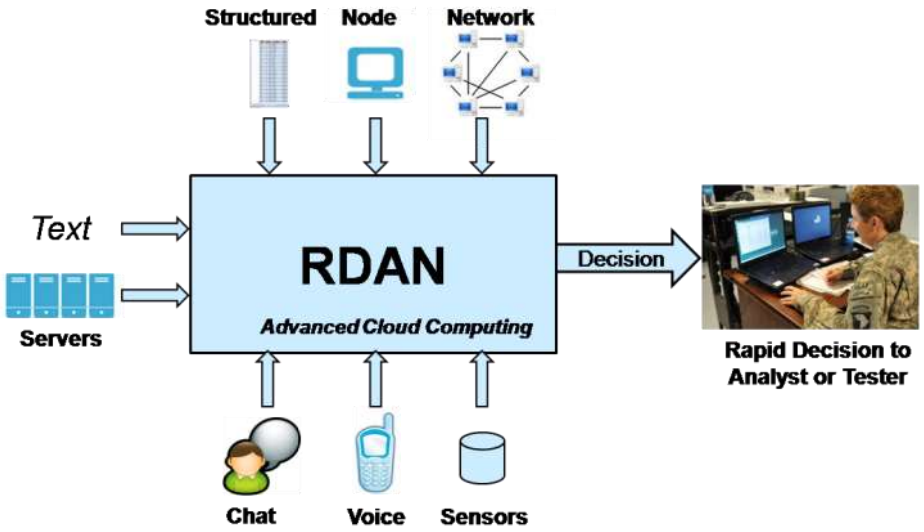
Current analysis techniques

Data to Decision extremely slow (hrs/days/wks)

- **High Volume Data Collection**
  - Unable to efficiently collect and analyze large unstructured data (i.e. text, voice, chat) and structured data across multiple sources and environments
  - Lack of capability to quickly review past historical data limits value of collected test records
- **Adaptability and Scalability**
  - Difficult to scale T&E systems as data grows

- **Analysis Tools**
  - Data is often manually processed with slow response time from Data to Decision
  - Manual processing introduces human error
  - T&E systems lack automated tools to analyze large volumes of unstructured text, voice, and structured data
  - T&E systems lack ability to perform deep analysis on test event as it occurs

RDAN System Overview



- Collect and analyze high-volume data from multiple data sources
  - Unstructured text data (Phase 1)
  - Voice data (Phase 2)
  - Structured data & TENA (Phase 3)

- Automatically analyze and index data using cloud technologies to support rapid search and analysis operations on large data sets
- Provide custom parallel algorithms and architecture to reduce time from Data to Decision from hours/days/weeks to seconds

RDAN applies automated analysis using cloud computing technologies to reduce the time from Data to Decision

Cloud Background: Private Big Data Processing Cloud

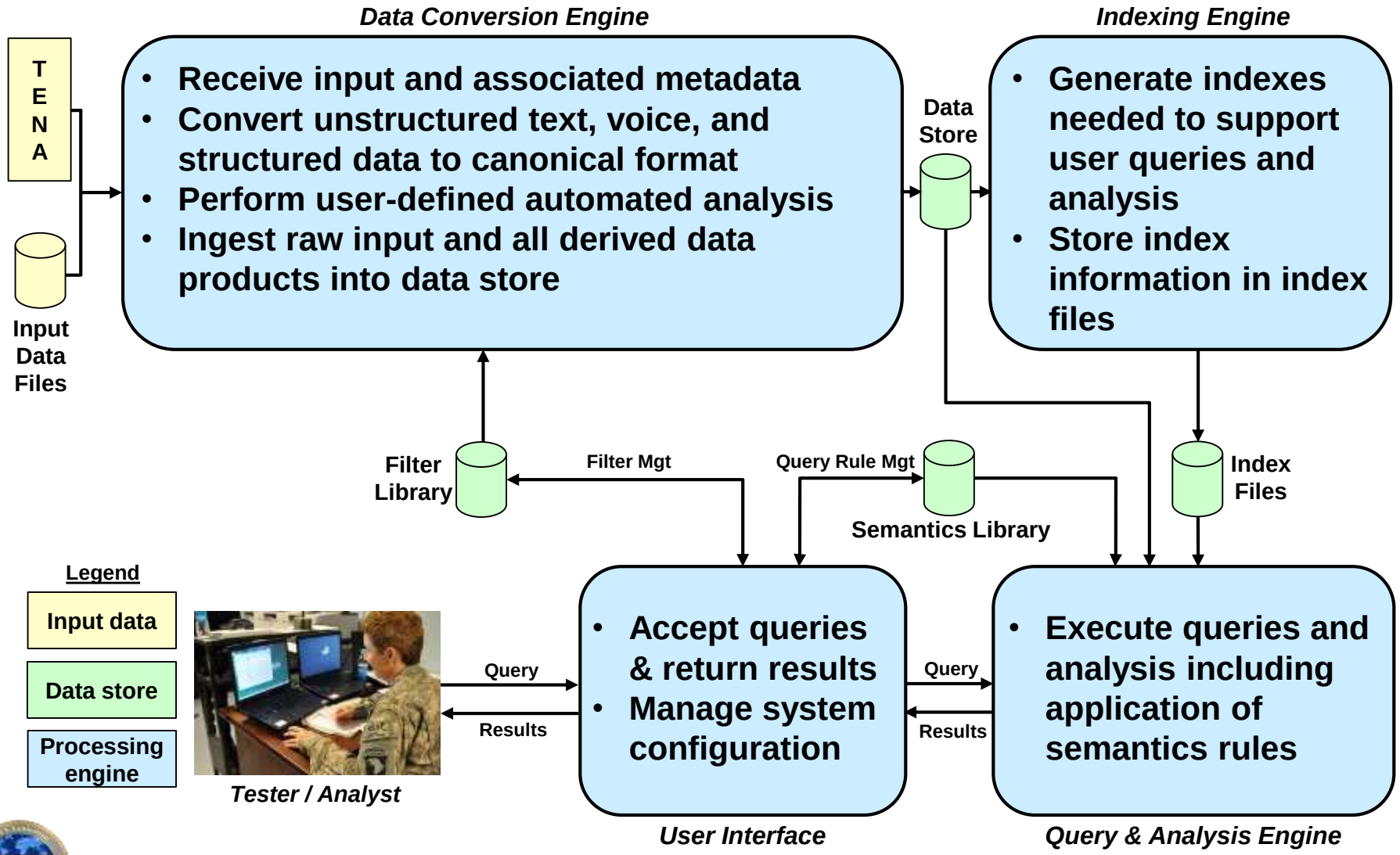
- **New cutting edge technology**
  - Hadoop publicly available in 2009
  - Apache Accumulo available in 2012
- **Optimized for processing big data**
  - Software framework designed for scalability & fault tolerance
  - Storage of data on compute nodes eliminates I/O bottlenecks
  - Large clusters with inexpensive commodity components support massive aggregate I/O, CPU, and network capacity
  - System hardware can be precisely tuned for workload

RDAN Private Cloud System

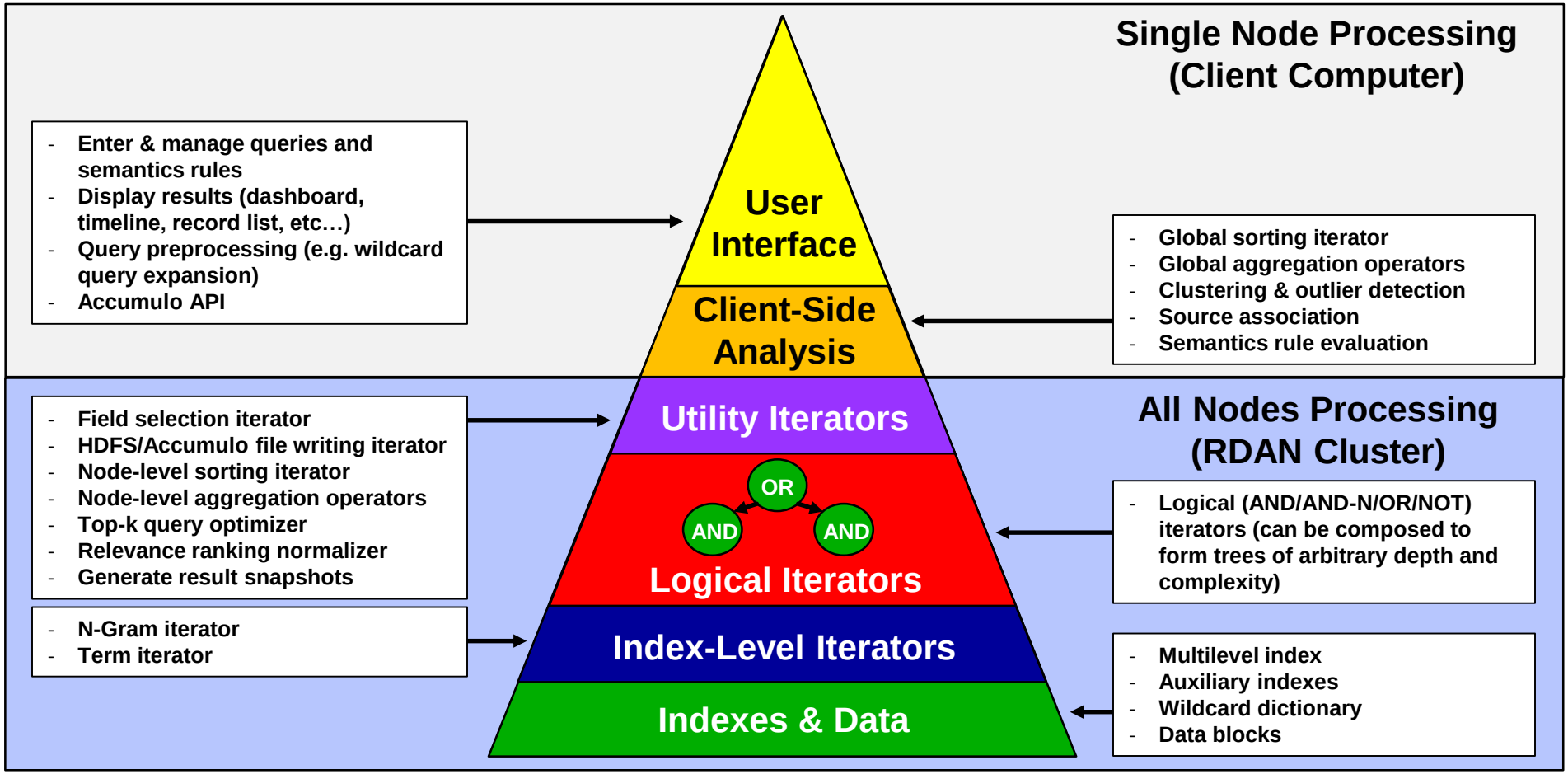
Apache hadoop accumulo™

Private big data processing cloud architecture is optimized for processing large volumes of data

RDAN Functional Diagram

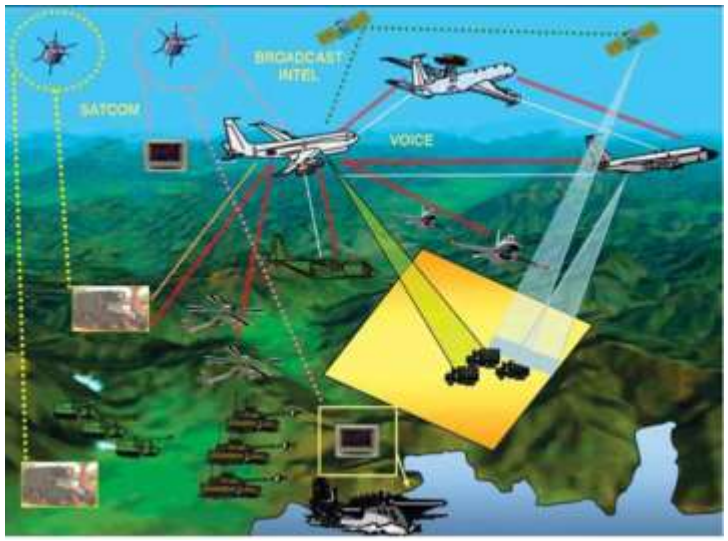


RDAN Query & Analysis Tools



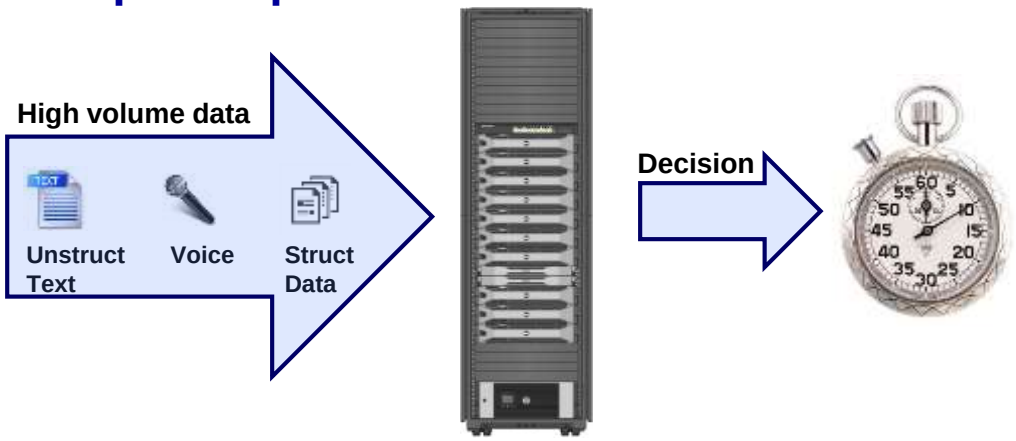
RDAN supports a diverse set of query and analysis tools that can be combined to support automated analysis

Key S&T Innovations



RDAN allows testers and analysts to perform near real-time analysis of complex distributed T&E events

Rapid response time from Data to Decision



Diverse data formats collected using a single system

New automated analysis and data fusion

Data to Decision is rapid (seconds)

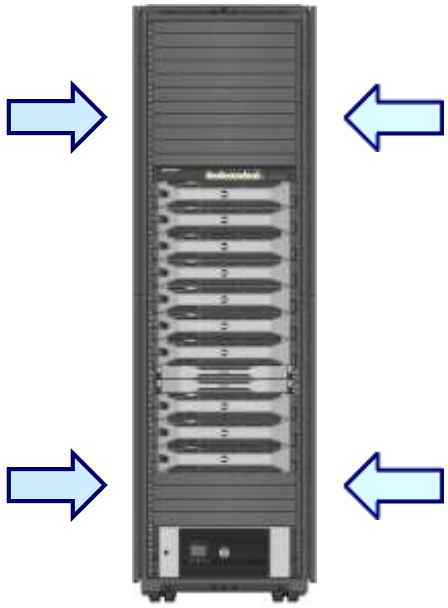
- **Novel Data Ingestion Architecture**
  - Custom parallel algorithms provide scalability, high throughput, and low latency for storage, indexing, and analysis using the latest cloud technologies
- **Pipelined Indexing Architecture**
  - New data structures and algorithms significantly improve indexing throughput while maintaining low query latency
- **Extensible Canonical Data Format**
  - Self describing data format allows new sensors and analysis modules to be added to system without modifying system architecture
- **Flexible Search and Analysis Tools**
  - New semantics rules allow analysts to search and analyze data using high-level constructs

Potential RDAN Applications

Analyze large unstructured and structured audit logs to rapidly detect cyber vulnerabilities

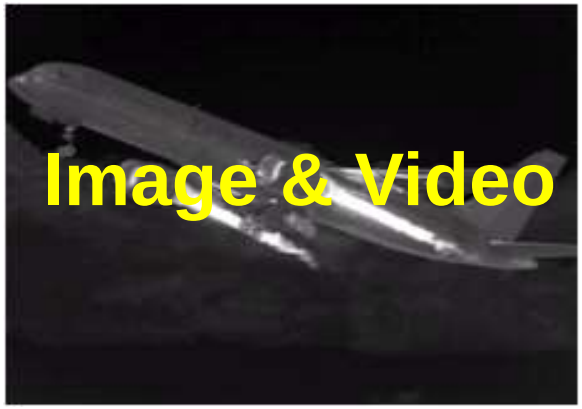
Provide near real-time feedback about test events to improve test range utilization

Rapidly collect, filter, store, and analyze diverse high volume data collected during T&E events



RDAN

RDAN can support multiple T&E needs



Automate analysis of large volumes of image and video data collected during T&E events

Summary and Future Work

- **RDAN is a prototype end-to-end system that utilizes the latest breakthroughs in cloud technologies to automatically analyze large volumes of unstructured, voice, and structured test data**
  - Reduces the time from Data to Decision from hours/days/weeks to seconds
  - Offers interactive and automated analysis of both live and recorded data
  - Supports near real-time analysis of T&E events as they are taking place
  - Scalable and highly configurable to support multiple T&E programs
- **RDAN mitigates risk during T&E events by providing near real-time analysis of data of different types, structures, and sizes**
  - Near real-time analysis of test event saves time and money
  - Automated processing of data minimizes human errors
  - Supports review of collected historical data during test
- **Proposed extensions to RDAN system**
  - Add support for high data rate image and video processing
  - Add support for cell-level security
  - Increase system performance and processing density using Graphics Processing Units

Points of Contact

Mr. Gil Torres
Net-Centric System Test Executing Agent
gilbert.a.torres@navy.mil

Ms. Kathy Smith
Net-Centric System Test Deputy Executing Agent
ksmith@gblsys.com

Mr. Bruce Einfalt
Principal Investigator (PI)
bte2@arl.psu.edu

Mr. Andrew Shaffer
Technical Lead
aps148@arl.psu.edu

Questions?