



CLEARED
For Open Publication

May 09, 2019

Department of Defense
OFFICE OF PREPUBLICATION AND SECURITY REVIEW

SLIDES ONLY
NO SCRIPT PROVIDED

JSF KM / RAPIDS Lessons Learned

Tracy Mullendore
Project Manager
JSF-KM / RAPIDS

Tracy.Mullendore@tena-sda.org

Bill Williams
Systems Engineering Lead
Joint Mission Environment Test
Capability (JMETC)

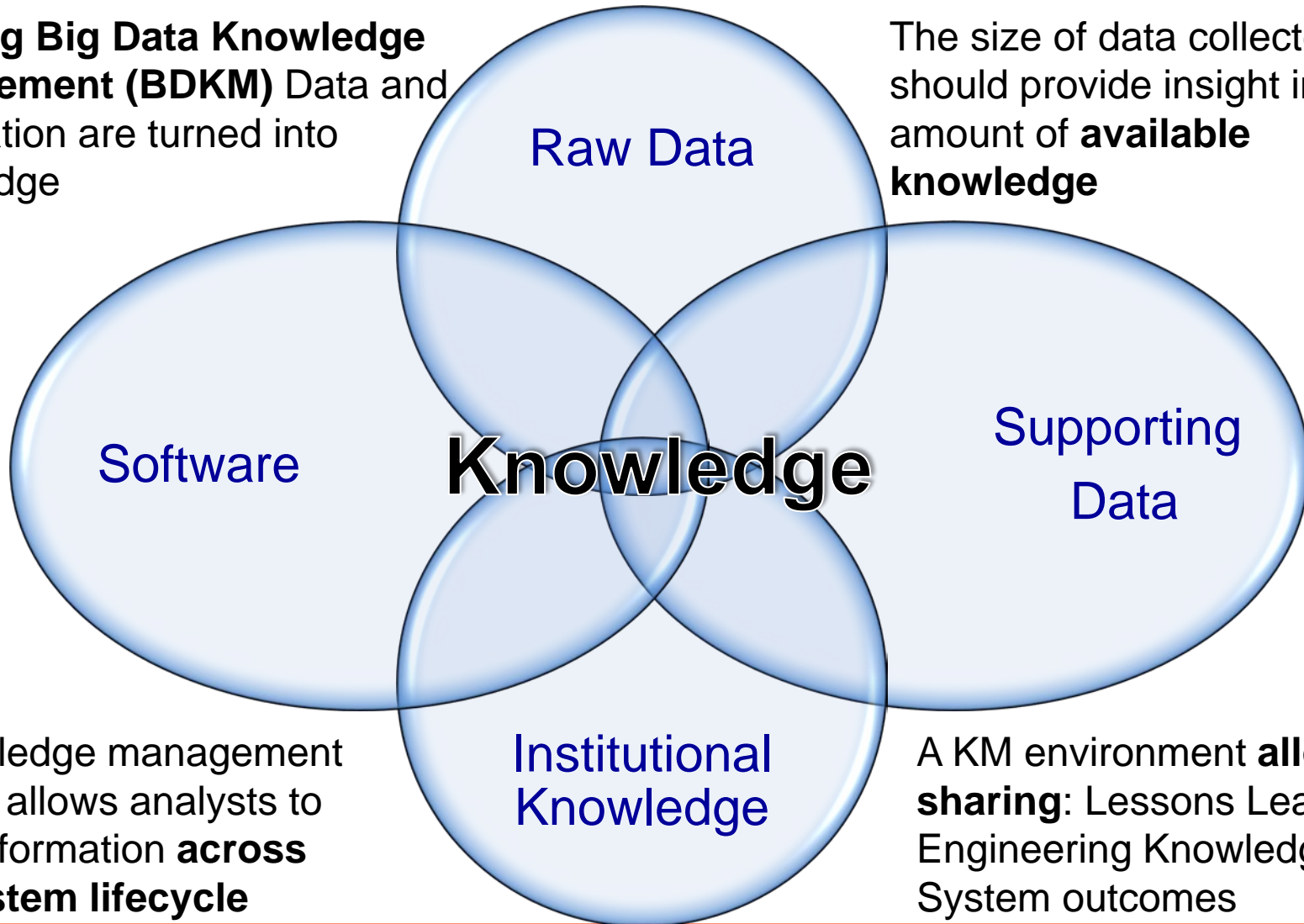
William.Williams@tena-sda.org



What is Knowledge Management?

Utilizing Big Data Knowledge Management (BDKM) Data and information are turned into knowledge

The size of data collected should provide insight into the amount of **available knowledge**



A knowledge management system allows analysts to have information **across the system lifecycle**

A KM environment **allows sharing**: Lessons Learned, Engineering Knowledge, System outcomes



What is “Big Data”?



- **What is Big Data?**

- Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis
- It’s not about the size of the data
- But it’s not the amount of data that’s important. It’s what organizations do with the data that matters
- Big data can be analyzed for insights that lead to better decisions and strategic business moves

- **I don’t have “Big Data.” How would BDKM really help me?**

- Are you efficiently creating quick-look reports?
- Are you able to repeatedly analyze all of your data or just your test points?
- Are you able to get your test reports out as quickly as you would like?

- **Big Data goals**

- Streamline processes to meet the operational need
- To empower a more efficient delivery of capability to the warfighter
- Instead of the systems engineering “V” utilize an iterative and agile process that has the flexibility to ensure all of the appropriate engineering and validation steps are met without the need to check boxes but deliver capability when its ready



What about Analytics?

- **My analysis tools work fine, why do I need analytics?**
 - The *difference between data analysis and data analytics* is that *data analytics* is a broader term of which *data analysis* forms a subcomponent
 - *Data analysis* refers to the process of compiling and *analyzing* data to support decision making, whereas *data analytics* also includes the tools and techniques use to do so
 - Continue to use your existing tools while utilizing technology to improve and expedite reporting to provide decision quality knowledge
- **Advantages in the use of BDKM and Analytics expands the Evaluation capability of the Range community**
 - “Every time we analyze existing test data, we find Operational test points that have already been met if data was shared across program test phases”
 - By using KM and analytics we can evaluate and review executed tests for existing data rather than utilizing Monte Carlo testing

Analysis tools are part of the analytic environment



Analytics Methodology



Roles



Data developer

- Build new data workflows
- Monitor data processing errors / backlog



Data SME

- Validate data against source
- Manage dictionary



Data Scientist / test engineer

- Anomaly detection / numerical workflows
- Build models against use cases
- Communicate results



Analytics Performance Manager

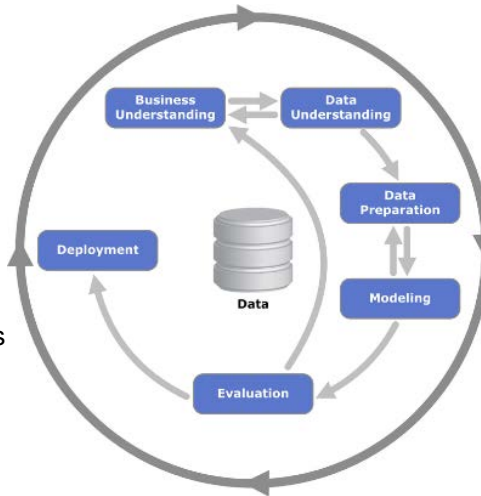
- Measure analytics performance
- Order new models / re-build old ones
- Communicate results



Engineers / designers

- Consume analytics outcomes
- Take actions

Process



Cross industry standard process – data mining (CRISP-DM)

Tools

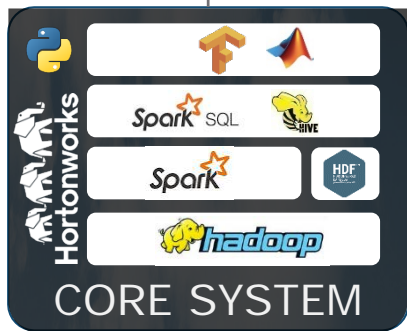
- Live and interactive dashboards
 - Data summaries
 - Analytics
 - Use cases
- Views into raw data for ad-hoc analysis
 - Exploratory visualization
 - Data transformations
- Integrated computational analytics
 - Scalable computing
 - Comprehensive library
 - Diverse levels of development effort
- Analytic reporting
 - Easily shareable
 - Easily reproducible / replicable
 - Automated

Knowledge Management
(Technology platform)

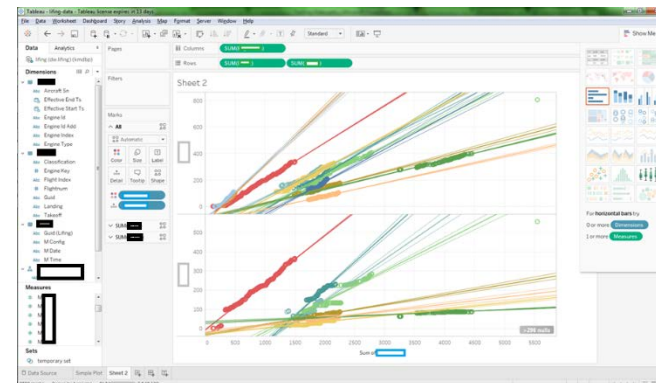


Analytics Tools

Overview of toolset for analytics



- Simple drag-and-drop plotting capability
- Easily switch between plot types
- Dashboarding with multiple linked plots



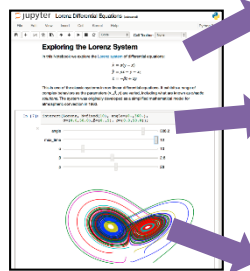
Description: Browser IDE and reporting tool incorporating analytic code (Python / R)



- Mix logical blocks with descriptive text and markdown
- Interactive computing
- Reproducible results
- Export via PDF, HTML, Python script, slide show

Use cases:

- Reproducible reports for specific use cases ("shot-sheet")
- Create technical documentation of model description for proposals of new models
- Sharing code and output in one file among a team or to one's manager



Data workflow script

Data integrity report

Analytic model documentation

Slide show

- Export as pure script for execution
- Export as PDF for full description

- Export as PDF for full documentation
- Export as HTML for interactivity

- Include widget elements
- Export as slide show

Reference: jupyter.org/



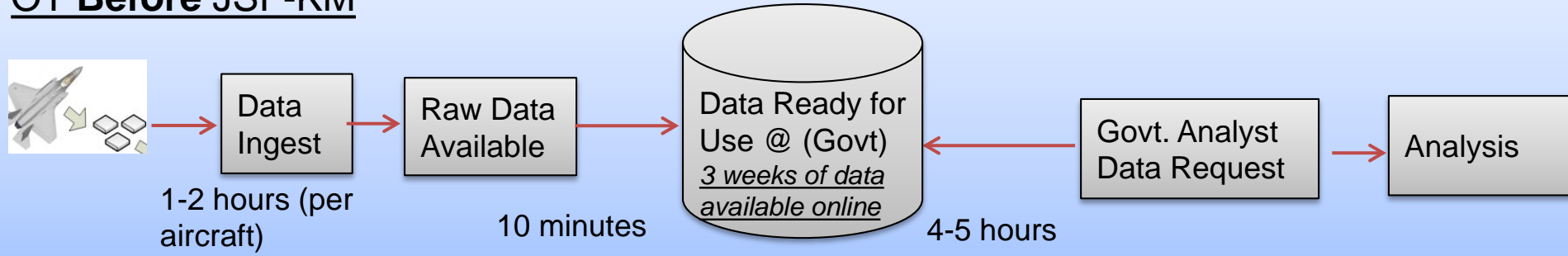
Achieved Optimizations Utilizing BDKM for JSF



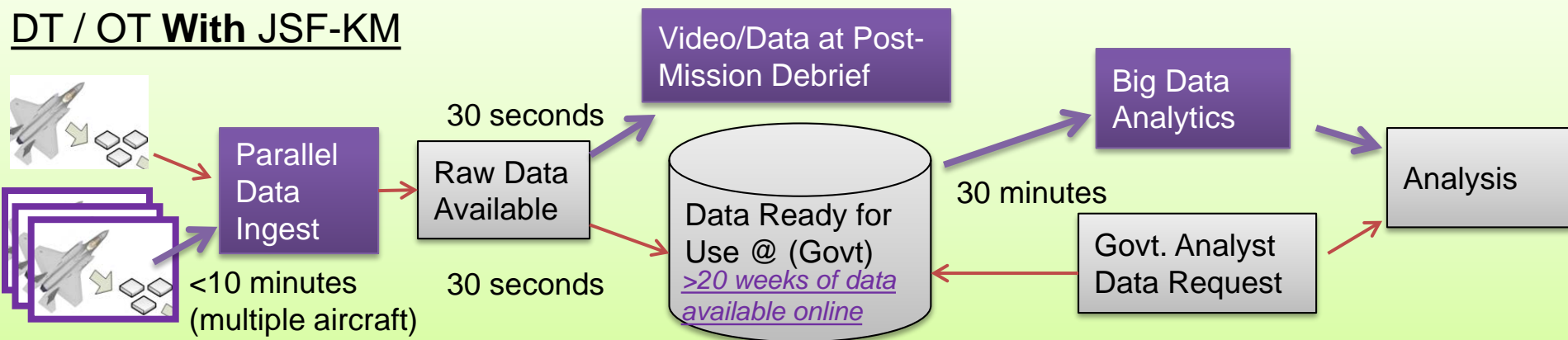
DT Before JSF-KM



OT Before JSF-KM



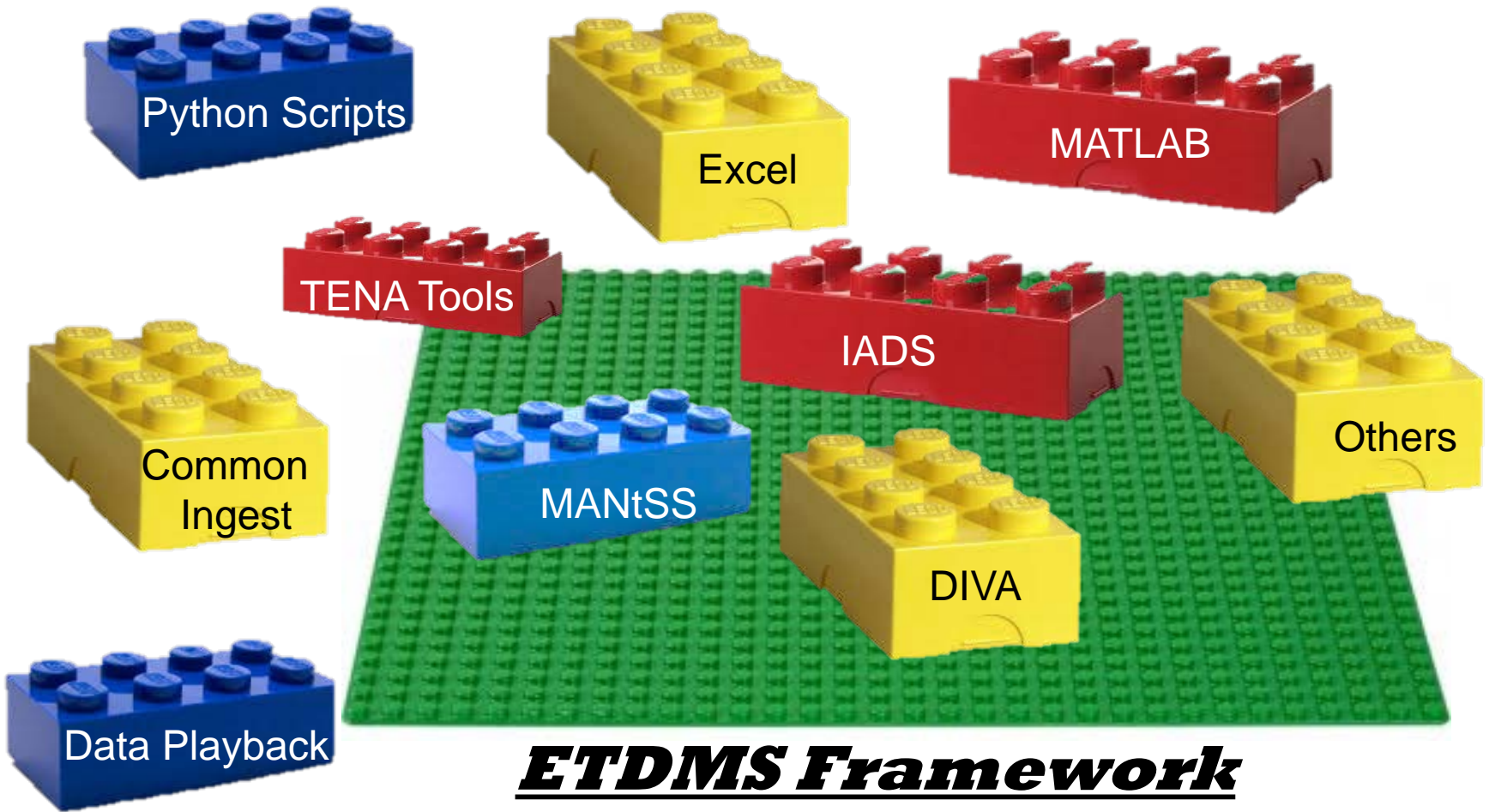
DT / OT With JSF-KM



Note: Numbers reflect single 2 hour flight mission



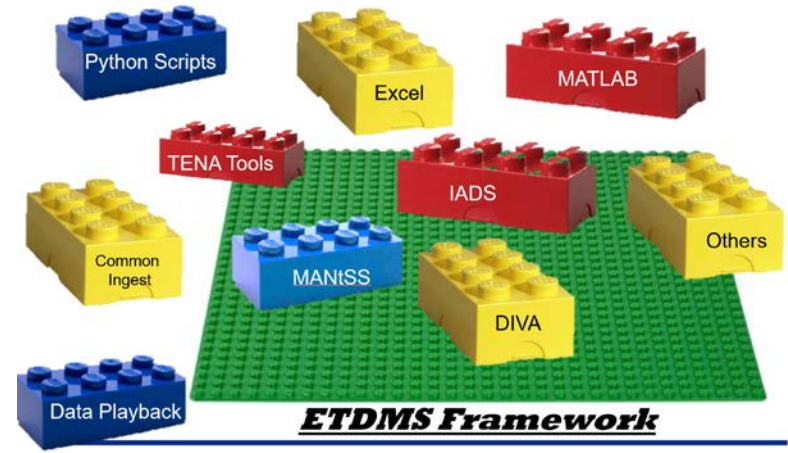
JSF-KM & RAPIDS Software Building Blocks



JSF-KM is a GOTS modularized system enabling customization



JSF-KM & RAPIDS Hardware Building Blocks



Laptop/PC



Small Mobile Analytic Platform



Small Computer Cluster
(Mixed or Common)



GPU Enabled Hyper Converged Cluster

RAPIDS/KM footprint is dynamic to processing requirements



Representative Hardware Footprint



RAPIDS

Parts List

Hardened Rack x2

Network Switch

Cartridge Reader

VxRail: 80 Cores, 2TB
RAM, 27TB Virtual Storage

Unity Storage: 60 TB All
Flash Data Storage

Tape Backup Library

Wyse Terminals with
Monitors

Power Backup Unit



KM

Parts List

Hardened Rack x2

3 Network Switches

2 Multi-Cartridge Readers

VxRail: 192 Cores (Physical), 12.2TB
RAM, 200TB SSD Virtual Storage

Isilon Storage: 1PB Hybrid
Storage (SSD and Spinning)

Wyse Terminals with Monitors

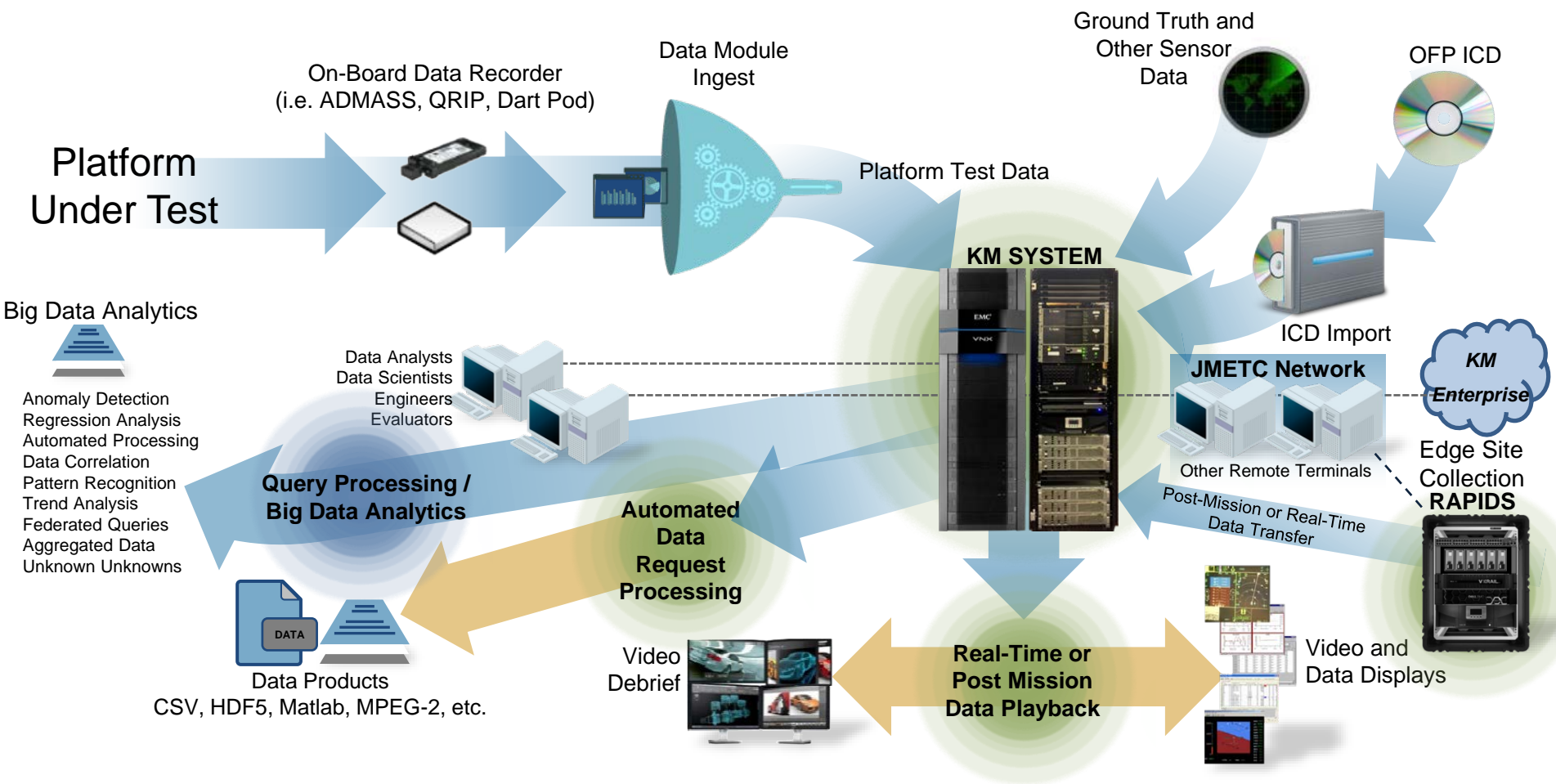
Power Backup Unit



Network



JSF-KM & RAPIDS Processing CONOPS



Processing flexibility to support the RDT&E environment



TRMC Proven Experience Implementing BDKM



TRMC will leverage lessons learned to implement the BDKM enterprise from the JSF-KM proof-of-concept which supports day-to-day platform analysis:

- **Over 3000 missions utilizing QRIP**
- **Over 300 missions utilized KM and RAPIDS for analytics and debrief**
- **RAPIDS system has supported multiple safari test/detachment missions**
- **BDKM software (ETDMS) has enabled analysts to evaluate system performance across platform builds/ICD's**
- **Hands-on Training developed to provide to the community quarterly**



JSF-KM Successes



- **Improved speed of analysis**

- Pilot data and video replay available during post-mission pilot debrief
- Reduced 9 hour MATLAB vibration analysis process to 23 ms
- Reduced data profile time from 5+ hours to 47 seconds per query
- AFOTEC SBIRS workflow optimizations reduced analysis process from weeks to days
- Enabling analysts to spend more time analyzing data – and less time gathering data

- **Improved quality of analysis**

- Identified 2 engines that consistently performed differently than others
- Identified a faulty/noisy ground sensor
- Found anomalous points and pattern within inconsistent sensor data sampling rates
- DT tools supported analysis within OT event during weapons testing

- **Automated Analysis Enables Creation of Predictive Models**

- Data scientist identified 70% of flights with engine issue and created a predictive model for identifying future failures
- Machine learning system able to identify combination of engine parameters which could be an early indicator for an uncommanded thrust event that occurred in the field

- **JSF Analysis & Workflow Enhancements**

- RAPIDS portable hardware configuration enabled safari testing to support test at multiple DET missions
- JSF complete ICD import only takes KM ~15 minutes vs. 24-48 hours for DPGS
- KM enabled viewing a complete ICD message near instantly vs. DPGS taking 30+ minutes
- Virtual desktop provisioning enables a new analyst desktop standup in minutes rather than days



Lessons Learned



- **ATO / IA for SAP/SAR environment**
 - Having dedicated, experienced team member for IA suggested for future efforts
 - Building relationships with IA POC's within the project/site are crucial to success
- **Personnel Management**
 - Difficult to find experienced and trained Data Scientists who can be cleared
 - Limited program “ticket” availability makes it difficult to get developer access
 - More work than cleared people on development team
- **Contracting**
 - Must have relevant contract vehicle/mechanism for personnel to be read-in
 - Overly optimistic award date estimates makes funding labor challenging without breaks in support or contractor working at risk
 - Large hardware buys can be difficult if not planned for the contract
 - Pass-through costs can be very high
 - Consider all options when determining contract strategy



Lessons Learned



- **Software / Hardware Configuration Management**
 - Implementation of Agile software process greatly assisted stability of codebase and increased team and customer's confidence in developers
 - Due to IA delays, hardware did not meet emerging requirements once ATO received
- **Concurrently: Building, Debugging, Deploying, Redesigning, Securing, Redeploying, Training, Documenting**
 - Documenting and agreeing on customer requirements imperative even on R&D efforts
 - Technical implementation choices should be delayed until requirements are fully vetted and agreed to by both gov and contractor
- **Proof of Concept / Prototype Expectation Management**
 - It is challenging to overhaul the system when utilizing the pathfinder technology in the day-to-day mission
 - Visibility can quickly turn into exposure
 - Growing interest risks alienation of other efforts or stovepipe developments
 - SME is needed to inform the data scientist to create quality and timely reports and analytics



Lessons Learned



- **Rapidly changing landscape of analytic tools**
 - Immature field of products led to rearchitecting backend
 - What works for industry doesn't always work for DoD
 - Truly bleeding edge - JSF data rates and amount of data drastically exceeded past DoD research efforts
 - A test environment would have alleviated a lot of trial and error work on production systems
 - Some open source software requires or fires an ad-hoc webserver and needs to be staged to ensure remote access to computer is not created
 - Depending on the env it is installed, most help documentation redirects to an online document store/webserver
- **SUT Platform Configuration Control**
 - Lack of documentation for SUT changes made initial understanding of data and related behavior difficult for developing and verifying the ingest capability



Future of T&E and BDKM



- **As a way to get new capabilities to the warfighter the Services are working to combine testing (Contractor/Development/Operational)**
 - As new platforms are being introduced to the Department, technologically advanced solutions are developed based on emerging technologies
 - The cost and amount of data collected continues to rise while timelines to meet operational requirements continue to get shorter
- **BDKM use within RDT&E can support combined testing (Contractor/Development/Operational):**
 - The development test data now needs to be a fidelity and quality that allows the operational test points to be evaluated
 - This combination of test more readily empowers a Program to utilize a single KM environment
 - The ranges can provide a more consistent infrastructure for test that doesn't require multiple investments for evaluation and reporting of data



Summary



TRMC will continue to collaborate with the community to stay current with the RDT&E BDKM challenges, requirements and lessons learned so we may inform investments to address those needs:

- **Big Data is about your ability to process your data – not the amount or size**
- **Analytics include the tools to optimize, automate, and expedite analysis**
- **JSF-KM and RAPIDS is modular software that will be shared with the community**
- **TRMC has viable lessons learned to share with the community**
- **TRMC has a vision to implement a BDKM enterprise for the RDT&E community**
- **TRMC is committed to funding the BDKM software baseline O&M for the community**



Questions?



ETDMS User Training



- **Trained potential users to access the entire flight database and the methodologies for getting datasets out of the system based on relevant criteria**
- **This gave current and potential users the opportunity to build ad-hoc visualization and dashboards**
 - Dashboard Ex. Flight path, drag and drop to a map; enter a tail number and plot flight path being able to see information in seconds (summary)
 - Visualization Ex: Plot of a single sensor output for further analysis and reporting (details)
- **Users went from novice SME's - to having an understanding of how to navigate the EDTMS framework to utilize the tools**



JSF-KM FY19 Success Stories



- **JSF JPO inserted TRMC KM system into JSF Block 4 TEMP**
- **Unknown data behavior discovered while developing shot sheet providing a better understanding of system behavior**
 - KM tools discovered that data was jumping between channels as part of the time outs
 - Issue was not showing up within LM tool suite and may not have been discovered without KM
- **Creation of an automatic time correct algorithm reduced analysts data processing time from 8 hours for a single flight to minutes for multiple flights**
 - Faster analyst access to time correlated data without having to wait for previous process to complete
 - Without JSF-KM analysts would still be spending 8 hours per flight correcting time issues
- **ICD comparison tool**
 - Prior to KM analysts had no way to know what changed within the new ICDs from LM