A Reference Architecture for Big Data Systems in the National Security Domain

John Klein (CMU SEI)

Ross Buglak, David Blockow, Troy Wuttke, Brenton Cooper (D2D CRC)

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213
Copyright 2016 Carnegie Mellon University

This material is based upon work funded and supported by the Data to Decisions CRC (D2D CRC), the Australian government’s Cooperative Research Centres Programme, and by the United States Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center sponsored by the United States Department of Defense.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN “AS-IS” BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

>Distribution Statement A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

DM-0003630
What is a reference architecture?

Knowledge transition tool - capture domain and solution knowledge

• Problem domain-specific
• Solution context-specific

The ”Art of the RA” – scope the problem domain and solution detail to optimize cost/value

Uses of our RA

• More descriptive than prescriptive
• Define concepts, functions, and solution elements to enable specification and comparison of systems for acquisition
• Framework for identification and analysis of tradeoffs
• Design template annotated with expert guidance
RA Contents

Motivating use cases - define problem scope

Architecture decomposition based on grouping of related concerns into architectural modules

Mapping of current technologies onto the concerns

Demonstration of how to use the RA to create big data system architectures
RA Module Decomposition

Big Data Application Provider

Application Orchestration

Collection  Preparation  Analytics  Visualisation  Access

Big Data Framework Provider

Processing

Messaging

Data Storage

Infrastructure

Legend

Module  Cross-Cutting Module  External System

DATA  Data Flow  Service Use
Lessons Learned

Use the RA to define several system architectures
  • Do this while you are developing the RA - validates content and usability

Ask “what is different about big data?”
  • Heuristic to keep you from filling the RA with general software knowledge

Provide mapping to specific products/technologies
  • Cut through hype, show where products fit into overall solution and limits of the contribution of each product
  • Separate section to allow easy periodic updates