BDEC Platform Demonstrator: A Global Data Logistics Platform for the Digital Contiuum

Micah Beck, Terry Moore, Piotr Luszczek   Ezra Kissel, Martin Swany
University of Tennessee   Indiana University

Introduction — People in the BDEC community are well positioned to understand the obstacles to research cooperation that different kinds of boundaries—national, physical, social, organizational, and technological—can produce. From years of experience building cyberinfrastructure for science and engineering, we know that, inevitably, the individuals who make up the international scientific community are embedded in a network of complex, multi-dimensional relationships that constrain their freedom to control or share their computing and data resources with potential collaborators across these boundaries. We also know that for the past three decades, the effects of such borders were mitigated and partly overcome by the TCP/IP+Unix/Linux platform paradigm, whose near universal adoption and use provided the foundation for interoperability and software portability throughout the community. But as described in the BDEC Pathways to Convergence report [1], this paradigm is becoming progressively more inadequate: swamped by the ongoing data tsunami; overwhelmed by swarms of new digital devices proliferating in the surrounding environment; and, most of all, rendered increasingly irrelevant and impervious to innovation for the purposes of science by its subjugation to the private ends of a few global cloud service corporations. Against this background, the BDEC demonstrator described below—*a Data Logistics platform (DLP) for the digital continuum*—expresses our conviction that the way forward to a next generation cyberinfrastructure for science is through a radically new platform design, one that meets the application challenges of the our transformed world while still achieving the same community-wide levels of adoption and use (of "deployment scalability") that the legacy paradigm enjoyed.

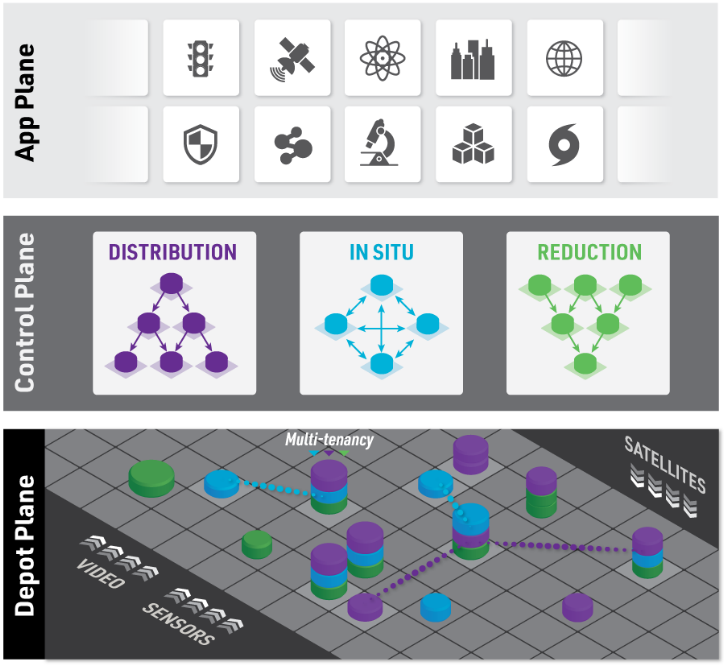

Figure 1: Simplified view of the *exposed buffer processing* (EBP) architecture for the DLP demonstrator: the local node layer composed of passive resource nodes (called *depots*) accessed via EBP protocols; the control layer with diverse local area or global service managers (e.g., file system, CDN, in transit/in situ data processing, etc.); and the application layer, where applications use the control plane managers to perform work on data in the node plane.

Vision of a Data Logistics platform Spanning the Digital Continuum— A data logistics platform is a distributed system in which the system's intermediate nodes, called *depots*, are used in a general way. That is, they don't just forward packets, but instead expose all three fundamental resources—communication, storage/buffer, and processing—to explicit programming. The BDEC DLP we propose to demonstrate is based on the idea that, in order to achieve the continuum-spanning interoperability and portability we seek, we have to build on a common service virtualization, or *spanning layer*, that is designed to be as simple, general and limited as possible while still supporting all necessary applications [2]. These requirements are fulfilled in the specification of a converged service that represents the "greatest common divisor" of storage, networking and computation, namely, *the allocation of, transfer of data between, and transformation of data in* ***buffers***, including both memory and storage and regardless of implementation. Accordingly refer to this design as an *exposed buffer processing* (EBP) architecture [3].

Figure 1 provides an overview its basic structure of this architecture and the BDEC DLP demonstrator. Using the Data Logistics Toolkit (*DLT*, [4]), it will be build on a collection of *DLT* depots deployed at a set of locations throughout the global research network. These depots, which constitute the data plane of the DLP give clients the ability to perform a set of low level fundamental operations, to whit: **1)** allocate a buffer, **2)** move data between buffers on a single or between LAN-connected depots or **3**) apply an operation to a set of buffers that serve the roles of inputs and/or outputs according to the definition of the operation. The protocol and service model are uniform across all depots in the DLP , whether they are providing access to the resources of nodes located at the very edge, the middle (or core) or at a data center. Thus access to the resources of all nodes that comprise the computing continuum are represented with as much uniformity and interoperability as possible. Nodes with greatly varying attributes may be differentiated by the publication of appropriate metadata (for instance distinguishing a node resource that exposes line buffers for transient buffer of network transfers from one that provides long lived storage cells for implementing file abstractions). But DLP data transfer must be interoperable between any two nodes that are adjacent in the LAN topology, with no intervening translators or gateways. Routing along paths is within the architecture, but gateways to overcome non-interoperability are not. Operations on data may similarly be implemented on some nodes, as described in appropriate metadata, but the same operation on two different nodes must be semantically indistinguishable.

Higher level services will be implemented by aggregating the fundamental operations that the depots support. These services are implemented by processes running in a control plane that has access to the entire data plane that models the computing continuum. In simple cases an entire service can be created by running a single process that issues commands to the nodes of the control plane, centralizing the logic of the service while distributing its execution to the data plane. In cases where such extreme physical centralization of the control plane causes problems (typically performance or reliability) a single process may be implemented by a set of processes that communicate using both the depots of the data plane and the Internet.

Having deployed a data plane, a particular (set of) demonstration application(s) will be chosen and a set of control plane services defined that are sufficient to implement it. A natural choice would be a prototype of the Earth Observing Data Network, a generalized Content Distribution Network that distributes a stream of satellite images emanating from one or a small set of orbiting instruments. The generalized nature of EODN allows it to support not only timely high performance access to satellite data in real time, but also allows light processing of those images using the nodes of the data plane as well as download to client facilitites. EODN also supports the upload by clients of secondary data products into EODN for further processing either in the data plane or after download by other clients. No such general, open, *cooperatively managed and operated* CDN currently exists for the global distribution of filtered, location-specific data from one nation's satellite, let alone a CDN for the distribution of relevant data from the satellites of international partners. The appetite for such data around the world, both inside and outside the scientific community, is substantial.

The Data Logistics Toolkit (*DLT*) as the Basis of the BDEC DLP The data storage and transfer functions of the DLP data plane will be implemented by the Internet Backplane Protocol (*IBP* depot, [5]), which implements the EBP spanning layer as an overlay on the legacy platform. A production quality, packaged implementation of the *IBP* depot is available in the NSF funded Data Logistics Toolkit (*DLT*). Light computing on data stored in *IBP* depots have been implemented experimentally, but this function is not currently part of the DLT. It will have to be reimplemented. The data storage, distribution and access functions of EODN are supported by the Intelligent Data Movement Service (IDMS) utility, and the IDMS policy can be controlled by the Flange languageṪhese capabilities are included in the current DLT distribution. Light processing of data within the data plane will require the augmentation of both IDMS and Flange with appropriate control functionality. Access to IDMS from an appropriate end-user interface, commands or client API will also be also required, as well as a means to harvest the stream of satellite images in real time.

Innovations of the BDEC DLP — The main innovations of the DLP derive from its architecture. The key one—a spanning layer that provides buffer-based interoperability and portability across the entire continuum—has been highlighted above. Some other unique attributes are briefly discussed below. It should be noted, however, that since the BDEC demonstrator will be implemented as an overlay on the legacy paradigm, it inherits the liabilities of that paradigm, specifically as regards security. The benefits of "Inherent support for role based, federated security" described below will have to await a native, or non-overlay implementation; the need for a new way to address such issues with the legacy paradigm is one the primary motives finding path to transition away from it. The BDEC DLP demonstrator will open that path.

- *Exposed topology for effective data logistics:* Approaches that hide topology from their clients are inherently inadequate platforms on which to build high traffic, globally distributed systems. To achieve efficient and performant data logistics, the EBP spanning layer exposes topology, including the placement and allocation of bandwidth, storage/memory and processing across the system. However, such exposed topology is generally too volatile to be "source scheduled" at one place in the network. A well tested solution is to create tightly controlled subdomains (i.e., Autonomous Systems (AS) and subnets) that peer at their boundaries. Defining subdomains allows for aggregate characterizations of platform topology that permit the level of exposure to be varied as the situation requires.
- *Decentralized service model enables local autonomy:* The creation of peering subdomains that control their own node layer resources has the added advantage decentralizing administration, giving localities the freedom of manage how their resources are shared. Since higher level functions do not need to have the same deployment scalability as the depot plane infrastructure, they may be centralized where necessary. This enables heterogeneity in the creation and management of higher level functions, allowing nodes to be very close (in network topology) to sensors, actuators and other edge devices. This combination of autonomous localities connected together by services based on weak assumptions will enable logistical services that efficiently make use of all the resources of the continuum.
- *Inherent support for role based, federated security:* The DLP spanning layer is local to the depot and doesnt export a global service; services that use it cannot assume the global reachability of any given local node. By minimizing the "target" that the node infrastructure offers to any external input or signal, DLP can create a kind of "white list" of allowable global services, as defined by the privileged control plane, leaving the node as impervious as possible to communication that is not part of such an authorized service. Then each higher level service can define its own strategy for authenticating, protecting and allowing access to the service it creates.

International Participation and Feasibility of the BDEC DLP Demonstrator — The fact that depots in the data plane can be deployed and used in a completely decentralized fashion makes participation the most basic form of participation in the DLP demonstrator straightforward: install one or more IBP depots on available hardware and give the result a suitable network connection. As descibed in [4], to achieve such easy deployment and configuration, DLT software has been packaged and documented for a number of operating system distributions; an accompanying meta-package resolves and installs any necessary dependencies needed to deploy an IBP depot node, as well as other companion DLT modules(e.g., Phoebus WAN accelerator, Periscope instrumentation, etc.), and it registers the new depot with the IDMS (see below). Available versions also include appliance images (VMs) that can be deployed on OpenStack- and Emulab-based rack technologies, as well as containerized versions with supporting documentation.

At the level of the control plane, we expect the *DLT*'s IDMS and policy engine to work for the BDEC DLP out of the box. Since the DLP depot plane provides a "bits are bits" infrastructure, the same functionality will be available to other application communities (e.g., microscopy, astronomy, etc.). But as Figure 1 suggests, various other more powerful services can be implemented on the control plane of the DLP , given availability operations installed on the depots. Whether the demonstrator plan is expanded in this way will likely be a function of community interest and available resources.

References

[1] M Asch, T Moore, R Badia, M Beck, P Beckman, T Bidot, F Bodin, F Cappello, A Choudhary, B de Supinski, E Deelman, J Dongarra, A Dubey, G Fox, H Fu, S Girona, W Gropp, M Heroux, Y Ishikawa, K Keahey, D Keyes, W Kramer, J-F Lavignon, Y Lu, S Matsuoka, B Mohr, D Reed, S Requena, J Saltz, T Schulthess, R Stevens, M Swany, A Szalay, W Tang, G Varoquaux, J-P Vilotte, R Wisniewski, Z Xu, and I Zacharov. Big data and extreme-scale computing: Pathways to convergence-toward a shaping strategy for a future software and data ecosystem for scientific inquiry. *The International Journal of High Performance Computing Applications*, 32(4):435–479, 2018. doi: 10.1177/1094342018778123. URL https://doi.org/10.1177/1094342018778123.

[2] Micah Beck. On the Hourglass Model, End-to-End Arguments, and Deployment Scalability. *Communications of the ACM*, to appear, July 2019.

[3] Micah Beck, Terry Moore, Piotr Luszczek, and Anthony Danalis. Interoperable convergence of storage, networking, and computation. In Kohei Arai and Rahul Bhatia, editors, *Advances in Information and Communication*, pages 667–690, Cham, 2019. Springer International Publishing. ISBN 978-3-030-12385-7.

[4] Ezra Kissel, Micah Beck, Nancy French, Martin Swany, and Terry Moore. Data Logistics: Toolkit and Applications. *GOODTECHS 2019 - 5th EAI International Conference on Smart Objects and Technologies for Social Good*, 2019. URL http://bit.ly/DLT-GoodTechs. submitted.

[5] A. Bassi, M. Beck, G. Fagg, T. Moore, J. S. Plank, M. Swany, and R. Wolski. The Internet Backplane Protocol: A study in resource sharing. In *Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on*, pages 194–194, May 2002. doi: 10.1109/CCGRID.2002.1017127.