FERMILAB-SLIDES-18-060-CD

BigData Express: Toward Predictable, Schedulable, and High-performance Data Transfer

Wenji Wu wenj@fnal.gov

Internet2 Global Summit

May 8, 2018

BigData Express

- Funded by DOE's office of Advanced Scientific Computing Research (ASCR)
- Collaborative effort by Fermilab and Oak Ridge National Laboratory
  - KISTI joined as a unfunded partner at 2017
  - ESnet provides WAN service
- A three-year research project
  - Start: Oct 1, 2015
  - End: Sep 30, 2018
- http://bigdataexpress.fnal.gov

Fermilab

OAK RIDGE National Laboratory

ESnet ENERGY SCIENCES NETWORK

KiSTi www.kisti.re.kr

BigData Express Research Team

- FNAL
  - Wenji Wu (PI)
  - Qiming Lu
  - Liang Zhang
  - Amy Jin
  - Sajith Sasidharan
  - Phil DeMar
- ORNL
  - Nageswara Rao
  - Gary Liu
- KISTI
  - Syed Asif Shah
  - Seo-Young Noh
  - Jin Kim

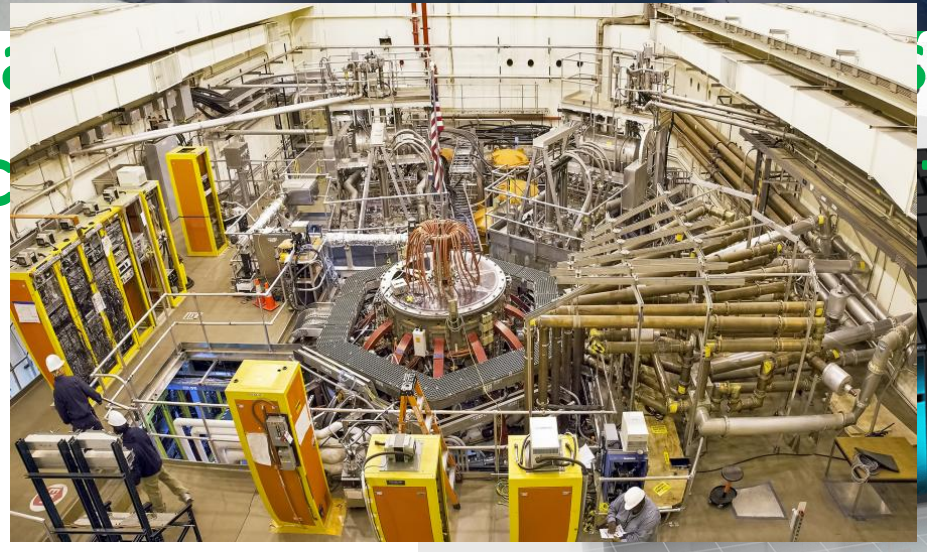
Note: KISTI and ESnet are unfunded project partners

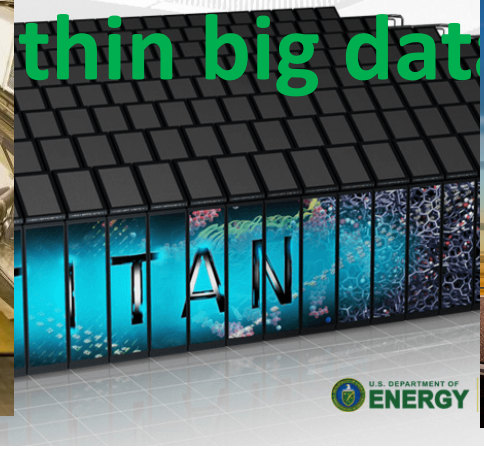
Big data enables scientific discoveries

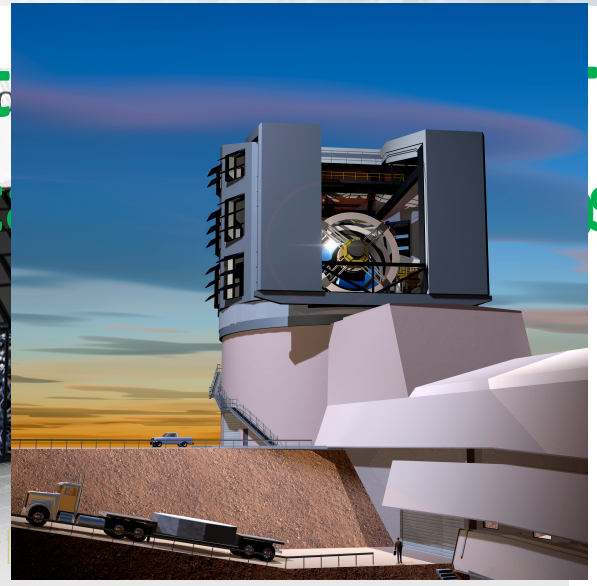
DOE Leadership Computing facilities offer computing and storage resources needed to process and analyze science data

The efficient movement of science data from their sources into processing and storage facilities and ultimately on to user analysis is critical to the success of any such endeavor.

Data [illegible] for the [illegible]-scale disc[illegible]rs.







Why BigData Express?

- **Targeted at optimizing data transfers in high-speed networks**
  - Large-scale data movement of Big Data Science
  - High-speed network environments (40/100GE+)
- **Builds on Multicore-Aware Data Transfer Middleware (MDTM)**
  - mdtmFTP: a high-performance data transfer tool
    - Pipelined I/O-centric design to streamline data transfer
    - MDTM optimizes use of underlying multicore system
    - Extremely efficient in transferring of Lots Of Small Files (LOSF)
  - http://mdtm.fnal.gov

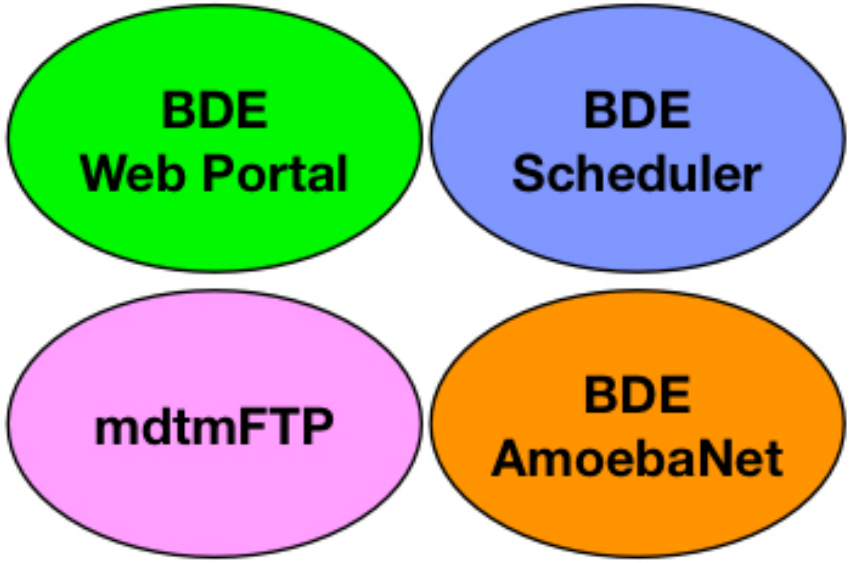
- **Orchestrates system (DTN), storage, & network (SDN) resources**
  - To provide full end-to-end performance optimization

BigData Express versus SENSE

- **BigData Express is data transfer middleware**
  - Uses SENSE for WAN SDN services
- **SENSE is a network service**
  - Provides higher-level applications with SDN-type services
  - BigData Express is an application to SENSE
- **BigData Express and SENSE are each stand-alone services in their own right**
  - BigData Express works fine without SENSE
    - WAN component is simply Best Effort
  - SENSE is agnostic to higher-level applications using its services

BigData Express Major Components

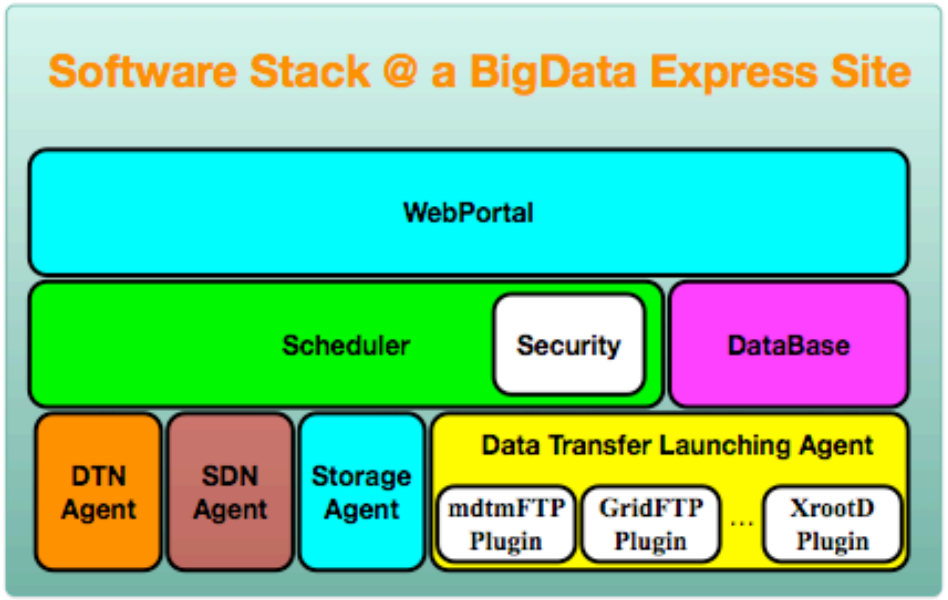
- **BDE Web Portal**
  - Allow users to access BigData Express data transfer services
- **BDE Scheduler**
  - DTN as a service
  - Co-scheduling of DTN, storage, and network
- **BDE AmoebaNet**
  - Network as a service
- **mdtmFTP**
  - a high-performance data transfer engine
  - http://mdtm.fnal.gov

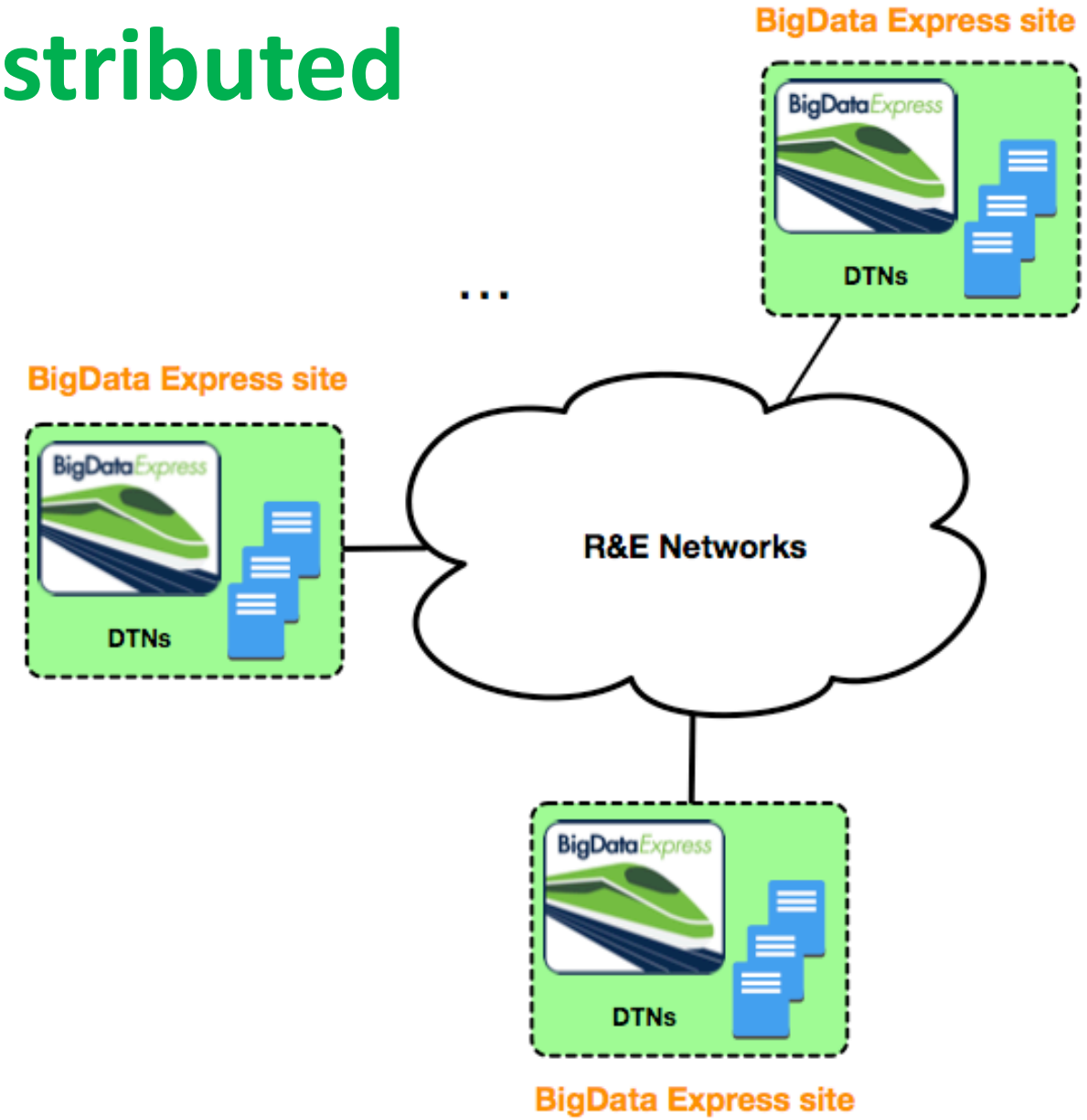


BigData Express Major Components (cont.)

- **DTN Agents**
  - Manage and configure DTNs
  - Collect and report the DTN configuration and status
- **Storage Agents**
  - Manage and configure storage systems
- **Data Transfer Launching Agent**
  - Launch data transfer jobs
  - Support different data transfer protocols

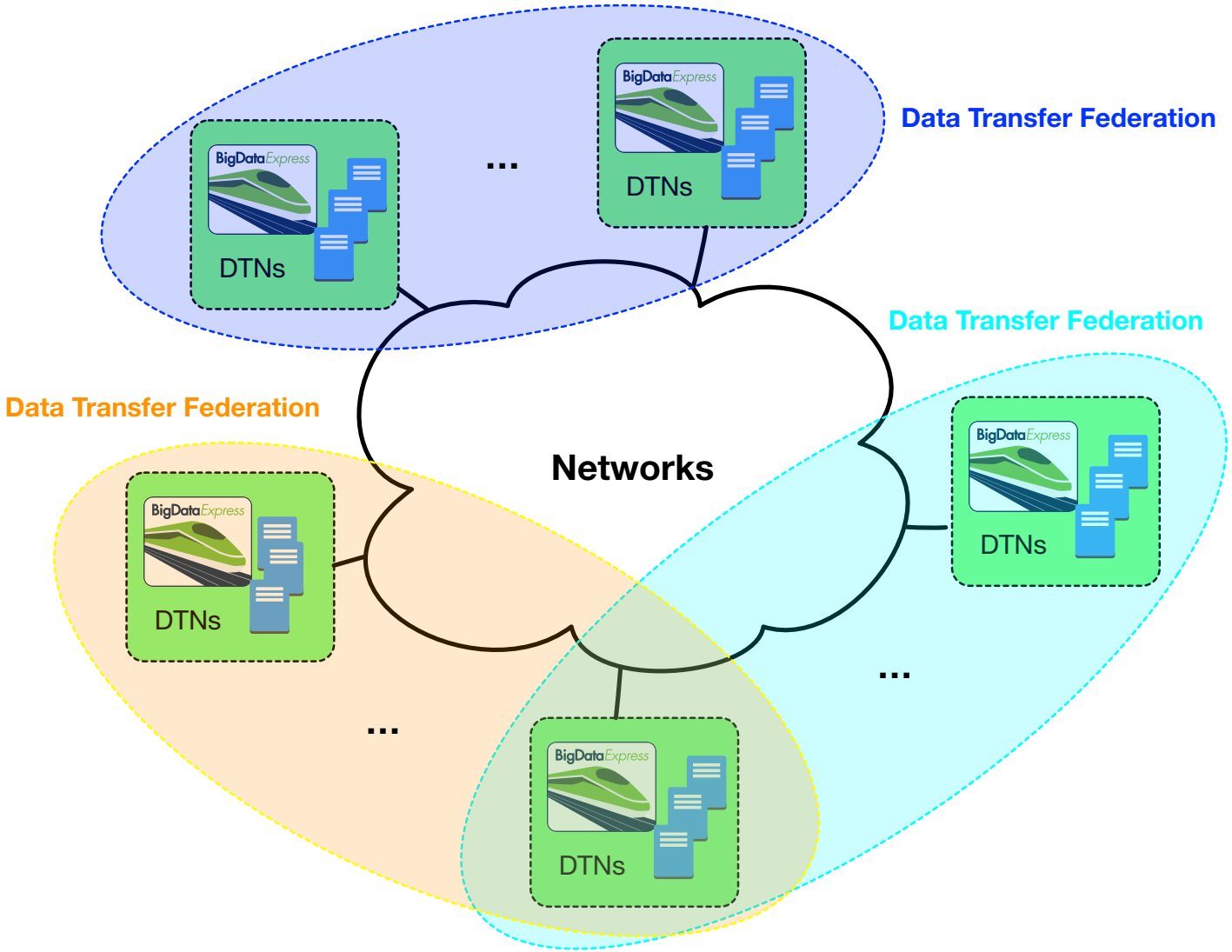
BigData Express -- Distributed





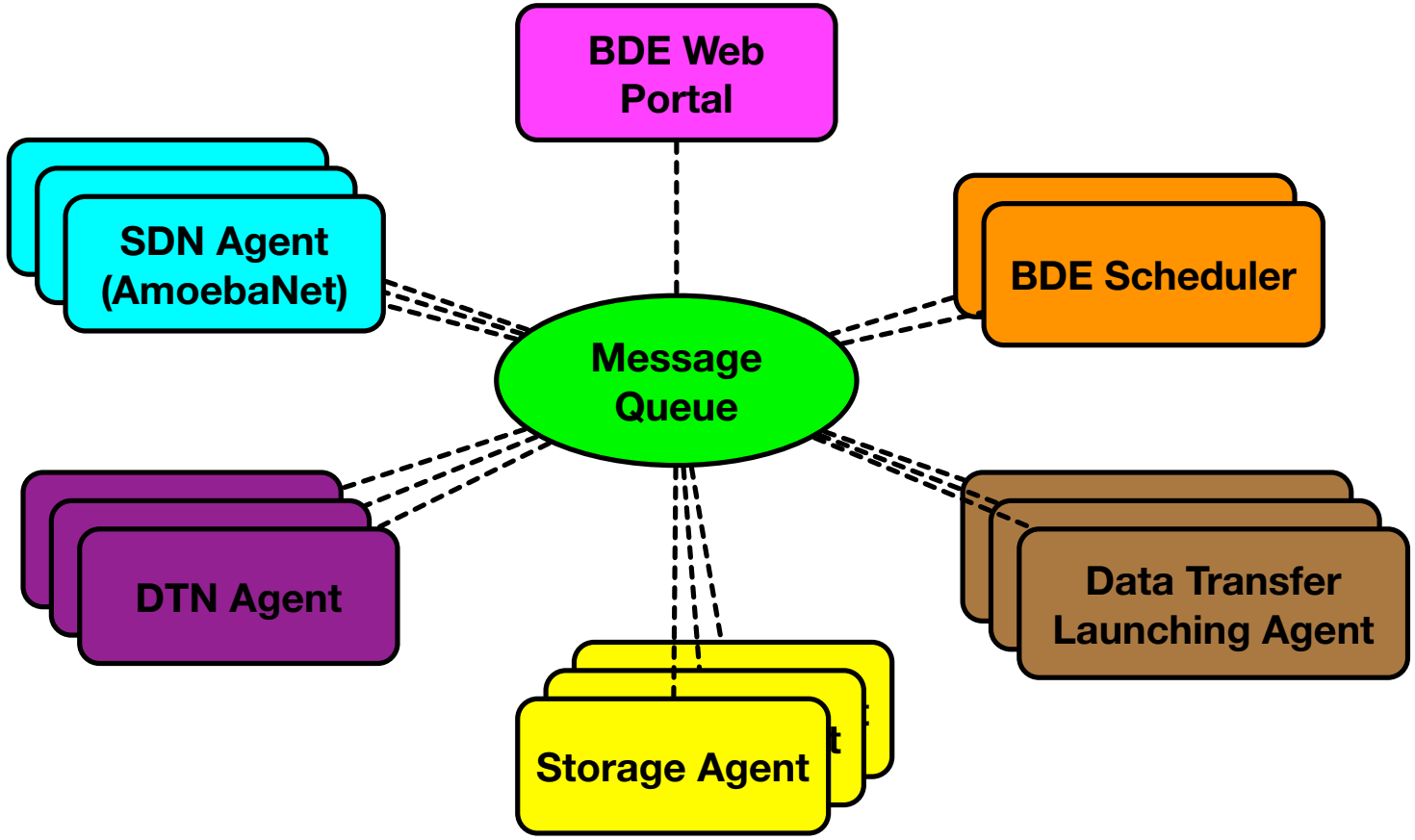
A Peer-to-Peer model

BigData Express -- Flexible



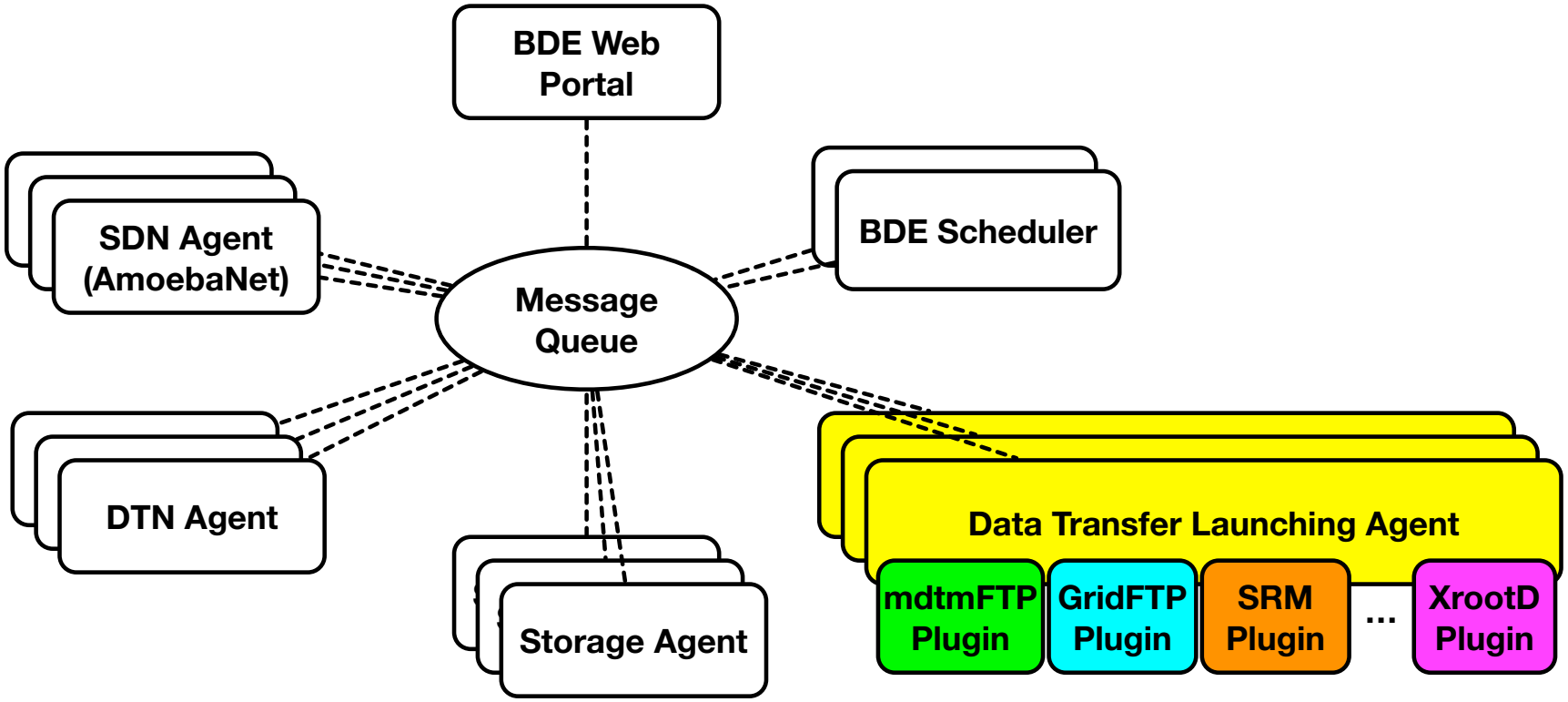
- **Flexible to set up data transfer federations**
- **Providing inherent support for incremental deployment**

BigData Express -- Scalable



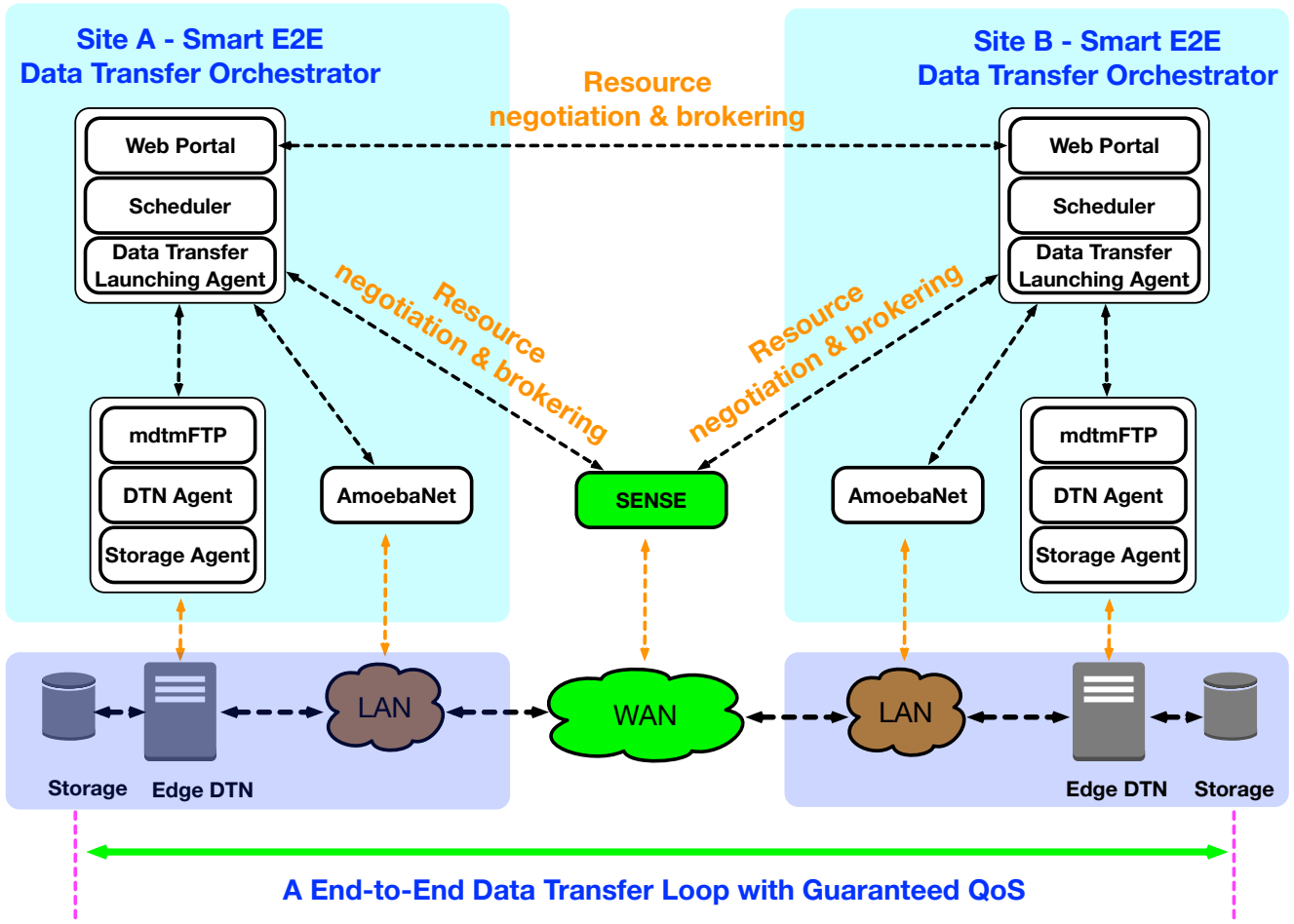
- BigData Express scheduler manages site resources through agents
- Use RabbitMQ as message bus

BigData Express -- Extensible



- Extensible Plugin framework to support various data transfer protocols
  - mdtmFTP, GridFTP, SRM, XrootD, …

BigData Express -- End-to-End Data Transfer Model



- Application-aware network service
  - On-demand programming
- Fast-provisioning of end-to-end network paths with guaranteed QoS
- Distributed resource negotiation & brokering

BigData Express -- Three Types of Data Transfer

- Real-time data transfer
- Deadline-bound data transfer
- Best-effort data transfer

BigData Express vs. Globus Online

Features	BigData Express	Globus Online
Architecture	• Distributed service • Flexible to set up data transfer federations	• Centralized service

Supported Protocols	• Extensible plugin framework to support multiple protocols:
	  ○ mdtmFTP
	  ○ GridFTP, XrootD, SRM (coming soon)
		• GridFTP

SDN Support	• Yes, Network as a service • Fast-provisioning end-to-end network paths with guaranteed QoS	• No
Supported Data Transfers	• Real-time data transfer • Deadline-bound data transfer • Best-effort data transfer	• Best-effort data transfer
Error Handling	• Checksum • Retransmit	• Checksum • Retransmit

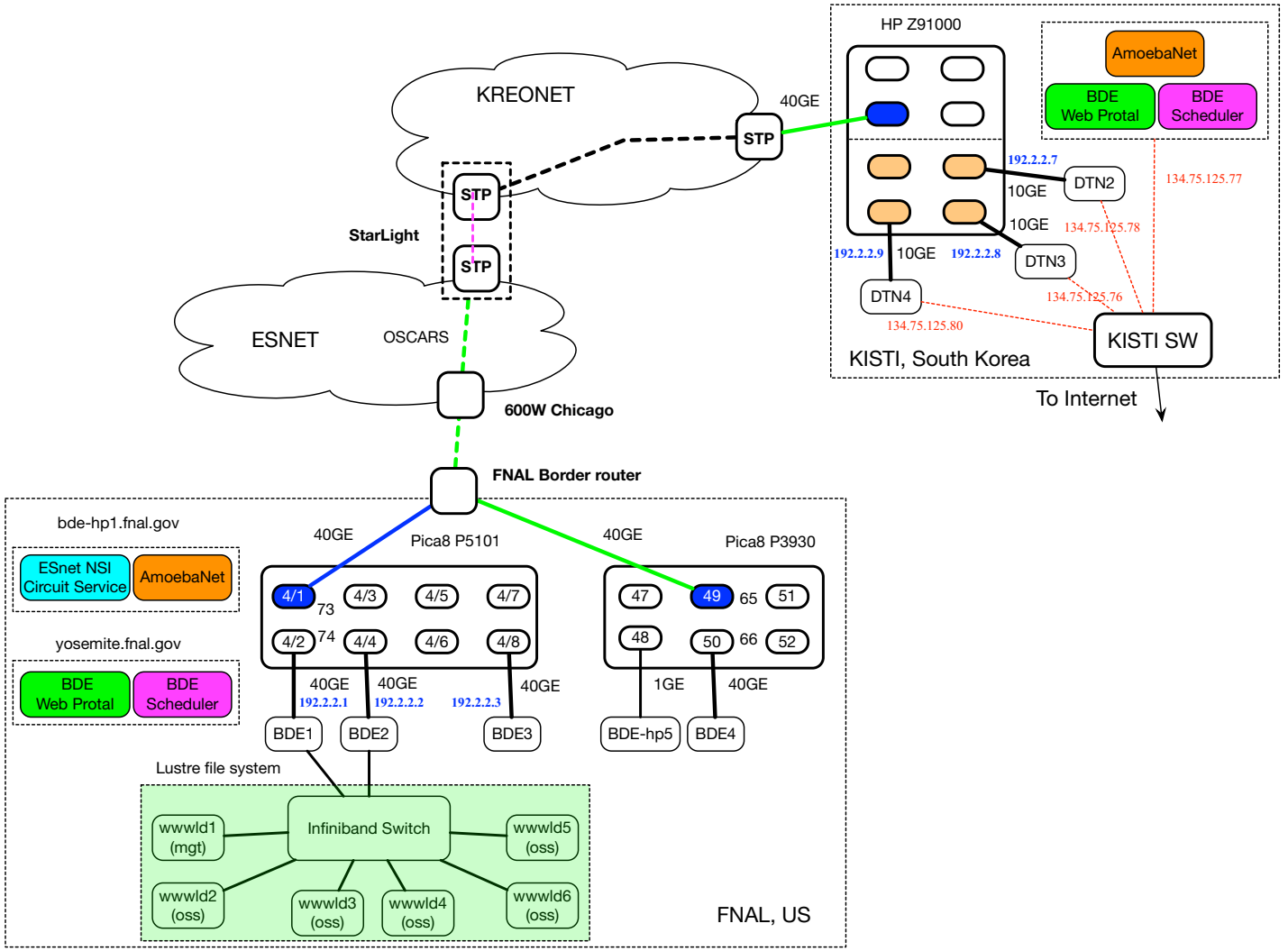
BigData Express SC'17 DEMO

- BigData Express: a schedulable, predictable, and high-performance data transfer service
  - QoS-guaranteed data transfer
  - DTN as a service
  - Network as a service
  - Distributed resource brokering/matching

A DOE/SC/ASCR-sponsored research project

Software is available at: http://bigdataexpress.fnal.gov

A Cross-Pacific SDN Testbed



BigData Express Deployment

- Completed deployment: KISTI, UMD, StarLight, FNAL
- Ongoing deployment: KSTAR, ESnet
- Work with StarLight to deploy BDE at XRPs
  - Pacific Research Platform (PRP)
  - National Research Platform (NRP)
  - Global Research Platform (GRP)
  - The European Research Platform (ERP)
  - Asia Research Platform (ARP)
- Collaborate with SENSE for BDE+SENSE deployment
- Work with US CMS to deploy BDE at US CMS sites

Support Science

- Fusion community
  - Work with KSTAR, KISTI, PPPL, and ORNL to transfer/stream data from KSTAR to US research institutions
- XRPs (PRP, NRP, GRP, ERP, ARP)
  - Work with StarLight to deploy BDE at XRPs to support various science
- HEP community
  - Work with US CMS to deploy BDE at US CMS sites
    - PI has been invited to give a BDE demo for US CMS
    - Tentatively scheduled for the last week of May, 2018

More information about BigData Express

http://bigdataexpress.fnal.gov

PI: Wenji Wu, Fermilab

wenji@fnal.gov