Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Belle II Computing and Networking Requirements

MALACHI SCHRAM

Pacific Northwest National Laboratory

LHCOPN-LHCONE Meeting University of Michigan (US)

Overview

- *Computing & Networking Requirements*
- *Belle II Networking Workshops*:
  - Pacific Network and Computing Requirements Workshop hosted by PNNL October 2012
  - European Networking Workshop hosted in Vienna – October 2013
  - General Networking Workshop during SC14 – November 2014
- *Data Challenges and VC setups:*
  - *General*
  - *ANA-100*
  - *KEK-PNNL*
- Plan Moving Forward

Belle II Collaboration

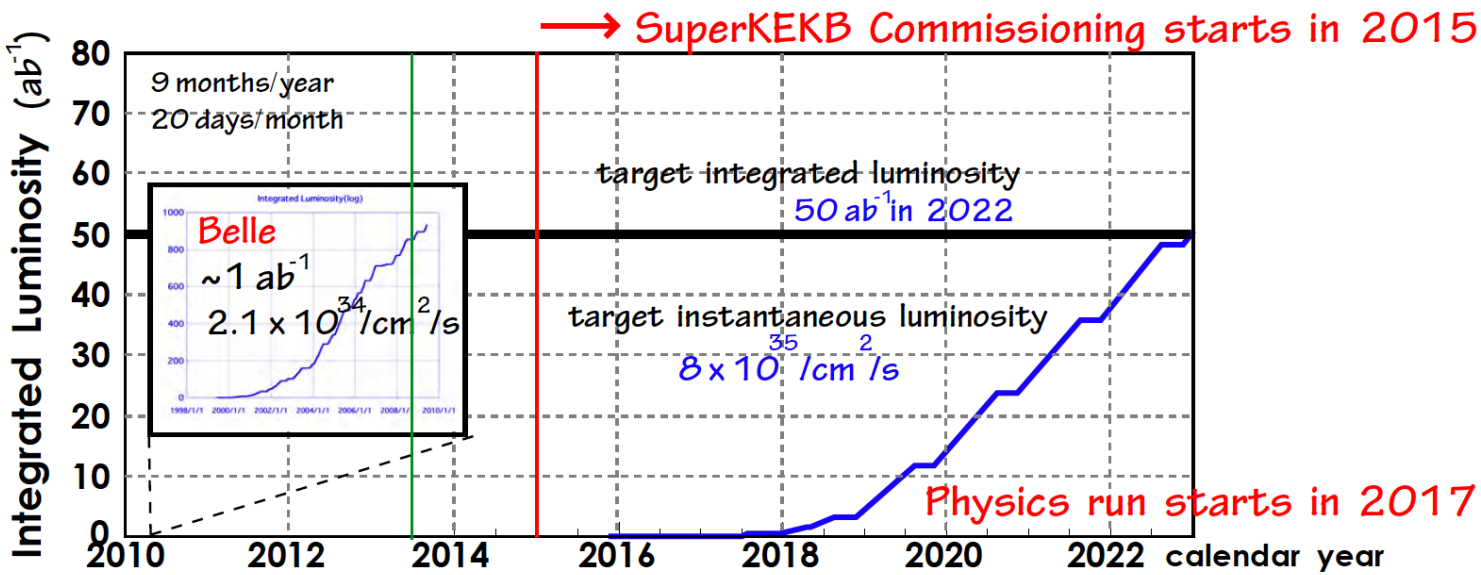
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Asia : ~45%
- Japan : 137
- Korea : 34
- Taiwan : 22
- India : 20
- China : 15
- Australia : 18

N. America : ~15%
- US : 63
- Canada : 17

Europe : ~40%
- Germany : 83
- Italy : 59
- Russia : 37
- Slovenie : 14
- Austria : 14
- Poland : 11

Belle II Luminosity and Data Rate



Experiment	Event size	Rate @ Storage	Rate @ Storage
	[kB]	[event/sec]	[MB/sec]
Belle II	300	6,000	1,800
ALICE (Pb-Pb)	50,000	100	4,000
ALICE (p-p)	2,000	100	200
ATLAS	1,500	600	700
CMS	1,500	150	225 (<~1000)
LHCb	55	4,500	250

(@ max. luminosity)

(LHC experiments : as seen in 2011/2012 runs)

Belle II Computing Model

- The Belle II Computing model will manage in a geographically distributed environment the following main tasks:
  - RAW data processing.
  - Monte Carlo Production
  - Physics analysis
  - Data Storage and Data Archiving

- On going activities
  - Resource Estimation
  - Define strategy for analysis and data distribution
  - Evaluating technologies

Site Categorization

- The Belle II Computing Sites are categorized as follow:
  - **Raw data Center:** Stores the RAW Data and data processing and/or data reprocessing.
  - **Regional Data Center**: Large data center that stores mDST and participates at the Monte Carlo production
  - **MC Production site**: Data Center that produces and stores Monte Carlo simulations, that included:
    - Grid Site
    - Cloud Site
    - Computing Cluster Site

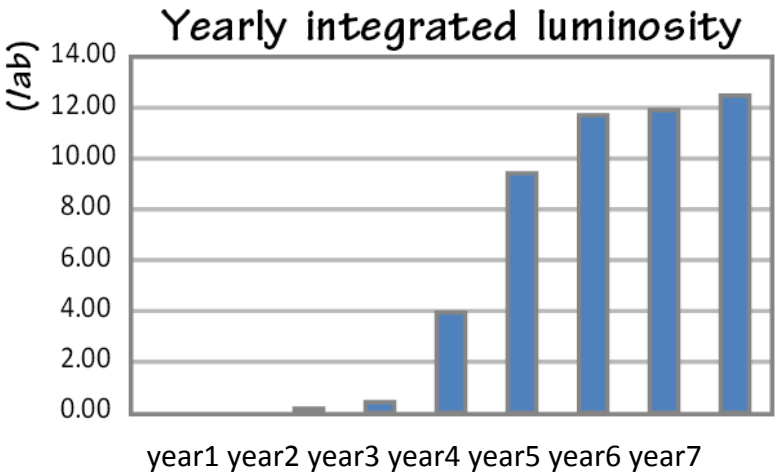
Data Volume Estimation

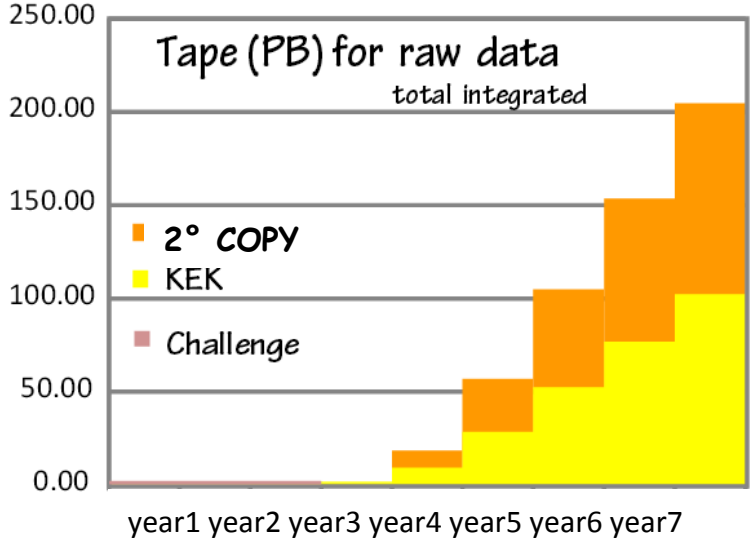
- Storage estimation based on:
  - RAW data
  - mDST after data taking
  - mDST during data reprocessing
  - mDST-Monte Carlo related the data
  - mDST-Monte Carlo related data reprocessing

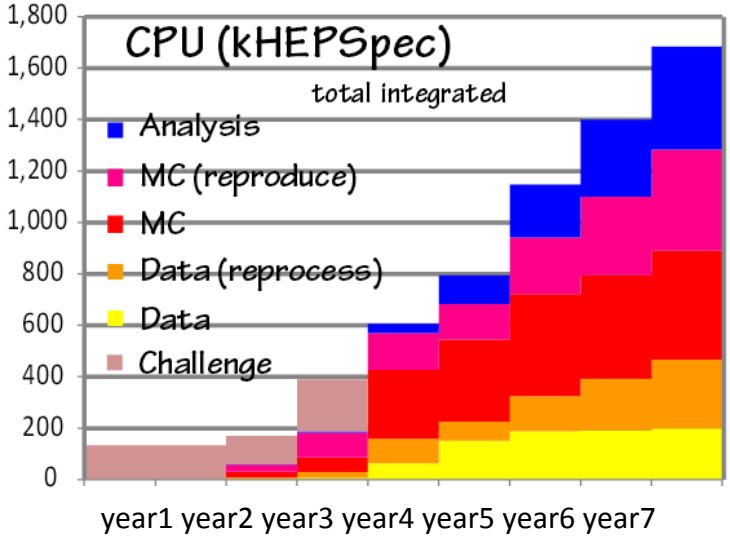
- The current parameters for data estimation are
  - Event Size x RAW data: 300Kb
  - Event Size x mDST : 40Kb

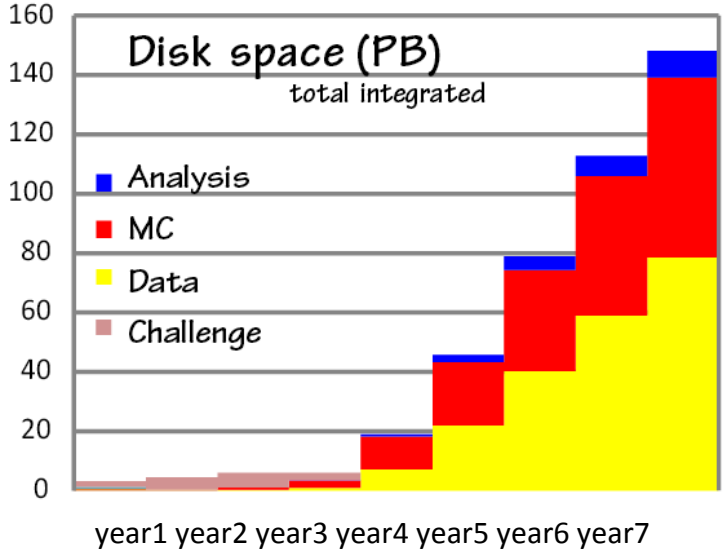
X10^9	Year1	Year2	Year3	Year4	Year5	Year6	Year7
Event/year	1.2	3.1	29.6	70.7	87.8	89.3	93.5
Integrated	1.2	4.3	33.9	104.6	192.4	281.7	375.2

Belle II Computing Resource Estimation

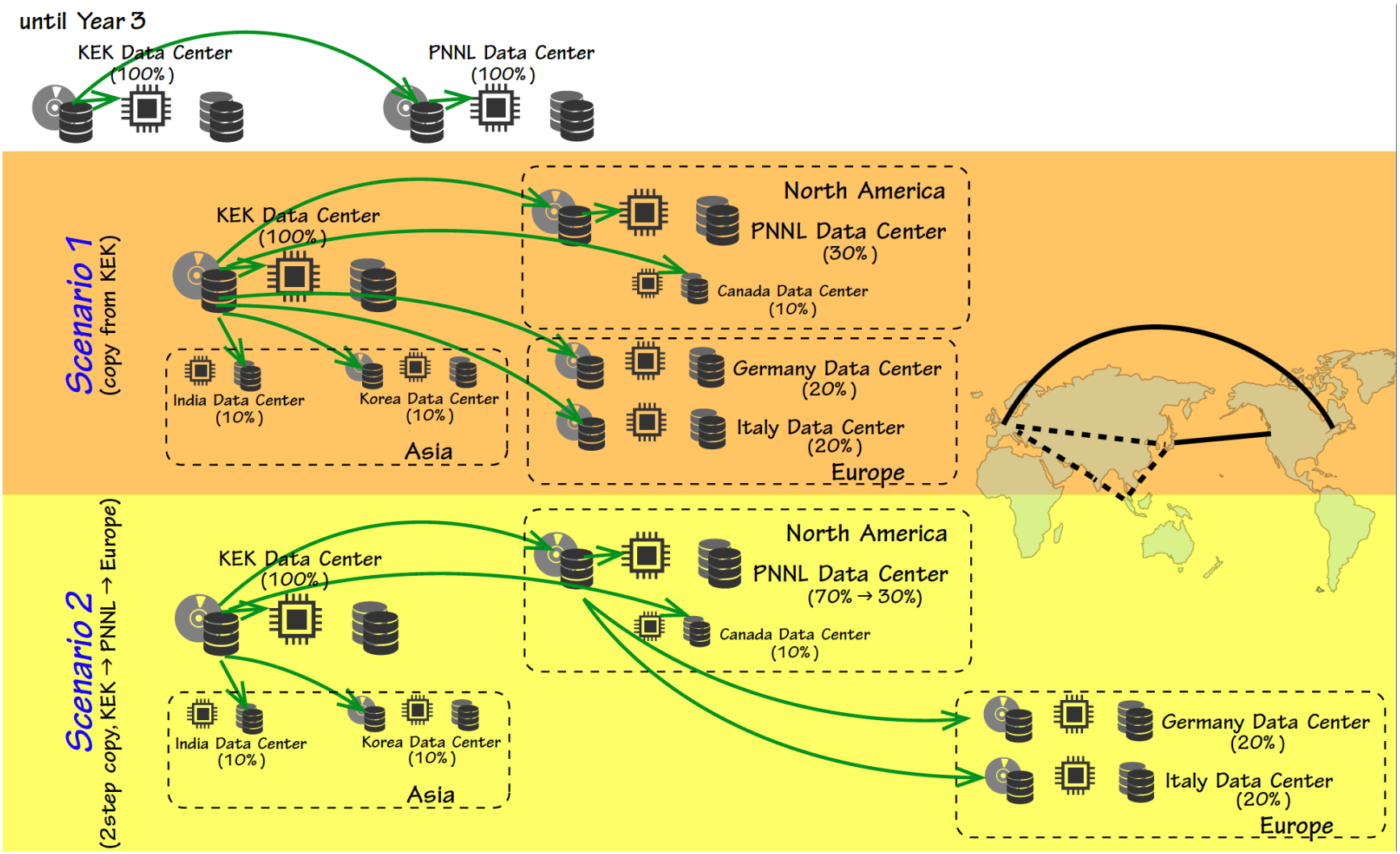








Belle II Raw Data Distribution

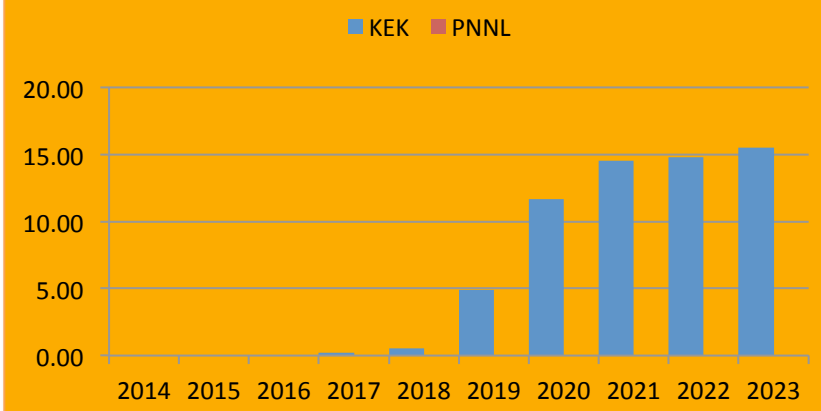


Belle II Raw Data Network Requirements

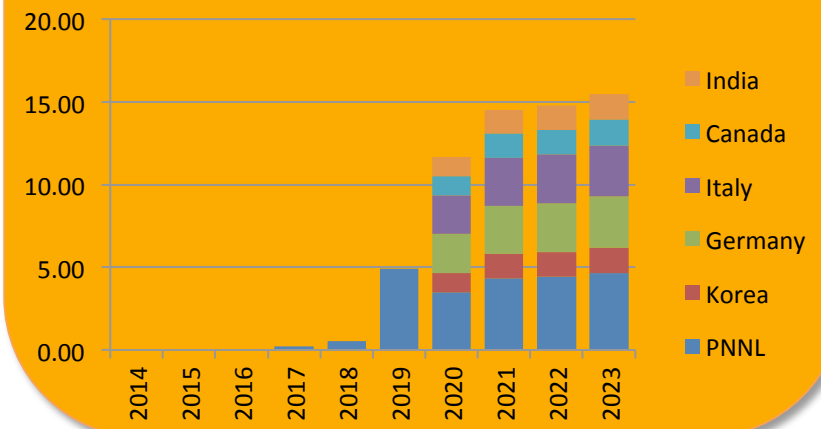
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Scenario 1

RAW Data - Outbound

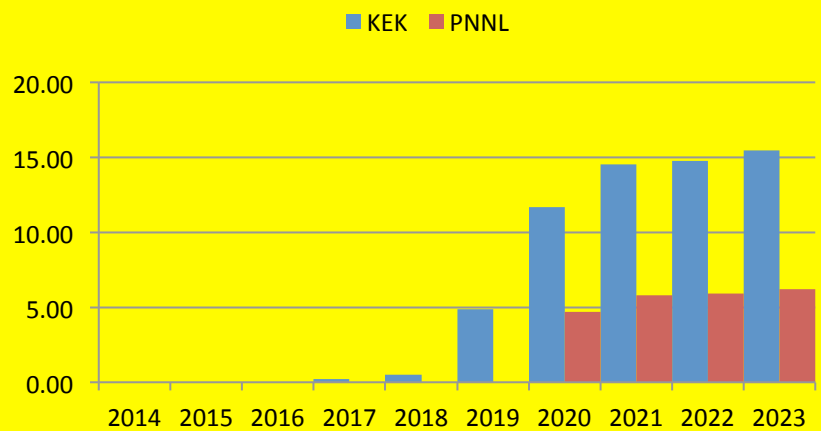


RAW Data – Inbound

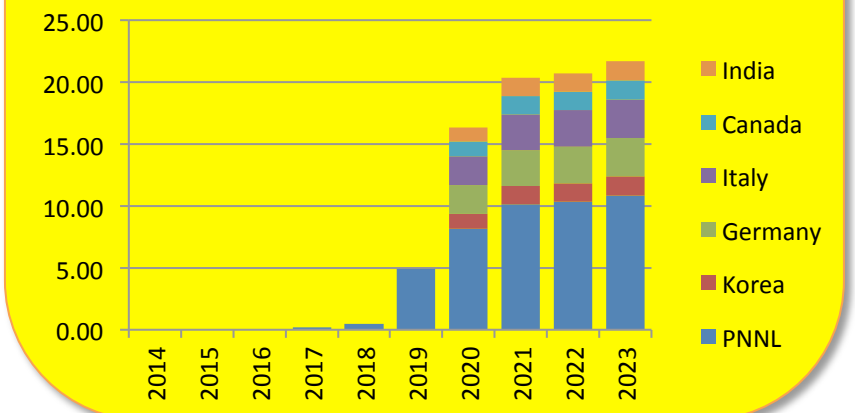


Scenario 2

RAW Data - Outbound

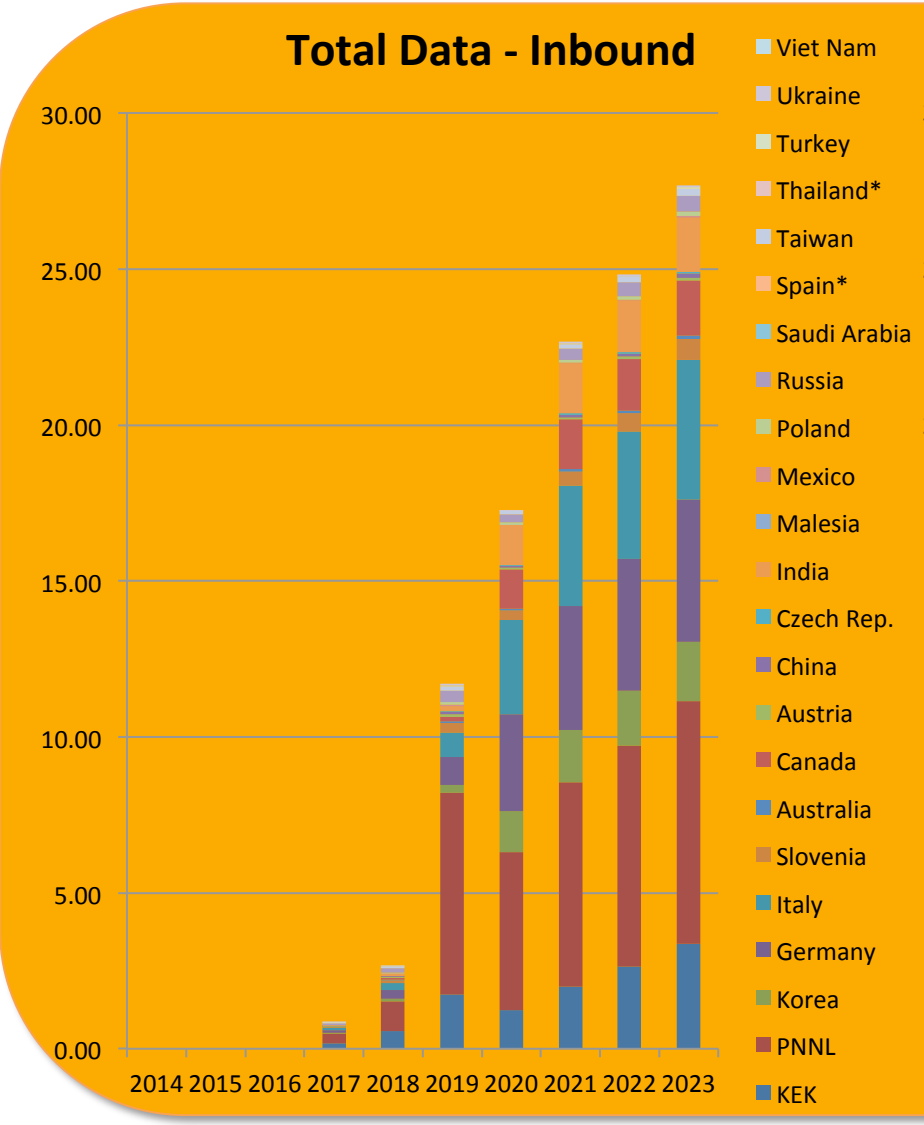


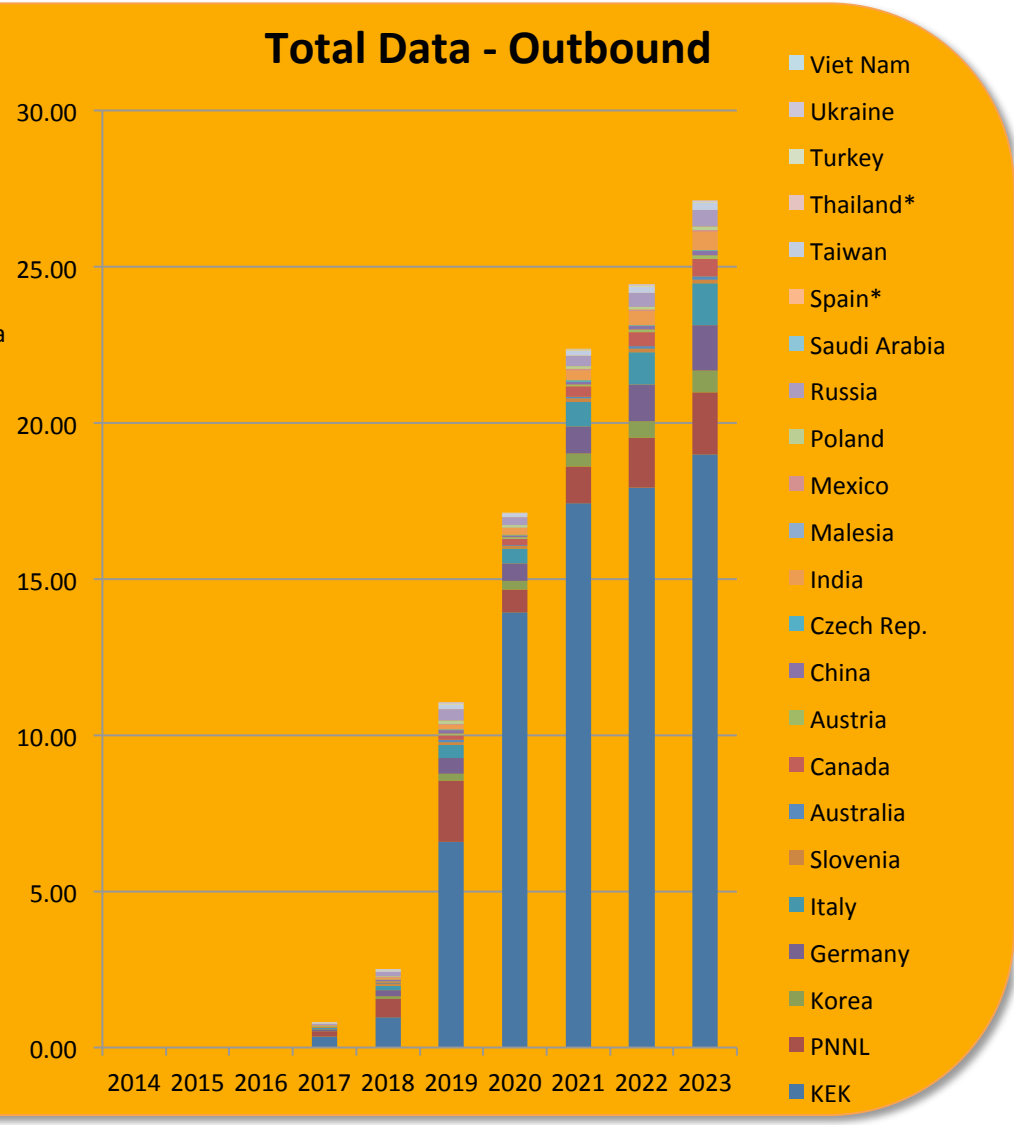
RAW Data – Inbound



Belle II Total Data Network Requirements Scenario 1

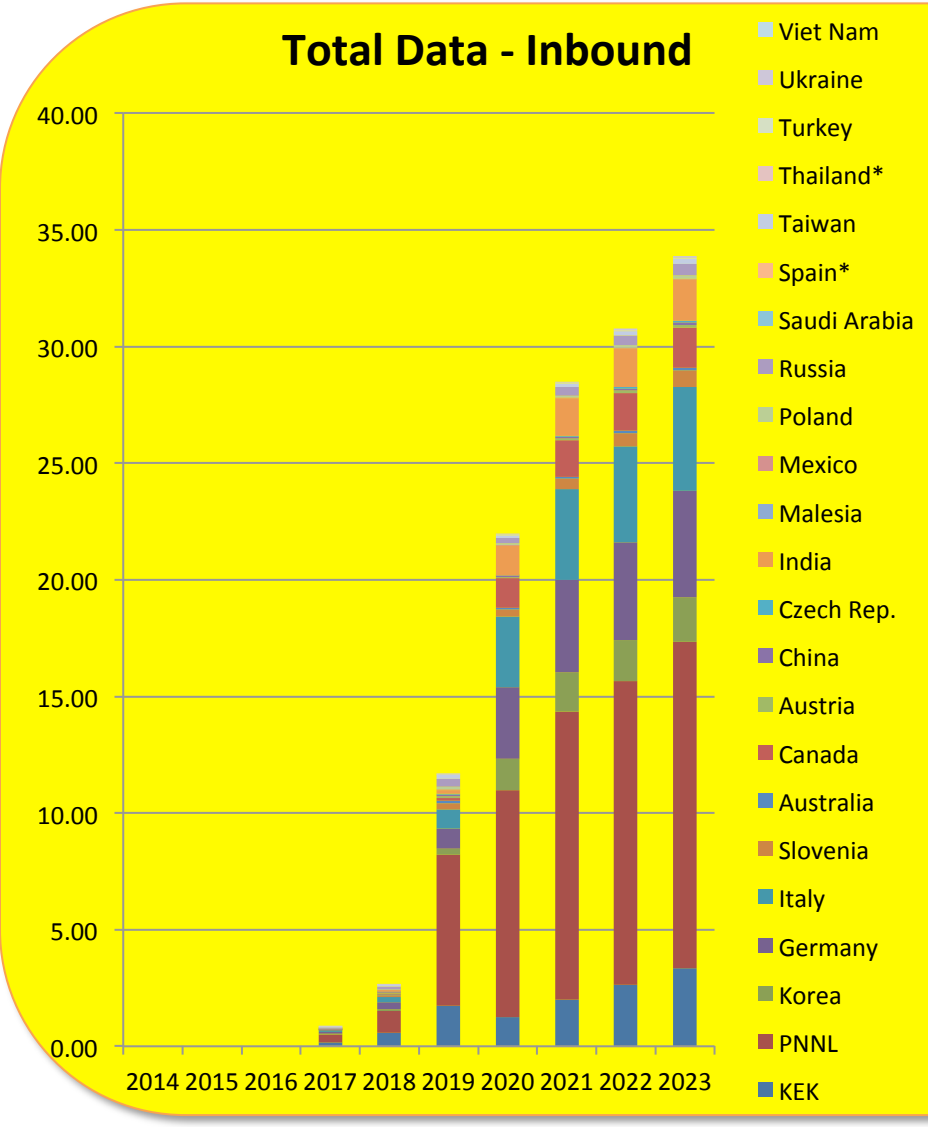
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

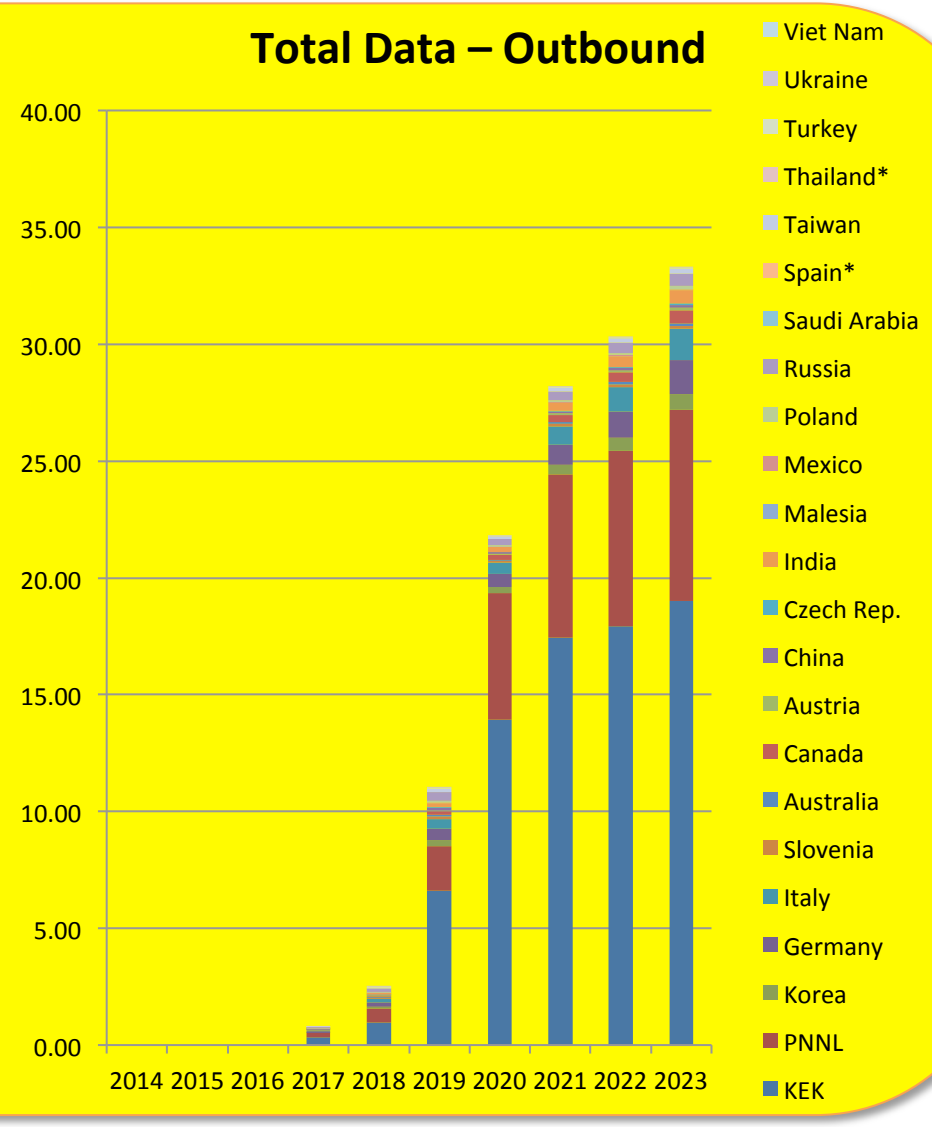




Belle II Total Data Network Requirements Scenario 2

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965





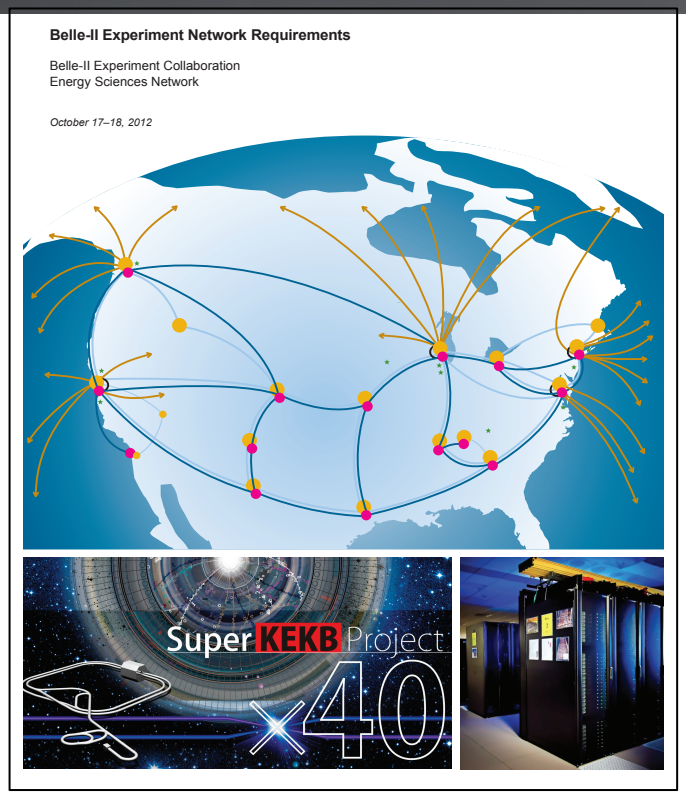
"Pacific Network and Computing Requirements" Workshop hosted by PNNL - Oct 17-18, 2012

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- The purpose of this workshop was to begin preparation for addressing the wide-area networking requirements for science in general and of the Belle II experiment in particular.

- Report can be found at:
  http://www.es.net/assets/pubs_presos/Belle-II-Experiment-Network-Requirements-Workshop-v18-final.pdf

- Various goals were defined, for example preliminary data challenge goals are:



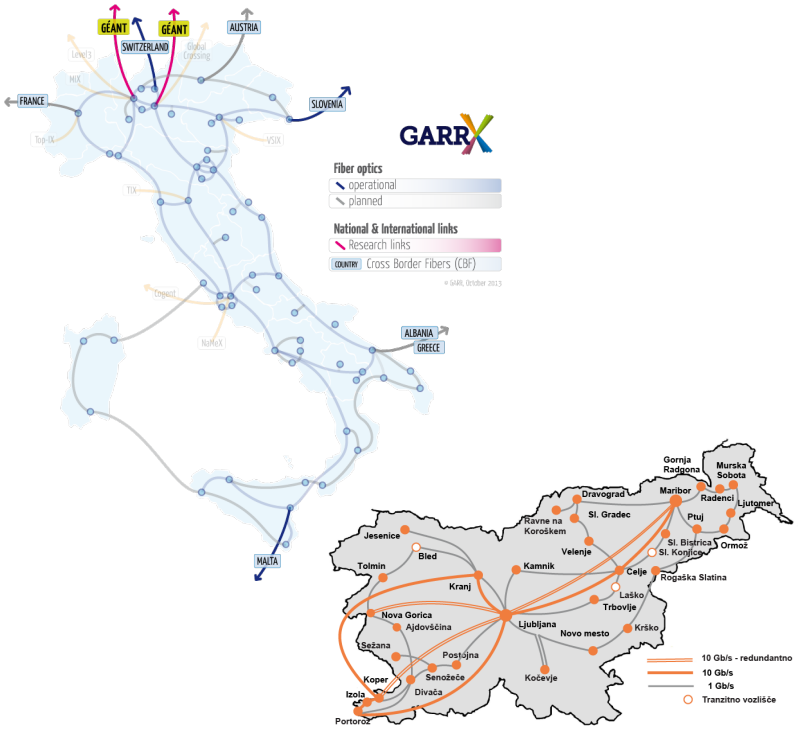
Date	Summer 2013	Summer 2014	Summer 2015
Rate	100MB/sec	400MB/sec	1000MB/sec
Duration	24 hours	48 hours	72 hours

European Networking Workshop hosted in Vienna – October 2013

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- The purpose of this workshop was to begin preparation for addressing the wide-area networking requirements for science in general and of the Belle II experiment in Europe.

- Report is ongoing, potential milestones:

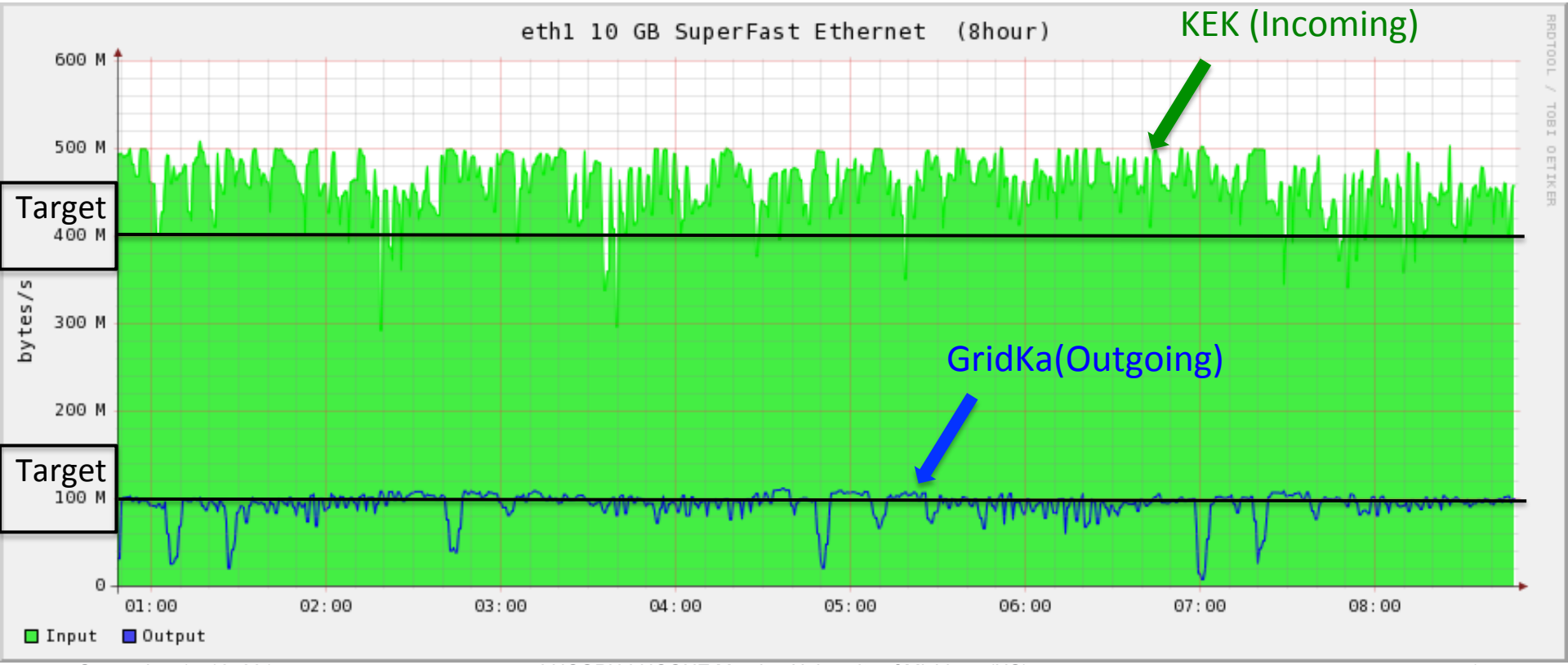


Date	Winter 2013	Summer 2014	Summer 2015
Rate	100MB/sec	200MB/sec	400MB/sec
Duration	24 hours	48 hours	72 hours

First Belle II Data Challenge
KEK to PNNL & PNNL to GridKa

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

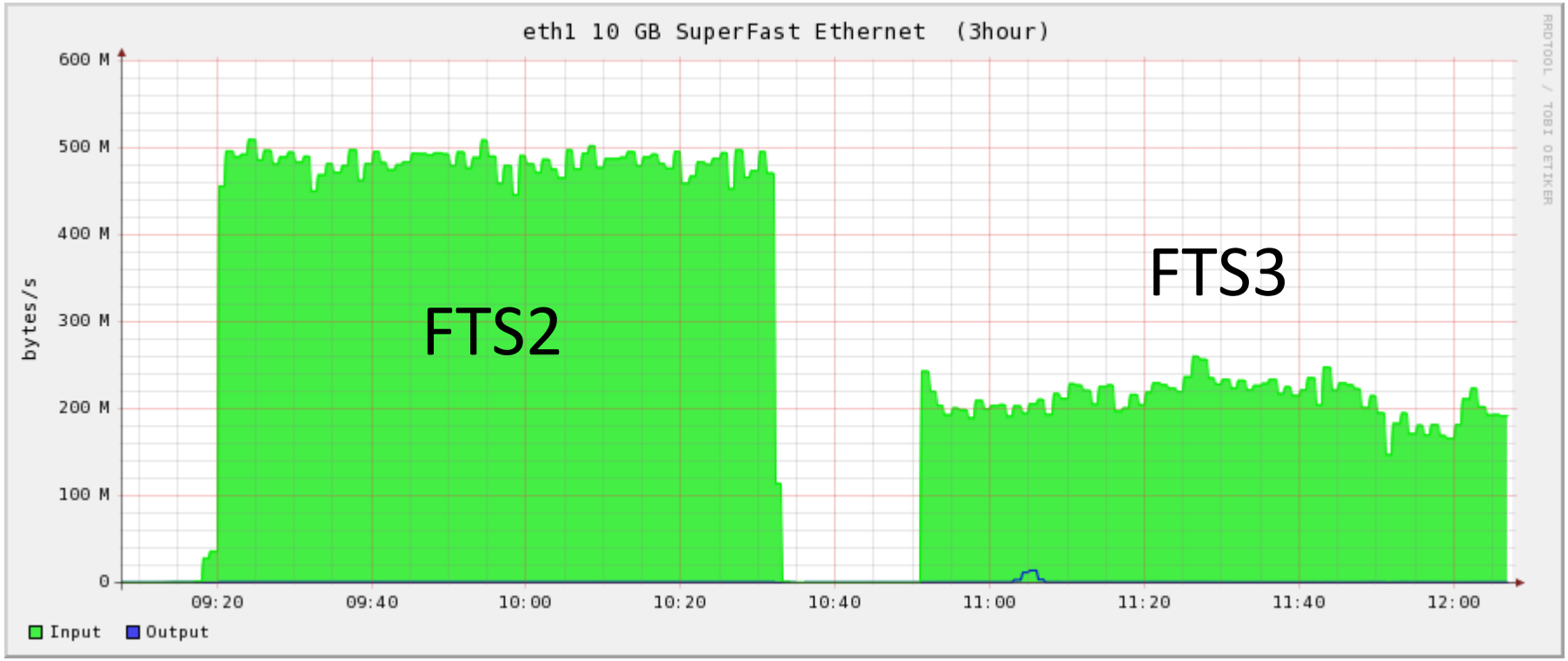
- Transfer rate from KEK to PNNL during 48hr stability test meet summer 2014 goal
- Transfer rate from PNNL to GridKa during 24 hr stability test was >100MBps (only 8hrs shown in figure).



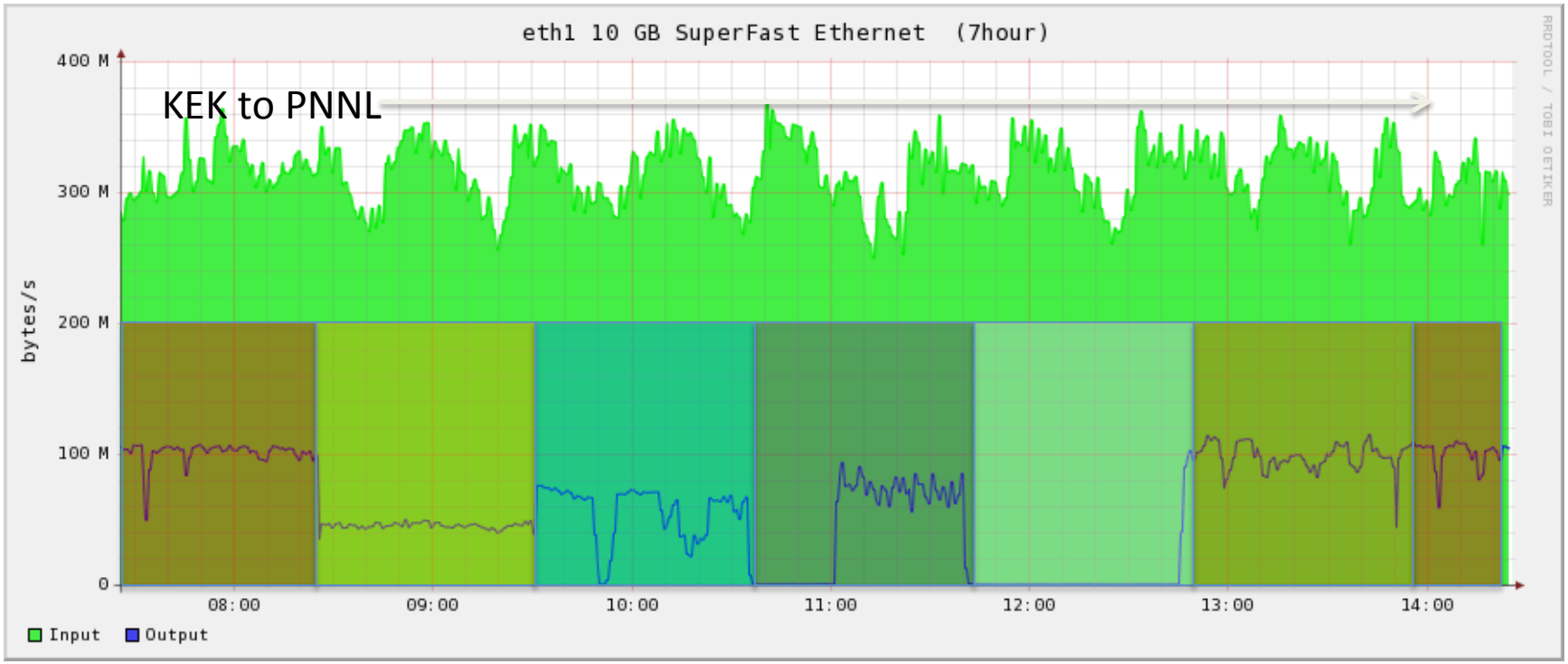
General Network Data Challenge

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Deploy new FTS3 server at PNNL
- Evaluating FTS2 vs. FTS3
- KEK to PNNL throughput for FTS3 was initially half that of FTS2
- Fine tuning FTS3 is ongoing



FTS3 Throughputs



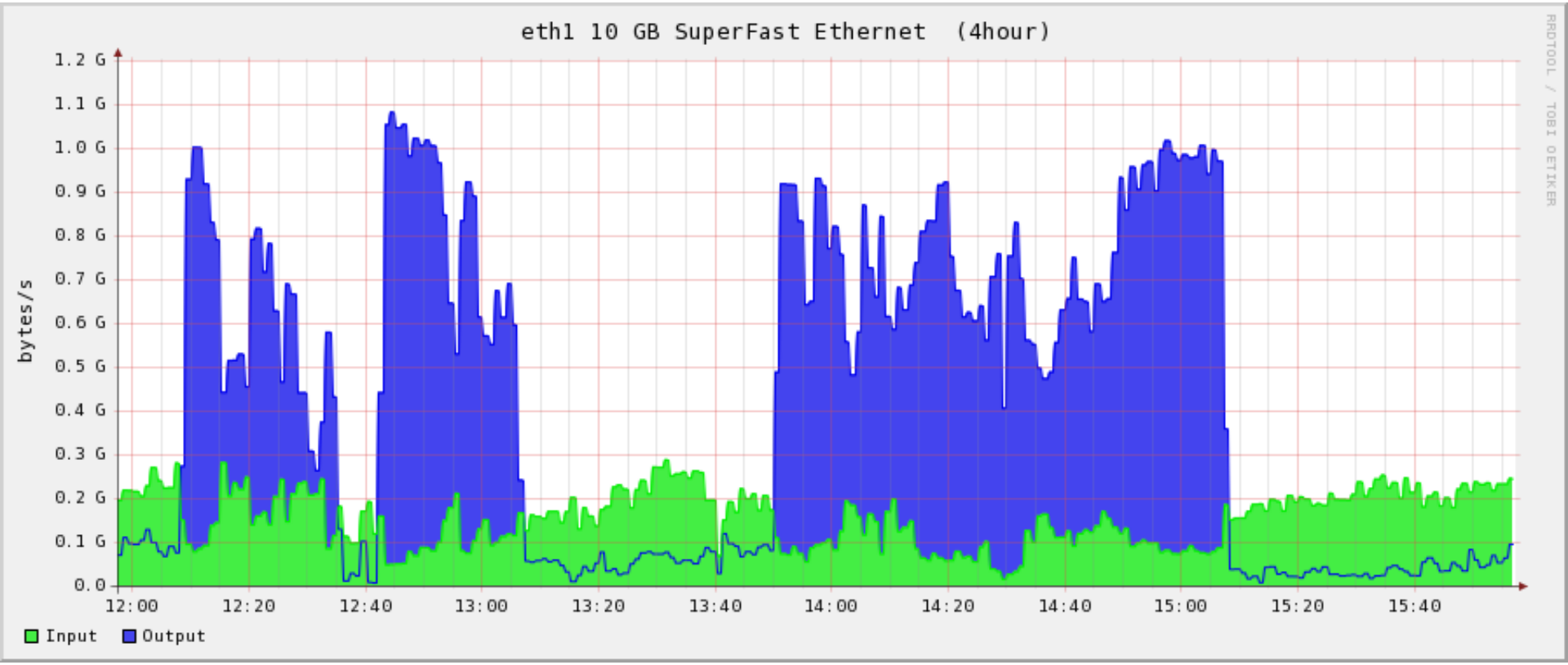
PNNL to …

- GridKa (Germany)
- DESY (Germany)
- SiGNET (Slovenia)
- U of Victoria (Canada)
- INFN Torino (Italy)
- U of Melbourne (Australia)

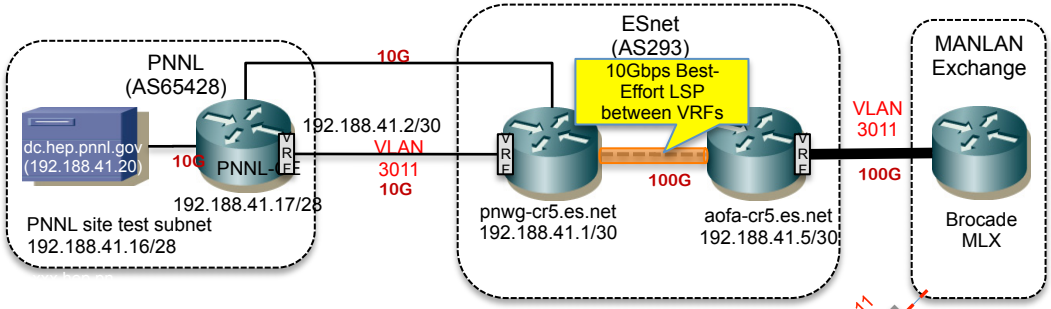
KEK to PNNL FTS3 Data Transfers

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Tuning FTS3 allows us to reach ~8.8 Gbps transfer rates from KEK to PNNL
- Working on cloud implementation to provide load balanced FTS3 services



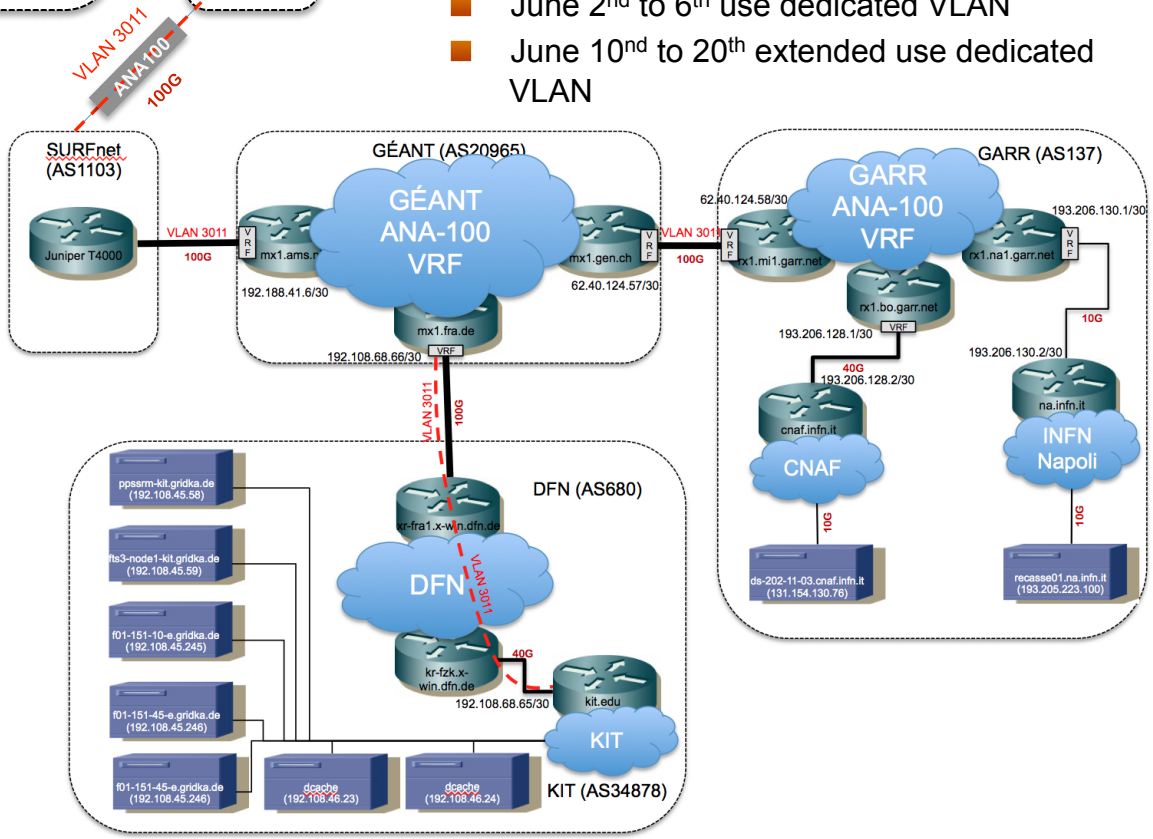
Belle II ANA-100 Setup



- Goal
  - Test the ANA-100 Trans-Atlantic link
  - Test/tune/profile the performance of current Belle II data transfer tools
- Dates:
  - June 2nd to 6th use dedicated VLAN
  - June 10nd to 20th extended use dedicated VLAN

- Network Setup:
  - Network providers (Geant, ESnet, GARR, DFN, etc.) setup the VLAN
  - Local network providers and sites coordinated final configurations
  - Sites must configure hardware interface to match destinations

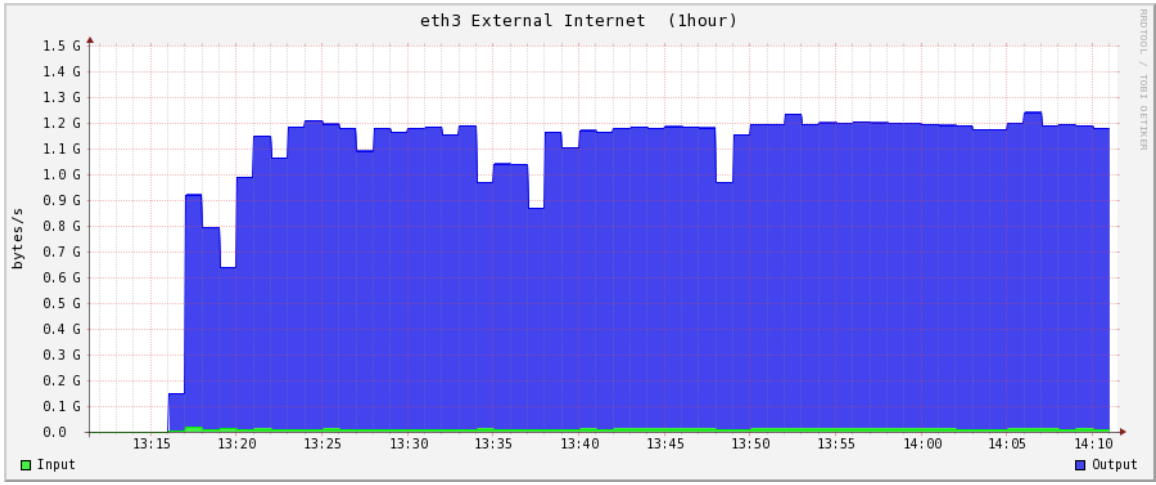
- Testing Tools:
  - *Traceroute* was used to confirm the routing to each DTN
  - *Iperf* was used to do initial network transfer rate tests
  - *gridftp* and/or *srm-copy* was used to test site
  - *FTS3 server* at GridKa was used to schedule data transfers

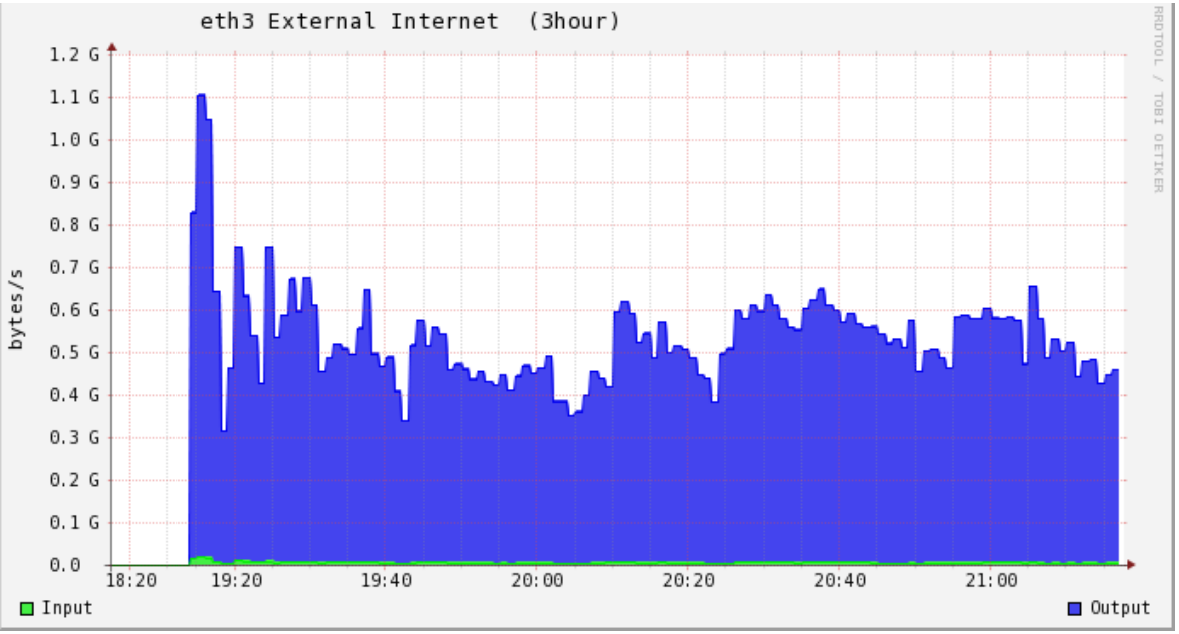


Belle II ANA-100 Results

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- First few days we conducted tests using *iperf* for true network testing
- Required several parallel transfers to reach network saturation
- Reached ~9.6GBps (>2x the Tier-1 EU site requirements)



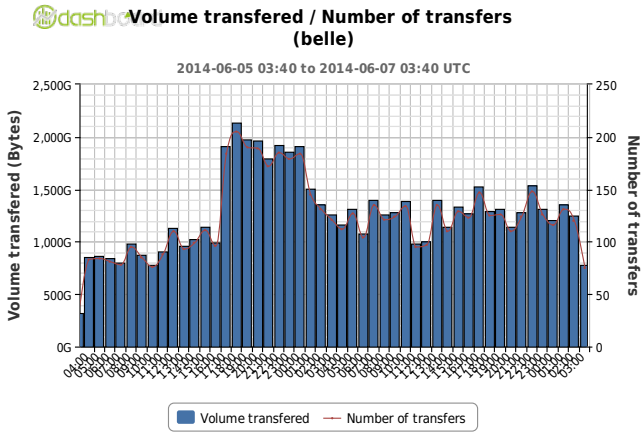


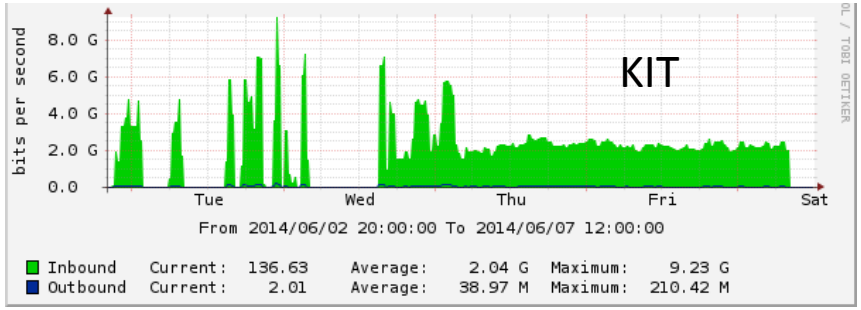
- Transitioned to FTS3 server
- Reached network saturation but rates fell very quickly
- Large amount of drop packets
- Satisfies the incoming network requirements for Tier1 EU sites up to calendar year 6

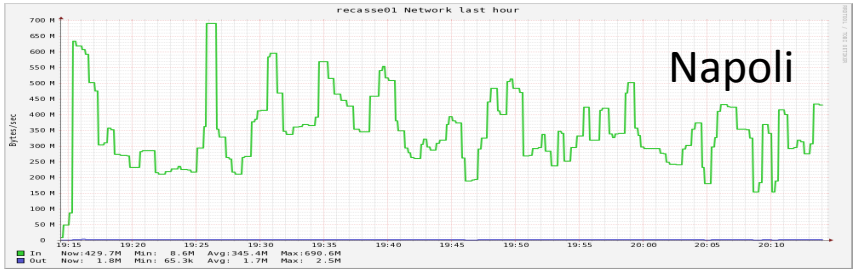
Lessons Leaned

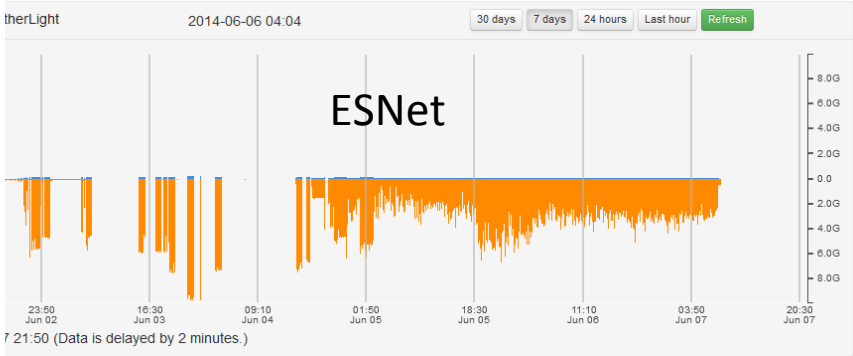
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Challenges encountered:
  - The main issue was the configuration of the local network apparatus.
  - Having all the servers at each site using/checking the proper network route
  - Hardware limitation (router, storage, etc.)
  - Not having dedicated setups (shared with ATLAS, etc.)
- Modification to sites to accommodate the increased rates:
  - Modification of TCP windows was performed at PNNL and Italy
  - Routing hardware interface
  - Configure/tune network interrupts for multicore
  - Modification of the FTS3 optimization & global-timeout



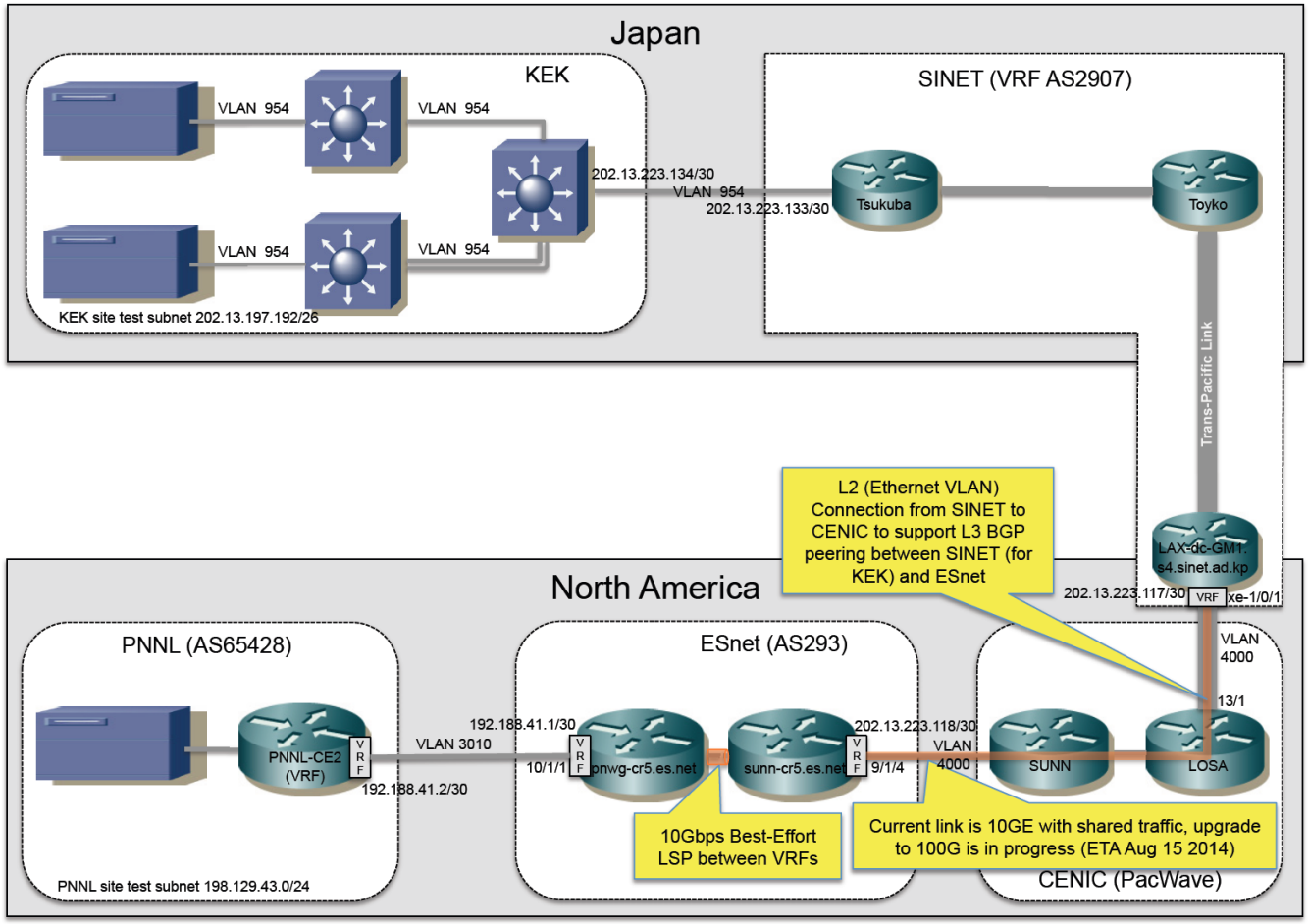






Belle-II Testing between KEK and PNNL
(Setup to stay in place thru 30 June 2016)

- The KEK-PNNL VC and endpoints are configured
- Iperf tests were performed yielding a 1-3Gbps throughput
- Additional testing required
- Ideally, we would like to setup a gridftp server on KEK PC and start FTS3 transfers



Plan Moving Forward

- LHCONE is for LHC experiments
  - Canadian and European sites are already part of LHCONE
  - KEK and PNNL, key sites, are not part of LHCONE
- Belle II thoughts and consideration:
  - Belle II would like to have a closed network similar to LHCONE
  - Configure LHCONE-like VRF layer?
    - Complicates configurations and operations for sites that are already part of LHCONE
  - Can Belle II join LHCONE?
    - Is it difficult to expand LHCONE to non-LHC experiments?
    - Belle II traffic would be shared/compete with LHC experiment?
    - Easier to coordinate sites under one umbrella
- Your comments/suggestions are invaluable to find the best solution!