

NDN-Hadoop: Exploring Applicability of NDN for Big-Data Computing

Datacenter Networks Are Challenging



- Densely connected
- Lots of engineering to maximize performance before hitting limits
- Alternative protocols may simplify system and applications



Telephony: name the path

IP: name the endpoint



Name Data Network (NDN): name the data

- Developed for the Internet
- Efficient multicast
- In-network caching
- Failure recovery



NDN Benefits for Big Data

- NDN provides network optimizations • Simplifies development of Big Data applications
- NDN provides data multicast
 - Benefits data replication in Big Data storage
- NDN provides in network caching
 - Improves performance in Big Data computing
- NDN retrieves data from alternate sources
 - Transparent failure recovery in Big Data systems



Chris Gniady and Beichuan Zhang **Department of Computer Science, University of Arizona** gniady@cs.arizona.edu, bzhang@cs.arizona.edu

Exploring NDN in a Datacenter





Multicasting in Datacenter



Class	Classification	\pm Lines of Code	Description
ipc.Server	Hadoop-specific	+12	RPC server
DataXceiver		-100	Core class for data transfer
DFSInputStream		-4	Core input class
DFSOutputStream		-126	Core output class
DFSOutputStream.Packet		-150	Removed
All other classes		-482	28 other classes modified
NDNSocket	Glue code	+476	Generic wrapper class
NDNServerSocket		+137	Generic wrapper class
NDNBufferConsumer	NDN transport layer	+159	Base NDN data receiver
NDNBufferProducer		+186	Base NDN data sender

- 800 lines of code eliminated
- Several corner cases

Simpler Code

 Minimal changes in current code for comparison • Further code simplification in optimized applications

More Efficient Replication

