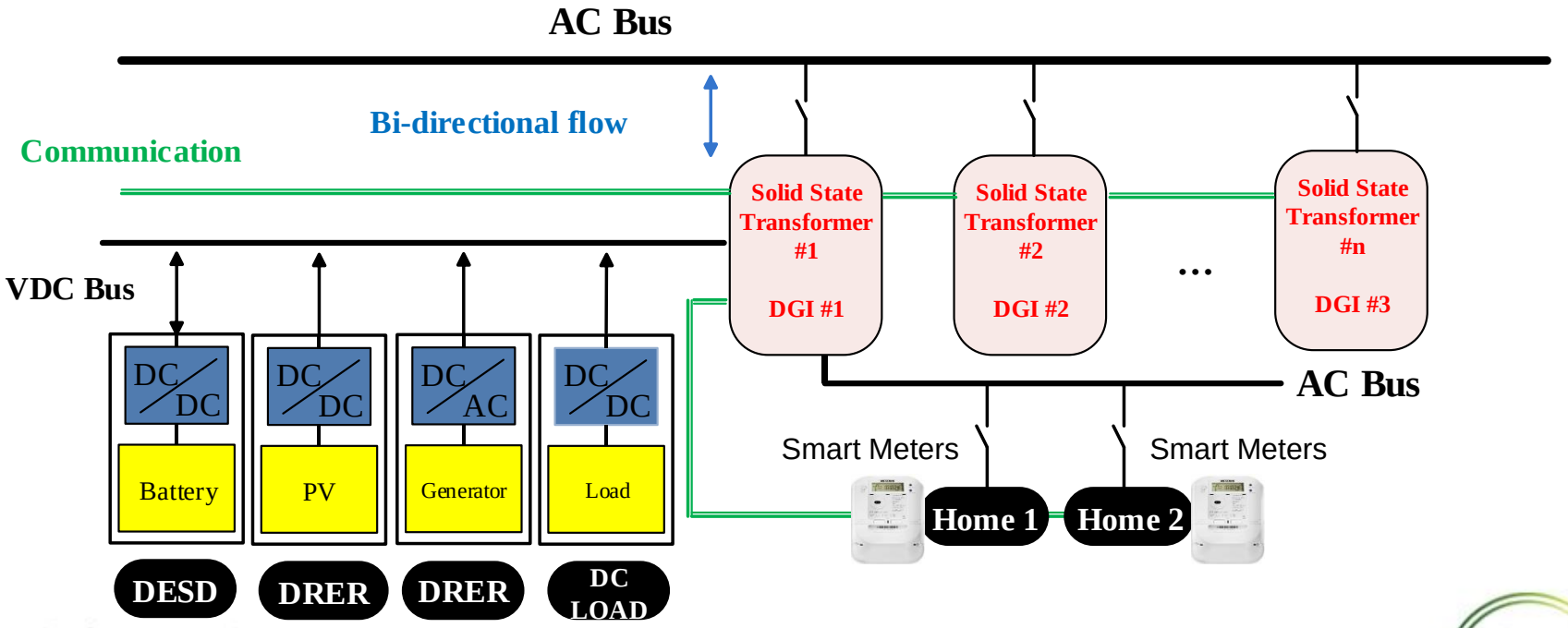
A High-availability and Fault-tolerant Distributed Data Management Platform for Smart Grid Applications

Ni Zhang, Yu Yan, and Shengyao Xu, and
Dr. Wencong Su
Department of Electrical and Computer Engineering
University of Michigan-Dearborn

Objectives

- To investigate a radically different approach through ***distributed software agents*** to store and process a massive amount of smart grid data in a timely and reliable fashion
- To substantiate the proposed distributed data management systems for smart grid implementation on a ***proof-of-concept demonstration***

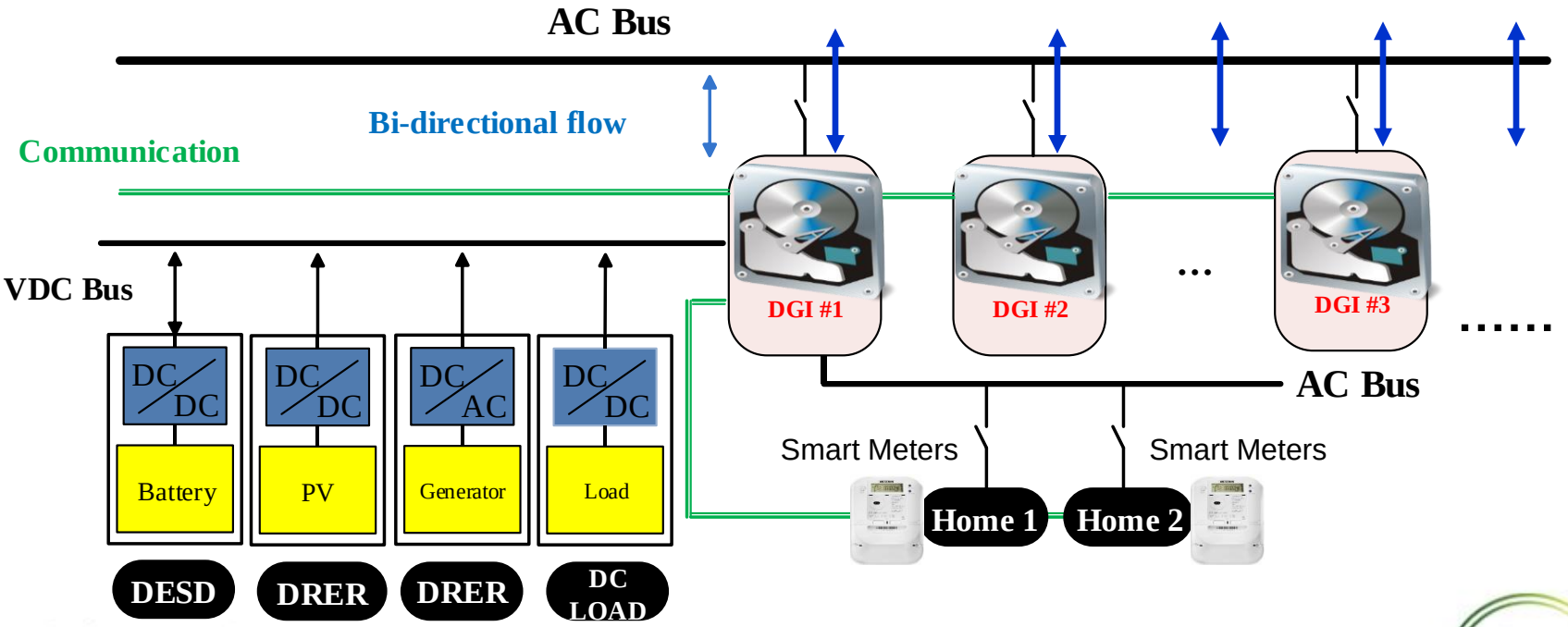


DGI: Distributed Grid Intelligence

Major Challenges

- High-volume
- High-velocity
- Expensive
- Single point of failure
- Not easy to expand
- Highly complex

Centralized Data Center



DGI: Distributed Grid Intelligence

galvin center for electricity innovation at ILLINOIS INSTITUTE OF TECHNOLOGY

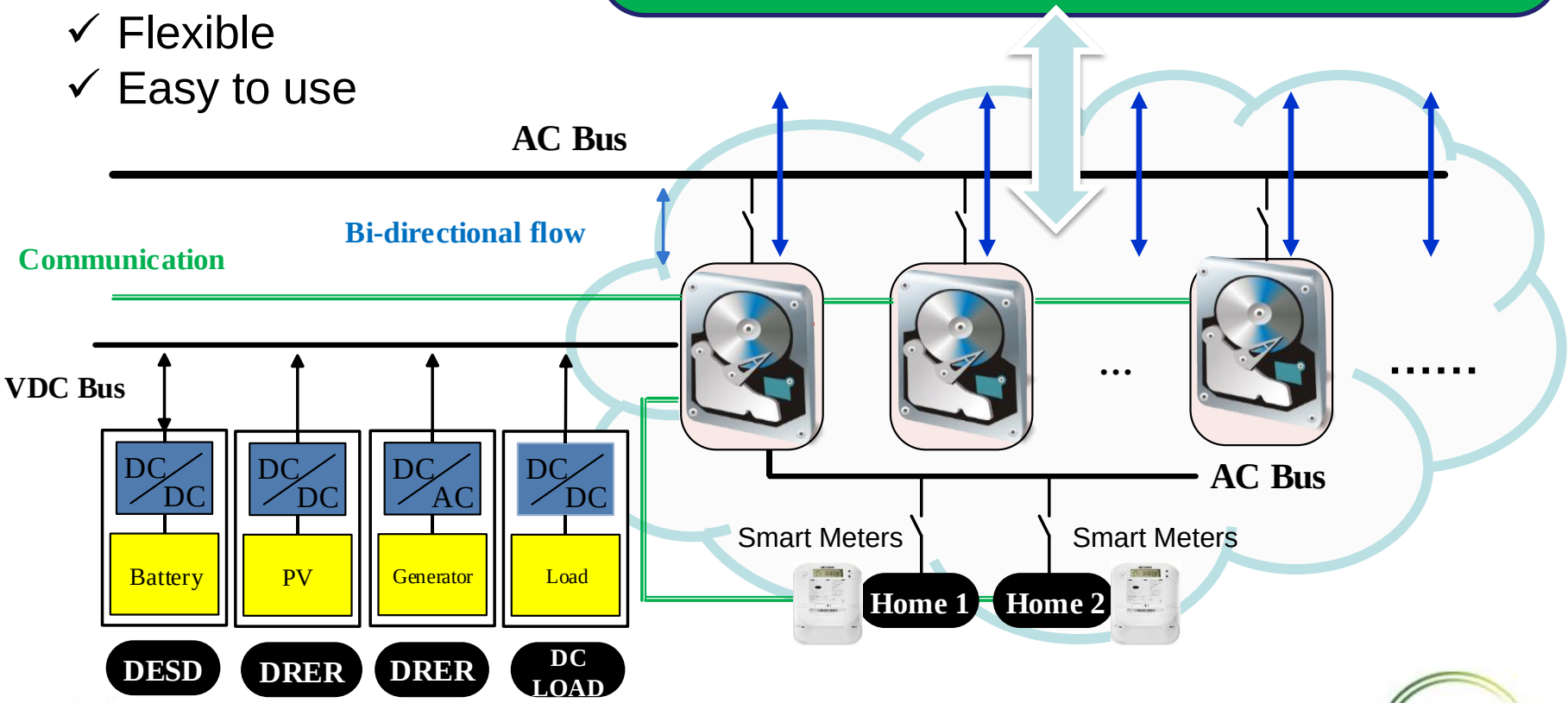
IEEE PES Power & Energy Society®

Distributed Data Storage and Processing

- ✓ Cost-effective
- ✓ Scalable
- ✓ Fast
- ✓ Resilient to failure
- ✓ Flexible
- ✓ Easy to use

Operation View

1. Data Collection: Open Standard and Plug-and-Play
2. Distributed Data Storage and Management
3. Open-source Programming & Open API



DGI: Distributed Grid Intelligence

galvin center for electricity innovation at ILLINOIS INSTITUTE OF TECHNOLOGY

IEEE PES Power & Energy Society®

Proof-of-Concept Demo

DGI Layer

A cluster of low-cost single board PCs (e.g., CubieBoard)

- High-availability distributed object-oriented platform
- Google's MapReduce and Google File System

. . . . . .

Smart Meter Layer

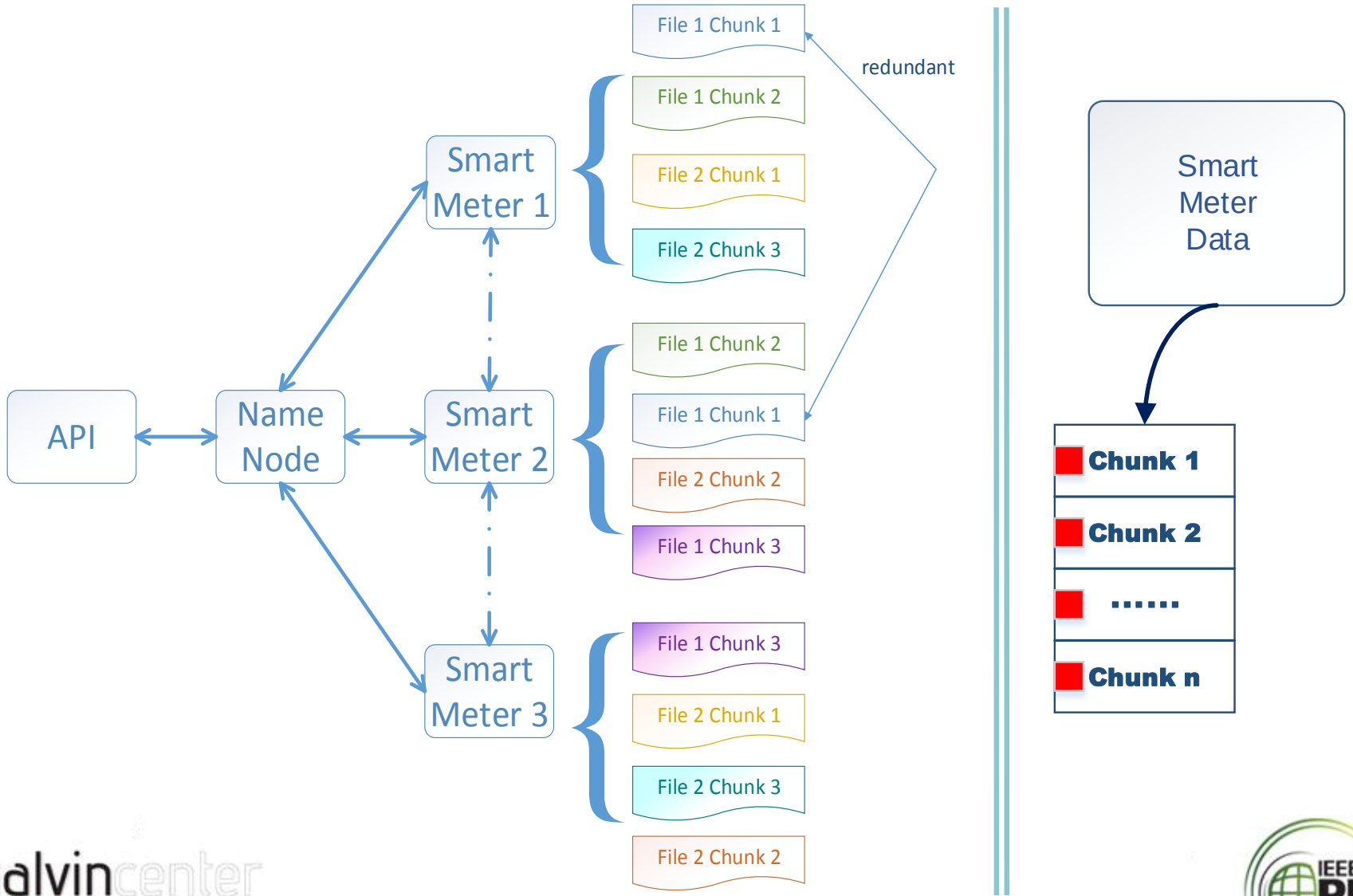
. . . . . .

. . . . . .

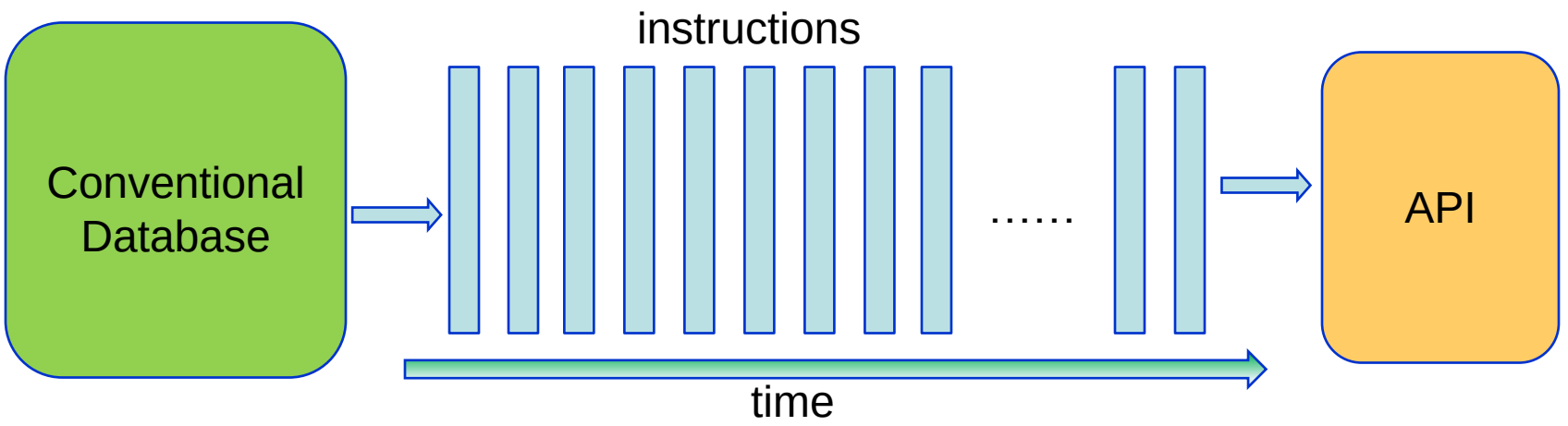
. . .

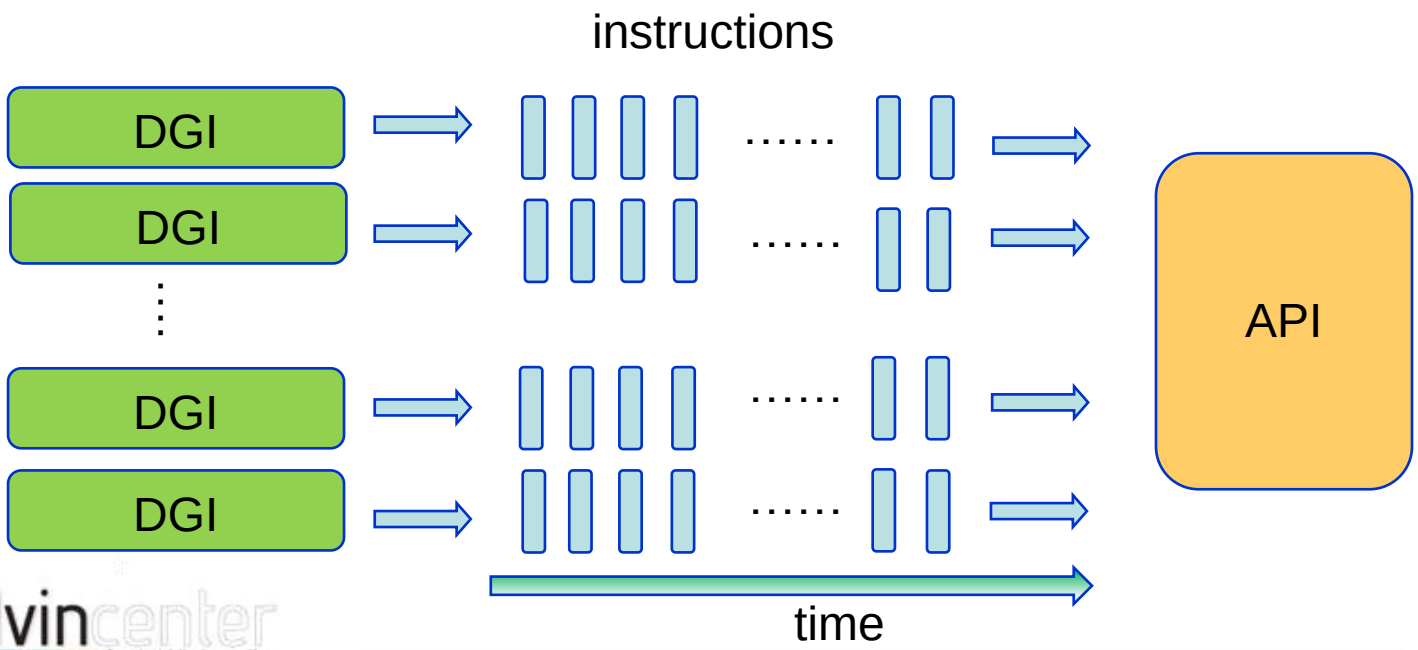
Simulated smart meter data is implemented into Lego Mindstorm EV3 Intelligent Brick (ARM 9 processor with Linux-based operating system)

Hadoop Distributed File System (HDFS)



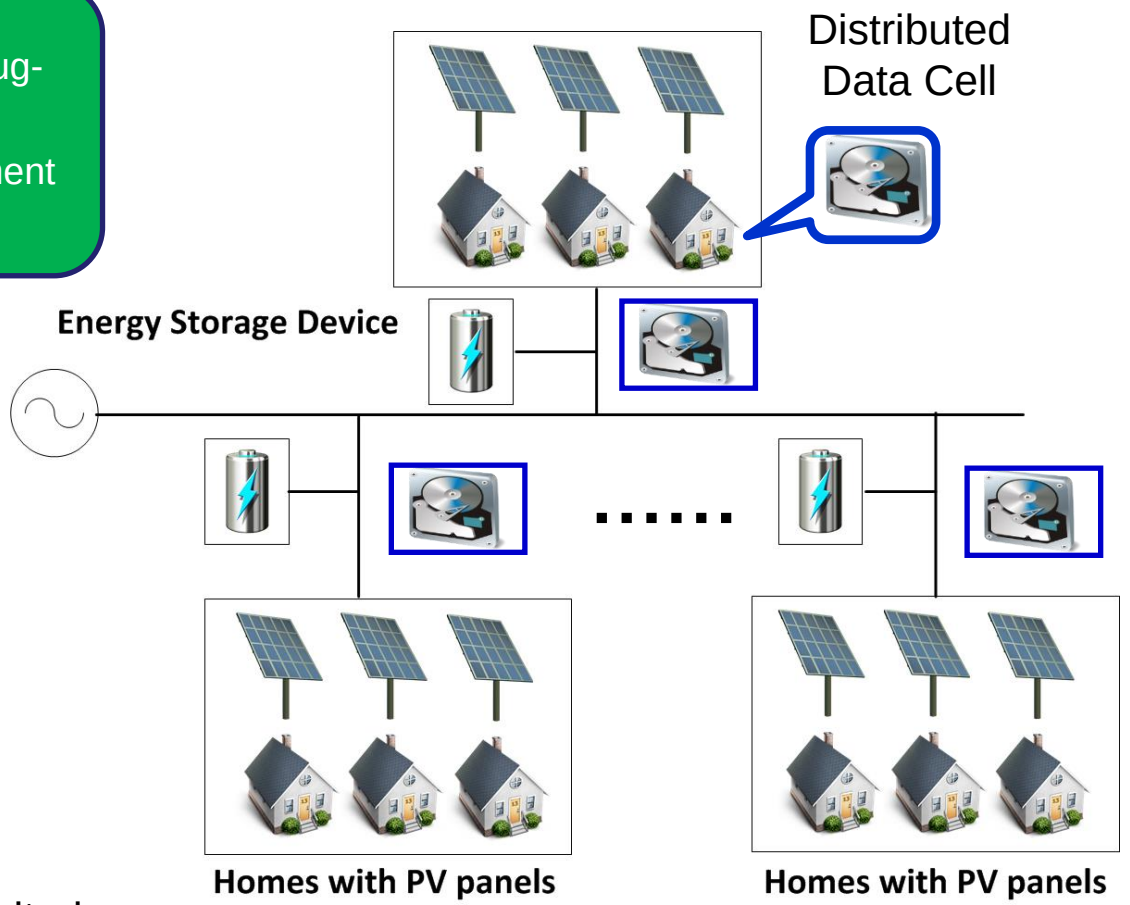
Parallel Computing





Operation View

1. Data Collection: Open Standard and Plug-and-Play
2. Distributed Data Storage and Management
3. Open-source Programming & Open API



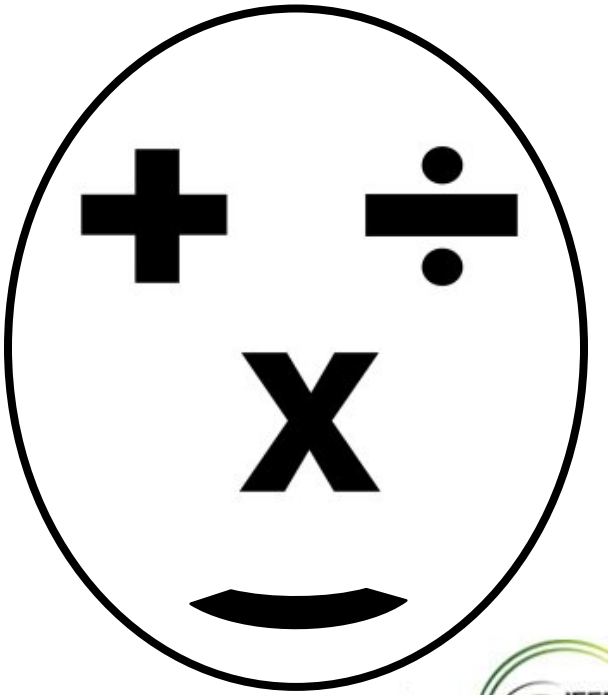
Centralized Data Center

- Real-time fault and performance monitoring
- Distributed data storage and processing
- Automated altering system (e.g., battery state-of-charge, inverter failure, device disconnection, voltage sags/surges, low performance yield)

All-in-one Solution

Scalable **Reliable** **Efficient** **Low-cost**

A Proof-of-Concept Demo



Plug-and-Play

Highly Scalable Data Storage System:

- Easy to add commodity servers and disks to scale up storage
- Bandwidth scales linearly with the number of DGI nodes and disks
- Fast to rebalance the entire data file

Cluster Summary

263 files and directories, 252 blocks = 515 total.

Configured Capacity	:	1.69 TB
DFS Used	:	17.57 GB
Non DFS Used	:	126.76 GB
DFS Remaining	:	1.55 TB
DFS Used%	:	1.01 %
DFS Remaining%	:	91.68 %
Live Nodes	:	12
Dead Nodes	:	3
Decommissioning Nodes	:	0
Number of Under-Replicated Blocks	:	0

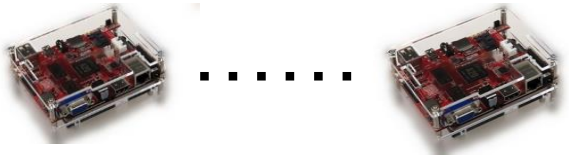
Cluster Summary

258 files and directories, 243 blocks = 501 total.

Configured Capacity	:	1.69 TB
DFS Used	:	17.46 GB
Non DFS Used	:	126.76 GB
DFS Remaining	:	1.55 TB
DFS Used%	:	1.01 %
DFS Remaining%	:	91.68 %
Live Nodes	:	15
Dead Nodes	:	0
Decommissioning Nodes	:	0
Number of Under-Replicated Blocks	:	0

Reliable

API #1:
Find abnormal values
59.8 < frequency < 60.2

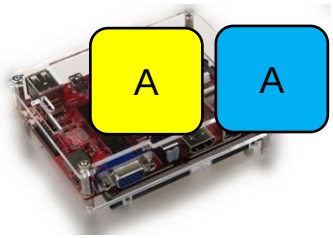


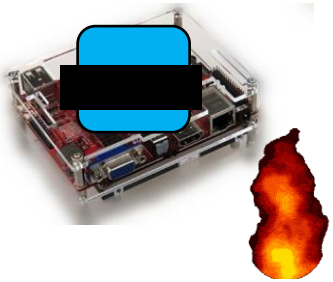
DataNodes (smart meters)

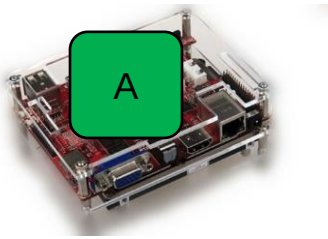
"Replication Factor" = 3

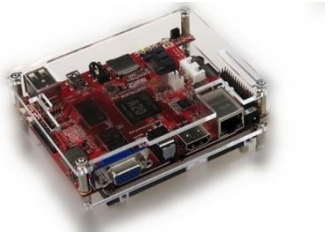
Data Format: ID; Voltage; Current; Temp; Frequency

234,000,000 rows ~10GB









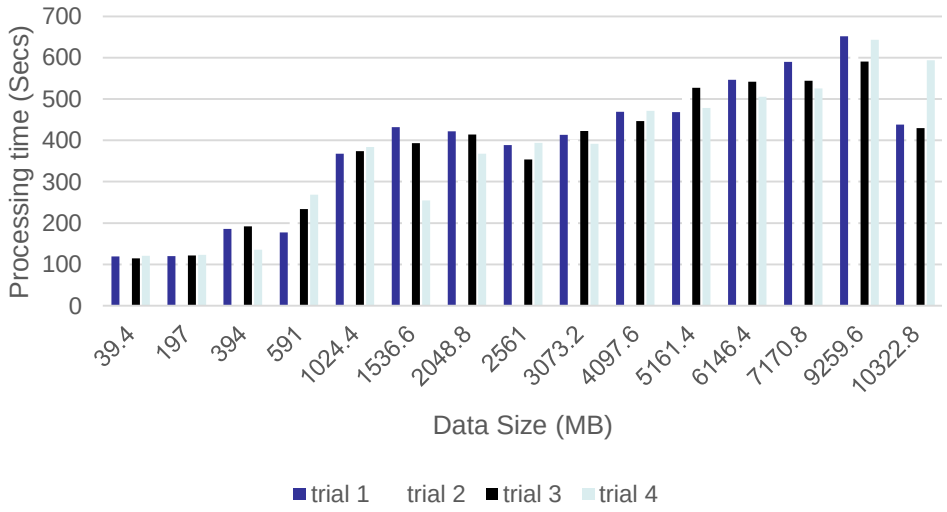
```
hive> select * from problem limit 10;
OK
2       109.994 21.3076 28.3818 59.7789
3       110.088 18.6367 24.1194 59.7355
6       110.395 21.5609 29.2089 60.2764
8       109.763 21.9666 23.5411 59.657
10      109.821 18.8199 27.3487 60.3424
13      110.346 18.4925 30.3939 60.2807
14      109.813 19.466  23.1061 60.2381
16      110.092 21.2459 27.1855 60.2211
17      110.03  20.2868 23.5379 59.6388
19      109.755 21.1212 31.0692 59.7266
Time taken: 0.077 seconds, Fetched: 10 row(s)
```

Reliable and Fault-tolerance Data Storage:

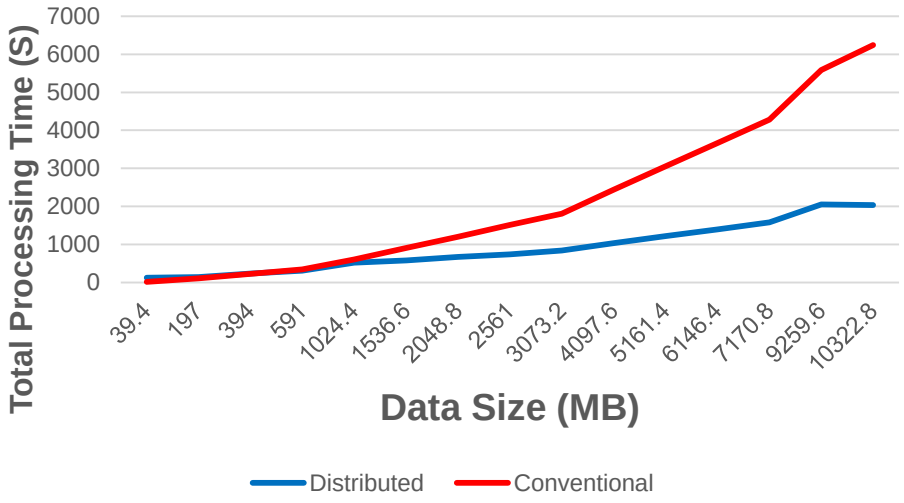
- ✓ No need to repair a failed DGI node immediately
- ✓ The entire data file is always accessible

Efficient

- Allows parallel data processing over a large amount of smart grid data
- Reduce overall processing time through parallel & faster execution



Total processing times using the distributed data management systems over different data size and trials

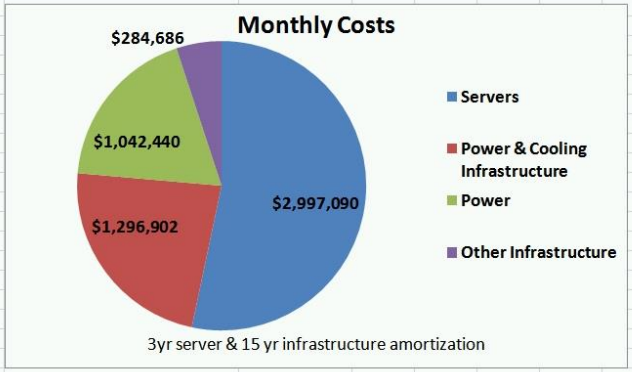


Comparison on average data processing time using distributed and conventional data management systems

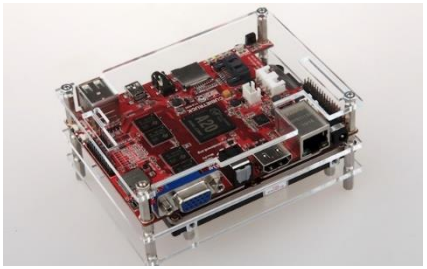
Low-Cost

Conventional database replies on very expensive, proprietary hardware and different systems to store and process data.

Assumptions			
	Cost of power ($/kwh):	$0.07	
	Cost of Facility ($):	$200,000,000.00	
	Facilities Amortization:	180	(15 years)
	Number of Servers:	50,000.00	
	Cost/Server ($)	$2,000.00	
	Server Amortization (months)	36	(3 years)
	Size of Facility (Critical Load MW):	15,000,000.00	(15MW)
	Annual Cost of Money (%):	5%	
	Average Power Usage (%):	80%	(Average % of provisioned power used)
	Power Usage Effectiveness	1.7	
	Power and Cooling Infrastrucure (%)	82%	(% of infrastructure that is power & cooling)
	Network egress charges not included (workload dependent)		
Calculations			
	Infrastructure	$1,581,587	[=-PMT(CostOfMoney/12,FacilityAmortization,FacilityCost,0)]
	Servers	$2,997,090	[=-PMT(CostOfMoney/12, ServerAmortization, ServerCount*ServerCost, 0)]
	Power & Cooling Infrastructure	$1,296,902	[=InfrastructureMonthly*PowerAndCoolikngInfrastructurePercentage]
	Power	$1,042,440	[=MegaWattsCriticalLoad*AveragePowerUsage/1000*PUE*PowerCost*24*365/12
	Other Infrastructure	$284,686	[=InfrastructureMonthly-PowerAndCoolingInfrastructureMonthly]
	Full burdened Power	$2,339,342	[=PowerAndCoolingInfrastructureMonthly+PowerMonthly]
	Total:	$5,621,117	



Source: James Hamilton's Blog



- No need to built high-cost data center
- The proposed framework can be easily extended and tailored to the existing smart meter infrastructure.
- ~46 million smart meters in U.S. as of 2013.

Future Work

- Design a user-friendly graphical user interface (e.g., real-time alerting; interactive mapping tool) for residential owners, utilities, utility commission, and other entities
- Develop a massive array of mini PC-based distributed data storage and processing testbed for a large-scale system
- Design a user-friendly graphical user interface
- Customize the MapReduce algorithms to meet specific needs of smart grid applications
- Apply it to support data-intensive Smart Grid and other big data applications (e.g., intelligent transportation systems)
- Collaboration with industry partners
- ……

Acknowledgement

1. This work was partially supported by the National Science Foundation (NSF) under Award Number EEC-0812121.
2. This work was also supported by the New Faculty Start-Up Fund at University of Michigan-Dearborn.

N. Zhang, Y. Yan, S. Xu, and **W. Su***, “A Distributed Data Storage and Processing Framework for Next-Generation Residential Distribution Systems”, *Electric Power Systems Research*, vol.116, pp.174–181, November 2014. (**Link**)