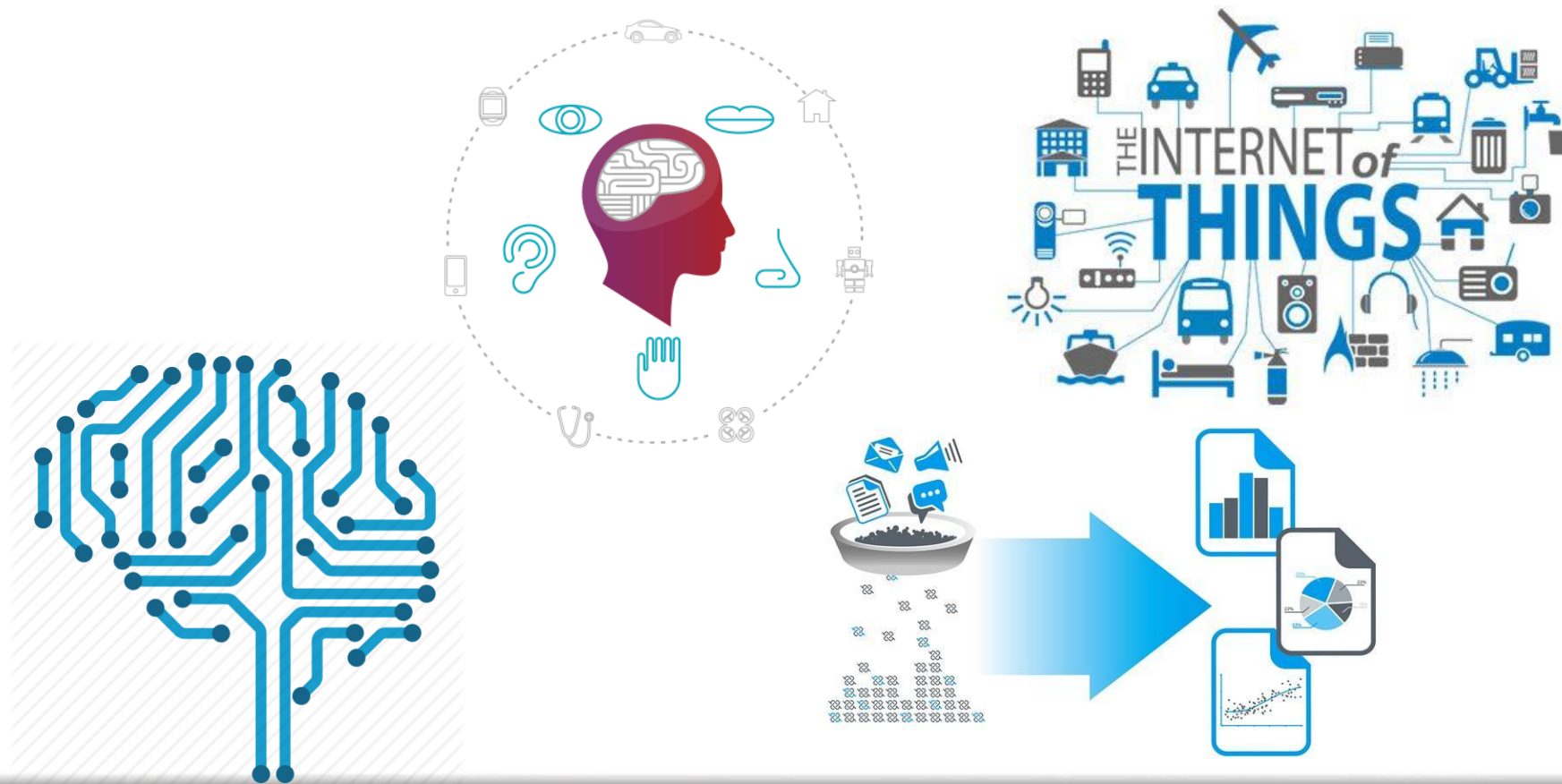


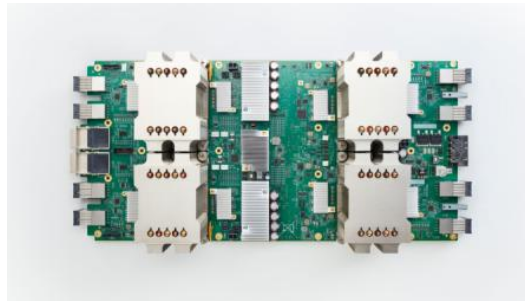
# Тенденции развития технологий Big Data и высокоскоростная сеть Ангара

А.С. Семенов, А.С. Фролов



Обучение

Применение



Google TPUv2



NVIDIA Volta



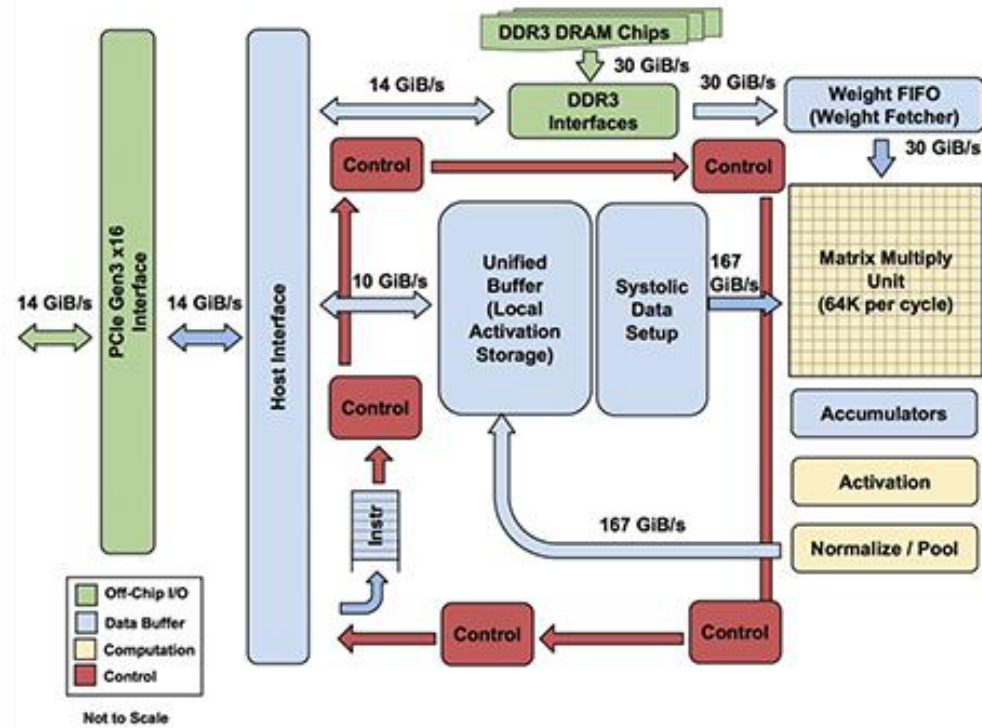
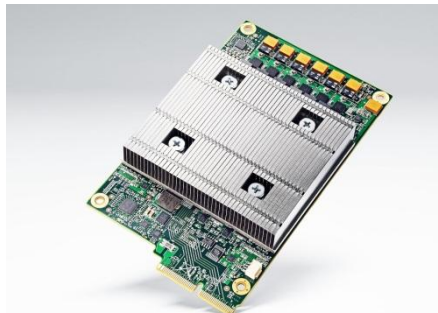
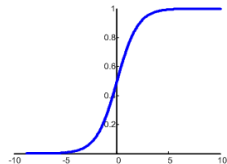
Google TPU

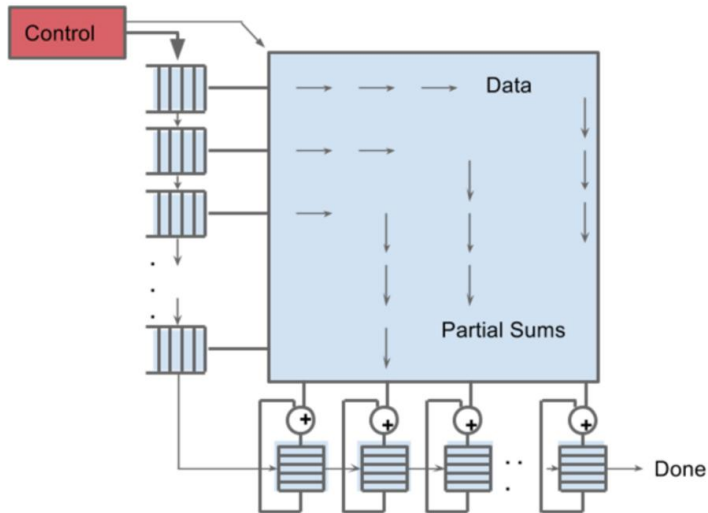


Функциональность vs Производительность

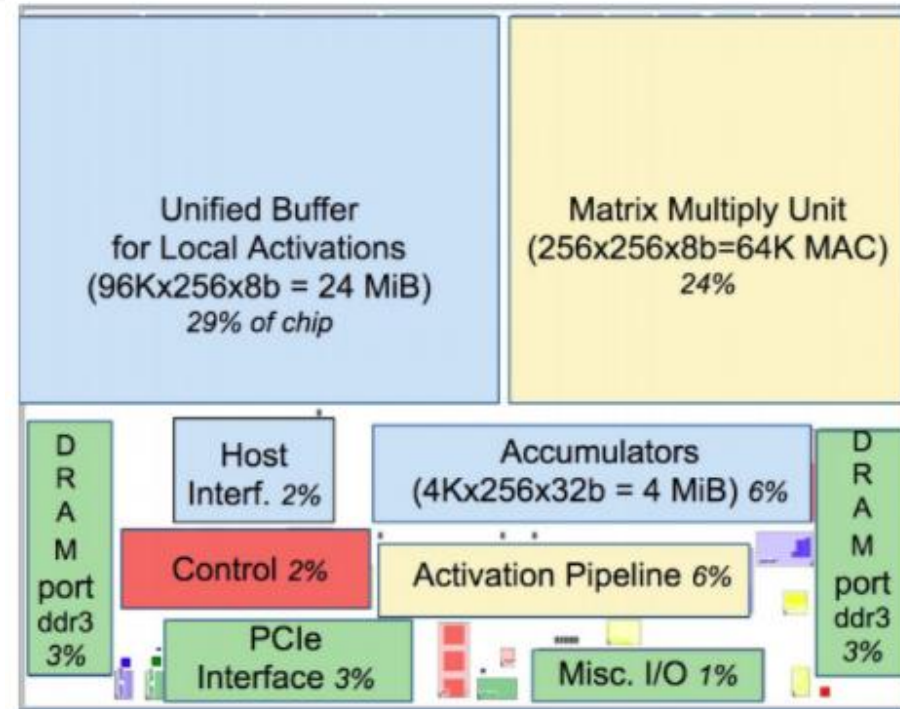
- Matrix Multiply Unit (MXU) – 64K 8 bits fma
- Activation Unit – функции активации
- ~ 12 инструкций CISC
  - Read\_Host\_Memory
  - Read\_Weights
  - MatrixMultiply/Convolve
  - Activate
  - Write\_Host\_Memory

$$f(x) = \frac{1}{1+e^{-x}}$$





- 28 nm, 700 MHz, 92 Tops, 40W
- **Производительность/W:**  
83X vs CPU, 29X vs GPU
- Unified Buffer – 24 MB
- 8 GB DDR3, 34 GB/s

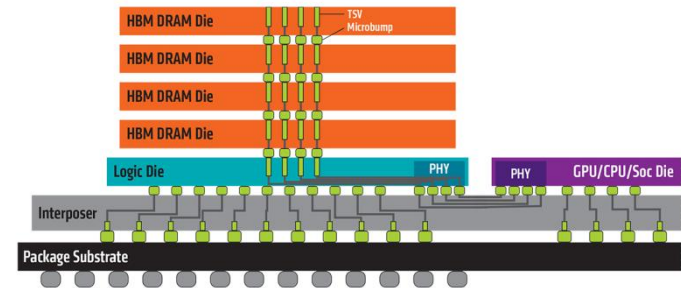


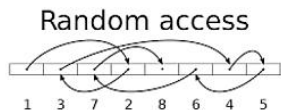
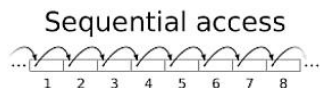


- TPU v1 – **bandwidth limited**

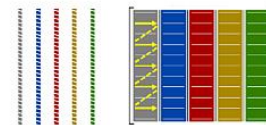
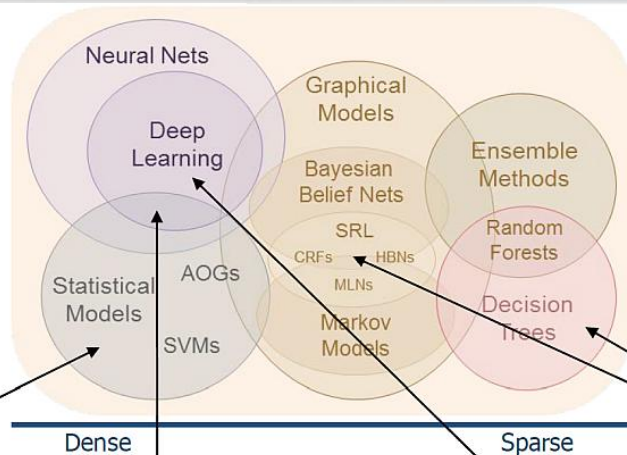
## TPU v2

- HBM 16 GB – 600 GB/s
- 32 bits floating point
- Обучение и применение
- 45 TFlops



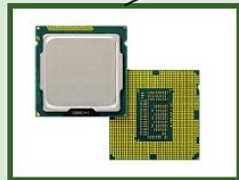


Sequential access is good for dense data  
Sparse data requires random access



Lower level primitives  
(5x5 Matrix)

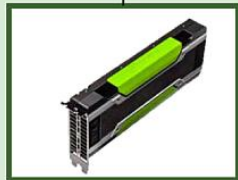
- 25 Scalar operations
- 5 Vector operations
- 1 Matrix operation



### Intel CPU

- Sequential processing
- Sequential memory access
- Slow (20GB/s) to memory
- Limited scalability (16GB/s)
- Optimized for Statistics

Source: Intel



### Nvidia GPU

- Parallel processing
- Sequential memory access
- Faster (288GB/s) to memory
- Limited scalability (20GB/s)
- Used for CNNs

Source: Nvidia



### Google TPU

- Parallel processing
- Sequential memory access
- Slow (20GB/s) to memory
- Limited scalability (16GB/s)
- Optimized for DNNs

Source: Google



### HIVE

- Parallel processing
- Parallel memory access
- Fastest (TB/s) to memory
- Higher scalability (TB/s)
- Optimized for Graphs



Graphcore®

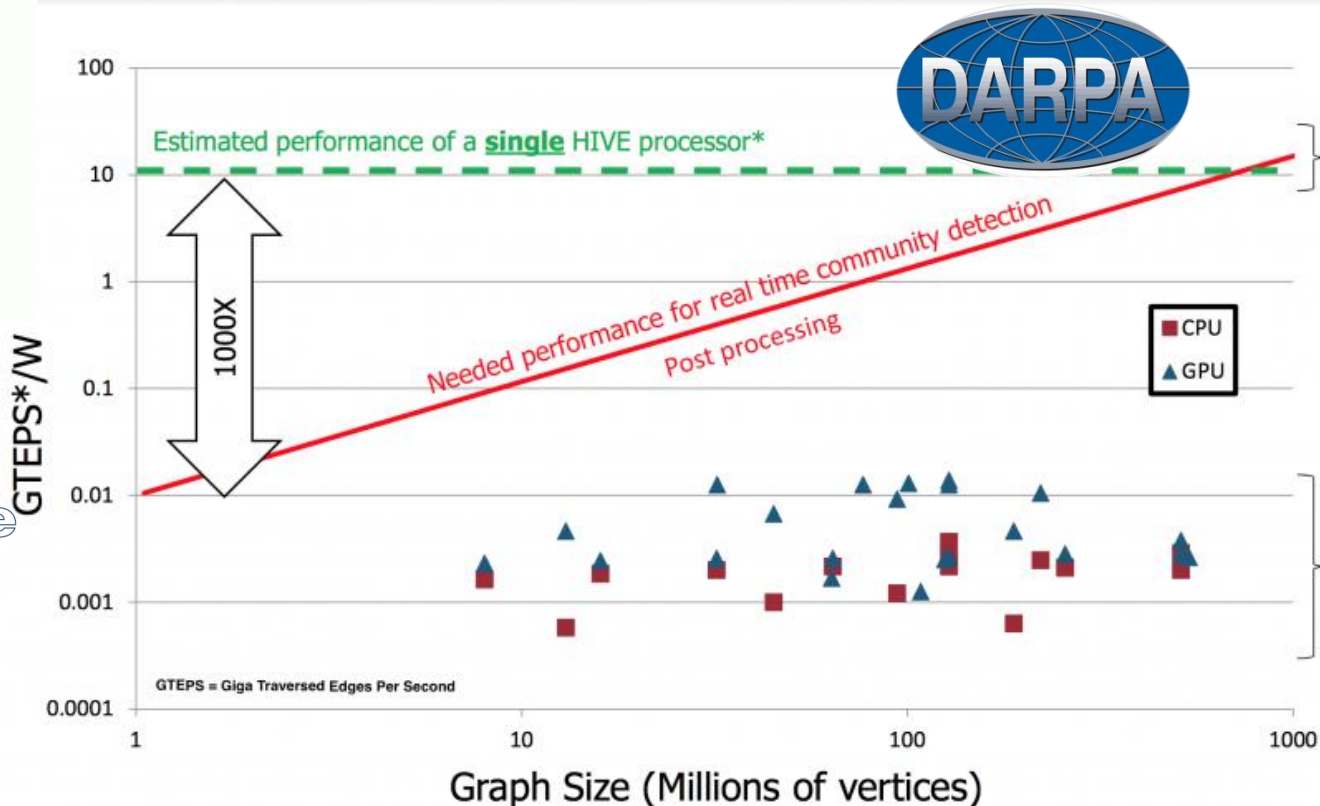
EmuTechnology



Pacific Northwest  
NATIONAL LABORATORY



**NORTHROP  
GRUMMAN**



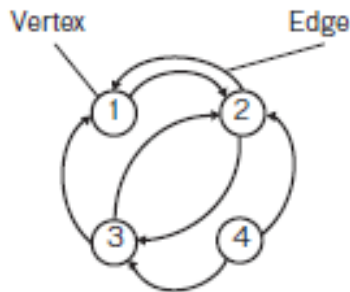
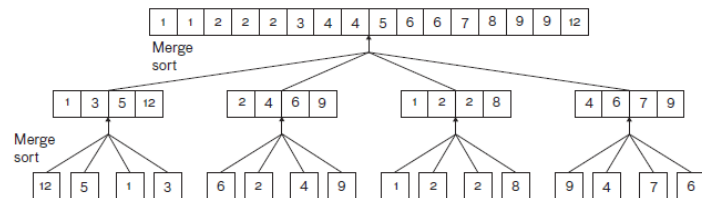


**GraphBLAS**

Операция	Описание
$A * x, A * B$	Обход графа (BFS), поиск кратчайших путей
$A + B, \dots$	Объединение, пересечение графов

**95% операций – сортировка**

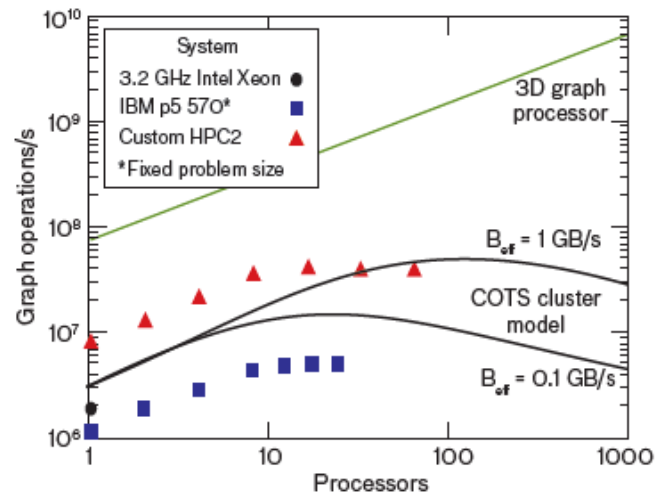
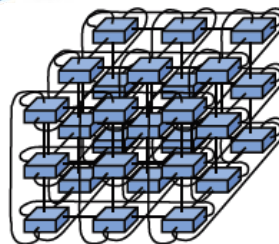
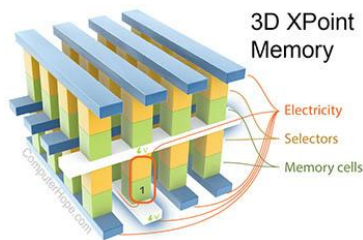
**k-way merge sort**



	12					
21		23				
	32		34			
42	43					

Coordinate format  
storage (row based)

Data value	12	21	23	32	34	42	43
Row index	1	2	2	3	3	4	4
Column index	2	1	3	2	4	2	3



# BIG DATA LANDSCAPE 2017

## INFRASTRUCTURE

**HADOOP ON-PREMISE**  
 Cloudera  
 Hortonworks  
 MAPR  
 Pivotal  
 IBM InfoSphere  
 bluedata  
 jethro

**HADOOP IN THE CLOUD**  
 Amazon  
 Microsoft Azure  
 Google Cloud Platform  
 IBM InfoSphere BigInsights  
 treasure data  
 aletriscale  
 CAZENA  
 CenturyLink

**STREAMING / IN-MEMORY**  
 Amazon Kinesis  
 databricks  
 confluent  
 stream  
 GridGain  
 METAMARKETS  
 DATATORRENT  
 dataArtisans  
 hazelcast  
 TERRACOTTA

**NOSQL DATABASES**  
 Google Cloud Platform  
 ORACLE  
 Amazon DynamoDB  
 Microsoft Azure  
 MarkLogic  
 mongoDB  
 DATASTR  
 REDIS  
 Couchbase  
 redislabs  
 InfluxData

**NEWSQL DATABASES**  
 SAP  
 Clustrix  
 Pivotal  
 nuodb  
 Cockroach Labs  
 memsql  
 splice  
 Voltron  
 citusdata  
 Trifolium  
 Seasdb  
 paradigm4

**GRAPH DBS**  
 neo4j  
 IBM ORACLE  
 N1 METEORA  
 Ochon  
 logunit  
 ASO  
 Informatica Objectivity

**MPP DBS**  
 VERTICA  
 N1 METEORA  
 Ochon  
 logunit  
 ASO  
 dremio

**CLOUD EDW**  
 Amazon  
 Google Cloud Platform  
 Microsoft Azure  
 Pivotal  
 snowflake  
 InfoWorks

**DATA TRANSFORMATION**  
 talend  
 pentaho  
 alteryx  
 TRIFACTA  
 tamr  
 Paxata  
 StreamSets  
 UNIFI

**DATA INTEGRATION**  
 informatica  
 snapLogic  
 MuleSoft  
 TEALUM  
 enigma  
 aloma  
 ZALONI  
 splint  
 import  
 Stitch

**DATA GOVERNANCE**  
 informatica  
 IBM skyhigh  
 colibra  
 Alation  
 Waterline

**MGMT / MONITORING**  
 Amazon CloudWatch  
 New Relic  
 APPDYNAMICS  
 Octrino  
 WAVEFRONT  
 unravel  
 splunk  
 Tronco  
 Numerity

**STORAGE**  
 Amazon S3  
 Google Cloud Platform  
 Microsoft Azure  
 ALLUXIO  
 nimblestorage  
 Qumulo  
 COMO  
 panteras

**CLUSTER SERVICES**  
 Amazon EMR  
 Kuber netx  
 mesos  
 docker  
 MESOSPHERE  
 CoreOS  
 prepodata

**APP DEV**  
 lightbend  
 rainforest  
 CRSK

**CROWDSOURCING**  
 amazon mechanical turk  
 upwork  
 WorkFusion  
 CrowdFlower

**HARDWARE**  
 Google TPU  
 ARM  
 MYTHIC  
 NVIDIA  
 Movidius  
 SCORTEX

## CROSS-INFRASTRUCTURE/ANALYTICS

amazon web services
 Google Cloud Platform
 Microsoft
 IBM
 SAP
 Hewlett Packard Enterprise
 sas
 JOJO data
 vmware
 TIBCO
 TERRADATA
 ORACLE
 NetApp

## ANALYTICS

**DATA ANALYST PLATFORMS**  
 Microsoft  
 pentaho  
 alteryx  
 guavus  
 AYASDI  
 WATTIVO  
 Datameer  
 Quid  
 ClearStory  
 OrigamiLogic  
 interana  
 Bottlenose  
 ARIMO  
 ENDOT  
 MODE

**DATA SCIENCE PLATFORMS**  
 IBM  
 KNIME  
 data iku  
 DOMINO  
 yhat  
 CONTINUUM ANALYTICS  
 Alpine  
 ALGORITHMIA  
 rapidminer  
 Anqoss

**BI PLATFORMS**  
 Microsoft  
 amazon  
 looker  
 Veeva Analytics  
 ARC4DATA  
 GoodData

**VISUALIZATION**  
 tableau  
 Google Cloud Platform  
 qlik  
 CELONIS  
 Purviscope  
 CHARTIO  
 plotly

**VERTICAL ANALYTICS**  
 PREDIX  
 CAPE  
 UPTAKE  
 TACHYUS  
 Aluminum  
 Autonoma

**STATISTICAL COMPUTING**  
 sas  
 SPSS  
 MATLAB

**DATA SERVICES**  
 Palantir  
 CONFERENCE OPERA  
 DATA SCIENCE kaggle  
 FUSICA  
 DataKind  
 FF

**MACHINE LEARNING**  
 Google Cloud Platform  
 H2O  
 DataRobot  
 context relevant  
 VISION ZEE  
 bonsai  
 databrain  
 Celonis

**HORIZONTAL AI**  
 IBM Watson  
 Cortana  
 Facet  
 Voyager  
 Affective  
 vtronicon  
 OSAI  
 CURIOUS AI  
 BLUE VISION

**SPEECH & NLP**  
 Google Cloud Platform  
 twilio  
 semantic machines  
 Vocalabs  
 ARRIA  
 Tolkit  
 snips  
 Soundbound Inc.

**SEARCH**  
 elastic  
 CRACKLE BRICKS  
 ThoughtSpot  
 U2 Lucidworks  
 swiftype  
 MAANA  
 alphasense  
 SearchNix  
 SINEOIA

**LOG ANALYTICS**  
 splunk  
 sumologic  
 loggly  
 librato  
 logzio

**SOCIAL ANALYTICS**  
 Hootsuite  
 sprinklr  
 NETBASE  
 DATA SIFT  
 synthesio  
 simple reach  
 bitly  
 predata

**WEB / MOBILE / COMMERCE ANALYTICS**  
 Google Analytics  
 mixpanel  
 sumali  
 retention  
 graniify  
 AMPLITUDE  
 Airtable  
 SIGOPT  
 custora

## OPEN SOURCE

**FRAMEWORK**  
 Hadoop  
 Amazon EMR  
 Hadoop  
 Flink  
 YARN  
 MEZOS  
 Spark  
 CDAP

**QUERY / DATA FLOW**  
 Spark  
 SQL  
 Presto  
 SLAMDATA  
 Google Cloud Dataflow  
 CouchDB  
 riak  
 HARISE  
 Spanner  
 SCOUTMULO

**DATA ACCESS**  
 nifi  
 mongoDB  
 cassandra  
 ZOOKEEPER  
 CouchDB  
 riak  
 HARISE  
 Spanner  
 SCOUTMULO

**COORDINATION**  
 talend  
 Apache Zookeeper  
 Apache Ambari

**STREAMING**  
 Spark  
 Flink  
 beam  
 kafka  
 druid  
 STORM

**STAT TOOLS**  
 python  
 ScalaLab  
 SciPy

**AI / MACHINE LEARNING / DEEP LEARNING**  
 TensorFlow  
 Caffe  
 CNTK  
 DM  
 Keras  
 VELES  
 DIMSUM  
 DS2NE  
 milib  
 DL4J  
 Aerosolve

**SEARCH**  
 elasticsearch  
 Solr

**LOG ANALYSIS**  
 kibana  
 logstash

**VISUALIZATION**  
 BEAKER  
 Rodeo

**COLLABORATION**  
 Jupyter  
 Zepplin  
 ANA CONDA

**SECURITY**  
 Apache Ranger  
 KNOX  
 Pentry

## DATA SOURCES & APIS

**HEALTH**  
 JAWBONE  
 VALIDIC  
 practicefusion  
 fitbit  
 GARMIN  
 Human API  
 kinsa

**IOT**  
 GE Digital  
 UPDATE  
 helium  
 samsara

**FINANCIAL & ECONOMIC DATA**  
 Bloomberg  
 THOMSON REUTERS  
 DOW JONES  
 S&P CAPITAL IQ  
 CB INSIGHTS  
 xignite  
 quandl  
 YODLEE  
 PREMISE  
 estimote  
 Eagle Alpha  
 StackTwits  
 PLAID  
 mastermark

**AIR / SPACE / SEA**  
 PLANET  
 Airware  
 spire  
 SKYWATCH  
 AEROBOTICS  
 INTELLIGENT  
 TELLUS LABS  
 WINOWARD  
 BroneDeploy  
 MasterCall

**PEOPLE / ENTITIES**  
 axlomo  
 Experian  
 EPILSON  
 Crimson Hexagon  
 BASIS  
 SAFEGRAPH

**LOCATION INTELLIGENCE**  
 FOURSQUARE  
 Sense  
 esri  
 CARTO  
 STREETLINE  
 MapAnything  
 PlaceIQ  
 factual  
 Mapillary

**OTHER**  
 qualtrics  
 DATA.GOV  
 data.world  
 panjiva  
 enigma

## APPLICATIONS - ENTERPRISE

**SALES**  
 Salesforce  
 CHIRUS  
 INSIDESALES.COM  
 conversica  
 clari  
 TACT  
 fuse/machines  
 TROOPS

**MARKETING - B2B**  
 RADIUS  
 EVERSTING  
 infer  
 sense  
 tubular  
 DataFox  
 JENAGIO

**MARKETING - B2C**  
 App Annie  
 Lattice  
 mintigo  
 Reflection  
 Zeta  
 bloomreach  
 PERSADO  
 ACTIONIQ  
 bluecore  
 SAITHRU  
 quantifind  
 Amplero

**CUSTOMER SERVICE**  
 MEDALLIA  
 zendesk  
 CLARIBRIDGE  
 CLICKFOX  
 DigitalGenius  
 AUTOMATIX  
 maga  
 INTERCOM  
 Gainsight  
 NGDATA  
 appturi  
 Frame.ai

**HUMAN CAPITAL**  
 hivIQ  
 entelo  
 hiQ  
 GIGSTER  
 Pacta  
 RESTART  
 WadeWendy  
 Quester  
 Stella  
 pymetrics

**LEGAL**  
 RAVEL  
 Seal  
 uora  
 tdemart  
 SAHAMA  
 TRADESHIP  
 ROSS  
 casetext

**FINANCE**  
 anaplan  
 Zuora  
 tdemart  
 SAHAMA  
 TRADESHIP  
 ROSS  
 casetext

**ENTERPRISE PRODUCTIVITY**  
 slack  
 Facebook  
 ORACLE  
 lumina  
 Clara Talia  
 butter  
 KASISTA

**BACK OFFICE AUTOMATION**  
 HyperScience  
 Coprlicity  
 AppZen

**SECURITY**  
 TANIAN  
 CYLANE  
 illumio  
 StackPath  
 DARKTRACE  
 ThreatMetrix  
 DataGravity  
 VETCTA  
 CyberCore  
 Guardian Analytics  
 ANOMALI  
 Shift science  
 SCIFYO  
 SentinelOne  
 Recorded Future  
 SecurityScorecard  
 feedzot  
 scout24  
 PORTSCALE  
 Logixia  
 CyberArk

## APPLICATIONS - INDUSTRY

**ADVERTISING**  
 AppNexus  
 criteo  
 XAD  
 Integral  
 OpenX  
 MOAT  
 theTradeDesk  
 drabridge  
 DataXu  
 DYNAMIC YIELD  
 Yieldmo

**EDUCATION**  
 KNEWTON  
 Clever  
 Declera  
 kidaptive  
 PANDORA  
 know2  
 gradience

**GOVERNMENT**  
 Socrata  
 OPENGOV  
 mark43  
 FNI FiscalNote  
 OpenDataSoft

**FINANCE - LENDING**  
 OnDeck  
 Kreditech  
 AVANT  
 finance  
 INSIKT  
 MoneyLion  
 TrueAccord  
 cignifi  
 Active AI

**FINANCE - INVESTING**  
 Dataminr  
 KENSIC  
 QUANTONIQ  
 NUMERA  
 ISENTUM  
 clearlymymoney  
 Algorix  
 ReverPack

**REAL ESTATE**  
 Opendoor  
 VTS  
 CREDIFI  
 reonomy  
 COMPISTAT

**INSURANCE**  
 Insuramile  
 Lemonade  
 CYENCE  
 Shift Technology  
 Tractable

**HEALTHCARE**  
 FLATIRON  
 GINGER  
 COTI  
 zebra  
 Acluro  
 imagine  
 entatic  
 Qventus  
 BAYLABS  
 HABEN  
 eviden

**LIFE SCIENCES**  
 color  
 oerz  
 zymogen  
 BenevolentAI  
 ZEPHYR  
 Clear Labs  
 Protonica  
 Citrine  
 twoAR  
 Aerovision  
 pop genomics

**TRANSPORTATION**  
 UBER  
 TESLA  
 CLEAR  
 drive  
 nato  
 pilipai  
 OTO  
 OPTIMUS  
 FLEXOR  
 comma  
 ntradyne  
 Civi Maps  
 NIC

**AGRICULTURE**  
 FARMERS  
 FarmersEdge  
 FarmLogs  
 BLUE DRIVER  
 maxiv  
 Terraviva  
 prospera

**COMMERCE**  
 Instacart  
 STITCH FIX  
 RetailNext  
 HowGood

**OTHER**  
 e-Harmony  
 stem  
 rethink robotics  
 Happi  
 BOX EVER  
 collect  
 duet  
 Unidrol  
 Second Spectrum

## DATA RESOURCES

**INCUBATORS & SCHOOLS**  
 PLURALSIGHT  
 DataCamp  
 DataElite  
 INSIGHT  
 The Data Incubator  
 galvanize  
 MEV

**RESEARCH**  
 facebook research  
 OpenAI  
 MIRI  
 Allen Institute of Artificial Intelligence  
 AI2

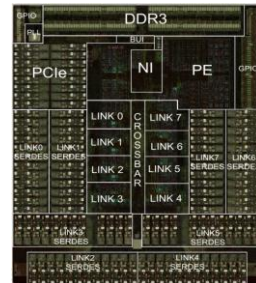
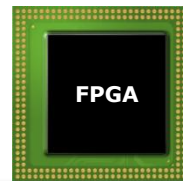
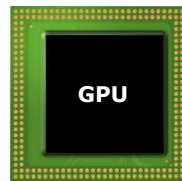
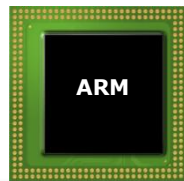
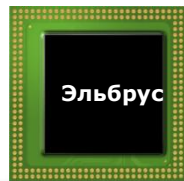
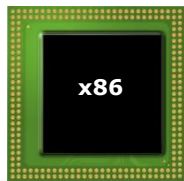
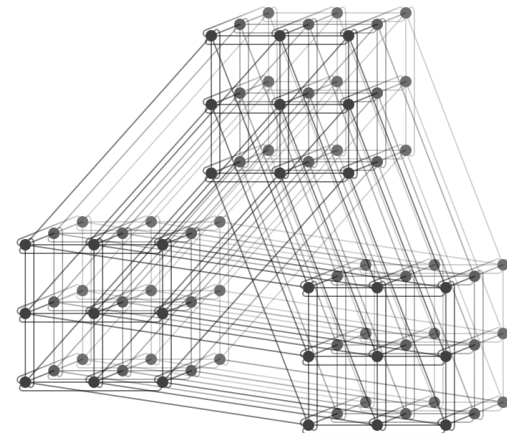
source: <http://www.christophersherwin.com/>

	HPC	BigData
Языки программирования	C/C++, перекомпиляция	JVM, Python, переносимость
Типы нагрузки	Изначально вычисления	Много и сложных данных
Масштабируемость	Изначально вертикальная	Горизонтальная
Управление выполнением	На каждом узле – MPI	Раздача заданий – Spark jobs
Файловые системы	Lustre, ...	HDFS, локальные данные
Аппаратура	High-end, RDMA	Commodity, Ethernet
Отказоустойчивость	Необходима обработка	Встроенная



### Ключевые особенности:

- Топология сети: 1D..4D-тор
- Адаптер на базе СБИС (65 нм, АО «НИЦЭВТ»)
- До 8 каналов связи с соседними узлами
- Прямой доступ в память удаленного узла (RDMA)
- Поддержка многоядерности
- Адаптивная передача пакетов
- Задержка на MPI ping-pong: 0,85/ 1,54 мкс (x86/Эльбрус-8С)
- Задержка на хоп: 130 нс
- Масштабирование: до 32К узлов
- Энергопотребление до 20 Вт
- Различные физические среды передачи данных





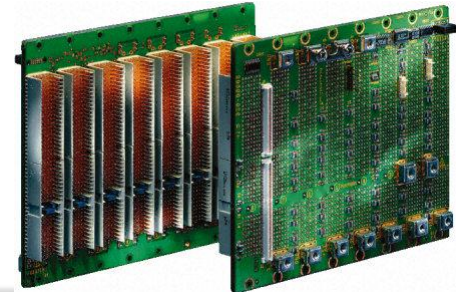
**1. Высокопроизводительное решение** на базе FHFL адаптера и Samtec кабеля



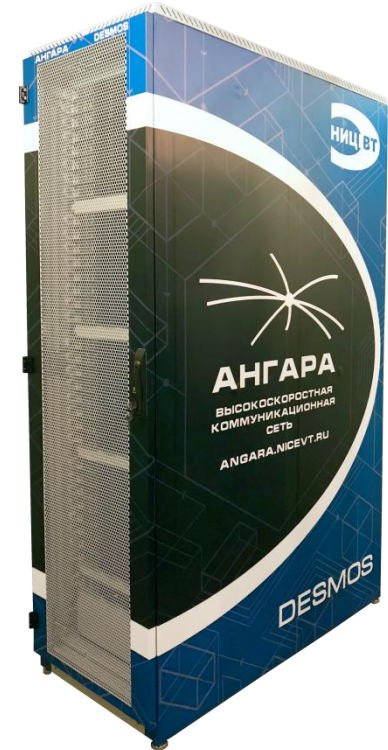
**2. Универсальное решение** на базе 24-портового коммутатора, low-profile адаптера и CXR кабеля



**3. Заказное решение** на базе объединительной платы и оптических кабелей



- **ОИВТ РАН: 32 вычислительных узла**
  - 1 процессор Intel Xeon E5-1650 v3 (6 ядер, 3.0 ГГц)
  - Nvidia GeForce GTX 1070
  - DDR4 16 ГБ
  - 4D-тор 4x2x2x2
- **Ангара-К1: 36 вычислительных узлов**
  - 12 узлов с 1 процессором Intel Xeon E5-2660 (8 ядер, 2.2 ГГц)
  - 24 узла с 2 процессорами Xeon E5-2630 (6 ядер, 2.3 ГГц)
  - 64 ГБ
  - 3D-тор 4x3x3





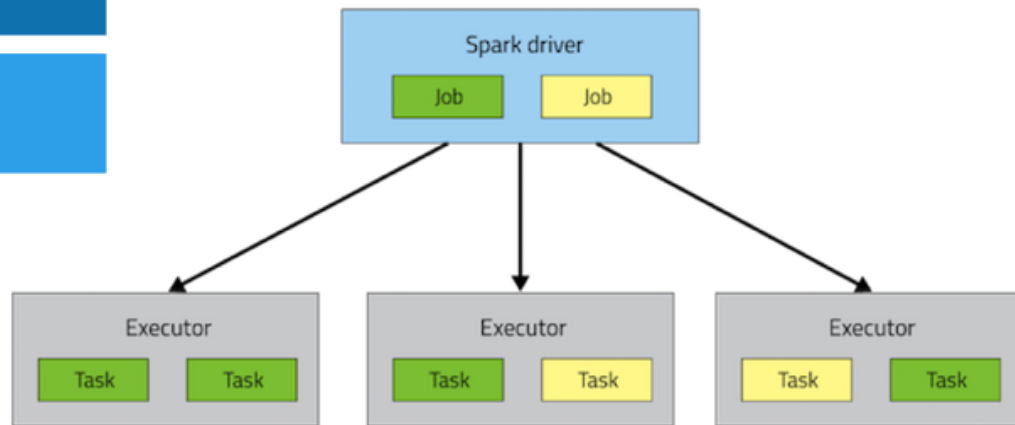
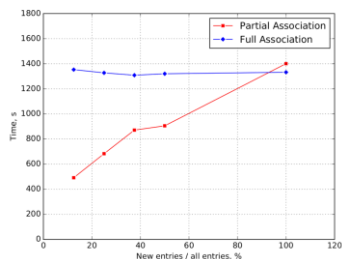
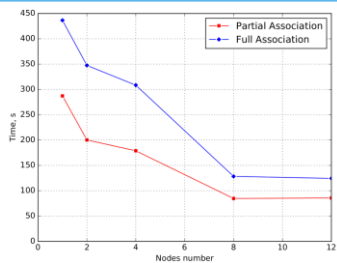
Kernel-space

User-space

- Поддержка ОС: Astra Linux SE 1.3-1.5, ОС «Эльбрус», OpenSUSE/SLES 11 SP3/4, CentOS 6.0-7.3, Версия ядра Linux от 2.6.21 до 3.16.0
- MPICH 3.0.4, 3.2, OpenMPI 1.10.2
- Поддержка компиляторов языков Fortran 77/90/95 (GNU, Intel), C/C++ (GNU, Intel)

Spark  
SQLSpark  
StreamingMLlib  
(machine  
learning)GraphX  
(graph)

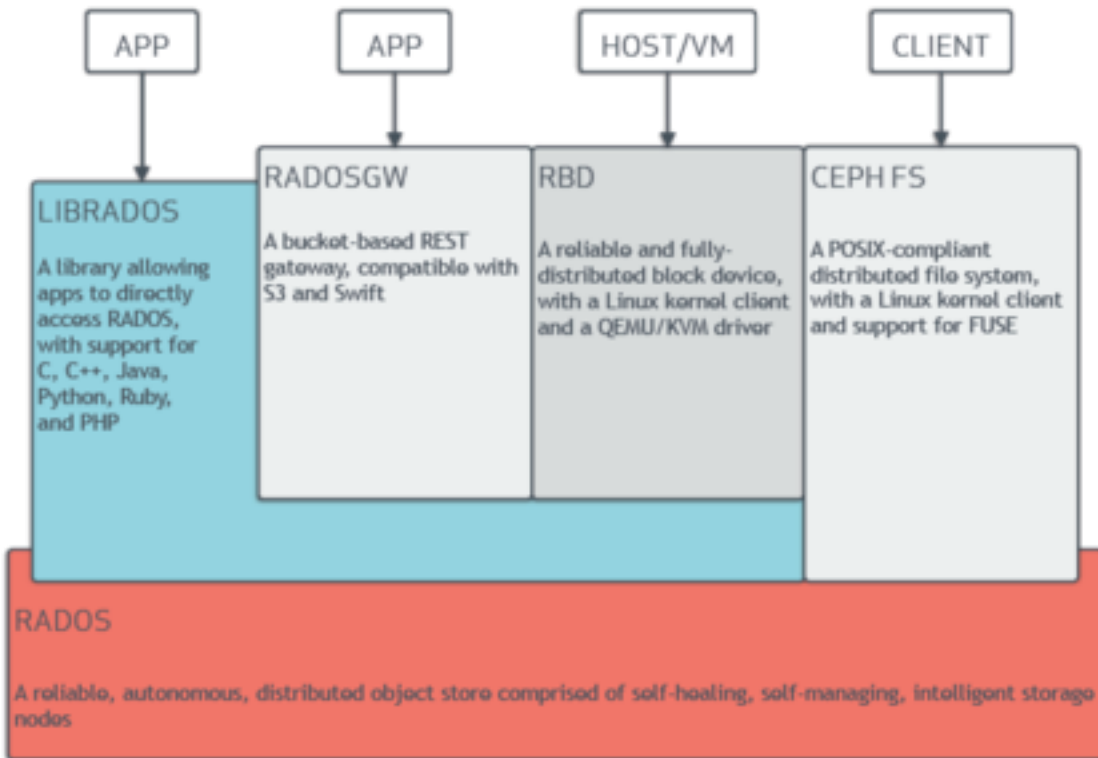
Apache Spark





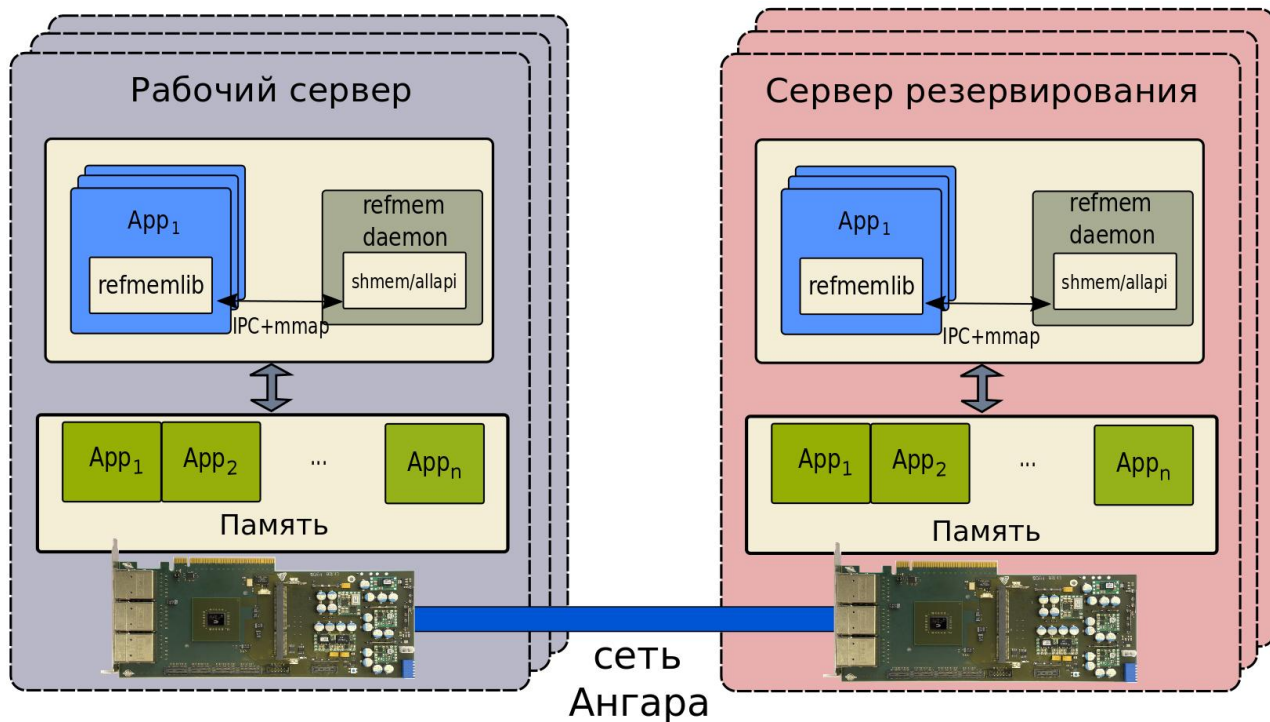


ceph



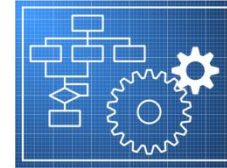
```
$# rados bench -p scbench 10 rand
sec Cur ops started finished avg MB/s cur MB/s last lat(s) avg lat(s)
0 0 0 0 0 0 0 - -
1 16 258 242 967.695 968 0.142073 0.0610817
2 16 487 471 941.789 916 0.0234243 0.0647762
3 15 739 724 965.153 1012 0.145909 0.0643161
4 15 1049 1034 1033.83 1240 0.0233676 0.0603486
5 16 1361 1345 1075.84 1244 0.0055336 0.0579456
6 16 1714 1698 1131.84 1412 0.0299221 0.0556169
7 16 2065 2049 1170.7 1404 0.012719 0.0536391
8 16 2419 2403 1201.34 1416 0.0165833 0.0523875
9 16 2754 2738 1216.73 1340 0.0138274 0.0517339
10 15 3103 3088 1235.04 1400 0.0764744 0.0510114

Total time run: 10.090779
Total reads made: 3104
Read size: 4194304
Object size: 4194304
Bandwidth (MB/sec): 1230.43
Average IOPS: 307
Stddev IOPS: 49
Max IOPS: 354
Min IOPS: 229
Average Latency(s): 0.0512704
Max latency(s): 0.22856
Min latency(s): 0.00462222
```



- Область памяти, которая автоматически реплицируется на другие узлы, на которых работает задача
- Возможен автоматический и программный режимы управления

- Настройка программного обеспечения на вычислительных системах, в том числе MPI
- Оперативная поддержка пользователей  
[angara.nicevt.ru](http://angara.nicevt.ru)  
[support@angara.nicevt.ru](mailto:support@angara.nicevt.ru)
- Профилирование и адаптация прикладного ПО



## Контакты:

117587, Москва, Варшавское ш, 125

[angara@nicevt.ru](mailto:angara@nicevt.ru)

