



Microsoft Big Data

Александр Гвоздев

Microsoft Россия



А что есть "Big Data"?

Massive collections of records – the
10's of PB

Для некоторых, "Big Data" это синоним NoSQL систем или параллельных реляционных DBMS типа



но дешевых узлов
обрабатывающих данные!!

Немного статистики Big Data



Sources:
"Big Data: The Next Frontier for Innovation, Competition and Productivity."
US Bureau of Labor Statistics | McKinsey Global Institute Analysis

= 1 billion terabytes
= 1 trillion gigabytes

Big data by industry



Financial Services

Risk and Monte Carlo;
Anti Money Laundering; Trend Analysis; Forecasting



Coms and Media

CDRs; Web Analytics; Social Media;
Advertising Analysis; Image analytics



Retail

Customer Analytics; Sales Analysis; Web Retail Analysis



Government and Academia

National Labs; Defence and Intelligence;
Research Data Mining



Online Gaming

Anti Fraud



Pharmaceutical

Genomic research; field trail analysis

Как работать с “Big Data”

«Старая
гвардия»

**Использовать параллельную
СУБД**

eBay – 10PB on 256 nodes

«младая
поросль»

Использовать NoSQL систему

NOSQL

Что значит...

~~NO to SQL~~

Это не значит что SQL больше не будет
Использоваться или что SQL умер...

NOT Only SQL

Это признание того факта, что
есть задачи для которых
иные решения возможно лучше!

Why?

NOSQL?

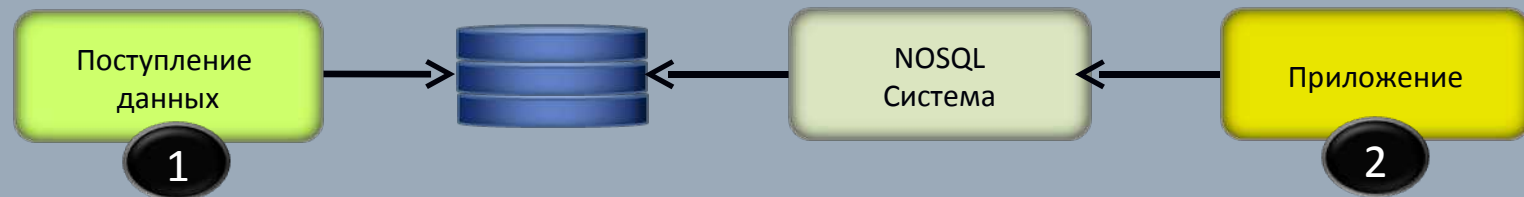
- More data model flexibility
 - JSON as a data model
 - No “schema first” requirement
- Relaxed consistency models such as *eventual consistency*
 - Willing to trade consistency for availability
- Low upfront software costs
- Never learned anything but C/Java in school
 - Hate declarative languages like SQL
- Faster time to insight from data acquisition

Быстрее получить <хоть какой-то> результат

SQL:



NoSQL:



Иногда называют "Schema Later"

- ❑ Нет очистки!
- ❑ Нет ETL!
- ❑ Нет загрузки!
- ❑ Анализировать данные там где они живут!

Two Major Types of NOSQL Systems

- **Key/Value Stores**

- Examples: MongoDB, CouchBase, Cassandra, Windows Azure, ...
- Flexible data model such as JSON
- Records “sharded” across servers
- Single record retrieval

- **Hadoop**

- Scalable fault tolerant storage for MASSIVE data sets
- Typically no data model
- Records stored in distributed file system

Think NOSQL
OLTP

Think NOSQL
data warehousing

2 вселенных стали новой реальностью



Структурированные

&



Неструктурированные

Реляционные СУБД

NoSQL Системы

Structured data w/ known schema

(Un)(Semi)structured data w/o schema

ACID

No ACID (but BASE*)

Transactions

No transactions

SQL

No SQL

Rigid Consistency Model

Eventual consistency

ETL

No ETL

Longer time to insight

Faster time to insight

Maturity, stability, efficiency

Flexibility

*BASE -basically available, soft state, eventually consistent

И ЧТО...?

- I believe the world has truly changed
- Relational DB systems no longer the only game in town
- As SQL “guys” we must accept this new reality and understand how best to deploy technologies like Hadoop

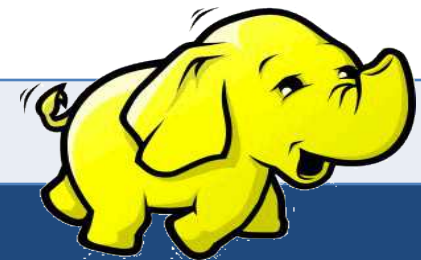


David J. DeWitt
Microsoft Jim Gray Systems Lab

This is NOT a paradigm shift
RDBMS will continue to dominate
transaction processing and ALL small to
medium sized data warehouses
But many businesses will end up with
data in both universes

Сравнение RDBMS и MapReduce (взгляд с другой стороны)

	Traditional RDBMS	MapReduce
Data Size	Terabytes	Petabytes
Access	Interactive and Batch	Batch
Updates	Read / Write many times	Write once, Read many times
Structure	Static Schema	Dynamic Schema
Integrity	High (ACID)	Low
Scaling	Nonlinear	Linear
DBA Ratio	1:40	1:3000



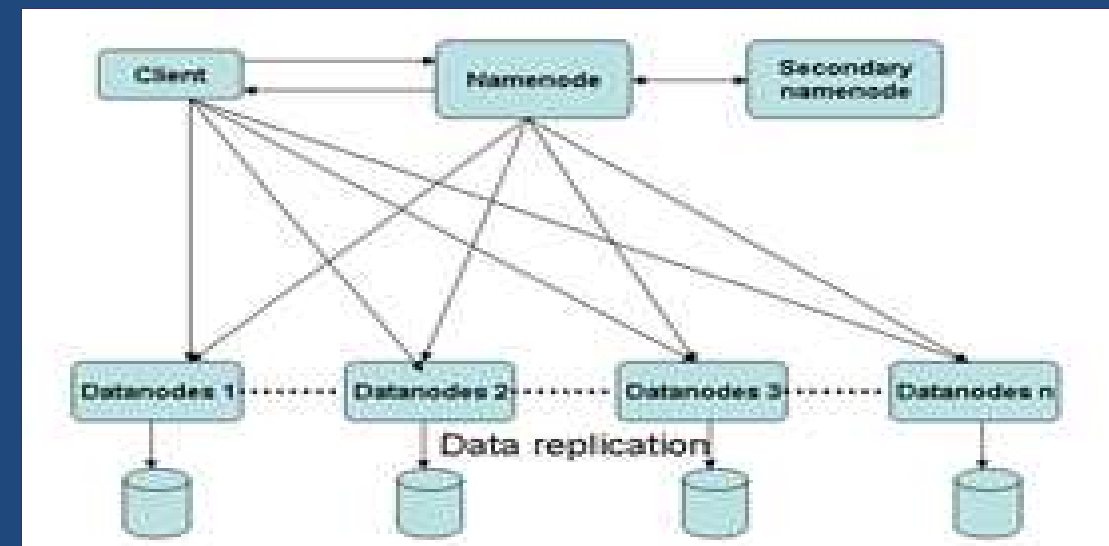
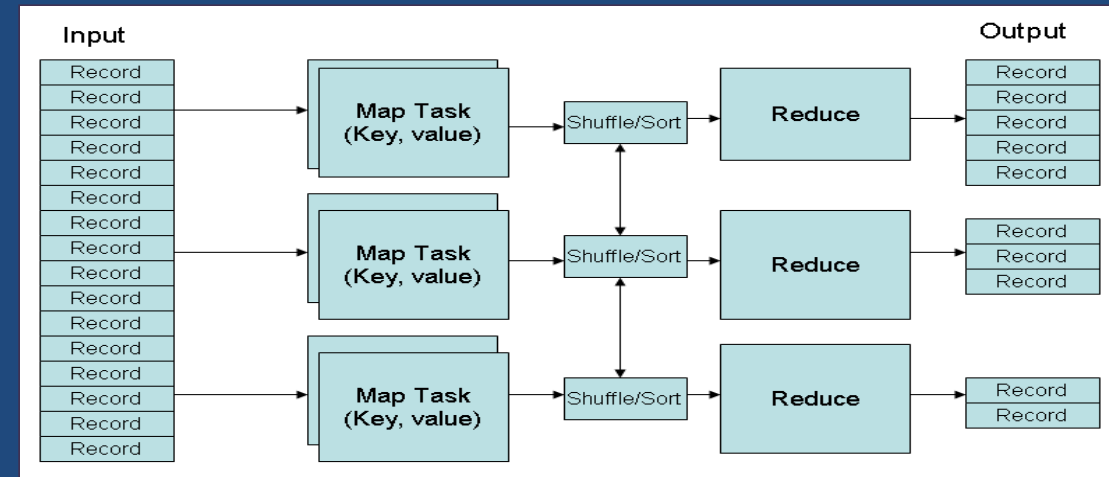
Reference: Tom White's Hadoop: The Definitive Guide

Hadoop: The most visible face of Big Data

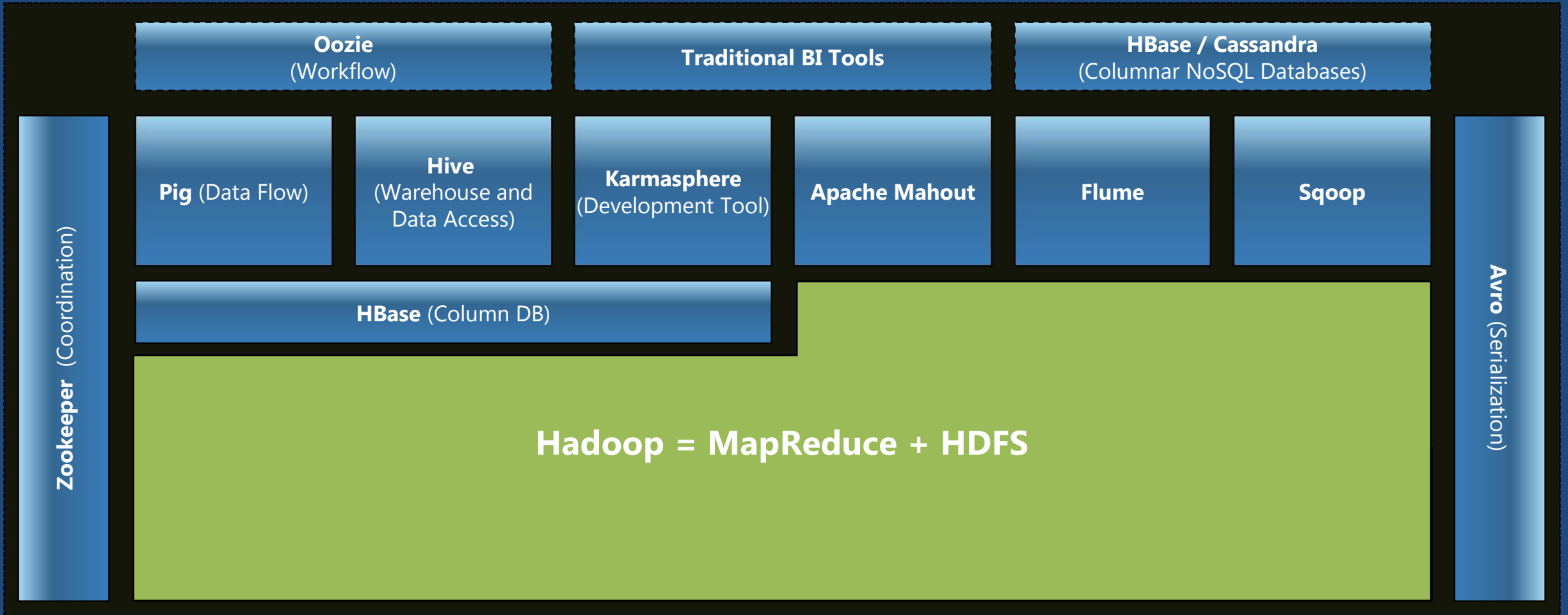
Enables analysis of semi and unstructured data distributed across a commodity cluster

Infrastructure to **automatically distribute and replicate data across multiple nodes** and **execute and track map reduce jobs** across all of those nodes

- Based on Google's MapReduce paper and Google File system (GFS)



Hadoop Ecosystem





Cassandra
Hive
Scribe
Hadoop



Hadoop
Oozie
Pig (-latin)



BackType
Hadoop
Pig / Hbase
Cassandra



MR/GFS
Bigtable
Dremel
...



SimpleDB
Dynamo
EC2 / S3
...

Microsoft | Cosmos | Daytona | LinqToHPC | Viola-James | F# | Azure

NoSQL ecosystem | open source, commodity

Mahout	Scalable machine learning and data mining
MongoDB	Document-oriented database (C++)
Couchbase	CouchDB (doc dB) + Membase (memcache protocol)
Hbase	Hadoop column-store database
R	Statistical computing and graphics
Pegasus	Peta-scale graph mining system
Lucene	full-featured text search engine library

Microsoft и Big Data

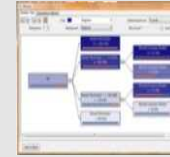
ЗНАКОМЫЕ
ИНСТРУМЕНТЫ



Power View



Excel with PowerPivot



Predictive Analytics



Embedded BI

BI ПЛАТФОРМА



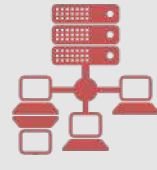
SSAS



SSRS



Hadoop On
Windows Azure



Hadoop On Windows
Server

Connectors



Microsoft EDW

ДАННЫЕ



Sensors



Devices



Bots



Crawlers



ERP



CRM



LOB



APPs

MARKETPLACE

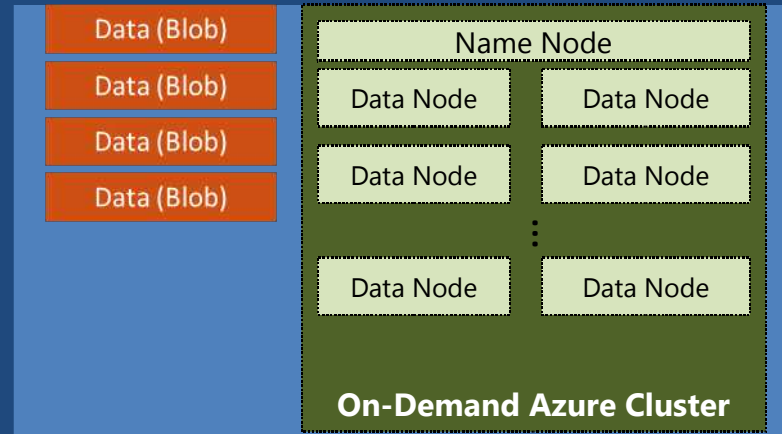
External Data and Services

Microsoft Hadoop Strategy

Make Hadoop Enterprise Ready

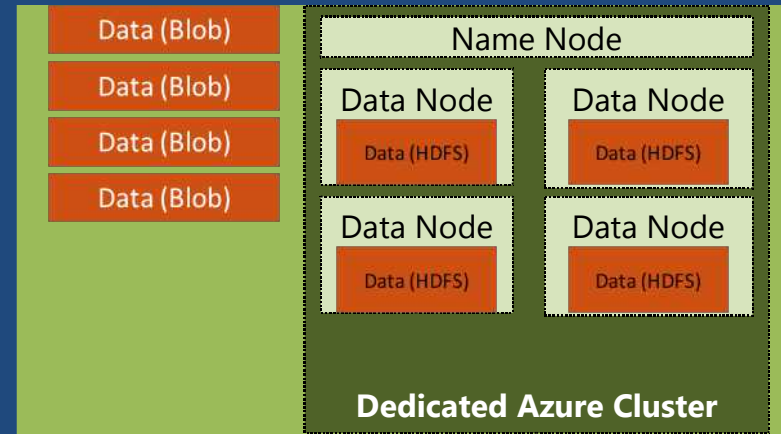
-
- Our own distribution of Hadoop
 - Submit changes back to Apache Foundation
 - Download for free
-
- Optimized for Windows & Azure
 - AD & Systems Center integration
 - Hadoop-as-a-service-on-Azure
-
- Focus on .NET Developers
 - Integration with Visual Studio
 - Support for C#
-
- Differentiation through Enterprise Readiness
 - Performance and Scale
 - High Availability
 - Ease of use
-

Microsoft Hadoop Deployment Models



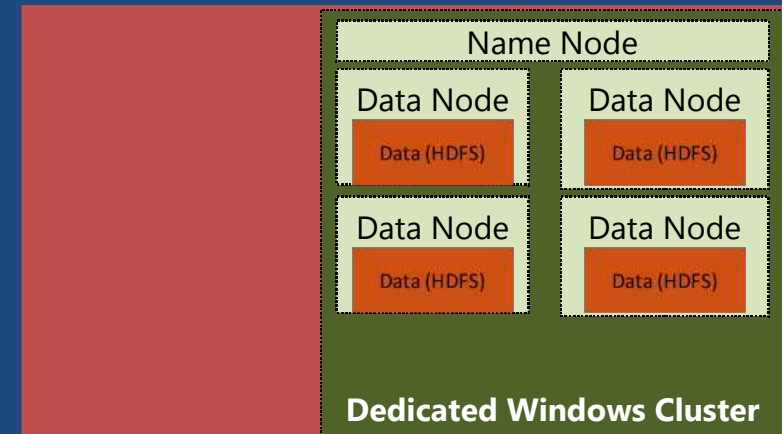
- On-demand elastic Hadoop cluster
- Bursty / batchy workloads
- Simplified management and deployment
- Complete elasticity

Elastic MapReduce



- Dedicated Hadoop cluster in the cloud
- Dedicated cloud workloads with more stringent storage / latency requirements
- Some elasticity (depending on HDFS storage and partitioning)

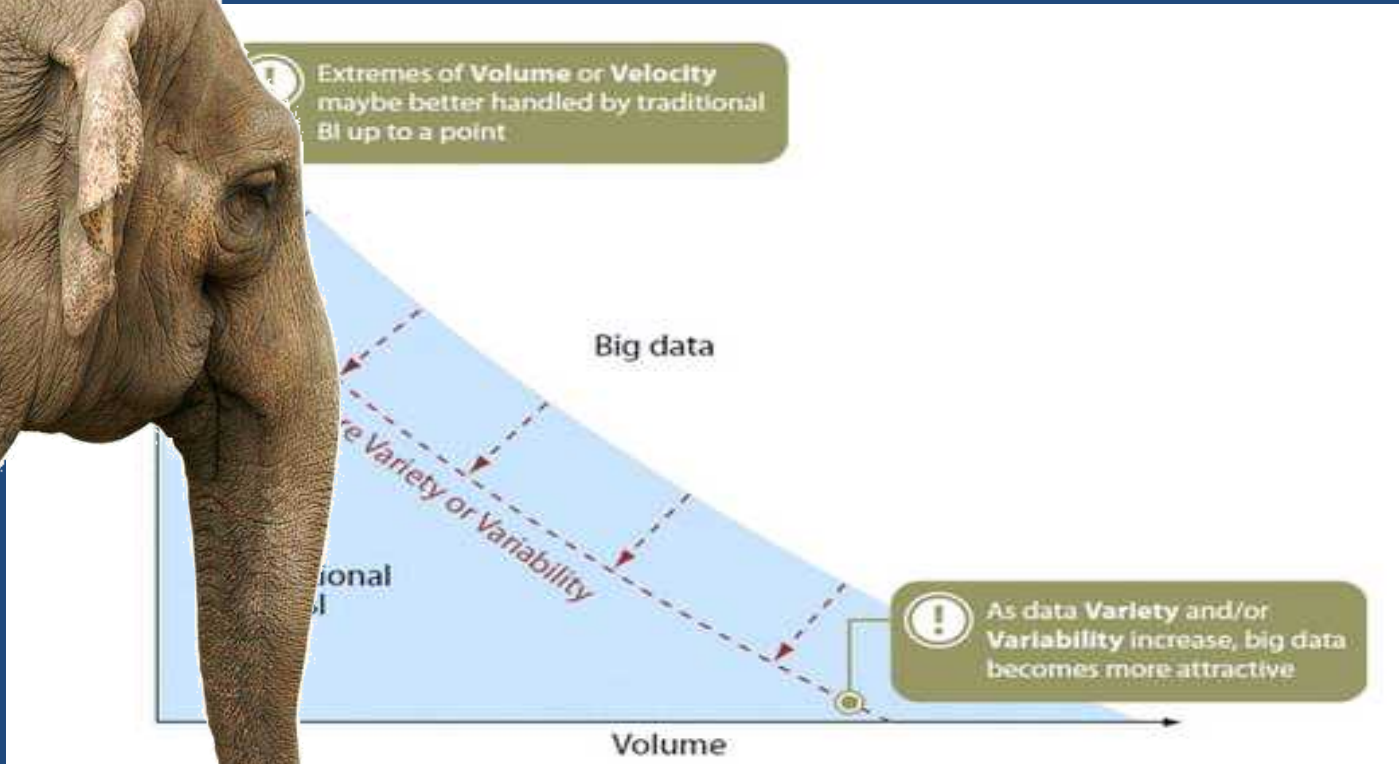
Dedicated Cloud Cluster



- Dedicated Hadoop cluster on Windows Server
- On premise workloads
- Management integration

Dedicated On Premise Cluster

The 4Vs of Big Data



Volume

Exceeds physical limits of vertical scalability.

Velocity

Decision window small compared to data change rate.

Variety

Many different formats make integration expensive

Variability

Many options or variable interpretations confound analysis

<http://nosql.rospescu.com/post/9621746531/a-definition-of-big-data>

Common Big Data Patterns



all potentially
ble ambient
ze later



Information Production

Mine the digital shoebox for insight
Feed downstream systems



Information Factory

Produce consumable feeds
Discover, enrich and publish

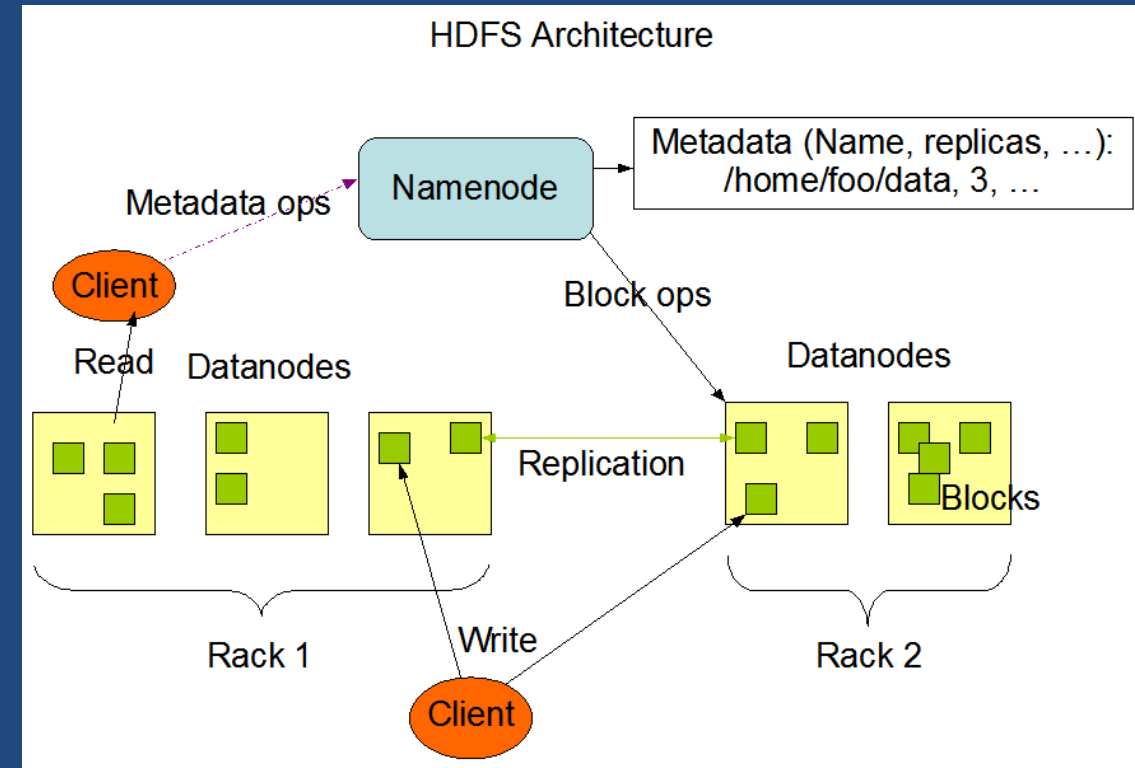


Optimization Loop

Analyze ambient data
Monitor system
Optimize behavior

HDFS Architecture

- Hadoop Distributed File System (HDFS) is a **scalable, distributed file** system
 - Single **namenode** manages **metadata** to enable **locality**
 - Many (1000's) **datanodes** store **data**
- Data is:
 - **Write once, read many**
 - **Replicated** to multiple data nodes



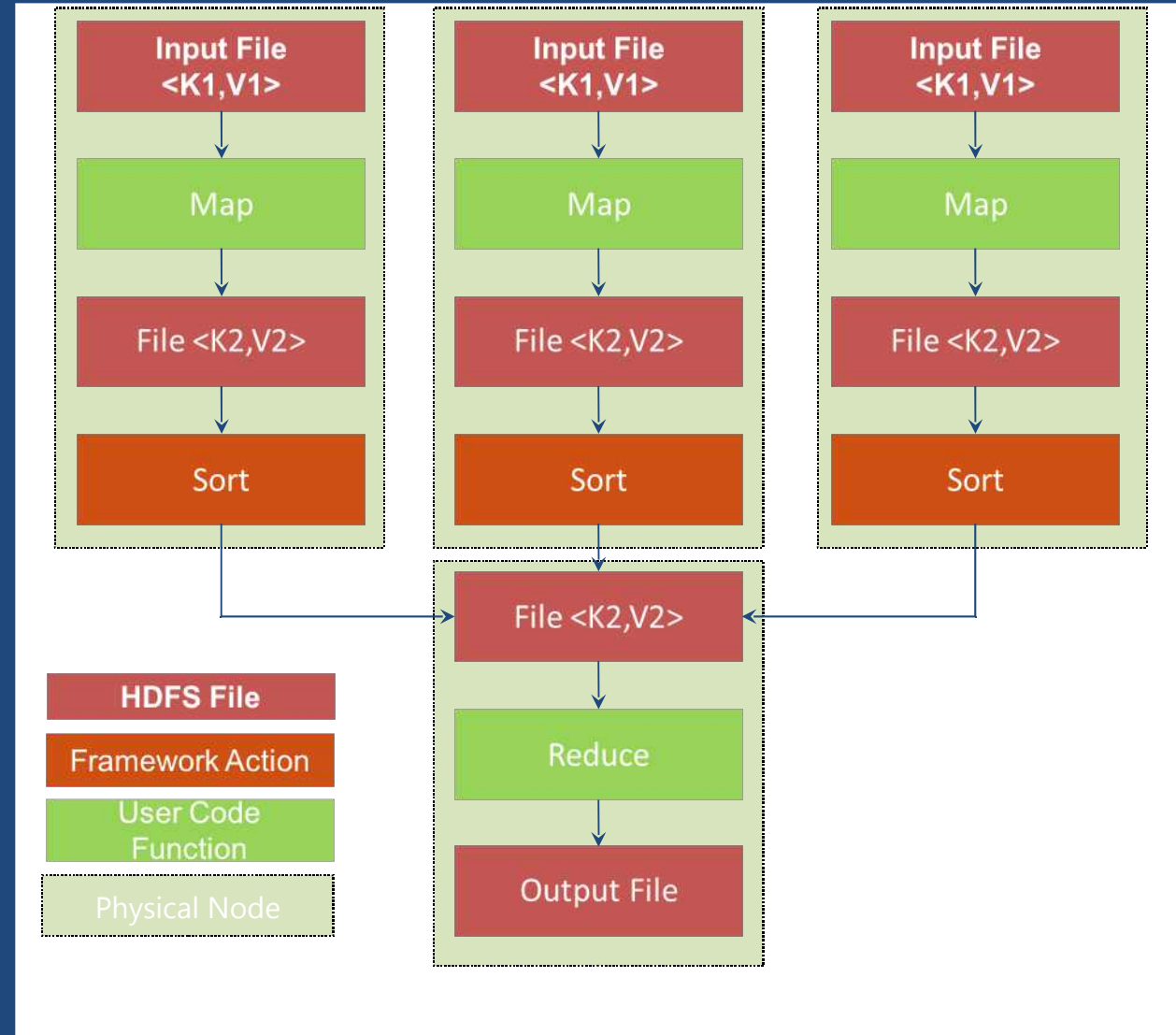
Courtesy of
http://hadoop.apache.org/common/docs/current/hdfs_design.html

HDFS, the Comic
Maneesh Varshney



MapReduce Workflow

- Three phases:
 - **Map**. Split the data set into fully independent data chunks.
 - **Shuffle**. Sort the output of the maps into coherent data sets [framework]
 - **Reduce**. Reduce or aggregate each shuffled data set.
- The framework handles:
 - **Scheduling** and **allocating** tasks (map execution locality)
 - **Monitoring** and **restarting** failed jobs (reallocating to other nodes)
- MapReduce jobs can be chained together to deliver more sophisticated analytics



Browser window showing the URL <https://www.hadooponazure.com/Account>. The address bar contains search, lock, and refresh icons. The page title is "Apache™ Hadoop™-based Services for Windows Azure". The browser's taskbar shows several open tabs: "Apach...", "Which on...", "Gender In...", "Interactiv...", "Manage...", and "Microsoft...". The Windows taskbar includes icons for Bing, Outlook, and various applications. The search bar contains the text "read".

Use Hadoop 1.0 – there will be NoSQL!

Request a new cluster

DNS name

DNS name
 Available
<http://sqlgal1.cloudapp.net>

Cluster size

- | | | | |
|---|---|---|--|
| <input checked="" type="radio"/> Small
4 nodes
2 TB disk space
Available | <input type="radio"/> Medium
8 nodes
4 TB disk space
Available | <input type="radio"/> Large
16 nodes
8 TB disk space
Available | <input type="radio"/> Extra large
32 nodes
16 TB disk space
Available |
|---|---|---|--|

Cluster login

Username

Password

Confirm password



Browser window showing the URL <https://www.hadooponazure.com/Account>. The address bar contains the text "read". The browser interface includes navigation buttons, a search bar, and a taskbar with various application icons.



Use Hadoop 1.0 – there will be NoSQL!

Allocation in progress

Your cluster **sqlgal.cloudapp.net** is being allocated. This will take a few minutes.

You may close the browser and return to this page to check the status of your cluster.

Role Instance	Status
IsotopeWorkerNode_IN_0	Allocating node...
IsotopeWorkerNode_IN_1	Allocating node...
IsotopeWorkerNode_IN_2	Allocating node...
IsotopeWorkerNode_IN_3	Allocating node...
IsotopeHeadNode_IN_0	Allocating node...

Manage your account

Summary of account management metrics displayed in orange boxes: 57 and 9.

Your Tasks

+ New

Create Job

Your Cluster

JavaScript: Idle
Hive: Idle

Interactive Console

Status: OK

Remote Desktop

FTP
ODBC Server

Open Ports



0.00 TB used

Manage Data

Manage your account

Billing History

4

Job History

Samples

Downloads

Browser window showing the URL <https://www.hadooponazure.com/Cluster/ManageData>. The browser includes a search bar with the text "read", navigation buttons (Previous, Next), and a dropdown menu for "Options". The browser's taskbar shows several open tabs and various application icons.

Manage Cluster

0.00 TB out of 1.89 TB



Disaster Recovery

Disaster recovery is configured

Set up ASV

Use your Windows Azure Blob Storage account.

DataMarket

Import data from DataMarket.

Set up S3

Use your Amazon S3 account.

Browser window showing the URL <https://www.hadooponazure.com/Cluster/ConfigurePorts>. The browser includes a search bar with the text "read", navigation buttons (Previous, Next), and a dropdown menu (Options). The taskbar shows various open applications and system icons.



← Configure Ports

By default all of the ports on your cluster are closed. For more complex configuration, use the remote desktop connection.

Remember that opening ports can present a security risk.

Name	Port #	Status	Toggle
FTPS	2226	Open	<input type="checkbox"/>
ODBC Server	10000	Open	<input type="checkbox"/>

Interactive JavaScript

JavaScript

Hive

```

js> #ls
Found 2 items
drwxr-xr-x - larar supergroup          0 2012-02-23 02:43 /user/larar/.oink
drwxr-xr-x - larar supergroup          0 2012-02-23 02:41 /user/larar/examples

js> fs.put()
File uploaded.

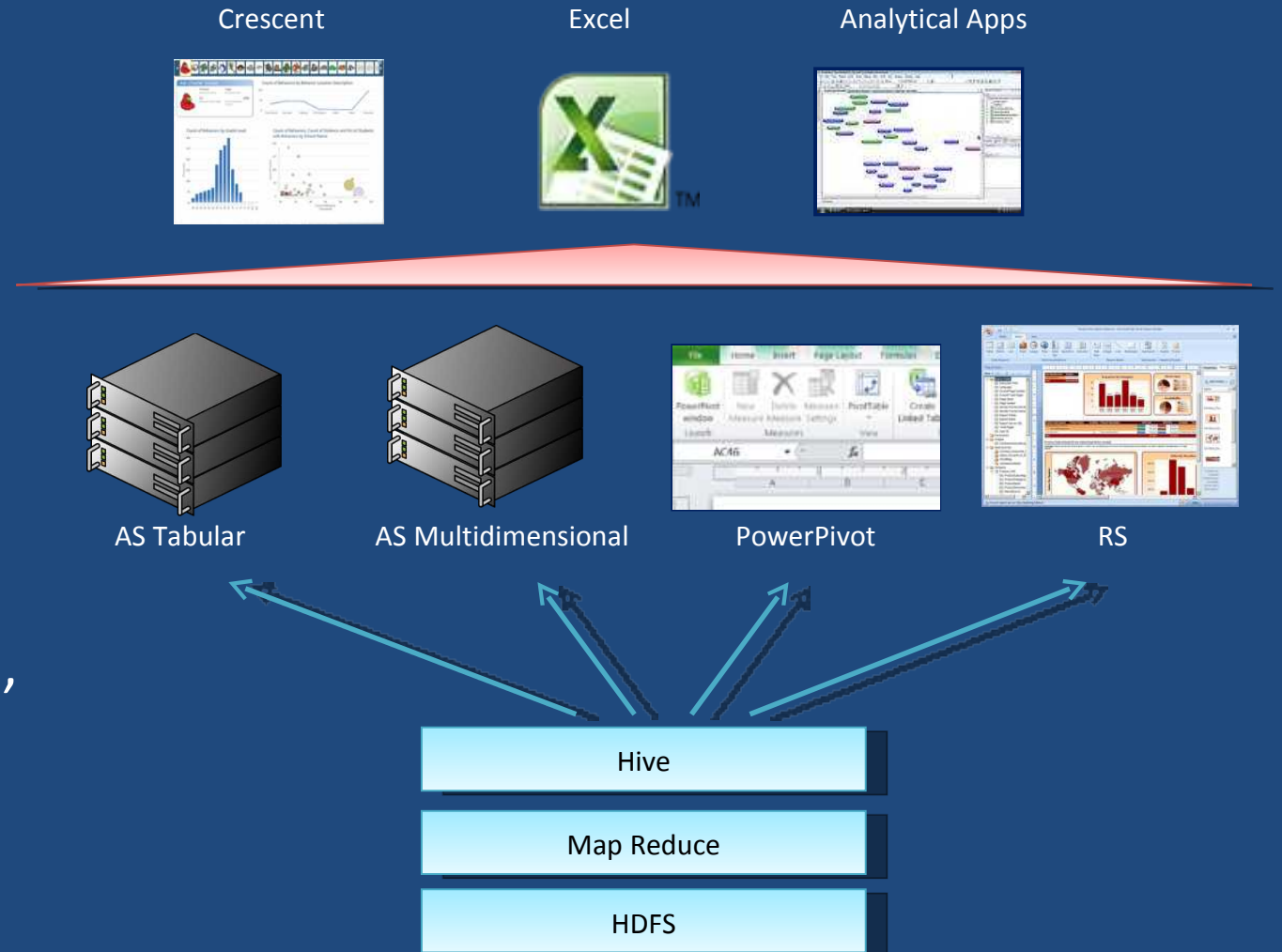
js> #ls
Found 3 items
drwxr-xr-x - larar supergroup          0 2012-02-23 02:43 /user/larar/.oink
-rw-r--r-- 3 larar supergroup        405 2012-02-23 03:49 /user/larar/WordCount.js
drwxr-xr-x - larar supergroup          0 2012-02-23 02:41 /user/larar/examples

js> #cat WordCount.js
var map = function (key, value, context) {
  var words = value.split(/[^\a-zA-Z]/);
  for (var i = 0; i < words.length; i++) {
    if (words[i] !== "") {
      context.write(words[i].toLowerCase(), 1);
    }
  }
};var reduce = function (key, values, context) {
  var sum = 0;
  while (values.hasNext()) {
    sum += parseInt(values.next());
  }
  context.write(key, sum);
};

js> pig.from("/example/data/davinci.txt").mapReduce("WordCount.js", "word, count:long").orderBy("count DESC")
2012-02-23 03:53:04,030 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceI

js>
    
```

Hive Connector: первые шаги по интеграции MS BI Platform



Hive ODBC driver

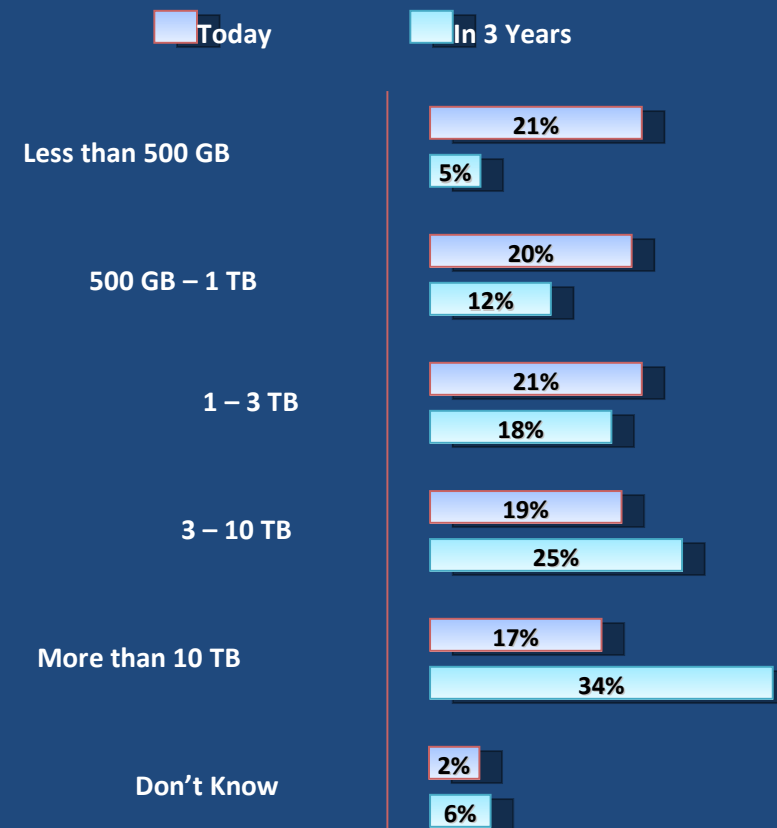
Использование Hadoop для Map Reduce, text mining, statistical analysis и т.д.

Получение Hadoop данных в AS, RS, PowerPivot через HiveQL

Ожидания клиентов

- Увеличение объемов данных

Approximate data volume managed by data warehouse

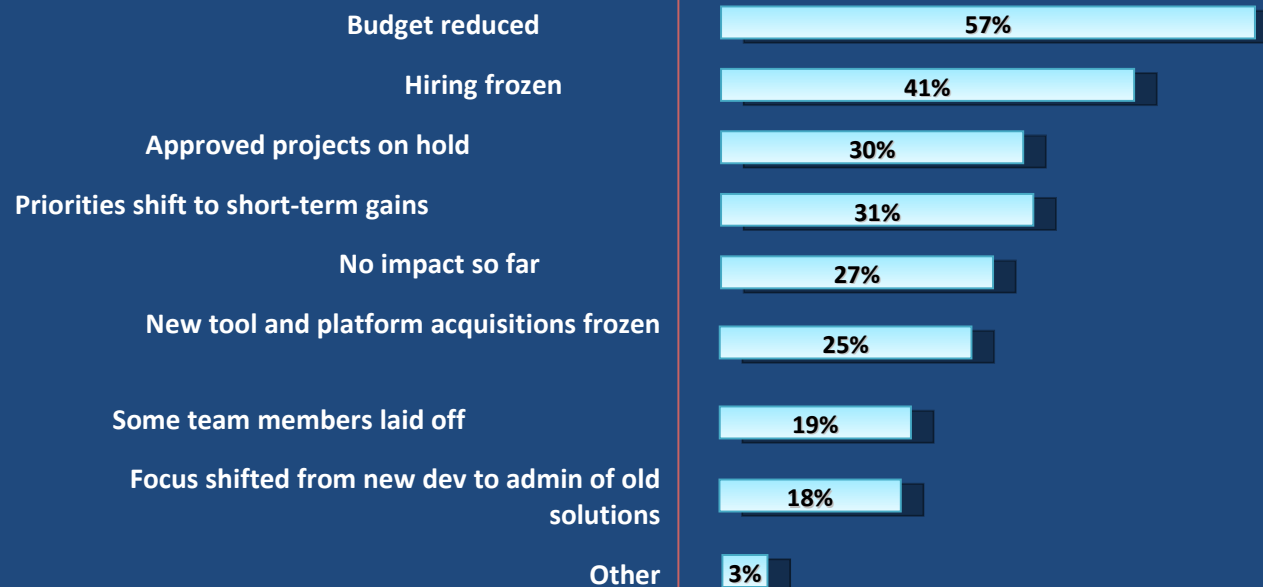


Source: TDWI Report – Next Generation DW

Ожидания клиентов

- Увеличение объемов данных
- Сокращения бюджетов

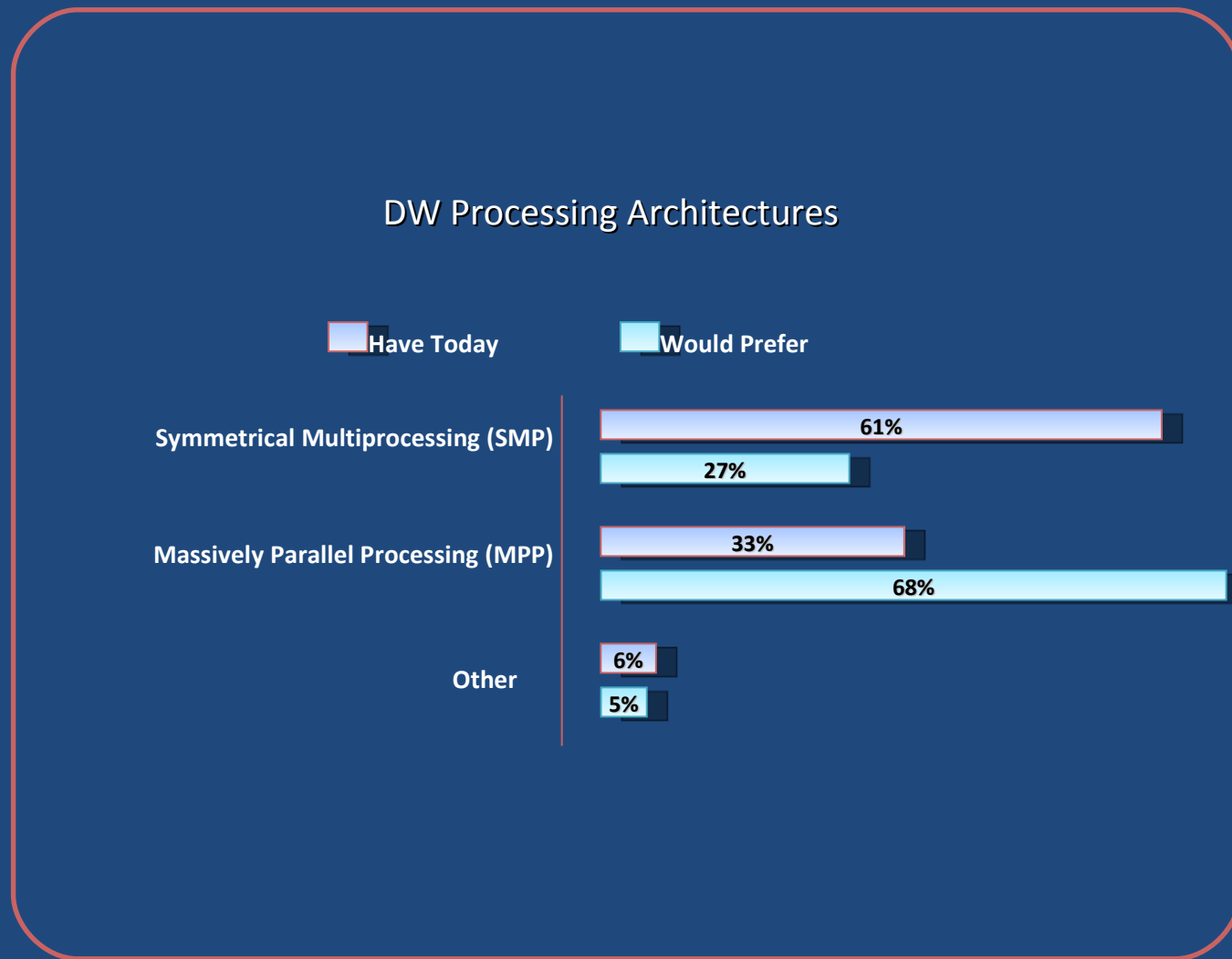
Effect of current recession on DW teams and projects



Source: TDWI Report – Next Generation DW

Ожидания клиентов

- Увеличение объемов данных
- Сокращения бюджетов
- Освоение appliances
- Переход на MPP



Source: TDWI Report – Next Generation DW

Классические решения для хранилищ данных



Построенные самостоятельно



Business DW Appliance



DL370

DL380/5



Microsoft®
SQL Server™ 2008 R2
Fast Track Data Warehouse

DL580/5

DL980/5



Enterprise DW Appliance (PDW)

Усилия на построение	Очень высокие	Очень малые	Средние	Средние	Средние	Средние	Очень малые
Вместимость	Различная	5 TB	14 TB	20 TB	40 TB	80 TB	500 TB
Параллелизм	Различный	Легкий	Легкий	Средний	Средний	Высокий	Очень высокий
Сложность запросов	Различная	Средняя	Средняя	Средняя	Средняя	Высокая	Очень высокая
Архитектура	SMP	SMP	SMP	SMP	SMP	SMP	MPP
Тип решения	-	Программно-аппаратный комплекс	Ссылочная архитектура	Ссылочная архитектура	Ссылочная архитектура	Ссылочная архитектура	Программно-аппаратный комплекс

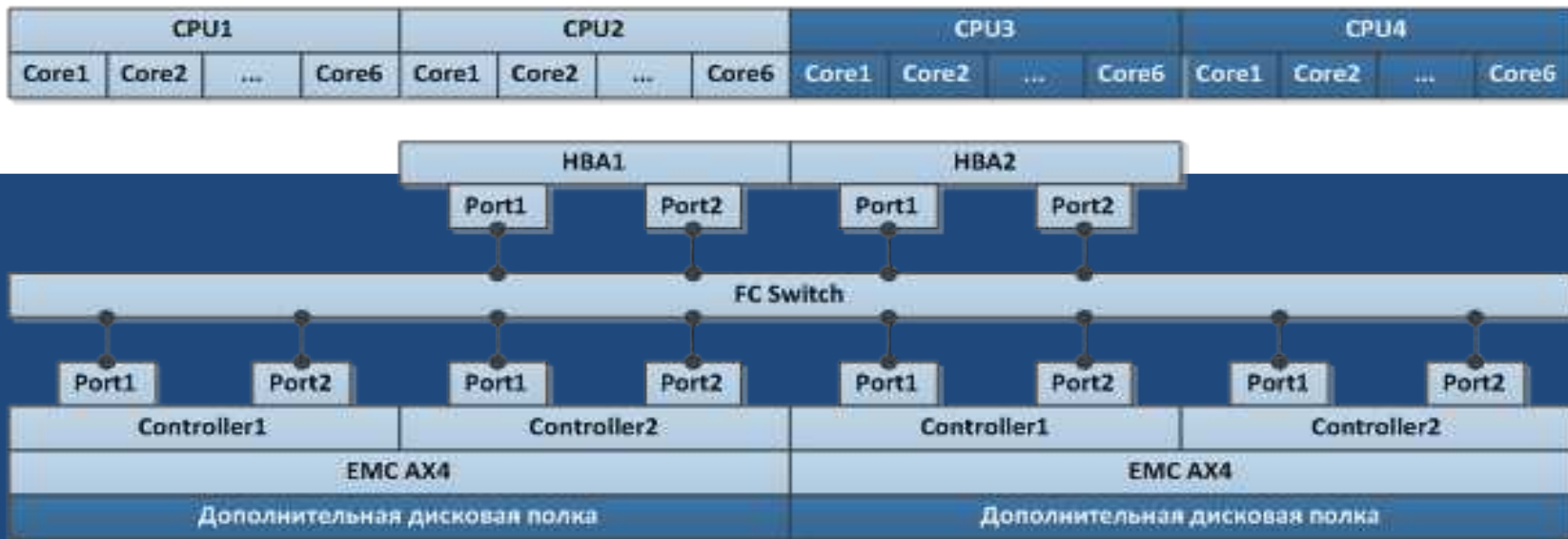
Что такое FastTrack Data Warehouse?

- **Эталонные** «железные» **конфигурации** разработанные с HW vendors
- **Метод** построения эффективной по затратам, сбалансированной системы для нагрузки, типичной для хранилищ данных
- **Рекомендации** размещения, загрузки и управления данными



Performance guide for Microsoft SQL Server
2008 Fast Track Data Warehouse on HP
ProLiant DL385





Компонент	Кол-во компонентов	Подкомпонент	Кол-во подкомпонентов в компоненте	Пропускная способность подкомпонента			Пропускная способность компонента		
				Гбит/с	Гбайт/с	Мбайт/с	Гбит/с	Гбайт/с	Мбайт/с
CPU	4	Ядро	6	1,5625	0,1953125	200	37,5	4,6875	4800
HBA	2	Порт	2	8	1	1024	32	4	4096
FC switch	1	Порт	24	8	1	1024	192	24	24576
AX4 контроллер	4	Порт контроллера	2	4	0,5	512	32	4	4096
Дисковая полка	4	1 HDD x 2 (RAID 1)	4	2,34375	0,29296875	300	37,5	4,6875	4800

Excel калькулятор для помощи в выборе конфигурации

Enter your factors into the green fields and the results will be calculated in the pink cells.

The spreadsheet uses a weighted average to determine the number of cores required based on your inputs.

		Adjust for workload mix	Estimated % of workload	Estimated % data found in SQL Server cache	Estimated Query Data Scan Volume MB (Uncompressed)	Desired Query Response Time (seconds) (under load)	Estimated Disk Scan volume MB (Uncompressed)
Anticipated total number of users expected on the system	3 000 users						
Estimated percent of actual query concurrency	1% concurrency	Simple	70%	10%	8 000	25	7 200
Fast Track DW CPU max core consumption rate (MCR) in MB/s of page compressed data per core	200 MB/s	Average	20%	0%	75 000	180	75 000
Estimated compression ratio (default = 2.5:1)	2,5 :1	Complex	10%	0%	450 000	1 200	450 000
Estimated drive serial throughput speed in compressed MB/s	100 MB/s		100%				
Number of data drives in single storage array	8 drives						
Usable capacity per drive	272 GB						
Space Reserved for TempDB	26%						

Calculations and Results

	% of core consumption rate achieved	Expected per CPU core consumption rate (MB/s)	Calculated Single Query Scan Volume in MB (compressed)	Calculated Target Concurrent Queries	Estimated Target Queries per Hour	Required IO Throughput in MB/s	Estimated Number of Cores Required	Estimated Single Query Run Time (seconds)
Simple	100%	200	2 880	21	3 024	2 419	12,10	0,5
Average	50%	100	30 000	6	120	1 000	10,00	9,4
Complex	25%	50	180 000	3	9	450	9,00	112,5
				30	3 153	3 869	32,00	

Arrays Required based on throughput requirements	Single Array Throughput in MB/s	Throughput in MB/s for All Required Arrays
5	800	4 000

Suggested Fast Track RA Server Requirements

No of CPU cores	Number of arrays	Total Compressed Data Capacity (TB)	Max achievable IO Throughput in MB/s	Max achievable CPU consumption in MB/s	Required IO Throughput in MB/s
32	8	16	6 400	6 400	3 869

Fast Track 3.0 – партнеры по Hardware



Solution Partners for Fast Track 3.0

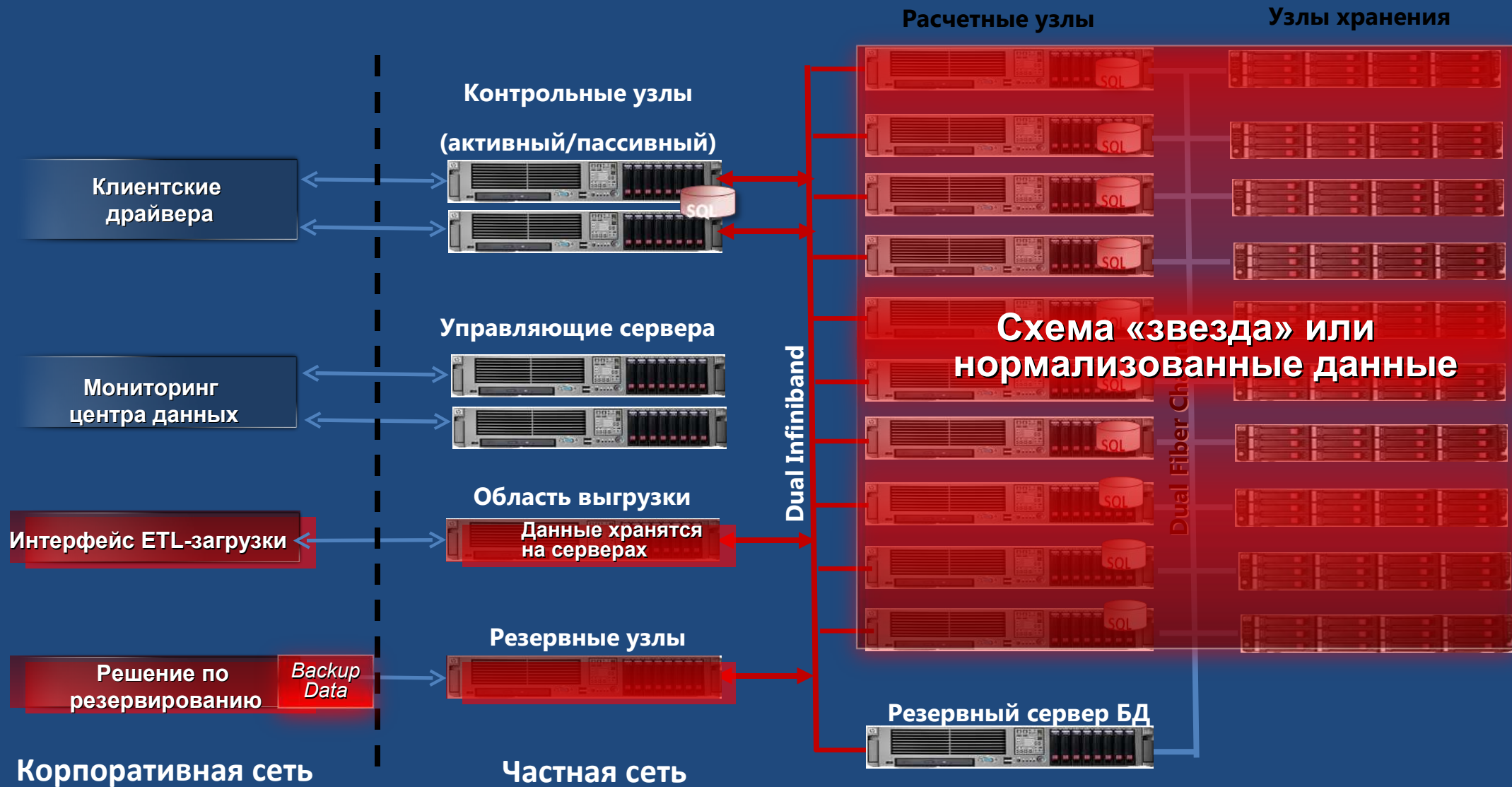


WhereScape **RED**
Builds data warehouses, fast.





Parallel Data Warehouse Appliance



Microsoft[®]

© 2011 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.
MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.