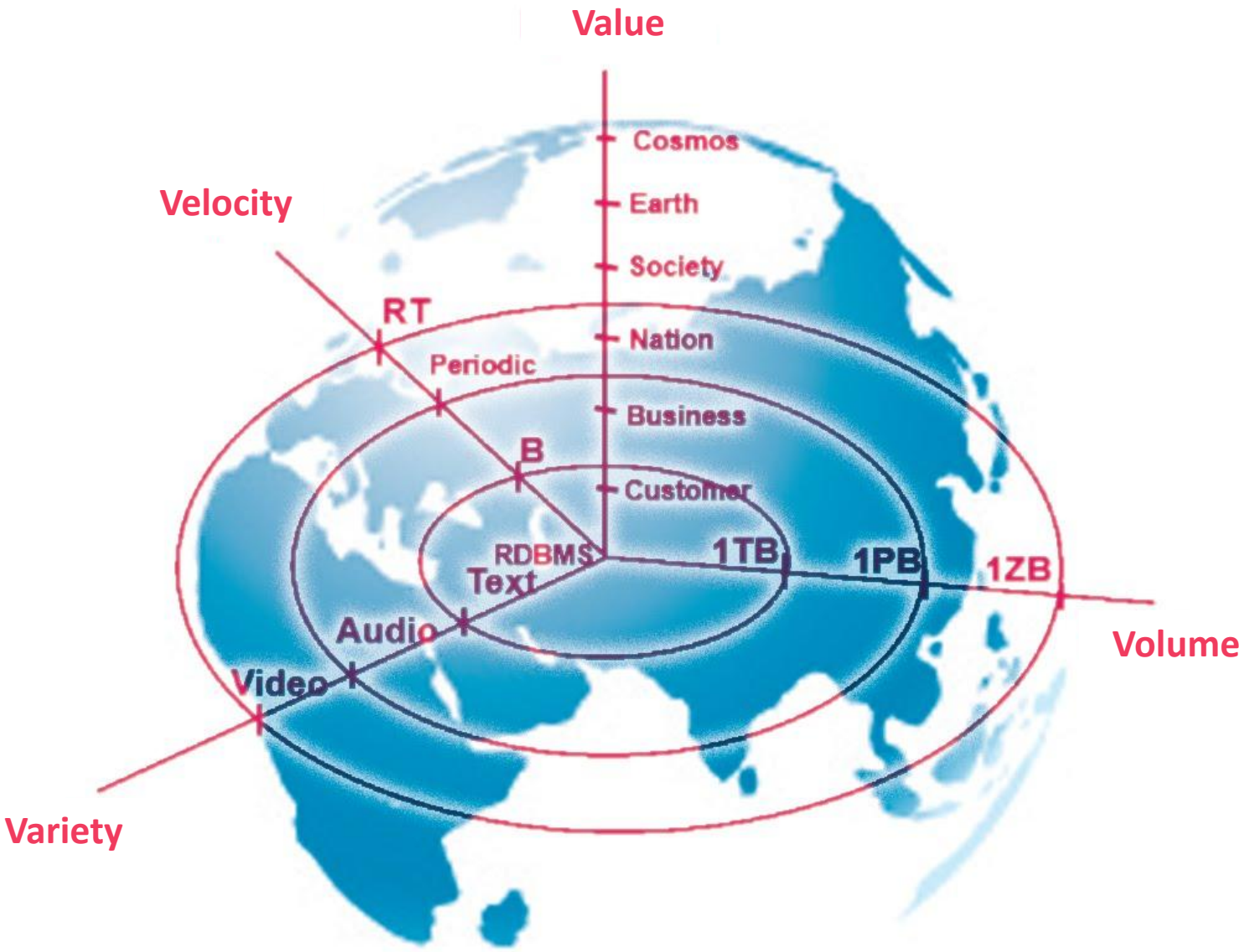
Big Data Analytics and Applications

Pavlovskiy E.N., Ph.D.
head of the Stream Data Analytics and Machine Learning lab NSU
http://bigdata.nsu.ru


Value
Velocity
Volume
Variety
Cosmos
Earth
Society
Nation
Business
Customer
RT
Periodic
B
RDBMS
Text
Audio
Video
1TB
1PB
1ZB

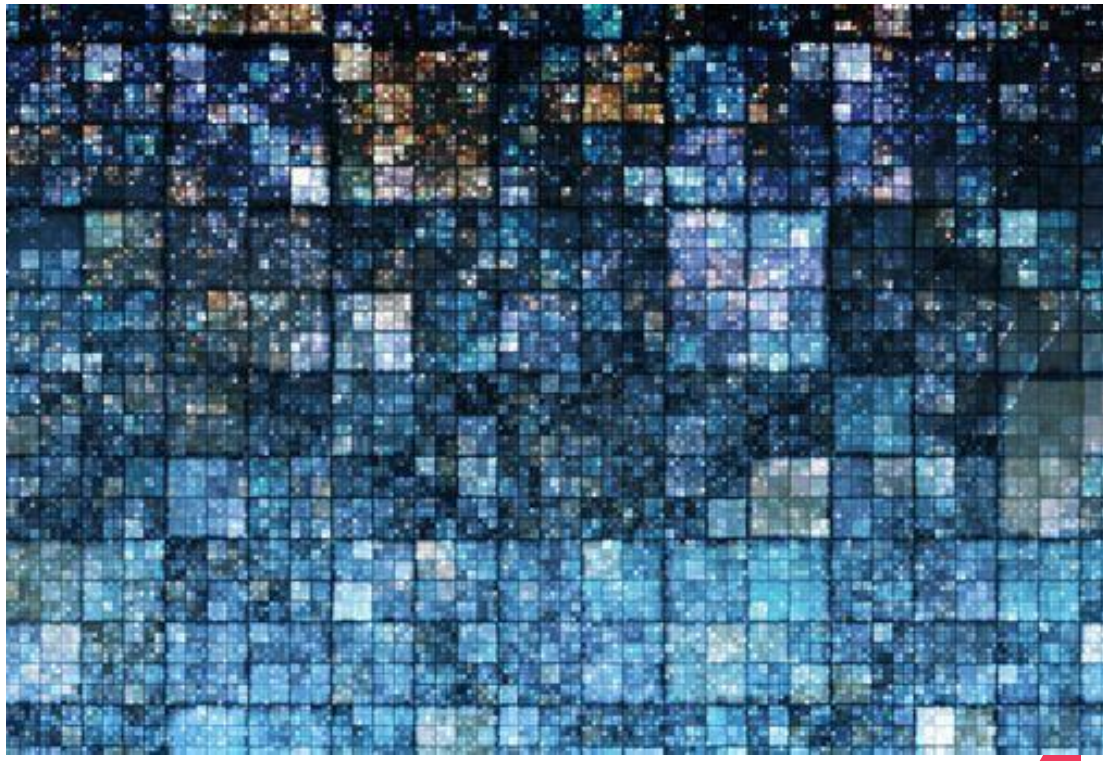
Big Data are not data!

- Technology for gathering, storage, processing, and utilize
- Method of data processing and representation
- Problem of resource lack
- Social phenomenon

- Data of big **v**olume, **v**ariety, **v**elocity, distributed
- Big potential value

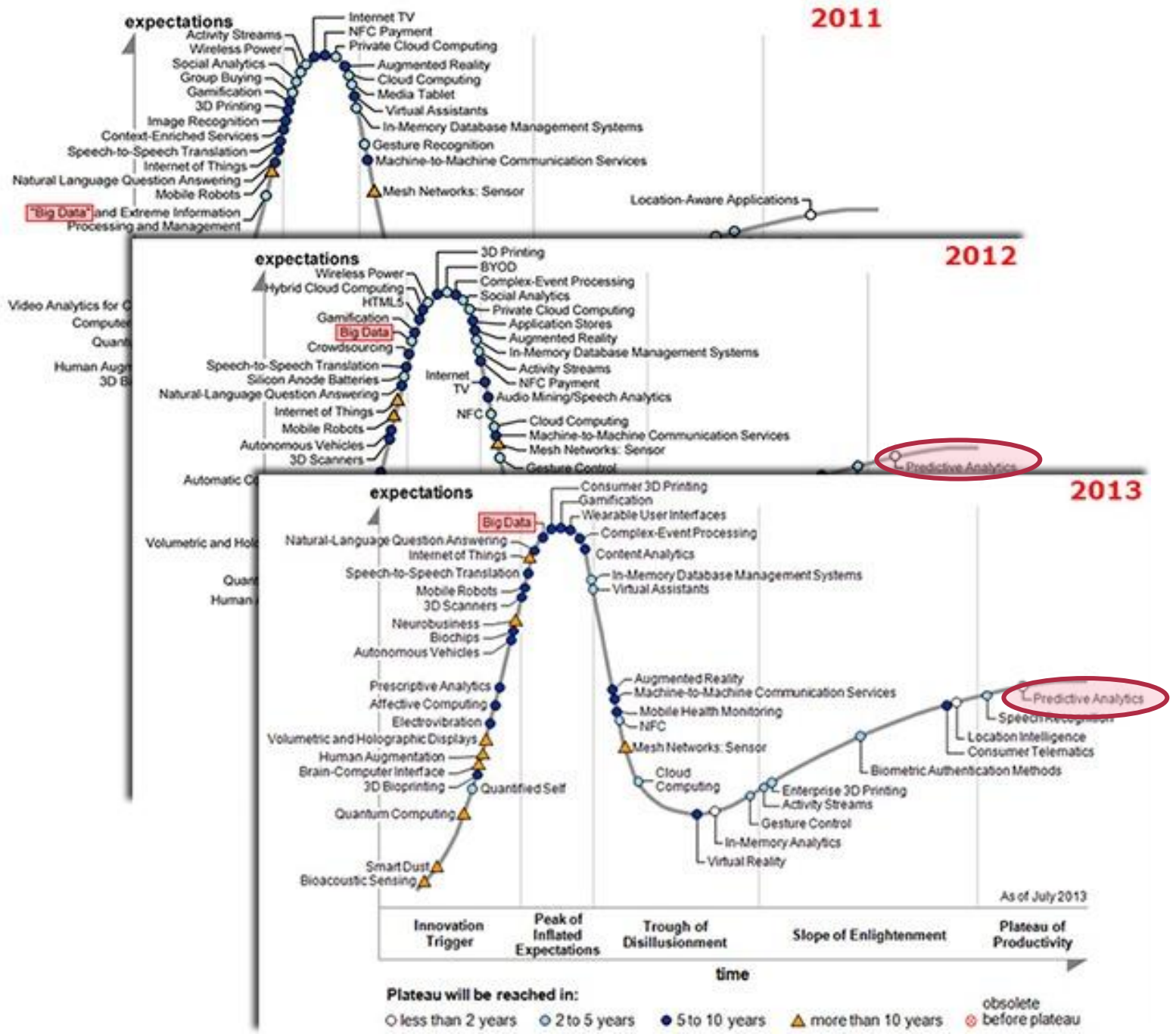
Paradigm shift

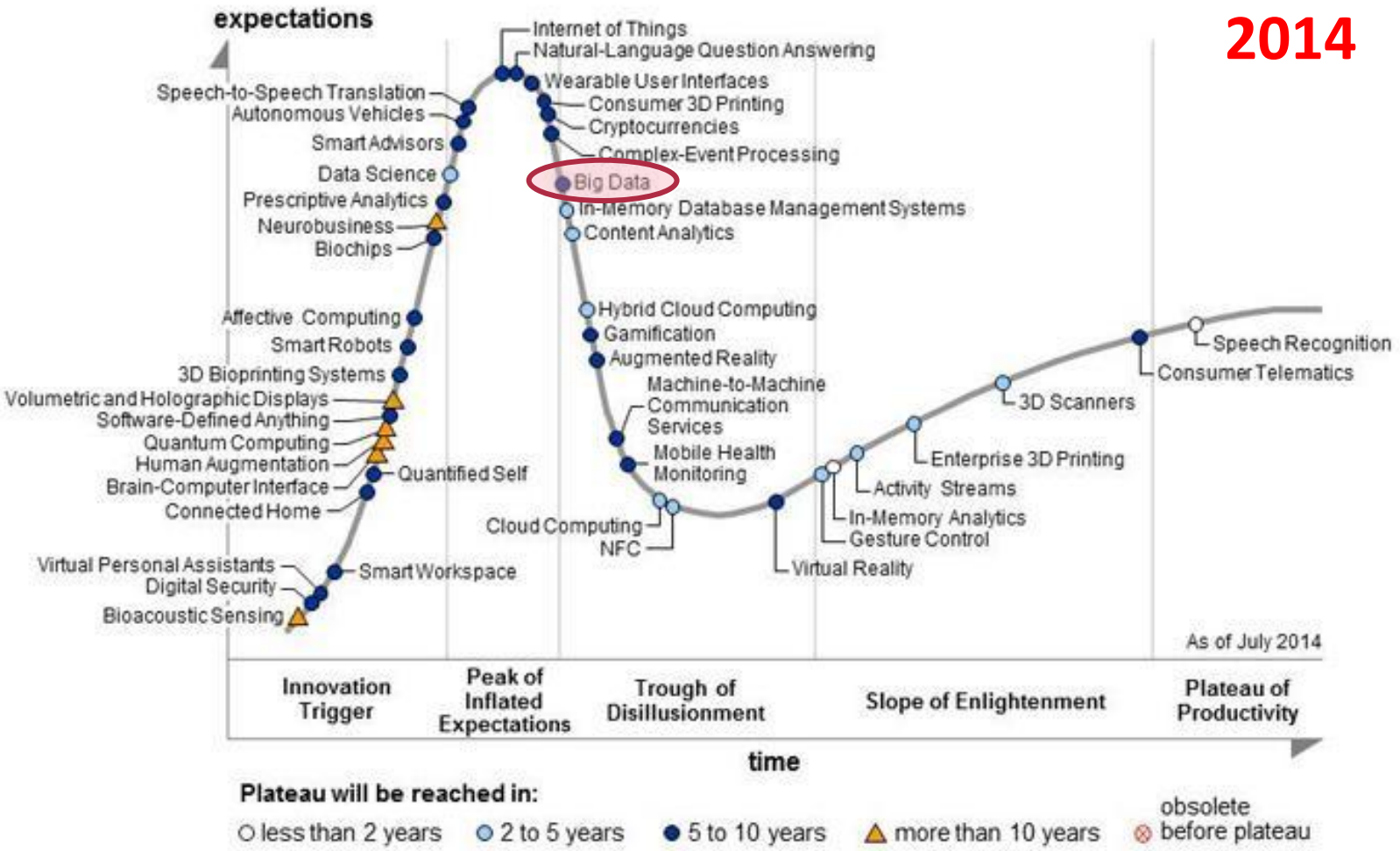
- Subject of labour is not a program, but hypothesis and data



Paradigm shift

- More sources – higher veracity
- More data – higher accuracy
- More data – lower quality requirements
- High-speed algorithms: O(N) or O(NlogN)
- Unmovable data => parallelism and map reduce
- Structure decline => information extraction


2011
expectations
Internet TV
NFC Payment
Private Cloud Computing
Activity Streams
Wireless Power
Social Analytics
Augmented Reality
Group Buying
Cloud Computing
Gamification
Media Tablet
3D Printing
Virtual Assistants
Image Recognition
In-Memory Database Management Systems
Context-Enriched Services
Gesture Recognition
Speech-to-Speech Translation
Machine-to-Machine Communication Services
Internet of Things
Natural Language Question Answering
Mesh Networks: Sensor
Mobile Robots
Location-Aware Applications
"Big Data" and Extreme Information Processing and Management
2012
expectations
3D Printing
Wireless Power
BYOD
Hybrid Cloud Computing
Complex-Event Processing
HTML5
Social Analytics
Gamification
Private Cloud Computing
Big Data
Application Stores
Crowdsourcing
Augmented Reality
In-Memory Database Management Systems
Speech-to-Speech Translation
Activity Streams
Silicon Anode Batteries
Internet TV
NFC Payment
Natural-Language Question Answering
Audio Mining/Speech Analytics
Internet of Things
NFC
Cloud Computing
Mobile Robots
Machine-to-Machine Communication Services
Autonomous Vehicles
Mesh Networks: Sensor
3D Scanners
Gesture Control
Predictive Analytics
2013
expectations
Consumer 3D Printing
Gamification
Big Data
Wearable User Interfaces
Complex-Event Processing
Natural-Language Question Answering
Content Analytics
Internet of Things
In-Memory Database Management Systems
Speech-to-Speech Translation
Virtual Assistants
Mobile Robots
3D Scanners
Neurobusiness
Biochips
Autonomous Vehicles
Augmented Reality
Prescriptive Analytics
Machine-to-Machine Communication Services
Affective Computing
Mobile Health Monitoring
Electrovibration
NFC
Volumetric and Holographic Displays
Mesh Networks: Sensor
Human Augmentation
Cloud Computing
Brain-Computer Interface
Enterprise 3D Printing
3D Bioprinting
Quantified Self
Activity Streams
Quantum Computing
Gesture Control
In-Memory Analytics
Smart Dust
Virtual Reality
Bioacoustic Sensing
Biometric Authentication Methods
Predictive Analytics
Location Intelligence
Consumer Telematics
As of July 2013
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time
Plateau will be reached in:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau


2014
expectations
Internet of Things
Natural-Language Question Answering
Speech-to-Speech Translation
Autonomous Vehicles
Wearable User Interfaces
Consumer 3D Printing
Cryptocurrencies
Smart Advisors
Complex-Event Processing
Data Science
Big Data
Prescriptive Analytics
In-Memory Database Management Systems
Neurobusiness
Content Analytics
Biochips
Affective Computing
Hybrid Cloud Computing
Smart Robots
Gamification
Augmented Reality
3D Bioprinting Systems
Machine-to-Machine Communication Services
Volumetric and Holographic Displays
Software-Defined Anything
Quantum Computing
Human Augmentation
Quantified Self
Mobile Health Monitoring
Brain-Computer Interface
Connected Home
Cloud Computing
NFC
Virtual Personal Assistants
Smart Workspace
Digital Security
Bioacoustic Sensing
Virtual Reality
In-Memory Analytics
Gesture Control
Activity Streams
Enterprise 3D Printing
3D Scanners
Speech Recognition
Consumer Telematics
As of July 2014
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time
Plateau will be reached in:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau

2015


expectations
Advanced Analytics With Self-Service Delivery
Autonomous Vehicles
Internet of Things
Smart Advisors
Speech-to-Speech Translation
Machine Learning
Micro Data Centers
Digital Dexterity
Wearables
Software-Defined Security
Cryptocurrencies
Neurobusiness
Consumer 3D Printing
Citizen Data Science
Natural-Language Question Answering
Biochips
IoT Platform
Connected Home
Affective Computing
Smart Robots
3D Bioprinting Systems for Organ Transplant
Hybrid Cloud Computing
Volumetric Displays
Human Augmentation
Brain-Computer Interface
Augmented Reality
Quantum Computing
Enterprise 3D Printing
Gesture Control
Bioacoustic Sensing
Virtual Reality
Cryptocurrency Exchange
Autonomous Field Vehicles
People-Literate Technology
Digital Security
Virtual Personal Assistants
Smart Dust
As of July 2015
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time
Plateau will be reached in:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau

Problems in Russian Big Data

- No depersonalization culture (FL-152)
- No understanding of potential value
- Insufficient competence in statistics
- Absence of data brokers
- Highly risked data analytics projects
- Lack of data

Big Data education in Russia

10

Master programs

HSE:

- Big Data Systems
- Data Sciences

MSU:

- «Intellectual analysis of big data»
- «Big Data: infrastructure and solution technique»

NSU

- Big Data Analytics
- Computer modeling

Online

1 week to 1 year

- Coursera, edX (http://rusbase.com/list/bigdatye-kursy/)
- Intuit (Introduction to Big Data Analytics) http://bit.ly/IntuitBDA

ИНТУИТ
НАЦИОНАЛЬНЫЙ ОТКРЫТЫЙ УНИВЕРСИТЕТ

Additional education

1 week - 3 month - 2 years

- Yandex Data Analysis School – https://yandexdataschool.ru/

- Digital October – http://newprolab.ru

- Beeline - http://bigdata.beeline.digital/databa

- Expasoft – http://expasoft.com/edu/

NSU Big Data Strategy

Online courses	• Wide audience • Leads to offline
Master of Sciences (10-20 per year)	• Mobility • For industry and science
Ph.D. (5-10 per year)	• Scientific schools
Additional study (20 – 100 per year)	• MBA

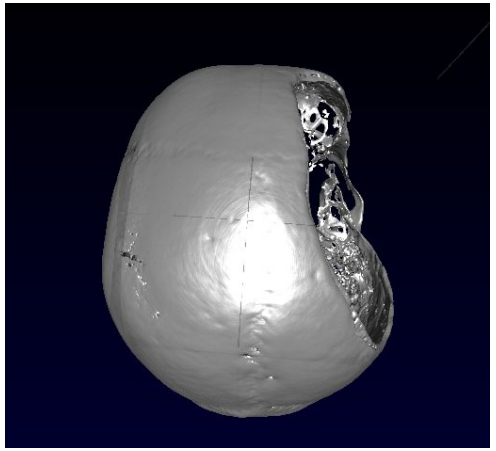
Syllabus of Master program

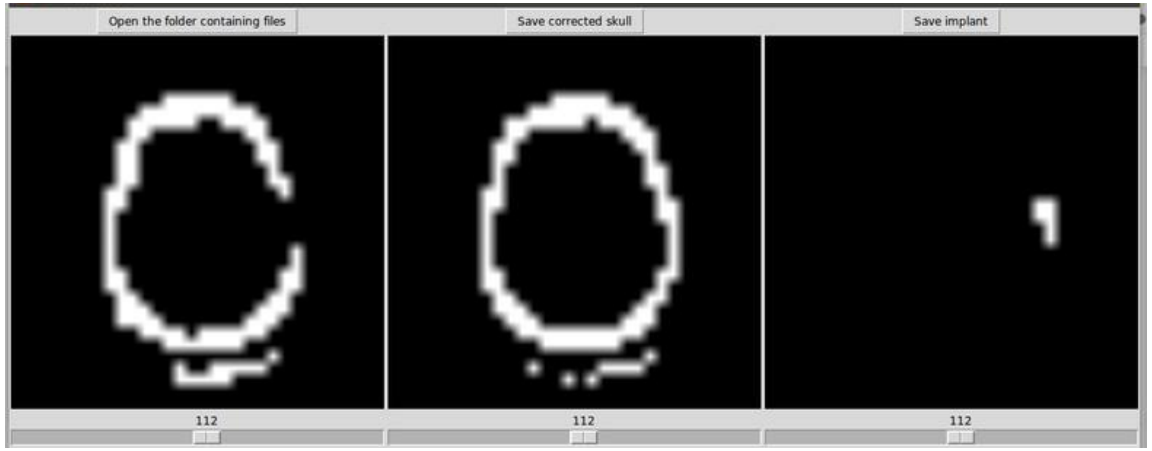
	1st year. Known tech						2nd year. Innovate		
	1 Business			2 Science			3 Management		4 Thesis
	Business understanding	Ready solutions	Scripting	Access to data	Mining	Presenting	Deployment	Scaling	Final State Certification
BA	Business Analy BGoals, Commu		Business Analy Requirement	Business Analys		Presenting to SH		Marketing	
Engineering	Business Cases	Excel, Gree Deductor, V	SE: Programm (Python, R), Prot	Storage technol	Big Data Develop Environmets	Processing big data w cloud-based technolog	Virtualization an Consolidation	Couds	
Management	Project management							Product management	
Math	Knowledge presentation	Machine Lear	Operations Res		Machine Learni	Visualization methods and tools w practice	Fuzzy logic and rou computing		
Advancing	Theory of Constra		Enterpreneursh	Juridical issues	Decision making th		Enterpreneurship 2	Technology trer	
Elective business domain	Social networks analysis / Bioinformatics / Cognitive Data Mining / Instrumentation / Healthcare / Telecom								

Challenges

- 1st place, 2015, AVITO
- 1st place, 2015, eKapusta
- 4th place among 619 teams, 2009, Data Mining Cup

Skull surface restore





No formulae
No negative examples
Neural networks,
autoencoders

Deep learning

Unsupervised

Semantic segmentation

http://arxiv.org/pdf/1511.00561v2.pdf

Van Gogh Iv

Alex J. Champandard. Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks. 2016

Paintings

http://tinyclouds.org/colorize/

Articles for revision

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the compariciely in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ??. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M^\bullet} = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example ??. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma ??. Namely, by Lemma ?? we see that R is geometrically regular over S.

Lemma 0.1. *Assume (3) and (3) by the construction in the description.*

Suppose $X = \lim |X|$ *(by the formal open covering* X *and a single map* $\underline{Proj}_X(\mathcal{A}) = \mathrm{Spec}(B)$ *over* U *compatible with the complex*

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

When in this case of to show that $\mathcal{Q} \to \mathcal{C}_{Z/X}$ *is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition ?? (without element is when the closed subschemes are catenary. If* T *is surjective we may assume that* T *is connected with residue fields of* S*. Moreover there exists a closed subspace* $Z \subset X$ *of* X *where* U *in* X' *is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem*

(1) f *is locally of finite type. Since* $S = \mathrm{Spec}(R)$ *and* $Y = \mathrm{Spec}(R)$.

Proof. This is form all sheaves of sheaves on X. But given a scheme U and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\ldots,n} U_i$ be the scheme X over S at the schemes $X_i \to X$ and $U = \lim_i X_i$. □

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{\mathcal{X},\ldots,0}$.

Lemma 0.2. *Let* X *be a locally Noetherian scheme over* S, $E = \mathcal{F}_{X/S}$. *Set* $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. *Since* $\mathcal{I}^n \subset \mathcal{I}^n$ *are nonzero over* $i_0 \leq \mathfrak{p}$ *is a subset of* $\mathcal{J}_{n,0} \circ \overline{A}_2$ *works.*

Lemma 0.3. *In Situation ??. Hence we may assume* $\mathfrak{q}' = 0$.

Proof. We will use the property we see that $\mathfrak{p}$ is the mext functor (??). On the other hand, by Lemma ?? we see that

$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where K is an F-algebra where δ_{n+1} is a scheme over S. □

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Pushkin A.I.

Зафонствуя попруг,
Ивисшивый чела,
На воспопе днего,
Я могина бесслужел,
Катирей свети довой,
Из увядебиле меня,
И на гразой шле, далодной
Вольностью примстают;
Я, водешил перцов миренья?

N.I. Putincev, stream data analytics and machine learning lab

Thank you!

http://bigdata.nsu.ru

Evgeniy Pavlovskiy,
head of the SDAML
N*SU
e@expasoft.ru,
+79139117907