INFOCOMP 2014

Achievements and Successful Solutions with Big Data and Computing Challenges:

National and International Perspectives

July 23, 2014, Paris, France

The Fourth International Conference on Advanced Communications and Computation (INFOCOMP 2014)

IARIA

INFOCOMP / DataSys
July 20–24, 2014 - Paris, France

INFOCOMP International Expert Panel: Big Data & Computing Challenges

Panelists

- *Claus-Peter Rückemann* (Moderator), Westfälische Wilhelms-Universität Münster (WWU) / Leibniz Universität Hannover / North-German Supercomputing Alliance (HLRN), Germany
- *John Ashley,* NVIDIA, U.K.
- *Małgorzata Pańkowska,* Department of Informatics, University of Economics in Katowice, Poland
- *Janna L. Maltseva,* Lavrentyev Institute of Hydrodynamics, Novosibirsk State University, Russia
- *Shangping Ren,* Illinois Institute of Technology, U.S.A.
- *Elena A. Troubitsyna,* Åbo Akademi University-Turku, Finland

INFOCOMP 2014: `http://www.iaria.org/conferences2014/INFOCOMP14.html`

INFOCOMP International Expert Panel: Big Data & Computing Challenges

Related Topics for Audience Discussion:

- Big Data,
- Advanced Applications,
- High End Computing,
- Scientific Computing, ...

- Long-term issues,
- Data structuring,
- Work in communities and disciplines,
- Big data and sustainability,
- Integrated systems, ...

- Knowledge discovery,
- Prosumption, ...

- Auditing of acquisition/operation,
- Operation and service,
- Quality of Service, ...

- Intelligent application components,
- Learning systems,
- Innovative infrastructure and network developments,
- Heterogeneous resources and architectures, ...

- Legal aspects and funding,
- Physical issues,
- Parallelisation and localisation,
- Collaboration and access,
- Documentation, learning, and teaching, ...

INFOCOMP International Expert Panel: Big Data & Computing Challenges

Panel Statements:

- **Key-terms:** Big Data, Volume, Variability, Velocity, Vitality, Veracity, . . . sustainable long-term knowledge discovery, multi-disciplinary context, . . .
- **High End Computing:** Close integration of different "V" requires a new type of resources.
- **Optimisation: Structure over computation!**
- **Long-term data resources:** Multi-disciplinary, universal classification, long-term vitality, sustainability, . . .
- **Resources management:** Reduce complexity from planning to operation, with hardware and software.
- **Data wall:** HPC and Big Data facing latency and bandwith more then math - attention to data structures.
- **Counter trend:** Increased complexity, increased math.
- **Scaling:** Move computing to the data, data/info management.
- **Parallelisation:** Become communication minimizing, latency-fault-missing data tolerant.
- **Legal:** Sarbanes-Oxley (SOX) & 'Security' acts, Basel, USA/Europe.

INFOCOMP International Expert Panel: Big Data & Computing Challenges

Pre-Discussion-Wrapup:

- **Definitions:** Big, Data, quality, quantity, HPC, ...?
- **Disciplines:** Algorithms, scenarios, ...?
- **Focus:** Data organisation or computing and algorithms?
- **Handling:** General solutions and recommendations?
- **Communities and differences:** Are there differences in national and international context?
- **Status:** Present and future development and deployment?
- **Architectures:** Shared, parallel, distributed, ...?
- **Application scenarios:** Developments, software, hardware?
- **Sustainability:** How can sustainable big data solutions be created?
- **Lessons learned:** Results, operated solutions, approaches in development?
- **Call for Collaboration and Networking:** Open research calls?

INFOCOMP International Expert Panel: Post-Panel-Discussion Summary

Post-Panel-Discussion Summary (2014-07-23):

- No general solutions available (whatever said). No application-ready toolbox.
- Elementary technical components for different ("V"-scenarios) in development.
- ... strong demands from research and business.
- Knowledge is important long-term value. Structure, conceptual knowledge and understanding, and information management of increasing importance.
- Data, information, knowledge - individual algorithms required for value creation.
- Characteristics of "data" and applications make logical/physical preconditions. Technical, communication, application environments contribute. Mostly all scenarios require to consider hardware and software.
- HPC and Big Data facing latency and bandwith "Data wall" more then math – attention to data structures and localities. On other scenarios increased complexity and increased math.
- Scaling: Move computing to the data. Information management. Parallelisation: Become communication minimizing, "latency-fault-missing data" tolerant.
- Tradeoff between usability/performance acceptable in many cases. High level environments can provide good, effective solutions ("solution over performance").
- Agreements, legal regulations have to support Big Data solutions internationally.
- Education required on understanding aspects behind Big Data.
- Collect general solutions on application scenarios (next collaborations).

INFOCOMP International Expert Panel: Table of Presentations, Attached

Panelist Presentations: (presentation order, following pages)

International Expert Panel INFOCOMP 2014

Achievements and Successful Solutions with Big Data and Computing Challenges: National and International Perspectives

Big Data – Bigger Knowledge – Biggest Cognition: Optimising Organisation & Structure for Exa-Scale

The International Conference on Advanced Communications and Computation (INFOCOMP 2014)
July 23, 2014, Paris, France

IARIA

Dr. rer. nat. Claus-Peter Rückemann[1,2,3]

[1] Westfälische Wilhelms-Universität Münster (WWU), Münster, Germany
[2] Leibniz Universität Hannover, Hannover, Germany
[3] North-German Supercomputing Alliance (HLRN), Germany

ruckema(at)uni-muenster.de

Westfälische Wilhelms-Universität Münster

GEXI

ACTMAP

Status: Exa-Scale High End Data vs. High End Computing

Gilding data: Wish and realities?

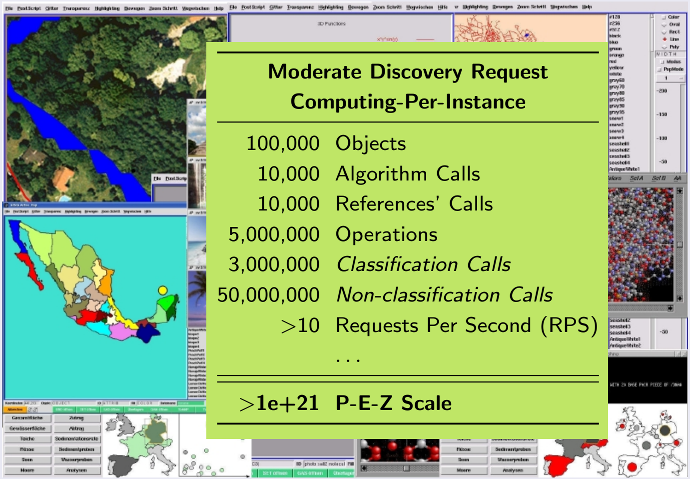
Big Data	–	Bigger Knowledge	–	Biggest Cognition?
Big Data	–	Bigger Efficiency	–	Biggest Cost Reduction?
Big Data	–	Bigger Shares	–	Biggest Profits?

High End Data ...

- **Big data**
  - **Long-term content and knowledge aspects**:
    Vitality, ...
  - **Scenario and application aspects**:
    Volatility, ...
  - **Physics, technical aspects**:
    Volume, Velocity, Variability, Veracity, ...
- **Long-term storage, archiving, and resources**
- **Sustainable efficiently computation-aware knowledge**

Big Data: Science, Content, Implementation, Computation, Operation

Example: Integrated Information and Computing Systems IICS (GEXI case study)



Moderate Discovery Request
Computing-Per-Instance

100,000	Objects
10,000	Algorithm Calls
10,000	References' Calls
5,000,000	Operations
3,000,000	*Classification Calls*
50,000,000	*Non-classification Calls*
>10	Requests Per Second (RPS)
	...
>1e+21	**P-E-Z Scale**

Data objects

- vector
- raster
- aerial
- photo
- spatial
- calculation
- measurement
- processing
- meta objects
- interactive
- commercial
- license
- ...

Big Data: Universal Decimal Classification

Example excerpt of Universal Decimal Classification (UDC) codes:

UDC Code	Description (English)
UDC:55	Earth Sciences. Geological sciences
UDC:56	Palaeontology
UDC:911.2	Physical geography
UDC:902	Archaeology
UDC:903	Prehistory. Prehistoric remains, artefacts, antiquities
UDC:904	Cultural remains of historical times
UDC:25	Religions of antiquity. Minor cults and religions
UDC:930.85	History of civilization. Cultural history
UDC:"63"	Archaeological, prehistoric, protohistoric periods and ages
UDC:(7)	North and Central America
UDC:(23)	Above sea level. Surface relief. Above ground generally. Mountains
UDC:(24)	Below sea level. Underground. Subterranean
UDC:=84/=88	Central and South American indigenous languages

- Multilingual Universal Decimal Classification Summary (2012). UDC Consortium, 2012, Web resource, v. 1.1. The Hague: UDC Consortium (UDCC Publication No. 088), Retrieved January 12, 2014, from http://www.udcc.org/udcsummary/php/index.php.
- Creative Commons Attribution Share Alike (2012). Creative Commons Attribution Share Alike 3.0 license, Retrieved January 12, 2014, from http://creativecommons.org/licenses/by-sa/3.0/.

Big Data: Knowledge Top Level

What is knowledge :: Universal Decimal Classification (UDC) excerpt:

UDC Code	Description (English)
UDC:0	Science and knowledge. Organization. Computer science. Information.
UDC:00	Prolegomena. Fundamentals of knowledge and culture. Propaedeutics
UDC:001	Science and knowledge in general. Organization of intellectual work
UDC:002	Documentation. Books. Writings. Authorship
UDC:003	Writing systems and scripts
UDC:004	Computer science and technology. Computing
UDC:004.2	Computer architecture
UDC:004.3	Computer hardware
UDC:004.4	Software
UDC:004.5	Human-computer interaction
UDC:004.6	Data
UDC:004.7	Computer communication
UDC:004.8	Artificial intelligence
UDC:004.9	Application-oriented computer-based techniques
UDC:005	Management
UDC:005.1	Management Theory
UDC:005.2	Management agents. Mechanisms. Measures
UDC:005.3	Management activities
UDC:005.5	Management operations. Direction
UDC:005.6	Quality management. Total quality management (TQM)

Vision – Optimise the Optimisation for Exa-Scale

Optimisation of Data Structures and Knowledge Resources!

- In Disciplines, Big Data should result in Bigger Knowledge instead of Bigger Resources or flat technical optimisation only,
- Fostering content vitality,
- Optimising data structures and organisation,
- Universal classification,
- Deployment of systematics and methodologies with content,
- Long-term sustainability of knowledge discovery, multi-disciplinary, multi-lingual content.

National and international means . . .

- Knowledge recognition (expertise) and decision making,
- Knowledge documentation, general plan for classification,
- Mandatory data centric tasks combined with project funding,
- Funding of researchers instead of institutions and resources' providers,
- Multi-disciplinary, multi-lingual collaboration,
- Long-term efforts on education (science and disciplines).

Conclusions

Optimisation of organisation and structure . . .

Selected challenges and deficits: (as identified by the INFOCOMP 2012 Panel on High End Systems and last years' INFOCOMP Panel on Exa-Intelligence)

- *Intelligent components, knowledge discovery, learning systems!* in the context of:
- Integrating hardware / software solutions, outlasting High Performance Computing solutions!
- Scalability, fast and massive I/O and communication solutions, advanced networks!
- Automation and autonomous components! Reliability and efficiency in operation!
- Education and teaching!

Optimisation of organisation and structure, long-term multi-lingual universal knowledge and computation:

- **Knowledge recognition** (expertise).
- **Knowledge documentation** (for any aspect of nature and society).
- **Long-term means**.
- **Multi-disciplinary work** (content, context, knowledge).
- **Sciences and disciplines** (expertise in different disciplines).
- **Complexity and intelligence** (holistic knowledge and components).
- **Integrated Information and Computing System components**.

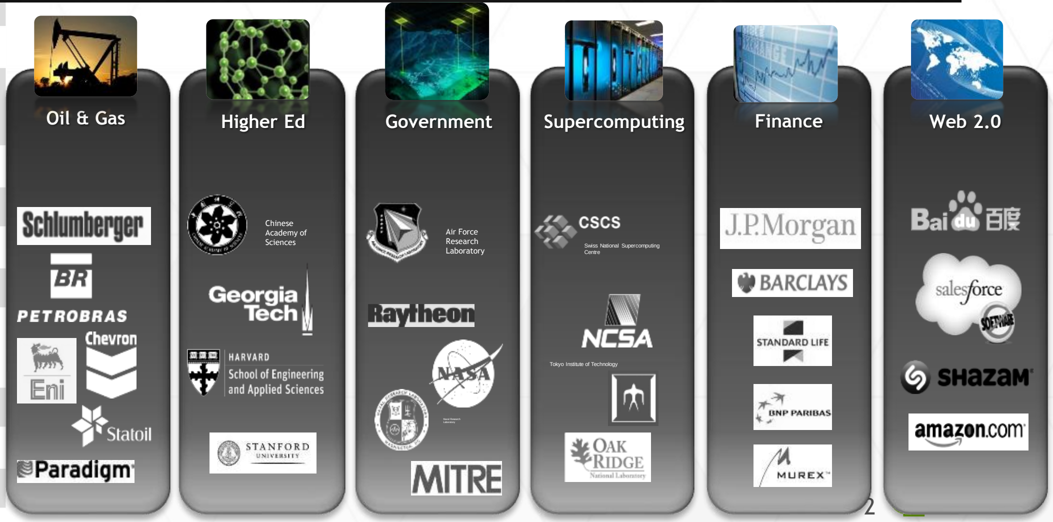
HPC & BIG DATA CHALLENGES

John Ashley, Senior Solutions Architect *

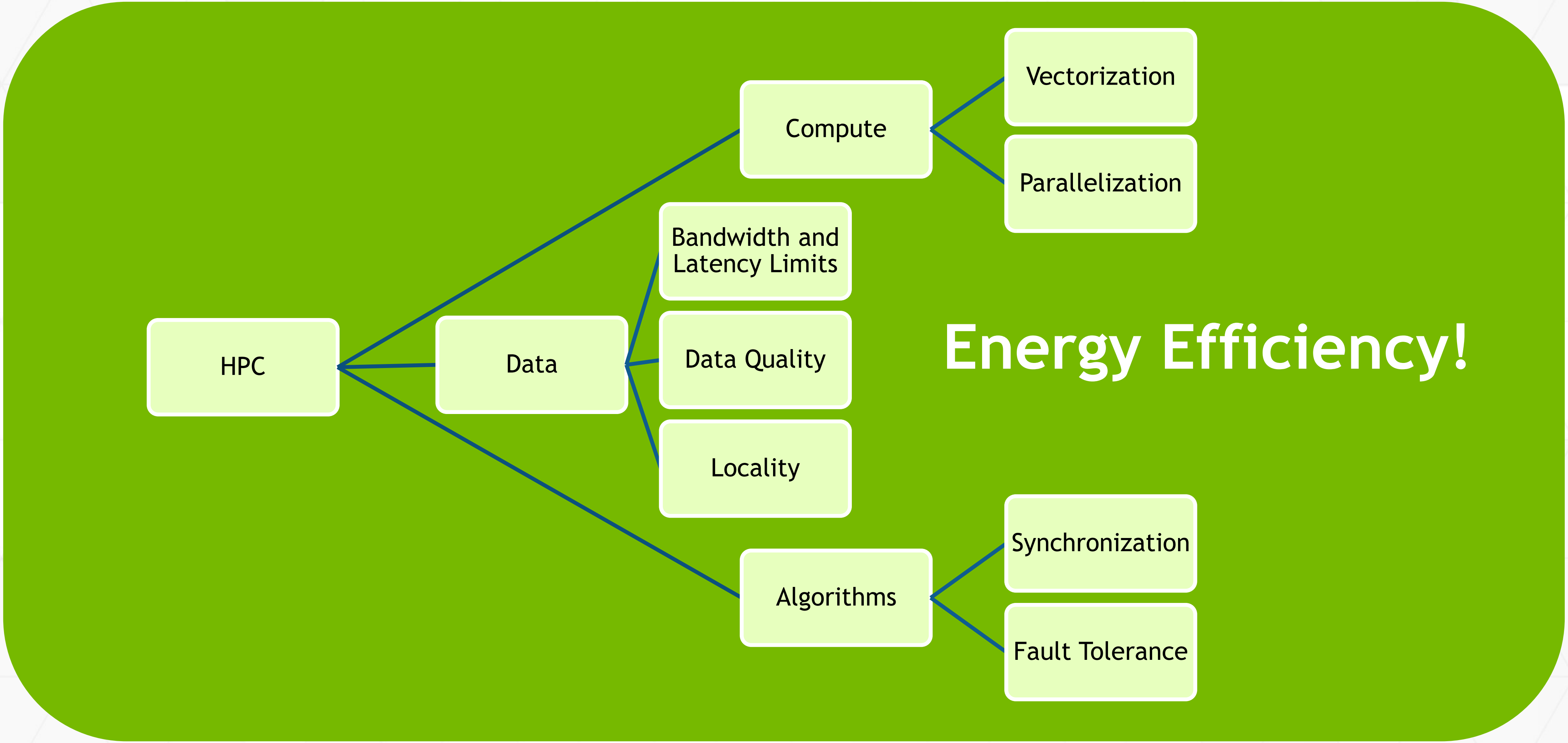
*These views are mine and do not necessarily reflect those of NVIDIA.

PERSPECTIVE

Green500 Rank	MFLOPS/W	Site
1	4,390	TSUBAME-KFC - GSIC Center, Tok Technology
2	3,632	Cambridge University
3	3,518	Center for Computational Sciences, University of Tsukuba
4	3,459	SURFsara - Netherlands
5	3,186	Swiss National Supercomputing Centre (CSCS)
6	3,131	ROMEO HPC Center - Champagne-Ardenne
7	3,020	CSIRO
8	2,952	TSUBAME 2.5 - GSIC Center, Tokyo Institute of Technology
9	2,813	ENI S.p.A. – Italian Energy Corporation
10 - 14	2,629	Financial Institutions
15	2,629	Max-Planck-Gesellschaft MPI/IPP



“BIG COMPUTE” & “BIG DATA” CHALLENGES



RESPONSES?

Code
- Refactor to expose existing parallelism
- Switch to vector friendly data structures

Algorithms
- Parallel
- Asynchronous
- Communications avoiding

Systems
- Move compute to data (over provison)
- Fault tolerant design

University of Economics in Katowice

your place

your space

your future

Service Science Facing Big Data

Malgorzata Pankowska
Member of the Board
ISACA Chapter
CGEIT Coordinator, Research Director

Big Data

Back office Applications

New Sources

- **Sensors producing readings**
- **Internet**
- **Social Media Services**
- **Customer Conversations**
- **Unstructured Public and Internal Sources**

Interactive Analytics, Exploration and Discovery

Planning, Acquisition, Processing and Integration

Visualization, Story Telling and Collaboration

Real-time Decision Management

Systems of Engagement, Sentiment Analysis

Big Data premises

- Business model transformation by globalization and connectivity
- Company movement from being product oriented to service oriented
- Service personalization as business maturity measure
- New sources of data
- Technology: mobile devices, large-scale networks, virtualization
- Translation user data management needs into service needs

↓

Big Data characteristics

data volume, velocity, variety, ambiguity, complexity

Big Data architecture management plan

- Focus on the enterprise's culture, maturity and strategy
- Work with key stakeholders to understand their data needs
- Improve the analytical support team
- Translate user requirements into services
- Develop Big Data architecture
- Invest to improve technology, staff, systems and processes
- Consider Big Data as a Service
- Implement Service Level Management

Servitization

- present transdisciplinary and multisectoral perspectives on the nature of service systems, on research and practice in service, and on the future directions to advance service science [Demirkan et al., 2011]
  - **service** = value co-creation
  - **services** in a service-oriented environment are jobs, duties, tasks and activities that a business or a service provider offers to customers or consumers.
  - **in technology aspect, services** mean software applications, methods, operations or communications between two computing objects, or the interface between two software components

Service Level Management

- process forms a crucial link between the service provider and the business.
- Service Level Agreements (SLAs) with the Business, Operational Level Agreements (OLAs) with internal organizations, and Underpinning Contracts (UCs) with third party suppliers
- **Service innovation life cycle**
  - Service Exploration: Possible scenarios are generated and considered to analyze changes from a business strategy viewpoint.
  - Service Engineering: One or more options are explored in more detail.
  - Service Management: IT Infrastructure and testing phase.

References

- **Ballard Ch., Compert C., Jesionowski T., Milman I, Plants B., Rosen B., Smith H.: Information Governance Principles and Practices for a Big Data Landscape, IBM, NY, March 2014.**
- **Zikopoulos P.C., Eaton Ch., deRoos D., Deutsch T., Lapis G.: Understanding Big Data , Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill NY 2012**
- **Krishnan K.: Data Warehousing in the age of Big Data, Morgan Kaufmann, Elsevier, Amsterdam, 2013**
- **Uden L., L.S.L Wang, J.M.Corchado Rodriguez, H-Ch. Yang, I-H. Ting. Eds. The 8th International Conference on Knowledge Management in Organization, Social and Big Data Computing for Knowledge Management, Springer Dordrecht, 2014**
- **M.Chen., S.Mao., Y.Zhang., V..M.Leung , Big Data, Related technologies, challenges and future prospects, Springer Cham Heidelberg, 2014,**
- **Gottlob G., Grasso G., Olteanu D., Schallhart Ch (eds.) Big Data, Springer, Berlin, 2013,**
- ***http://myria.cs.washington.edu/projects.html J. Ortiz, V.T. Almeida, M. Balazinska.* A Vision for Personalized Service Level Agreements in the Cloud, Workshop on Data Analytics in the Cloud (DanaC), with SIGMOD/PODS 2013**
- **http://www.emc.com/collateral/software/white-papers/h10839-big-data-as-a-service-perspt.pdf Big Data-as-a-Service: A Marketing and Technology Perspective, EMC Solutions Group, July 2012.**

Mathematics&Mathematica

Janna L. Maltseva,
Lavrentyev Institute of Hydrodynamics,
Novosibirsk State University,
Russia

Janna Maltseva

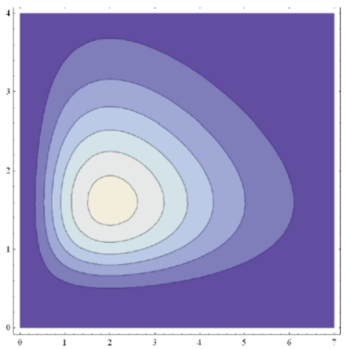
- **Areas of Interest**
  - Fluid Mechanics
  - Stratified Flows
  - Nonlinear Waves
  - Mathematical Modelling

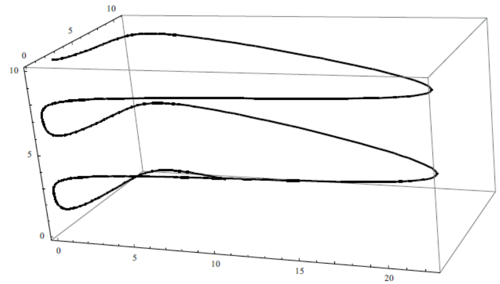
…….

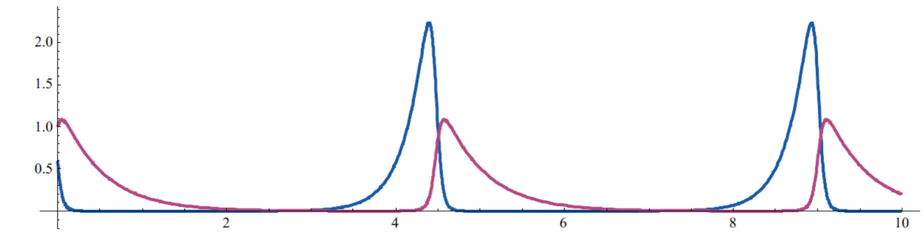
- **Senior Researcher** (Lavrentyev Institute of Hydrodynamics)
  Semi-analytic Methods (Wolfram Mathematica / Asymptotic methods)
- Associate Professor (Novosibirsk State University),
  Department of Natural Sciences (Chemistry and Biology)

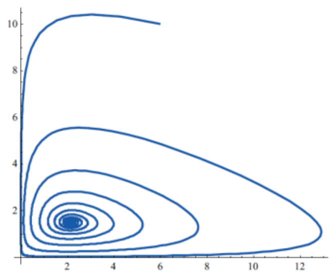
- **Topics for students**:
  - Conservation Equations in Chemical Kinetics,
  - Math Models of Chromatography,
  - Differential equations of Math physics (Wave eqs., Thermal Conductivity, Theory of Oscillations …)
  - Simulation of Populations

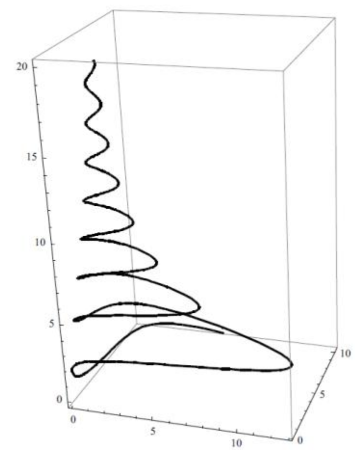
Population models "predator – prey"

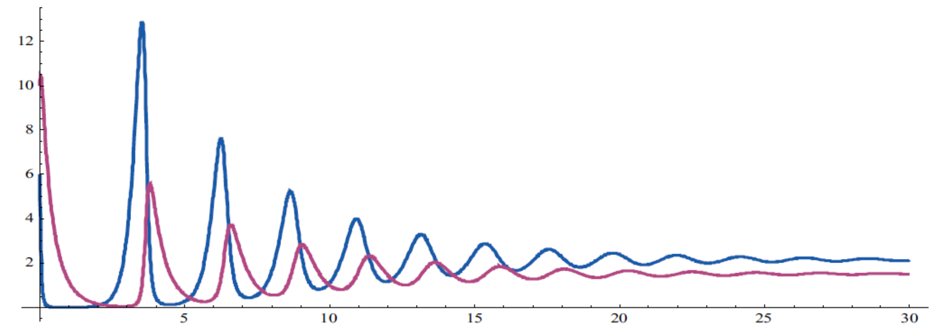


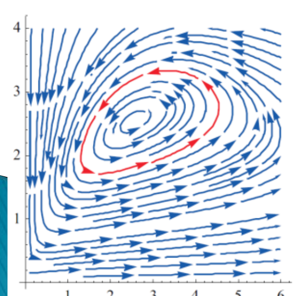










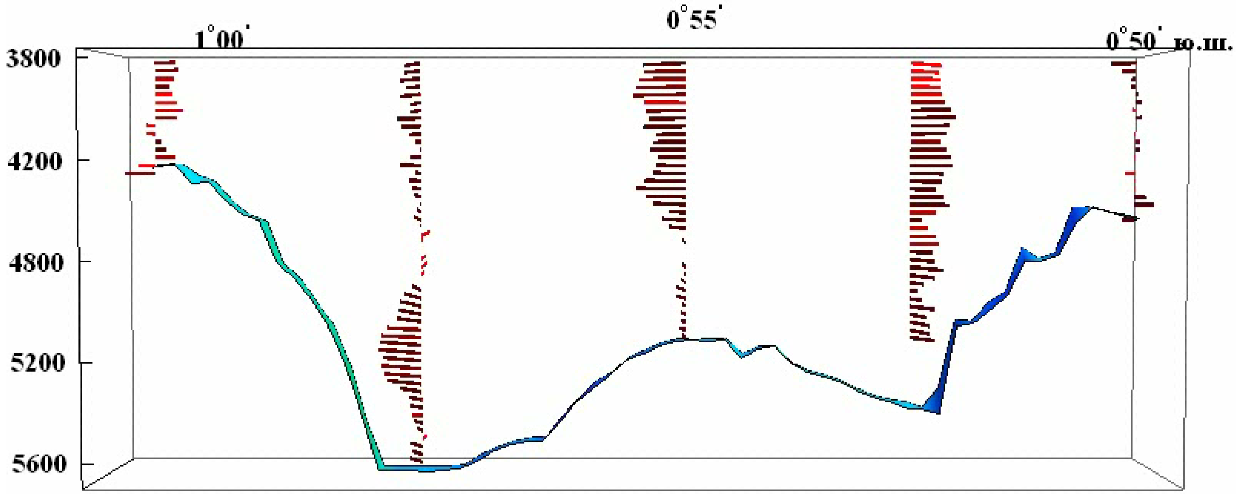


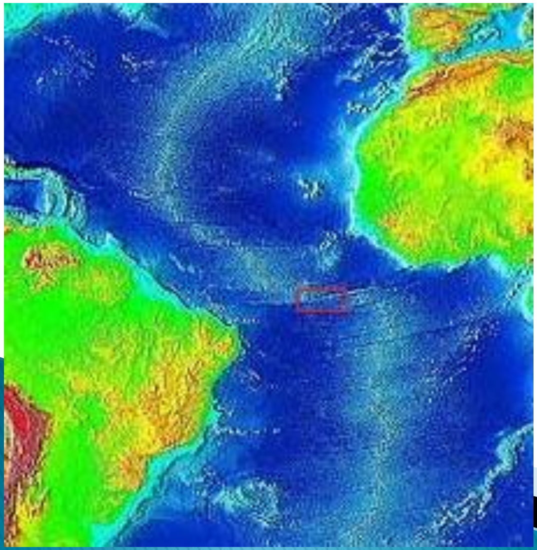
– Only for Mathematica!

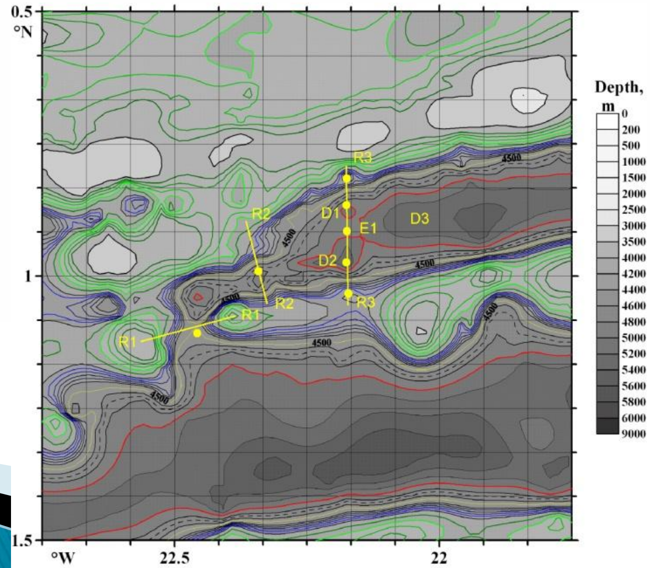
Benefits of Mathematica:

- Professor's view:
  - Easy learning, even for non-specialized students
  - Elegant Interface, natural representation of Formulae
  - Smart helps and Virtual Book
  - Lot of Examples
  - Additional - set of possibilities for advanced students!
- Researcher's view:
  - Capacity for solving scientific problems with Big Data (Meteorology, Oceanography, etc.)
  - Consume a lot of data formats (almost any binary formats and smart tuning for others)
  - Small tricks of Big Data calculations (multiparametric problems - Manipulate, 3D Vector Fields)
  - …..
  - …..

Near-bottom stratified Flows
(Atlantic Ocean, equator, Romanche Channel)







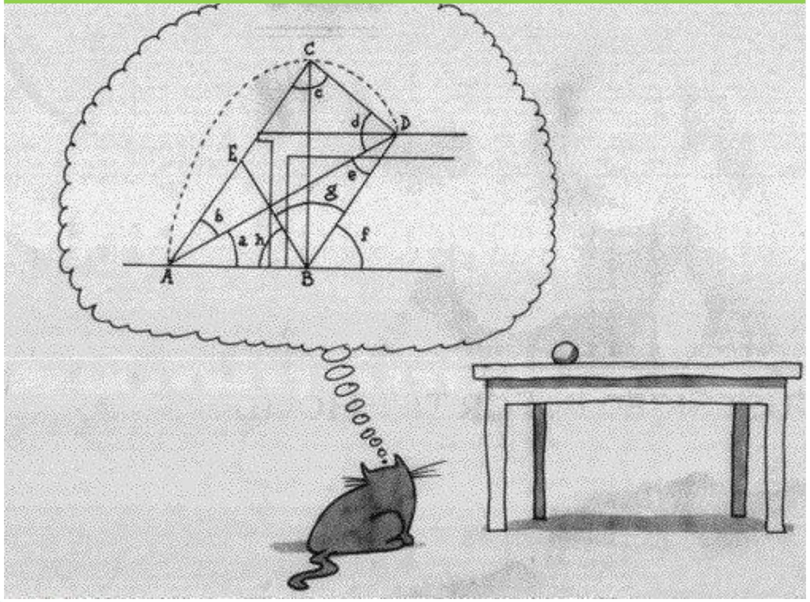
Summary

- **Math is everywhere! Study Math&Math.**

Wolfram *Mathematica*

Simplifies the reality

SO… Jumps not immediately!



Big Data and Computing Challenges

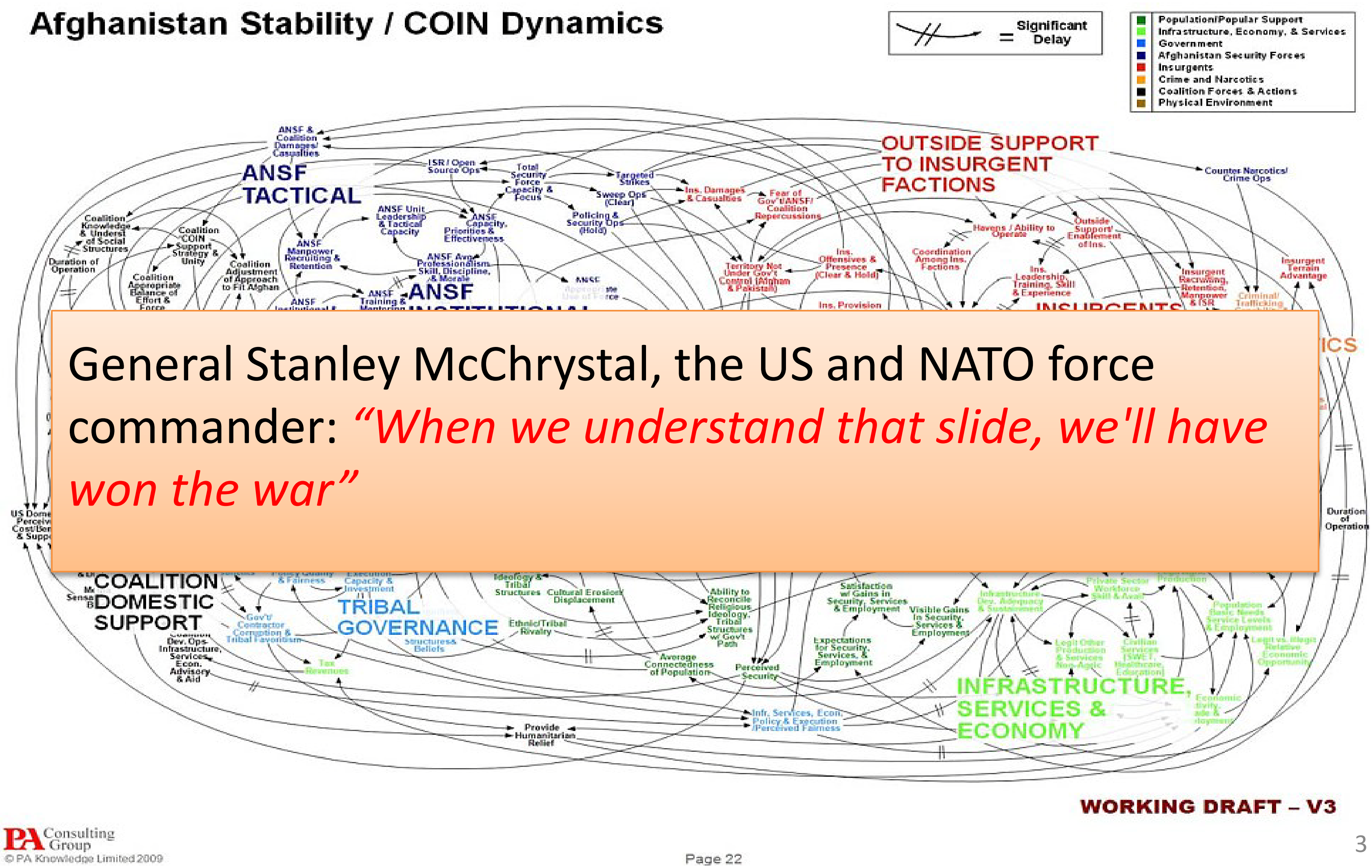
Shangping Ren

Illinois Institute of Technology
U.S.A.

What is Big Data?

- In 2012, Gartner updated its definition as follows: "Big data is high Volume, high Velocity, and/or high Variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."
- A new "Veracity" is added by some organizations to describe it.
- Additional dimensions: Variability and Complexity

Afghanistan Stability: Where are the Challenges?



General Stanley McChrystal, the US and NATO force commander: *"When we understand that slide, we'll have won the war"*

Where are the Challenges?

Big Data:

Digital footprints vs information

Computing Challenges:

Processing digital footprints vs

Information understanding and processing

Complex correlations; and evolving correlations

Dealing with the Challenges

Will traditional engineering principles work for big data computing?

Principle 1: separation of concerns?

Principle 2: hierarchical approach?

Thank You!

Questions?

Big data : success and challenges

Elena Troubitsyna
Åbo Akademi University, Finland

Big data perspective

- Pervasive monitoring and sensing: maturity of technology, low cost
  - Sensing in broad sense

- Web-services – user behaviour
- Internet of Things – environment, critical infrastructures, production environment

- Vision: Big Data for better life, environment, industy

Big data for mercury business

- Data worth much more than the actual application
- Quick learning of user feedback, gaining customer insight

- Challenge:
- build highly services that capitalise on the knowledge of customer behaviour
- Enable proliferation of services driven by customer-needs
- Create highly efficient development environments for data-driven services

Big data for cyber-physical world

- System diagnostics and providing preventive maintenance
- Efficient logistics and management of assembly lines
- Energy grids and energy distribution
- Environment: agriculture, natural disasters response

- Challenge:
- create an eco-system for data sharing, service provisioning while preserving safety and privacy

BIG DATA and BIG QUESTION:

How to reap the great potential without compromising privacy and safety?