

IBM BusinessConnect
Vernetzter, intelligenter und informierter denn je



Big Data



Big Data – Vom Hype zur Realität

Dipl.Ing. Wolfgang Nimführ



Fundamentale Änderungen finden statt

Stetige Transformation ist die neue Norm

Neue Geschäftsmodelle



Konvergenz von Technologien



Stetig wachsender Druck nach Veränderungen

Nichts tun ist die schlechteste Option



Mit weniger
mehr erreichen



Die Macht der
Konsumenten



Eine neue
vernetzte Welt

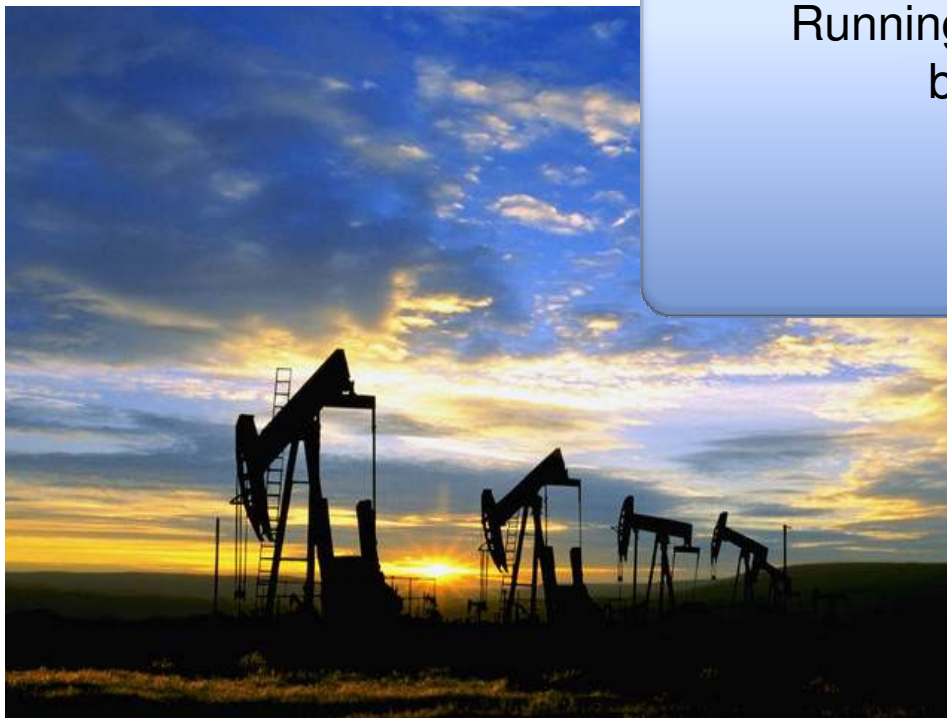


Big Data ist die nächste natürliche Ressource

Nutzbarmachung erfordert Aufbereitung

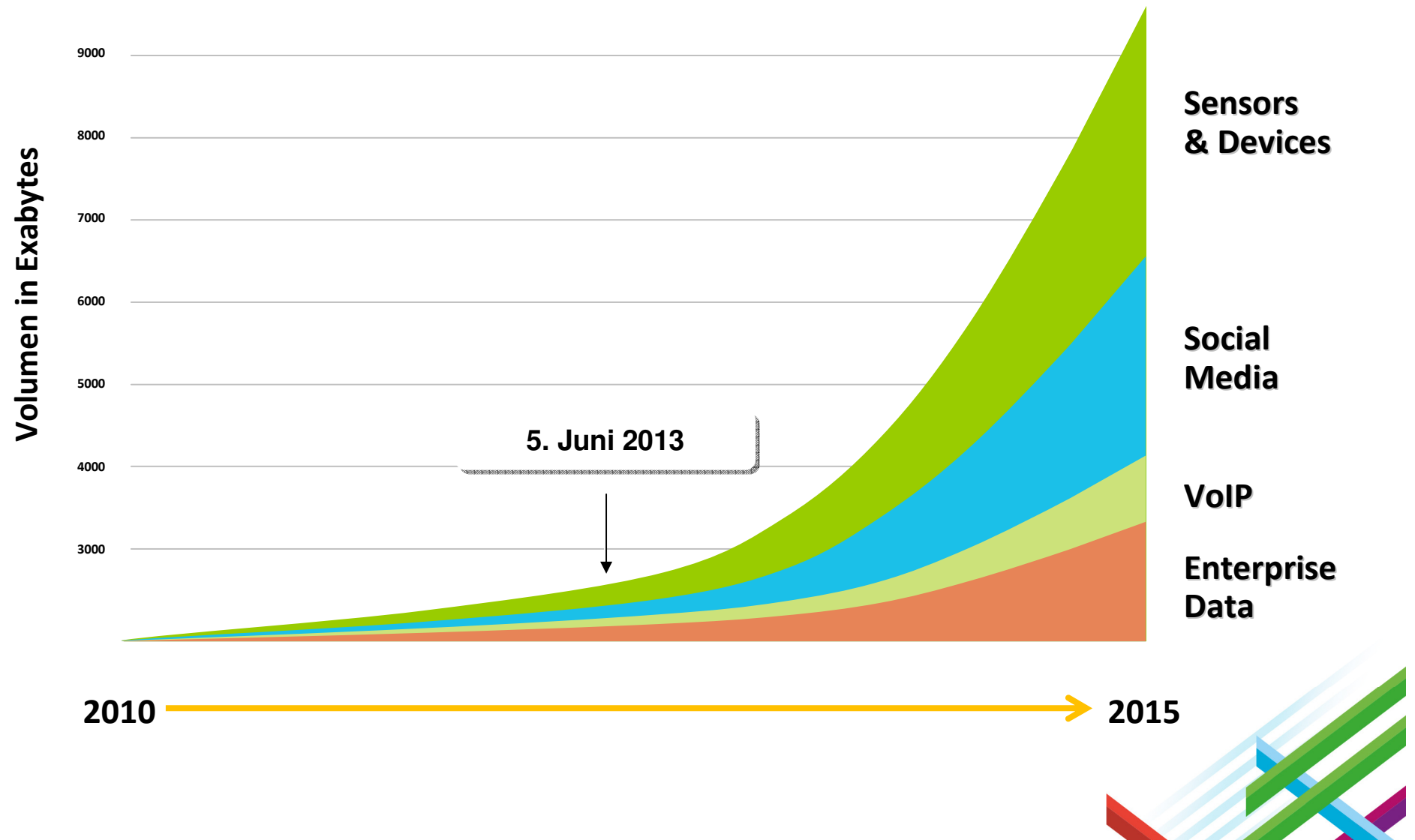
“We have for the first time an economy based on a key resource [Information] that is not only renewable, but self-generating. Running out of it is not a problem, but drowning in it is.”

– John Naisbitt



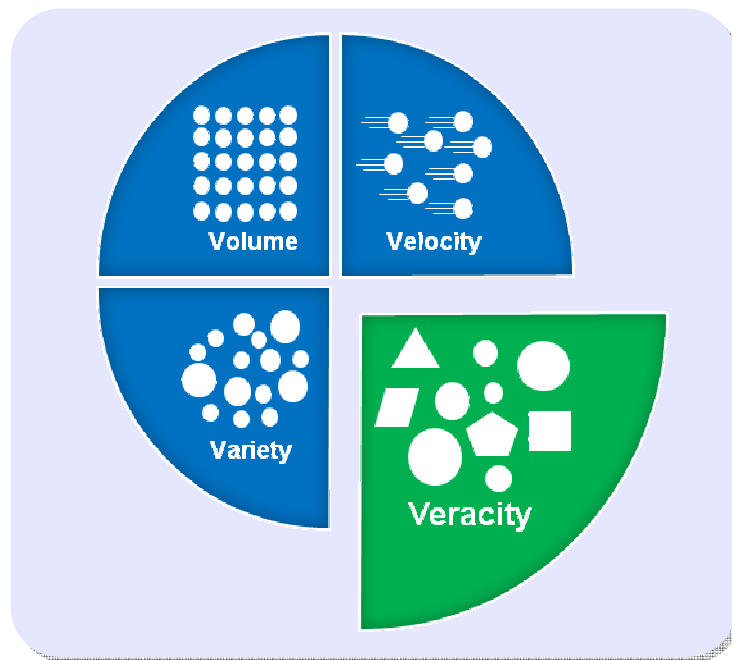
Der Aufbruch zu Big Data

Wir stehen erst am Anfang

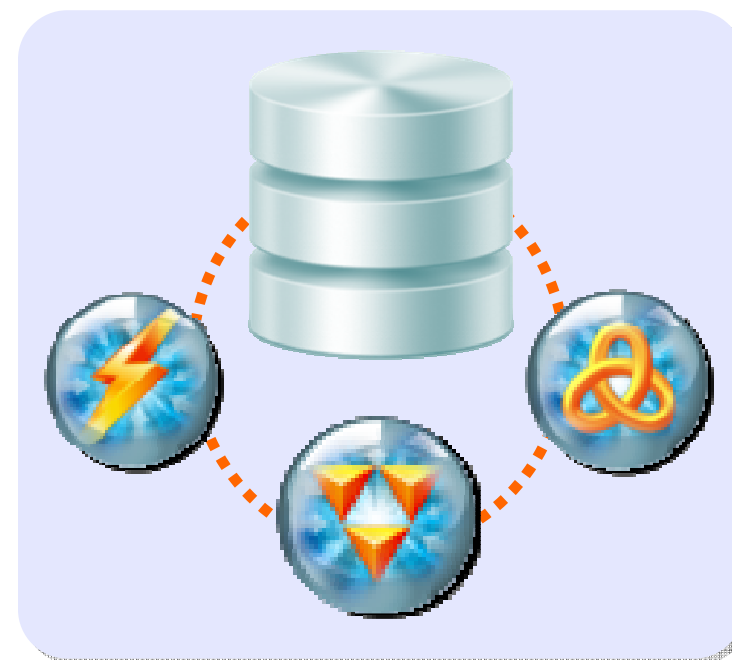


Was ist Big Data wirklich?

Warum ist Big Data gerade jetzt wichtig?



Die Macht der Daten



Die Kraft der Technologie



Signifikante Optimierungen für jedes Unternehmen

Auf der Schwelle zum kognitiven Zeitalter



Jede Branche hat Szenarien für Big Data

Konkrete Fallbeispiele als Referenzen



Banking



Insurance



Telco



Retail



Travel & Transport



Consumer Products



Automotive



Chemical & Petroleum



Aerospace & Defense



Energy & Utilities



Media & Entertainment



Electronics



Government



Healthcare



Life Sciences



Die 5 besten Big Data Use Cases

Mit konkreten Ausprägungen für jede Branche



Big Data Exploration



Complete View of the Customer



Security/Intelligence



IT Operations Analysis

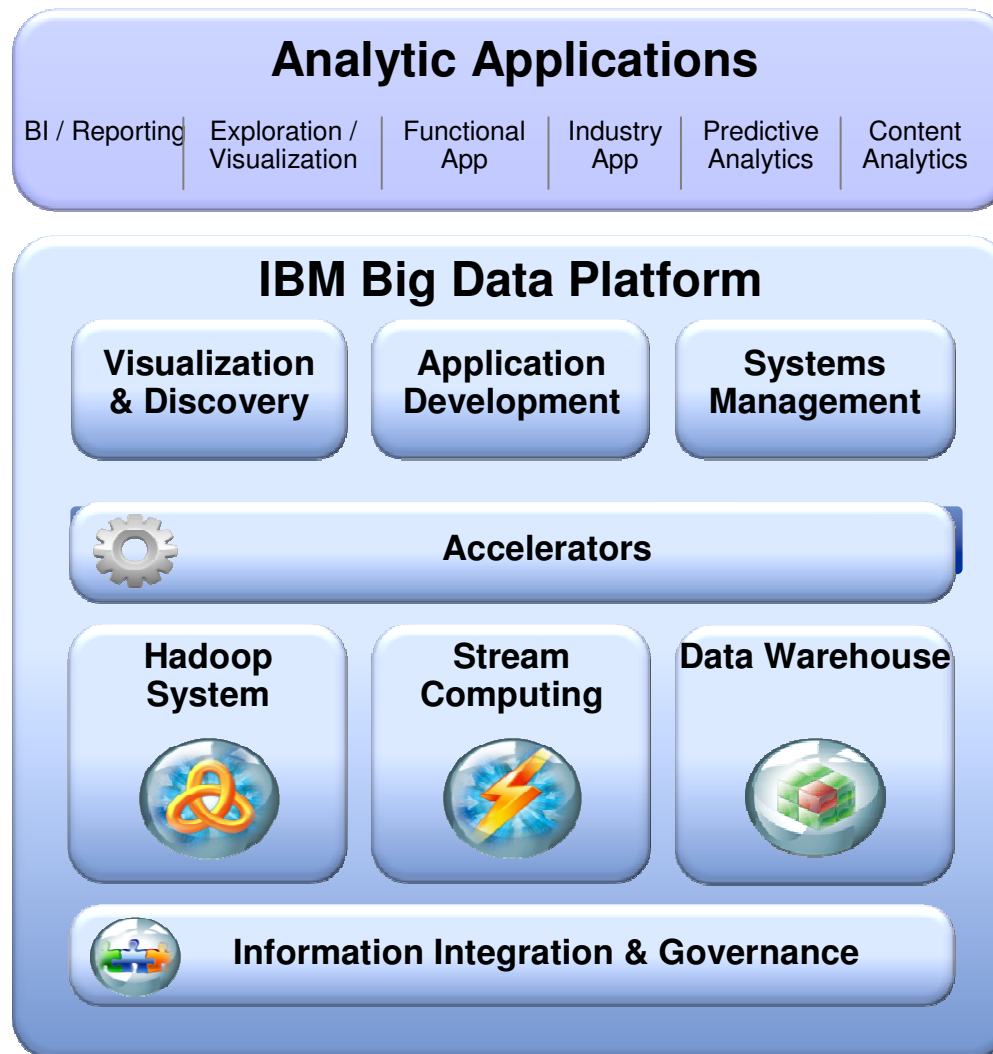


Data Warehouse Augmentation

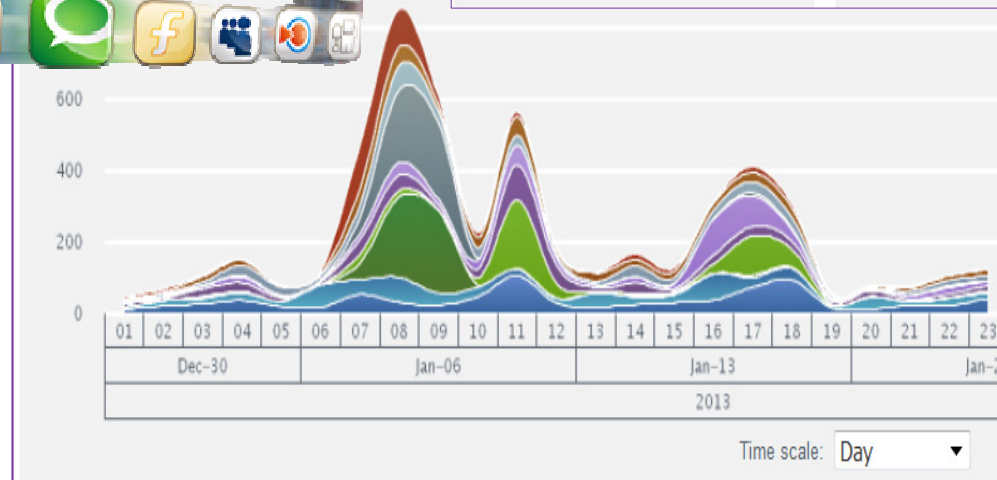


Ein ganzheitlicher Ansatz von Big Data und Analytics

Die solide Basis zum Erfolg



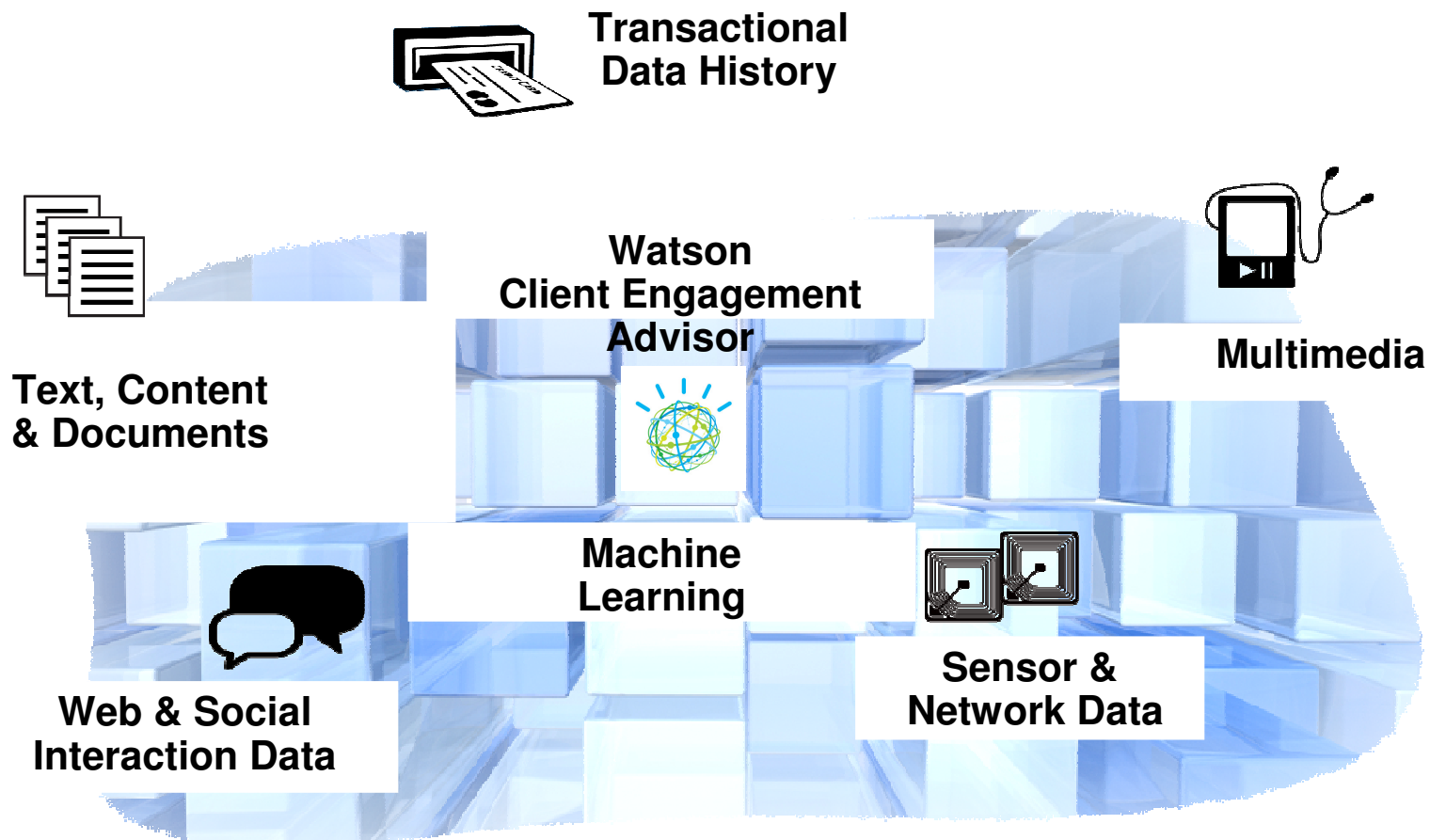
Big Data verlangt nach neuen Visualisierungsmethoden





Big Data erfordert neue Algorithmen & Analysetechniken

IBM Watson



Big Data Beispielapplikation zur Lead-Generierung

APPLICATION: [Lead Generation](#) [Reputational Risk](#) [Disease Tracking](#)

Lead Generation Real Time Dashboard

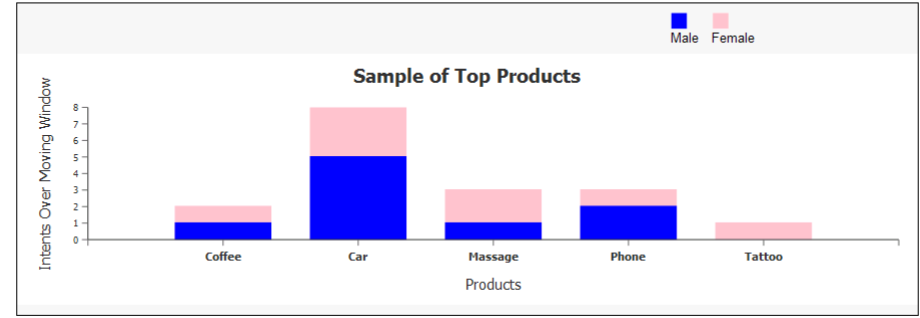
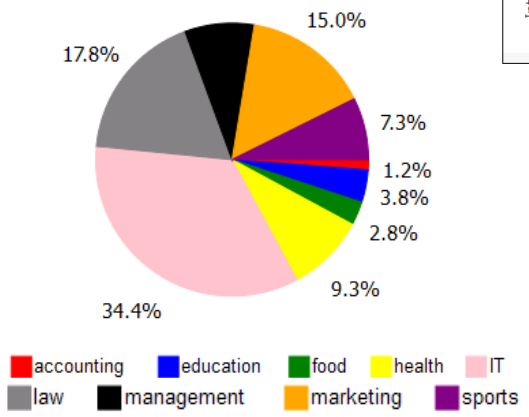
Actions: [Resume](#) [Microsegmentation Information](#)

Name	Sources	Intent	Gender	Location	Interests	Tweet
	Twitter, Tumblr	Laptop	Female			I need a new laptop.
	Twitter, Facebook	Phone			Foodies	I need a phone.
Abdul Yab	Twitter, Tumblr	Car	Male	Cairo		Need a car.
kesha rose	Twitter	Trousers	Female	Bristol		omg i cant find any black trousers. i need some for my trial at john lewis tomorrow.
vanessa natal	Twitter	Conditioner	Female			I need to find a good conditioner...my hair is growing so thin...:(
Chris Castle	Twitter	Juicer		Seattle	Sports	@bdtrimberger whoa! I've been thinking about getting a juicer. I might have to try yours out!
Matt Arlauckas	Twitter	Diner	Male	Rochester	Foodies	@shotbykim Holy Moley! I need a diner. STAT!
Jodi	Twitter	New Car	Female			Thinking about test driving cars this weekend. When's the best time of year to buy a new car? (best=cheapest)
chris	Twitter	Shoes				@TGODcaponi nah just for chillin. i need a new pair of blue shoes.
Courtney Morrison	Twitter	Ice Cream				I really want some ice cream. I shoulda got that instead of these cookies.

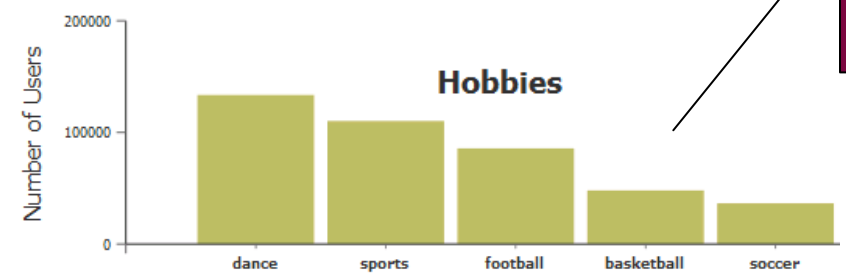
Real-time product intents enriched with consumer attributes

Micro-segmentation of product intents by occupation

Phone Intents by Occupation

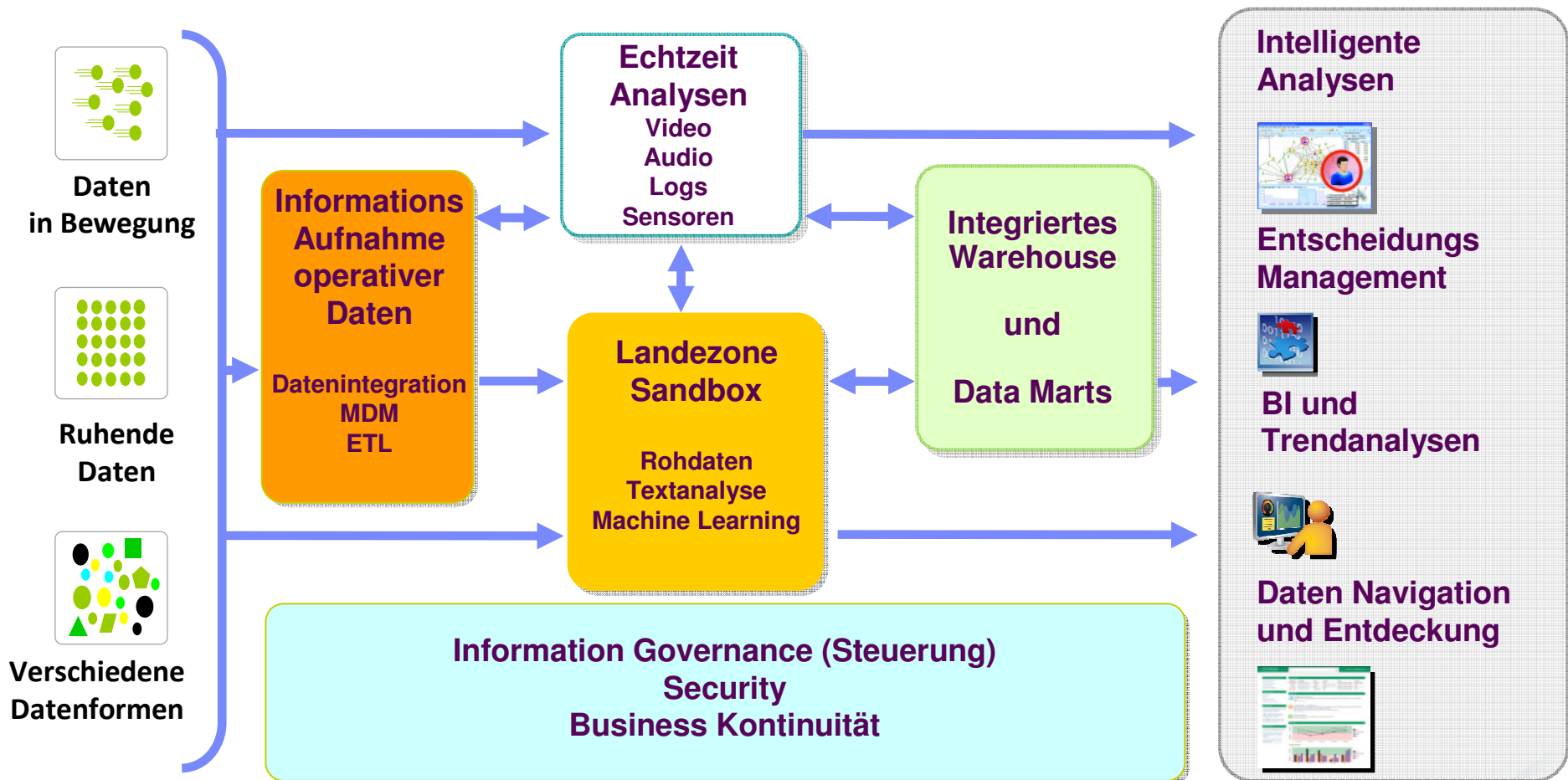


Real-time tracking by micro-segmentation



Micro-segmentation of consumers by hobbies

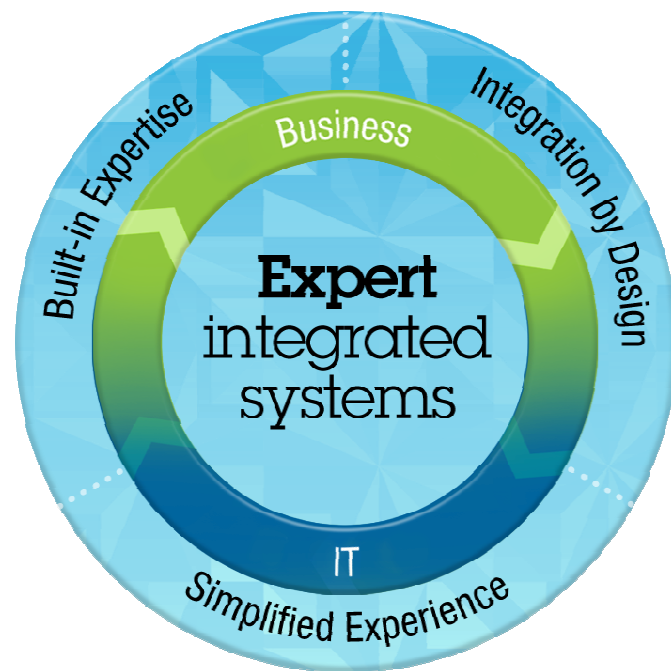
Big Data wird auf Basis von Referenzarchitekturen implementiert





IBM PureData Systems Appliances sind startbereit

Optimierte Systeme für Big Data



PureData

For Hadoop

Optimized system to accelerate analytics on big data and online archive with appliance simplicity



For Analytics

Optimized system delivering data services for analytics and reporting

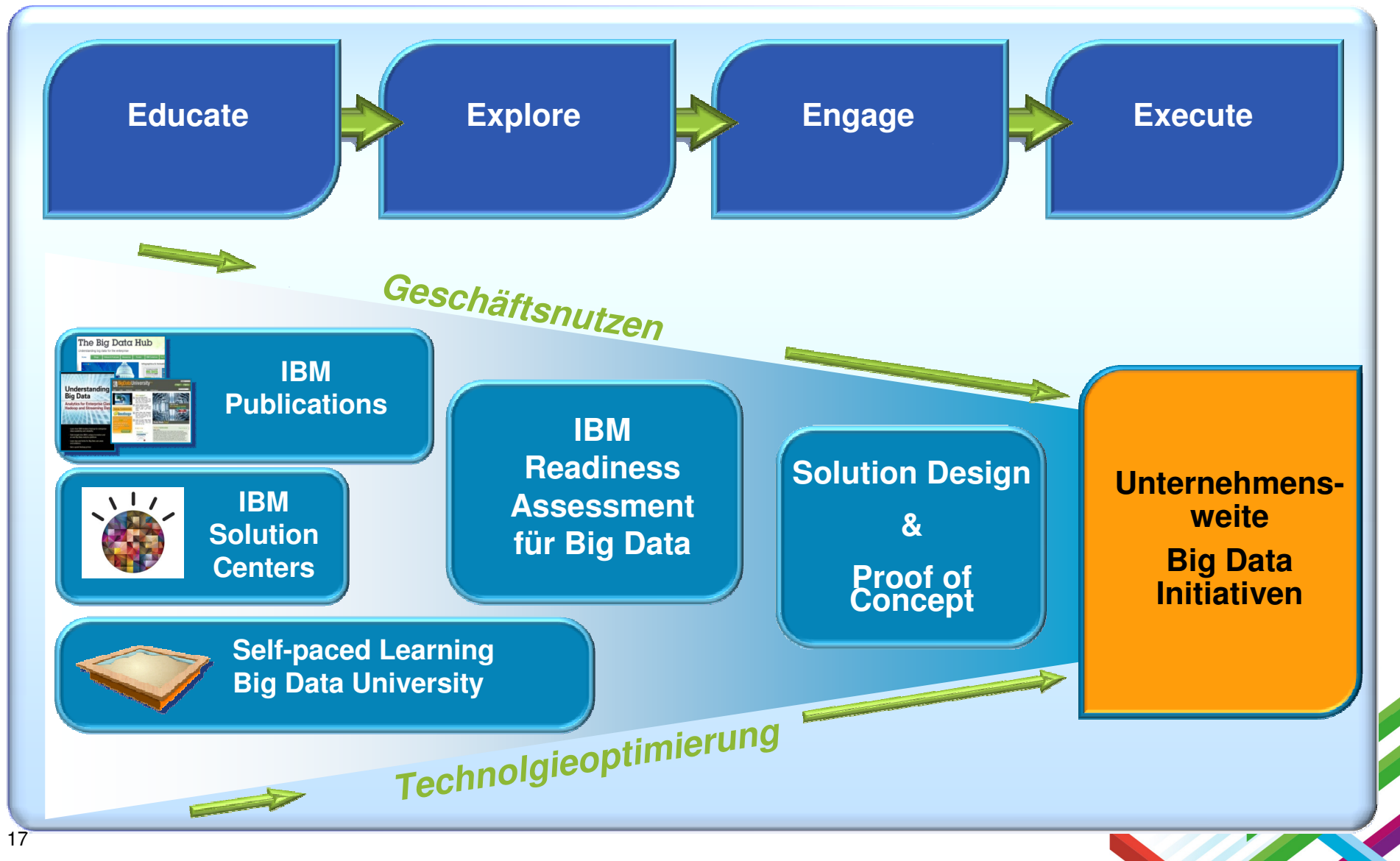
For Operational Analytics

Optimized system delivering data services for operational analytics



Einführung von Big Data

gemeinsam mit dem IBM Big Data Industry Solutions Team





Big Data University BETA

Learn from the industry's best

[go](#)
[sign up](#) [login](#)

HOME COURSES DOWNLOADS RESOURCES

[Share This](#)

 Like 2.5k
 Tweet 1,277

Join **77921** registered students and put your career on the right track. [Sign up now](#)

BigData University



Akmal Chaudhri on Hadoop & Big Data

Why register?

- Easy and Affordable**
Learning Hadoop and other Big Data technologies has never been more affordable! Many courses are FREE!
- Latest industry trends**
Acquire valuable skills and get updated about industry's latest trends right here. Today!
- Learn from the Experts!** Big Data University offers education about Hadoop and other technologies by the industry's best!
- Learn at your Own Pace!** Find everything right here when you need it and from wherever you are.

▶ **sign me up**





Hadoop for Dummies

NOW AVAILABLE

Sign up for a **FREE** softcopy.

Sign Up

Study Made Easy!

Student Testimonials

▶ [go to sign up](#)

Camelia

I found the tutorial very useful, with step-by-step explanations. The course material manages to cover a large spectrum of aspects, both architectural and operational, in dense short lessons. The labs provide practical end-to-end examples .

Roman

Hadoop on the IBM SmartCloud Enterprise



Learn to create your own Hadoop cluster on the IBM SmartCloud Enterprise with this **FREE** course.

Enroll now!



youtube.com/user/ibmbigdata

Bringing big data to the enterprise

IBM Big Data

Subscribe 2,976

Big Data at the Speed of Business
472 views 1 week ago
<http://ibm.co/BigDataSpeed> - Organizations are discovering that their continued relevance, and even survival, depends on harnessing big data to better understand their customers, reduce risk, and discover entirely new opportunities for growth in a changing world. On April 3rd, IBM announced exciting innovations that help organizations gain advantage from all data -- including data that was once considered too big or too complex to affordably manage and analyze.
Join us on April 30th to learn from IBM clients and big data experts how IBM ...

On April 30, 2013 IBM Will Announce Major Enhancements to the Big Data Platform

What's New

- Big Data Use Case #2 - Enhanced 360 degree vi...**
by ibmbigdata
529 views
- Big Data - IBM's Machine Data Accelerator**
by ibmbigdata
538 views
- Big Data and UCLA tackle brain trauma**
by ibmbigdata
946 views
- Big Data at the Speed of Business**
by ibmbigdata
206 views
- Big Data Use Case #1 - Exploration**
by ibmbigdata
1,015 views
- InfoS...**
by ibmbigdata
641 views

Big Data Client Success Stories

- Sprint - IBM Smarter Computing Client Suc...**
3:08
- NCSU - IBM Smarter Computing Client Suc...**
3:47
- IBM Big Data Solutions: Redefining email Marketing**
3:15
- IBM Big Data Analytics Help NYSE Improve For...**
2:06
- IBM Sentiment Analysis: USC gains Efficiency**
3:20
- IBM e...**

The Big Data Hub

Understanding big data for the enterprise

Home

Blogs

Videos & Podcasts

Resources

Events

IBM Solutions

For Developers



Complete Book—Now Available!

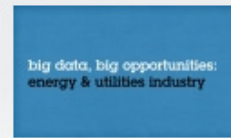
Boost your big data IQ, learn about the most common use cases, discover how to deploy projects faster and with less risk—all from the authors of "Understanding Big Data." **Download now** →

Infographics & Animations

→ View All



Tuning Into Big Data as the Buzz Gets Louder



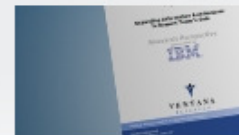
Big Data, Big Opportunities: Energy & Utilities



Big Data: The New Natural Resource

Featured Whitepapers & Reports

→ View All



Getting Big Value from Big Data: A Ventana Analyst Report



Big Data: New Insights Transform Industries



Analytics: The real-world use of big data



Competitive green energy through HPC and Big Data

Anders Rhod Gregersen, Senior Specialist
Vestas Wind Systems A/S



Who is Vestas?

- Market leader in Wind turbines
- Wind only company
- Install base 50+GW / 50.000+ wind turbines
- 18.000 employees
- World wide
- Wind on par with oil and gas

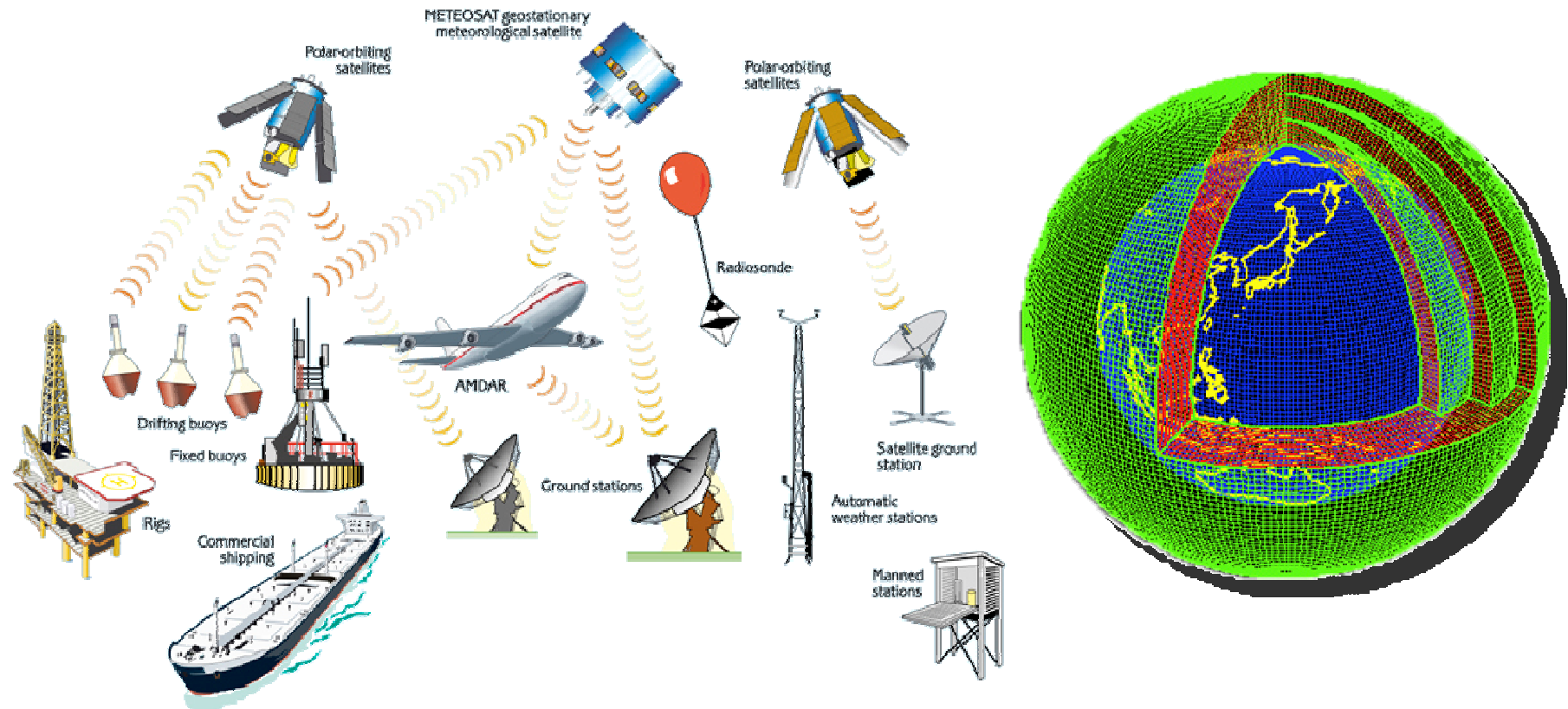


What matters to our customers?

- Cost of Energy
- Derived from a complex interplay
 - Wind resource
 - Wind turbulence
 - Service costs
 - Site complexity....
- Point measurement
 - MET- MAST
 - Months of measuring time



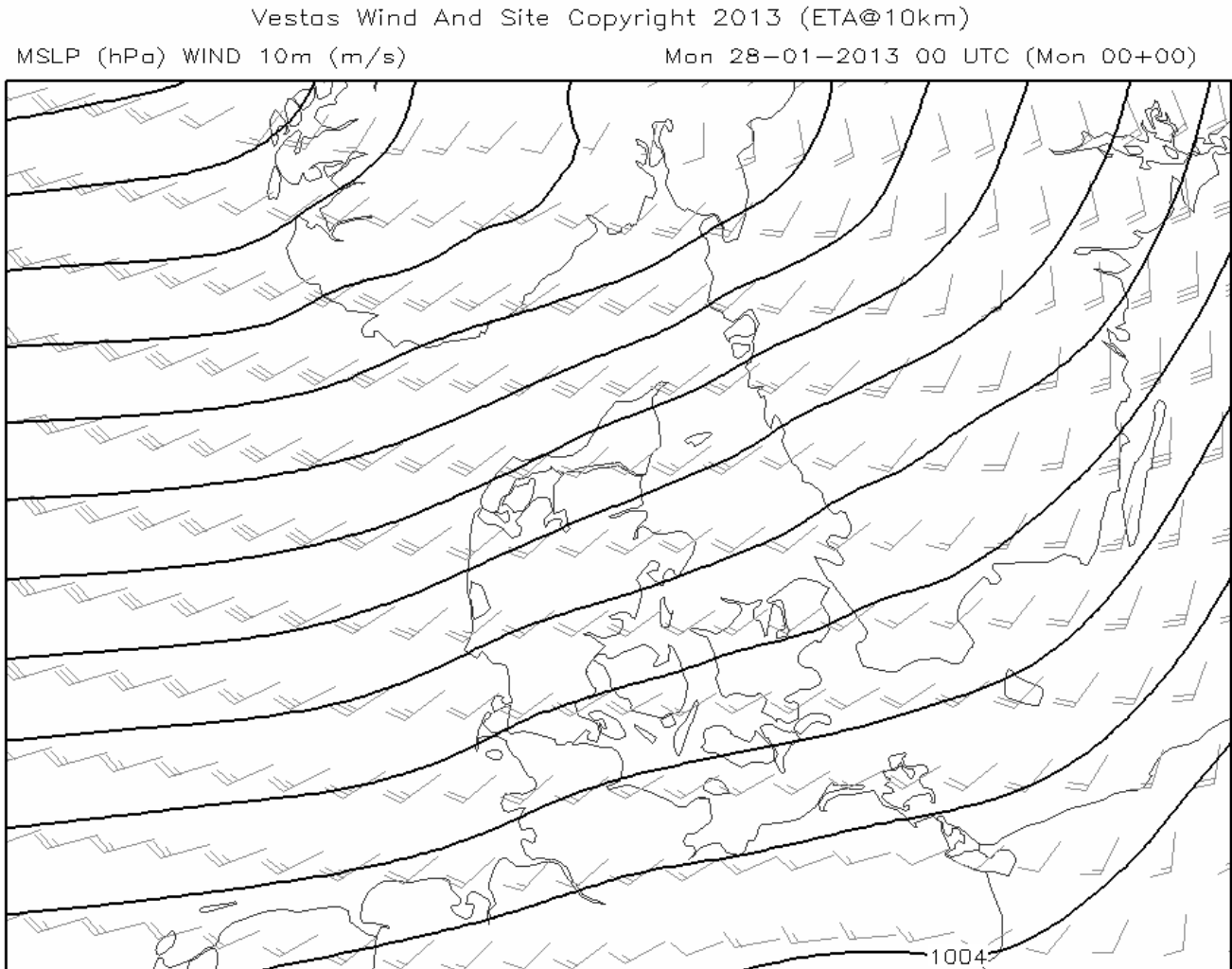
Global Weather Model



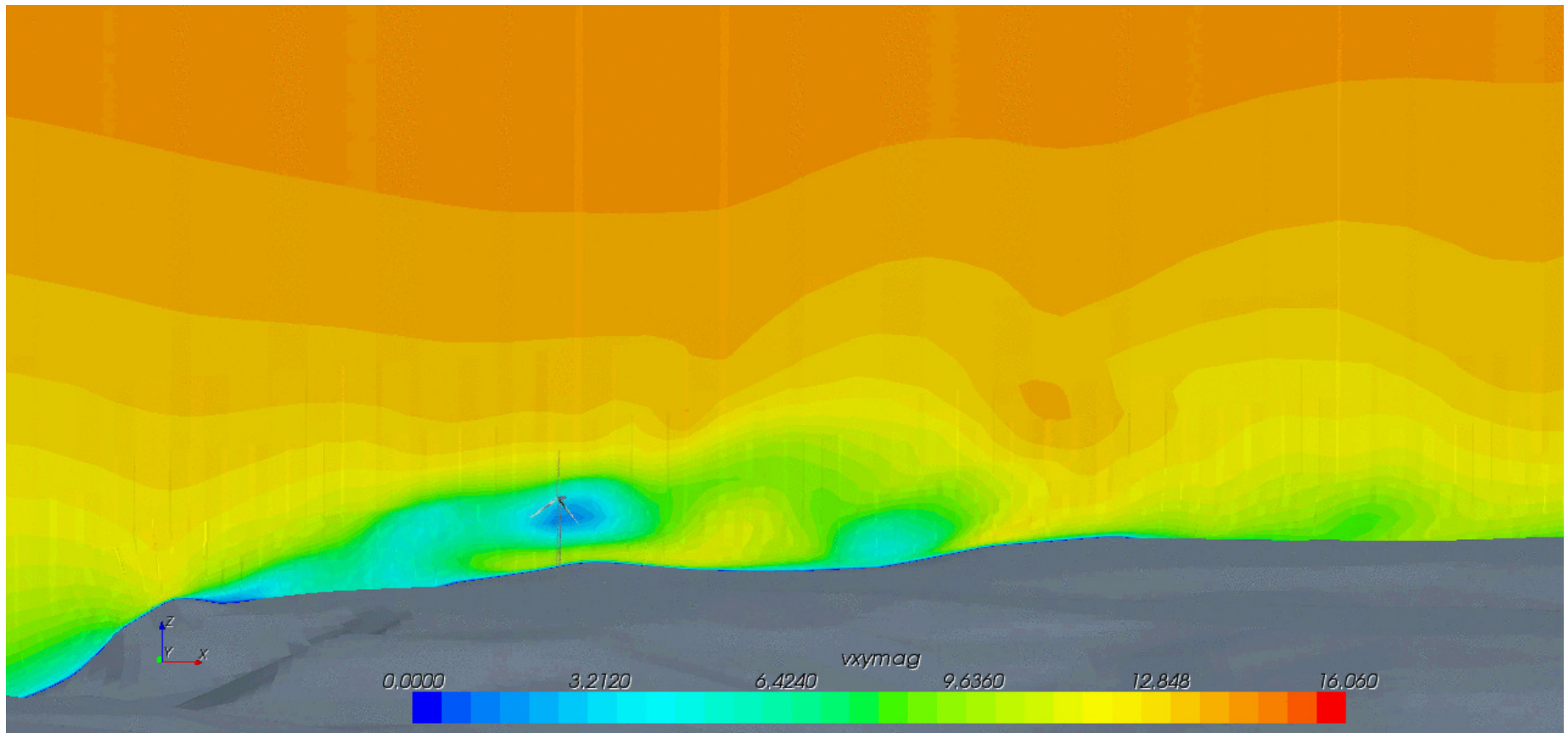
Input from 35.000 met-stations every 6 hours from National Weather Centers in UK, US or Japan



Example of numerical weather forecasting (re-analysis)



Example of a fluid flow simulation



The Vestas approach

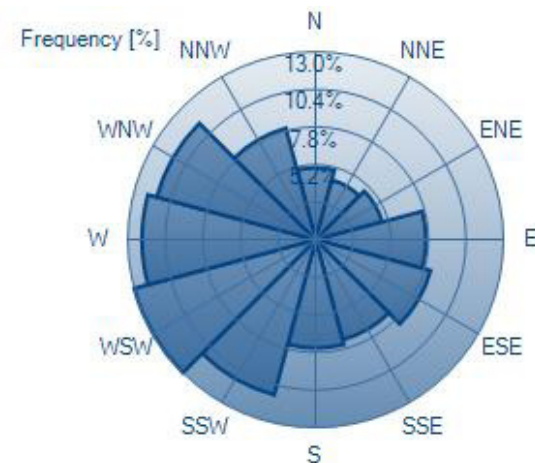
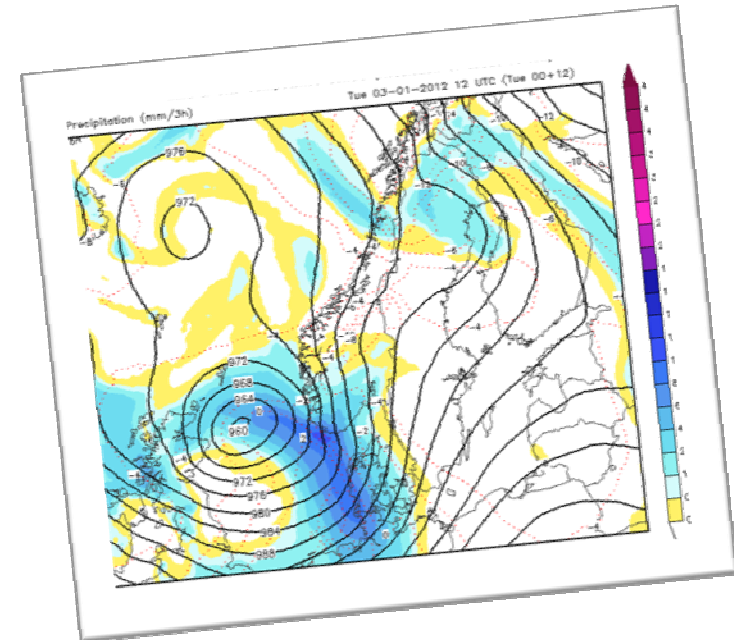
Globally

- A database with atmospheric data (petabytes)
- Down to a square km scale
- With hourly measurement since 2000
- With approx. 200 parameters

On a per site-by-site basis:

- Flow understanding of the site
- Pin-point turbulence
- Understand in-flow angles
- Understand wind shear

**With this we can minimize the cost of energy
for our customers while increasing business case certainty**



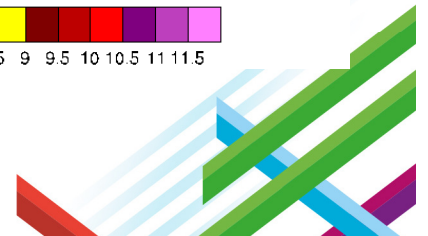
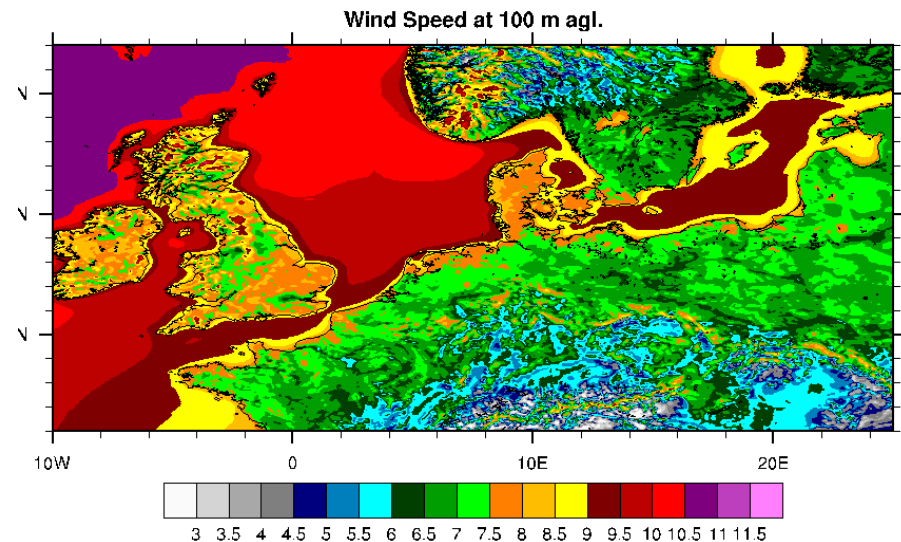
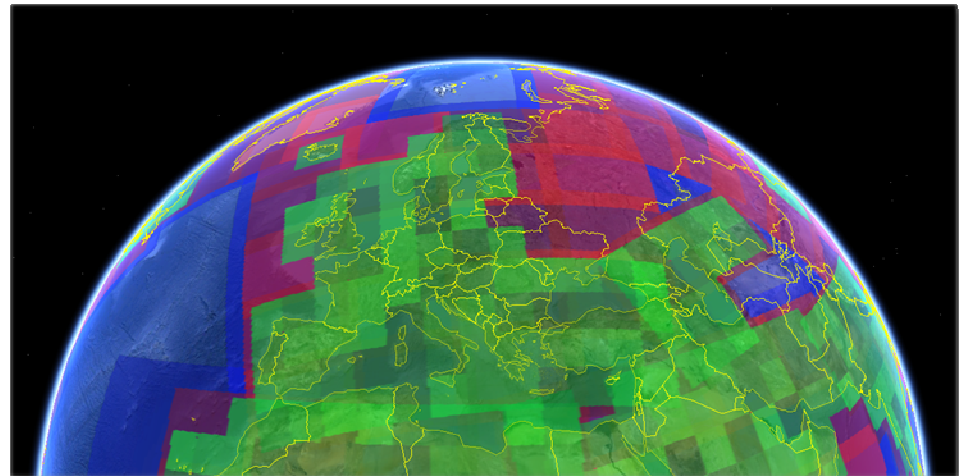
Why IBM InfoSphere BigInsights?

Old system

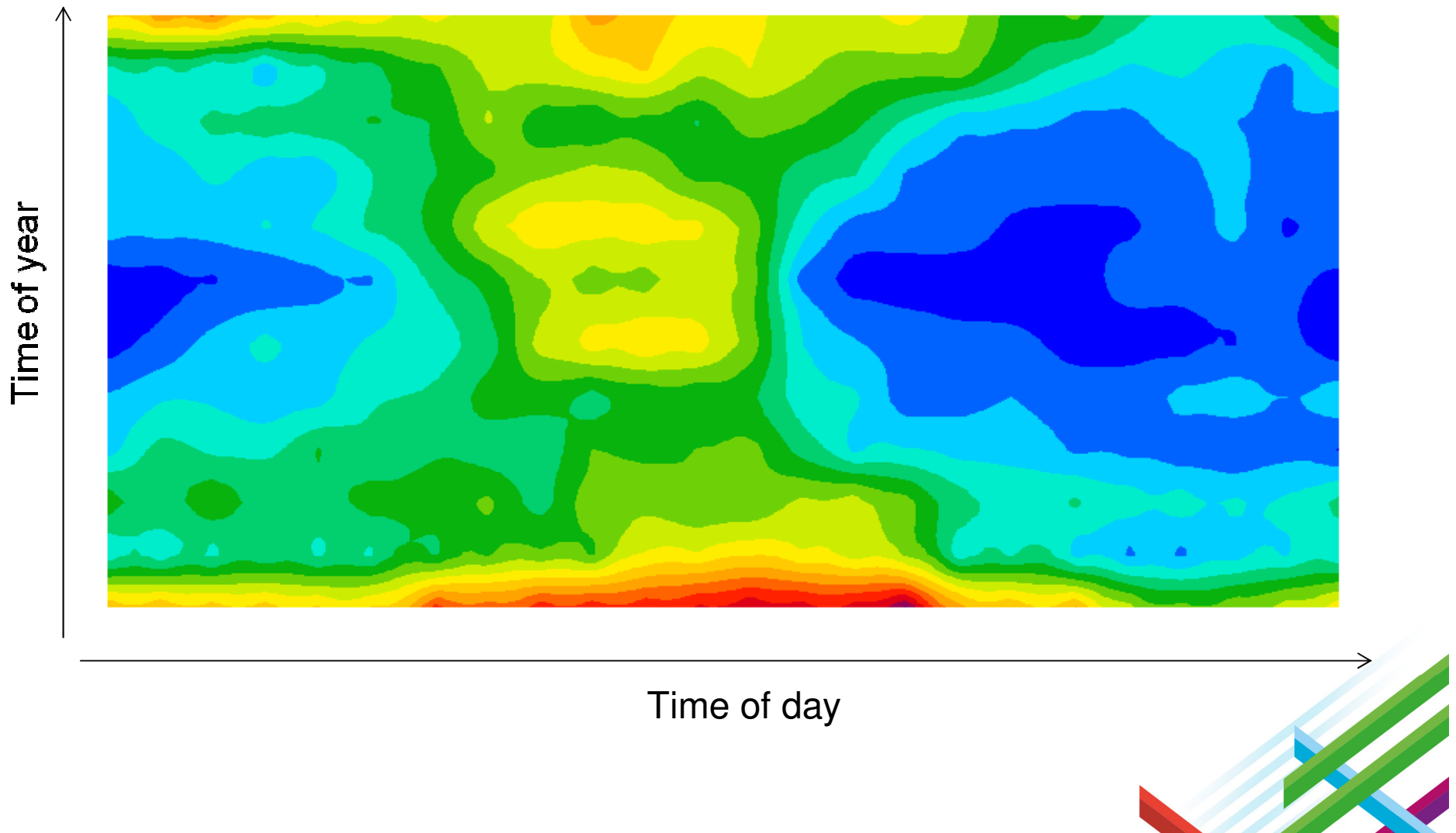
- Every request was programmed
- Response time in weeks
- Limits to request complexity

IBM BigInsights system

- SQL-like query language
- Point queries in low minutes
- Supports queries that scale beyond PB
- Product lifecycle handled by IBM
- Scale-out solution, limited only by I/O



Quantifying the Climates impact on the business case



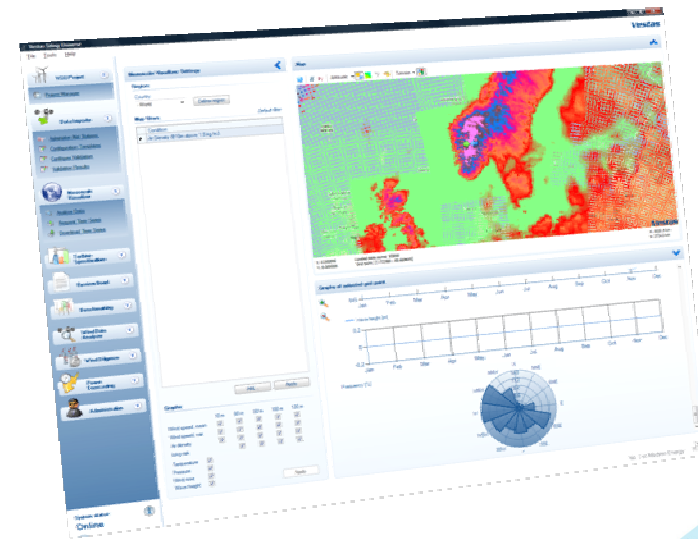
How do we interact with big data?

- High Performance Computing (HPC) grew out of academia
- What to expect from users?
- Users range from PhDsto sales force
- One common tool for common useenables traceability
- Point and click HPC for regular users
- Terminal based interaction for the power user
- SQL-like query language for power users

```

top - 09:21:20 up 28 days, 12:55, 1 user, load average: 12.00, 12.00, 11.55
Tasks: 253 total, 13 running, 240 sleeping, 0 stopped, 0 zombie
Cpu0 : 98.7%us, 1.0%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.3%hi, 0.0%st, 0.0%wt
Cpu1 : 99.0%us, 1.0%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu2 : 99.3%us, 0.7%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu3 : 99.0%us, 1.0%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu4 : 98.7%us, 1.3%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu5 : 98.7%us, 1.3%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu6 : 98.7%us, 1.3%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu7 : 98.7%us, 1.3%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu8 : 98.7%us, 1.3%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu9 : 98.7%us, 1.3%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu10 : 98.7%us, 1.3%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Cpu11 : 99.0%us, 1.0%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st, 0.0%wt
Mem: 24659948k total, 9333028k used, 15326920k free, 161628k buffers
Swap: 33551744k total, 0k used, 33551744k free, 730264k cached

  PID USER      PR  NI  TSTT  RES   SHR  S  %CPU  %MEM    TIME+  COMMAND
 27872 mazag    25   0 1190m 624m  27m  R 100.1  2.6  51:23.16 wrf.exe
 27874 mazag    25   0 1226m 618m  28m  R 100.1  2.6  51:24.65 wrf.exe
 27875 mazag    25   0 1193m 620m  28m  R 100.1  2.6  51:23.61 wrf.exe
 27876 mazag    25   0 1158m 602m  27m  R 100.1  2.5  51:23.20 wrf.exe
 27881 mazag    25   0 1157m 614m  28m  R 100.1  2.6  51:23.97 wrf.exe
 27871 mazag    25   0 1059m 563m  26m  R 99.8  2.3  51:20.97 wrf.exe
 27872 mazag    25   0 1113m 605m  28m  R 99.8  2.5  51:23.60 wrf.exe
 27876 mazag    25   0 1080m 593m  25m  R 99.8  2.3  51:23.60 wrf.exe
 27877 mazag    25   0 1156m 566m  27m  R 99.8  2.4  51:20.38 wrf.exe
 27878 mazag    25   0 1160m 622m  27m  R 99.8  2.6  51:23.15 wrf.exe
 27880 mazag    25   0 1152m 621m  27m  R 99.8  2.6  51:23.16 wrf.exe
 27895 mazag    25   0 1122m 562m  27m  R 99.8  2.3  51:23.56 wrf.exe
 614 agrep    15   0 17072 1316  896  R  0.3  0.0   0:00.13 top
  
```



Building up HPC capability

- **Q1/2003**
First commercial license for
Computational Fluid Dynamics (CFD)
- **Q4/2006**
First cluster (40 cores)
- **Q3/2007**
CFD model validated
- **Q4/2008**
Second cluster (15 TeraFLOPS, TiBs)
- **Q2/2011**
Third cluster (3rd largest commercial HPC, PiBs)



Results

- Time to solution down from months to minutes
- Part of a solution which completely reframes wind turbine siting
- Flexible solution: SQL-like interface
- Fast implementation: from nothing to science fiction in 6 months
- Queries that span datasets of petabytes
- Queries which enables new discoveries

Competitive advantages:

- More productive sales force
- Refinement of existing products (lower cost of kWh produced)
- Increased business case certainty (customer understands ROI early)
- Democratic, a tool used by sales force not only specialists
- Data-driven product refinement



Suchen Sie mit Ihren Worten

z.B. Über Ostern nach Mallorca

Klassische Suche

- Pauschalreisen Last Minute Hotel
 Flug Ferienhaus Städtereisen

Abflughafen: Top-Ziel auswählen:

oder Reiseziel eingeben:

von bis

Reisedauer Erwachsene



Kinder (Alter bei Reiseantritt)

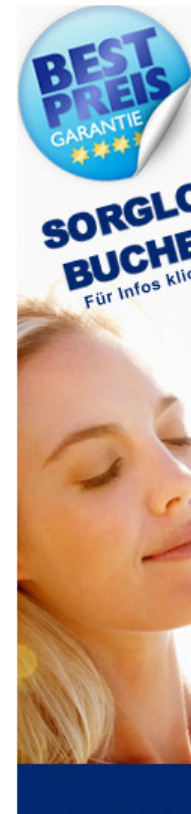


Traumurlaub auf Mallorca
Entdecken Sie das Mittelmeer
5 Nächte / 4* / inkl. Flug **ab 237 €**

1 2 3

Diese Reisen wurden gerade gebucht!

-  **Wow Topkapi Palace**
★★★★★
Antalya & Belek **ab 510 €**
-  **Treff Münster City Centre**
★★★★
Münsterland **ab 83 €**

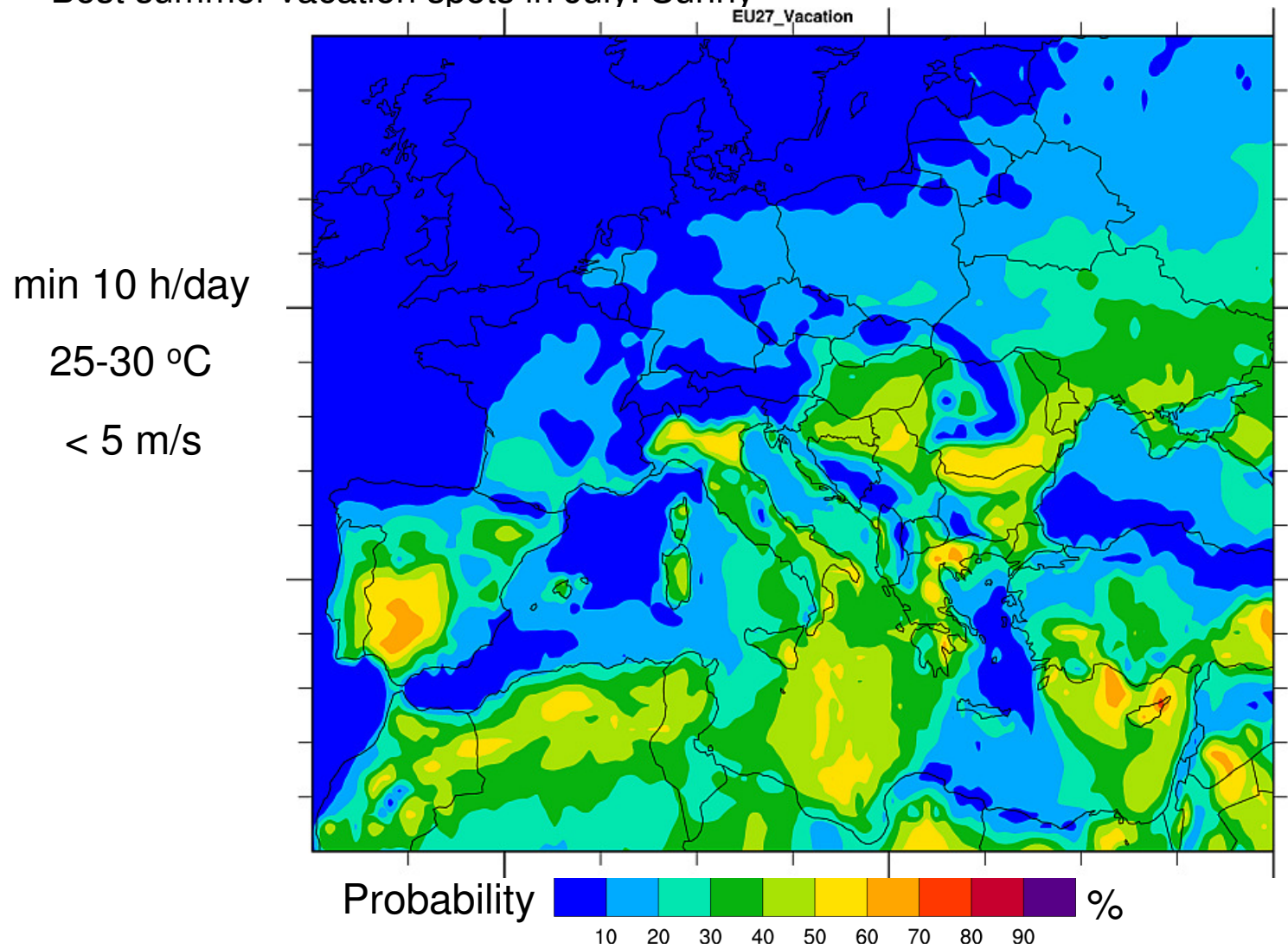



BEST PREIS GARANTIE
SORGLICHS BUCHEN
Für Infos klicken



Unparalleled fast data access

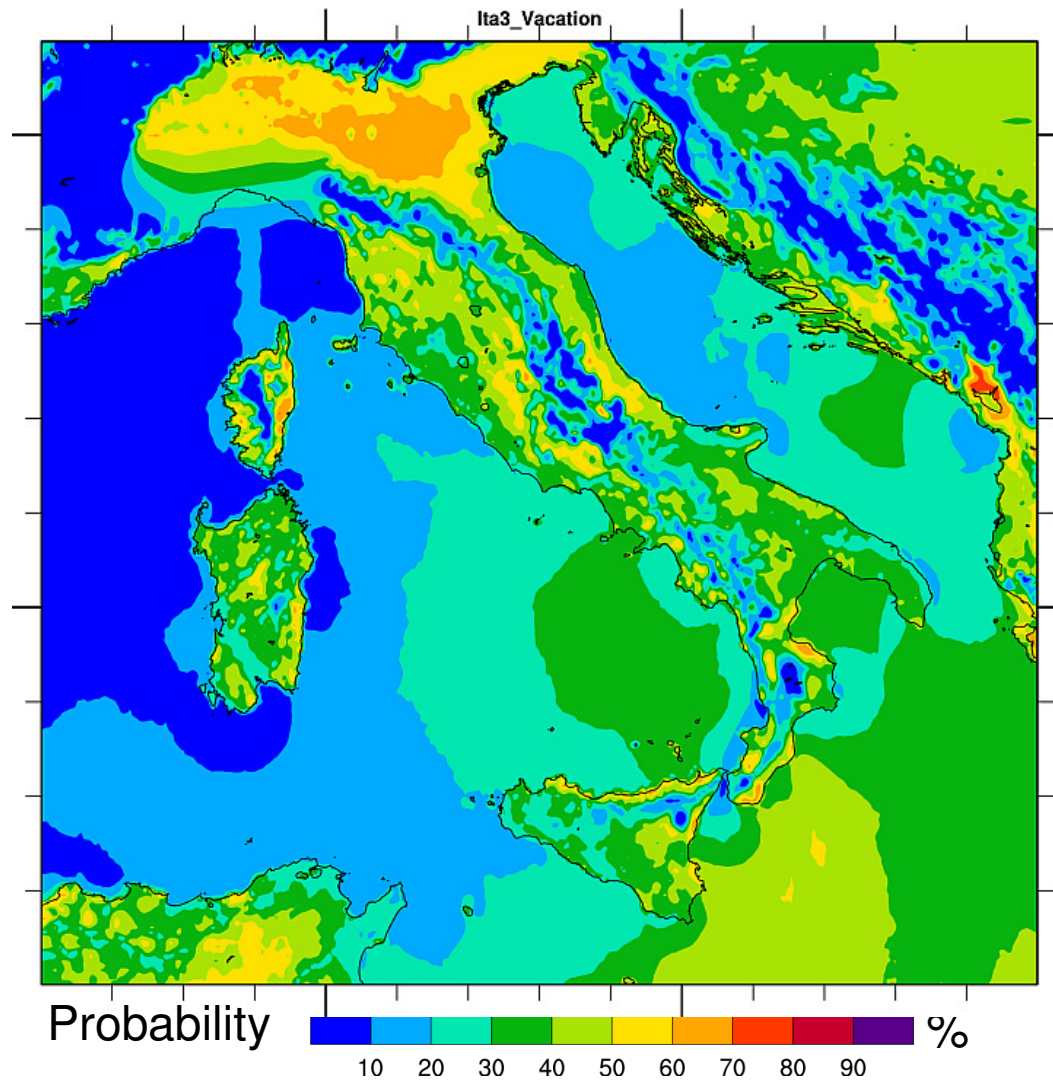
Best summer vacation spots in July: Sunny + Warm + Calm



Unparalleled fast data access

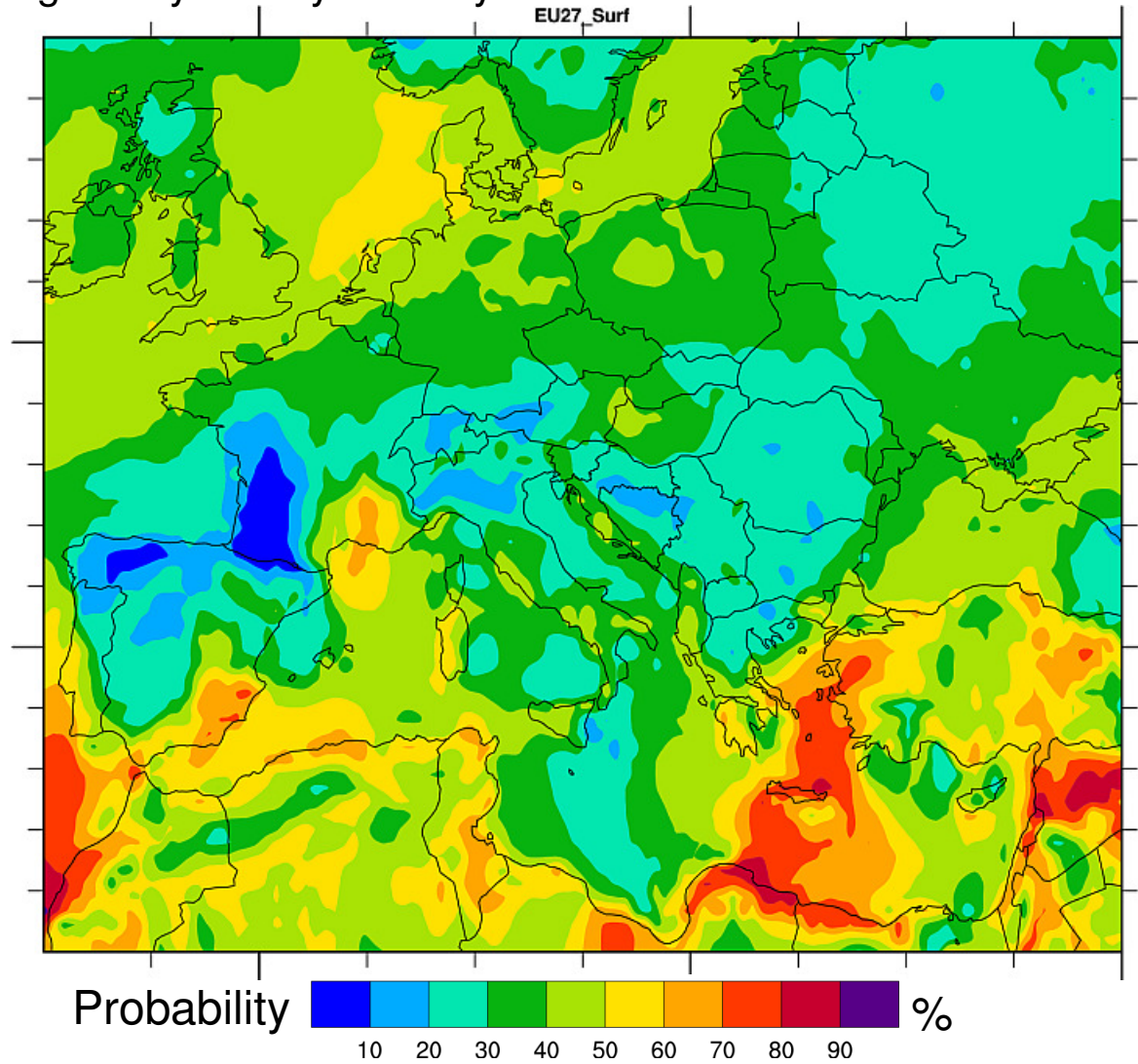
Best summer vacation spots in central Mediterranean

min 10 h/day
25-30 °C
< 5 m/s



Unparalleled fast data access

Best wind surfing in July: Sunny + Windy



Thank you!



Architecture



Architecture overview

- **Compute:**

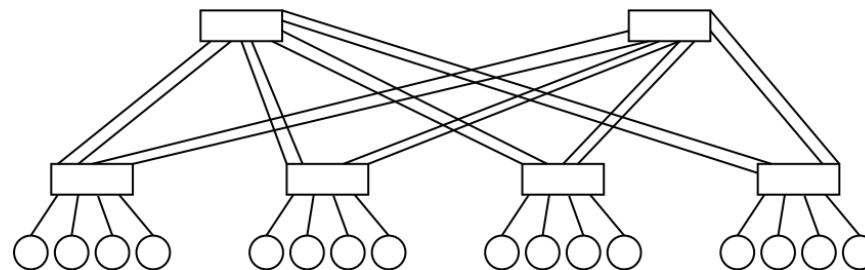
- 1.306 compute nodes
- 2.612 Intel X5670 CPUs
- 15.672 X5670 cores
- >34TiB RAM
- >161TFLOPS
- 10.752 M2070Q cores

- **Interconnects:**

- 4xQDR Infiniband (40Gb)
- 2:1 over-subscribed
- Ethernet

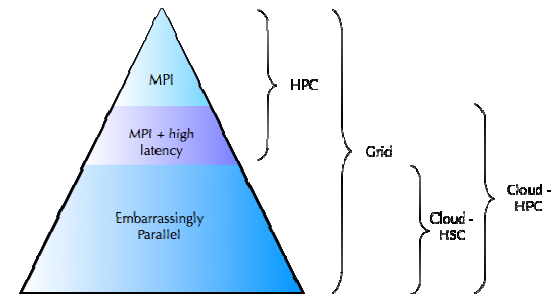
- **Storage:**

- 1.680 spindles
- 14 I/O servers
- >20GB/s



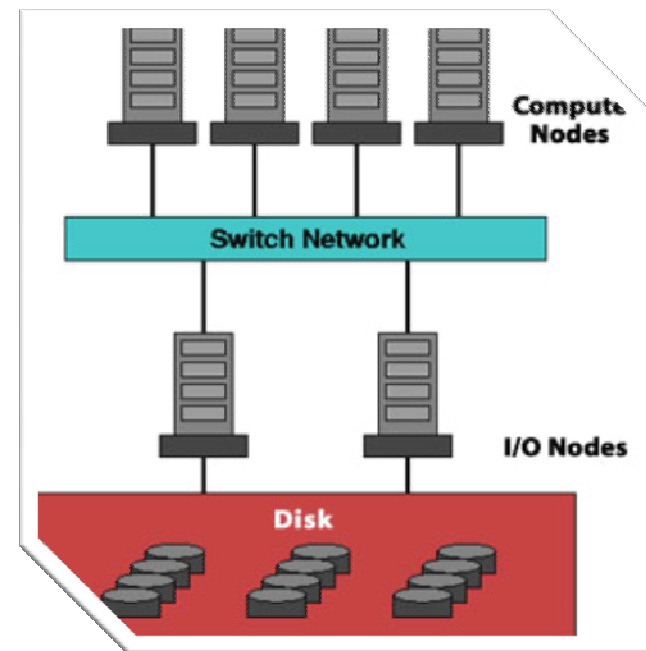
Primary memory

- Modeling fluid flow means
- Solving Navier-Stokes equations which means
- Memory bound code
- On die memory controller
- One bank, three ranks/CPU
- Remote memory via Message passing interface (MPI)
- MPI via Infiniband
- Low latency / high bandwidth



Secondary memory

- GPFS – general parallel filesystem
- Freedom in both bandwidth and capacity
- Singular name space
- I/O moves via Infiniband
- Exposed to the OS as POSIX
- Backup vs snapshots



Anders Rhod Gregersen

- **At Vestas:**

- Design, implementation, running the 3rd largest commercial HPC

- Impl. BigData database

- Won the Computerwoche “Das beste Big Data-Projekt 2012”

- Won the most innovative award at IOD 2011 for Vestas

- **Before Vestas:**

- Heading the data & analytics effort for the Nordic countries for the ALICE project (Big Bang simulation at CERN, Geneva)

- Independent HPC consultant

- Heading HPC efforts at Aalborg University, Denmark

- **Background:**

- M.Sc CS/Math + MBA

- **Contact:**

- E-mail: agreg@vestas.com

- Mobile: +45 2250 7205

