

ENDPOINT DETERMINATION USING VISUAL ANALYTICS

MARK LAMBRECHT, PhD

SAS HEALTH AND LIFE SCIENCES GLOBAL PRACTICE
PHUSE SINGLE DAY EVENT, FRANKFURT, APRIL 15TH 2015



HISTORY JOHN W. TUKEY

"The greatest value of a picture is when it forces us to notice what we never expected to see."

John Tukey



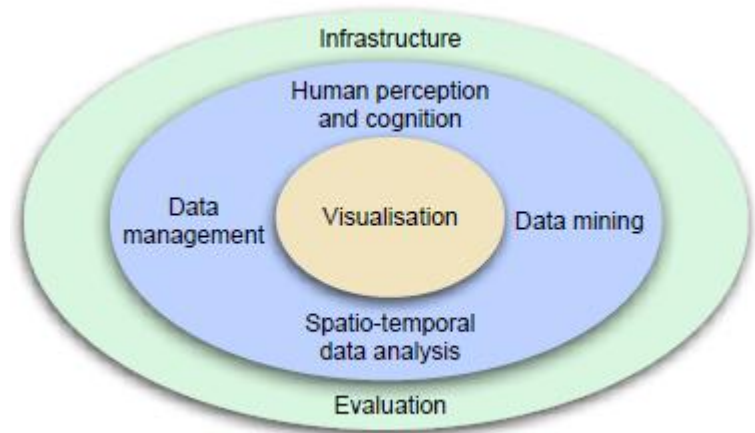
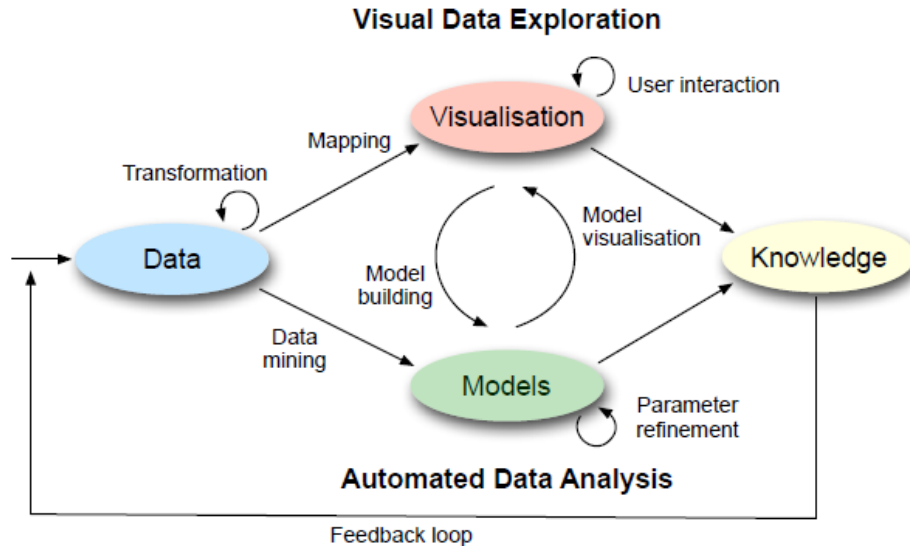
- Automatic analysis techniques such as statistics and data mining developed independently from visualization and interaction techniques
- Important shift from confirmatory data analysis (using exploratory data analysis charts and other visual representations to just present results) to exploratory data analysis (interacting with the data/results)



J.W. Tukey. Exploratory Data Analysis. Addison-Wesley, Reading MA, 1977.

DEFINITIONS VISUAL ANALYTICS

“The science of analytical reasoning facilitated by interactive human-machine interfaces”



“Visual analytics combines automated analysis with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets.”

BUSINESS VISUALIZATION

THE DIFFERENCE BETWEEN RAPID INSIGHT AND FAST INFORMATION

DATA VISUALIZATION

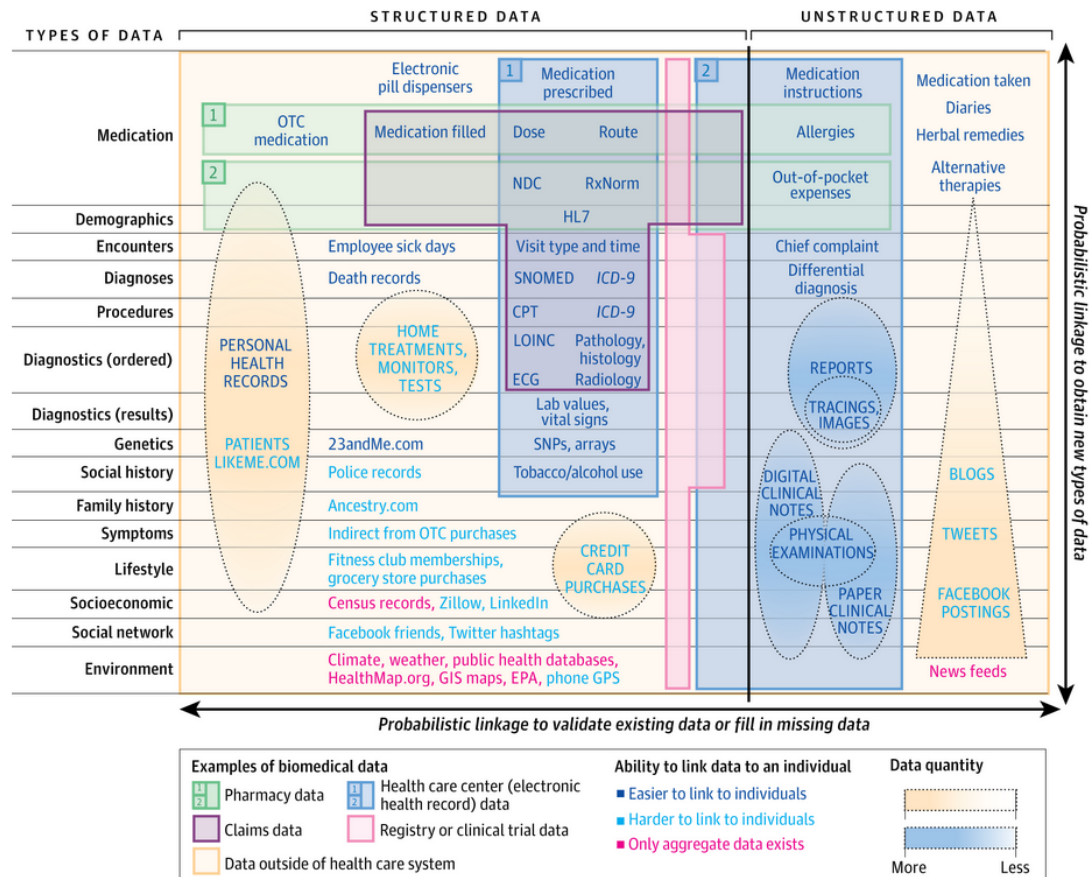


EXPLORATION

ANALYTIC VISUALIZATION



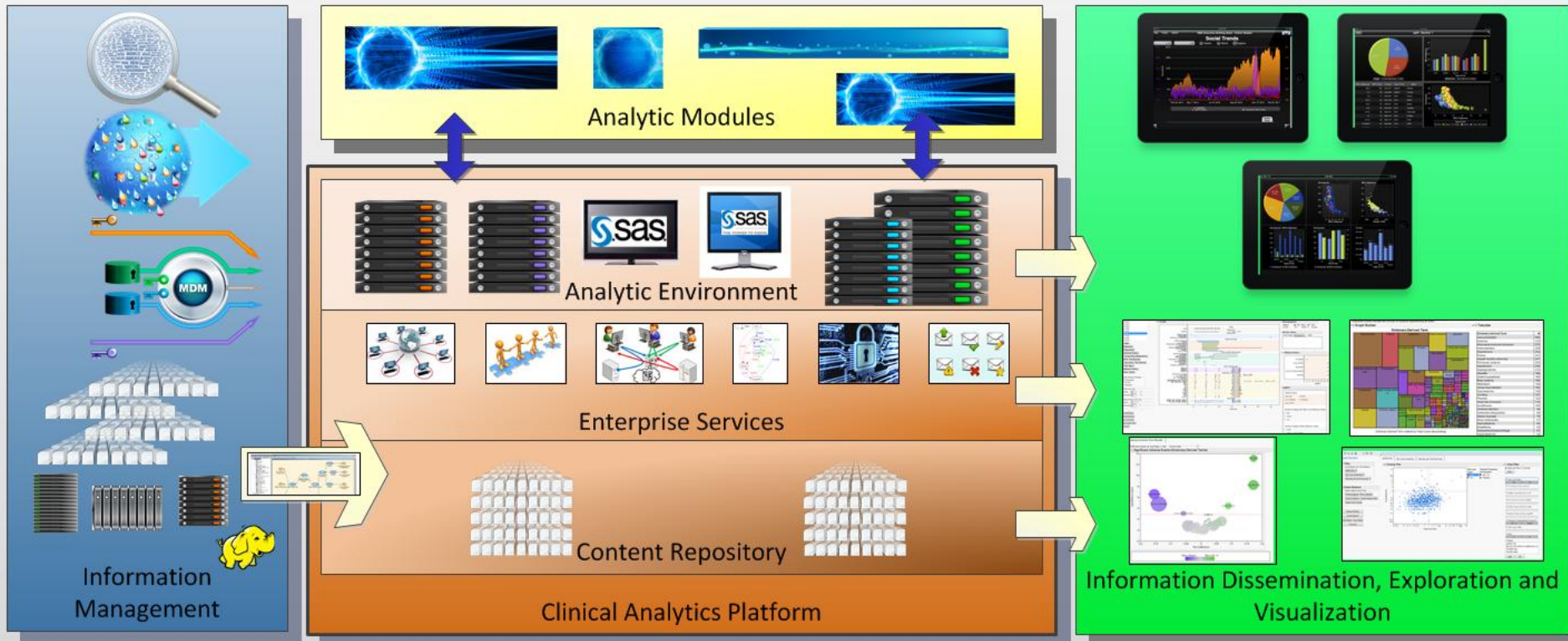
DISCOVERY



Source: "Finding the Missing Link for Big Biomedical Data", Weber *et al.*, 2014, Downloaded From: <http://www.jamafacial.com/> on 06/11/2014

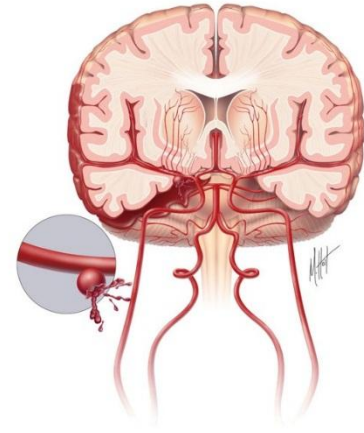
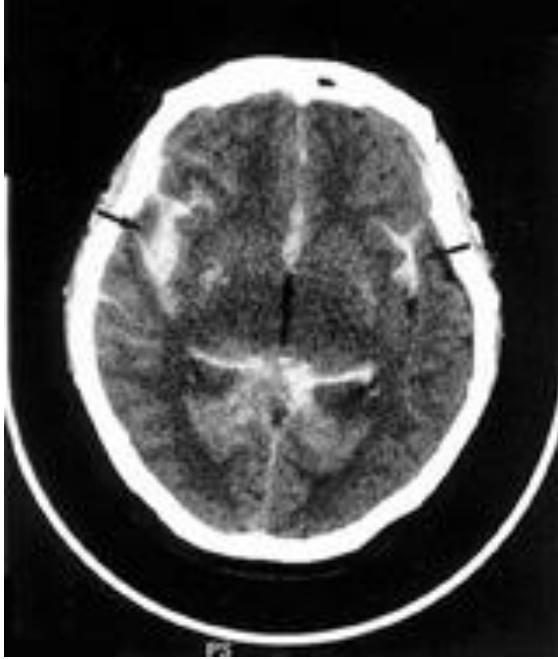
ENTERPRISE ANALYTICS

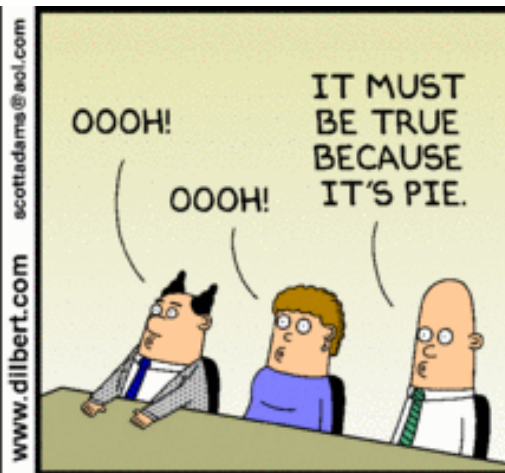
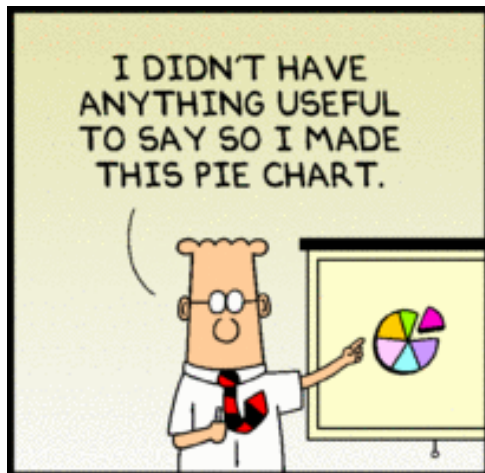
THE ANALYTICS LANDSCAPE





SUBARACHNOID HAEMORRHAGE : CLINICAL ENDPOINTS ?





www.dilbert.com

© 2009 Scott Adams, Inc./Dist. by UFS, Inc.

A NEW WORLD OF TECHNOLOGY TO IDENTIFY RELEVANT ENDPOINTS

SAS® Studio - SAS Programmer

Search

Folders

- Folder Shortcuts
- My Documents
- Desktop
- My Folders
- CA
 - \$Recycle.Bin
 - demodata
 - Documents and Settings
 - Intel
 - MSOCache
 - PerfLogs
 - Program Files
 - Program Files (x86)
 - ProgramData
 - Public
 - IMSTAT
 - Knowledge
 - data
 - data_analysis_usin
 - demo_short_with_
 - demo_short_with_
 - SAS® LASR™.pptn
 - Thumbs.db
 - VSTAT
 - Recovery
 - SASv9
 - Support
 - System Volume Information
 - temp
 - Users
 - V9setup
 - Windows

Tasks

Snippets

Libraries

File Shortcuts

*demo_short_with_ODS_10032013_v3_smpUSE.sas

CODE | LOG

```

proc imstat;
  /---MERGE---/
  table lasr.cardatap;
  schema carinfo (modelkey=modelkey / prefix=info) / mode=table;
run;

41 table lasr.carinfo;
42 fetch / from=1 to=5 format;
43 run;
44
45 /---EXPLORE---/
46 table lasr.cardatap;
47 distinct_all_ / save=tab1;
48 run;
49 * the name list in LIST1 is automatically created;
50 * no need to type in variable names one by one;
51 store tab1(where="Ndistinct le 5", 1)=list1;
52 run;
53 frequency list1;
54 crosstab wheelTypeID*wheelType;
55 corr mmr;
56 ods output frequency=work.freq1;
57 run;
58 table lasr.cardatap(tempnames=(avgOdo));
59 summary avgOdo / tempnames=avgOdo tempexpress="avgOdo = vehOdo / v
60 groupby=badbuy partition;
61 run;
62 quit;
63 ods graphics / height=4in width=12in;
64
65 proc spanel data=freq1;
66   panelby column / novarname columns=8 rows=1 uniscale=row;
67   vbar formattedvalue / group=column response=frequency datalabe
68 run;
69
70 proc imstat;
71 /---MERGE---/
72 table lasr.cardatap;
73 schema carinfo (modelkey=modelkey / prefix=info) / mode=table;
74 run;
75
76 /---CLEAN-UP---/
77 table lasr.&tempast;
78 promote cardata2;
79 where info.nationality ne 'NULL';
80 run;
81 table lasr.cardata2;
82 where;
83 compute delta1 "delta1 = vehboost - mmracquisitionretailaveragepr
84 compute delta2 "delta2 = vehboost - mmcurrentretailaveragepr
  
```

RESULTS

	30500	0	0	0
1				
2	0	33004	0	0
3	0	0	0	755
NULL	0	0	3159	0

Pairwise Correlations for Table WORK.CARDATAP

Column	Row	mmracquisitionauctionaverageprice	mmracquisitionauctioncleanprice	mmracquisitionretailaverageprice	mmracquisitionretailcleanprice	n
mmracquisitionauctionaverageprice	1	1.0000	0.9903	0.9104	0.9102	
mmracquisitionauctioncleanprice	2	0.9903	1.0000	0.9027	0.9182	
mmracquisitionretailaverageprice	3	0.9104	0.9027	1.0000	0.9902	
mmracquisitionretailcleanprice	4	0.9102	0.9182	0.9902	1.0000	
mmcurrentauctionaverageprice	5	0.9378	0.9239	0.8516	0.8459	
mmcurrentauctioncleanprice	6	0.9325	0.9321	0.8494	0.8557	
mmcurrentretailaverageprice	7	0.8718	0.8907	0.9131	0.9032	
mmcurrentretailcleanprice	8	0.8710	0.8712	0.9051	0.9058	

Summary Statistics for Table WORK.CARDATAP

auction	badbuy	Column	Min	Max	N	Sum	Mean	Std Dev.	Std Error	Coefficient of Variation	Number Missing
ADESA	N	avgOdo	2074.14	59757	12246	182126140	14873	5032.88	45.4769	33.8400	0
ADESA	Y	avgOdo	2061.14	38340	2193	20867437.5	13619	4298.17	91.7834	31.5561	0
MANHEIM	N	avgOdo	768.86	83729	36328	547241384	15064	5443.37	28.5593	36.1382	0
MANHEIM	Y	avgOdo	1870.13	47410	4716	62388270.7	13228	4292.05	62.5094	32.4528	0
OTHER	N	avgOdo	2118.89	48523	15433	256772573	16638	8907.80	46.7459	34.0059	0
OTHER	Y	avgOdo	1208.25	42628	2088	26776623.4	14399	4865.38	106.99	33.7903	0

Formatted



Clinical Project Manager



Business Analyst



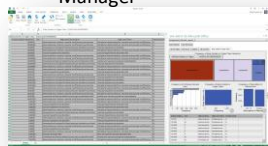
Business Analyst



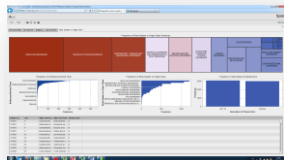
Data Scientist



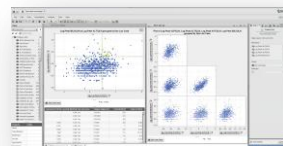
Epidemiological statistician



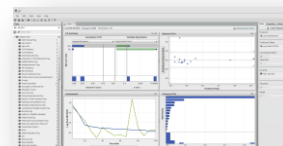
SAS® Office Analytics



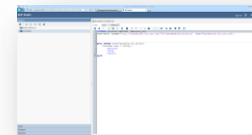
SAS® Visual Analytics Reporting



SAS® Visual Analytics Exploration



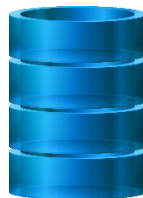
SAS® Visual Statistics



SAS® In-Memory Statistics



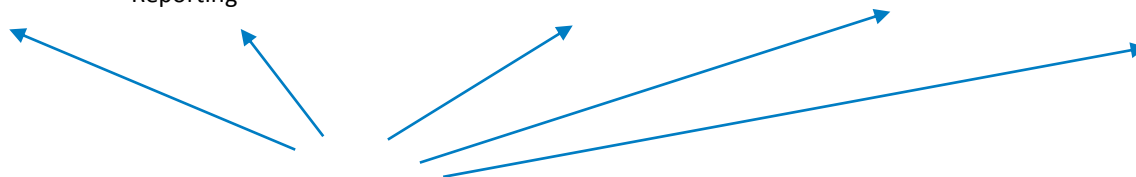
Oracle Hadoop



SAS LASR

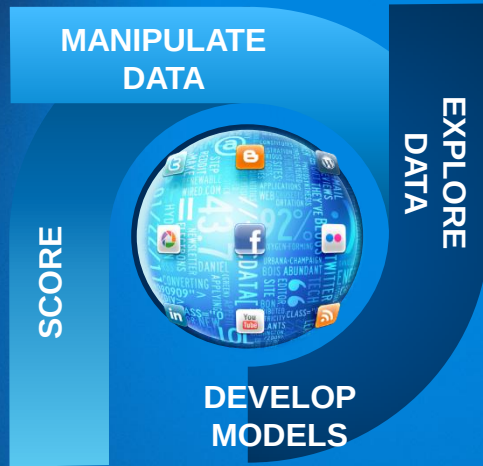


CSV Flatfiles



SAS IN-MEMORY STATISTICS FOR HADOOP SAS LASR ANALYTICS SERVER

- LASR Analytics Server = an analytic platform to provide a secure, multi-user environment for concurrent access to data that is loaded into memory
- PROC IMSTAT = server administration and analysis for programmers using the in-memory tables on the server



WHAT IS IT

Provides a single *interactive* programming environment to perform

- analytical data preparation
 - variable transformations
 - exploratory analysis
 - statistical modeling and machine learning
 - integrated modeling comparison and scoring
-
- Takes advantage of distributed in-memory computing optimized for analytical workloads

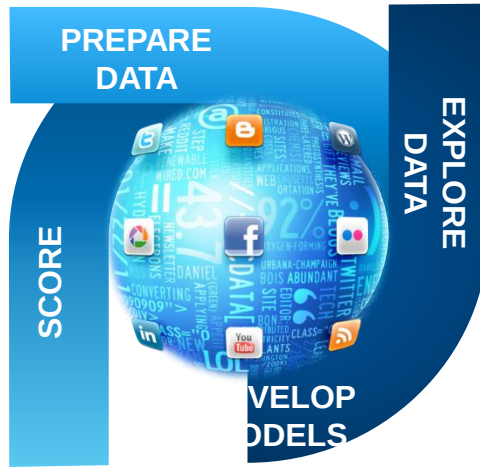
Data Manipulation

- Aggregate
- Compute
- Update
- Append
- Set
- Schema
- DeleteRows
- DropTables
- PurgeTempTables

Data Exploration

- Boxplot
- Corr
- Crosstab
- Distinct
- Fetch
- Frequency
- Histogram
- KDE
- MDSSummary
- Percentile
- Summary
- TopK

CAPABILITIES OF SAS® IN-MEMORY STATISTICS



Predictive Modeling

- Decision Tree
- Forecast
- Gen Linear Model
- Linear Regression
- Logistic Regression
- Random Forests
- Neural Networks

Descriptive Modeling

- Association
- Path Analysis
- Clustering (k-means)
- Clustering (DBSCAN)

Text Analytics

- Parsing
- SVD
- Topic generation
- Document projection

Recommendation Systems

- Association
- Clustering
- kNN
- SVD
- Ensemble

Evaluation, Deployment

- Assess
 - Misclassification matrix
 - Lift, ROC, Concordance
- Score
- Training / Validation

Utilities

- Where
- GroupBy
- TableInfo, ColumnInfo, ServerInfo
- Partition, Balance
- Store, Replay, Free
- Table, Promote

HDFS I/O

- Sasiola
- Sashdat
- Anyfile Reader

4 KEY POINTS | CLINICAL ENDPOINT DETERMINATION...

1. ...requires flexible visualization, advanced analytics and knowledge management, integrated
2. ...is an iterative process, before, during and after clinical trial
3. ...needs new technology that opens up unforeseen avenues to study structured and unstructured data, without limits related to volume, nature and correlation
4. ...foresees statistical calculation as an integral process of visualization -> surrogate endpoints' contribution to clinical endpoints

You are welcome at the SAS Forum Germany
2015 on June 9th in Bochum
Frühbuchertarif still available !

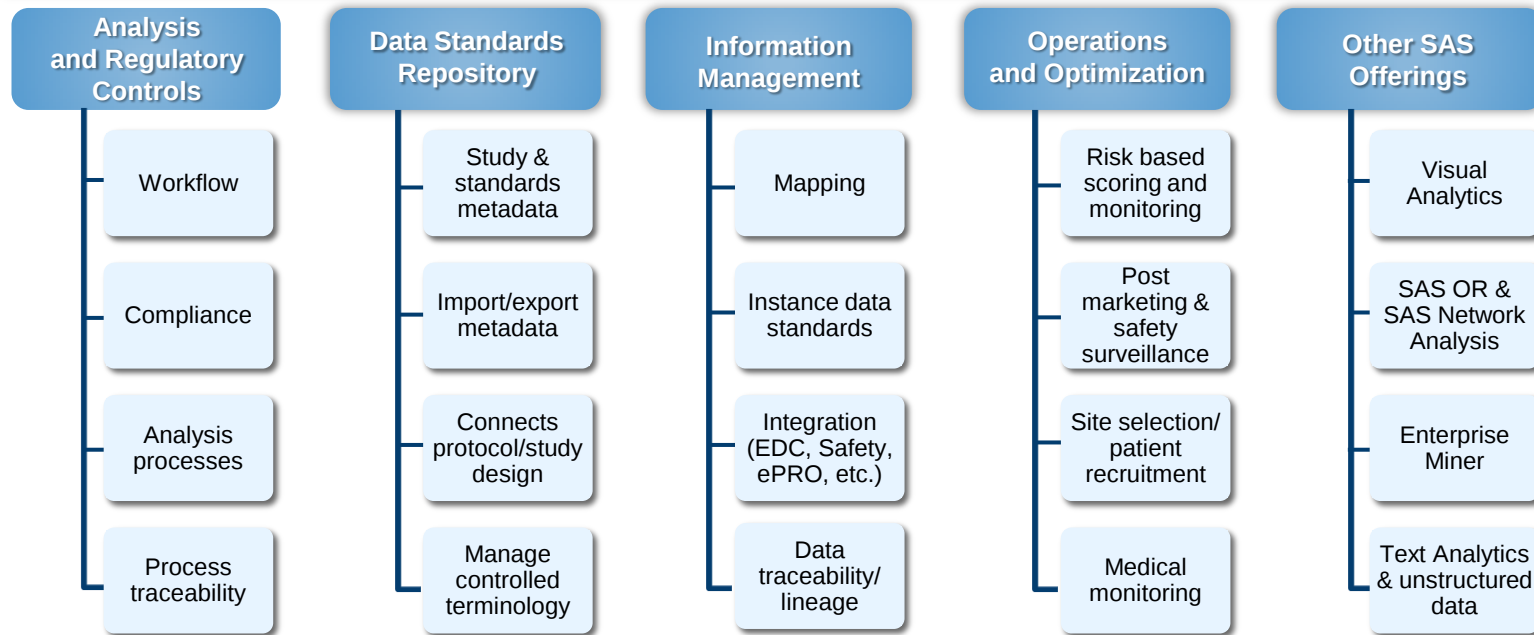
THANKS ! MARK.LAMBRECHT@SAS.COM

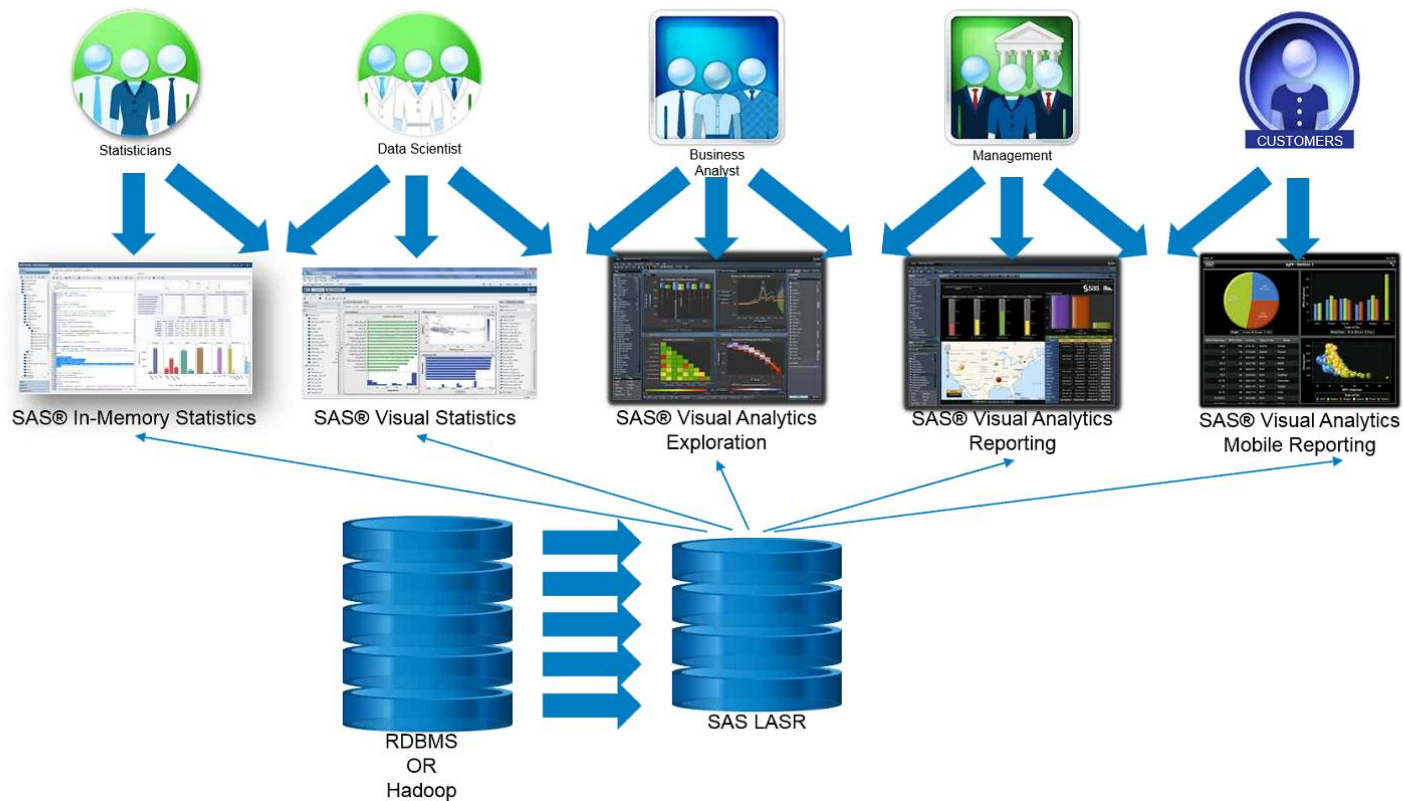


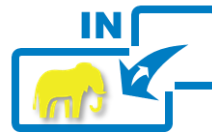
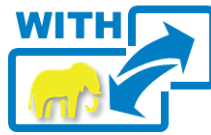
SAS® LIFE SCIENCE ANALYTICS FRAMEWORK

PRIMARY PLATFORM

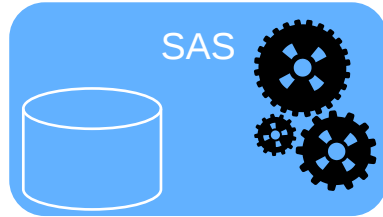
- Shared repository
- Single identity
- Partner collaboration
- Licensing
- Hosting model
- Access & audit trail



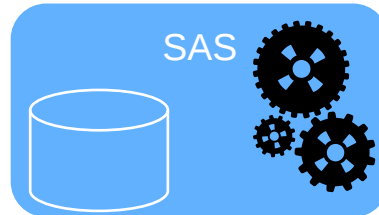




Regular SAS

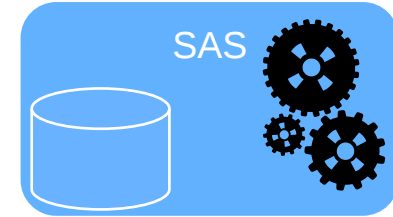
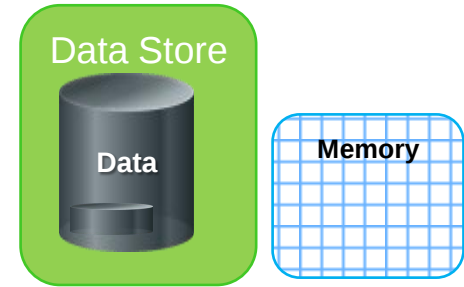


In-Database

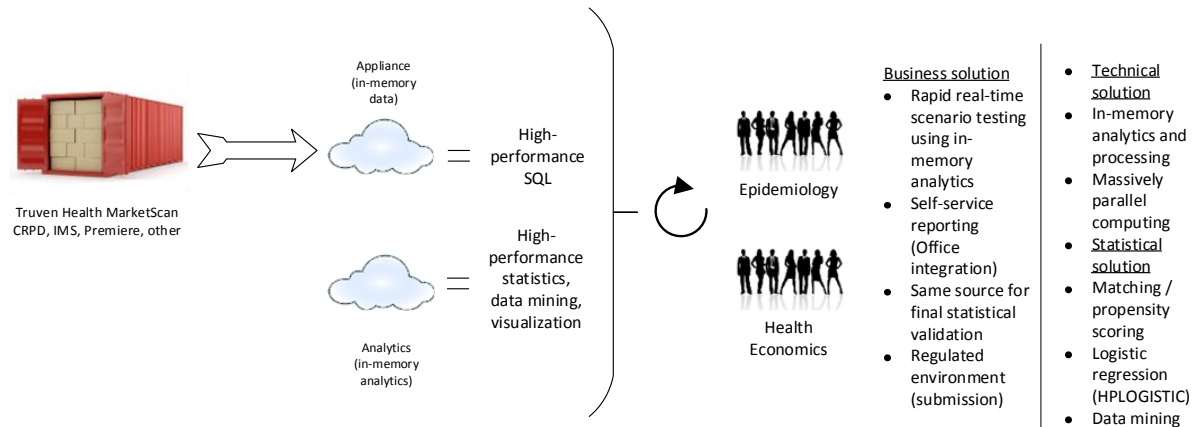
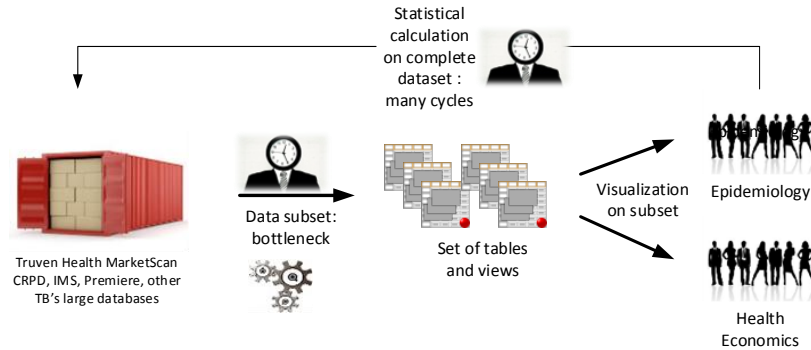


Even with In-Database processing there will still be some work performed on the SAS server

In-Memory

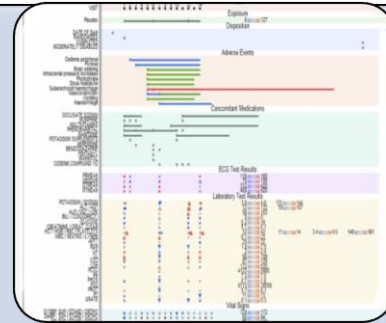
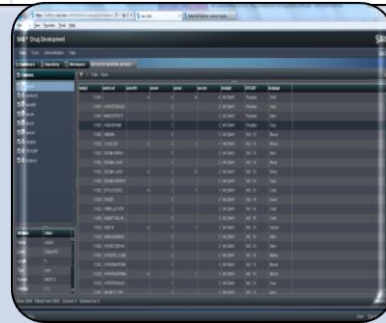
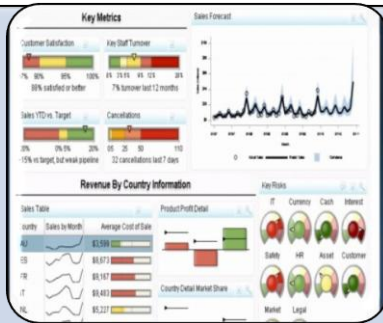


- These approaches are *complementary* & can be combined for maximum effect



VISUALIZATION

WHICH OPTION TO PROPOSE?



Business Intelligence

- **Personas:** General
- **Data:** General
- **Purpose:** Basic visualization/reporting
- **Value:** Easy understanding of information, trends and status

Data Explorer

- **Personas:** SAS Drug Development User
- **Data:** SAS Datasets in SDD
- **Purpose:** Quick view, filtering, structure exploration
- **Value:** Built in capability to browse and explore data sets

JMP

- **Personas:** Scientist, Medical Reviewer, Clinical Researcher
- **Data:** SDTM-like data sets
- **Purpose:** Clinical Trial specific visualization (e.g. Patient Profile, Patient Narratives, Medical Monitor Dashboard, Fraud Detection)
- **Value:** Wide variety of pre-built clinical solutions. Easy to use non-coder analytical viewer

Visual Analytics

- **Personas:** Senior Management, non statisticians
- **Data:** Large/Big data, Aggregated, Mixed purpose/secondary use
- **Purpose:** Ad hoc Exploration, Cohort Definition, Outlier Identification, Big Data Exploration
- **Value:** Intuitive, non-statistician explorer, Responsive large data exploration, Visually pleasing reporting and dashboards

ZIEKENHUIS GELDERSE VALLEI HOSPITAL

BUSINESS ISSUES

- Ziekenhuis Gelderse Vallei hospital measures performance for individual profit centers or units, but those units need to be able to analyze data to make informed decisions.
- Give units a way to measure performance and plan for the future.

SOLUTION

- [SAS® Visual Analytics](#)

RESULTS

- Physicians and other end users can now easily incorporate weather and socioeconomic data with care-related information to understand mortality risks, make predictions and inform diagnoses.
- Clinical personnel have the tools to produce reports, generate analyses, and find the answers themselves, thus freeing data staff.
- Analyze patient logistical processes to identify bottlenecks.
- Improved quality of care and a reduction in costs.

[Read the full story](#)

HEALTH CARE PROVIDER

“Previously, our colleagues on the floor could not produce reports or generate analyses. They used to ask us to do that for them. In our experience, every report we make generates 10 subsequent questions. With SAS® Visual Analytics, we give them the tools to create these reports themselves – and find the answers ... with a tool like SAS® Visual Analytics, it's easy to get people excited. It looks great and it is easy to use.”

Rik Eding
Data Specialist

GLOBAL PHARMACEUTICAL COMPANY HEADQUARTERED IN GERMANY

BUSINESS GOALS

- Provide access to clinical trial data ahead of potential European Medical Association (EMA) regulations.
- Take advantage of positive reputation impact from sharing clinical trial data.
- Ensure more positive patient outcomes through data sharing.
- Ensure security of clinical trial data.

SOLUTION

- [SAS® Clinical Trial Data Transparency](#)
- [SAS® Visual Analytics](#)

EXPECTED RESULTS

- Approved researchers can analyze clinical trial data and potentially benefit patients by gaining additional insights into clinical study results, how medicines work and how they can be further enhanced.
- SAS® Visual Analytics is packaged as front end visualization environment with potential internal use in IT, R&D, medical writing and regulatory affairs.
- Enhance the company's ability to provide the best patient outcomes based on its clinical studies.

LIFE SCIENCES

Ensure more positive patient outcomes by sharing clinical trial information from different pharmaceutical companies with researchers.

