

MIT HIGH PERFORMANCE IN DIE ANALYTIK VON SAS 9.4

GERHARD SVOLBA



AGENDA

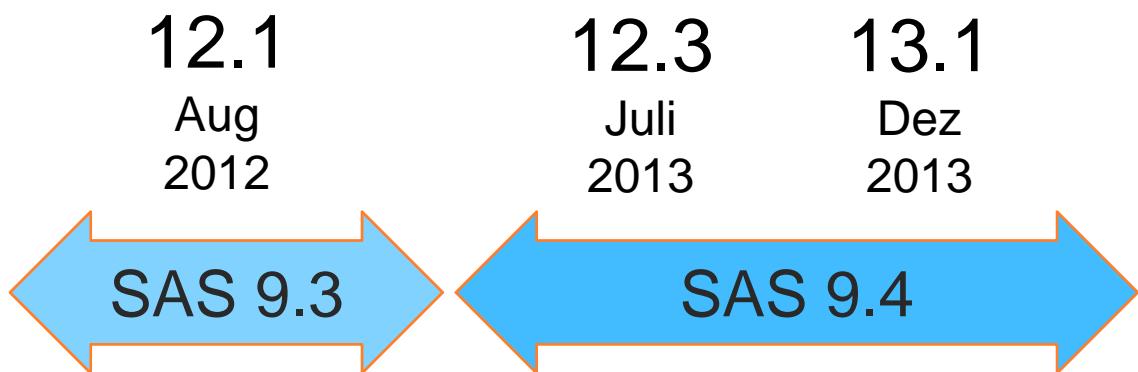
- SAS High Performance Analytics
- Neuigkeiten in den klassischen SAS Analytik Produkten



SAS 9.4

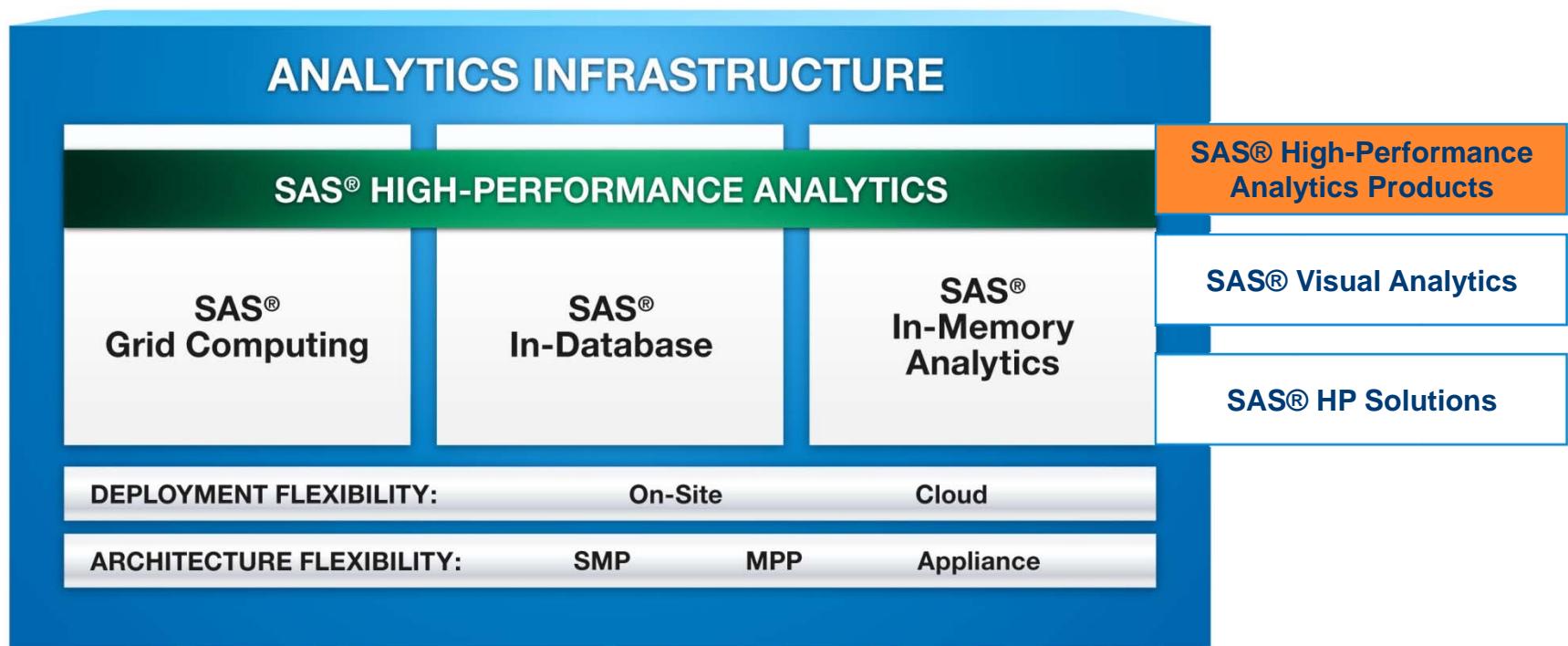
VERSIONSNUMMERN FÜR DIE SAS ANALYTIK-PRODUKTE

- SAS STAT
- SAS ETS
- SAS OR
- SAS QC
- SAS Enterprise Miner
- SAS Model Manager
- SAS Forecast Server
- SAS Text Miner
- SAS High Performance Analytics



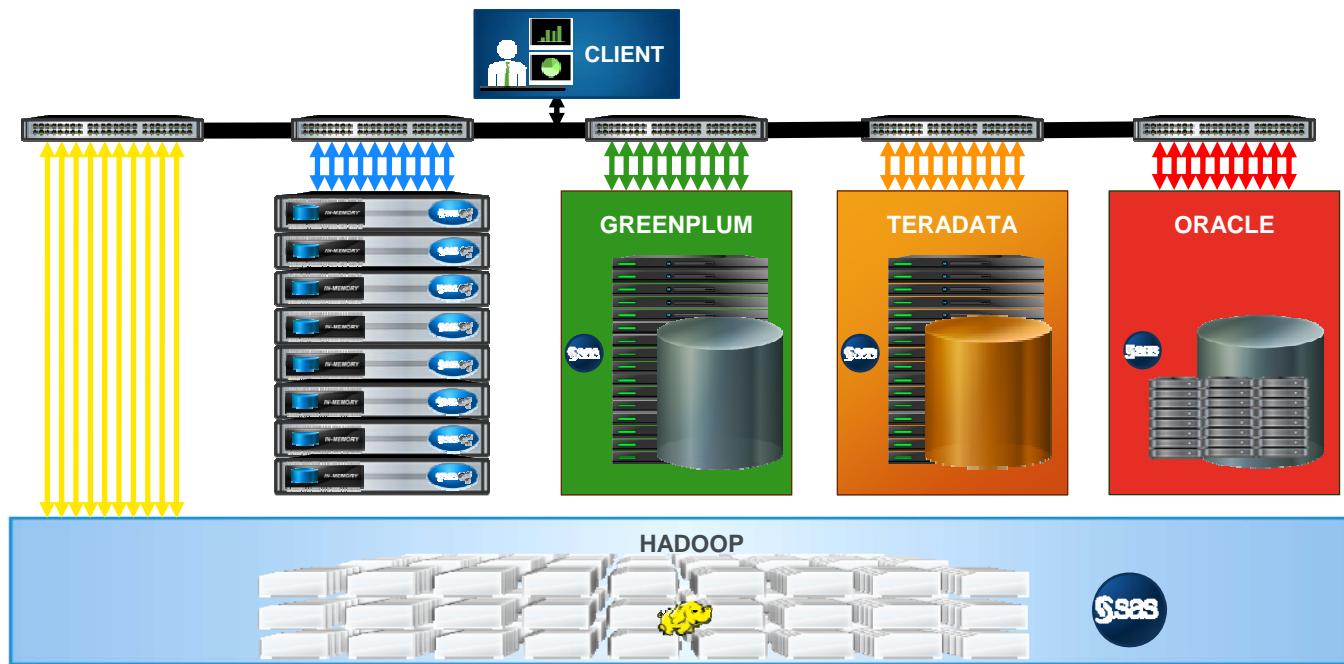
SAS HIGH PERFORMANCE ANALYTICS





SAS® HIGH-PERFORMANCE ANALYTICS

MÖGLICHE SAS COMPUTE APPLIANCE ARCHITEKTUR



- **Client oder lokale Daten:** Verteilung der Daten auf die Appliance / HDFS zur Laufzeit
- **Alongside the Database:** Die Daten liegen verteilt im DBMS vor und werden zur Laufzeit parallel in den Speicher geladen
- **Alongside HDFS:** Die Daten liegen verteilt im HDFS (SAS, Apache, Cloudera 4) und werden zur Laufzeit parallel in den Speicher geladen
- **Alongside LASR:** Die Daten liegen verteilt im LASR Server (in-Memory). Zur Laufzeit werden in-Memory Kopien genutzt

```
option set=GRIDHOST="hpa.sas.com";
option set=GRIDINSTALLLOC="/opt/TKGrid";
option set=GRIDMODE="„asym“ /* sym */;

libname local base compress=yes "c:\local_data";
```

```
libname applianc {greenplm, teradata, oracle, hadoop}
server ="hpa.sas.com" user =XXXXXX password=YYYYYY
database=mydb;

data applianc.simData;
  set local.simData;
run;
```

```
libname myhadoop sashdat hdfs_path="/hdfs";

data myhadoop.simData (replace = yes);
  set local.simData;
run;
```

```
proc lasr port=12345
  data=simData path="/tmp/";
run;
```

```
libname mylasr sasiola port=12345;
```

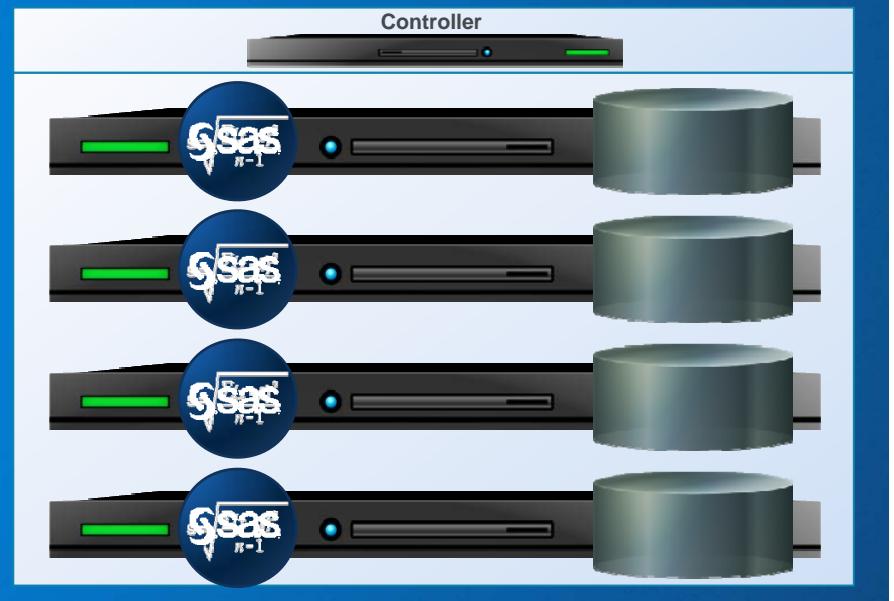
```
proc hplogistic
  data={local, applianc, myhadoop, mylasr}.simData;
  class a b c;
  model y = a b c x1 x2 x3;
run;
```

SAS® HIGH- PERFORMANCE ANALYTICS

IN-MEMORY (ALONGSIDE MODUS)



SAS® ANALYTICS
Client

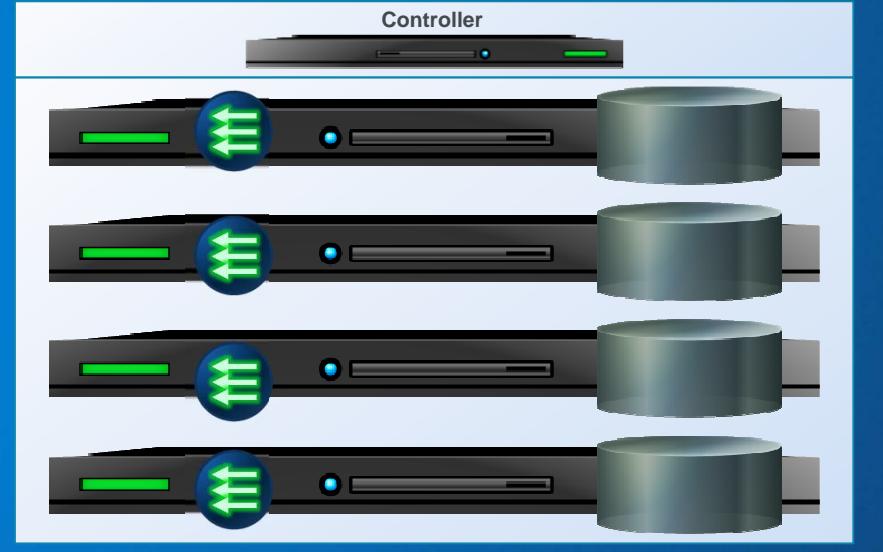


Teradata oder Oracle oder
Pivotal (Greenplum) Database oder
Hadoop auf Commodity Hardware



SAS® HIGH- PERFORMANCE ANALYTICS

IN-MEMORY (ASYMMETRISCHER MODUS)



SAS® HIGH-PERFORMANCE ANALYTICS

VERFÜGBARE SAS HIGH PERFORMANCE ANALYTICS PROCEDURES UNTER SAS 9.4 IN ANALYTICS 12.3

High-Performance Statistics	High-Performance Data Mining	High-Performance Text Mining	High-Performance Optimization	High-Performance Econometrics	High-Performance Forecasting
<ul style="list-style-type: none">HPLOGISTICHPREGHPLMIXEDHPNLMODHPSPLITHPGENSELECT	<ul style="list-style-type: none">HPREDUCEHPNEURALHPFORESTHP4SCOREHPDECIDE	<ul style="list-style-type: none">HPTMINEHPTMSCORE	<ul style="list-style-type: none">OPTLSOSelect features in<ul style="list-style-type: none">OPTMILPOPTLPOPTMODEL	<ul style="list-style-type: none">HPCOUNTREGHPSEVERITYHPQLIM	<ul style="list-style-type: none">HPFORECAST

Gemeinsame Prozeduren (HPDS2, HPDMDB, HPSAMPLE, HPSUMMARY, HPIMPUTE, HPBIN, HPCORR)[#]

- SMP Architektur: HP Prozeduren unter SAS 9.4 in Analytics 12.3 enthalten

SAS® HIGH- PERFORMANCE STATISTICS 12.3

HPLOGISTIC
HPREG
HPSPLIT
HPGENSELECT
HPLMIXED
HPNLMOD

Copyright © 2013, SAS Institute Inc. All rights reserved.

HIGHLIGHTS

- **Bestmögliche Performance** für Modellvorhersage und **Variablenelektion** auf großen Datenmengen
- **Produktivitätserhöhung** durch **Vereinigung** klassischer SAS/STAT **Funktionalitäten**, z.B.
 - HPREG = REG \cup GLM \cup GLMSELECT
 - HPGENSELECT = GENMOD + Variablenelektion
- **Neue Funktionalitäten**
 - Entscheidungsbäume

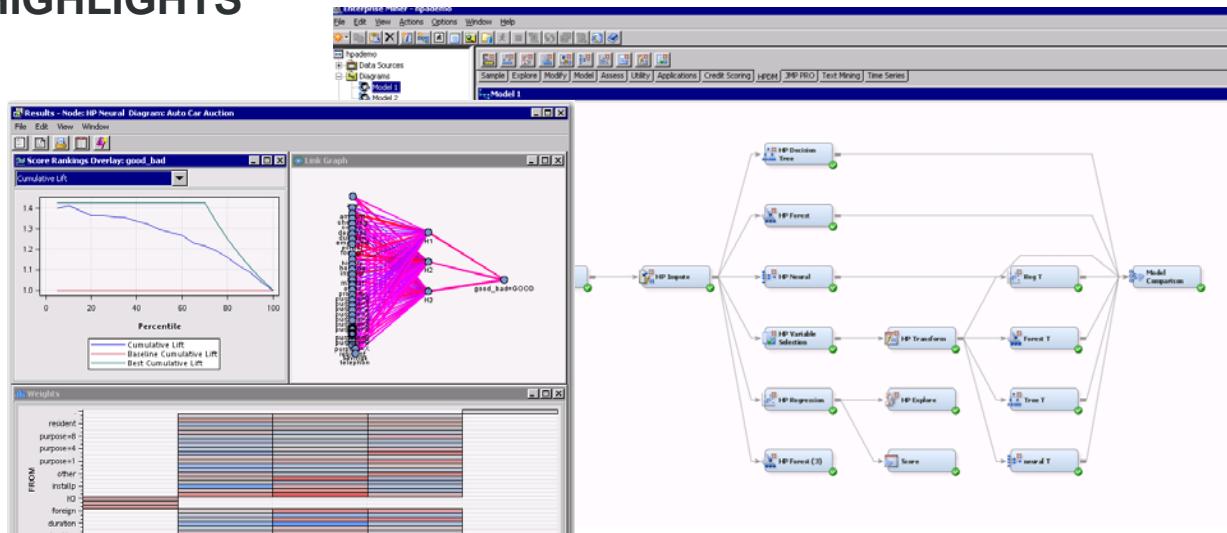


SAS® HIGH-PERFORMANCE DATA MINING 12.3

HP Data Partition
HP Explore
HP Transform
HP Impute
HP Variable Selection
HP Regression
HP Neural
HP Forest
HP Tree

Copyright © 2012, SAS Institute Inc. All rights reserved.

HIGHLIGHTS



- HP Data Mining beinhaltet HP Statistics
- 9 Knoten bündeln die Funktionalität von HP Statistics und zusätzlicher HP Data Mining Prozeduren
 - HPREDUCE, HPFOREST, HPNEURAL, HPDECIDE

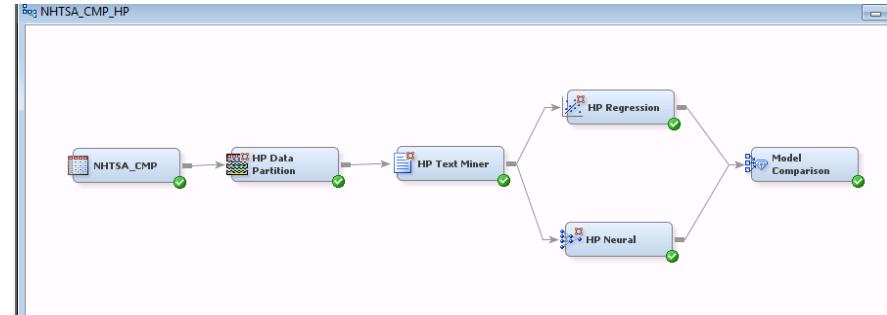
SAS® HIGH PERFORMANCE TEXT MINING 12.3

HPTMINE
TGPARSE
TMUTIL
SPSVD

HPTMSCORE
TGSCORE
SPSVD

Copyright © 2012, SAS Institute Inc. All rights reserved.

HIGHLIGHTS



- **Schnelles Parsing großer Textsammlungen**
 - Natural Language Processing
 - Term-Filterung & -Gewichtung
 - Dimensionsreduktion
- **Starke Prozessbeschleunigung** durch
 - Integration von Funktionalitäten
 - Multi-Threading und verteiltes Rechnen
 - Weniger I/O durch In-Memory Technologie
- Prozeduren **auch im Enterprise Miner™ nutzbar**

SAS® HIGH-PERFORMANCE OPTIMIZATION 12.3

OPTLSO

Features in
OPTMILP
OPTLP
OPTMODEL

Copyright © 2012, SAS Institute Inc. All rights reserved.

HIGHLIGHTS

- High-Performance Optimization zur
 - Lösung **großer**, strukturell **komplizierter Fragestellungen**
 - **Beschleunigung** diverser **Solver** (LP, MILP, QP, NLP)
- **Neu: High-Performance Local Search Optimization** für ableitungsfreie, gemischt-ganzzahlige Probleme
- **Neu: Ausgewählte Features**
 - **Multistart** Option für nicht-lineare Zielfunktionen
 - **Parametertuning** für gemischten-ganzzahligen Solver
 - **Dekompositionsalgorithmus** für große, LP & MILP mit speziellen Constraint-Strukturen



NEUIGKEITEN IN DEN KLASSISCHEN SAS ANALYTIK PRODUKTEN



ALLGEMEINE EIGENSCHAFTEN

- Pre-Built EM Templates für spezifische Analyseaufgaben
- Möglichkeit EM Diagramme direkt beim Start zu öffnen
- Support von PMML 4.1
- Verbesserte “Reporter” styles
- Survival Data Mining
 - Time varying covariates
 - Support für “user defined censored” dates und “truncation”
- Erweiterungen im Incremental Response Node
- Credit Scoring
 - Unterstützung von interval targets
- Enterprise Miner 4.3 wird nicht mehr ausgeliefert

DECISION TREE NODE UNTERSTÜTZT

- user-specified color schemes
- user-specified display of text and statistics in the nodes of a tree
- user-controlled sample size within interactive trees

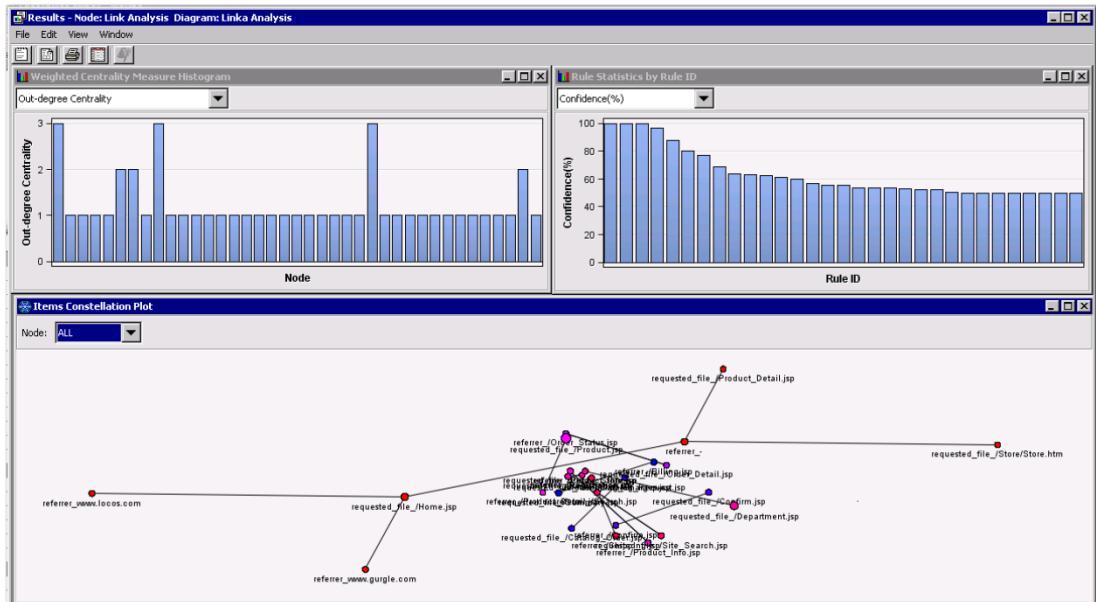
Report	
Precision	4
Tree Precision	4
Class Target Node Color	Percent Correctly Classified
Interval Target Node Color	Average
Node Text	...

Interactive Sample	
--Create Sample	Default
--Sample Method	Random
--Sample Size	10000
--Sample Seed	12345

Node Text	
Property	Value
Node ID	Yes
Show Validation	Yes
Count	Yes
Class Targets	
--Predicted Value	No
--Target Values	All Values
--Percent Correct	No
--Profit or Loss Values	None
Interval Targets	
--Average	Yes
--Root Average Squared Error	No

SAS ENTERPRISE MINER 12.3

NEW LINK ANALYSIS NODE



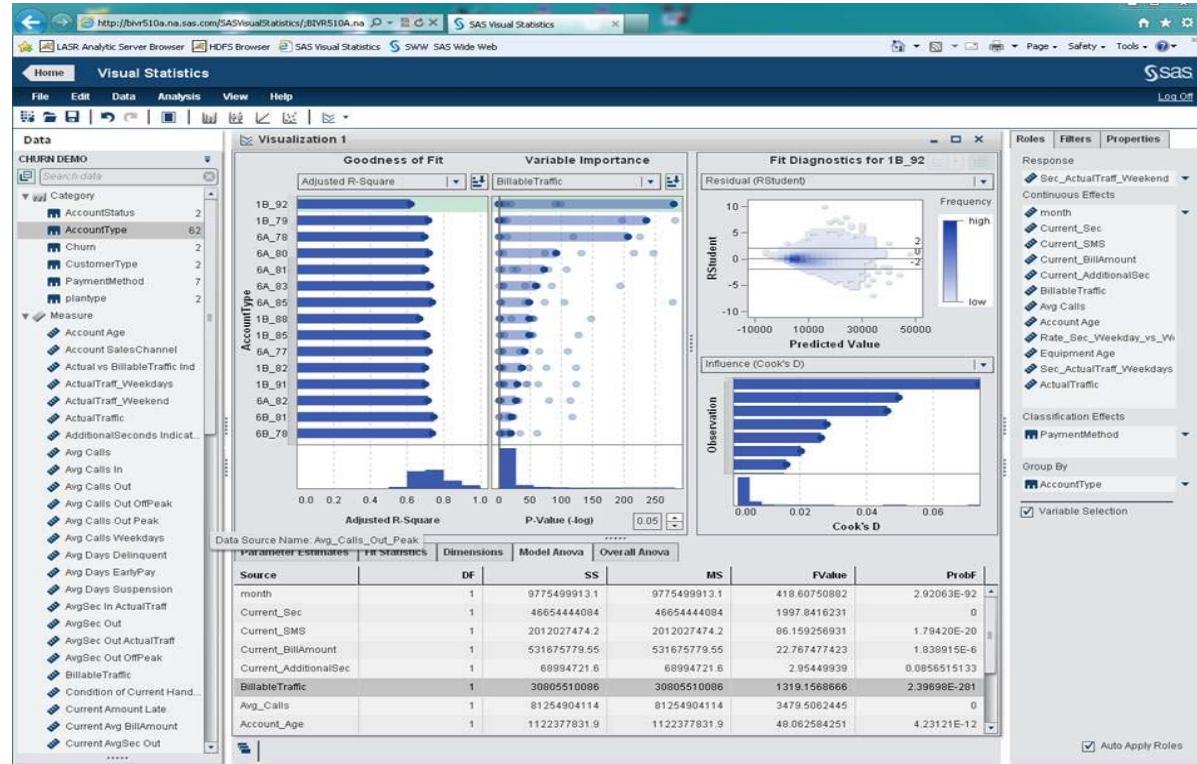
Train	
Variables	...
Minimum Confidence(%)	50
Association Support Type	Count
Association Support Percentage	5.0
Association Support Count	5
Sequence Property Settings	...
Binning	
Binning Method	Bucket
Number of Bins	5
Centrality Measures	
Centrality Measure Settings	...
Item-cluster	
Resolution	1.0
Item-cluster Detection Settings	...
Score	
Recommendation	
Generate Recommendation Table	No
Top N	1
Minimum Confidence(%)	.
Criterion Relation	AND
Export Recommendation Table	Full Recommendation T

ALLGEMEINE EIGENSCHAFTEN

- API für das R Modell-Training und die Ergebnis-Integration
- “Sticky notes”
- Neue HP nodes für Clustering und Principal Components
- Support Vector Machines
- Bayesian Belief Networks

NEUES USER INTERFACE FÜR DIE DATENANALYSE

Auch neue Web Clients für
SAS Enterprise Miner
SAS Forecast Studio
ab der Version 13.1



- SAS Visual Statistics <http://www.youtube.com/watch?v=G6bxjCC6IYc>

AUSGEWÄHLTE NEUIGKEITEN

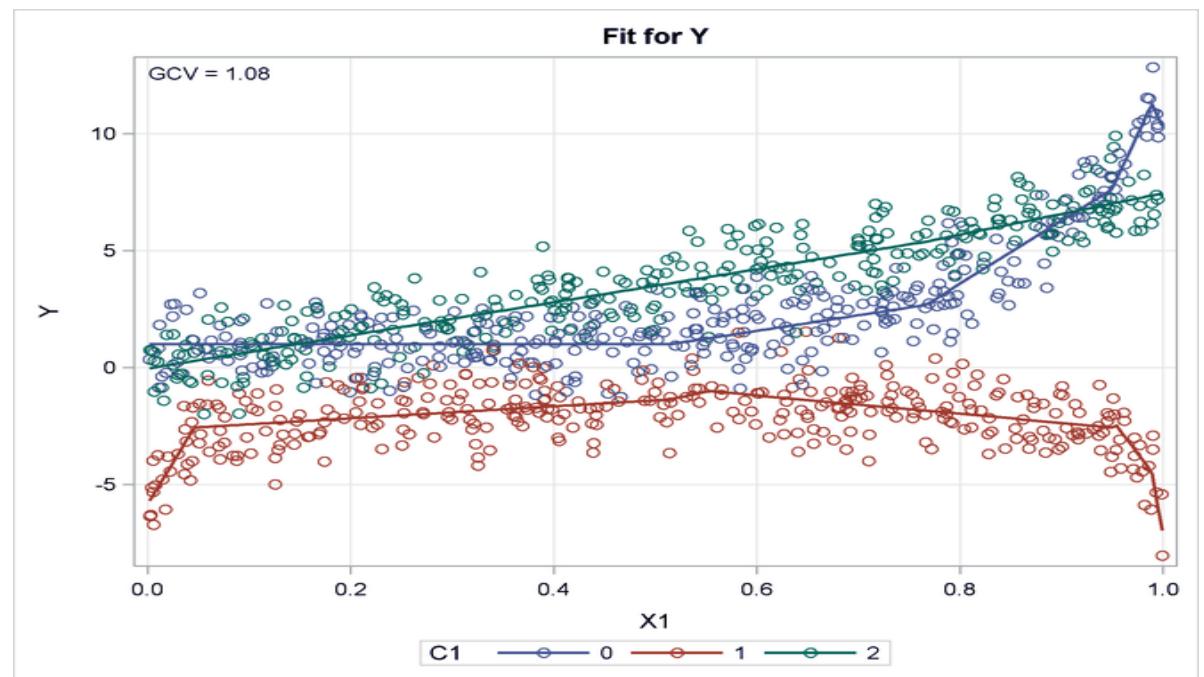
OR

- Parallel Solver (LP, MILP, QP, NLP) mit dem COFOR loop
- Threaded MILP solver
- Diverse Performance improvements

ETS

- High-performance Bayesian econometric modeling
- Combined frequency-severity model simulation procedure
- Neue Procedure für generalized exponential smoothing models

- Die neue ADAPTIVEREG Procedure ermöglicht einen non-parametrischen Modellierungsansatz für hoch-dimensionale Daten.
- PROC ADAPTIVEREG schätzt multivariate adaptive regression splines (siehe Friedman, 1991).



- Die neue PROC PLM stellt ein Rahmenwerk für alle post-fitting Analyse Statements dar – CONTRAST, LSMEANS, LSMESTIMATES, ESTIMATE
- Post-Fitting Inference (testing hypotheses, computing confidence intervals, producing prediction plots, scoring a new data set) kann betrieben werden, ohne das Modells nochmals anzupassen.

ZUSAMMENFASSUNG

- SAS 9.4 inkludiert per Ende 2013 folgende Analytik Versionen: 12.3 und 13.1
- SAS High Performance Procedures sind im Single Machine Modus nun in den jeweiligen Analytik-Produkten verfügbar
- Viele neue Features in den Analytik Modulen
- SAS Visual Explorer wird verfügbar



LINKS

- <http://support.sas.com/rnd/app/index.html>
- <http://support.sas.com/rnd/app/stat/papers/index.html>
- <http://support.sas.com/rnd/app/da/stat.html>