MarketShare

MarketShare Big Data Analytics

An Big Data Analytics architecture for the cloud

MarketShare

MarketShare: Modeling on Big Data

- Cloud architecture evolution
- Equations Compiler
- Distributed modeling on the cloud

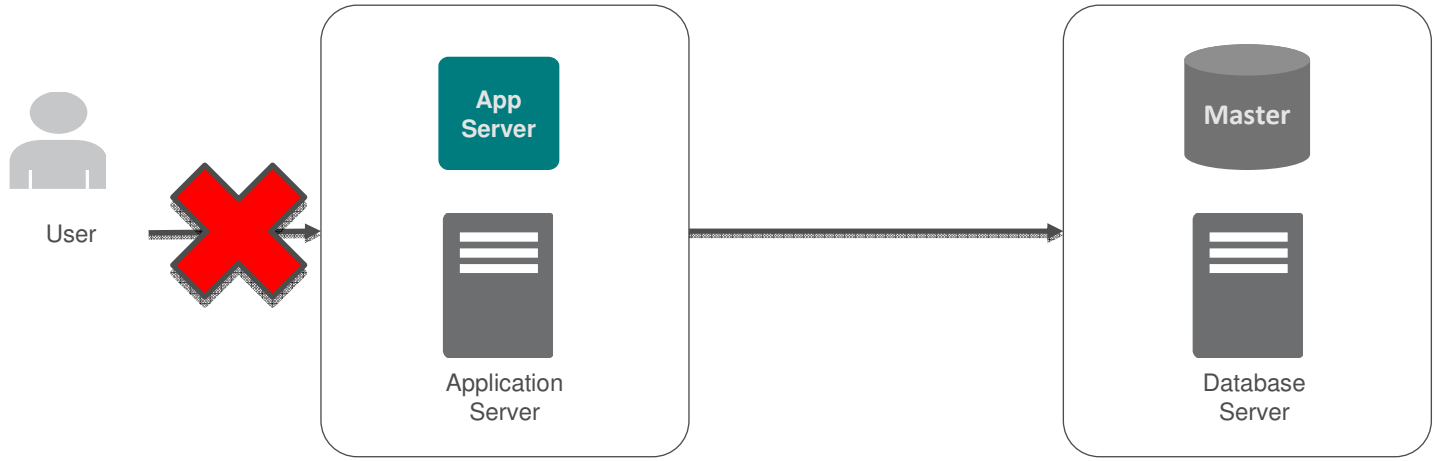
* Source: Interbrand 2011 report

MarketShare

Cloud + Big Data

MarketShare

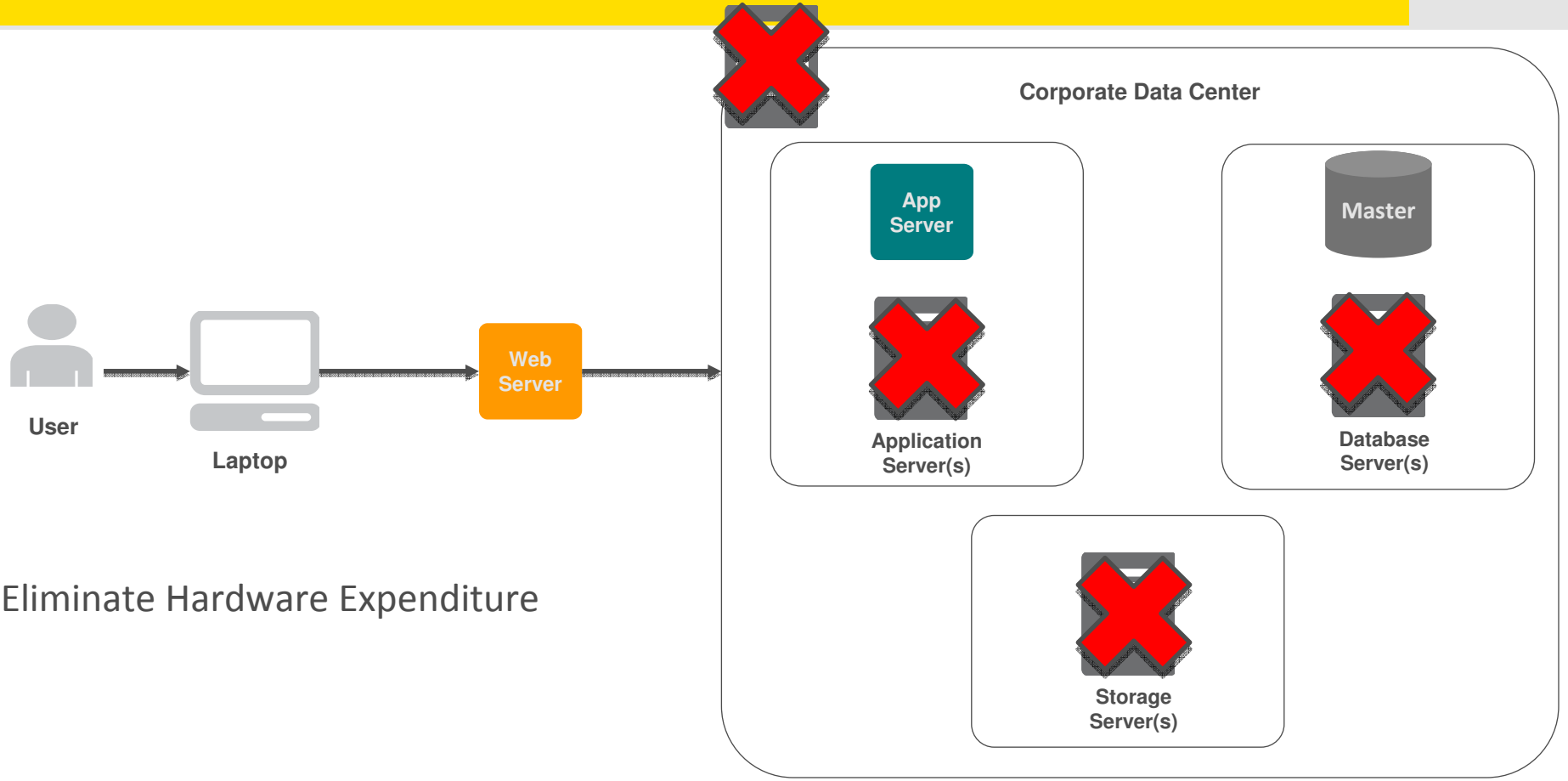
Traditional 3 tier architecture



Eliminate accessibility restrictions

MarketShare

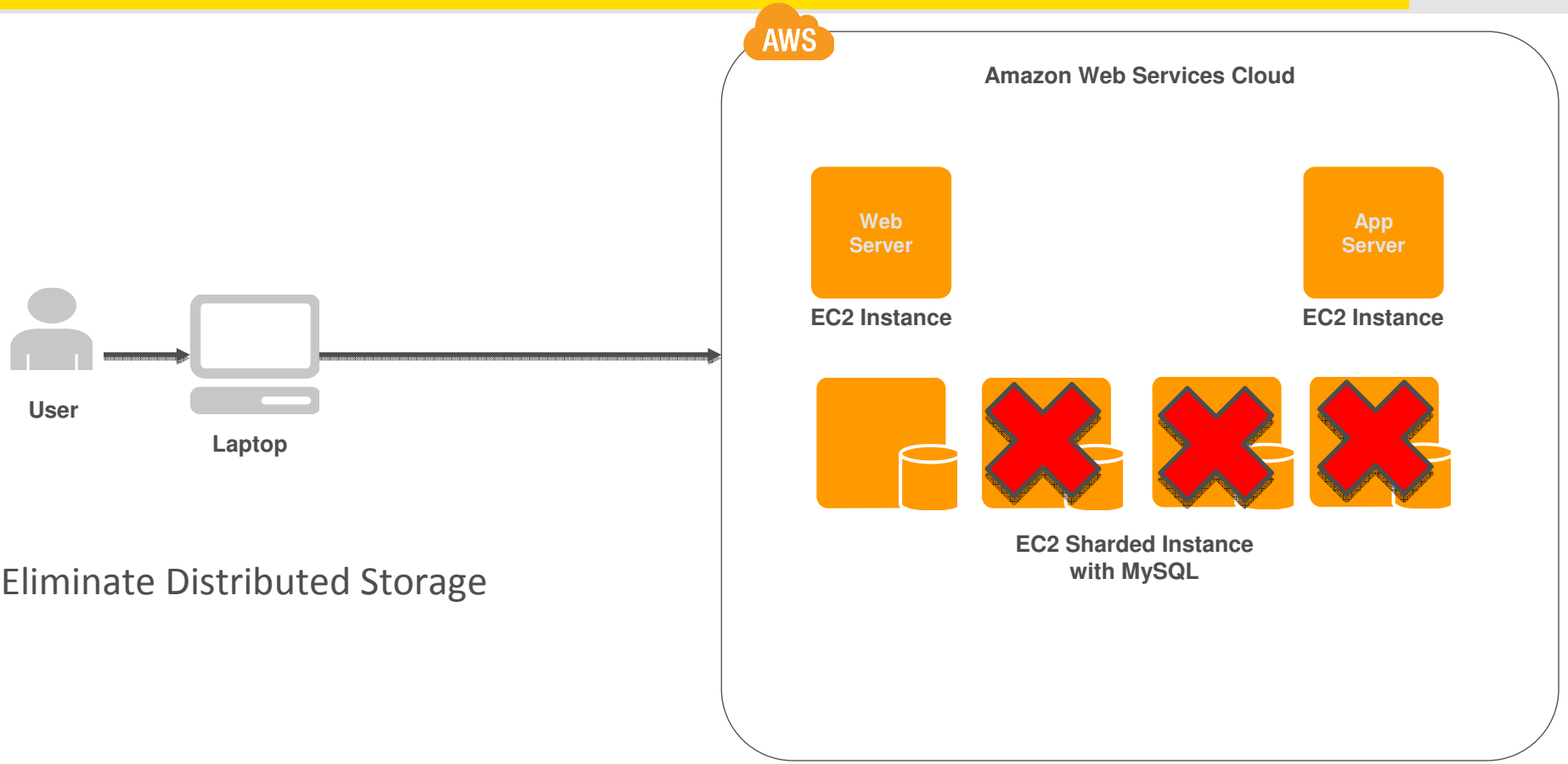
Moving to web based applications



Eliminate Hardware Expenditure

MarketShare

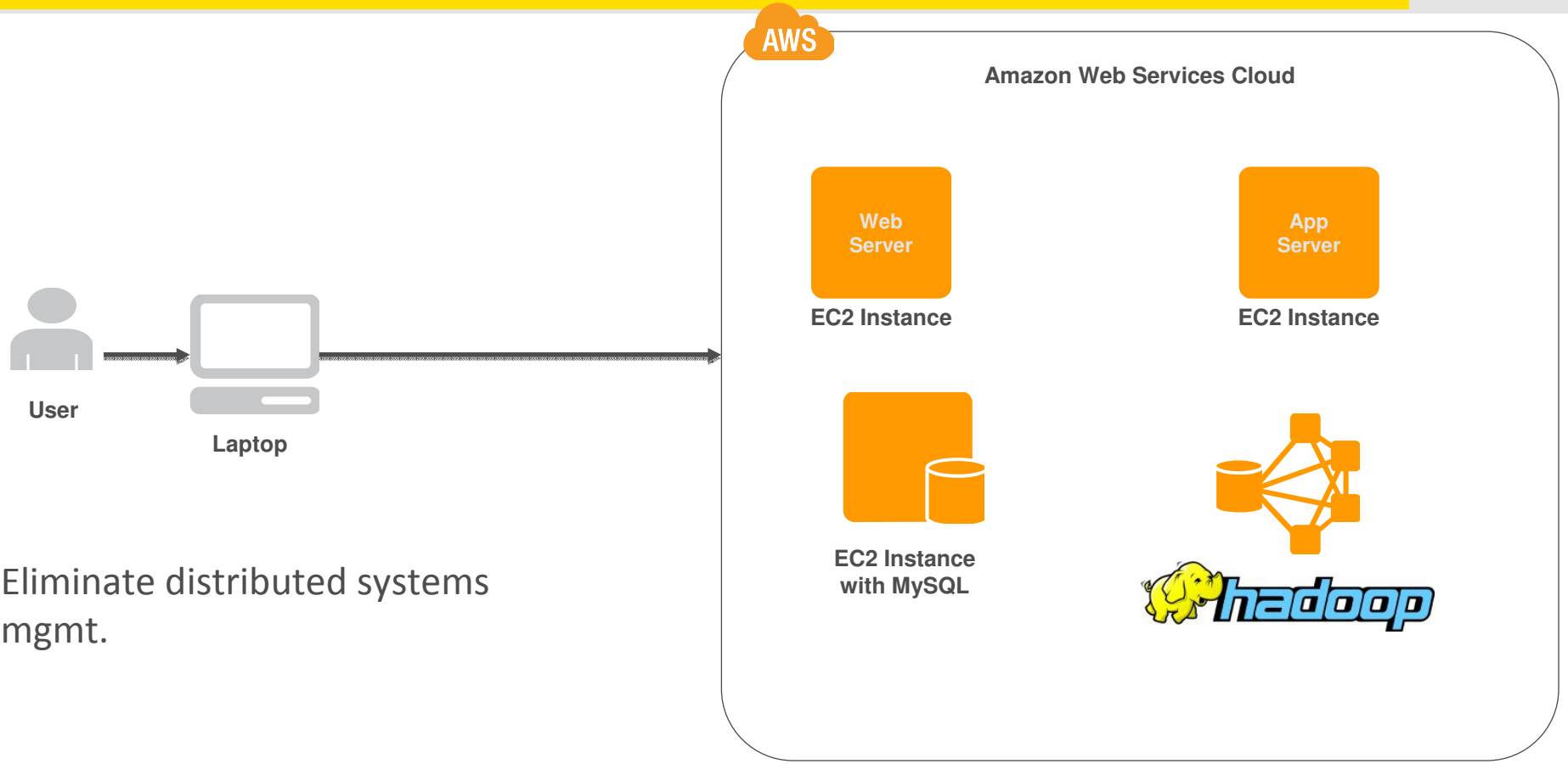
Moving to the cloud



Eliminate Distributed Storage

MarketShare

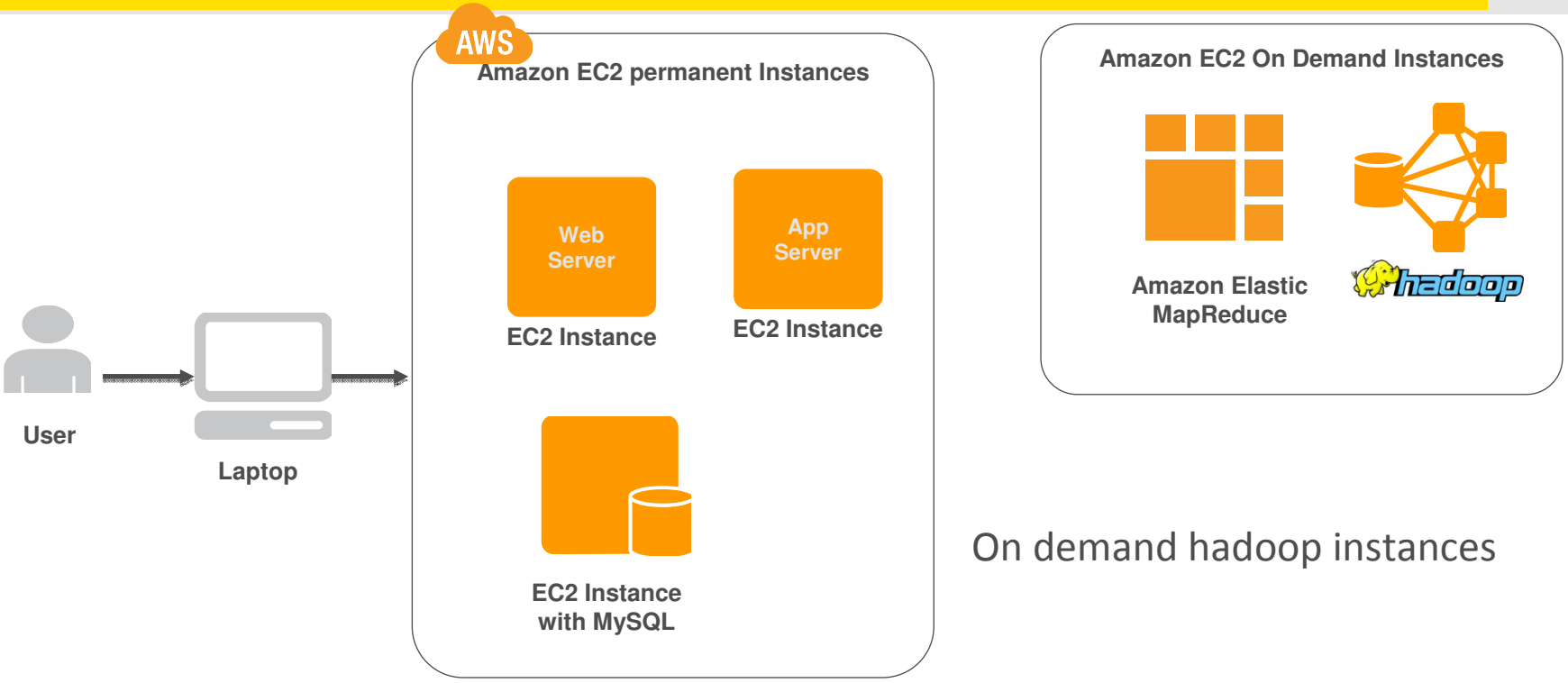
Moving big data to Hadoop



Eliminate distributed systems mgmt.

MarketShare

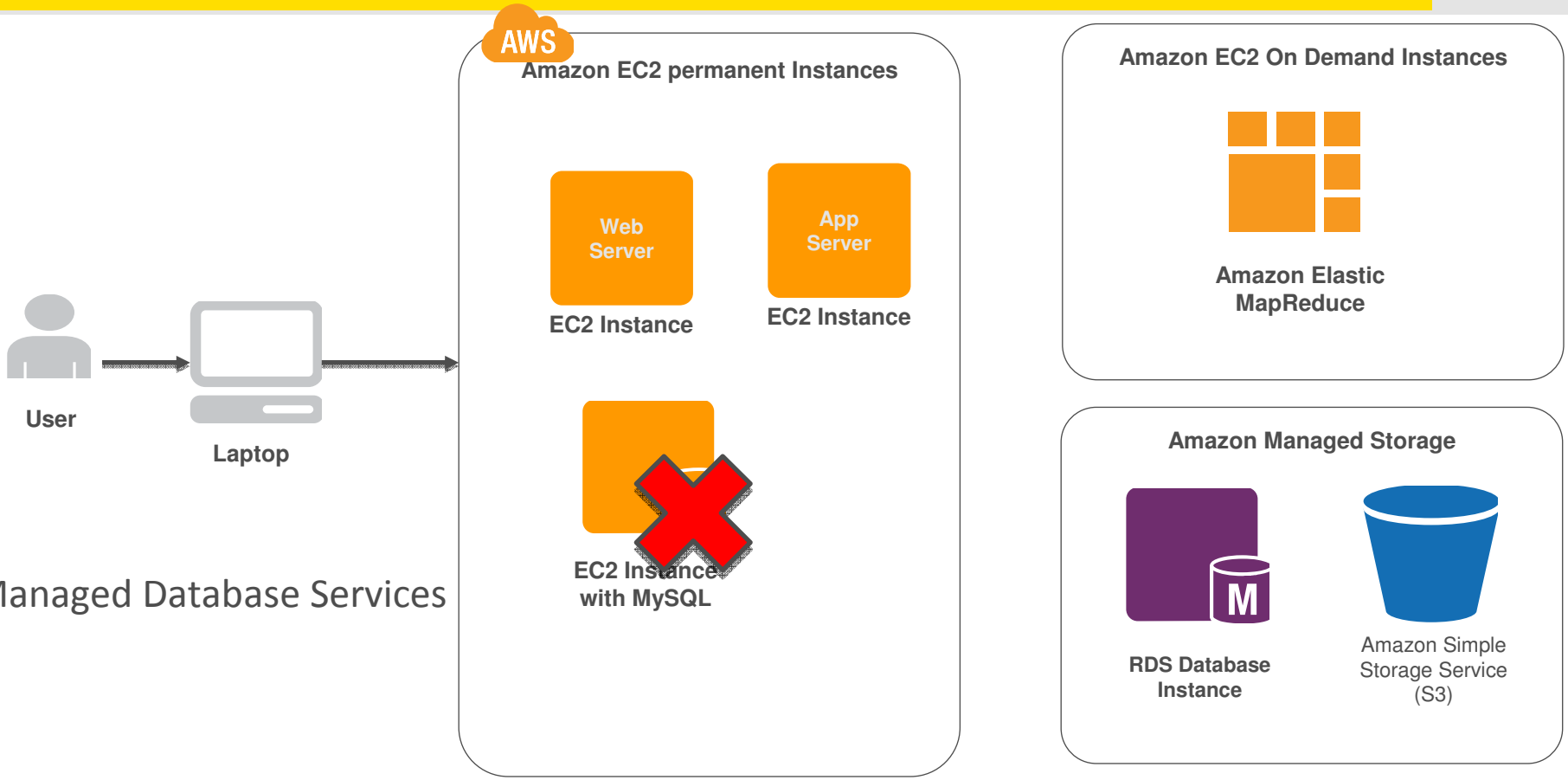
Compute Elasticity



On demand hadoop instances

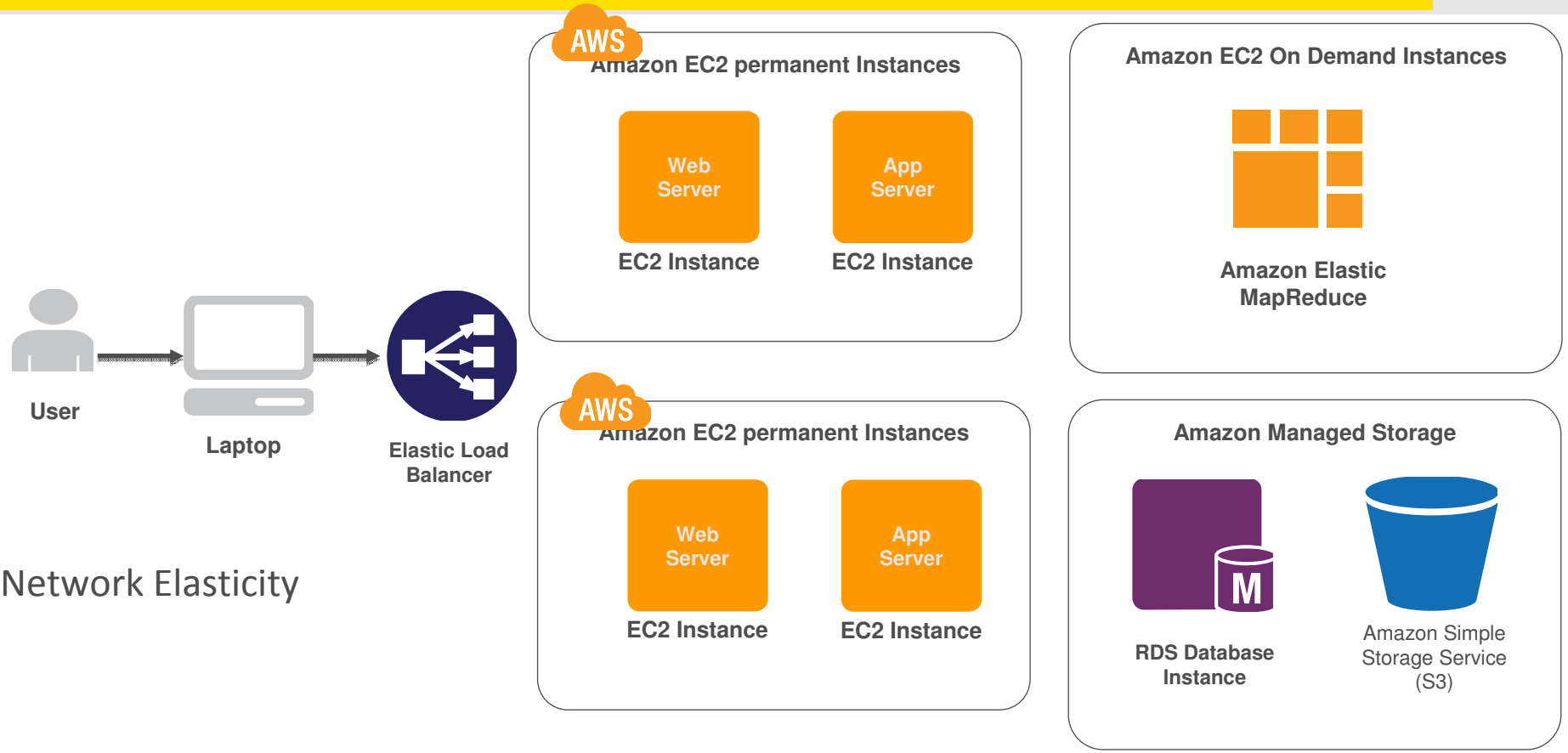
MarketShare

Storage Elasticity



MarketShare

Network Elasticity



Network Elasticity

MarketShare

Defining the Cloud

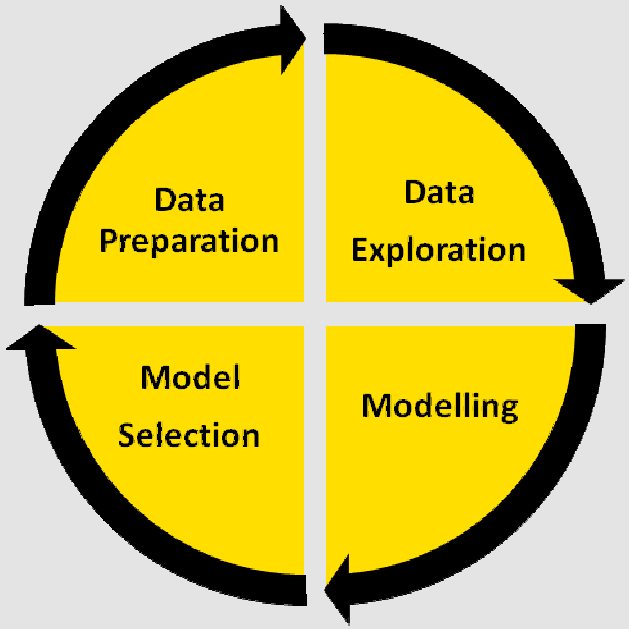
Cloud = Managed Storage + Network Elasticity + On Demand Compute

MarketShare

Cloud + Big Data + Modeling

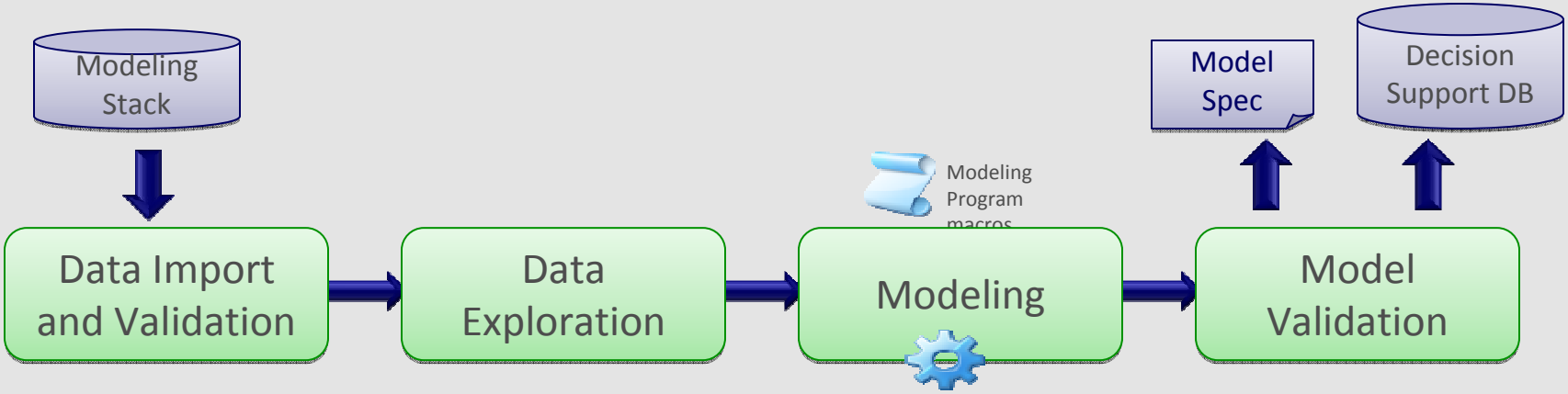
MarketShare

The Technology Puzzle



MarketShare

Modeling Process



Import to Modeling Program DB	Enhance Database	Modeling Demands	Confirm Models
• Time series of sales • Event, team variables • GQV, economic variables	• Variable generation E.g. game winner streak, % tickets sold • Exploration to understand and improve	• Regression based time series modeling • Initial starting variables • Statistical selection of best variables and weights • Iterative process	• Development and Holdout samples • Confirm model will work for future events

Modeling Objective: Find the best function:

Ticket Demand = F(time. event. team. GQV. economics. etc.)

MarketShare

Equations Compiler

MarketShare

An Equation

Dependent Variable: LOG(FULL_REV)
Method: Panel EGLS (Cross-section weights)
Date: 12/02/10 Time: 23:20
Sample: 1/15/2005 4/24/2010 IF PRODUCT="ACR"
Periods included: 276
Cross-sections included: 2
Total panel (balanced) observations: 552
Linear estimation after one-step weighting matrix

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.928059	0.581194	3.317407	0.0010
LOG(FULL_REV(-1))	0.435265	0.031105	13.99354	0.0000
D_JULY1407	-0.972718	0.123655	-7.866387	0.0000
NVER_ACR89S	0.088231	0.018782	4.697745	0.0000
NVER_ACR89W(5)	-0.539195	0.089515	-6.023482	0.0000
NVER_ACR89W(4)	-0.361140	0.090928	-3.971722	0.0001
NVER_ACR89W	0.374661	0.089790	4.172631	0.0000
NVER_ACR89W(-1)	-0.294214	0.092878	-3.167755	0.0016
M01	0.132766	0.031783	4.177300	0.0000
M02	-0.007974	0.029042	-0.274556	0.7838
M03	0.074205	0.029115	2.548652	0.0111
M04	-0.009924	0.029221	-0.339623	0.7343
M05	0.006072	0.030147	0.201425	0.8404
M06	-0.031082	0.033140	-0.937899	0.3487
M08	-0.027964	0.030574	-0.914643	0.3608
M09	-0.048330	0.029858	-1.618646	0.1061
M10	-0.019334	0.030224	-0.639684	0.5227
M11	0.128423	0.035376	3.630205	0.0003
M12	-0.043087	0.033052	-1.303608	0.1929
H_CHRISMAS	-0.417512	0.064325	-6.490661	0.0000
H_USTHANKS	-0.465110	0.073687	-6.311968	0.0000
H_MLKING	-0.136058	0.065313	-2.083173	0.0377
H_VET_REM	-0.152138	0.066112	-2.301234	0.0218
H_GOODFRI	-0.190309	0.091691	-2.075553	0.0384
LOG(O_STRONGFV+O_SLIGHTFV+1)	0.027277	0.009282	2.938598	0.0034
LOG(DISP_SPEND(-3)+1)	0.005074	0.002074	2.446742	0.0147
LOG(EMAIL_DIRE(-4)+1)	0.004605	0.003108	1.481605	0.1391
LOG(CLICK_GOOG(-1)+1)	0.008355	0.002984	2.799882	0.0053
LOG(TRIALS_QUN+1)	0.115813	0.020992	5.516966	0.0000
LOG(AVG_EXRATE)	1.120003	0.216736	5.167590	0.0000
LOG(CLOSESTOCK+1)	0.386930	0.063674	6.076777	0.0000

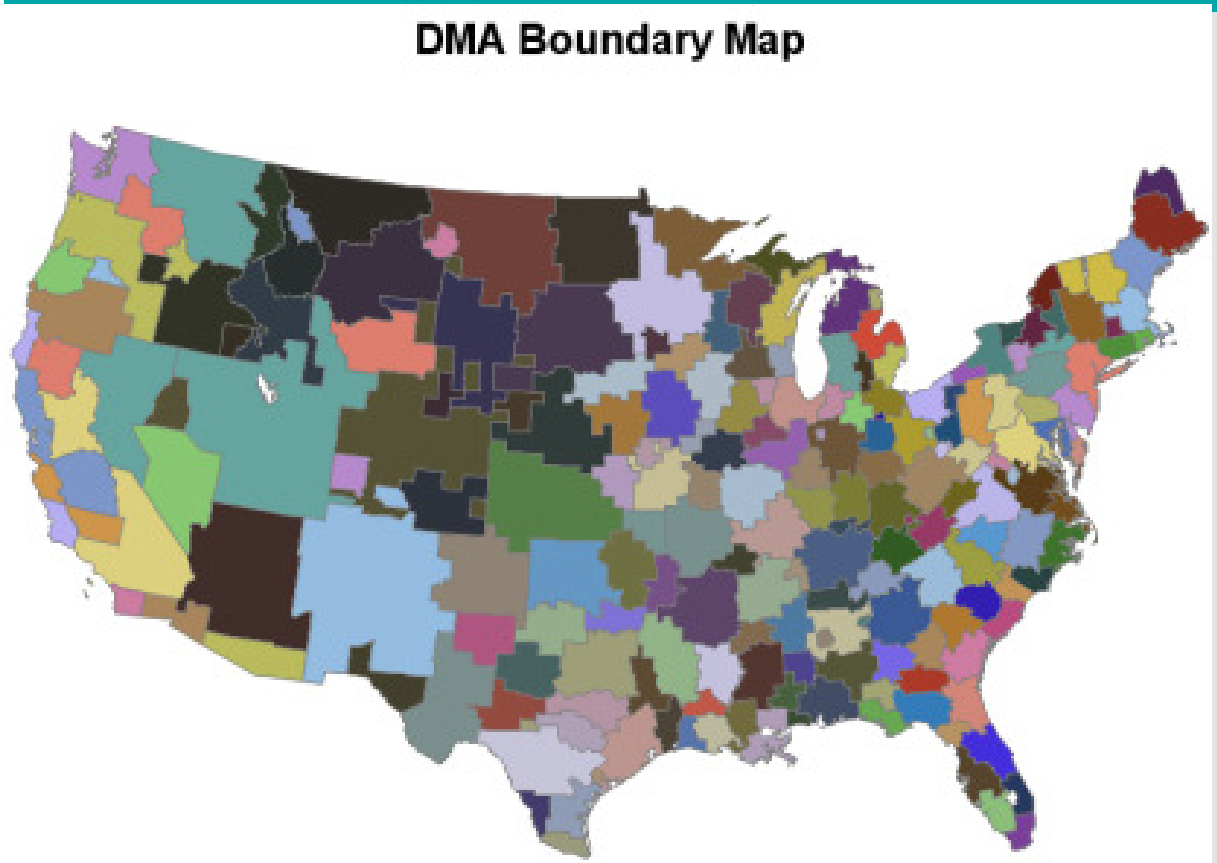
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.795063	0.606433	2.960034	0.0032
C(2)	0.411358	0.031245	13.16558	0.0000
C(3)	-0.998070	0.133955	-7.450769	0.0000
C(4)	0.086044	0.020100	4.280843	0.0000
C(5)	-0.606070	0.096889	-6.255279	0.0000
C(6)	-0.406790	0.098392	-4.134375	0.0000
C(7)	0.353382	0.097319	3.631175	0.0003
C(8)	-0.286982	0.100499	-2.855556	0.0045
C(9)	0.144991	0.034232	4.235475	0.0000
C(10)	-0.010716	0.031422	-0.341032	0.7332
C(11)	0.082793	0.031695	2.612185	0.0093
C(12)	-0.017799	0.031659	-0.562192	0.5742
C(13)	0.009877	0.032654	0.302482	0.7624
C(14)	-0.021096	0.035899	-0.587657	0.5570
C(15)	-0.037336	0.033121	-1.127258	0.2602
C(16)	-0.051434	0.032318	-1.591502	0.1121
C(17)	-0.018745	0.032728	-0.572732	0.5671
C(18)	0.130375	0.037506	3.476104	0.0006
C(19)	-0.057581	0.035689	-1.613407	0.1073
C(20)	-0.454027	0.069646	-6.519041	0.0000
C(21)	-0.461600	0.090634	-5.093014	0.0000
C(22)	-0.173996	0.075795	-2.295609	0.0221
C(23)	-0.152578	0.070989	-2.149320	0.0321
C(24)	-0.210900	0.086137	-2.448408	0.0147
C(25)	0.029704	0.010043	2.957655	0.0032
C(26)	0.004821	0.002251	2.141985	0.0327
C(27)	0.004250	0.003372	1.260550	0.2080
C(28)	0.007673	0.003230	2.375561	0.0179
C(29)	0.141108	0.022728	6.208503	0.0000
C(30)	1.043132	0.206418	5.053500	0.0000
C(31)	0.394039	0.073700	5.346492	0.0000
C(32)	-0.494569	0.738763	-0.669456	0.5035

Determinant residual covariance	0.032323

Equation: LOG(FULL_REV) = C(1)*(PROD_COUNTRY="ACR_US") + C(2)
*LOG(FULL_REV(-1)) + C(3)*D_JULY1407 + C(4)*NVER_ACR89S +
C(5)*NVER_ACR89W(5) + C(6)*NVER_ACR89W(4) + C(7)
*NVER_ACR89W + C(8)*NVER_ACR89W(-1) + C(9)*M01 + C(10)*M02
+ C(11)*M03 + C(12)*M04 + C(13)*M05 + C(14)*M06 + C(15)*M08 +
C(16)*M09 + C(17)*M10 + C(18)*M11 + C(19)*M12 + C(20)
*H_CHRISMAS + C(21)*H_USTHANKS + C(22)*H_MLKING + C(23)
*H_VET_REM + C(24)*H_GOODFRI + C(25)*LOG(O_STRONGFV

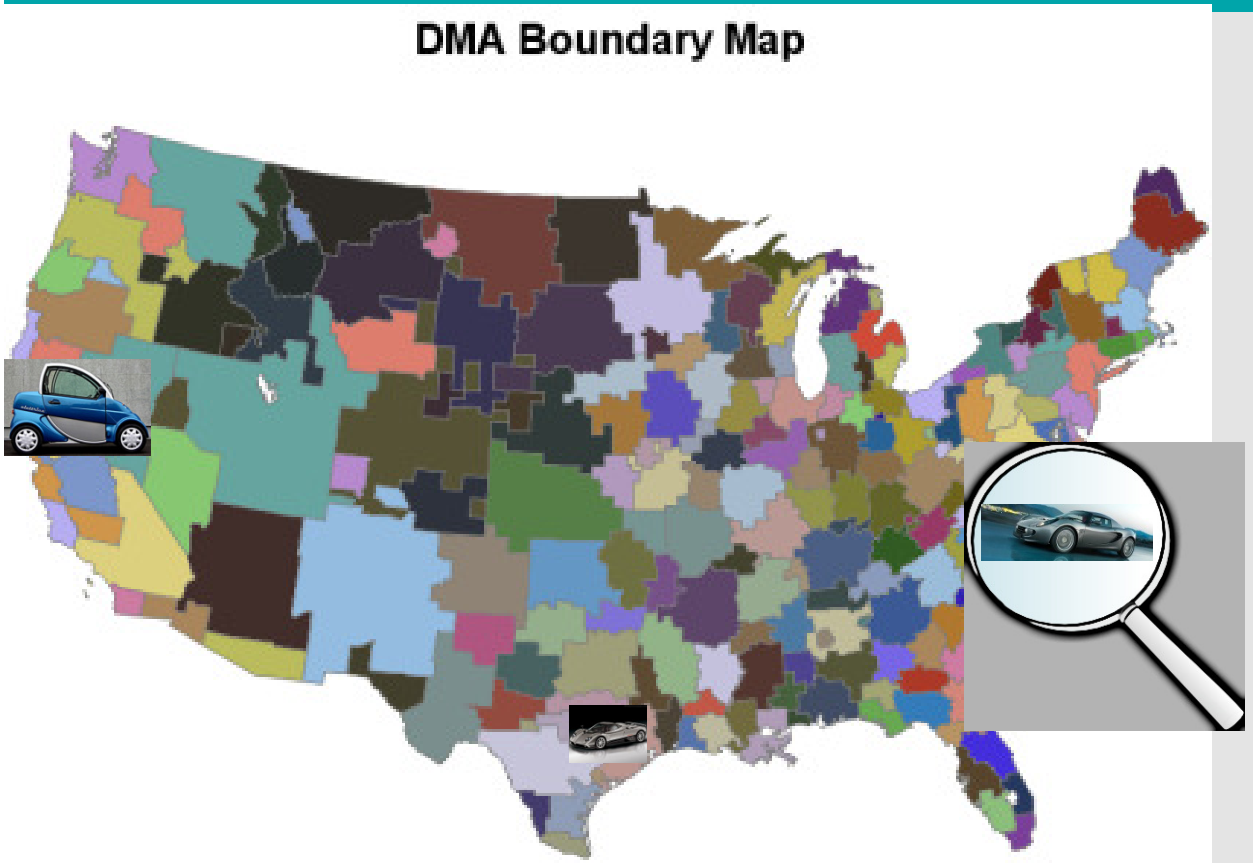
MarketShare

System of Equations = DMA



MarketShare

System of Equations = DMA x Product

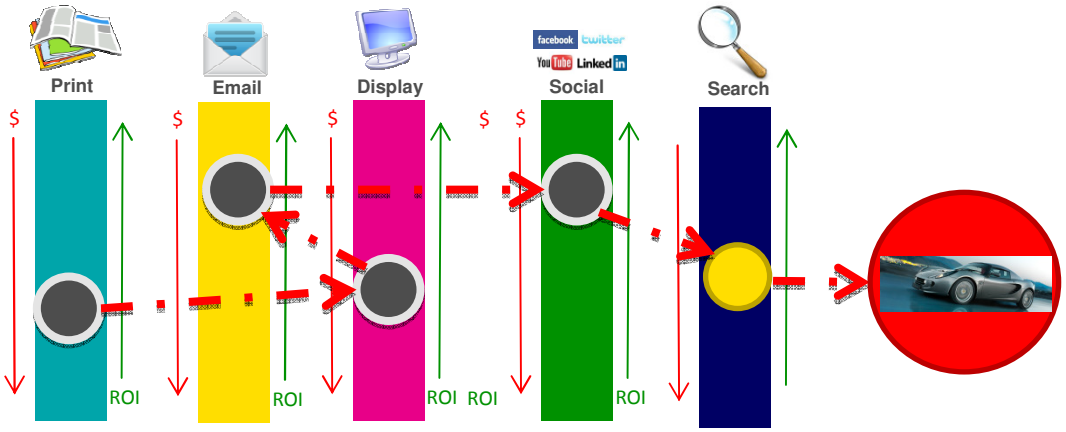


Product x DMA

DV : DLOG(GQV_BRND_CRD)
Date: 10/12/10 Time: 04:21
SAMPLE : 1/07/2007 4/25/2010 IF X_PID_KEEP AND X_ACTIVE=""Sales
PERIODS : 166
C_SECTION : 45
OBSERVATION :7470
C,-0.001662,0.003390,-0.490416
DLOG(DM_ACQ_PH_QP(-2)+1),0.003098,0.002026,1.529200
DLOG(MC2_OOH_CITI_SPD(-4)+1),0.009889,0.003011,3.284126
DLOG(MC2_TV_CITI_GRP(-4)+(MC2_TV_CITI_GRP(-4)=0)),0.003607,0.002298,1.569751
HOL_LABOR(1),-0.111736,0.066085,-1.690803
HOL_THANKS,0.079682,0.022311,3.571398
AR(1),-0.367022,0.077109,-4.759810
R-squared,0.224122
Adjusted R-squared,0.218893
F-statistic,42.86145
Durbin-Watson stat,2.058506
Prob(F-statistic),0.000000

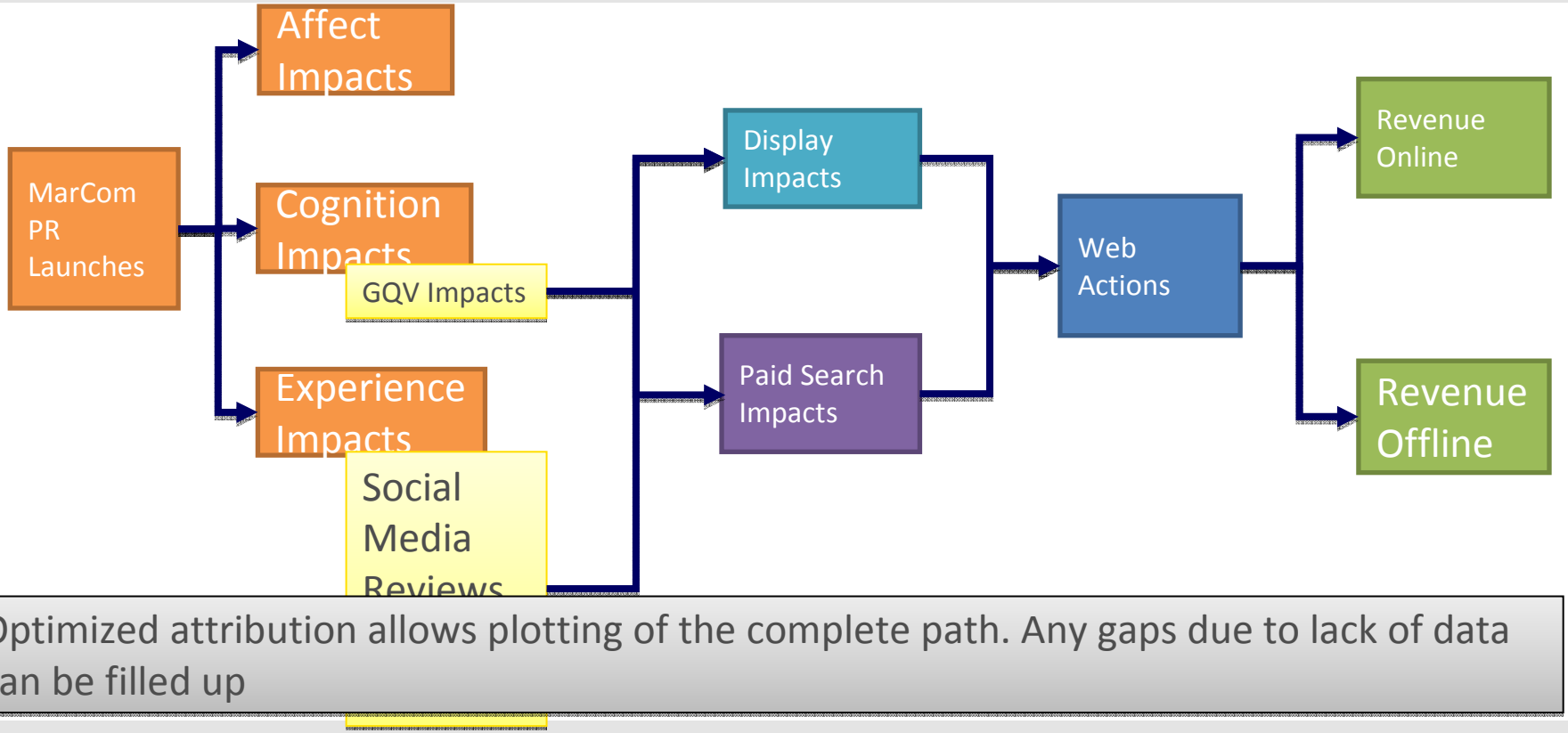
MarketShare

System of Equations = Product x DMA x Media Channels



MarketShare

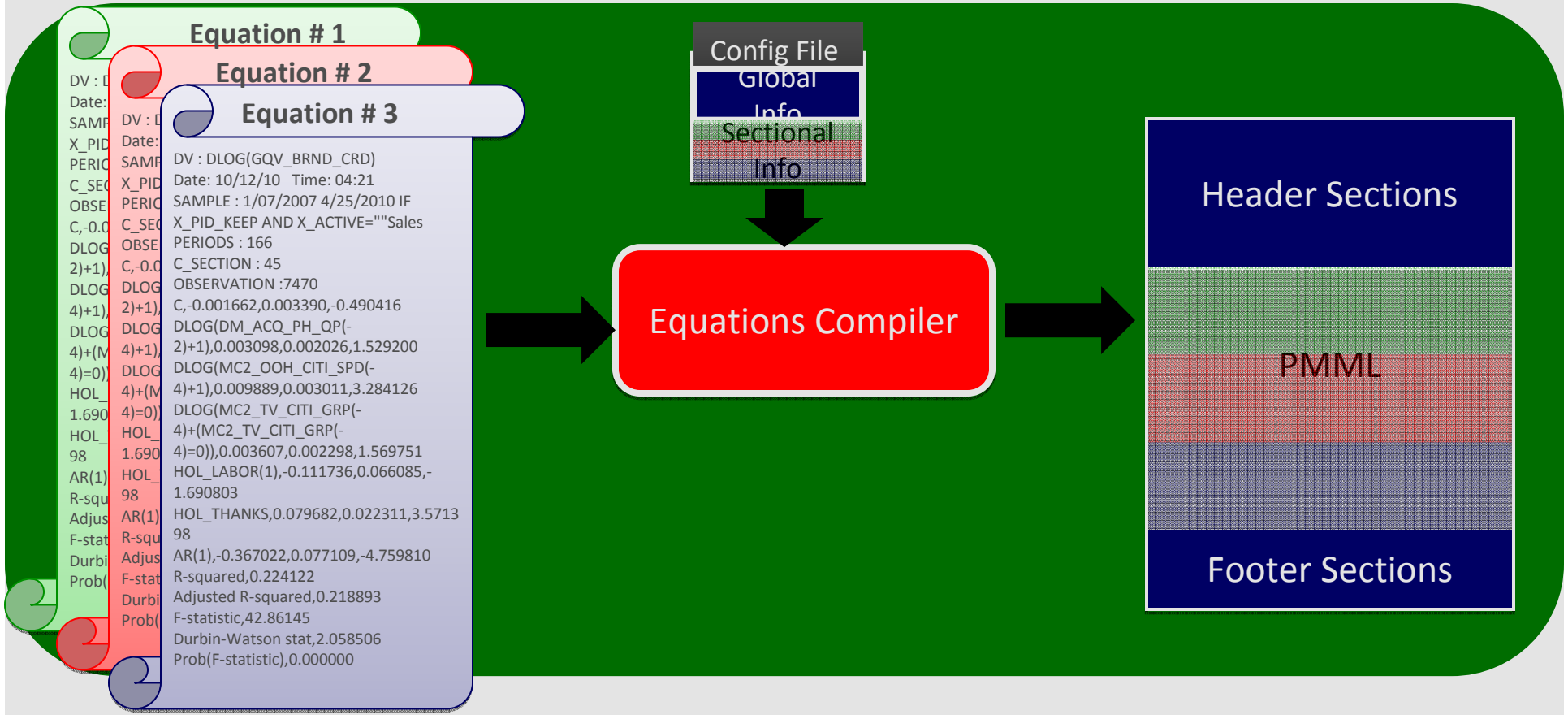
Purchase Paths are complex



Optimized attribution allows plotting of the complete path. Any gaps due to lack of data can be filled up

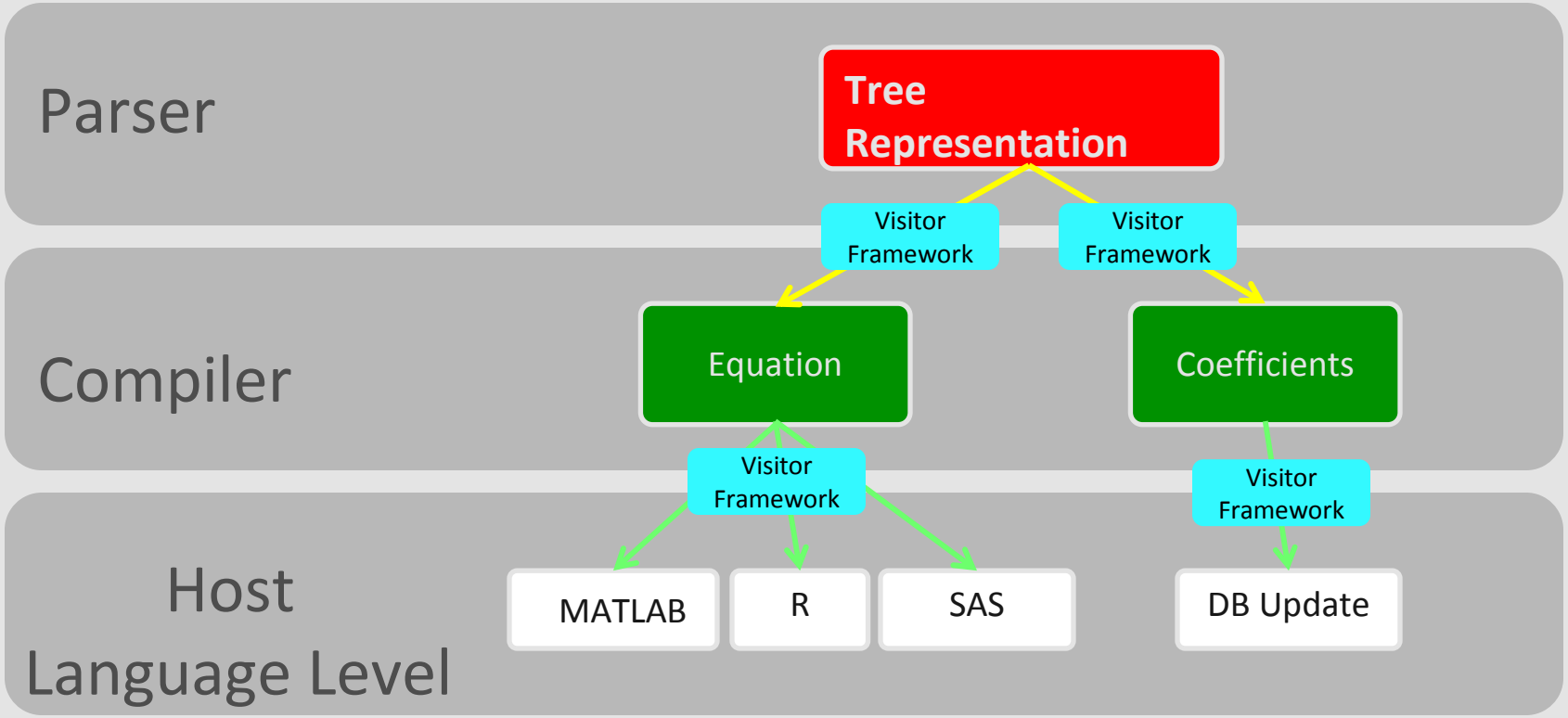
MarketShare

Equation Compiler maintains a System of Equations



MarketShare

Anatomy of an Equation

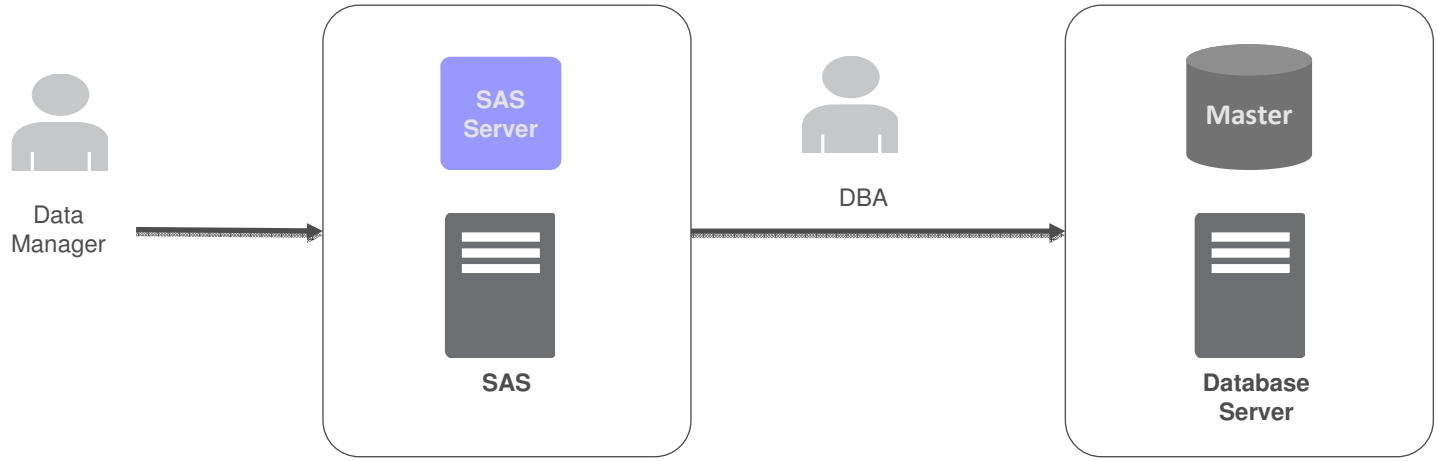


MarketShare

Elastic Modeling

MarketShare

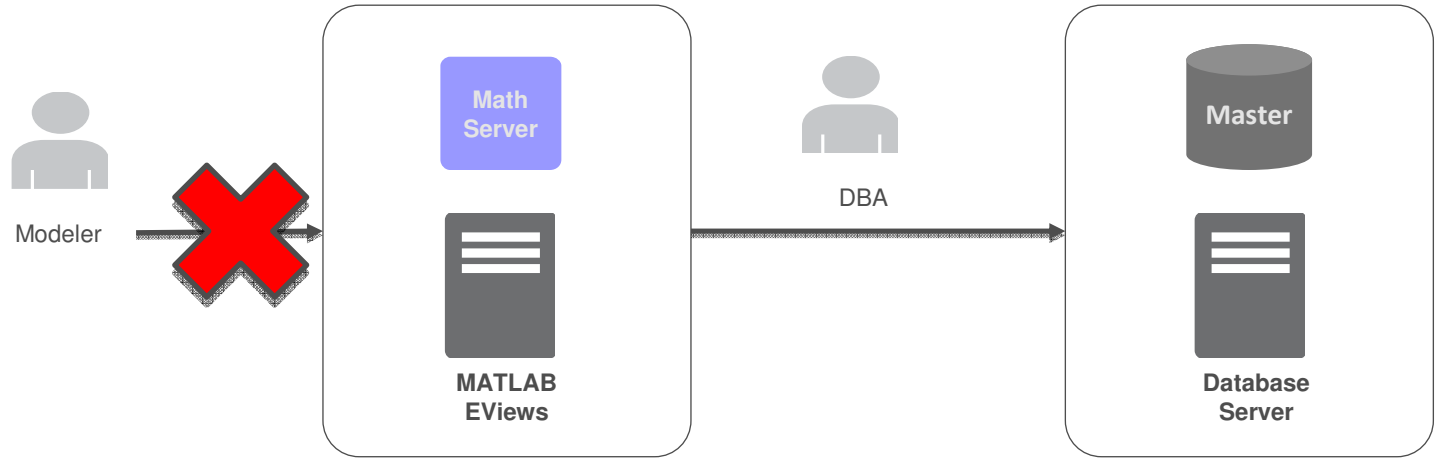
Traditional Data Preparation



Data Transformations

MarketShare

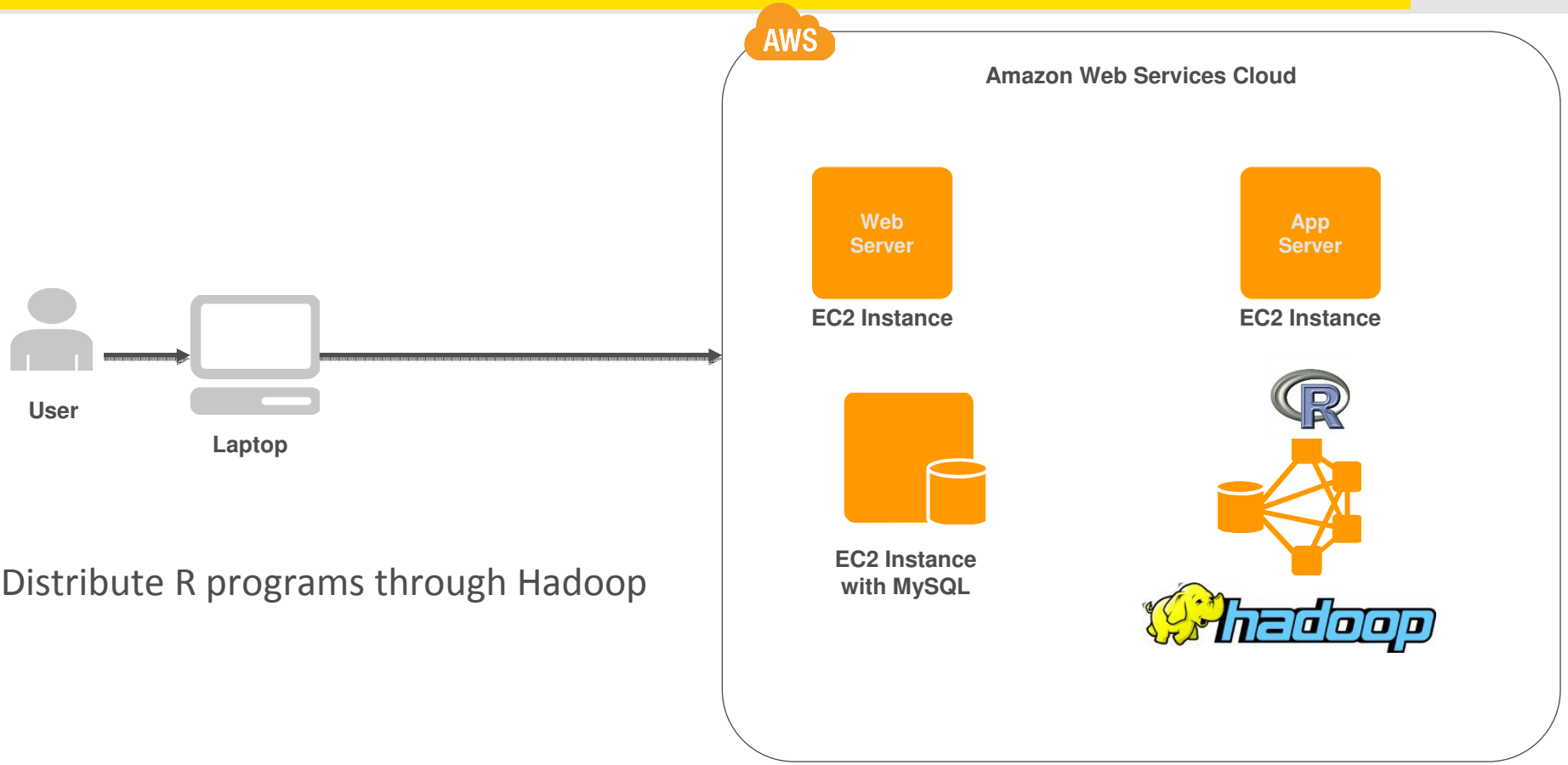
Traditional Modeling architecture



Eliminate accessibility restrictions

MarketShare

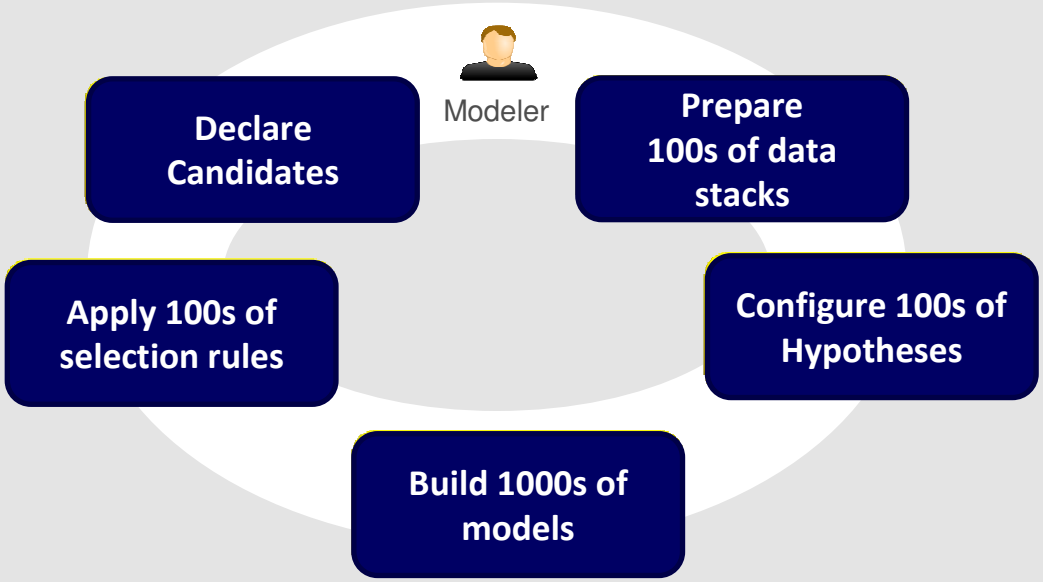
Distributed, Cloud based Modeling



Distribute R programs through Hadoop

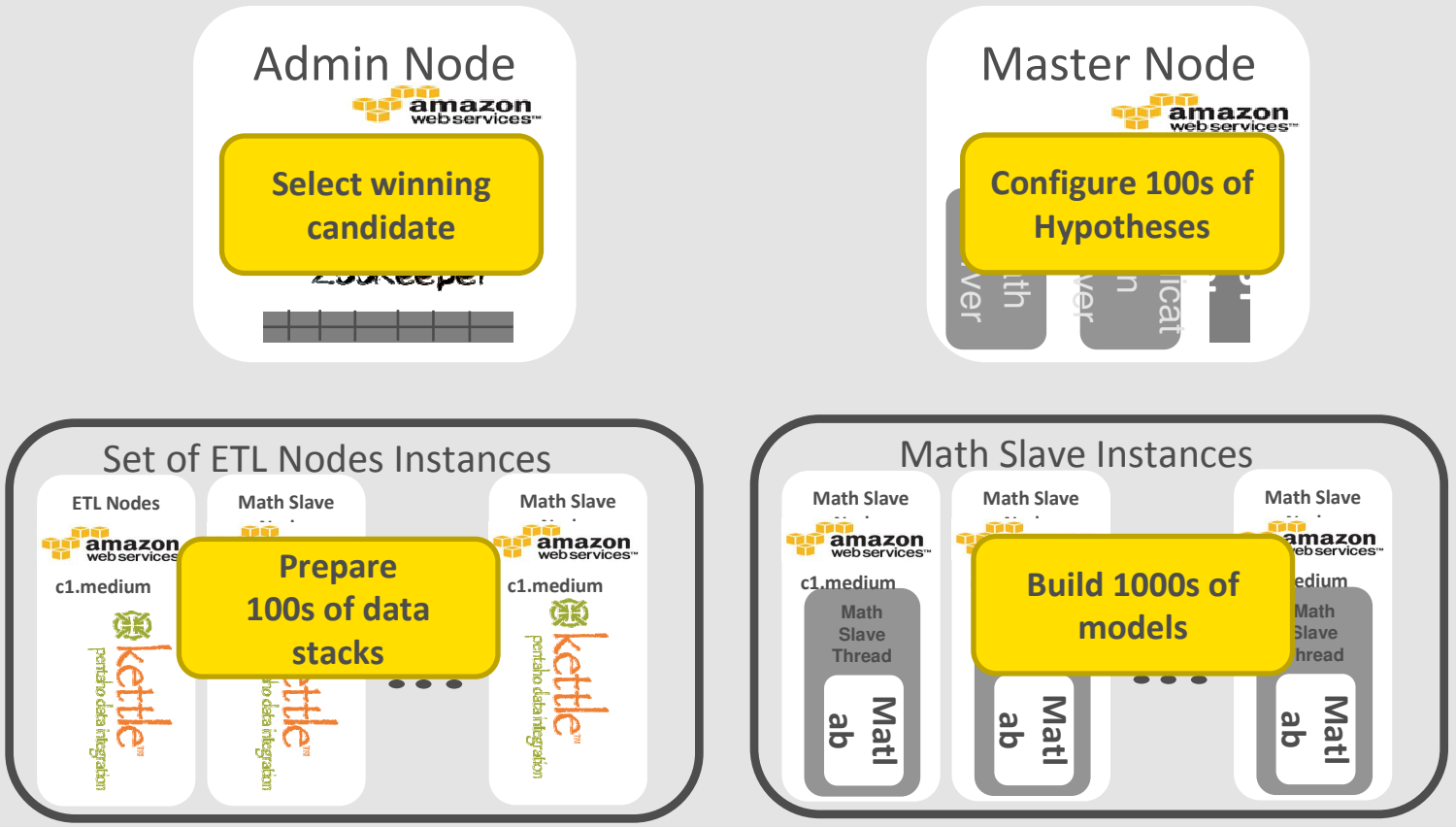
MarketShare

Moving modeling to the cloud

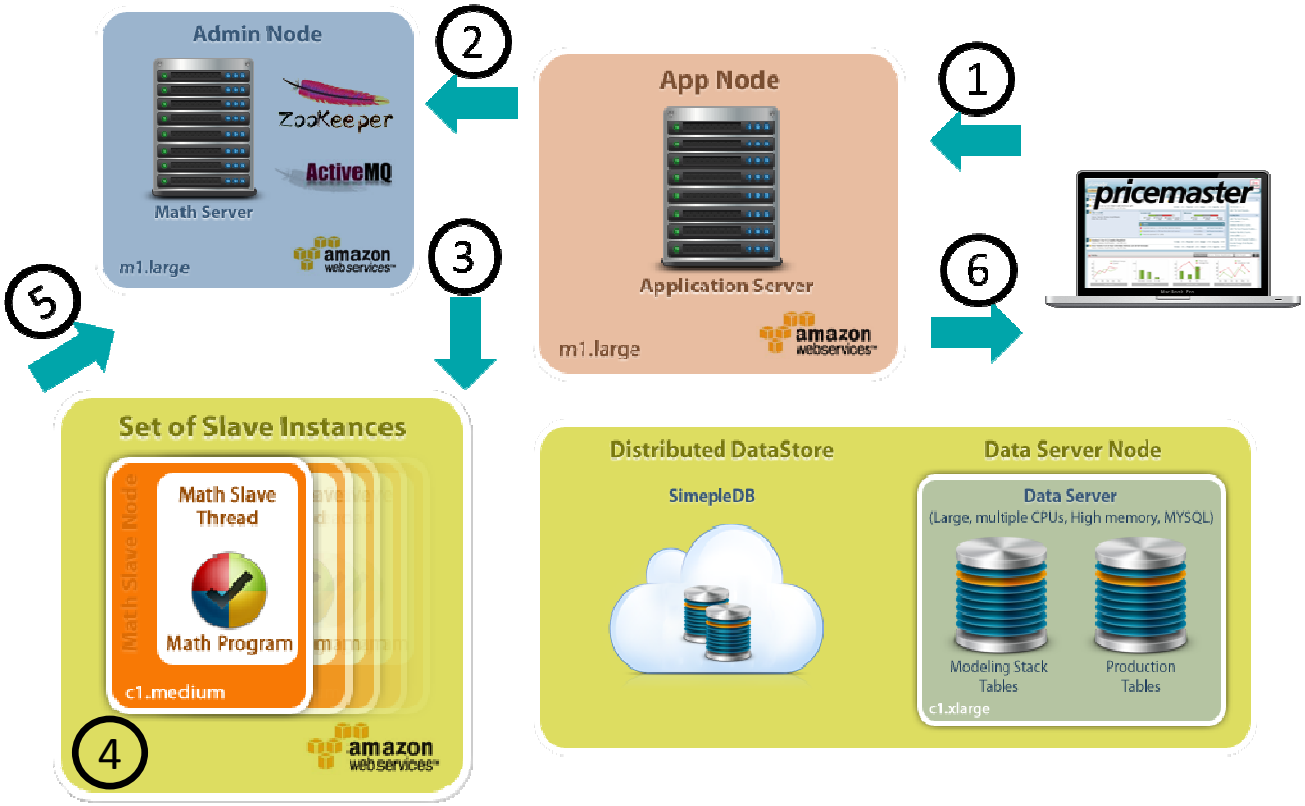


MarketShare

Underlying architecture



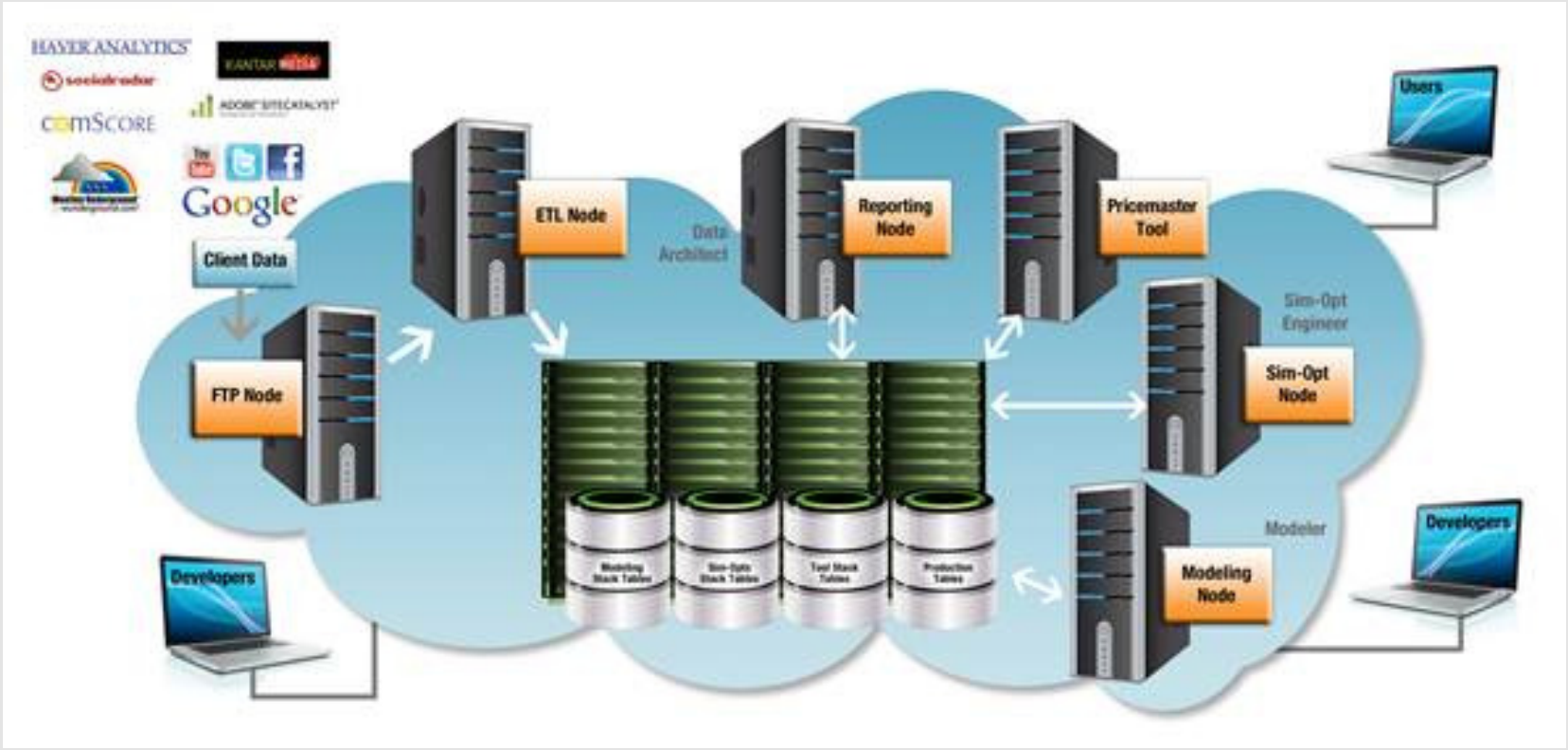
Distributed data flow enables unlimited scalability



1. User creates/refreshes a scenario
2. Application server creates a request and queues it with the messaging server
3. Math Slave reads the response
4. Math Slave calls Math Program programs and process the input and output
5. Math Slave queues response back with zookeeper
6. Application Server picks response and responds backs to UI

MarketShare

The big picture



MarketShare

Next Steps

- Lots of challenges in cloud + modeling
- Collaboration opportunities
- We are hiring!