

**Veriambarı projeleri artık hızlı,
kolay ve ekonomik: Netezza**

Feriye - İstanbul
7 Haziran 2012, Perşembe

İş Analitiği'ne Netezza ile Yüksek Performans Katın

Umut ŞATIR
İleri Analitik Çözüm Mimarı

Netezza and IBM Business Analytics

Baştan sona bir İş Analitiği çözümü

- Performans
- Kolaylık
- En Kapsamlı Analitik
- Tek IBM Çözümü



“Analitik” Derken Ne Kastediyoruz?

- **İşiniz ile ilgili sorular sormak**
- **Soruları veri ile yanıtlamak**

Analitik Türleri

İş Zekası

- Gerçekler
- Geçmiş ve Bugüne Yönelik Ölçümler

İleri Analitik

- İzler, Eğilimler, İlişkiler
- Gelecek davranışlarını tahmin etmek

İş Zekası

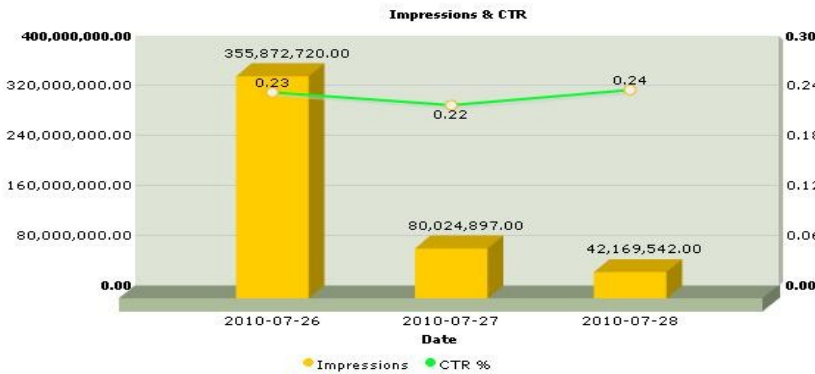
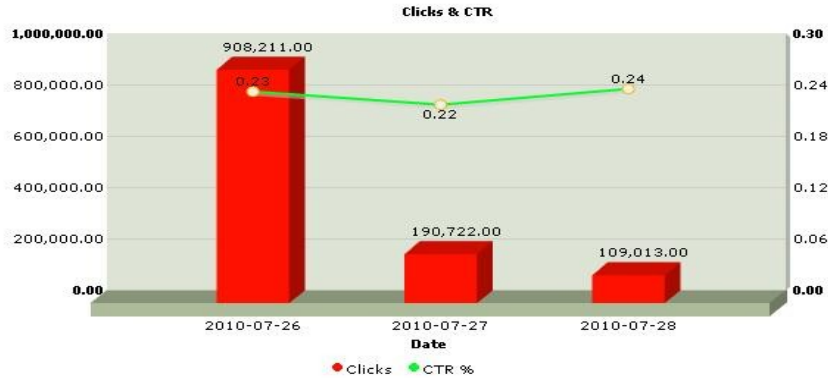
Date Range

Start date: 7/26/2010
End date: 7/28/2010

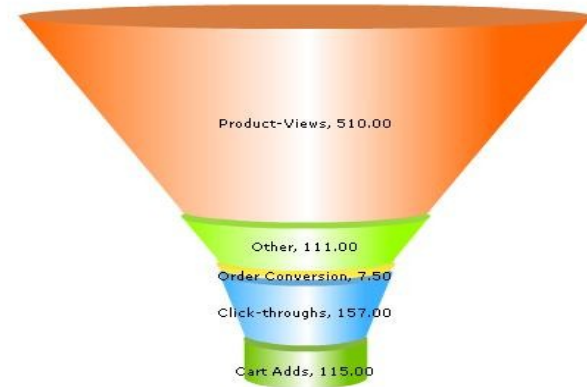
Submit

Cloudera Hadoop Actions

Select Oozie Action



1098933 Total Clickthroughs / 435898130 Total Impressions = 0.003 Average Clickthrough Rate

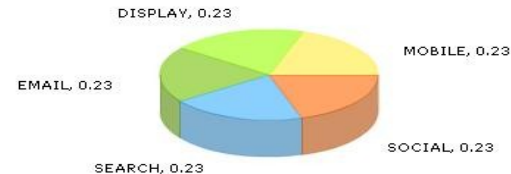


CTR by Devices



Sitemizi kaç kişi ziyaret etmiş?

CTR by Channel



İş Zekası

Healthcare Cost Spikes Dashboard

Healthcare Cost Spikes Analysis: Physician Clustering

Number of Clusters:

* 10

Select the desired number of clusters (from 4 to 16).

Sensitivity Threshold:

* 1

Default of 1.0, minimum of 0. Lower value yields more cost spike areas.

Finish

Display Clusters: Normal

Finish

Cost Spike Flag	Cluster ID	# Physician	Avg. Members per Physician	% Chg Total Members	% Chg Total Visits	% Chg Total Claims	% Chg Total Units	% Chg Total Billed	% Chg Total Allowed	% Chg Avg Billed per Member	% Chg Avg Allowed per Member	% Chg Avg Units per Member	% Chg Avg Billed per Unit	% Chg Avg Allowed per Unit	% Chg Avg Service Mix
High Cost Spike	7	1,053	1.41	9.02	7.69	13.20	47.10	47.02	37.60	58.05	50.51	56.89	59.88	75.81	99.41
	2	195	4.18	122.56	123.08	126.15	209.74	265.47	307.65	64.36	87.66	47.18	33.53	54.05	56.78
	3	173	3.81	220.23	281.50	293.64	795.95	852.96	834.15	259.02	215.45	241.62	70.48	66.90	86.85
	4	171	1.67	25.73	48.54	57.31	692.40	2,404.50	2,326.06	2,018.76	2,246.69	602.34	1,202.12	1,171.86	94.20
High Cost Spike - Total		1,592 (46.20%)													
Normal - Total		1,854 (53.80%)													
Overall - Total		3,446													

Hangi medikal uygulamalar daha yüksek sağlık sigortası taleplerine sebep oluyor?

İleri Analitik

Simulation Parameters	
Calculation Date	Fri 01/02/04
Risk Free Rate	2.6250%
Volatility Premium for Equity Options	37.5%
Number of Simulations	5,000

Calculate Correlation

Simulate

Portfolio	Longs		Shorts		Total # of Positions	Net Notional
	# Of Positions	Notional	# Of Positions	Notional		
Equity	4,328	2,058,853,578	5,247	2,538,244,351	9,575	-400,390,773
Equity Options	5,715	20,308,769	5,721	20,523,604	11,436	-214,835
Bond Portfolio	5,551	300,477,534	0	0	5,551	300,477,534
All Portfolios	15,594	2,379,639,881	10,968	2,558,767,955	26,562	-100,131,074

Simulated Future Values

Statistics	1 Day	2 Day	3 Day	4 Day	5 Day	1 Month	2 Month	3 Month	6 Month	9 Month	12 Month
1st Percentile	\$ (237,781,129)	\$ (246,867,833)	\$ (254,105,802)	\$ (257,951,400)	\$ (263,302,294)	\$ (238,622,863)	\$ (397,452,204)	\$ (444,434,569)	\$ (610,810,774)	\$ (774,922,228)	\$ (909,871,051)
5th Percentile	\$ (230,712,107)	\$ (237,117,512)	\$ (242,473,452)	\$ (245,978,413)	\$ (250,183,495)	\$ (297,780,927)	\$ (341,038,055)	\$ (383,640,887)	\$ (498,194,527)	\$ (596,333,143)	\$ (696,558,427)
10th Percentile	\$ (228,584,282)	\$ (232,801,997)	\$ (236,808,592)	\$ (238,223,344)	\$ (242,049,912)	\$ (280,294,073)	\$ (314,817,853)	\$ (346,105,246)	\$ (434,628,093)	\$ (515,503,325)	\$ (598,887,430)
25th Percentile	\$ (221,446,072)	\$ (224,654,455)	\$ (227,214,870)	\$ (228,864,011)	\$ (230,583,767)	\$ (253,185,311)	\$ (276,784,532)	\$ (294,089,165)	\$ (346,885,419)	\$ (391,113,939)	\$ (433,078,598)
50th Percentile	\$ (216,583,791)	\$ (216,824,349)	\$ (217,428,602)	\$ (217,773,974)	\$ (218,187,827)	\$ (224,748,154)	\$ (233,482,222)	\$ (242,328,086)	\$ (284,145,104)	\$ (284,040,533)	\$ (284,583,712)
Mean	\$ (216,587,317)	\$ (217,262,944)	\$ (217,728,832)	\$ (218,137,438)	\$ (218,430,043)	\$ (226,945,164)	\$ (238,482,795)	\$ (248,443,043)	\$ (275,754,315)	\$ (298,446,582)	\$ (309,891,590)
Std Dev	\$ 8,307,111	\$ 12,033,743	\$ 14,618,148	\$ 16,620,826	\$ 18,714,699	\$ 40,527,748	\$ 59,118,518	\$ 74,848,104	\$ 119,648,593	\$ 169,704,883	\$ 222,504,244
75th Percentile	\$ (211,659,591)	\$ (209,388,319)	\$ (208,048,523)	\$ (206,735,142)	\$ (205,560,420)	\$ (198,773,151)	\$ (197,053,863)	\$ (190,040,395)	\$ (194,333,417)	\$ (188,062,878)	\$ (185,022,797)
90th Percentile	\$ (206,045,518)	\$ (202,540,490)	\$ (199,247,166)	\$ (197,759,534)	\$ (195,028,915)	\$ (177,160,338)	\$ (165,301,545)	\$ (150,843,978)	\$ (135,268,227)	\$ (120,163,290)	\$ (81,316,788)
95th Percentile	\$ (204,194,520)	\$ (198,188,895)	\$ (194,514,501)	\$ (191,855,163)	\$ (188,346,357)	\$ (163,901,150)	\$ (148,486,803)	\$ (130,975,408)	\$ (102,254,453)	\$ (52,492,001)	\$ 16,792,555
99th Percentile	\$ (197,486,814)	\$ (188,088,316)	\$ (185,769,555)	\$ (179,754,354)	\$ (177,810,478)	\$ (138,887,869)	\$ (116,502,893)	\$ (85,073,815)	\$ (40,896,049)	\$ 68,598,763	\$ 192,153,373

Statistics	1 Day	2 Day	3 Day
1st Percentile	\$ (57,850,955)	\$ (66,538,758)	\$ (73,904,728)
5th Percentile	\$ (51,581,033)	\$ (59,886,437)	\$ (62,342,977)
10th Percentile	\$ (46,753,188)	\$ (52,470,923)	\$ (56,477,487)
25th Percentile	\$ (41,214,998)	\$ (44,533,381)	\$ (47,088,795)
50th Percentile	\$ (36,852,717)	\$ (36,893,275)	\$ (37,297,527)
Mean	\$ (36,776,243)	\$ (37,131,830)	\$ (37,598,758)
Std Dev	\$ 8,307,111	\$ 12,033,743	\$ 14,618,148
75th Percentile	\$ (31,538,487)	\$ (29,287,245)	\$ (27,435,416)
90th Percentile	\$ (26,714,444)	\$ (22,489,416)	\$ (19,116,092)
95th Percentile	\$ (24,063,446)	\$ (18,057,821)	\$ (14,724,099)
99th Percentile	\$ (17,325,740)	\$ (8,958,242)	\$ (5,838,481)

Pazar kötüye giderse ürün portföyümüze ne olur?

Click to add title

Click to add text

İleri Analitik

Bu müşteriye en etkileyecek öneri nedir?

İleri Analitik



**Bu yıl Tip 2 Diyabete
yakalanma olasılığı
olan hastalar**

Business Intelligence**Business Analytics**

Business Intelligence

Predictive & Advanced Analytics

Financial Performance & Strategy Management

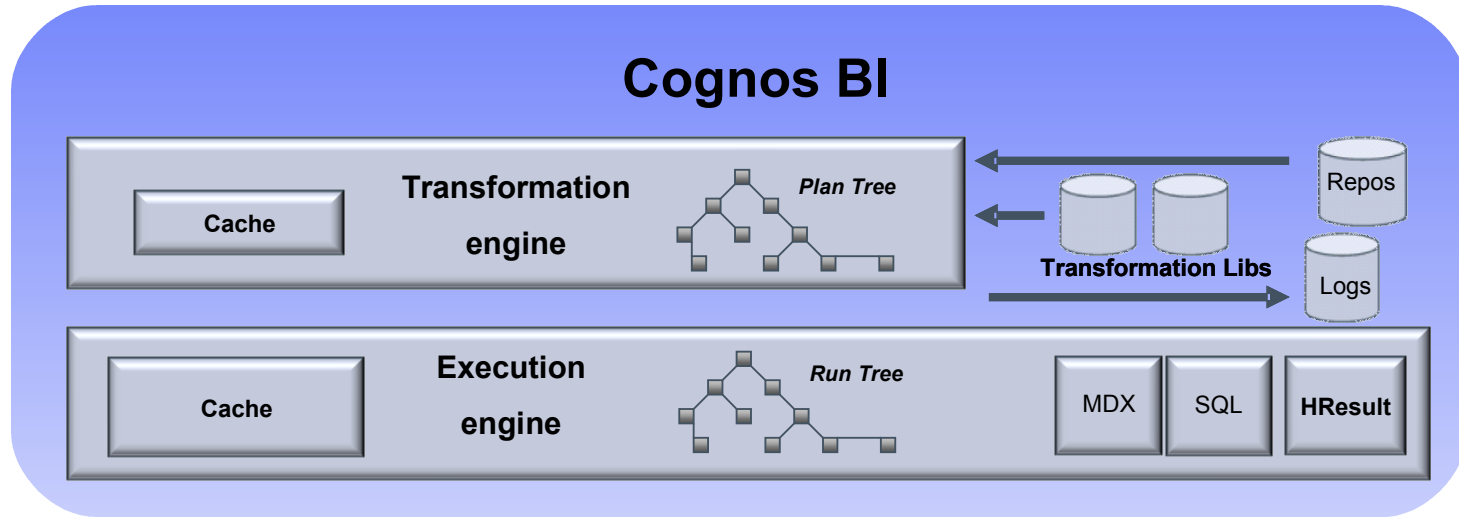
Governance, Risk & Compliance

Analytic Applications

**COGNOS®**

- Her karar verici için zengin BI özellikleri
- Sorgu, raporlama, analiz, scorecard ve dashboard
 - İstenilen yerde, zamanda, şekilde
- BI için en devrimci yol
- İş birliği özellikleri ile zenginleştirilmiş bütünleşik çalışma alanı
- Her zaman dilimine ait bilgi: geçmiş, şimdi & gelecek
 - Görüntülemeyen ileri analitiğe kolay ilerleme
- Kritik iş sorularına yanıtlar
 - “Ne durumdayım?”, “Neden”, “Ne yapmalıyım?”
 - Organizasyon çapında çözüm

Cognos BI Optimizations for Netezza



En Hızlı Sorgu Yanıtı

- Geliştirilmiş SQL
- Dinamik Sorgu – In memory cache
- Optimize Aggregation İşlemleri
- Multi-pass SQL özelliği ile multi-parallel işleme



Business Intelligence

Predictive & Advanced Analytics

Financial Performance & Strategy Management

Governance, Risk & Compliance

Analytic Applications

Tam Kapsamlı İleri Analitik

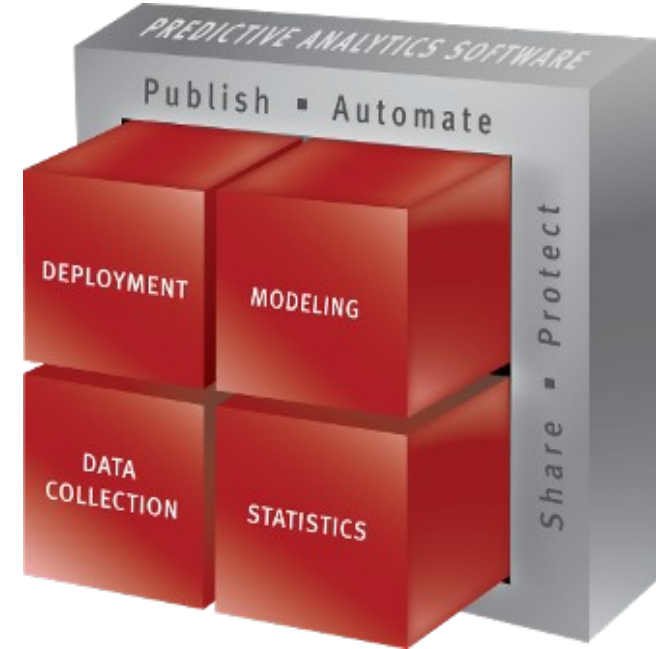
Veri toplama, istatistik, veri madenciliği, öngörülse modelleme, dağıtım...

Tahminleri iş kullanıcılarına dağıtmak

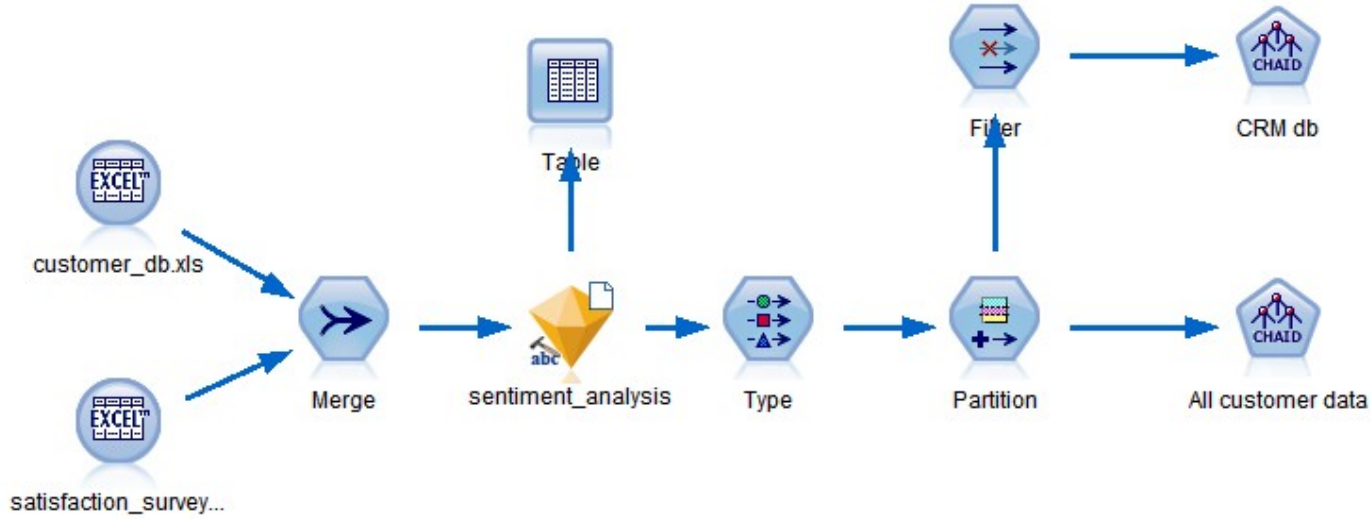
Karar Yönetimi

Daha iyi iş sonuçları için

Karlı müşterileri kazan ve elde tut
Hilekarlığı tespit et ve engelle
Operasyonel etkinliği geliştir



IBM SPSS Modeller Nedir?



IBM SPSS Modeller, görsel ve kullanıcı dostu arayüzü ile hızlı bir şekilde öngörülse modellerin oluşturulmasına imkan sağlayan, güçlü ve esnek bir veri madenciliği çözümüdür.

IBM SPSS Modeller

- **Kullanım Kolaylığı** – Görsel, sınıfında lider öngörüselleme
- **Esnek** – Farklı ortamlardaki verilere erişebilir
- **Otomasyon** – Deneyimli veri madencilerinden iş analistlerine
- **Güçlü** – Eksiksiz veri madenciliği yetenekleri
- **Dağıtım** – Hızlı model geliştirme & dağıtım
- **Ölçeklenebilir** – Veri tabanında veri işleme ve optimizasyon

The screenshot displays the IBM SPSS Modeller software interface. The main window shows a workflow diagram with nodes for 'CREDIT', 'Type', 'Select', 'Auto Data Prep', 'Type', and 'men'. The 'Auto Data Prep' node is highlighted, and a 'Details for Current Balance' window is open, showing 'Original' and 'Transformed' distribution plots. The 'Transformed' plot shows a normal distribution with a skewness of 0.94 and a standard deviation of 1.00. The 'Execution Feedback' window is also visible, indicating that the execution is complete. The interface includes a menu bar, a toolbar, and a status bar at the bottom.



Dahili İleri Analitik Kütüphanesi

Analitik Görevler

Fonksiyon Türleri

Veri Analizi ve Keşif

Veri Profilleme
Tanımlayıcı İstatistikler
Temel Bileşen Analizi
Bayes Ağları

Veri Hazırlama

Standardizasyon
Normalleştirme
Gruplama
Kayıp Değer Analizi

Model Geliştirme

Sınıflandırma: Karar Ağaçları, Naïve Bayes
Tahminleme: Doğrusal Regresyon, Regresyon
Ağaçları
Kümeleme: K-Ortalamlar, Hiyerarşik Kümeleme
Birliktelik Tespiti: APriori, FP-Growth

Model Değerlendirme

Ortalama/Bağlı Hata İstatistikleri
Hipotez Testi
Sınıflandırma/Yanlış Sınıflandırma İstatistikleri

Model Skorlama

Model Kurma İşlevleri ile Bağıntılı

IBM SPSS Modeller & IBM Netezza Analytics

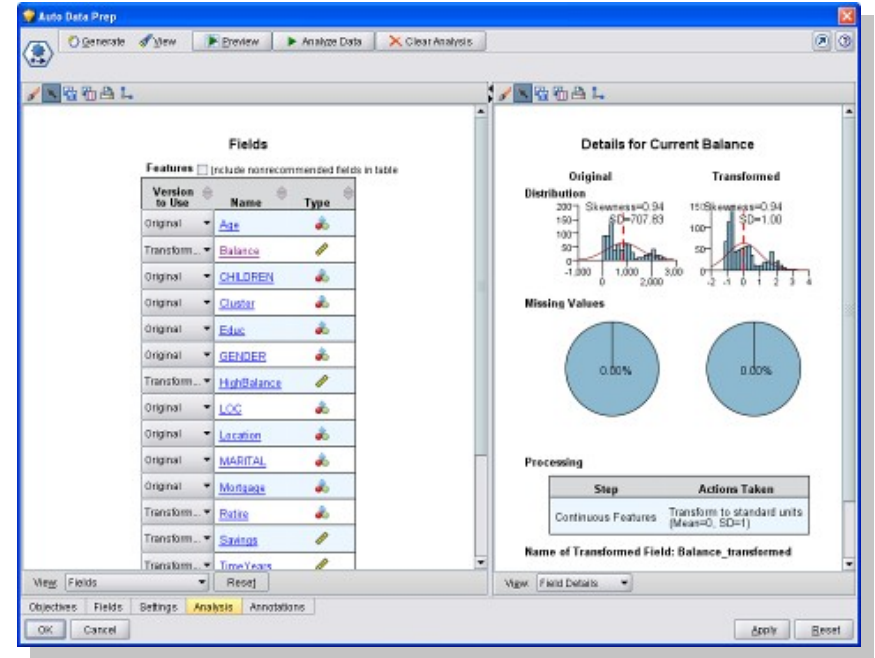
IBM SPSS öngörülse analitik konusunda sektör lideri bir ürün olarak:

Netezza müşterilerine, Netezza Analytics yeteneklerine görsel bir arayüzden ulaşma imkanı sunar.

Netezza Analitik özelliklerini, IBM SPSS Modeller'in zengin analitik kabiliyetleri ile birleştirme imkanı sunar.

Otomatik veri hazırlama, otomatik modelleme gibi özelliklere erişim imkanı sunar.

Model ve analitik varlık yönetimi özelliklerinden faydalanma imkanı sunar.





IBM SPSS Modeler & IBM Netezza

Yetkinlik	
SQL Pushback	Tier-1 seviyesinde SQL Push Back desteği. Veri madenciliği çalışması öncesi gerçekleştirilen veri hazırlama işlemlerinin veri tabanı içerisinde paralel bir mimaride çalıştırılabilmesi.
In-Database Data Mining	Kullanıma hazır, veri tabanı içi paralel çalışır veri madenciliği modelleri : <ul style="list-style-type: none">• Decision Trees• K-Means• PCA• Linear Regression• Regression Trees• Bayes Network• Naïve Bayes• K-Nearest Neighbors• Divisive Clustering• Time Series• GLM
Sampling	Örnekleme seçim işlemleri için optimize SQL
Scoring Accelerator	Veri tabanı için skorlama

IBM SPSS Modeler & IBM Netezza

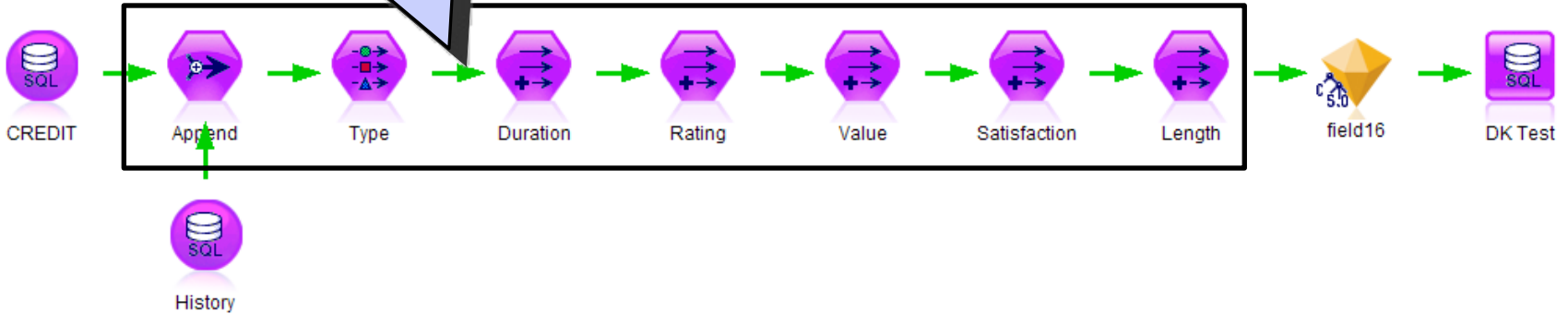
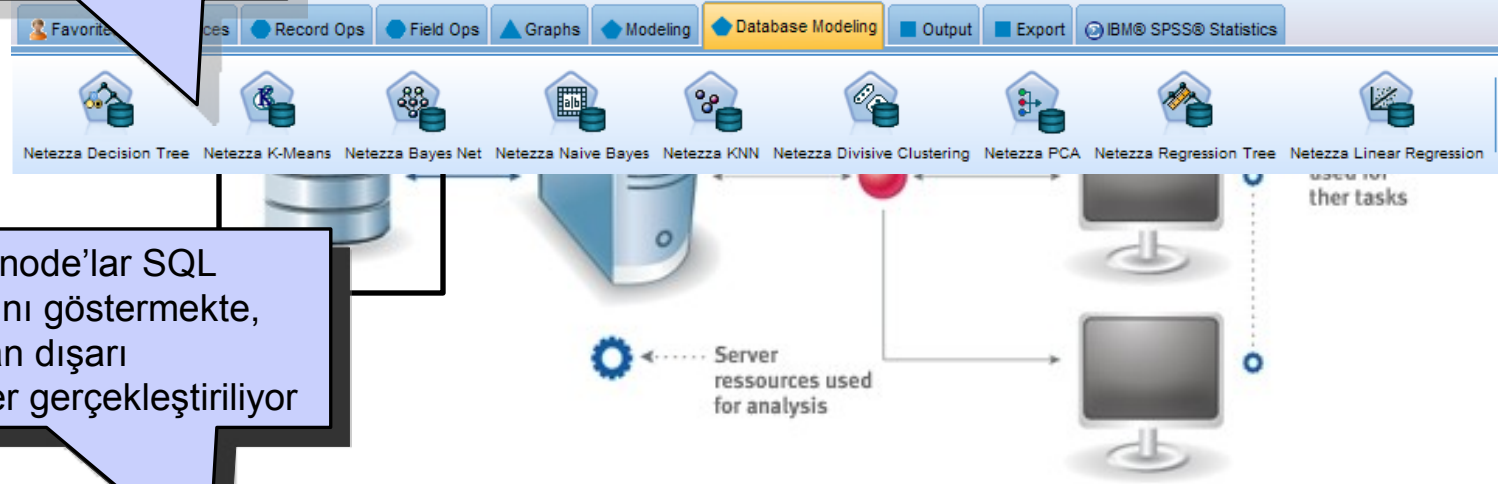
Araç kutusunda bulunan
Netezza veri madenciliği
fonksiyonları

Buradaki mor renkli node'lar SQL
Push back yapıldığını göstermekte,
veriler veritabanından dışarı
çıkarılmadan işlemler gerçekleştiriliyor

Heavy traffic limited to the
connection between IBM
SPSS Modeler Server
and the data

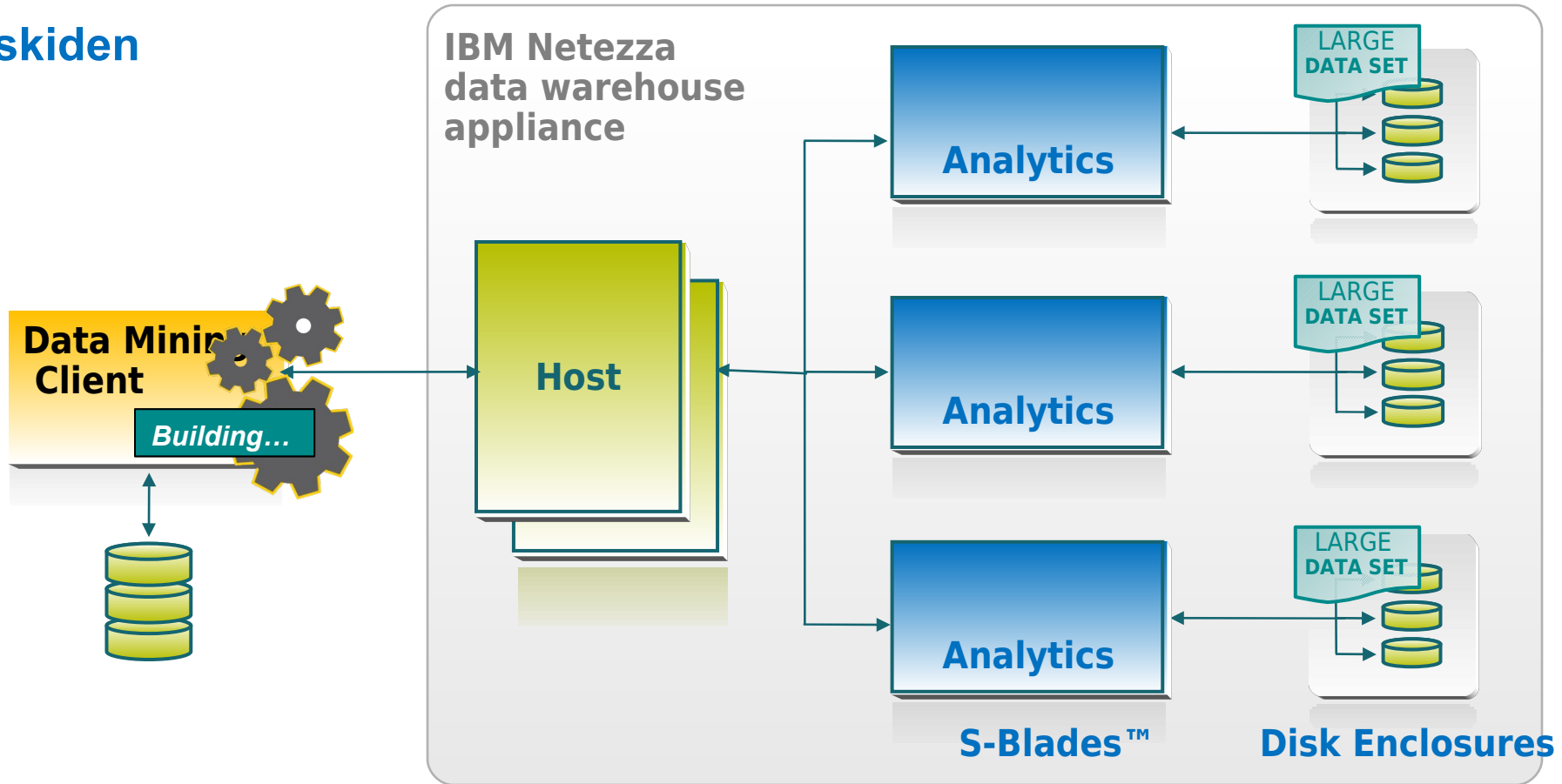
Greatly reduced network
traffic - now consisting
only of commands and
results

IBM SPSS Modeler desktop installations



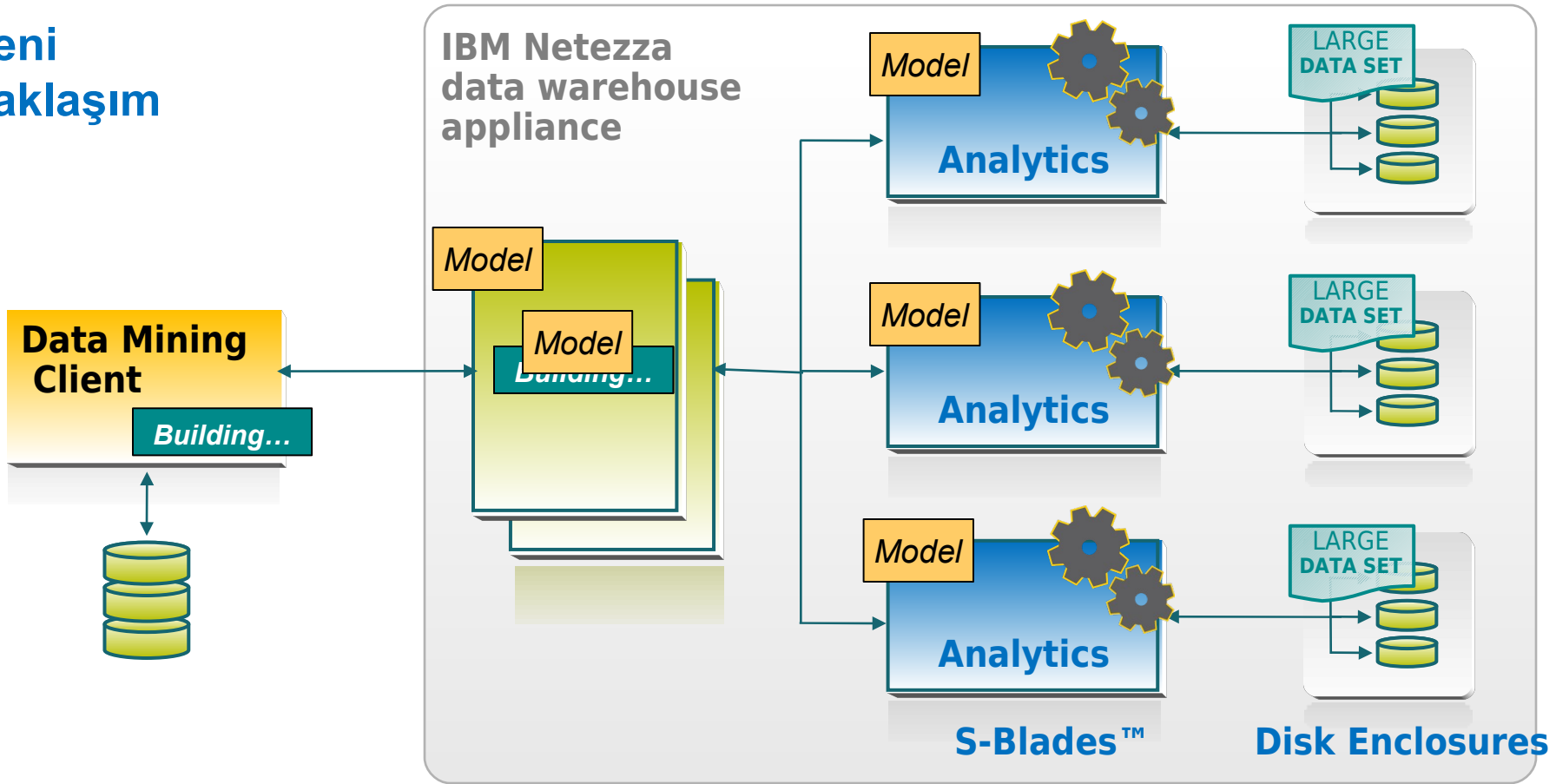
Büyük Verilerle Analitik Modeller Oluşturma

Eskiden



Veri Tabanı içi Analitik Yetkinlikler

Yeni Yaklaşım



Catalina Marketing

- Tüketicilere, alışveriş alışkanlıklarına göre en uygun tekliflerin yapılması
- Haftalık 300 milyon işlem
- 3 yıllık tüketici bilgisi saklanıyor
- Netezza üzerinde 2.5 petabyte veri – 600 milyar kayıtlık bir tablo
- Veri tabanı içi analitik yetkinlikler



Catalina Marketing: Müşteri Sadakati

	Rastgele Gönderim	Basit hedefleme; Köpek maması alanlara köpek maması indirim kuponu	Öngörüşel modelleme
Kupon Kullanımı	1%	6-10%	24%

- **195 milyon sadakat programı üyesi**

- Her kupon müşteriye özel
- 3 yıllık alışveriş davranışları göz önüne alınıyor

- **35 x kat daha fazla personel verimlilik**

- 2 aydan 2 güne düşen model geliştirme süreci
- Yıllık 90 modelden 900 modele geçiş
- 4.5 saatten 60 saniyeye düşen skorlama süresi
- 1 milyon ile 14.5 trilyon arası kayıt üzerinde çalıştırılan modeller



Teşekkürler...

Soru & Cevap