



# IBM Big Data Forum - Think Big

## Büyük Veri Perdesini Aralıyoruz

### Akan Veri İşleme ve Gerçek Zamanlı Analitik

A.Okay Akyüz  
Profesyonel Hizmetler Direktörü, OSiS

Hüseyin Yılmaz  
Bilgi Yönetimi Teknik Takım Lideri, IBM Türkiye



21 Şubat 2013

# Akan Veri Analizi ve Prematüre Bebekler



Ontario Institute of Technology (UOIT) Üniversitesi yeni doğan bebek ünitesinde sağlık kontrolü için Büyük Veri teknolojilerini kullandı

## Faydaları:

- Yaşamı tehdit eden koşulların 24 saat daha erken tespit edilebilmesi
- Gelişkin hasta hizmetleri ve düşük ölüm oranları

<http://www.youtube.com/watch?v=LW1WkkT-zAs>  
[http://www.youtube.com/watch?v=WNcclBzR\\_I4](http://www.youtube.com/watch?v=WNcclBzR_I4)



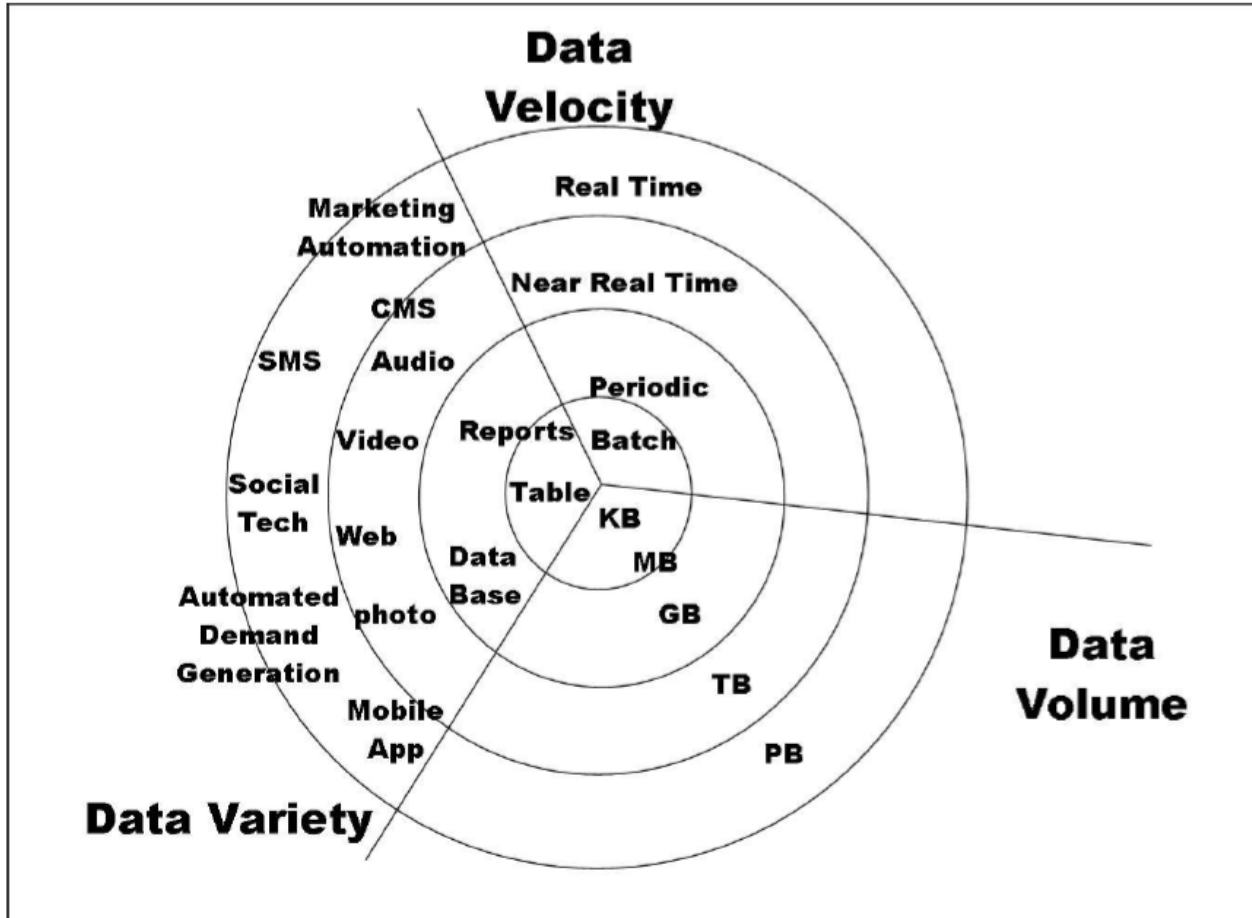
# Obama'ya kazandıran veri toplayıcılar find → predict → model

- Seçmenlerin bulunması ve dikkatinin çekilmesi
- Belirli etkilere cevap verecek kişilerin kimler olacağını öngörülmesi
- Davranış resmeden kişilerin modellenmesi



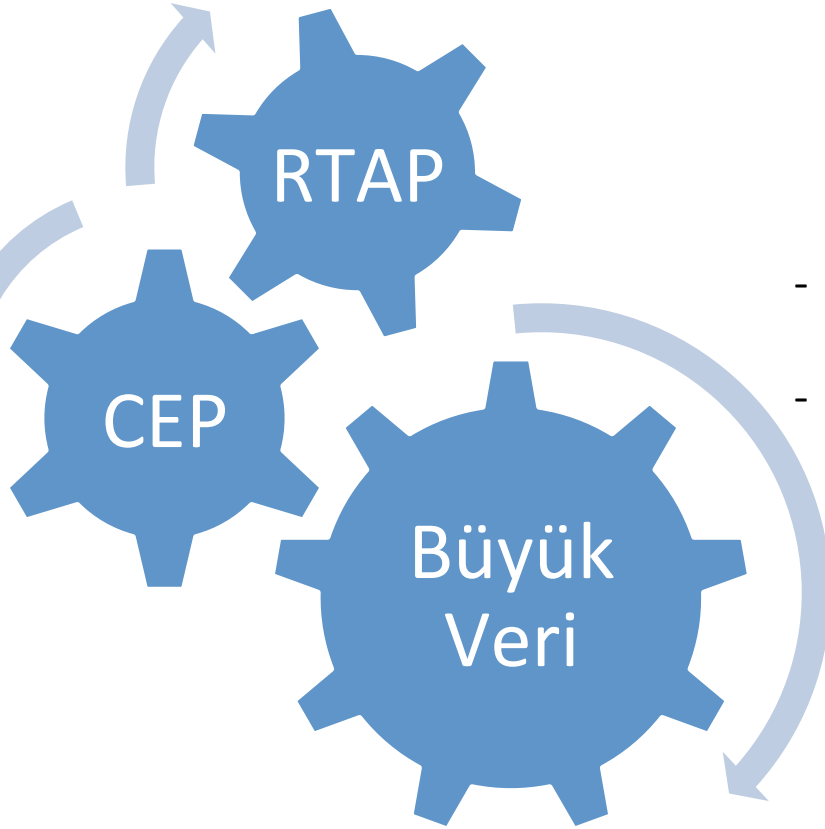
<http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/>

# Hangi akan veriler ile ilgilenmeliyiz?

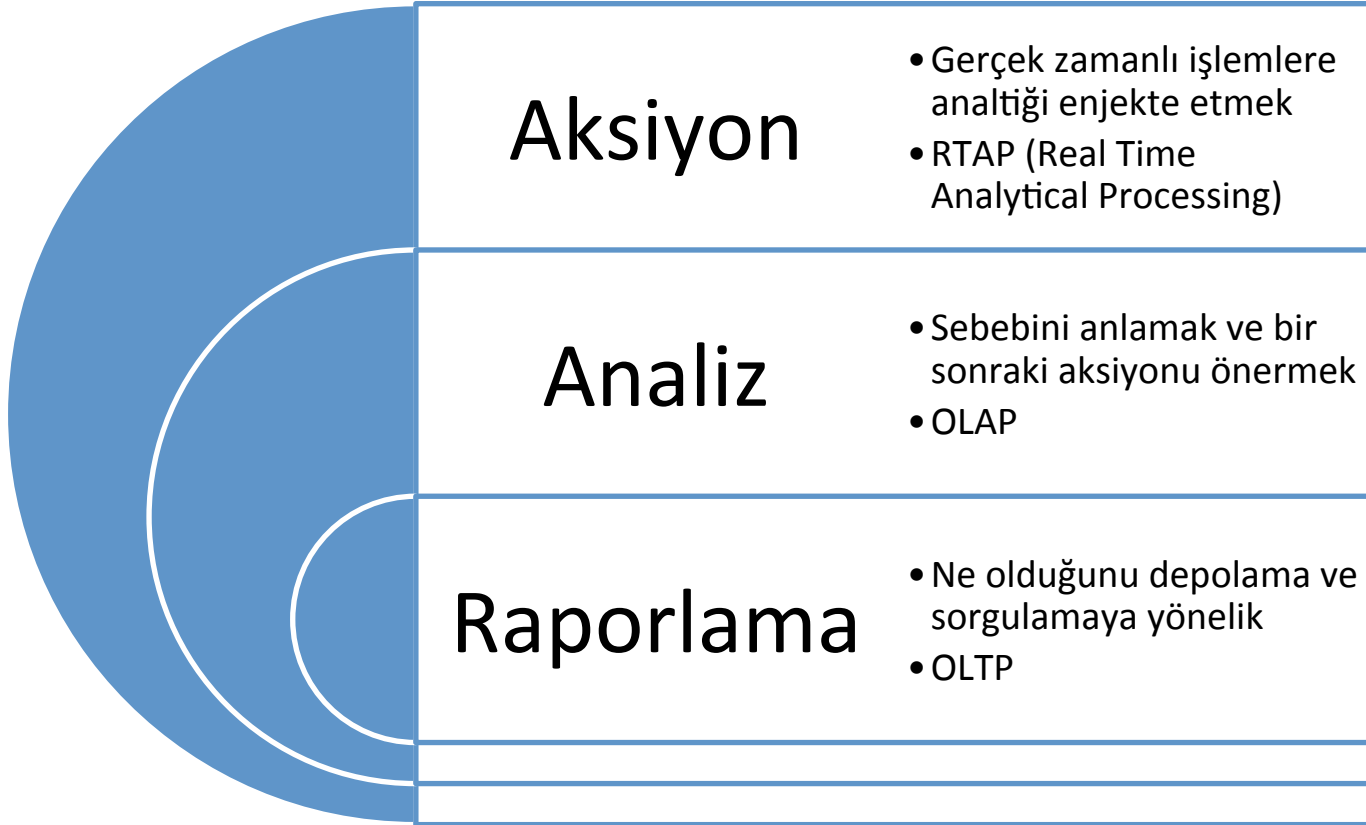


# Adaptif Analitik

- Farklı tip kaynaklardan gelen yüksek hacimli veri
- Geçmiş, Şimdi ve geleceğin kapsamlı görüntüsü
  - Çok kısa sürede netice veren gerçek zamanlı analizler
  - Derinlemesine analiz için tüm içeriğin kullanılabilmesi
- Akan ve ilgili diğer tüm verilerin bütünü için analitik
- Gerçek dünyada değişiklikler oldukça adaptasyon
  - Beklenmedik davranışlar gerçekleştiğinde
  - Önemli bir bilgiye rastlandığında derinlemesine analiz
  - İleride önemli olduğu anlaşılabilir şeyler
  - Önceden bilmediğimiz davranış ve aktiviteleri arama ve reaksiyon



# İş analitiğinin yarını: Gerçek zamanlı analitik işlem



# Complex Event Processing

## Geleneksel Yaklaşım

- Kesikli aktiviteler üzerinde analiz
- Aktiviteler arasındaki korelasyona bağlı olarak kural bazlı proses işleme
- Sadece yapısal veri tipleri desteklenir
- Daha düşük hız ve ölçekteki veri

## Akan Veri İşleme Yaklaşımı

- Sürekli akan veri üzerinde analitik
- Veri işleme proseslerinin içerisine basitten komplekse uzanan karmaşık analitik algoritmalar enjekte edilebilir.
- Yapısal ve yapısal olmayan veri tipleri desteklenir
- Daha hızlı akan veriler, yüksek ölçeklenebilirlik



# Akan Veri İşleme, Gerçek Zamanlı İleri Analiz ve Hesaplama Olanakları Sağlamaktadır

## Geleneksel İşleme



- Tarihsel verilerin analizi
- Disk üzerindeki verilerin taranması ve analizi
- Gecikmeli
- Sorgu-odaklı: sorguları durağan verilere gönderip sonuçlarını değerlendirme



## Akan Veri İşleme



- Anlık gerçeklerin işlenmesi
- Hareket halindeki verinin, diske yazılmadan analizi
- Düşük geçikme, anında aksiyon
- Veri tetiklemeli – verinin analitik süreçlerden geçmesi





# Akan Veri Teknolojileriyle Neler Yapılıyor?

## İletişim

- CDR işleme
- Sosyal Analiz
- Churn tahminleme
- Coğrafi Eşleştirme



## Finans Piyasaları

- Ortam değişkenlerinin hisse senedi fiyatlarına etkisi
- Pazar verilerinin anlık olarak analizi

## Savunma ve Siber Güvenlik

- Gerçek zamanlı çok boyutlu izleme
- Farkındalık
- Siber güvenlik taraması



## Ulaştırma

- Akıllı trafik yönetimi



## Akıllı Ağ & Enerji

- Kontrol
- Fazör izleme



## Sağlık Sektörü

- Yoğun Bakım Ünitesi İzleme
- Salgın Erken Uyarı Sistemleri
- Uzaktan Sağlık Durumu İzleme



## Sahtekarlık Engelleme

- Çok kişili sahteciliklerin tespiti
- Gerçek zamanlı sahtekarlık engelleme



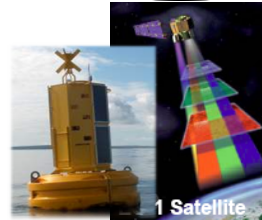
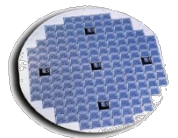
## Bilim

- Hava tahmini
- Geçici olay tespiti
- Sinkrotron atomik araştırmalar



## Diğer

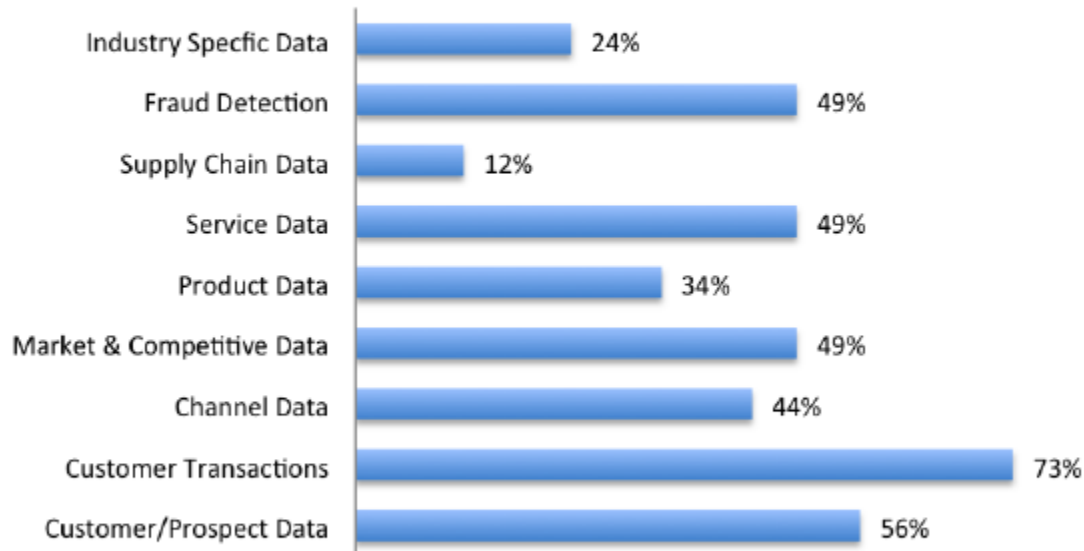
- Üretim Zinciri
- Text Analizi
- Kim kimle konuşuyor?
- ERP
- FPGA Hızlandırıcılar



## Doğal Yaşam Sistemleri

- Vahşi hayat yönetimi
- Su yönetimi

## What data domains are you most focused on in Big Data initiatives? (check all that apply)



### Big Data Executive Survey 2012

Consolidated Summary Report – Published  
December 31, 2012

# Akan Veri İşlemenin bir çok farklı uygulama alanı bulunuyor

## Müşterilerinizi daha yakından tanıyın

- Sosyal Medya, Müşteri Sentiment Analizi
- Promosyon Optimizasyonu
- Segmentasyon
- Müşteri karlılık
- Click-stream analizi
- CDR işleme
- Çok kanallı kampanyalar, etkileşimler
- Müşteri elde tutma programları
- Churn tahminleme



## Yeni Ürün ve Servisler Geliştirin

- Sosyal Medya – Ürün/Marka Duygu Analizleri
- Marka Stratejisi
- Pazar Analizi
- RFID izleme & analizi
- İç-görü tabanlı ürün / hizmetler oluşturmak için işlemlerin analizi

## Gecikmesiz operasyonlar

- Akıllı ölçüm cihazları
- Dağıtım yük tahmini
- Satış Raporlama
- Envanter ve satış optimizasyonu
- Finans piyasaları işlemleri
- Yoğun Bakım Ünitesi izleme
- Hastalık gözetim
- Ulaştırma ağı optimizasyonu
- Doğal ortam analizleri
- Deneysel araştırmalar



## Risk ve Sahteciliğin Anında Tespiti

- Multimodal surveillance
- Siber güvenlik
- Sahtekarlık modelleme
- Sahtekarlık tespiti
- Risk Modelleme ve Yönetimi
- Regülatif Raporlama

## Makine Verileri Analizi

- Network analitik çözümleri
- Varlık yönetimi ve akıllı sorun çözme
- Websitesi analitiği
- IT log analizleri

# Nowcasting zamanı

- Şimdigörü: Ekonomik ve hava tahminlerinde kullanılan yöntem, şimdi müşterinin yapacağı hareketleri görmemizde yardımcı oluyor.
- *Hal Varian and Google economist Hyunyoung Choi wrote in a [paper](#) that by using Google Trends data for a given geography, highly accurate predictions can be made for such major economic activities as car and home purchases. By modeling search data for, say, "real estate agents" in Texas, they are able to more accurately estimate home sale activity in the region.*

# Akan Veri ve Big Data ile ilgili yatırım yaptığımız konular

- Hareketleri anlık izleyerek bağlantılı süreçlerin optimizasyonu (Perakende)
  - Tahmin -> Sipariş -> Stok -> Fiyat Optimizasyonu
  - Müşteriye yönelik lokasyon ve segmente dayalı sosyal medya verisi ile destekli öneriler
  - İş Gücü Planlama
  - Suistimal tesbiti
- Gerçek zamanlı veri tutarlılığı kontrolü (Tüm sektörler)
- Sosyal Medya Monitör ve akıllı öneri (B2C)
- Suistimal Tesbiti (Bankacılık)
- Network Optimizasyonu (Telekomünikasyon)
- Belediyecilik ve şehir hizmetleri

# anlık satış



Satis

Mağaza

Bölge

Marmara

Ege

İç Anadolu

Akdeniz

Karadeniz

Doğu Anadolu

Güneydoğu An

Saatlere Göre Satış



Özet

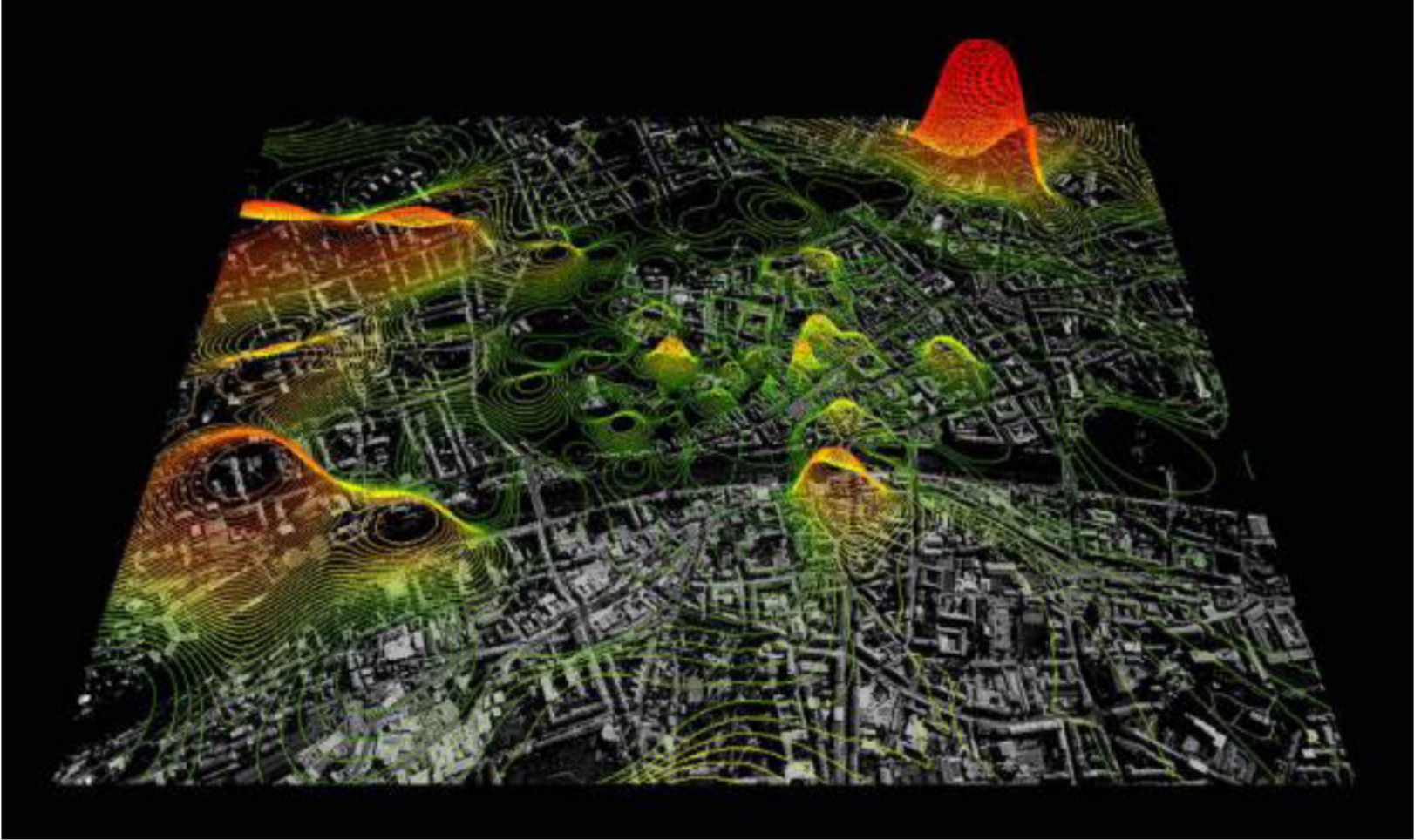
Anlık Satış: **75.937**

Günsonu Tahmini: **452.204**

Şehir Bazında Satış

Şehir	Anlık Satış	Dün Satış	Geçen Ay Satış	Geçen Yıl Satış
ANKARA	16.529	12.178	8.462	9.289
ANTALYA	2.637	1.476	1.033	1.804
BALIKESİR	6.410	2.069	1.714	1.947
DENİZLİ	1.418	1.151	709	688
ERZURUM	2.608	1.731	1.215	1.947
ERZURUM	2.608	1.731	1.215	1.947

# Kentlerde gerek zamanlı yoęunluk tahmini



The senseable project: <http://senseable.mit.edu/grazrealtime/>

# Güvenlik/İstihbarat Uzantısı: İhtiyaçlar



Güvenlik/İstihbarat Uzantısı, geleneksel güvenlik çözümlerini tüm veri tipi ve kaynaklarının analizi ile geliştirir.



Enhanced  
Intelligence &  
Surveillance  
Insight

## Akan & Duran veri analizi ile:

- İlişkileri/ortaklıkları bul
- Şablon 6 olguları ortaya çıkar
- Bilgi değerini koru



Real-time Cyber  
Attack Prediction &  
Mitigation

## Network trafiği analizi ile:

- Yeni tehditleri erken keşfet
- Bilinen karmaşık tehditleri tespit et
- Gerçek zamanlı aksiyon al



Crime prediction  
& protection

## Telco & sosyal veri analizi ile:

- Suç kanıtlarını topla
- Suç faaliyetlerini önle
- Proaktif olarak suçluları yakala



# Kamu alanında bazı uygulamalar

- Crowdsourcing
- Haiti depreminde SMS ile kalıntıların altında bulunması muhtemel insanları bildirme
- SMS sayısı ile hasar sayısı arasında korelasyon
- Boş park yeri bulma
- Gidilen bir restaurantın daha önce ceza alıp almadığı
- Google Trendlerinden istifade, yeni tür grip vak'alarının tesbiti

# Akan Veri Analizinin sağladıkları?

- Olaylar gerçekleşiyorken analiz edip, reaksiyon vermek
- Fazla sayıda kaynaktan gelen veriden gerçek zamanda istifade etmek
- Çok yakın gelecekte ya da bir sonraki aşamada ne olacağını söyleyen modellerin geliştirilmesi
- Streams sayesinde
  - Ölçeklenmek: saniyede birkaç event'den milyonlarca event'e kadar cevap verebilme
  - Reaksiyon zamanı: 1 saniye altında sürede aksiyon alınabilecek neticeler alınmasına imkan verir. (<20 mikro saniye)
- Gerçek manada durum tespiti



# InfoSphere Streams – Akan Veri İşleme Teknolojisi

## Analitik Uygulamalar

BI / Raporlama | Analiz/ Görselleştirme | Fonksiyonel Uyg. | Endüstri Uyg. | Veri Madenciliği | İçerik Analitiği

## IBM Big Data Platformu

Görselleştirme ve Keşif

Uygulama Geliştirme

Sistem Yönetimi



Hızlandırıcılar

Hadoop Sistemi

Akışkan veri işleme

Veri Ambarı



Veri Entegrasyonu ve Sahipliği

- **Hareket halindeki verilerin analizi**
  - Eş zamanlı çok kanallı akışkan veri girişi
  - Ölçeklenebilir mimari
- **Farklı tipten verilerin işlenebilmesi**
  - Yapısal / yapısal olmayan, video, ses
  - İleri Düzey Analitik operatörler
- **Adapte olan Gerçek Zamanlı Analitik**
  - Veri Ambarı ile birlikte
  - Hadoop ile birlikte

# Streams Nasıl Çalışır?

- Devamlı Besleme
- Sürekli Analiz

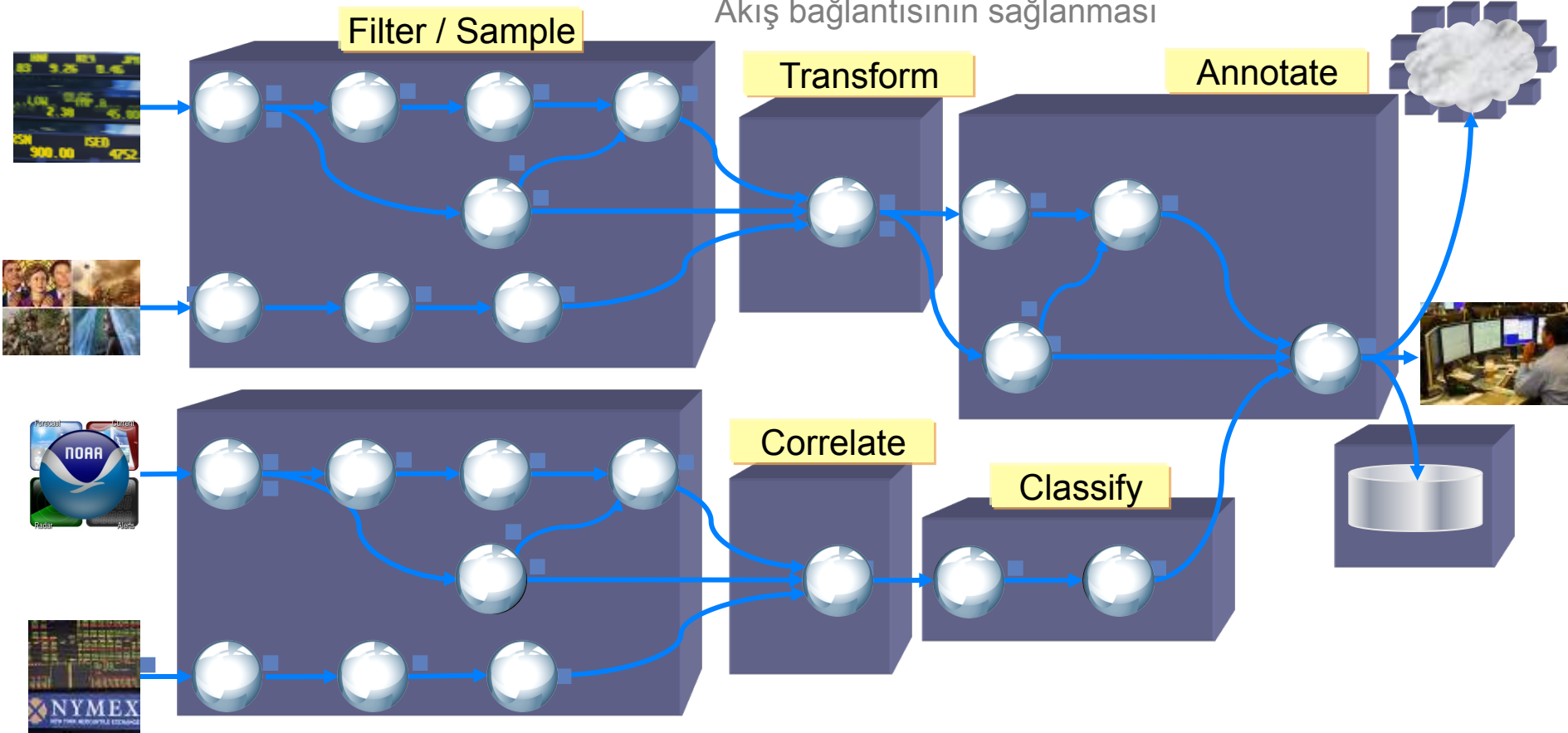


# Streams Nasıl Çalışır?

- Devamlı Besleme
- Sürekli Analiz

Altyapı şunlar için servis sunar;

Analitik işlerin kaynaklara dağıtımı, zamanlama,  
Akış bağlantısının sağlanması



Ölçek Kazanın:

Uygulamayı yazılım bileşenlerine bölerek  
Stream-bağlantılı donanımlar üzerine dağıtarak

Uygun olduğunda:

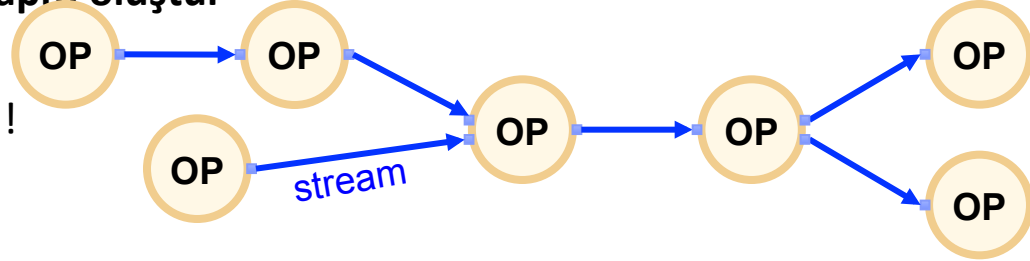
Elemanlar düşük gecikme için aynı  
ortamda bulunabilir

# Ölçeklenebilir Stream İşleme

- Streams programlama modeli: Çizge (Graph) oluştur

- Matematiksel Kavram

- line -, bar -, ya da pie chart değil!
- “Network” de denir
- Örneğin “Tree” bir çizgedir



- “**operator**”ler ile bunları bağlayan “**stream**”ler içerir

- Matematiksel Çizgenin Vertice ve Edge bileşenleridir
- Yönlendirilmiş (Directed) çizgedir

- Streams Çalışma zamanı modeli: Dağıtık İşleme

- Tek veya çoğul operatorler Processing Element (PE) oluşturur

- Derleme ve çalışmanı servisleri PE

- Bir makine üzerine
- Kümede birden çok makine üzerine yayılmasını kolaylaştırır

- Çalışmanı servisleri tüm bağlantı ve veri taşıma ihtiyaçlarını sağlar

- Otomatik
- Gerekli yerlerde manual yerleştirme direktifleri aracılığı ile

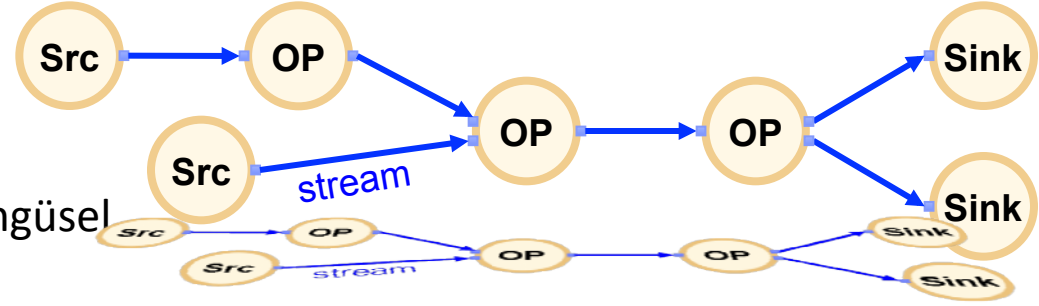
# InfoSphere Streams Çalışmaları

- *Küme üzerinde dağıtık çalışma*
  - Küme boyutunda limit yoktur
- *İşler arasında Stream Import/export*
  - Çeviklik: Uygulamaların dinamik tekrar konfigürasyonu
- *Performans ve Sağlık Çalışma (health) Metrikleri*
  - Operator seviyesine kadar
- *Herhangi bir anda işler gönderilebilir ya da iptal edilebilir*
- *Herhangi bir anda kümeye Host ekleme ya da kümeden çıkarma*
- *Yüksek devamlılık*
  - Uygulamaların belli parçaların Restart/Relocate edilebilmesi
  - Uygulama parçalarının “Host Tag”lerine göre Host havuzlarına atanması
- *Uygulama taşınabilirliği*
  - Basitleşmiş test ve işlemler

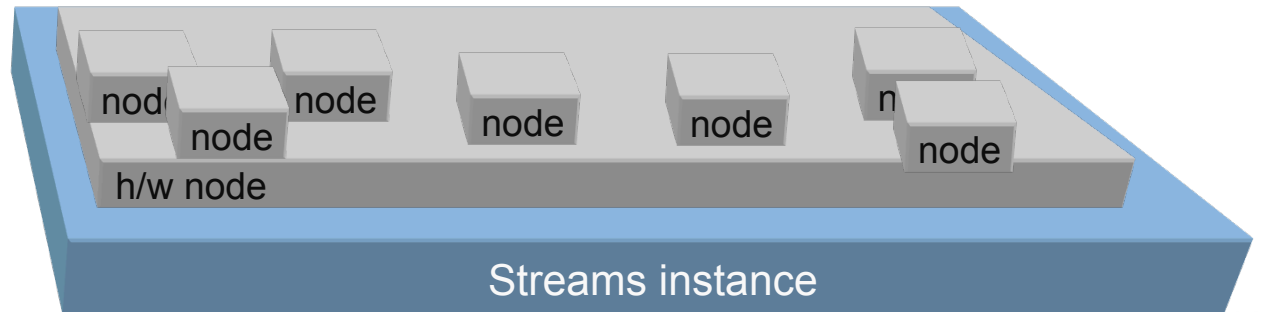
# Temel Öğelerden Çalışan İşlere

- Streams uygulama çizgesi:

- Yönlendirilmiş, muhtemelen döngüsel
- Operatorlerden oluşan
- Streamler ile bağlanmış



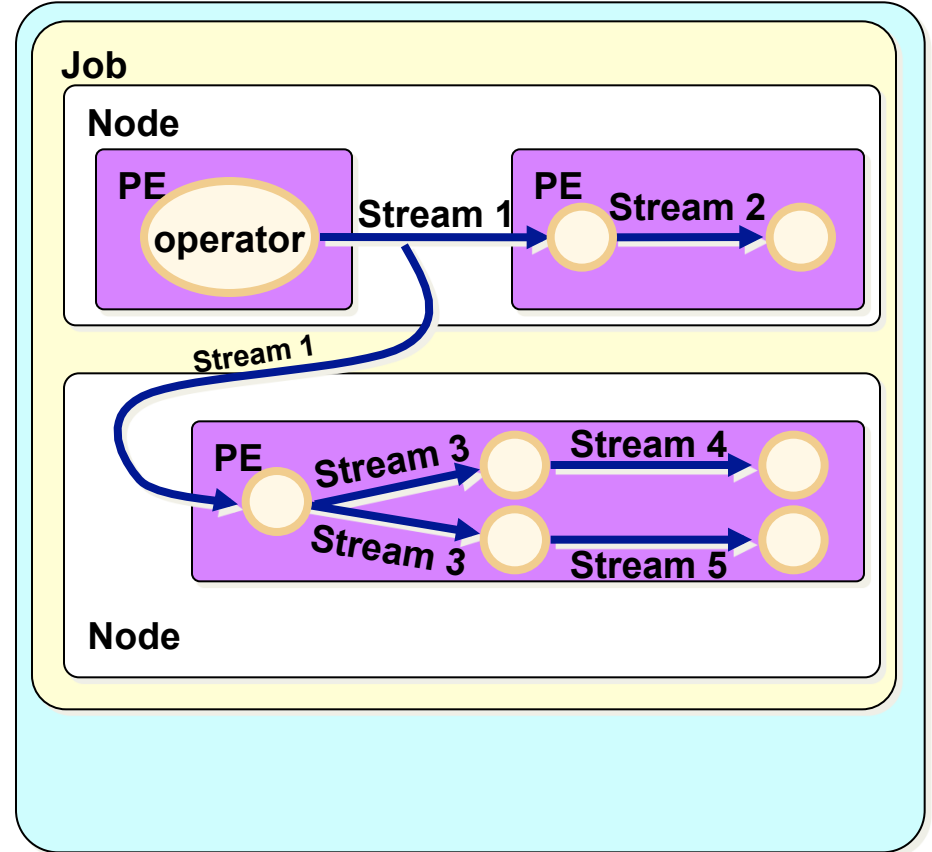
- Her tam uygulama, muhtemel konuşlandırılabilir iştir (job)
- İşler “Streams runtime environment” üzerine konuşlandırılır (Streams Instance)
- Instance tek bir işleme düğümünden (donanım)
- Ya da birden çok işleme düğümünden oluşabilir





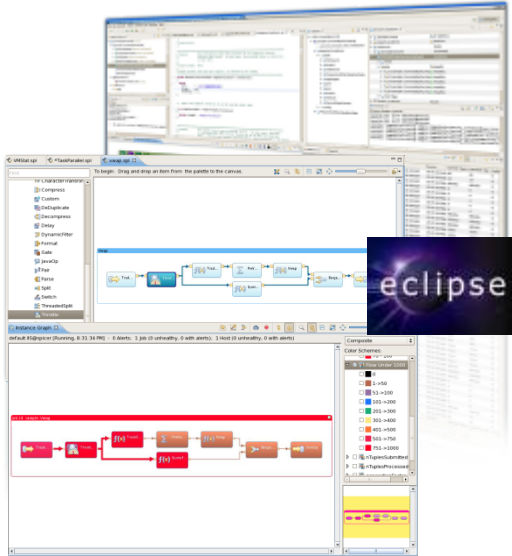
# InfoSphere Streams Nesneleri: Çalışmaanı Görünümü

- **Instance**
  - Bir veya daha çok host üzerinde işleyen InfoSphere Streams çalışmaanı örnekleme
  - Bileşen ve servis toplamıdır
- **Processing Element (PE)**
  - Streams Instance tarafından çalıştırılan temel işlem birimi
  - Bir ya da daha çok operator içerebilir
- **Job**
  - Instance içinde çalışan, Konuşlandırılmış Streams uygulamasıdır
  - Bir ya da daha çok PE içerir



# IBM InfoSphere Streams 3.0

## Kapsamlı Araçlar



- Eclipse IDE
- Web console
- **Drag & Drop editor**
- Instance graph
- **Streams visualization**
- Streams debugger

IBM Big Data Forum - Think Big

Büyük Veri Perdesini Aralıyoruz

26

## Ölçeklenebilir Mimari



- Limitsiz kapasite için kümeli çalışmaanı
- RHEL v5.3 ve üzeri
- CentOS v6.0 ve üzeri
- X86 & Power çok işlemcili donanım
- InfiniBand desteği
- Ethernet desteği

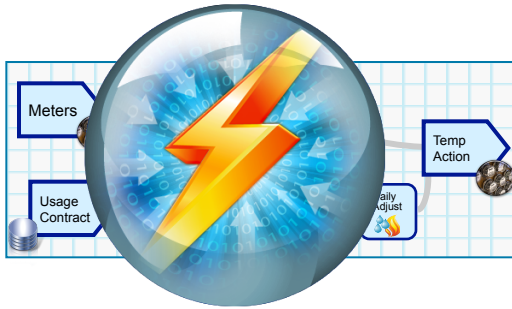
## Araç & Hızlandırıcılar ile Sofistike analitik



- **Big Data, CEP**, Database, **Data Explorer (Big Data)**, **DataStage**, Finance, **Geospatial**, Internet, **Messaging**, Mining, **SPSS**, Standard, Text, **TimeSeries** araçları
- **Telco & Social Media hızlandırıcılar**

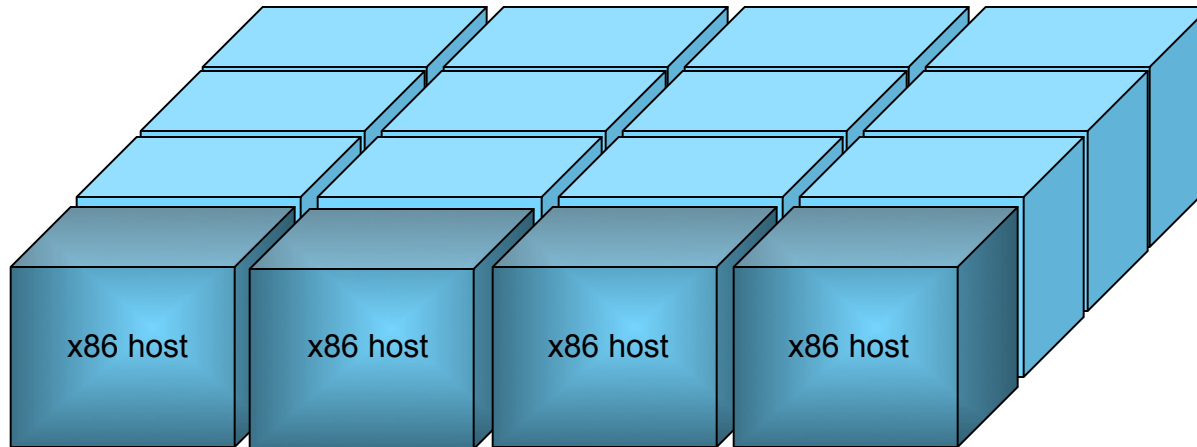


# Streams Runtime Resmi



Optimizing scheduler assigns jobs to hosts, and continually manages resource allocation

Commodity hardware – laptop, blades or high performance clusters



# Streams Runtime Resmi

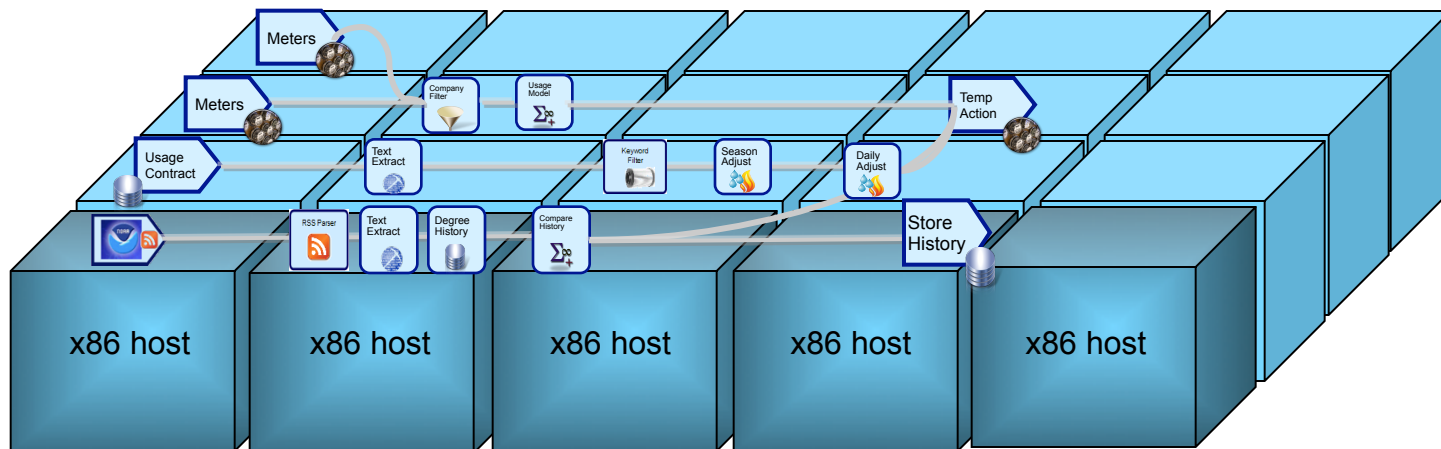


Optimizing scheduler assigns PEs to hosts, and continually manages resource allocation

Dynamically add hosts and jobs

Commodity hardware – laptop, blades or high performance clusters

New jobs work with existing jobs

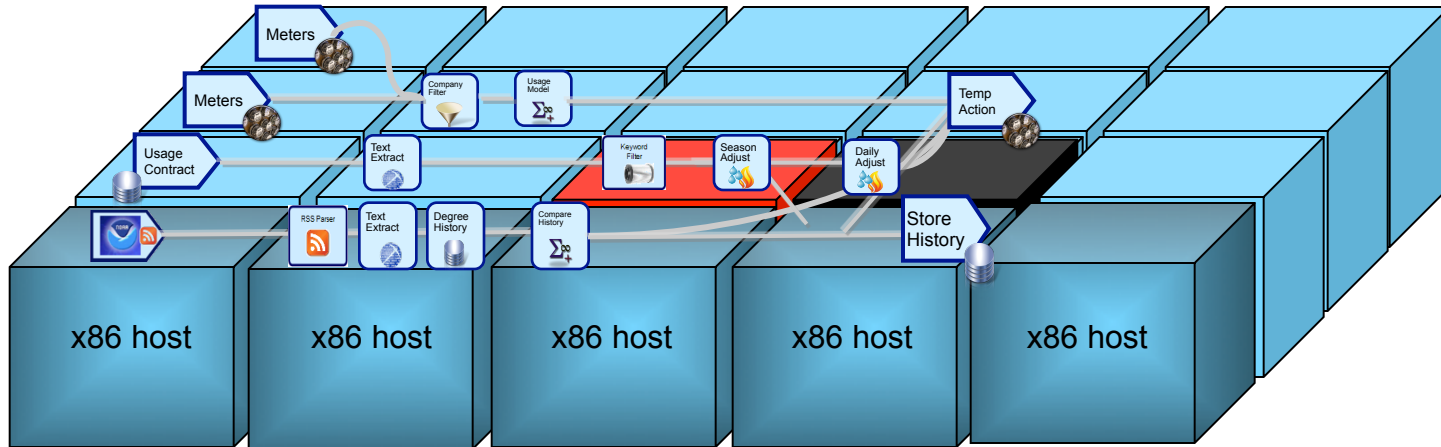


# Streams Runtime Yüksek Devamlılığı İçerir



PEs on busy hosts can be moved manually by the Streams administrator

A PE failing on one host can be moved automatically to another; communications are automatically rerouted



# Geliştirmeyi Basitleştirecek & Hızlandıracak Araç ve Operatorler

## Standart Toolkit

### Relational Operators

Filter	Sort
Functor	Join
Punctor	Aggregate

### Adapter Operators

FileSource	UDPSource
FileSink	UDPSink
DirectoryScan	Export
TCPSource	Import
TCPSink	MetricsSink

### Utility Operators

Custom	Split
Beacon	DeDuplicate
Throttle	Union
Delay	ThreadedSplit
Barrier	DynamicFilter
Pair	Gate
JavaOp	

Standart toolkit, ürün ile birlikte gelen operatorleri içerir.

## Internet Toolkit

### InetSource

HTTP	FTP	HTTPS
FTPS	RSS	file

## Database Toolkit

ODBCAppend	ODBCEnrich
ODBCSource	SolidDBEnrich
DB2SplitDB	DB2PartitionedAppend

Destekler: DB2 LUW, IDS, solidDB, Netezza, Oracle, SQL Server, MySQL

- Financial Toolkit
- Data Mining Toolkit
- Big Data toolkit
- Text Toolkit
- User-Defined Toolkits
  - Dili kullanıcı-tanımlı operator ve fonksiyonlar ekliyerek genişletin