Information Management

IBM

IBM Big Data Portfólió Áttekintés

Baranyi Szabolcs
+36 20 823 5619
Szabolcs.baranyi@hu.ibm.com

September 9, 2013

Tartalom

- Big Data Platform
- Big Insight
- InfoSphere BigInsights Quick Start Edition
- Streams
- InfoSphere Streams Quick Start Edition
- Data explorer

- PureData for Analytics (Appliance)

Big Data A technológia ami lehetővé teszi hogy minden adatot elemezünk

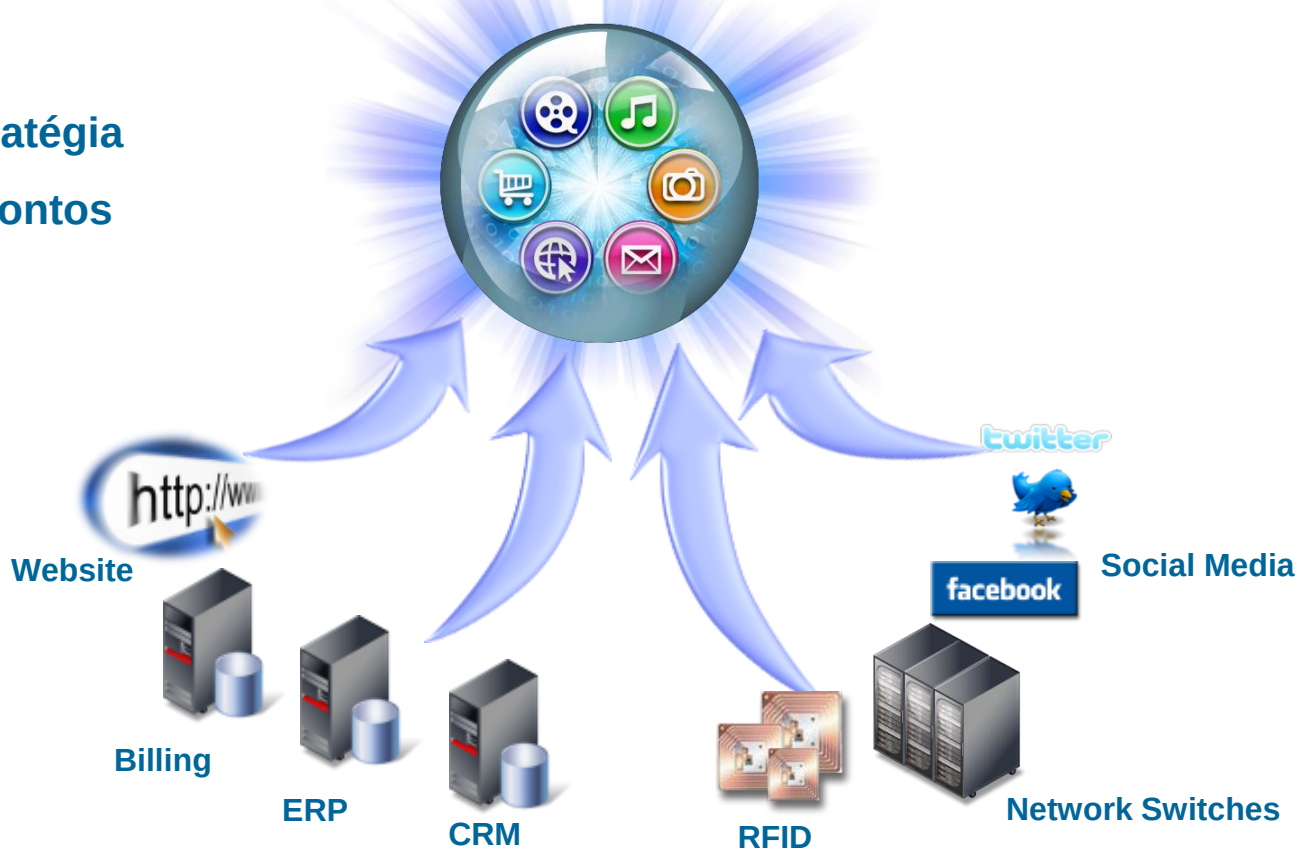
Költséghatékony menedzsmentje és elemzése

Struktúrált, struktúrálatlan és adatfolyamban

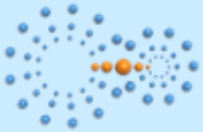
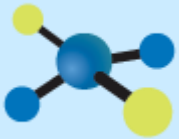
natív formában elérhető adatnak

BigData Stratégia

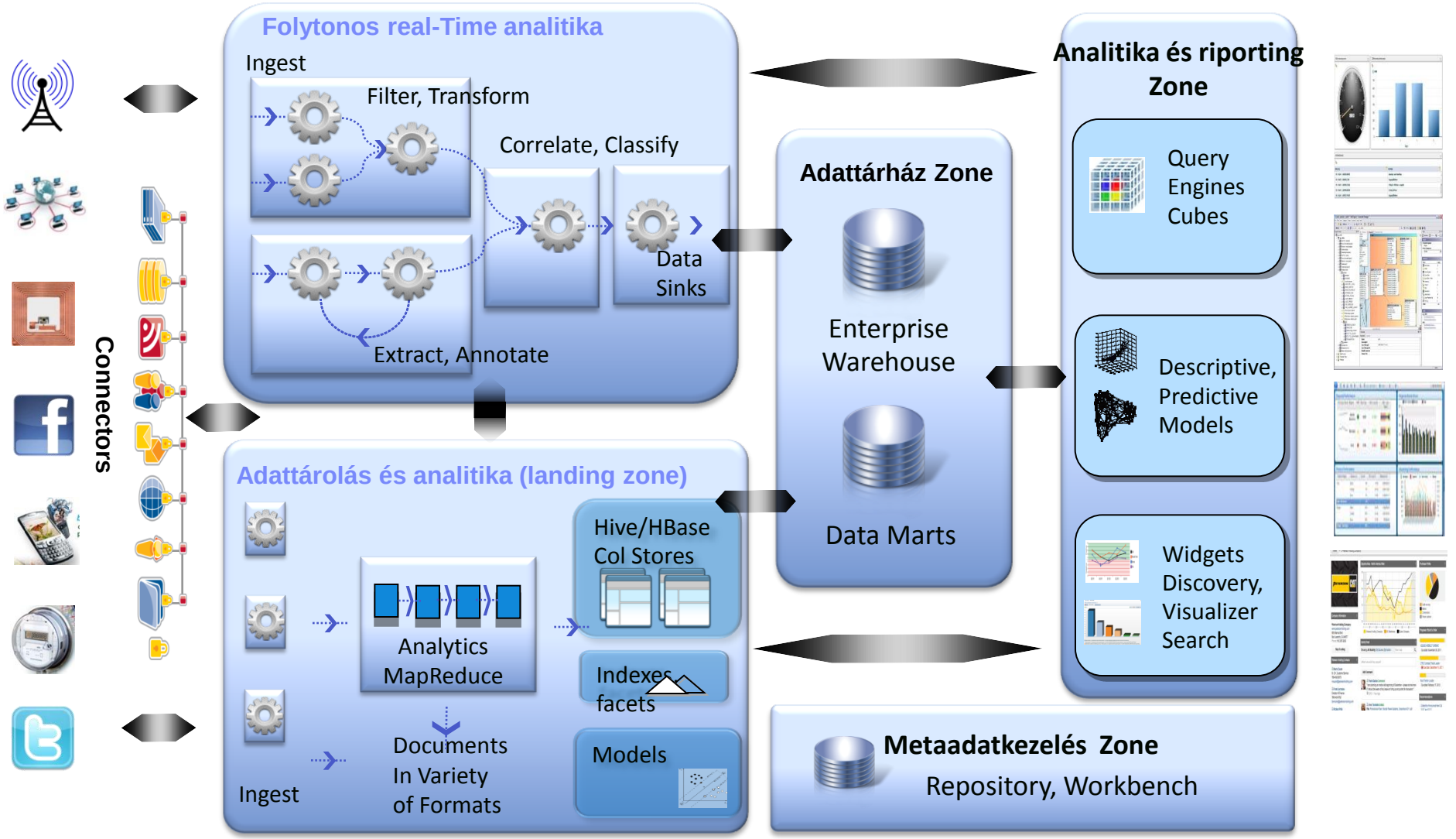
már most fontos



BIG DATA több mint HADOOP

Megtalálni, megérteni és navigálni az adathalmazban		Elosztott keresés és navigáció
Nagy mennyiségű adatkezelés		Hadoop elosztott file rendszer MapReduce: elosztott feladatok
Struktúrált adatok		Adattárház, Célhardverek
Adatfolyam, streaming média		Stream Computing: Adatfolyam feldolgozás
Nem struktúrált elemzés		Text Analytics Engine
Adatintegráció és követés sokféle adatforrásból		Integráció, Adatminőség, Biztonság, Életciklus

Big Data Platform funkciócsoportok



BigData Platform elemei

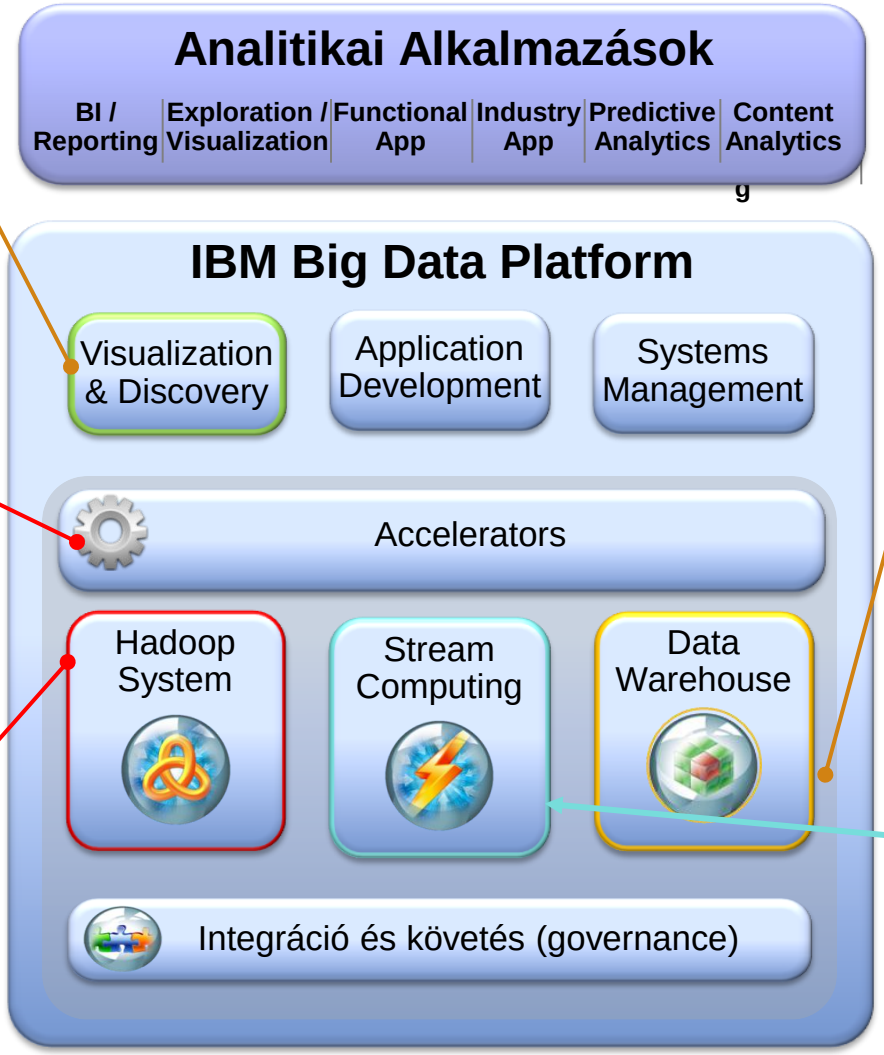
1 – Adatvisszanyerés, felfedezés

InfoSphere Data Explorer

Analytic:
Text, Geospatial, Time series,Data mining
Applications
Financial, Machine Data, Social, Telco

2 – Natív analízis
3 – Költséghatékony adattárolás

InfoSphere BigInsights



4 – Egyszerű, hatékony adattárház (célhardverek)

IBM Warehouse Solutions

5 – Adatfolyam feldolgozás, gyors válasz

InfoSphere Streams

Information Management

IBM

IBM InfoSphere BigInsights

Volumen és Variancia

September 9, 2013

IBM InfoSphere BigInsights v2.1 Enterprise Edition

Visualization & Discovery

- BigSheets
- Dashboard & Visualization

Applications & Development

- Big SQL
- Apps
- Text Analytics
- MapReduce
- Workflow
- Pig & Jaql
- Hive

Administration

- Admin Console
- Monitoring

Integration

- JDBC
- Netezza
- DB2
- Streams
- DataStage
- Guardium
- Platform Computing
- Cognos
- Flume
- Sqoop

Advanced Analytic Engines

- Adaptive Algorithms
- Text Processing Engine & Extractor Library)
- R

Workload Optimization

- Integrated Installer
- Enhanced Security
- Splittable Text Compression
- Adaptive MapReduce
- High Availability
- ZooKeeper
- Oozie
- Jaql
- Flexible Scheduler
- Lucene
- Pig
- H Catalog
- Index

Runtime

- MapReduce

Data Store

- HBase
- Hive

File System

- HDFS
- GPFS

Management

- Security
- Audit & History
- Lineage

Open Source

IBM

BigInsights és az adattárház

Big Data analytic applications

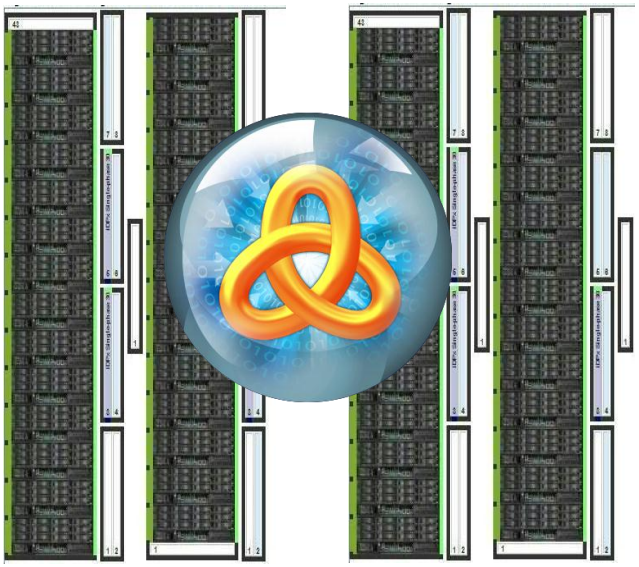
Tradícionális analitika

Adattárház

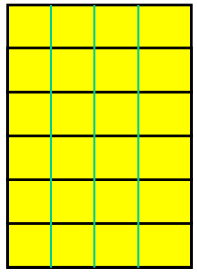
BigInsights

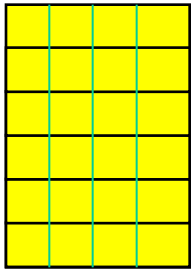
facebook

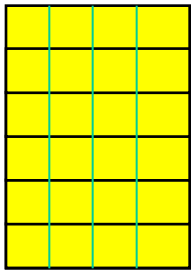
twitter



Filter *Transform* *Aggregate*







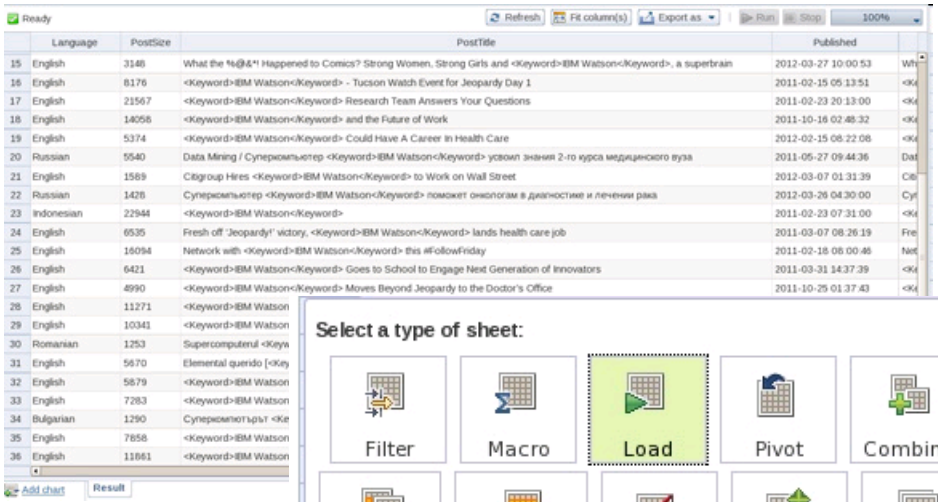
Táblázatos Analízis

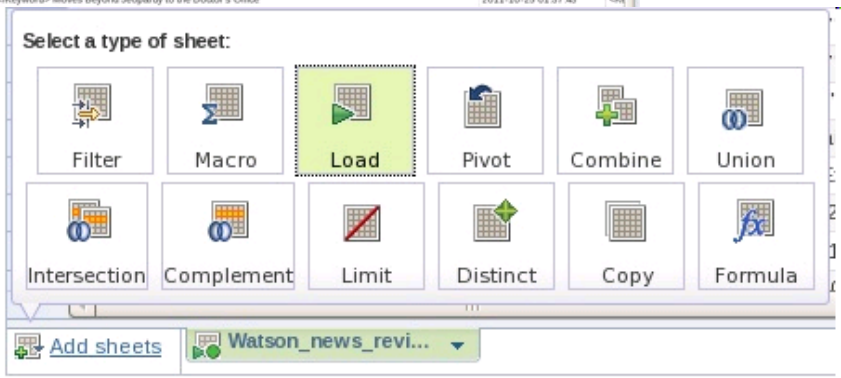
- **Webes analízis és vizualizáció**

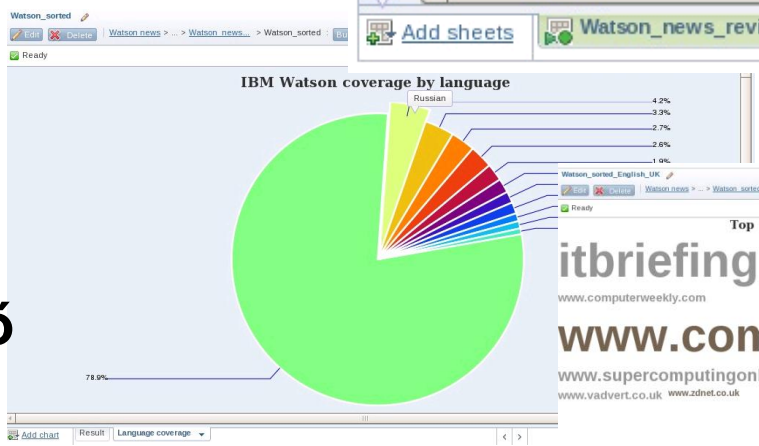
- **Táblázat alapú felület**
  - Táblázatos formában jobokat definiálunk
  - Visszaadott értékeket diagramokat elemezzük módosítjuk

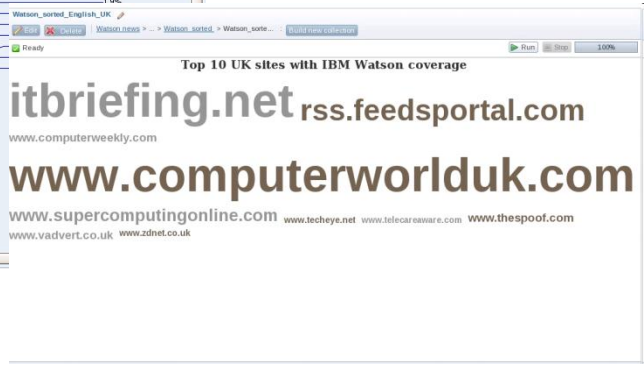
(Nagy excel)

- JAQL: **Speciális hierarchikus lekérdező nyelv hadoop környezethez**



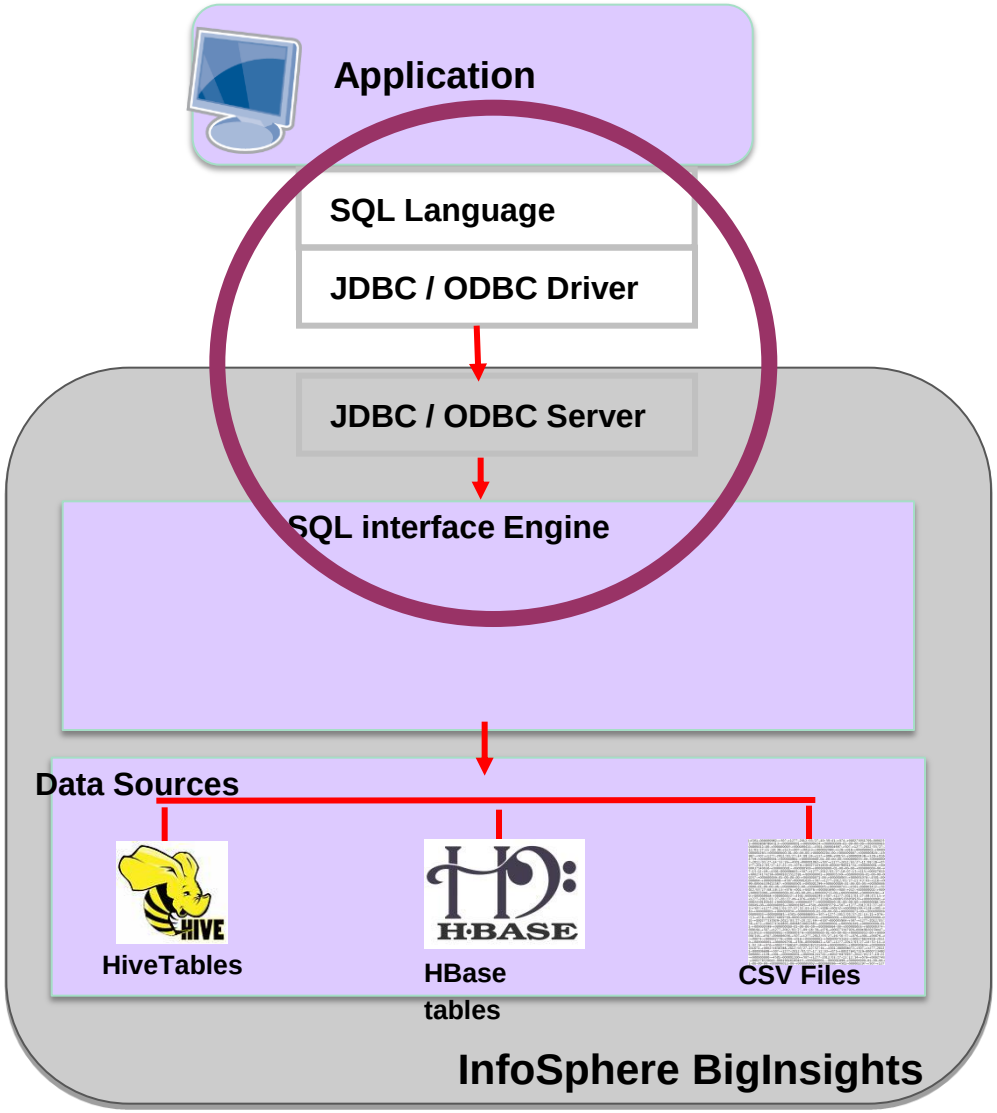






SQL interface . . . .

- **SQL lekérdezési lehetőség**
  - SQL '92 és 2011 opciók
  - Korellált subquery
  - Windowed aggregates
- **SQL elérés minden Big Insight beli strukturált adathoz**
- **JDBC/ODBC support**
- **MapReduce párhuzamosságának kihasználása**
- **BigSQL supports: create table ;data types including varchar, decimals, etc.**



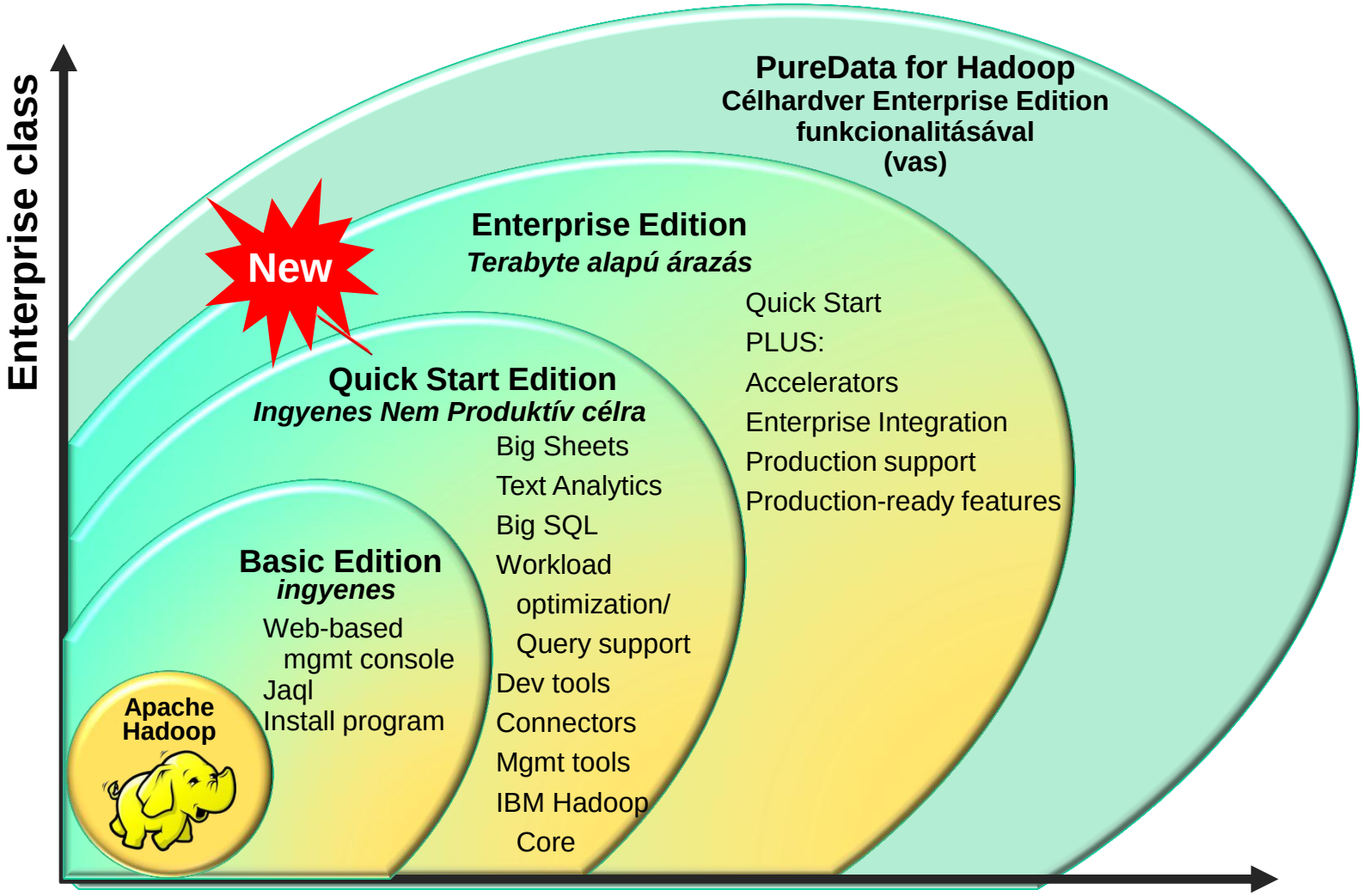
Cluster komponensek és monitorozásuk:

- **Cluster:**CPU/Disk/Memory/Netk ihasználtság, node életjel
- **HDFS:** File rendszer állapota, NameNode JVM, írás / olvasás statisztika
- **Mapreduce:** Jobok státusa, Mapper, Reducer, JobTracker
- **HBase**: lekérdezések állapota
- **Hive**: metadata store hívások gyakorisága
- **Oozie**: statistics
- **Zookeeper**: késleltetlés ,lekérdezések
- **Flume:** adatforrások, nyelők állapota

EXT E N S I B L E !!

Build your own Monitoring Dashboards,
with the key KPI that are of your interest!

Kezdetektől a nagyvállalatig



BigInsights Quick Start Edition contains most of the same features as the Enterprise Edition

Available

- Big Sheets
- Text Analytics
- Big SQL
- All Workload optimization/Query support
- Development tools
- Connectors
- Management tools
- IBM Hadoop Core

Unavailable

- Production support
- Production-ready features:
  - High Availability
  - GPFS
- Accelerators:
  - Machine Data
  - Social Data
- Limited use licenses:
  - Data Explorer
  - Cognos
  - Streams

IBM InfoSphere Streams 3.0

Agilitás, Gyorsaság

InfoSphere Streams

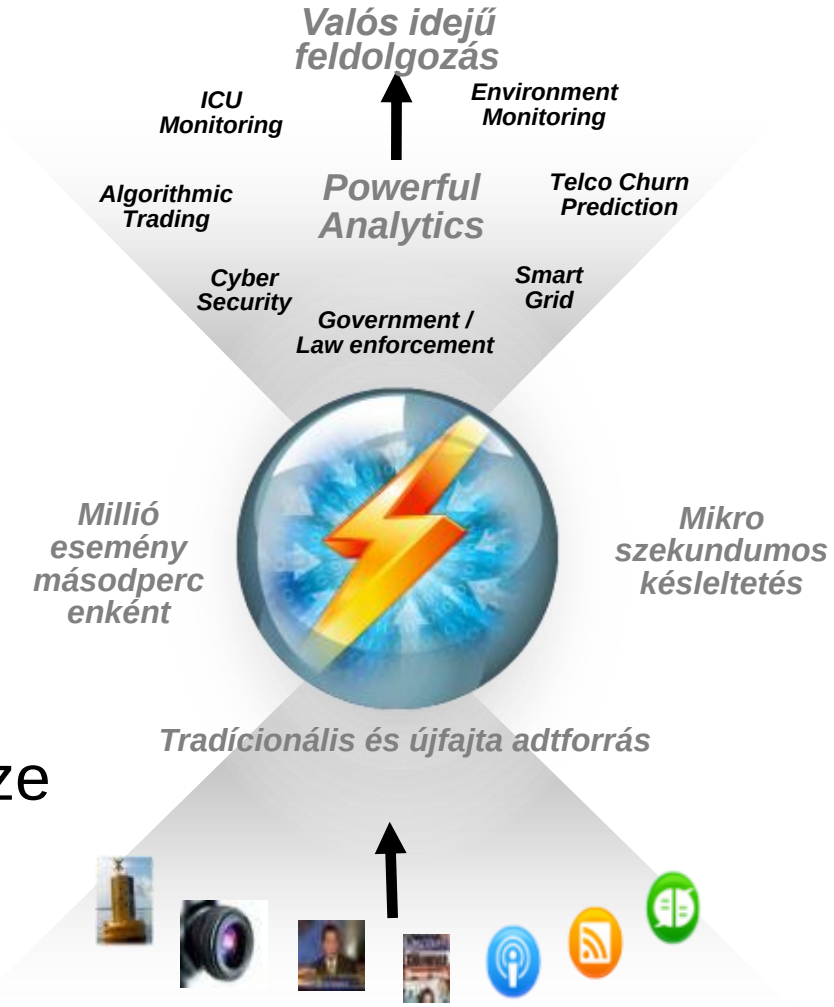
Valós idejű analititka BIG Data felett

Fókuszban a sebesség

Volumen Terabytes / sec
Petabytes / nap

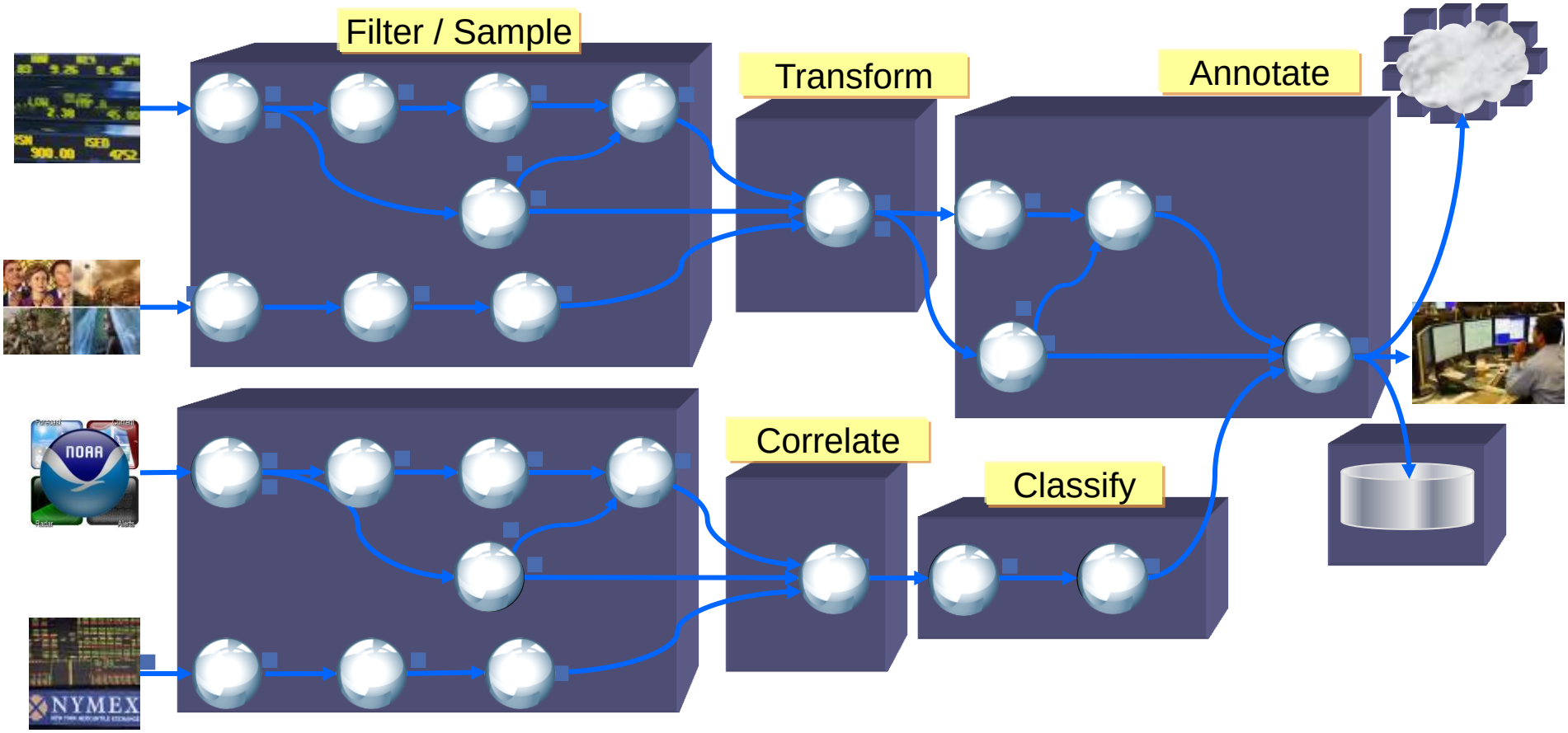
Variancia sokféle adat
sokféle elemzés

Sebesség Kiértékelés másodperc tört része alatt



Streams Működése

Streams infrastruktúra

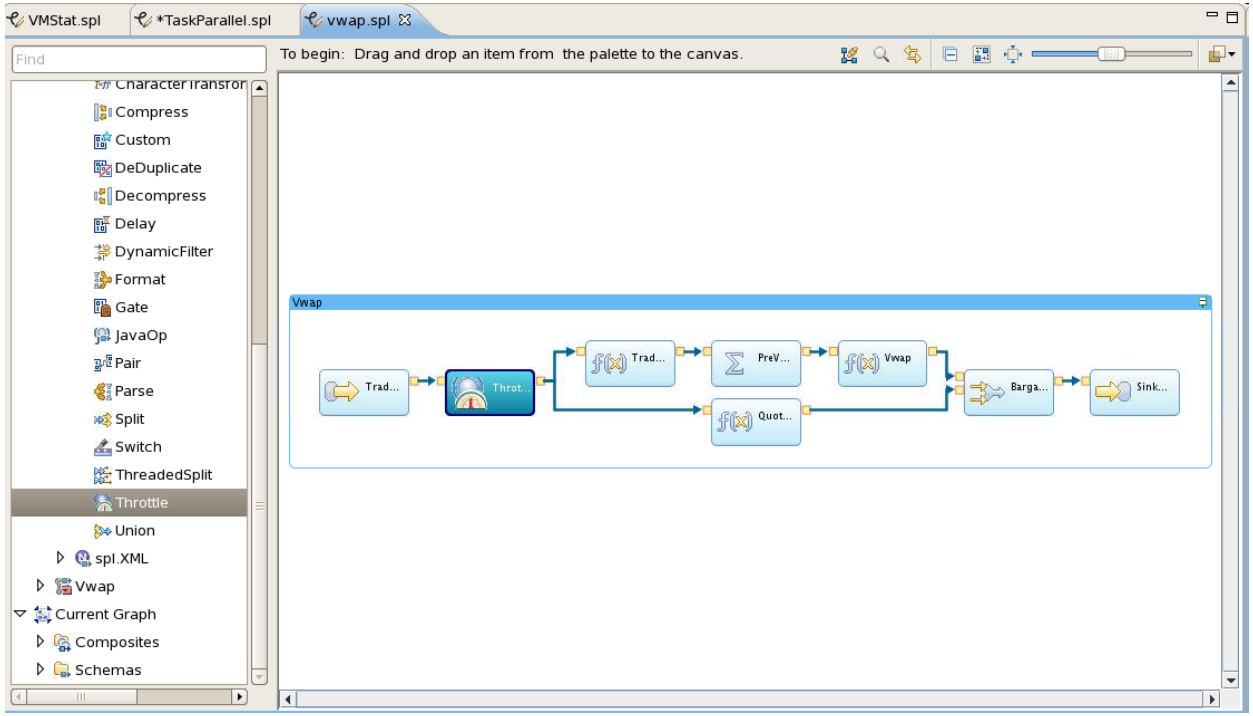


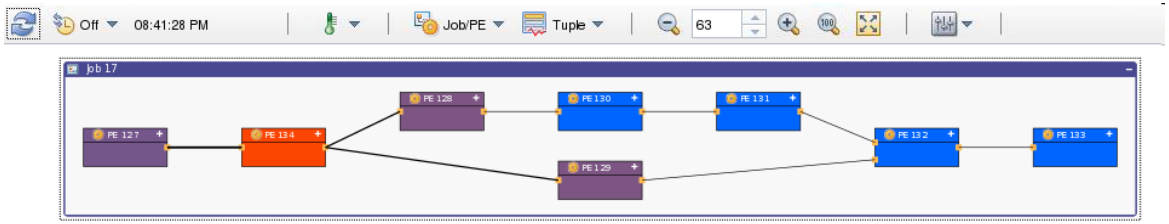
Egyedi gondolkodásmód

Folyam feldolgozó egységek

Grafikus szerkesztő és monitorozó

- Vizuális programozás
- SPL nyelv
- Hierarchikus vizuális monitoring

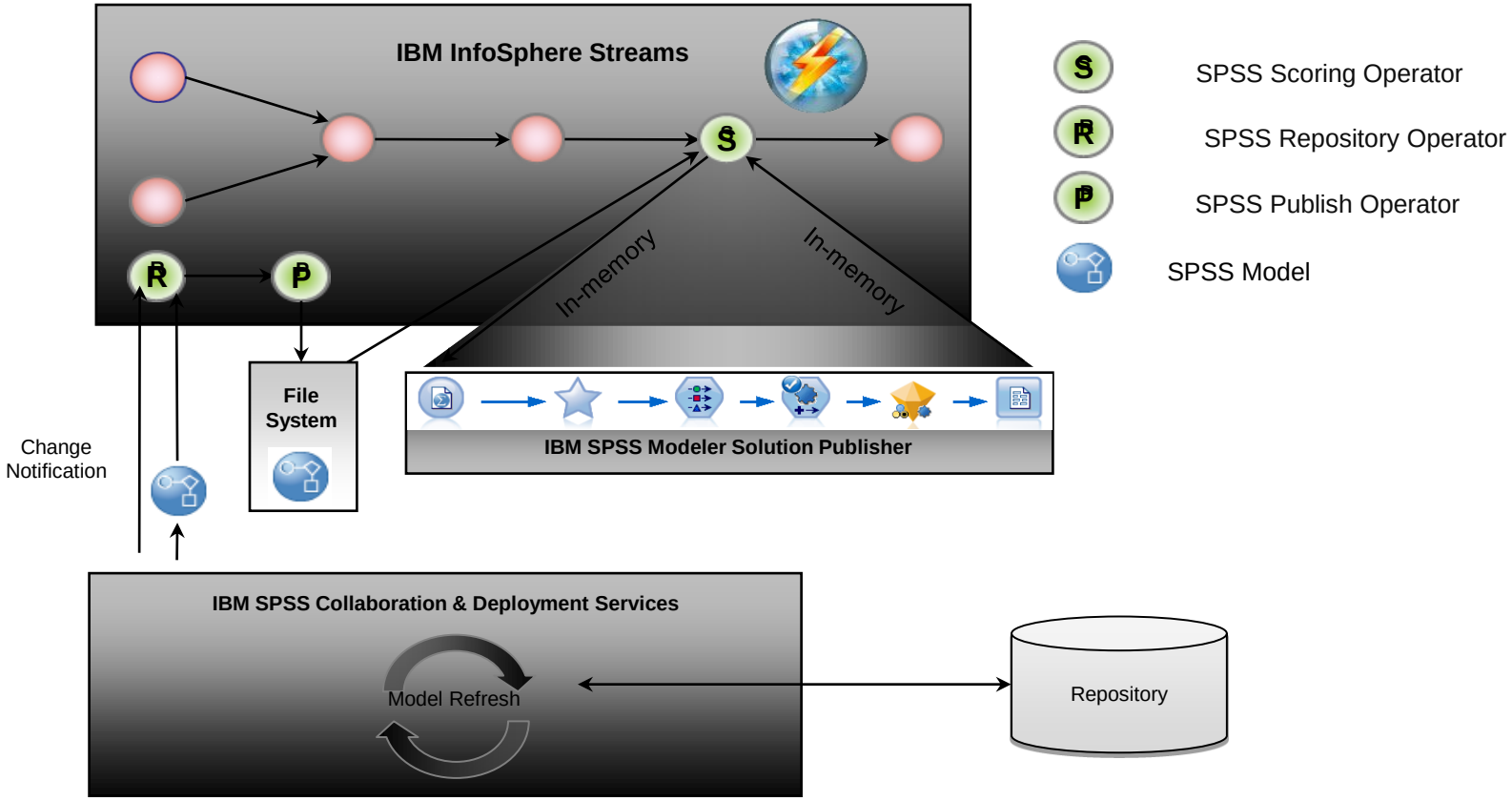




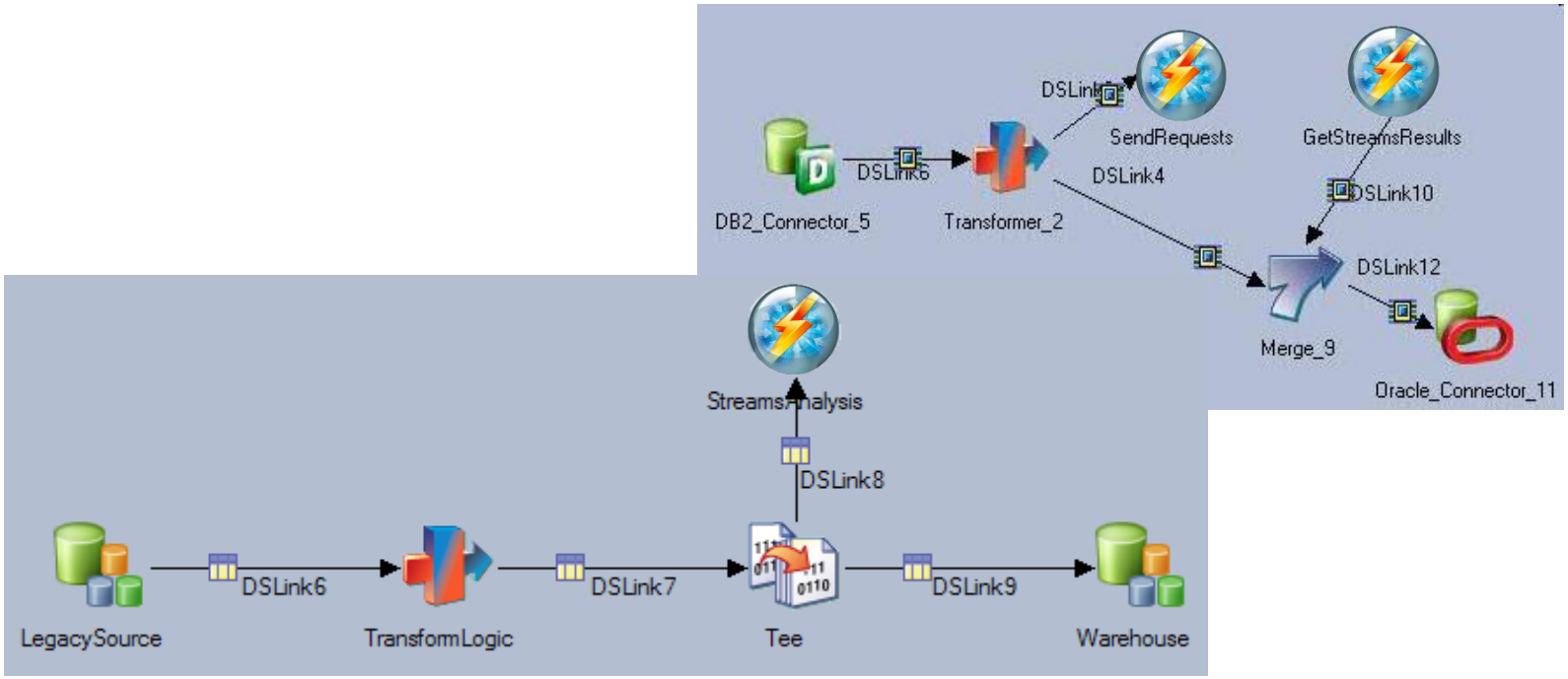
SPSS és Streams speciális kapcsolata

- SPSS modellek használata valós idejű döntéshozatalban
- SPSS Modeler generálta modellek közvetlen használata

- SPSS Modelek frissítése cseréje a Stream megállítása nélkül



IBM InfoSphere DataStage Integráció



- **Valós idejű feldolgozás és klasszikus ETL eszköz ötvözete**
  - Az adatáttöltés során „röptében” is tudunk elemezni
  - Adattárházat tudja analitikailag tehermentesíteni,
  - több napi riport
- **Streams ETL toolkit**
  - Streams and DataStage adatcsere adapterek
  - Integrációs kód

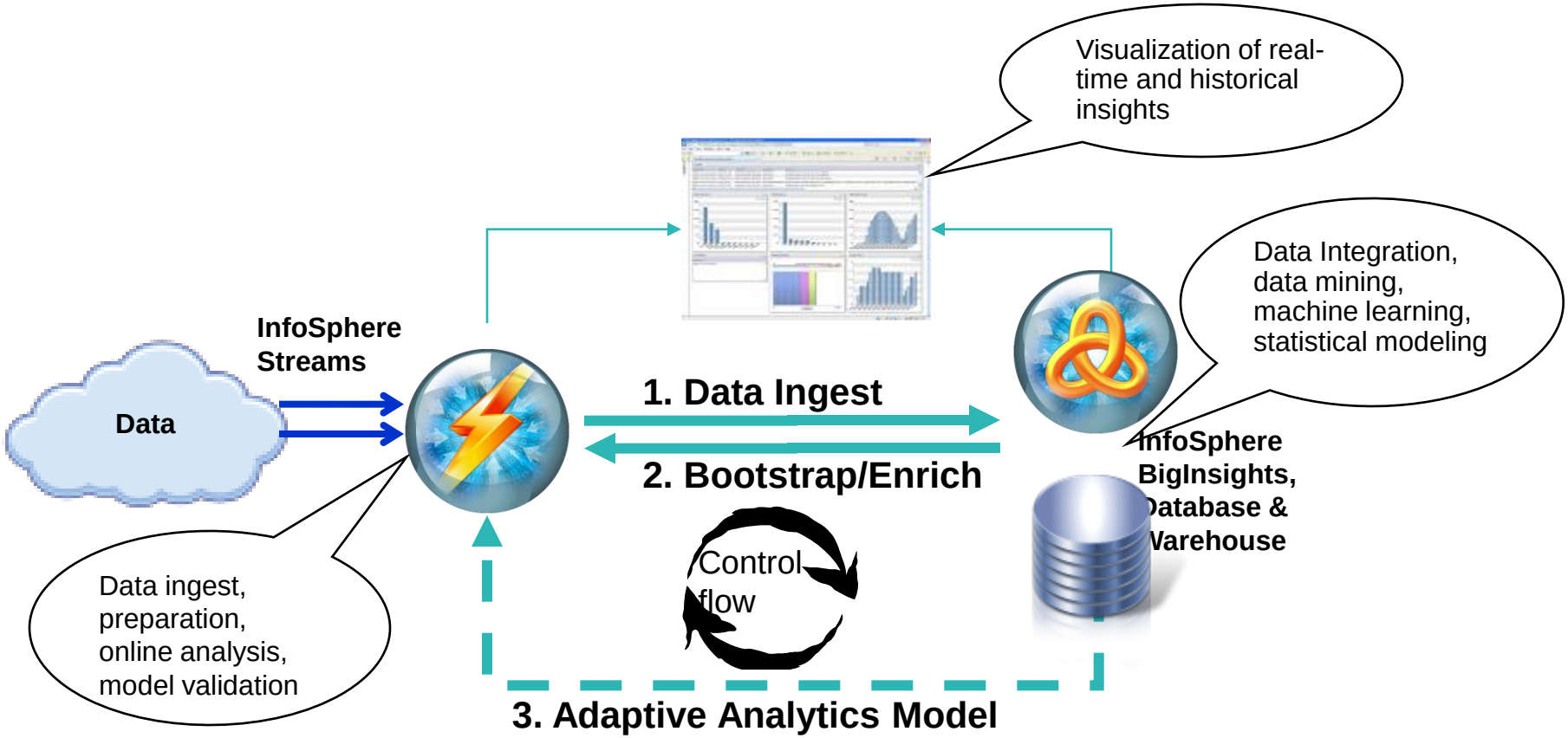
Streams Quick Start

- **Quick Start –ban elérhető**
  - **fejlesztőeszközök teljes készlete**
    - Grafikus editor, SPL nyelv
    - Adatvizualizáció
    - Vizuális monitoring
  - **Skálázható architektúra**
    - Elosztott platform
  - **Analitikai kiegészítések**
    - Time series analysis
    - Mining scoring
      - using PMML, R or SPSS
    - Complex Event Processing
    - Geospatial analysis
    - SPSS integráció

- **Nem elérhetők**
  - IBM InfoSphere BigInsights™ Enterprise Edition integráció
  - IBM DB2®
  - IBM Accelerator for Machine Data Analytics
  - IBM Accelerator for Social Data Analytics
  - IBM Accelerator for Telecommunications Event Data Analytics

21

Streams és BigInsights Interált folytonos feldolgozás és tárolás



Information Management

IBM

Data Explorer

Vizualizáció

September 9, 2013

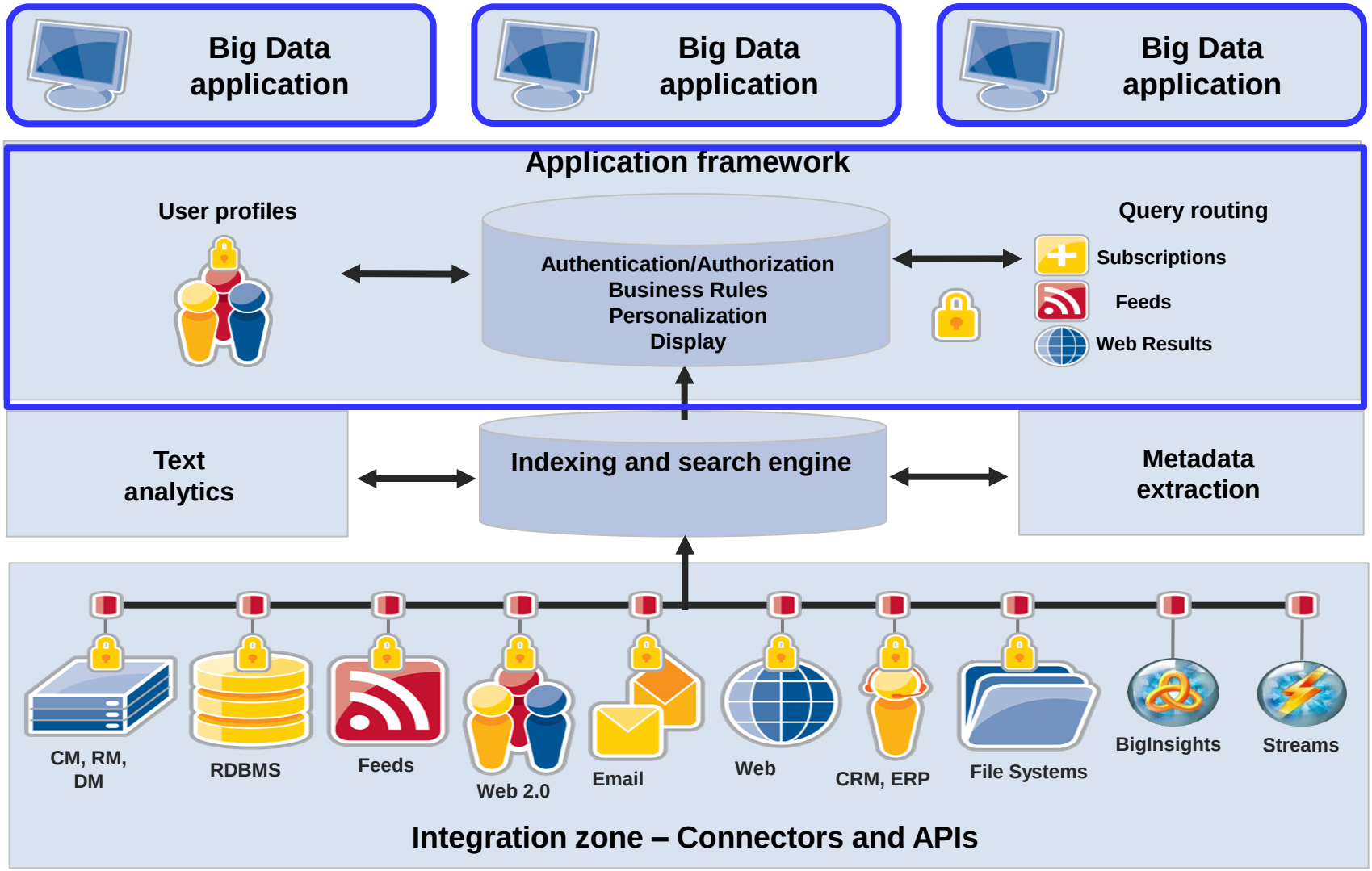
Data Explorer

Keresés, indexálás Adat vizualizáció

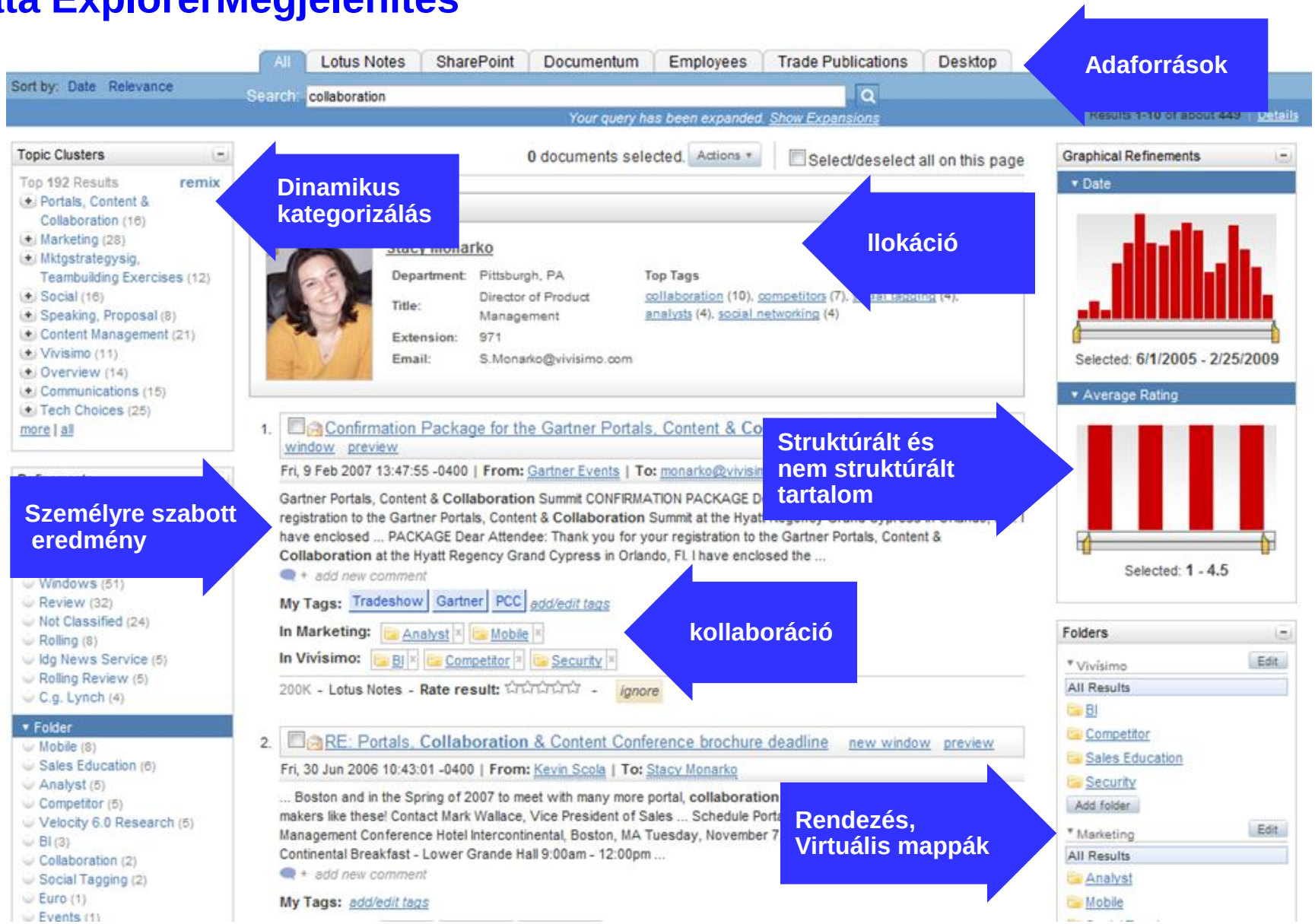
- **BigInsights, Streams, Adattárház, keimeneti adatainak webes fúziója**
- **A teljes képet mutatja minden kontextusban**

- **Sokféle adatforrásból származó adat egységes megjelenése**
- **Adatvagyon katalógus (glossary) szerinti csoportosítás**
- **Big Data Stratégi akezdő lépése**

InfoSphere Data Explorer Architektúra



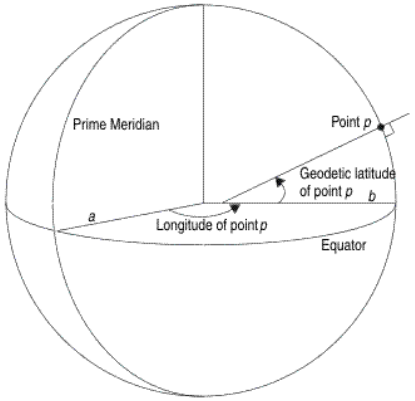
Data ExplorerMegjelenítés

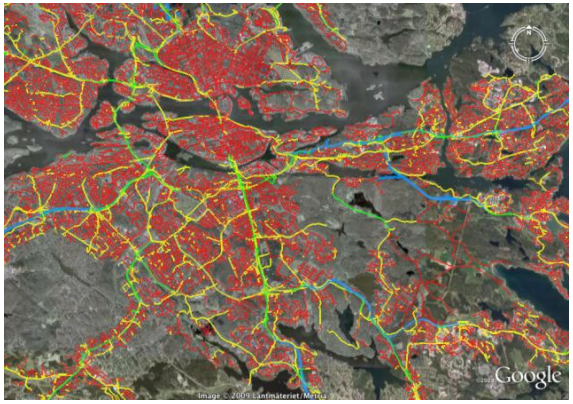


Egyedi értékek

Iparági megoldások

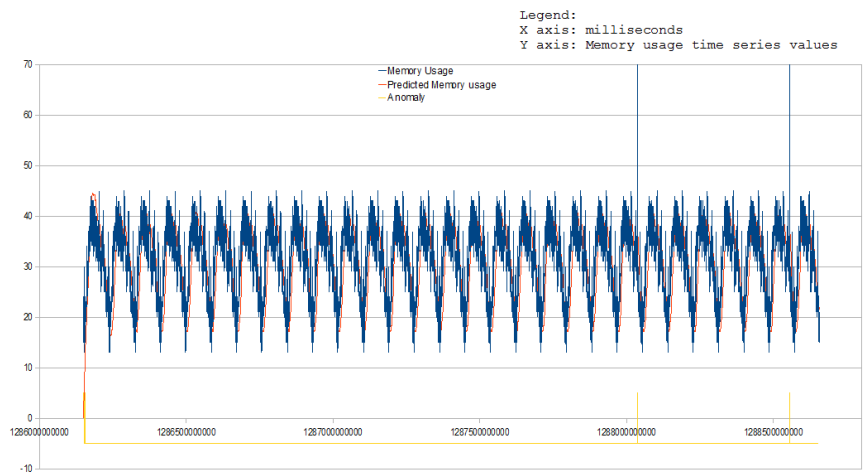
Geospatial Toolkit



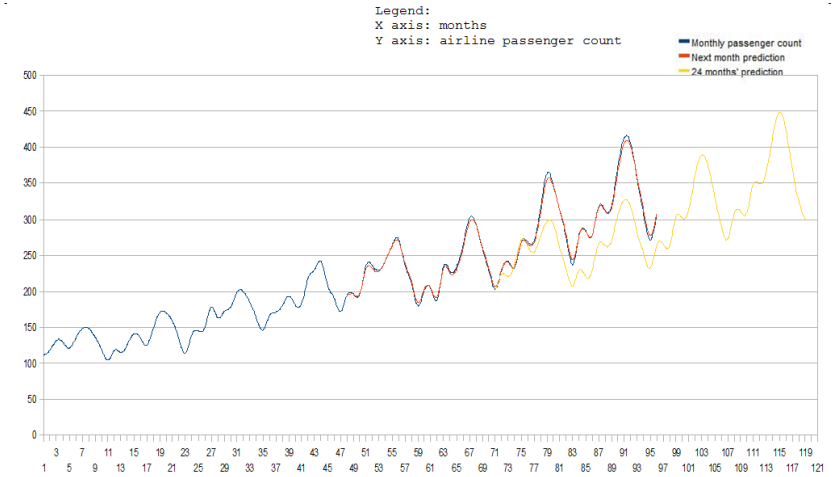


- **Nagyteljesítményű Térinformatikai modul**
- **Elosztott, rendszer, LoadBalance**
  - Smarter Transport
- **Térinformatikai adattípusok (Geospatial)**
  - e.g. Point, LineString, Polygon
- **Térinformatikai függvények**
  - e.g. Distance, Map point to LineString, isContained etc.

Time Series Toolkit



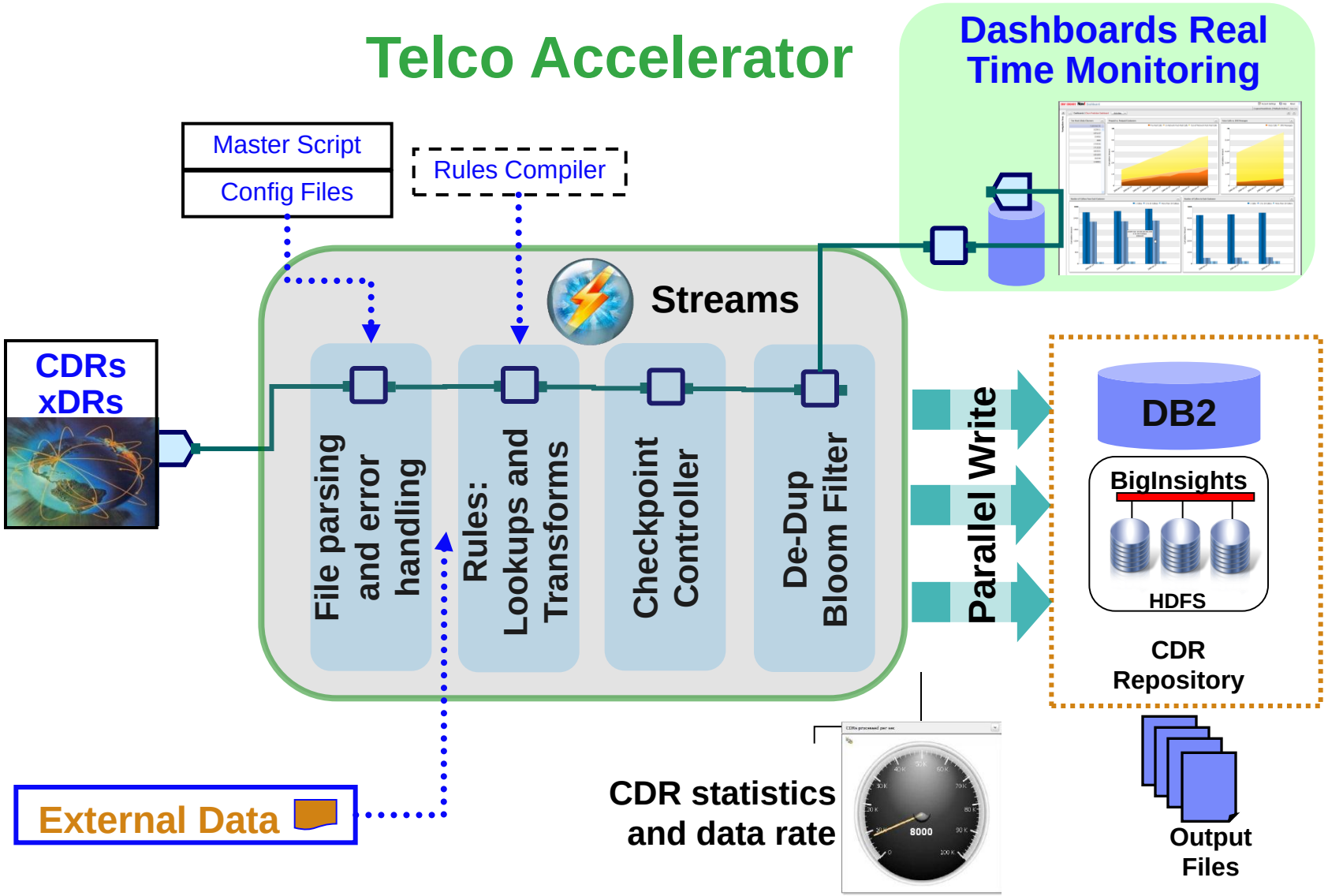
The time series simulates memory consumption from a computer . FMP is used for prediction and anomaly detection



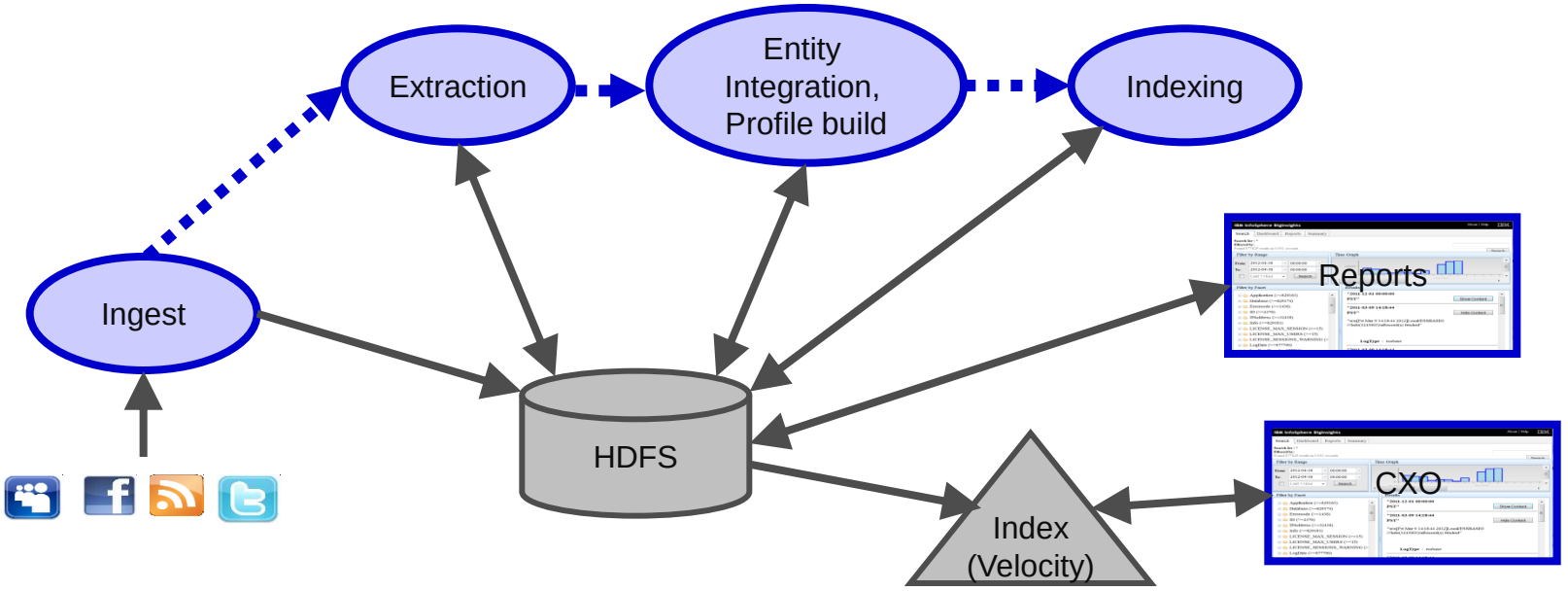
Holt Winters algorithm used for predicting next month and next 24 months ahead airline passengers count

- **Idősoros adatok elemzésére tervezve**
- **Gazdag funkciókészlet**
  - Adatsor generáció: függvény generátor
  - Feldolgozás : szűrés, aggregáció, mintavételezés (e.g. ReSample, Interpolate)
  - Analísis : korellációk anomáliák keresése
  - Modellezés : prediction, regression (e.g. Holt-Winters, GAMLearner)

Telco Accelerator



Social Media Accelerator



Integrated end user view

- **Üzleti lehetőségek feltárása**
- **Brand Management**
- **Mikroszegmentáció**
  - Személyes adatok
  - Érdeklődés,szokások, szosiális aktivitás, barátok,
- **Kimenet**
  - Sentiment analízis

Célhardverek

Pure Data for analytics
Pure data for hadoop

In October 2012

Adattárház Célgép

IBM Netezza átnevezve
IBM PureData System for Analytics

Adattárház Célgép

Igény ami életrehívta

- Adattárház teljesítmény igény
- DWH adminisztrációs költség csökkentés

Value statement

- **S**peed: 10 – 100x gyorsabb (mint alap tárház)
- **S**implicity: Alig igényel adminisztrációt (75% csökkentés)
- **S**calability
- **S**mart system
  - Adatbázison belüli párhuzamos analitika
  - Teljes SPSS integráció

Megoldás

- IBM Netezza *immáron:*
  - **PureData System for Analytics**


IBM big data • IBM big data • IBM big data
THINK
BIG
IBM big data • IBM big data • IBM big data
IBM big data • IBM big data
IBM big data • IBM big data

1 – Unlock Big Data

Customer need

- Understand existing data sources
- Search and navigate data within existing systems
    - **No copying of data**

Value statement

- Get up and running **quickly**
- Discover and retrieve big data
- Work even with big data sources – by business users

Solution

- Vivisimo Velocity *renamed to*
    - **IBM InfoSphere DataDiscovery**