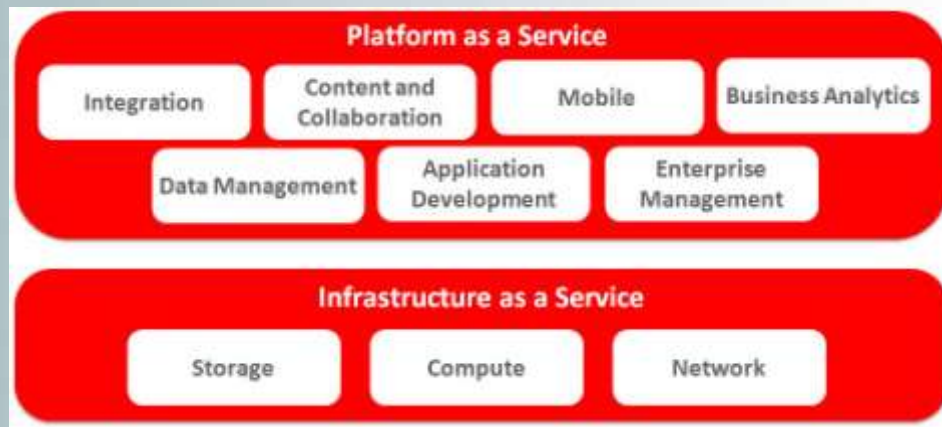


Adattárház = Aranytárház Adattárházak, Big Data és a dolgok 😊

Fekete Zoltán
principal presales consultant

2015. szeptember 10.

<https://blogs.oracle.com/zfekete/>



Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.



53%

**A vezetők fele
szerint túl sok
kritikus információ
érkezik későn.**

Source: Source: Delivering results: Growth and value in a volatile world, PwC 15th Annual Global CEO Survey, 2012

Információ/adat menedzsment modernizáció

Optimalizált adat menedzsment a valós-idejű betekintésig



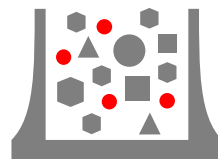
**Tároló
optimalizálás**



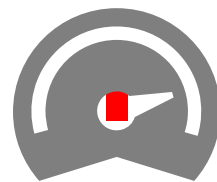
**Biztonság
növelése**



**Hozzáférés
kiterjesztés**



**Adatforrások
szélesítése**



**Gyorsabb
elemzések**



Around-the-clock
Availability



Consumer
Expectations



Real-time
Fulfillment



Flawless
Engagement



Personalized
Service



Information any
time, Anywhere



Insight

Digital Changes Everything



Employee
Expectations



Business Model



Speed to Market

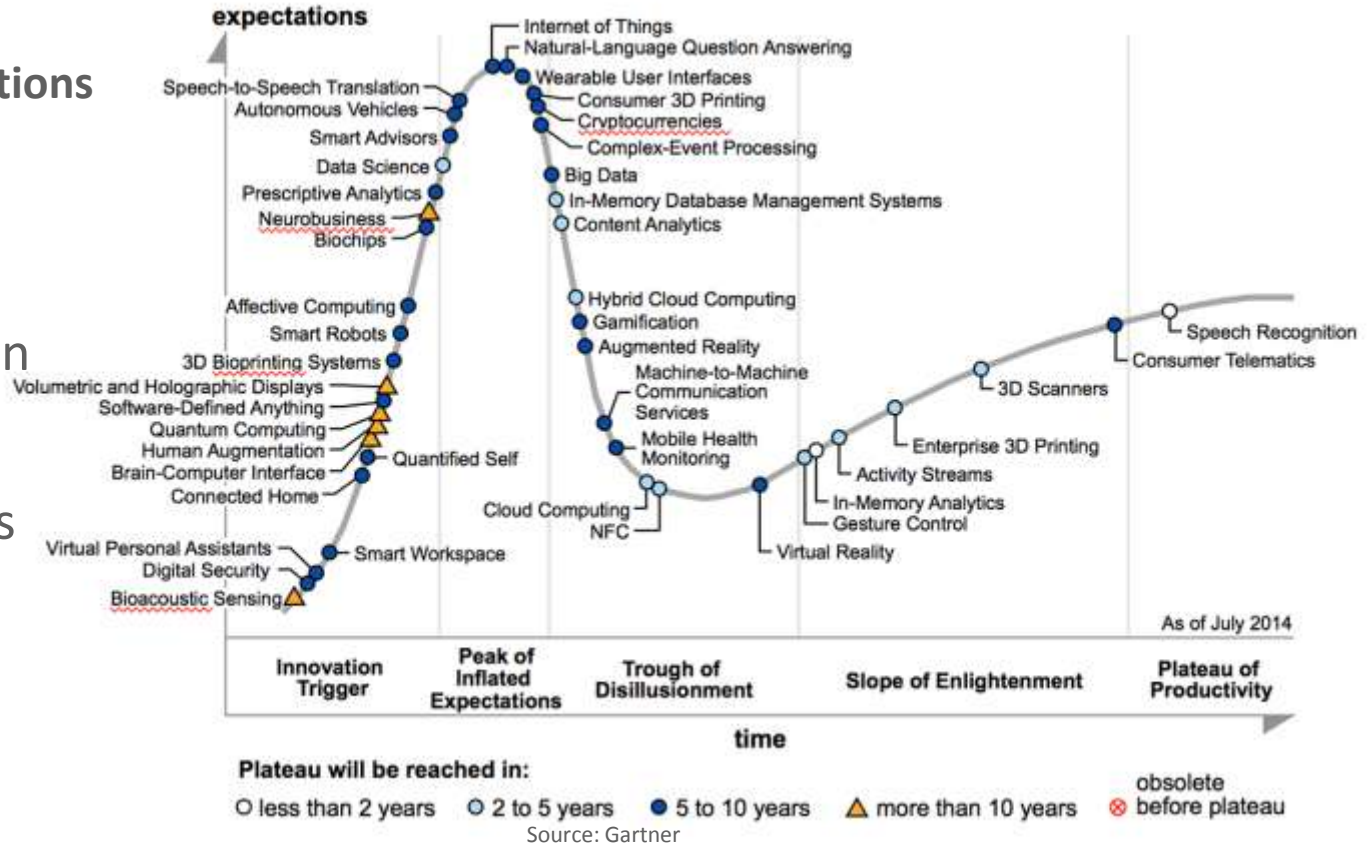


New Markets

Emerging Technologies

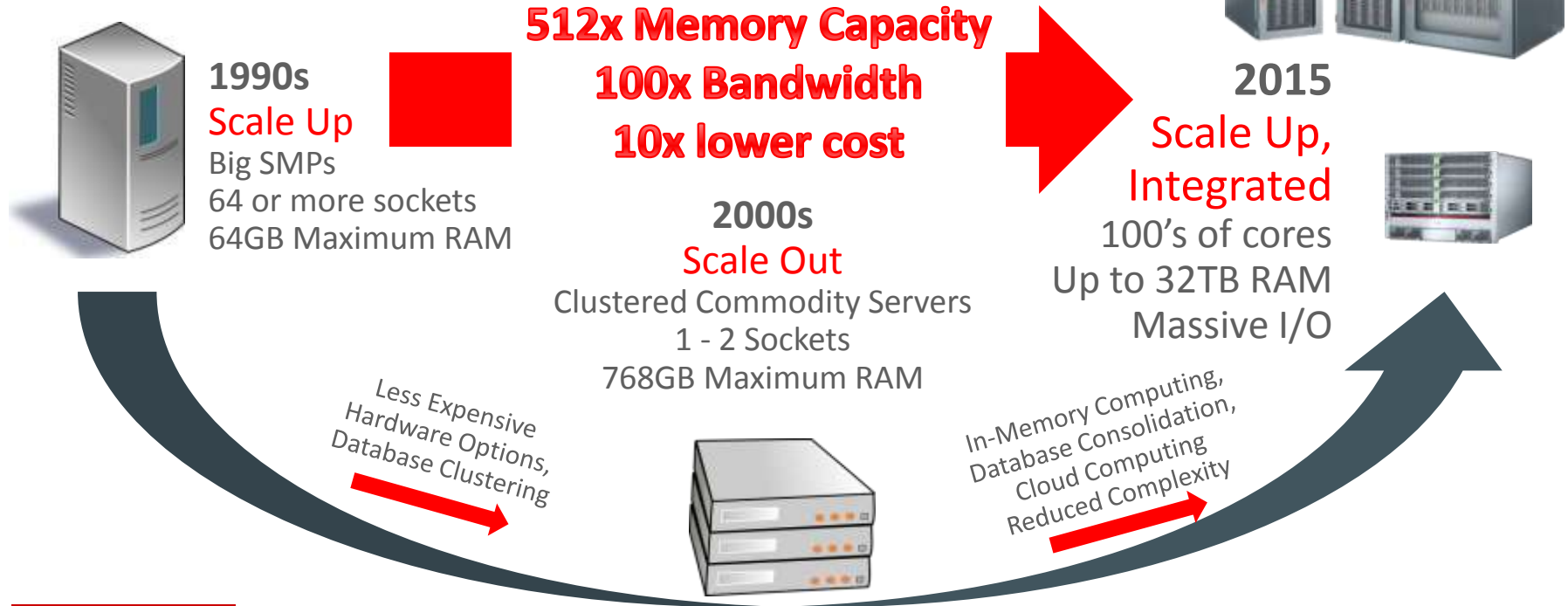
Enable IT Transformations

- Big Data
- In Memory
- Software in Silicon
- Hybrid Cloud
- Internet of Things



Integrated Systems

In-Memory Computing and Consolidation



Oracle Strategy

Complete Stack

- Best-of-Breed
- Open
- Vertical Integration
- Extreme Performance
- Engineered Systems



Complete Choice

- On-Premise
- Private Cloud
- Public Cloud
- Hybrid Cloud

Why Integrated Systems?

Oracle Engineered Systems can benefit Big Data, Cloud, Database and Applications

FIGURE 1

Typical Annual Benefits for a 1,000-User Organization
Using an Integrated System



Source: IDC, 2015

(\$ per 1,000-user organization)



1 **NEW**
DEFINITION
IS ADDED ON
urban

1,600+
READS ON
Scribd

13,000+ HOURS
MUSIC
STREAMING ON
PANDORA

12,000+
NEW ADS
POSTED ON
craigslist

370,000+ MINUTES
VOICE CALLS ON
skype

98,000+
TWEETS

320+
NEW
twitter
ACCOUNTS

100+
NEW
Linked in
ACCOUNTS

1 **NEW**
ARTICLE IS
PUBLISHED
associatedcontent

6,600+
NEW
PICTURES ARE
UPLOADED ON
flickr

50+
WORDPRESS
DOWNLOADS

695,000+
facebook
STATUS
UPDATES

125+
PLUGIN
DOWNLOADS

79,364
WALL
POSTS

510,040
COMMENTS

1,700+
Firefox
DOWNLOADS

694,445
SEARCH
QUERIES

168 MILLION
EMAILS
ARE SENT

60+
NEW
BLOGS

1,500+
BLOG
POSTS

70+
DOMAINS
REGISTERED

600+
NEW
VIDEOS

100+
Answers.com
40+
YAHOO! Answers

QUESTIONS
ASKED ON THE
INTERNET...

25+ HOURS
TOTAL
DURATION

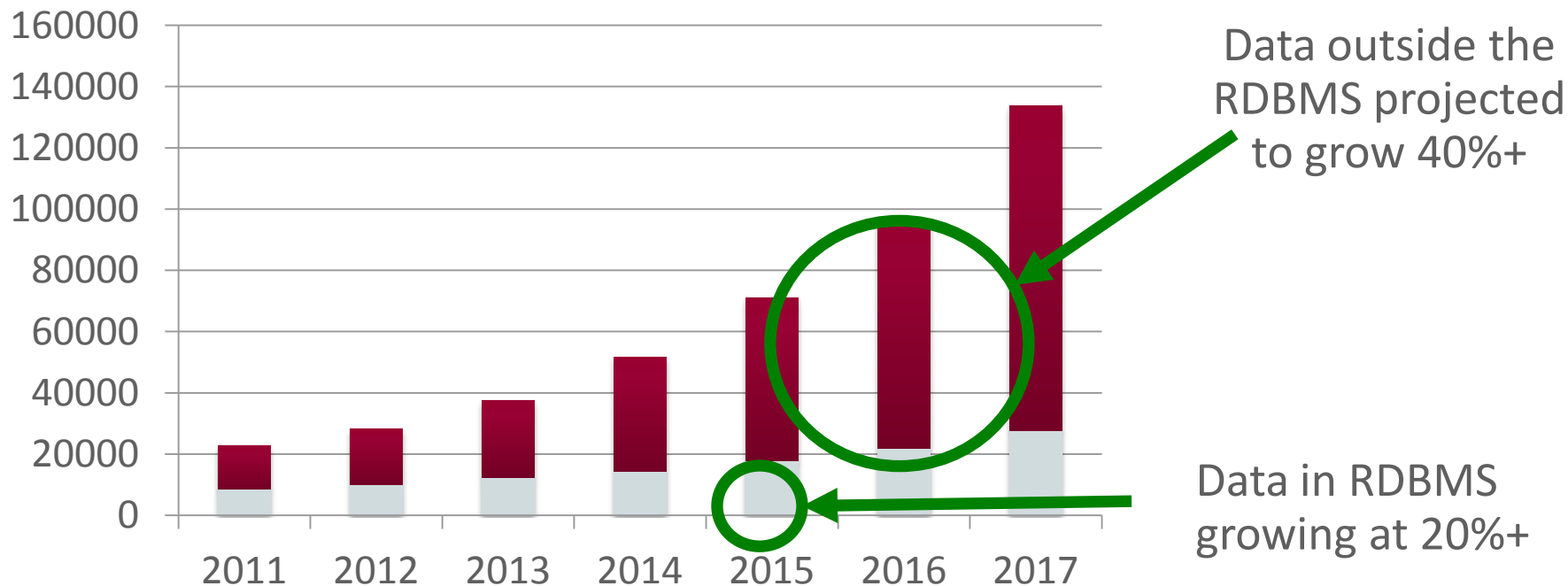


And Future Growth is Staggering



Több adatot tárolnak az RDBMS-en kívül

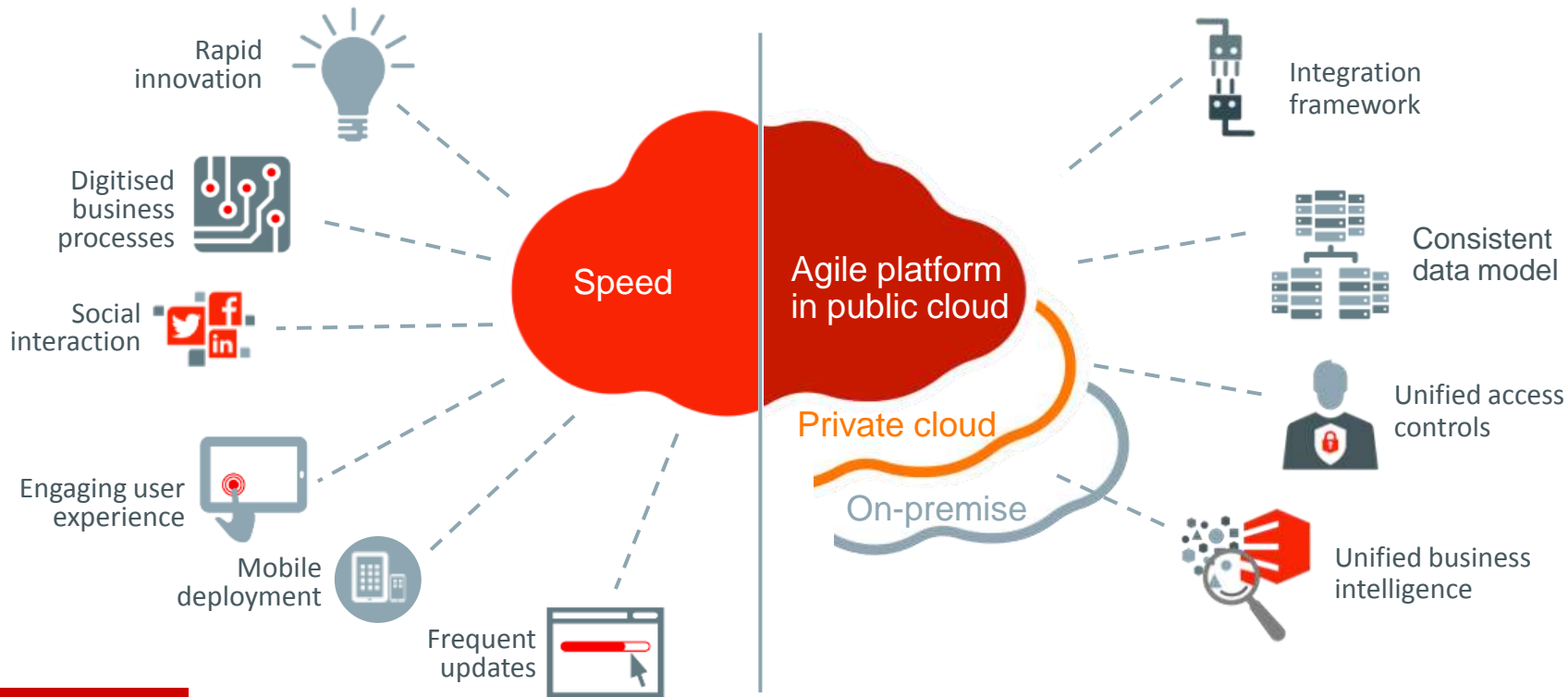
Big Data = Unstructured and Structured data combined



Source IDC, March 2014, Ashish Nadkarni & Natalya Yezhkova, ["Structured Versus Unstructured Data: The Balance of Power Continues to Shift"](#)

Integrated Cloud

Your Cloud, Oracle Cloud, Hybrid Cloud



Make a Leap Step Forward with Engineered Systems

Take Advantage of IT Trends

**Extreme
Performance**

**Low Risk
Deployment**

**Breakthrough
Efficiency**

Folyamatos Oracle Database innováció

Megőrizve az ügyfelek befektetéseit minden korszakban



Client-Server

Stored Procedures
Partitioning
Parallel Query
Unstructured Data



Internet

Resource Management
Real Application Clusters
Data Guard
XML

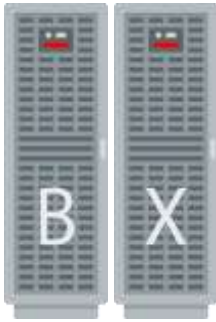


Big Data & Cloud

In-Memory
Multitenant
Big Data SQL
JSON

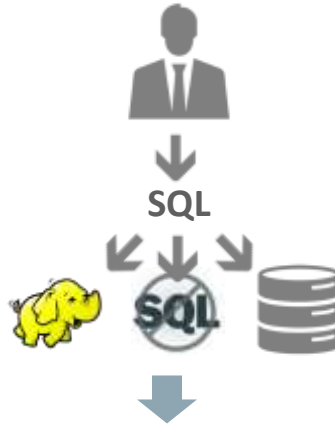
Az új Big Data technológiák **könnyebb bevezetése**

INTEGRATION



**Engineered
Systems**

SKILLS



**SQL on
All Data**

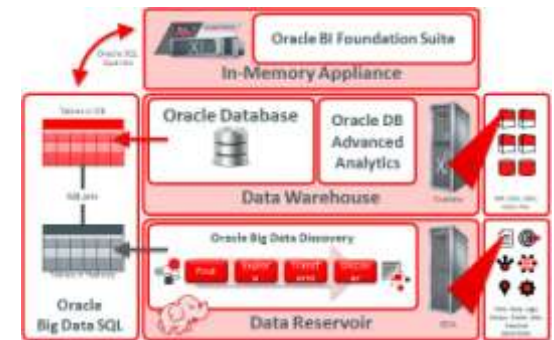
SECURITY



**Database
Security on
All Data**

Oracle Big Data

- 1 Oracle Big Data architektúra és a relációs világ: összekapcsolva
- 2 Oracle adatintegráció a Big Data világban
- 3 Oracle Big Data SQL és demó
- 4 Oracle Big Data Discovery
- 5 Oracle Advanced Analytics: R és Data Mining
- 6 Ügyfél felhasználási esetek és architektúrák



Adattárház definíciók: Inmon vs Kimball

- **Bill Inmon:**

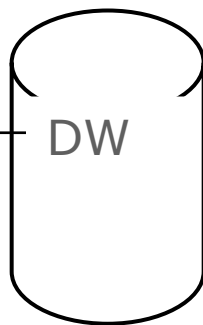
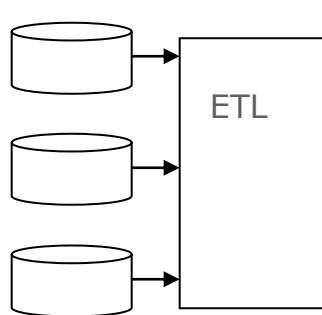
- Az adattárház egy általános BI rendszer
- Egy cégnek csak egy adattárháza van és az adatpiacok ebből kapják az inputot
- Az adattárházban az adatok 3. normálformában vannak

- **Ralph Kimball:**

- Az adattárház a cég összes adatpiacának összessége
- Az adat (információ) dimenzionális modellben van tárolva

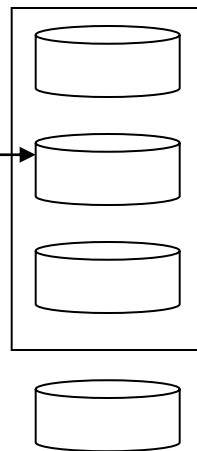
Inmon CIF architektúrájának leegyszerűsített nézete

Operatív
rendszerek



Integrált elemi
adatok
3NF formátumban

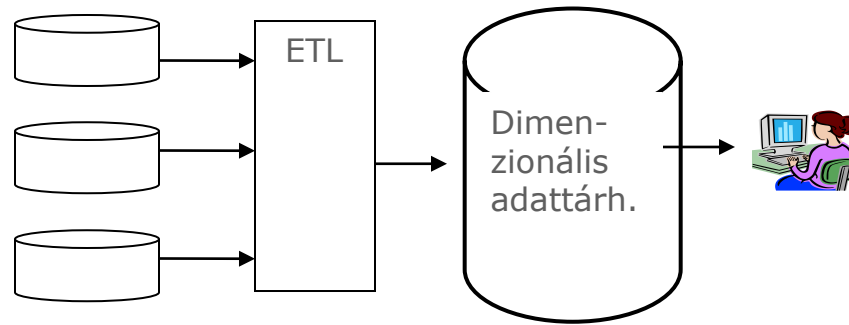
Adatpiacok és egyéb
alkalmazások



Adatpiacok
dimenzionális
formátumban

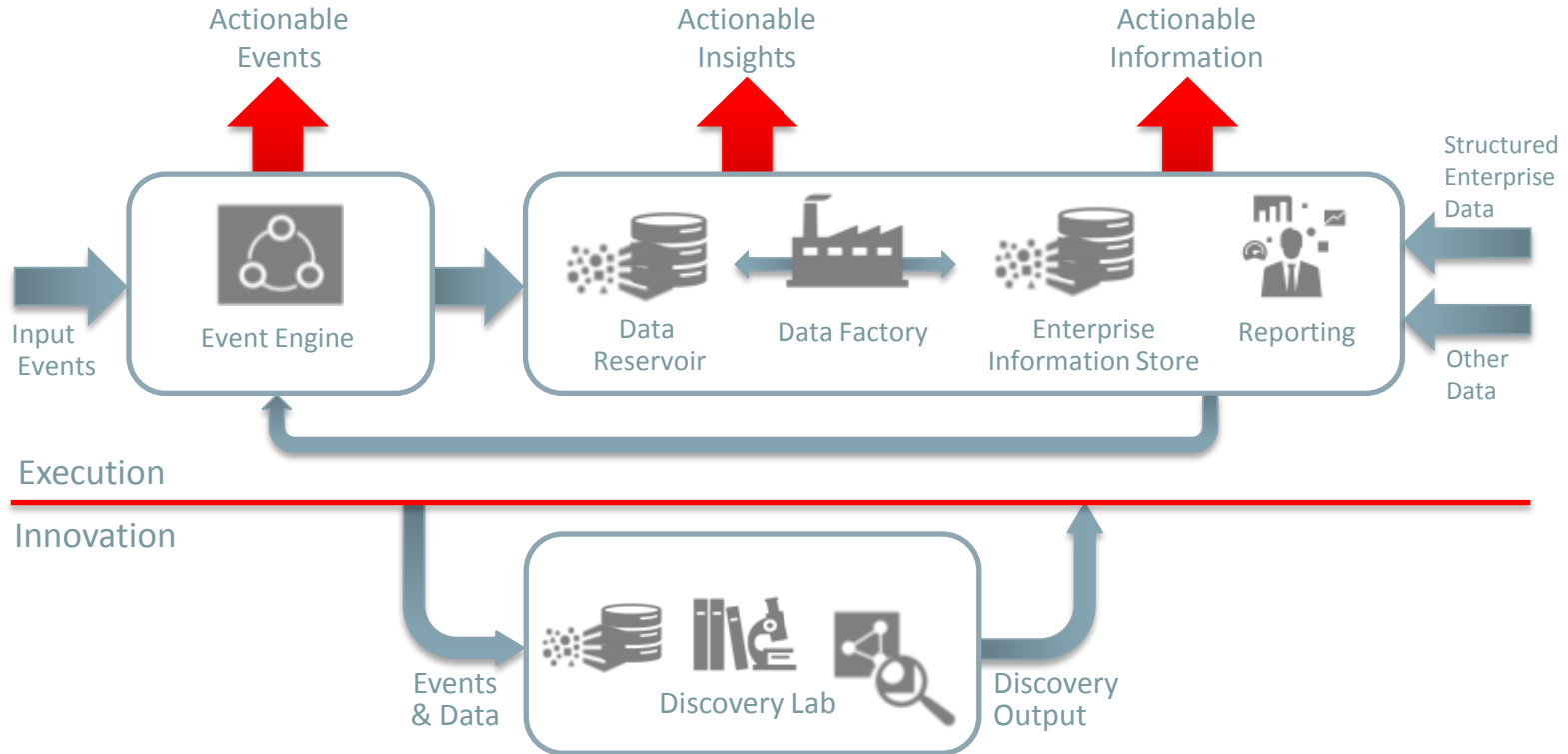


Kimball architektúrájának leegyszerűsített nézete



Integrált elemi adatok
dimenziális formátumban

Conceptual View



Component Outline

Data Engine	Respond to R/T events in appropriate and/or optimised fashion
Data Reservoir	Raw data Reservoir – typically event data at lowest grain
Data Factory	Managed ETL onto, within and between platforms
Enterprise Data	Data stores for Information Management
Reporting	BI tools and infrastructure components
Discovery Lab	Platform, data and tools to support discovery process

Execution – things you do every day

Innovation – innovation to drive tomorrows business

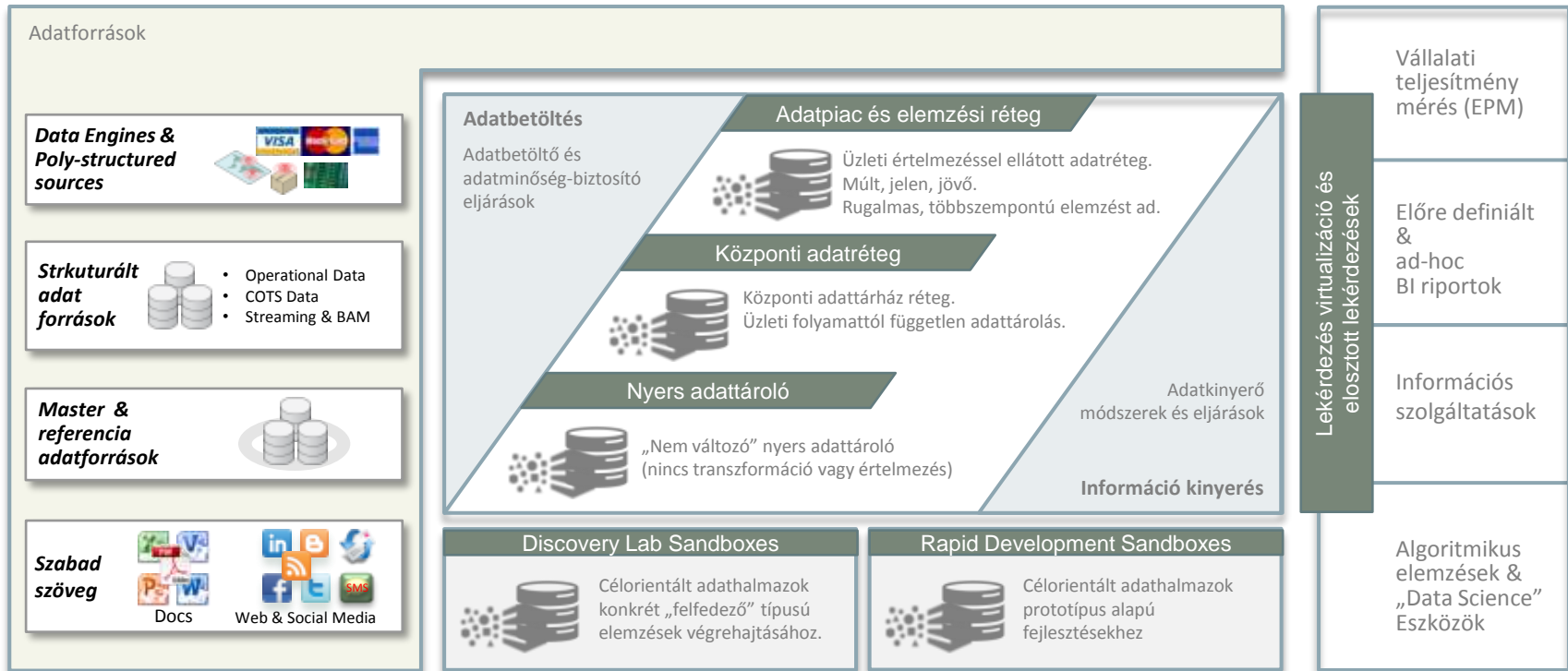
← Line of Governance!



– Possible outputs include new knowledge, mining models / parameters, scored data...

Information Management – Logikai nézet

„Data Reservoir” & Enterprise Information Store – teljes nézet



Egyensúly: a gyors fejlesztés / modellezés és elemzés között

Big Data és a relációs világ, jellemzők

Hadoop



- Adatok file-okban, MapReduce
- „Schema on read”
- Egyszerű programozási modell, nagy adatmennyiségekhez is
- Csak hozzáfűzés
- Blokkok soros elérése

NoSQL



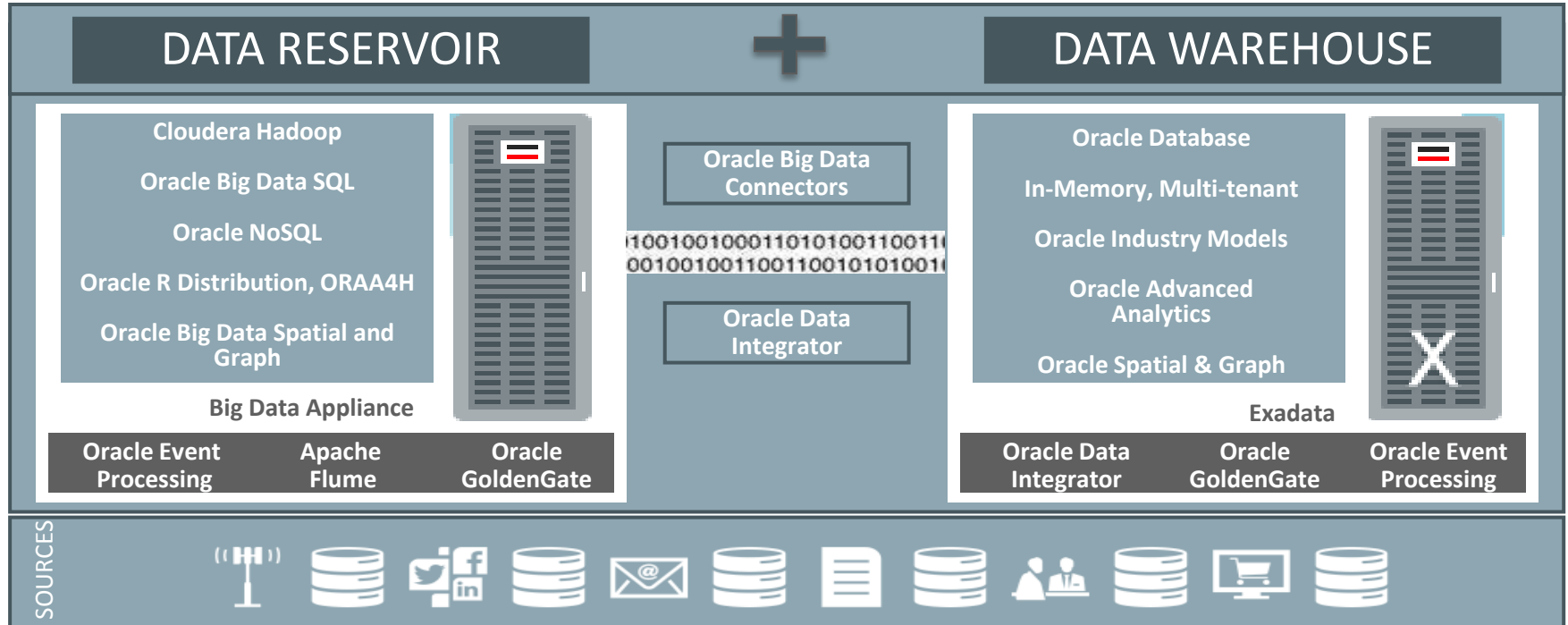
- Kulcs – érték párok; „BASE”
- Séma nélküli egyszerű modell
- Költséghatékony egyszerű lekérdezések
- Lineáris skálázhatóság
- Replikáció, sharding



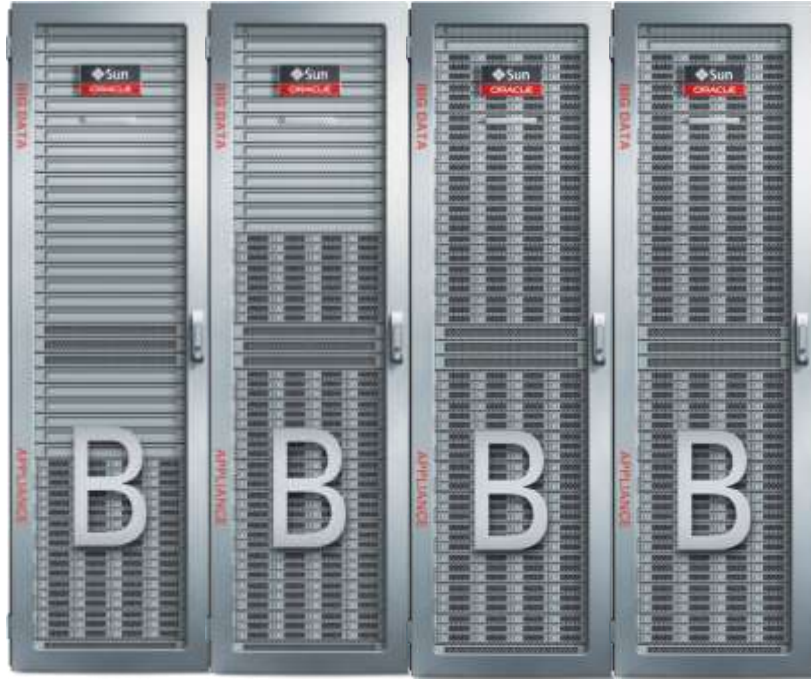
Relációs

- Szervezés: gyors lekérdezés
- Strukturált sémák
- Komplex programozási modellek
- Írás, olvasás, törlés, update
- Rekordok egyszerű elérése

The Big Picture – Oracle Big Data Management System



Oracle Big Data Appliance: Standard és moduláris



- Starter Rack: 6 node
- Elastic: node-onként bővíthető
- Akár 18 rackig egyszerűen összekábelezhető

Lekörözi a DIY clustereket:

- Bekerülési költség kisebb
- Gyorsabb bevezetés
- Nagyobb teljesítmény
- Skálázhatóbb
- Exadata kapcsolat
- Enterprise Manager framework

Big Data Appliance X5-2

Sun Oracle X5-2L Servers with **per server**:

- 2 * 18 Core Intel Xeon E5 Processors
- 128 GB Memory
- 12 * 8 TB = 96 TB Disk space

Integrated Software (4.1):

- Oracle Linux 6.5, Oracle JDK 7u72
- **Oracle Big Data SQL 1.1***
- **Cloudera Distribution of Apache Hadoop 5.3 – EDH Edition**
- **Cloudera Manager 5.3**
- **Oracle R Distribution 3.1.1-2**
- **Oracle NoSQL Database CE 3.2.4**
- ...

* Oracle Big Data SQL is separately licensed



Oracle Big Data Appliance kiforrott, bizonyított, **költséghatékony** megoldás

“Oracle Big Data Appliance is an excellent choice for customers looking to work with the full suite of Cloudera’s leading Hadoop-based technology. It’s **more cost-effective and quicker to deploy** than a DIY cluster.”

- Mike Olson, Cloudera founder, Chief Strategy Officer, and Chairman of the Board

cloudera



21%
Lower Costs



33%
Faster to Implement

Source: [ESG White Paper](#)



Oracle Big Data Light Virtual Machine 4.2

Ingyenes kipróbálási lehetőség, letölthető virtuális gép, adatokkal, gyakorlatokkal

- [Oracle Big Data Learning Library](#)
- [Analyze All Your Data with Oracle Big Data SQL](#)
- [Introduction to Oracle NoSQL Database](#)
- [Oracle NoSQL Database - Installation/cluster topology deployment \(pdf, scripts\)](#)
- [Access Data in Oracle NoSQL Database from Oracle Database](#)
- [Data Manipulation with Hive and Pig \(pdf | script\)](#)
- [Tame Big Data with Oracle Data Integration](#)
- [Integrate Hadoop Data with Oracle Database using Oracle Big Data Connectors](#)
- [Using SQL Pattern Matching](#)
- [Oracle Data Mining 12c Tutorial Series](#)
- [Oracle R Enterprise v 1.4 - Tutorial Series](#)

<http://www.oracle.com/technetwork/database/bigdata-appliance/oracle-bigdatalite-2104726.html>

Oracle Cloud Platform and Infrastructure

Complete Suite of Services for Building Modern Secure Cloud Applications

Platform as a Service

Integration

Content and
Collaboration

Mobile

Business Analytics

Data Management

Application
Development

Enterprise
Management

Infrastructure as a Service

Storage

Compute

Network

Nemsokára: Oracle Exadata Cloud Service és Oracle Big Data Appliance Cloud Service



- Exadata és Big Data Appliance rack-ek az Oracle Cloud-ban
- Customer is provided access to physical Exadata rack
 - Quarter Rack increments for Exadata
 - 6-node increments for Big Data Appliance
- Tenant Configuration isolated on separate VLAN
 - Prevent security breaches
 - InfiniBand will use IB partitioning
- Access via Secure Network
- DB provisioning + configuration based on Customer requirements

Oracle Big Data Cloud Service **Big Data Appliance Starter Racktől (6 szerver)**



Provisioning & Elasticity

On-demand

Provision fully configured Hadoop clusters on demand. Choose compute shapes and number of nodes based on your needs

Elastic

Scale worker nodes based on compute requirements



Comprehensive Software

Big Data Connectors

Seamlessly integrate with other Oracle services using big data connectors

Big Data Spatial and Graph

Simply process geospatial and graph data using Oracle's advanced capabilities



Management

Integrated security

Integrate with your enterprise security using pre-configured Kerberos authentication, LDAP based authorization, and robust centralized auditing

Automated Lifecycle

Manage clusters yourself with easy-to-use tooling and simplified patching and upgrades



Integration

Object Store Integration

Store all your data in the Oracle Storage Cloud Service, and process it with multiple Hadoop clusters

Big Data SQL

Seamlessly query data in Big Data Cloud Service and Database Cloud –Exadata Service

Oracle Big Data Cloud Service

- Starter Pack – 6 nodes
- additional nodes
- Dedicated instances in the Oracle Cloud, featuring 40 Gb/sec InfiniBand fabric inside Big Data Cloud Service and Database Cloud Service -- Exadata Edition.
- Simplified operations, and automated lifecycle management through a single command utility of the entire stack

Oracle Big Data Cloud Service

- Cloudera's comprehensive software suite including Cloudera Distribution including Apache Hadoop and Apache Spark
- **Big Data Connectors** delivers load rates of up to **15TB per hour** between
Big Data Cloud Service and Oracle Exadata Cloud Service
- **Big Data Spatial and Graph** provides cutting-edge tools for exploring and analyzing massive graphs and geo-locational data
- **Oracle Data Integrator** and **ODI Big Data** Module to simplify big data

Oracle Big Data Cloud Service



ORACLE[®] CLOUD My Services

nsa | ashok.shivarudraiah@... ▼

Instances

Users

Notifications



Oracle Big Data Cloud Service

Started Instances

1 of 1

Service Instances

Nodes

6

Memory

1.5 TB

HDFS Storage

288.0 TB

As of Jun 19, 2015 5:30:49 AM UTC ↺

all metrics
will be 0

no clusters in
beginning

Create instance

Create Instance



productionCa
READY

Version: 4.2.1

No of Nodes: 6

Created On: 2015-06-18T22:13:27+0000

Entitlement Id:

17C3A0BD4F5A51F3E050F40A04CE6185

Kerberos:



Live Chat | Contact Us

Oracle Big Data SQL Cloud Service

- Leverage existing SQL skills and tools to
 - Combine data from Oracle Database Cloud Service, Hadoop and NoSQL in a single SQL query
 - Query and analyze data in Hadoop and NoSQL
- Quickly integrate big data analysis into existing applications and architectures
- Easily extend security and access policies from Oracle database to data in Hadoop and NoSQL
- Get maximum query performance on all data using Smart Scan

BDPaaS

Big Data Preparation as a Service



Ingest

Import and Ingest

Import and ingestion of structured, semi-structured, and unstructured content from various sources

Cleanse and Normalize

Cleansing data such as removing non-essential characters, normalizing dates and phone numbers

Repair and Standardize

Identifying and fixing where possible inconsistencies in the data

Identify Duplicates

Powerful statistical similarity engine identifies duplicates in data



Enrich

Profile

Authoring process is assisted by statistical analysis of numerical data and frequency and term analysis of text data

Classify and Extract

Identifying categories in the data and identify characteristics of the data in terms of attributes, properties, schemata

Augment and Enrich

Knowledge based enrichments based on discovered relationships

Discover Metadata

Schema detection identifies the schema/metadata that is explicitly or implicitly defined in headers, fields, tags, and other information



Govern and Publish

Dashboard

Interactive dashboard provides a unified timeline of all processed datasets with associated metrics with drill-down capabilities for detailed auditing and analysis

Scheduling

Continuous pipeline execution support via scheduling service

Publishing Flexibility

Various cloud targets and publishing formats provide maximum flexibility



Intuitive Authoring

Recommendations

Machine Learning driven recommendations provide a highly intuitive and interactive user experience, guiding users thru every step in the authoring process

Visualizations

Interactive visualizations provide users with real-time previews of profile metrics and transformations on their datasets

Business User Friendly

Script free environment combined with videos and tutorials for ease of use

Big Data Discovery as a Service



Find

Catalog

Builds a rich, interactive catalog on all data in Hadoop

Browse

Familiar search and guided navigation for ease of use

Provision

Add personal or enterprise data to Hadoop via self-service



Explore

Assess

Understand the shape and quality of data

Correlate

Experiment with attribute combinations to understand correlations

Evaluate

Know if data set is worthy of further investment



Transform

Wrangle

Intuitive, user driven data wrangling

Improve

Leverage extensive library of data transformations

Enrich

Make the data better with location, language detection, text mining and classification



Discover

Blend

Combine data sets for deeper perspectives

Compose

Drag and drop discovery dashboard composition

Analyze

Interactive visual analytics, powerful search, intuitive navigation



Share

Collaborate

Share projects, bookmarks and snapshots with the team

Publish

Export transformed, enriched and blended data back to HDFS

Leverage

Use in any other product that talks to Hadoop

Oracle Big Data

1 Oracle Big Data architektúra és a relációs világ: összekapcsolva

2 Oracle adatintegráció a Big Data világban

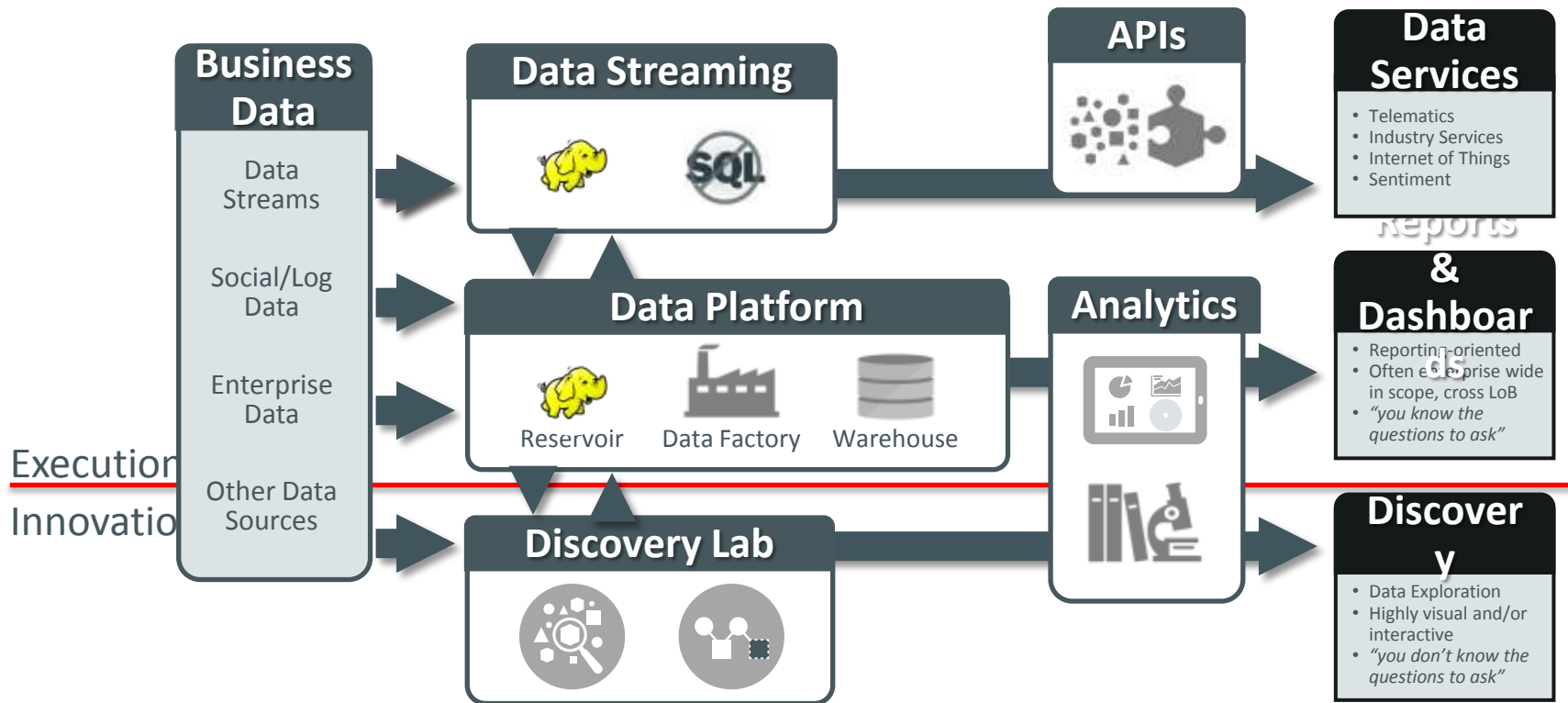
3 Oracle Big Data SQL és demó

4 Oracle Big Data Discovery

5 Oracle Advanced Analytics: R és Data Mining

6 Ügyfél felhasználási esetek és architektúrák

4th Generation **Data Architecture** for Big Data



Comprehensive Integration and Governance

Real-Time Data Movement

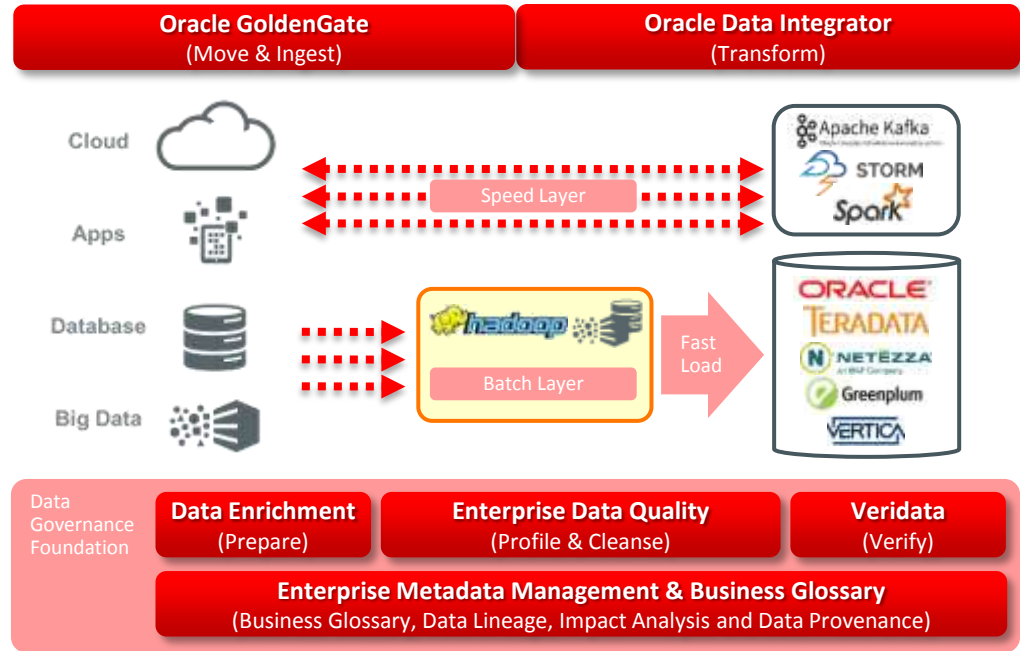
- Low impact capture, stage in Hadoop
- Continuous data availability

Data Transformation

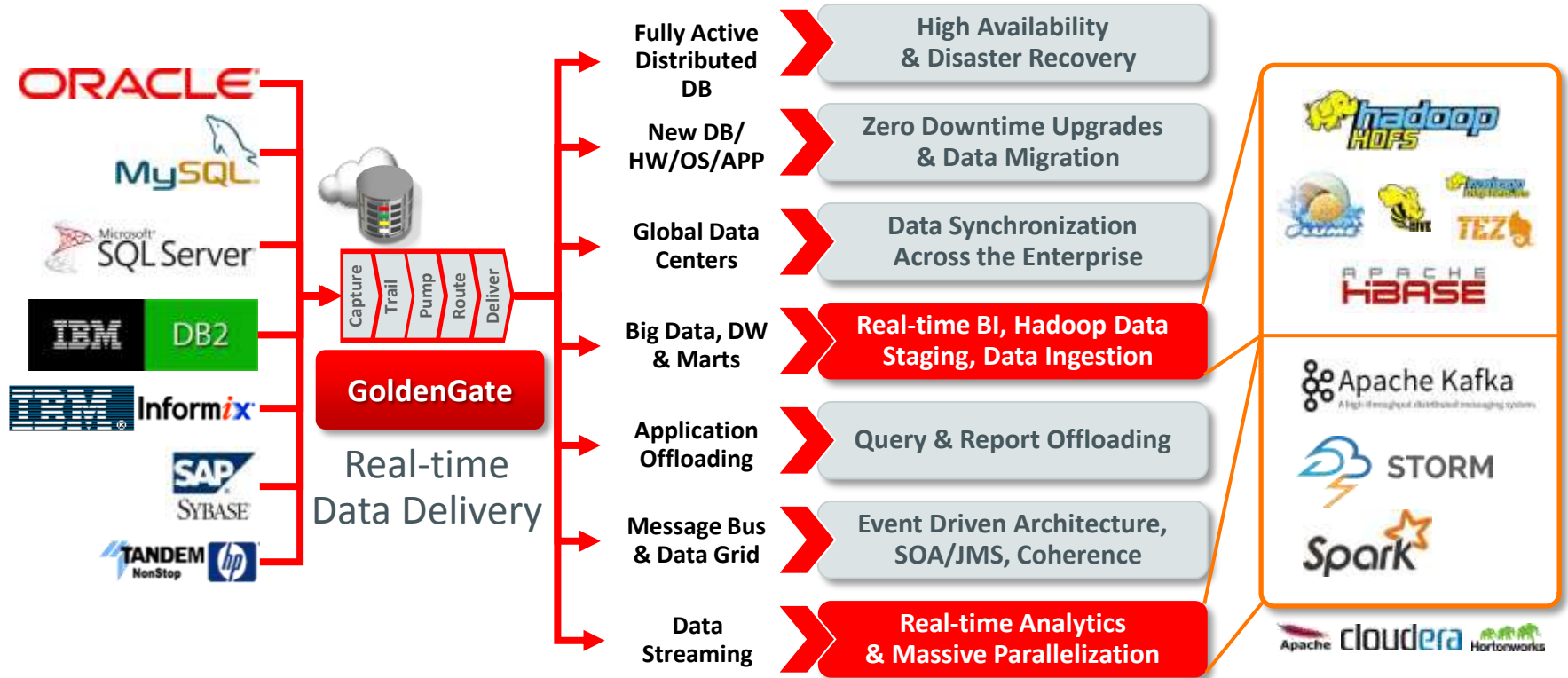
- Bulk data movement
- Pushdown data processing

Data Governance

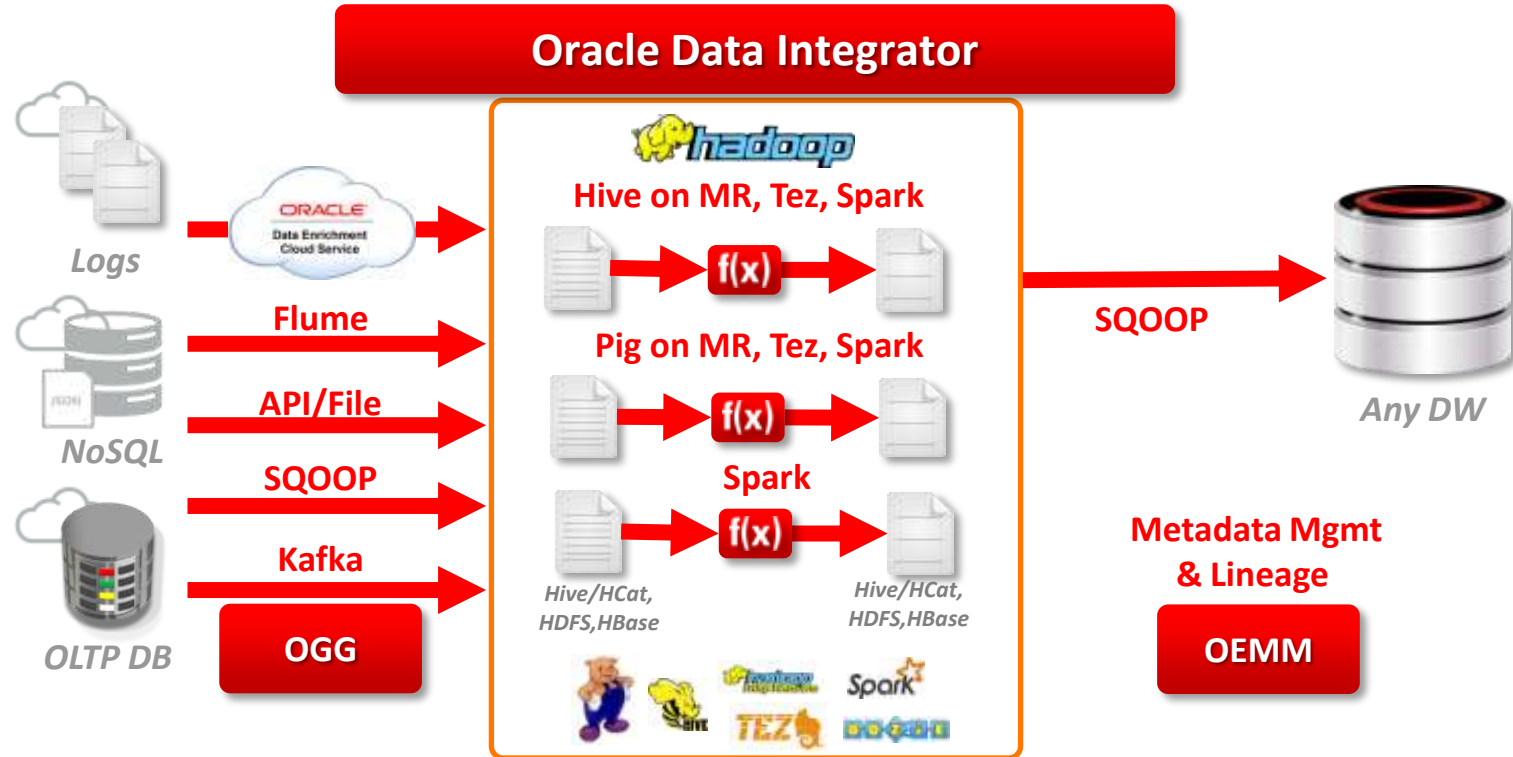
- Prepare unstructured data
- Profile data with sampling
- Clean data in real time or batch
- Verify data for consistency
- Trace lineage of all data
- Define glossary of business terms



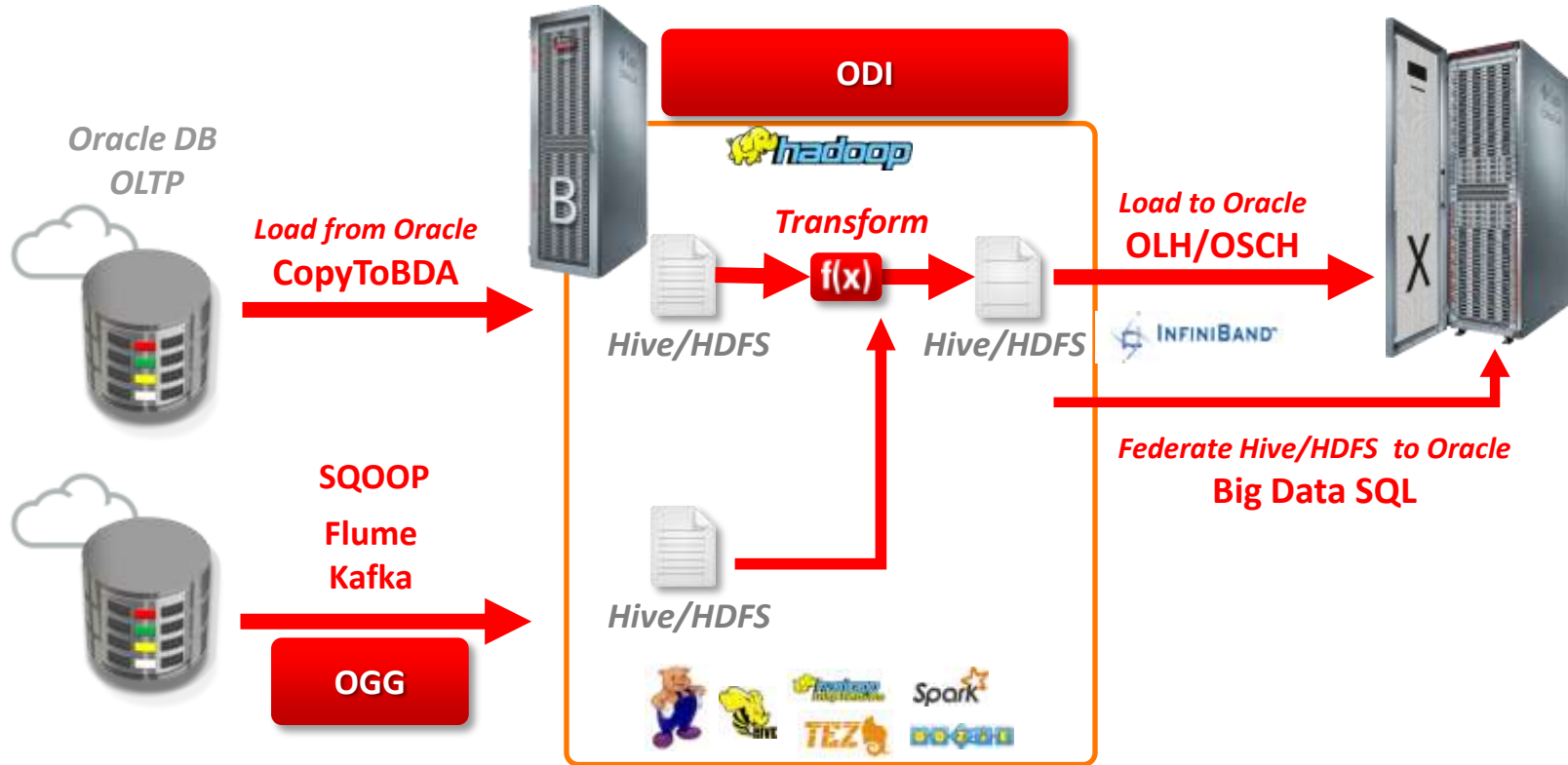
Oracle GoldenGate for Big Data



Oracle Data Integrator for **Transformation** on Big Data

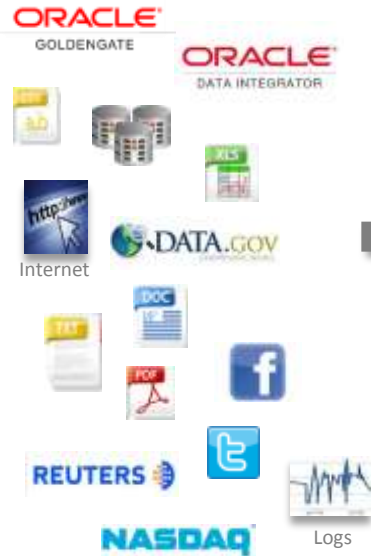


Oracle Data Integration on **Engineered Systems**



Operational **Data Preparation** for Data Without Schema

Unstructured & Structured Data



Enterprise ETL & Data Integration



Enterprise Reporting



Data Discovery & Visualization

Operational **Data Preparation** for Data Without Schema

Unstructured &
Structured Data



Enterprise
ETL & Data
Integration

Enterprise
Reporting

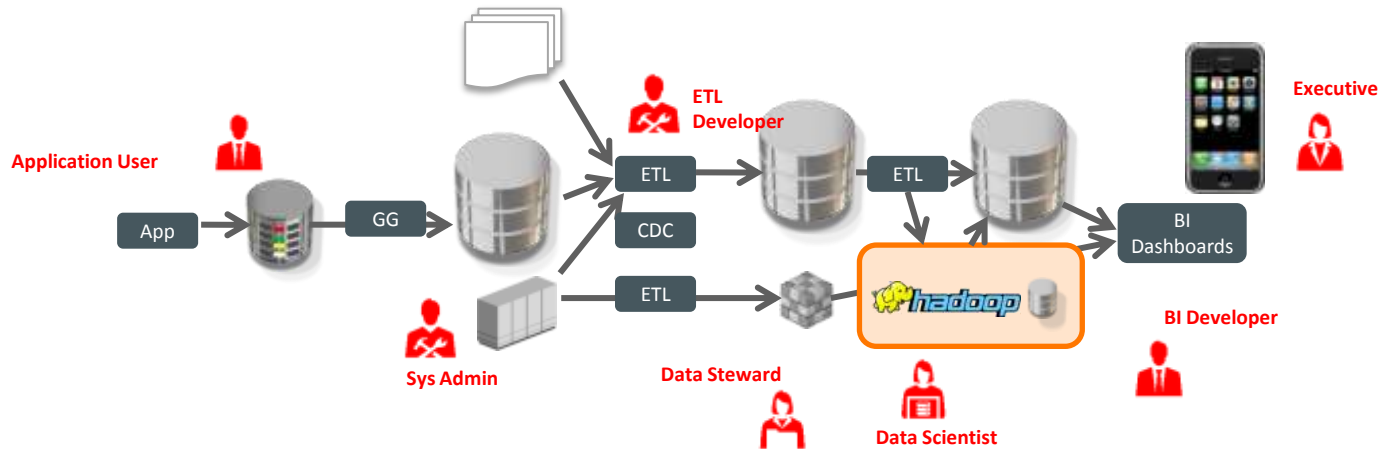
Data Discovery
& Visualization

Comprehensive **Metadata Management** for Big Data

Glossary & Catalog

Harvest & Stitch

Lineage



Oracle Enterprise Metadata Management



Most Open & Heterogeneous Big Data Solution

Operational Integration (Movement / Transformation)

- Hadoop HBase
- Hadoop Hive/Flume
- HP Enscribe
- HP NonStop
- HP Neoview
- Hypersonic SQL
- IBM DB2 i Series
- IBM DB2 UDB
- IBM DB2 z Series
- IBM Informix
- IBM Netezza
- JMS / MQ
- Microsoft Access
- Microsoft SQLServer
- MySQL
- Pivotal Greenplum
- PostgreSQL
- Salesforce.com
- SAP BW / BI
- SAP ERP / ECC
- SAS
- SQL/MP
- SQL/MX
- Sybase ASE
- Sybase IQ
- Teradata
- Oracle Database
- Oracle Exadata
- Oracle Big Data Appliance
- Oracle TimesTen
- Oracle OLAP
- Oracle Business Intelligence
- Oracle BI Applications
- Oracle E-Business Suite
- Oracle JD Edwards Enterprise One
- Oracle JD Edwards World
- Oracle Fusion Applications
- Oracle Governance Risk and Compliance
- Oracle Fusion AIA
- Oracle Retail Applications
- Oracle Agile BI / DW
- Oracle Agile PLM for Process
- Oracle iFlex FlexCUBE
- Oracle iFlex Mantas
- Oracle Hyperion Applications
- Oracle PeopleSoft
- Oracle Siebel CRM / OnDemand
- Oracle Communications
- Oracle WebLogic Server
- Oracle Coherence Data Grid
- Oracle SOA Suite
- Oracle Enterprise Service Bus

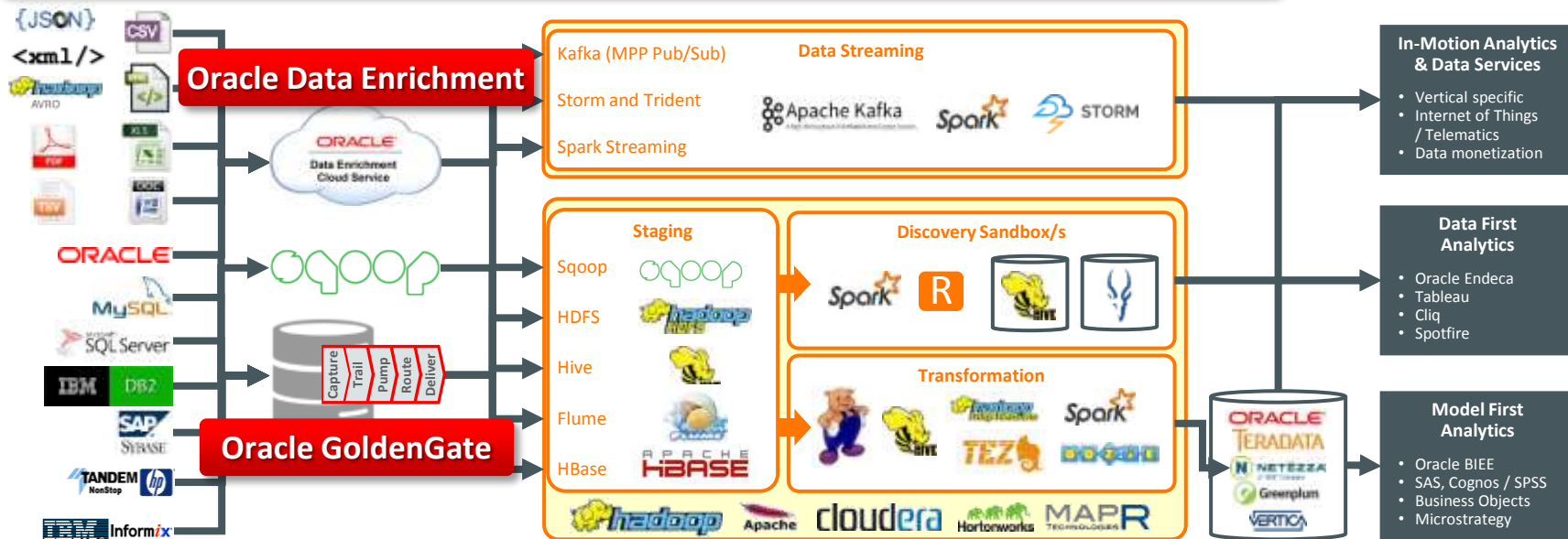
Metadata Harvesting (Glossary, Lineage & Impact Analysis)

- Adaptive
- Altova
- Apache Hcatalog
- Apache Hive/HQL
- Borland
- CA ERwin
- Cloudera Impala
- COBOL Copybook
- DataStax
- Embarcadero
- EMC ProActivity
- GentleWare
- Google BigQuery
- Grandite
- Hadapt Hive
- Hortonworks Hive
- IBM Cognos
- IBM DB2
- IBM DataStage
- IBM Discovery
- IBM Federation Server
- IBM Lotus Notes
- IBM Netezza
- IBM Rational Rose
- IBM Rational Architect
- Informatica Metadata Mgr.
- Informatica PowerCenter
- CoSORT
- ISO SQL Standard (DDL)
- MapR Hadoop Hive
- MicroFocus
- Microsoft Access
- Microsoft Office Excel
- Microsoft Visio
- Microsoft SQL Server
- Microsoft SSIS
- Microsoft Visual Studio
- Microstrategy
- Magic Draw
- OMG CWM Standard
- OMG UML Standard
- Oracle BI Answers
- Oracle BI Enterprise Edition
- Oracle BI Server
- Oracle DAC
- Oracle Data Integrator
- Oracle Data Modeler
- Oracle Database
- Oracle Designer
- Oracle Hyperion Applications
- Oracle Hyperion Essbase
- Oracle Warehouse Builder
- Pivotal Greenplum
- PostgreSQL
- QlikView
- SAP BO Crystal Reports
- SAP BO Designer
- SAP BO Desktop Intelligence
- SAP BO Repository
- SAP BO Data Integrator
- SAP BO Data Steward
- SAP Master Data Management
- SAP Sybase PowerDesigner
- SAP Sybase ASE Database
- SAS Data Integration Studio
- SAS BI Server
- SAS Information Map
- SAS Metadata Management
- SAS OLAP Server
- Select
- Sparx Architect
- Syncsort
- Tableau
- Talend
- Teradata
- Tigris
- Visible
- W3C DTD & XSD Schema

+ open APIs and standards based meta-model

Oracle Data Integration Can Help Right Now

Oracle Data Integrator



Oracle Data Governance

Oracle Big Data

1 Oracle Big Data architektúra és a relációs világ: összekapcsolva

2 Oracle adatintegráció a Big Data világban

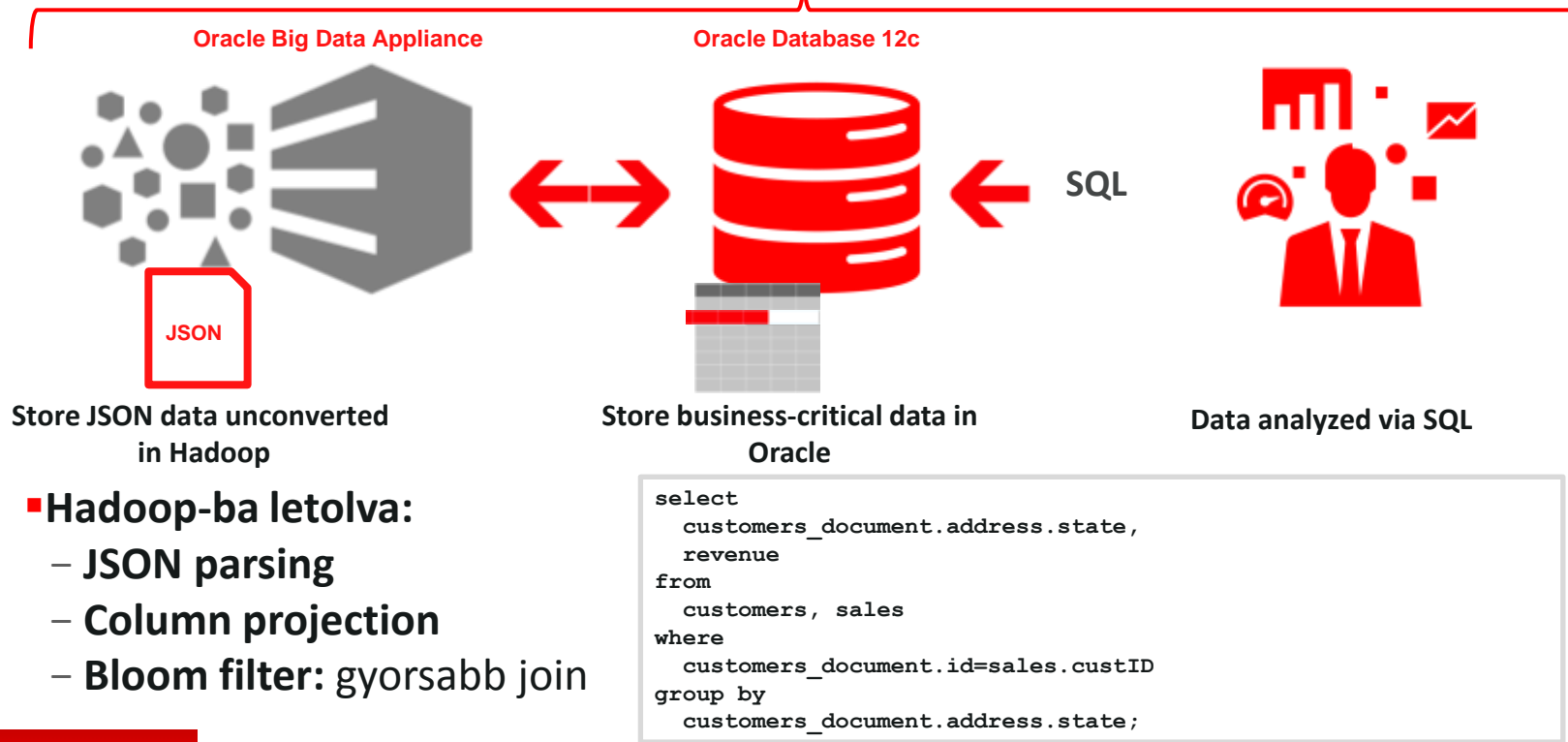
3 Oracle Big Data SQL és demó

4 Oracle Big Data Discovery

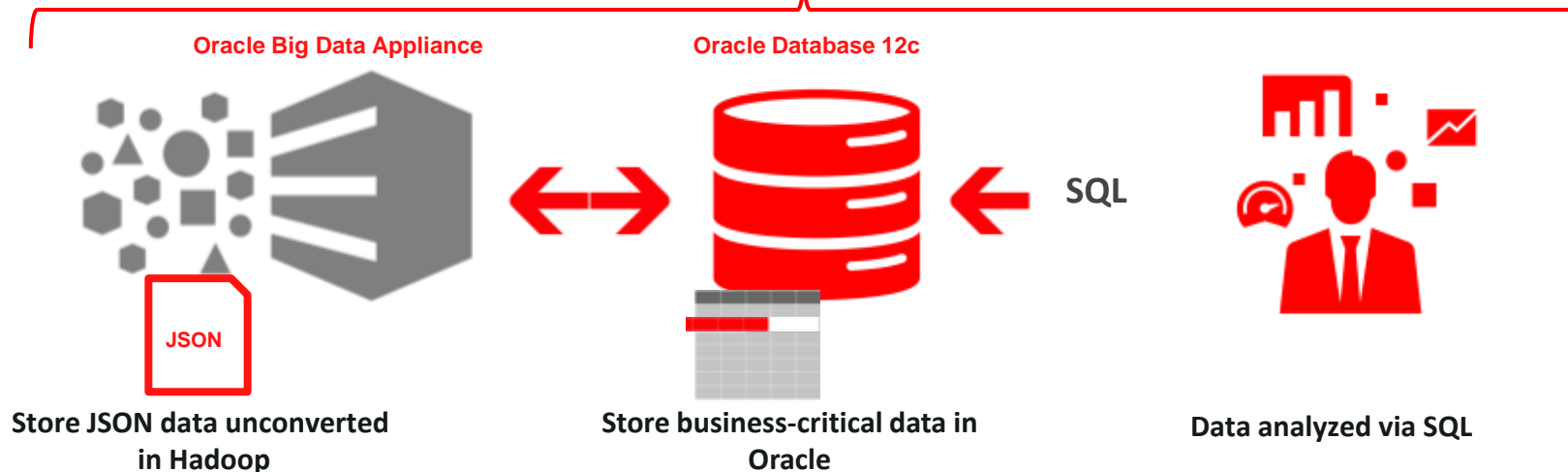
5 Oracle Advanced Analytics: R és Data Mining

6 Ügyfél felhasználási esetek és architektúrák

Minden adat lekérdezése—SQL & Oracle Big Data SQL



Minden adat szabályozása—SQL & Oracle Big Data SQL



- **Advanced Security Hadoop-on**
 - Masking/Redaction
 - Virtual Private Database
 - Fine-grained Access Control

```
DBMS_REDACT.ADD_POLICY(  
  object_schema => 'txadp_hive_01',  
  object_name => 'customer_address_ext',  
  column_name => 'ca_street_name',  
  policy_name => 'customer_address_redaction',  
  function_type => DBMS_REDACT.RANDOM,  
  expression => 'SYS_CONTEXT(''SYS_SESSION_ROLES'',  
    ''REDACTION_TESTER'')='''TRUE'''  
);
```

Oracle External Table kiterjesztés

```
CREATE TABLE movielog (  
  click VARCHAR2(4000))  
ORGANIZATION EXTERNAL (  
  TYPE ORACLE_HIVE  
  DEFAULT DIRECTORY DEFAULT_DIR  
  ACCESS PARAMETERS  
  (  
    com.oracle.bigdata.tablename logs  
    com.oracle.bigdata.cluster mycluster  
  ))  
REJECT LIMIT UNLIMITED;
```

```
CREATE TABLE ORDER (  
  cust_num VARCHAR2(10),  
  order_num VARCHAR2(20),  
  order_total NUMBER(8,2))  
ORGANIZATION EXTERNAL (  
  TYPE ORACLE_HIVE  
  DEFAULT DIRECTORY DEFAULT_DIR  
)  
PARALLEL 20  
REJECT LIMIT UNLIMITED;
```

- External Tables új típusok
 - ORACLE_HIVE (metadata öröklés)
 - ORACLE_HDFS (metadata megadása)
- Big Data paraméterek elérése
 - Hadoop cluster
 - Remote Hive database/table
 - DBMS_HADOOP Package, automatikus import
- Párhuzamosságra tervezve
 - Külső egységek leképezése az Oracle párhuzamosságra

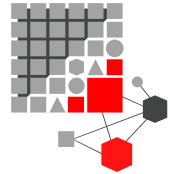
Oracle Big Data

- 1 Oracle Big Data architektúra és a relációs világ: összekapcsolva
- 2 Oracle adatintegráció a Big Data világban
- 3 Oracle Big Data SQL és demó
- 4 Oracle Big Data Discovery**
- 5 Oracle Advanced Analytics: R és Data Mining
- 6 Ügyfél felhasználási esetek és architektúrák

BD felhasználó szükséges fontos képességei



Raw Files
in HDFS



Ready Files
in HDFS

- Egyetlen termékkel a Hadoop nyers adatokba üzleti betekintést adjon, ami publikálható vállalati felhasználásra.



Business User



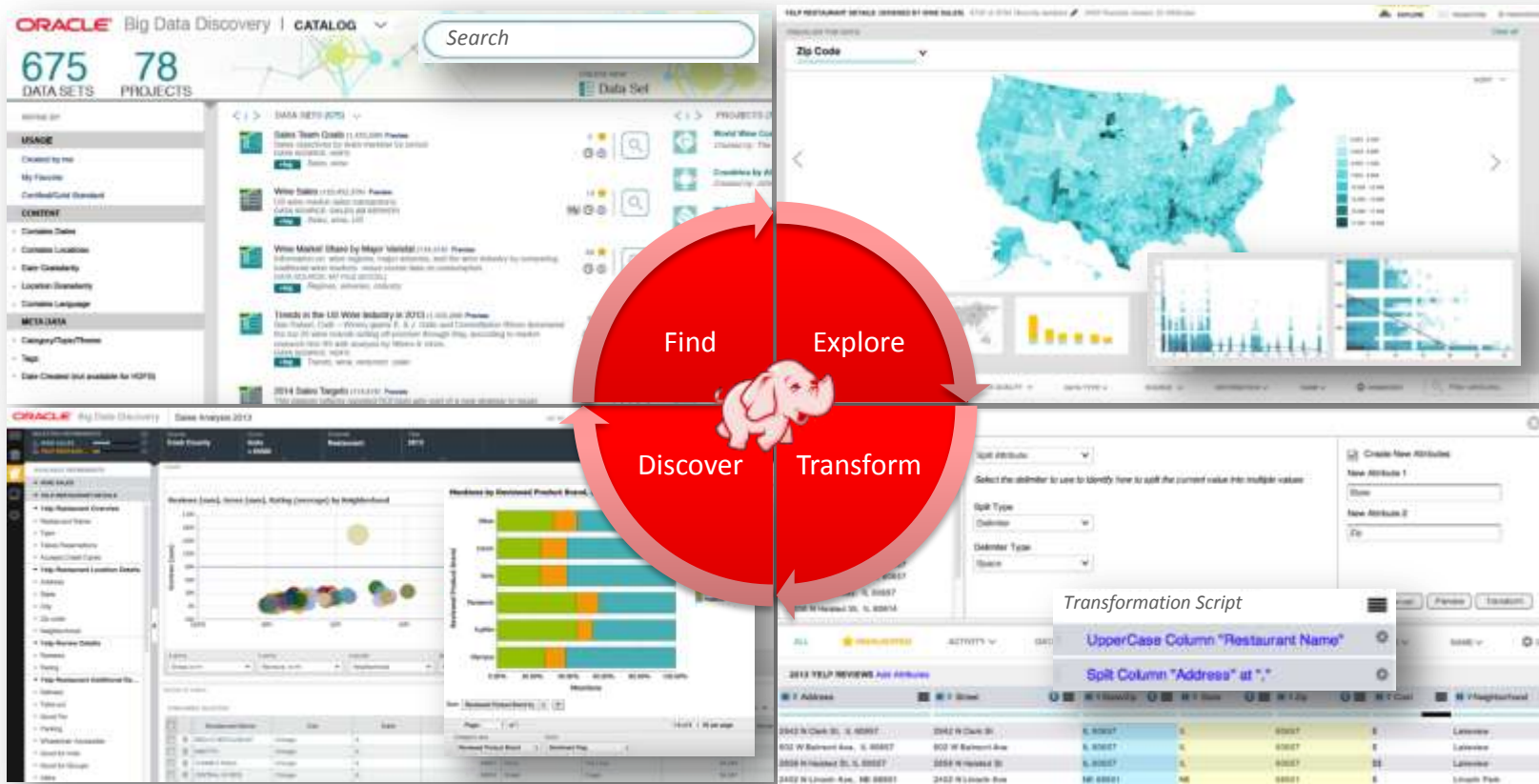
Business Analyst



Data Scientist

- Egyszerű vizuális felhasználói interfész:
mindenki elemezheti a Big Data adatokat, nem csak a Data Scientist

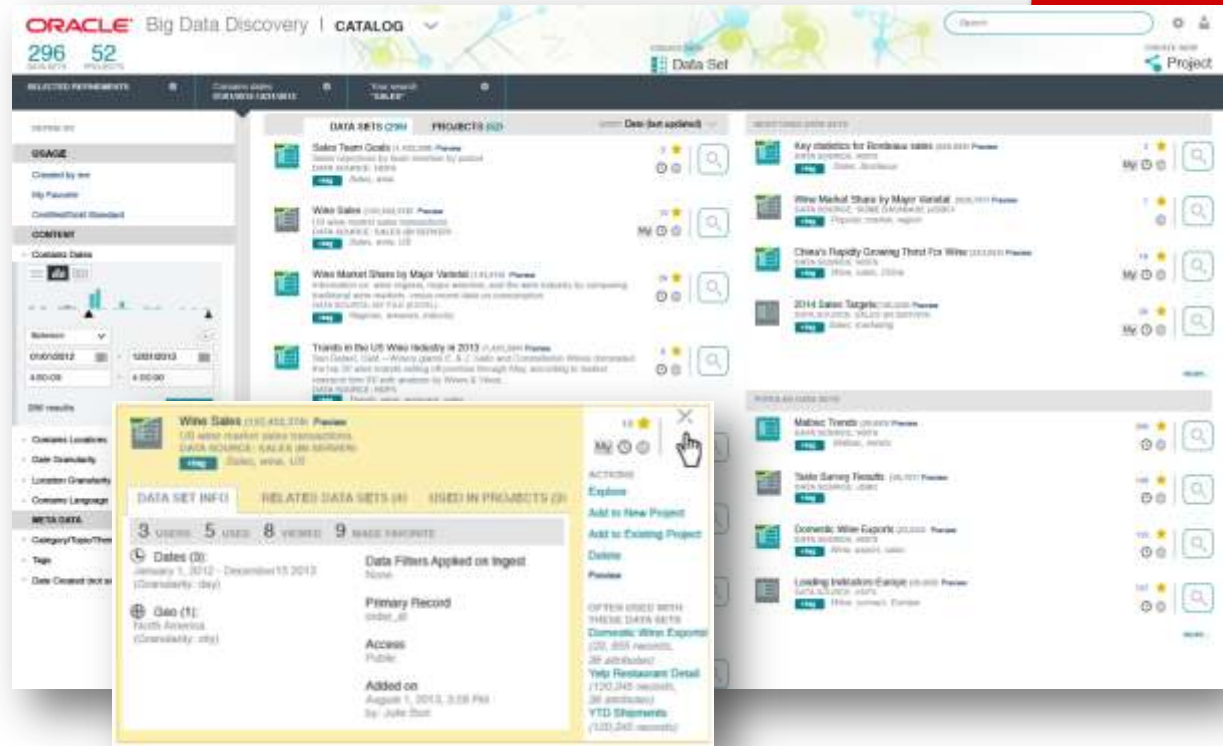
Oracle Big Data Discovery – The Visual Face of Hadoop



Találjuk meg a releváns adathalmazokat

Find

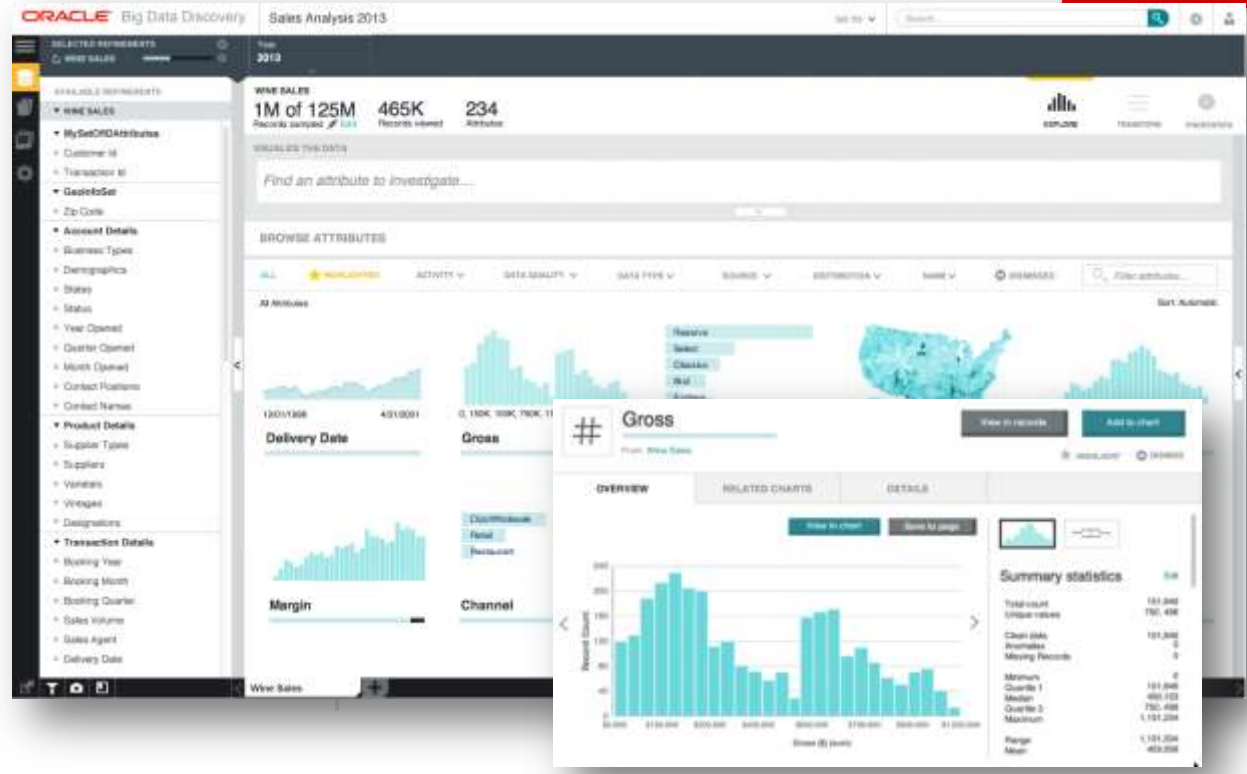
- Access a rich data set catalog:
The visual face of Hadoop
- Use a familiar **search and guided navigation** user interface
- Data set summaries, annotation, **recommendations** & projects
- **Bring your own data** through self-service upload
- Automatically turn files in HDFS into **in-memory indexes**
- Intelligently sample large data sets for **fast performance at scale**



Kísérletezve ismerjük meg a sokféle adat jellemzőit

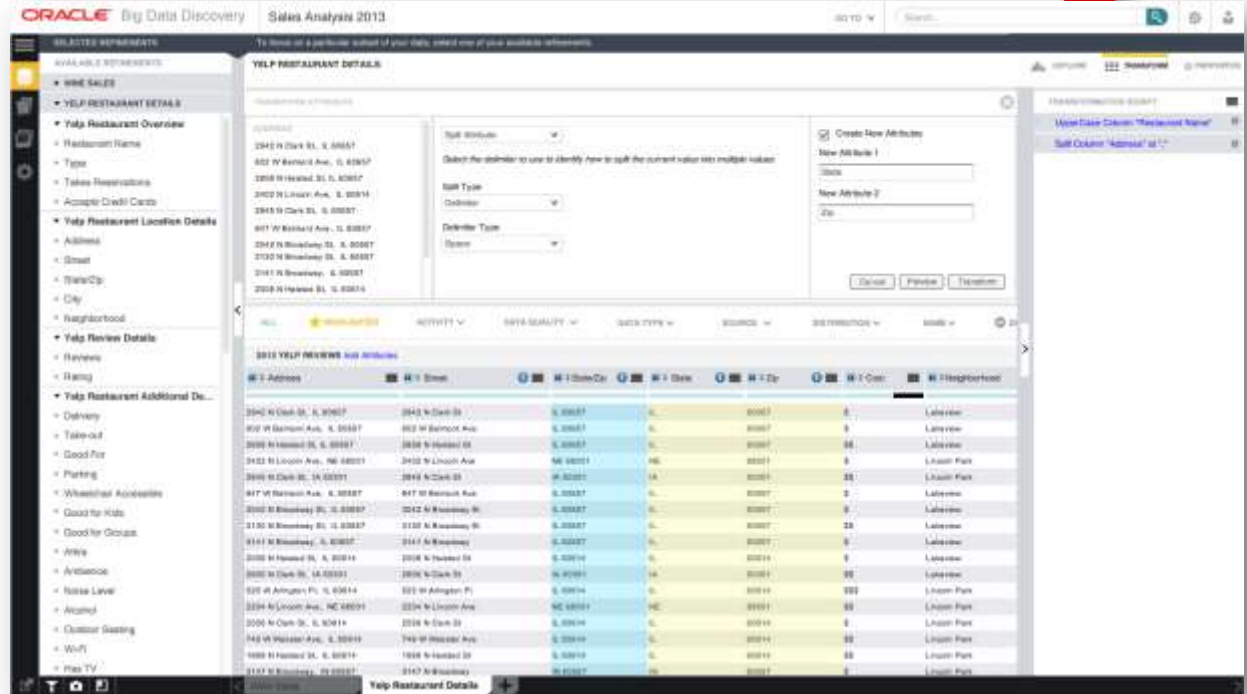
Explore

- Take a **guided tour** of your data
- Describe **shape of data** by visualizing value distribution based on type
- Highlight **statistical correlations** between attributes
- Understand **data quality** and outliers
- Highlight most important **attributes** with entropy-based sorting
- Evaluate whether a **data set is worthy of further investment**



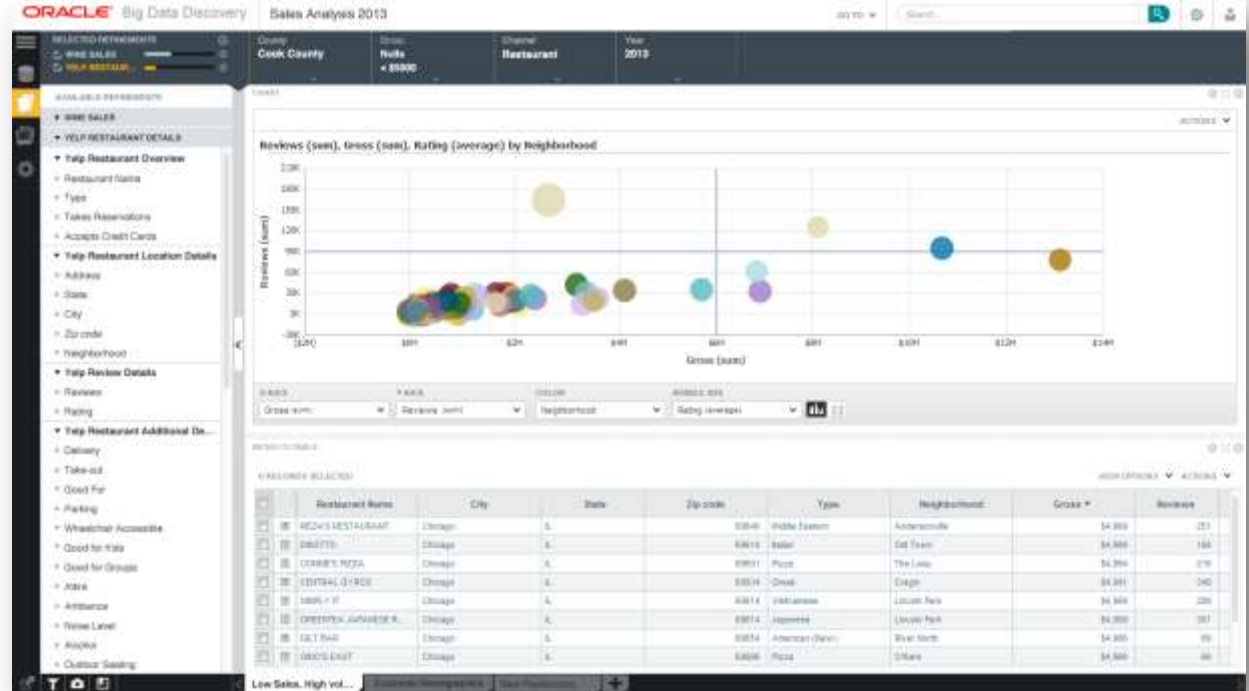
Transzformáljuk az adatokat elemzésre

- Empower users to **fix the data** themselves
- **Wrangle data** interactively to replace missing values, convert types, collapse, reshape, pivot, group, custom tag, merge and more
- **Enrich data interactively** with locations, terms, entities, sentiment
- **Preview** the result before applying, **rollback and replay** transforms
- **Execute transforms at scale** via Spark
- Write custom transformations in Spark for **extensibility**



Fedezzük fel az adatok sokféleségében rejlő mélységeket

- **Blend data sets** together for fresh perspectives
- Filter with powerful and familiar **search** and **guided navigation**
- **Drag and drop** dashboard from a rich palette of visualizations
- Create machine learning models to **predict outcomes**
- **Publish** data back to Hadoop
- **Share** projects with team members



Oracle Big Data Discovery – „megváltoztatja a játékot”

Üzleti előnyök

- **Előnyök gyorsabb elérése:** nyers adatok villámgyors üzleti feldolgozása a teljes szervezet számára
- **Többen vehetnek részt a munkában:** hatékonyabb Big Data project team: többszínű, további értékes tudás, képesség bevonása



IT előnyök

- **Előkészítés, megosztás, fejlődés:** Integrálás a Hadoop nyílt standardokkal, a teljes Hadoop környezet felhasználása
- **Meglévő technológiai korlátok lebomlása:** Natívan a Hadoop node-okon futnak a műveletek, maximális skálázhatóság és teljesítmény

Oracle Big Data

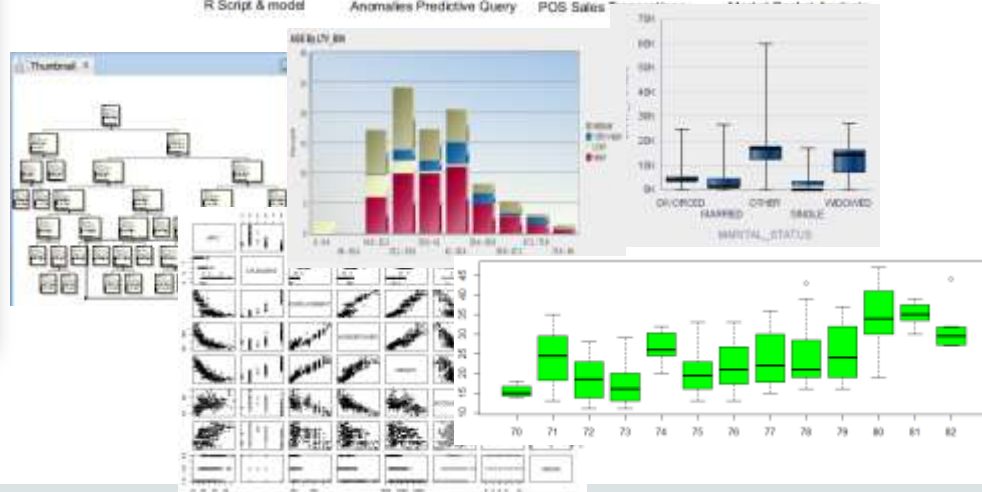
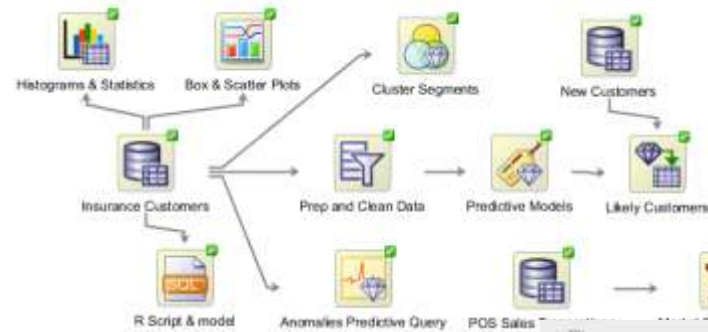
- 1 Oracle Big Data architektúra és a relációs világ: összekapcsolva
- 2 Oracle adatintegráció a Big Data világban
- 3 Oracle Big Data SQL és demó
- 4 Oracle Big Data Discovery
- 5 Oracle Advanced Analytics: R és Data Mining
- 6 Ügyfél felhasználási esetek és architektúrák

Oracle Advanced Analytics

A leggyorsabb út a skálázható széleskörű prediktív analitikához

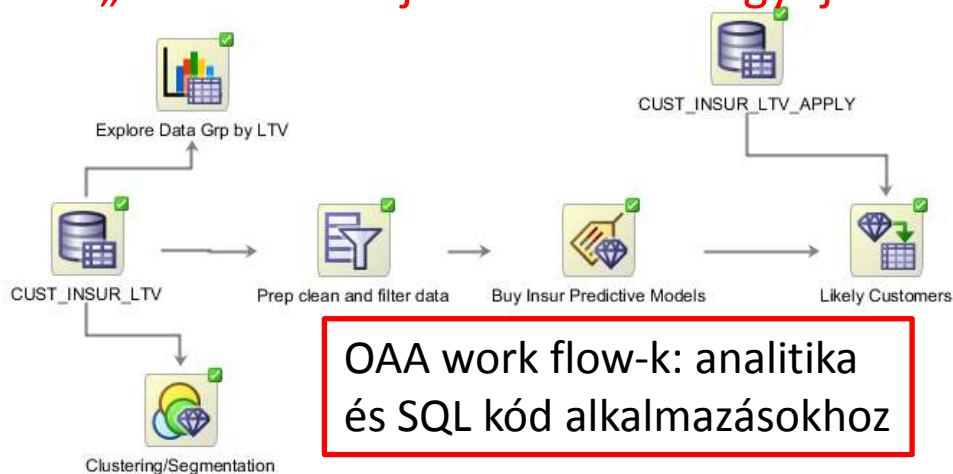
Legfontosabb jellemzők

- In-database adatbányászati és open source R algoritmusok
- Három Oracle Database felület:
SQL, **SQLDev/ODMr GUI**, **R**
- **Skálázható, párhuzamos in-database**
- Workflow GUI és IDE-k
- A Database integráns része
- Vállalati alkalmazásokba **beágyazható**



Példa: biztosítás

a „vásárlásra hajlamosak” összegyűjtése és profilozása



OAA work flow-k: analitika
és SQL kód alkalmazásokhoz

```
Rule Surrogates Target Values
If BANK_FUNDS > 225.5
And CHECKING_AMOUNT <= 207.5
And MONEY_MONTHLY_OVERDRAWN > 53.115
And T_AMOUNT_AUTOM_PAYMENTS > 8283.5
And N_TRANS_ATM > 6.5
Then Yes
```

Node: 7
Prediction: Yes
Support: 2617 (28.53%)
Confidence: 67.86%
No : 841 (32.14%)
Yes : 1776 (67.86%)
Split: MONEY_MONTHLY_OVERDRAWN

Node: 6
Prediction: Yes
Support: 2998 (32.69%)
Confidence: 61.54%
No : 1153 (38.46%)
Yes : 1845 (61.54%)
Split: T_AMOUNT_AUTOM_PAYMENTS

Node: 4
Prediction: Yes
Support: 3685 (40.18%)
Confidence: 54.84%
No : 1664 (45.16%)
Yes : 2021 (54.84%)
Split: MONEY_MONTHLY_OVERDRAWN

Node: 9
Prediction: No
Support: 381 (4.15%)
Confidence: 81.89%
No : 312 (81.89%)
Yes : 69 (18.11%)
Split: N_TRANS_ATM

Node: 23
Prediction: No
Support: 621 (6.77%)
Confidence: 71.98%
No : 447 (71.98%)
Yes : 174 (28.02%)
Split: N_TRANS_ATM

Node: 22
Prediction: No
Support: 66 (0.72%)
Confidence: 96.97%
No : 64 (96.97%)
Yes : 2 (3.03%)
Split: N_TRANS_ATM



Oracle Big Data

- 1 Oracle Big Data architektúra és a relációs világ: összekapcsolva
- 2 Oracle adatintegráció a Big Data világban
- 3 Oracle Big Data SQL és demó
- 4 Oracle Big Data Discovery
- 5 Oracle Advanced Analytics: R és Data Mining
- 6 Ügyfél felhasználási esetek és architektúrák

FIFA Early Warning System (EWS)

Big Data for Fraud Detection on Online Sports Bets

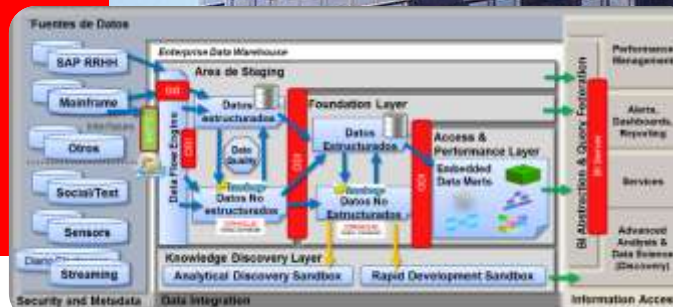


Customer Example: La CaixaBank

Big Data Warehouse using Oracle Data Integration and Hadoop

Big Data Reservoir

- Comprehensive modernization program for the bank's data warehouse infrastructure
 - Combining structured enterprise data with unstructured data for next gen insights
 - Variety of reporting and machine learning approaches used with different data types
- Accelerated delivery of business insights from use of Data Discovery layer
 - Data Factory concepts used for delivery



Customer Example: DirecTV

Oracle Data Integration and Kafka for Realtime Big Data



Realtime Big Data

Real-time Streaming Analytics for determining success rate of programs

- Real-time loading of the change data from receiver collected on Oracle GoldenGate to Kafka and HDFS
- 180M– 200M transactions
- Leveraged GoldenGate to Stream Change Data into Kafka and HDFS
- Downstream processing of data from Kafka and XML data using Data Torrent



További információ:

Linkek, anyagok, videók, virtuális gép

- <https://cloud.oracle.com/bigdata>
- <https://cloud.oracle.com/bigdatapreparation>
- <https://cloud.oracle.com/bigdatadiscovery>
- <https://cloud.oracle.com/nosqldatabase>
- Oracle Big Data: <https://www.oracle.com/bigdata/>, [Oracle Big Data Light VM letöltés](#)
- [Data Warehousing](#), [Engineered Systems](#), [Business Analytics](#)
- [Information Management and Big Data A Reference Architecture](#)

Videók:

- [Oracle Big Data csatorna](#), [Oracle Big Data Discovery csatorna](#)
- [Oracle Data Integrator Getting Started Series](#), [Oracle GoldenGate](#)
- [Oracle Advanced Analytics](#), [Using SQL Pattern Matching with Big Data Lite VM](#)

Integrated Cloud

Applications & Platform Services