

Target Reply

Esperienze Enterprise Data Lake



AGENDA

- Reply: breve introduzione
- BigData: Typical architectures
- Referenze e progetti



AGENDA

- **Reply: breve introduzione**
- BigData: Typical architectures
- Referenze e progetti



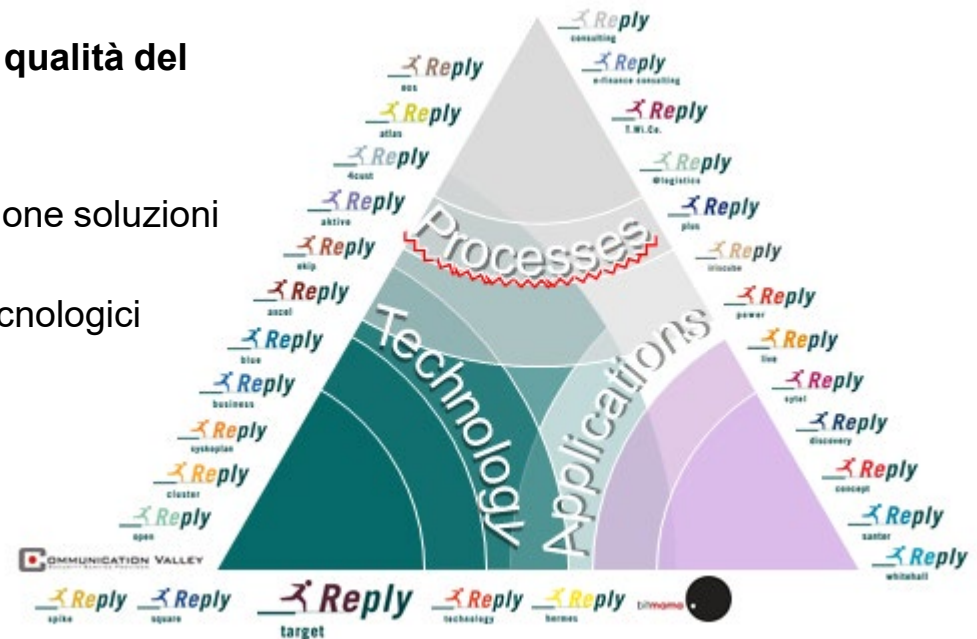
Il modello di business di Reply

Il modello di business di Reply si basa su una struttura a rete costituita da società controllate, specializzate per linee di offerta, che costituiscono centri di eccellenza in grado di posizionarsi come *best in class* nei rispettivi ambiti di competenza (Processi, Applicazioni, Tecnologie)

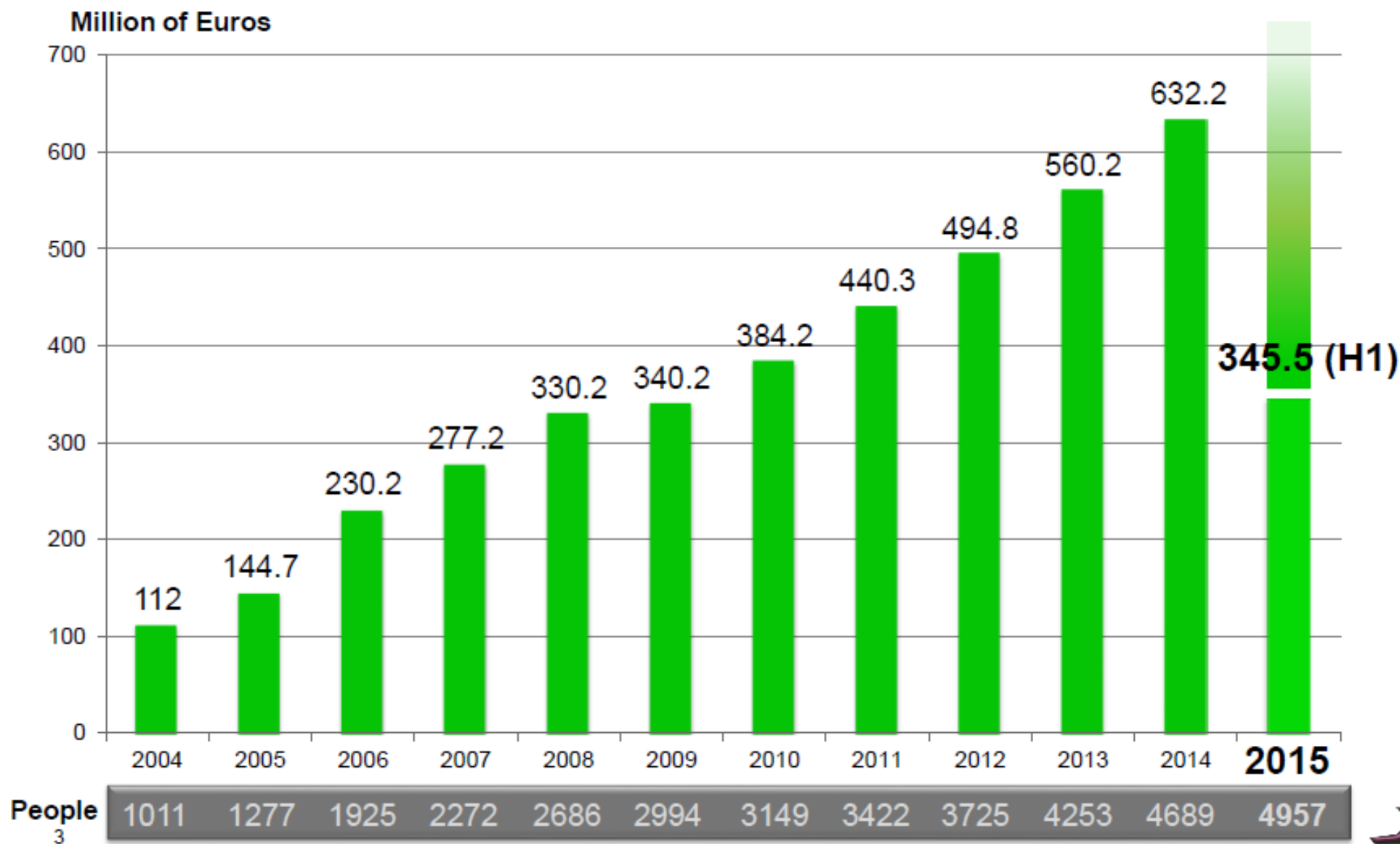
Questo modello coniuga la **flessibilità**, la **specializzazione** e il **dinamismo** tipici delle piccole strutture, con la **capacità progettuale** ed **organizzativa** di un'entità di grandi dimensioni

I principali **fattori competitivi** di Reply sono:

- riconoscimento da parte del mercato e dei clienti della **qualità del servizio offerto**
- esperienza consolidata nel **project management**
- **rapidità** nel progettare, realizzare e mettere in produzione soluzioni complesse
- **flessibilità** nell'adeguamento ai rapidi cambiamenti tecnologici
- **innovazione**

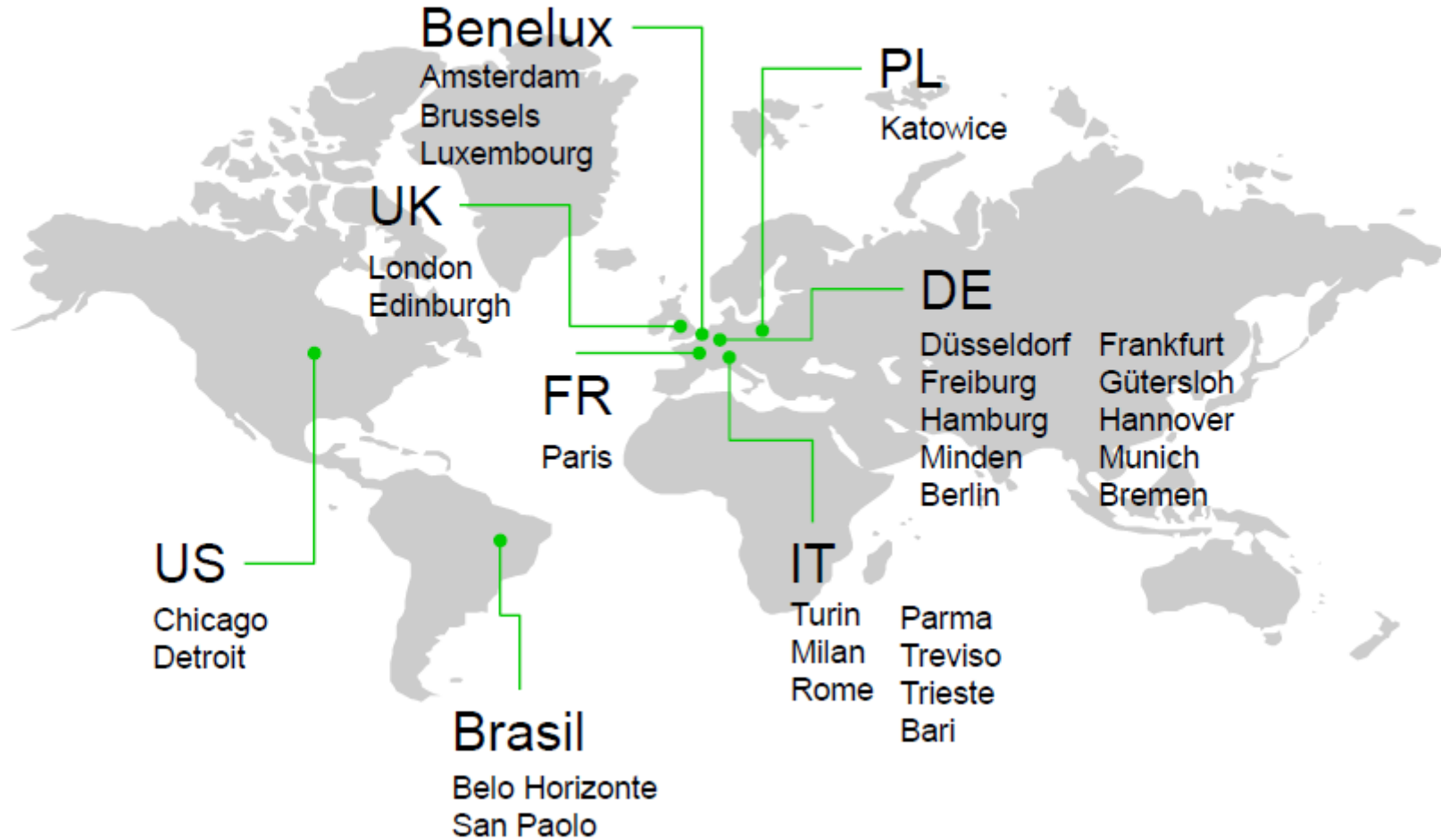


Reply – Revenue & People



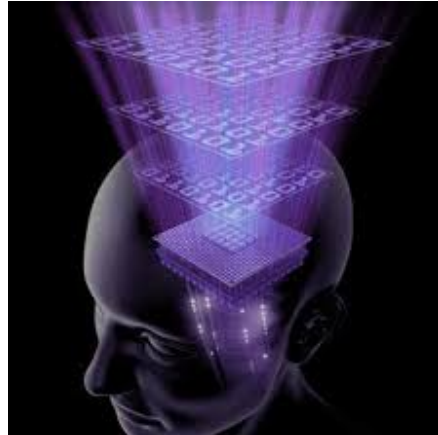
Reply in the world

—● Main Locations



Target Reply

- Target Reply is the Reply Group's company specialized in driving customers to build **Business Intelligence**, **Data Warehousing** and **Big Data** solutions.
- Target Reply is composed by **100 people** skilled in technologies, processes and methodologies.



- **Since 2007** Target Reply has been working with Italian and foreign leader companies.
- Target Reply operates with experience within all **major industries**.

Manufacturing

Retail

Energy

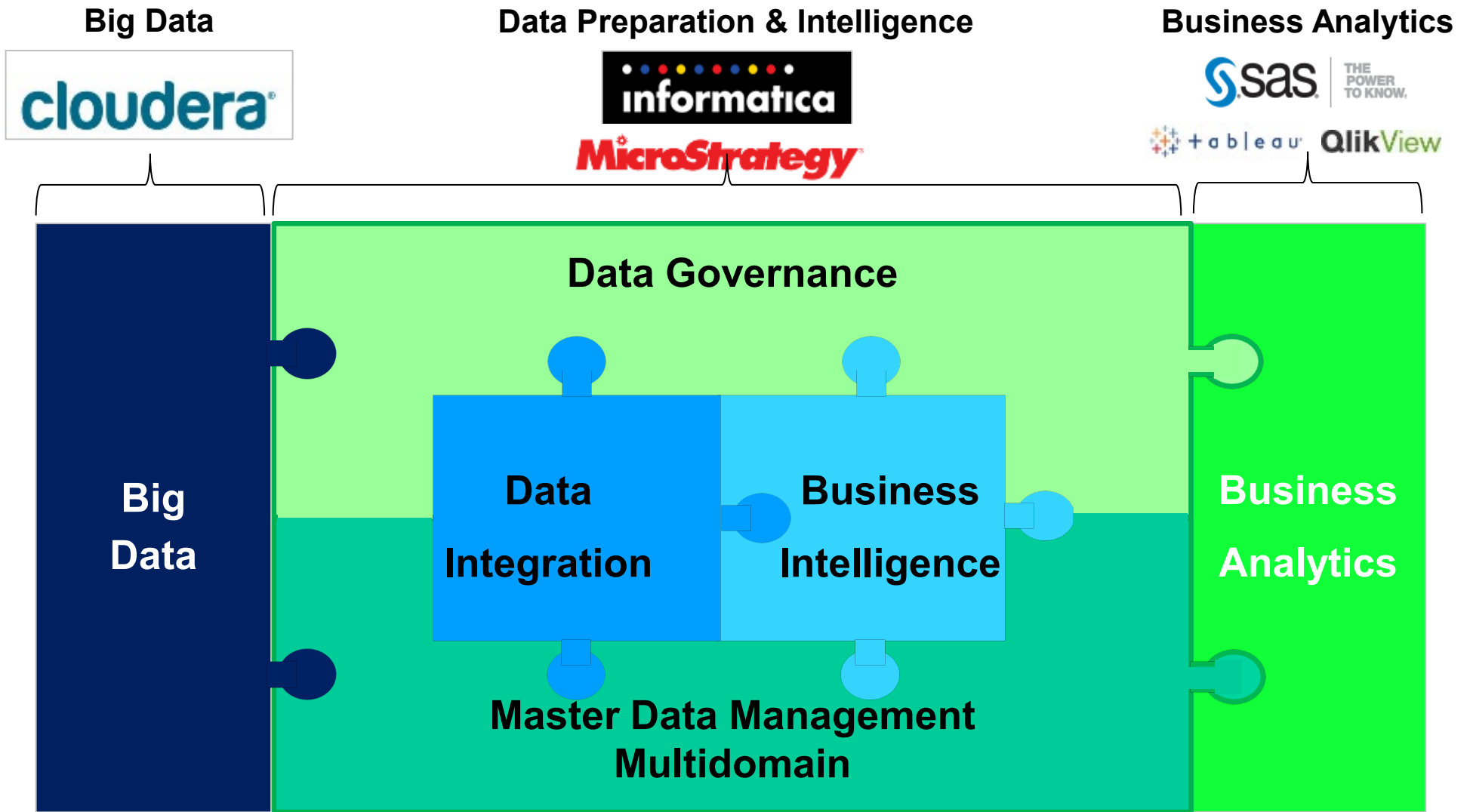
Telco

Finance

Media



360° View of Data



Main Skills

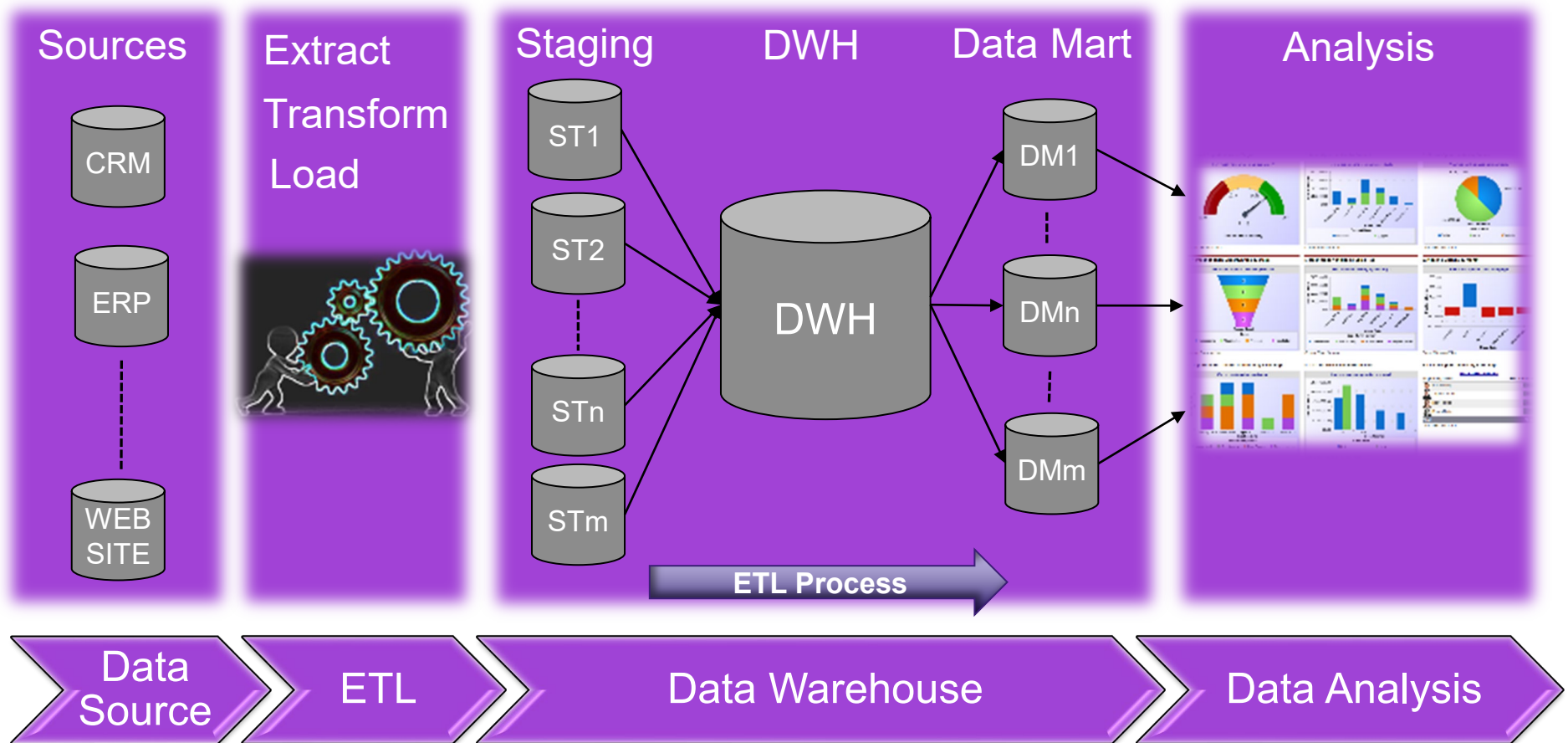


AGENDA

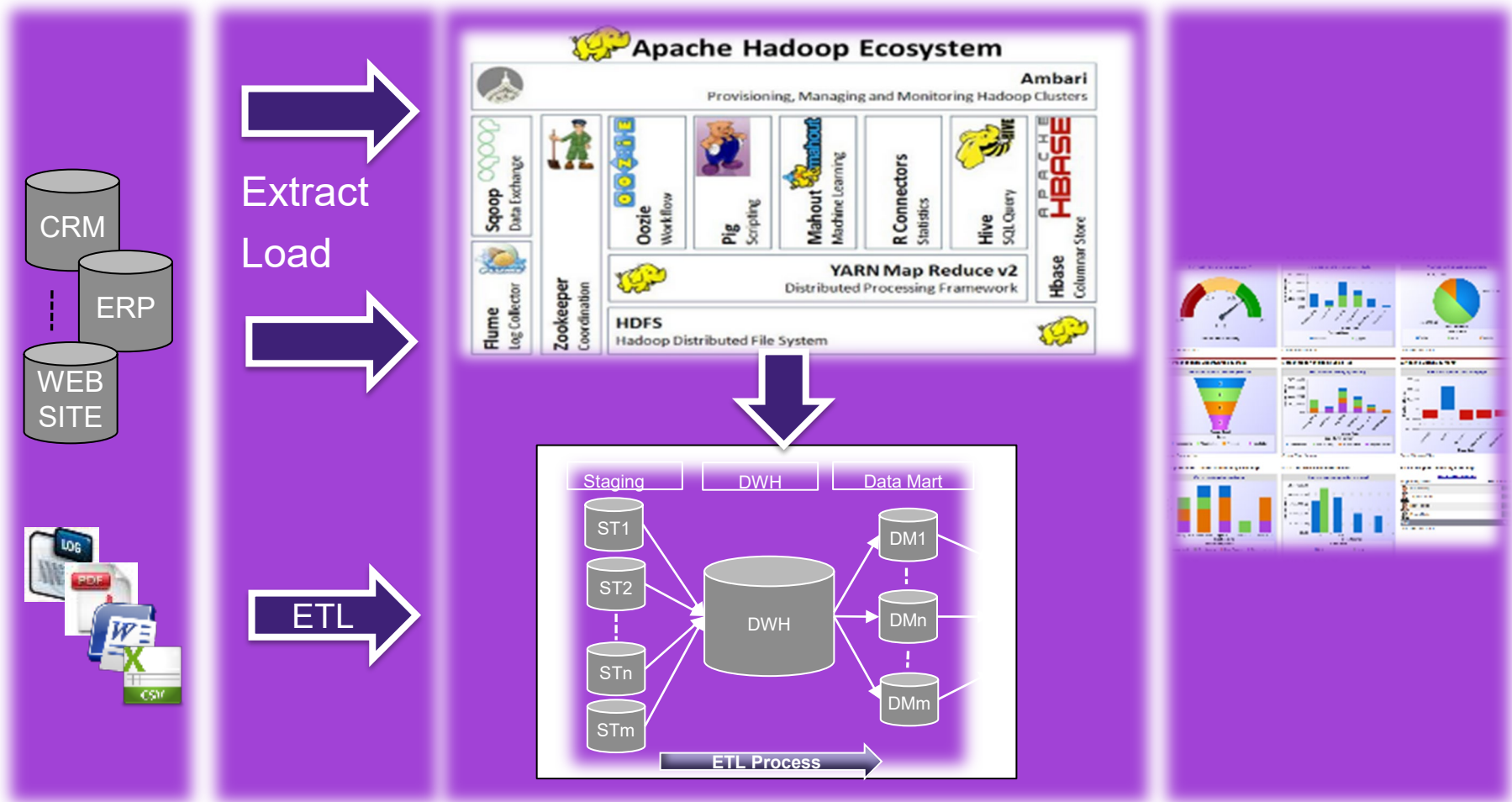
- Reply: breve introduzione
- **BigData: Typical architectures**
- Referenze e progetti



Traditional DWH Architecture



Typical Data Lake Architecture



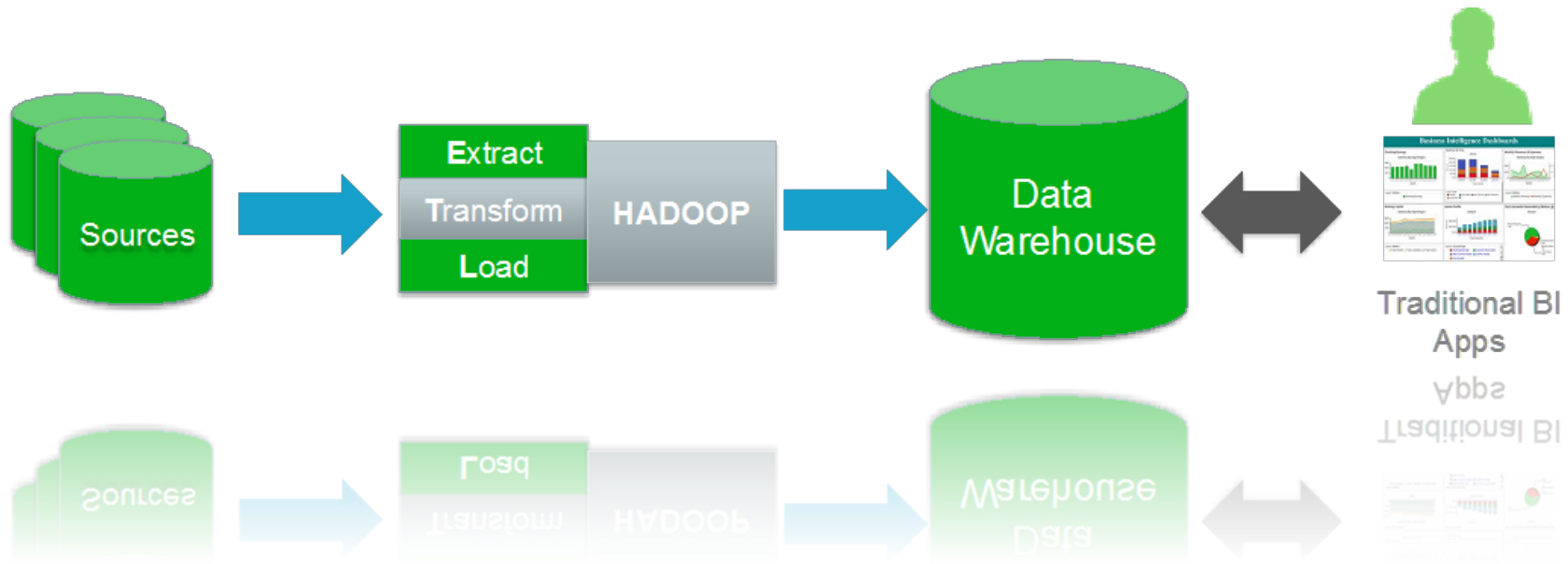
Use case #1 Operational Efficiency

ETL Acceleration

EDW Optimization

Active Archive

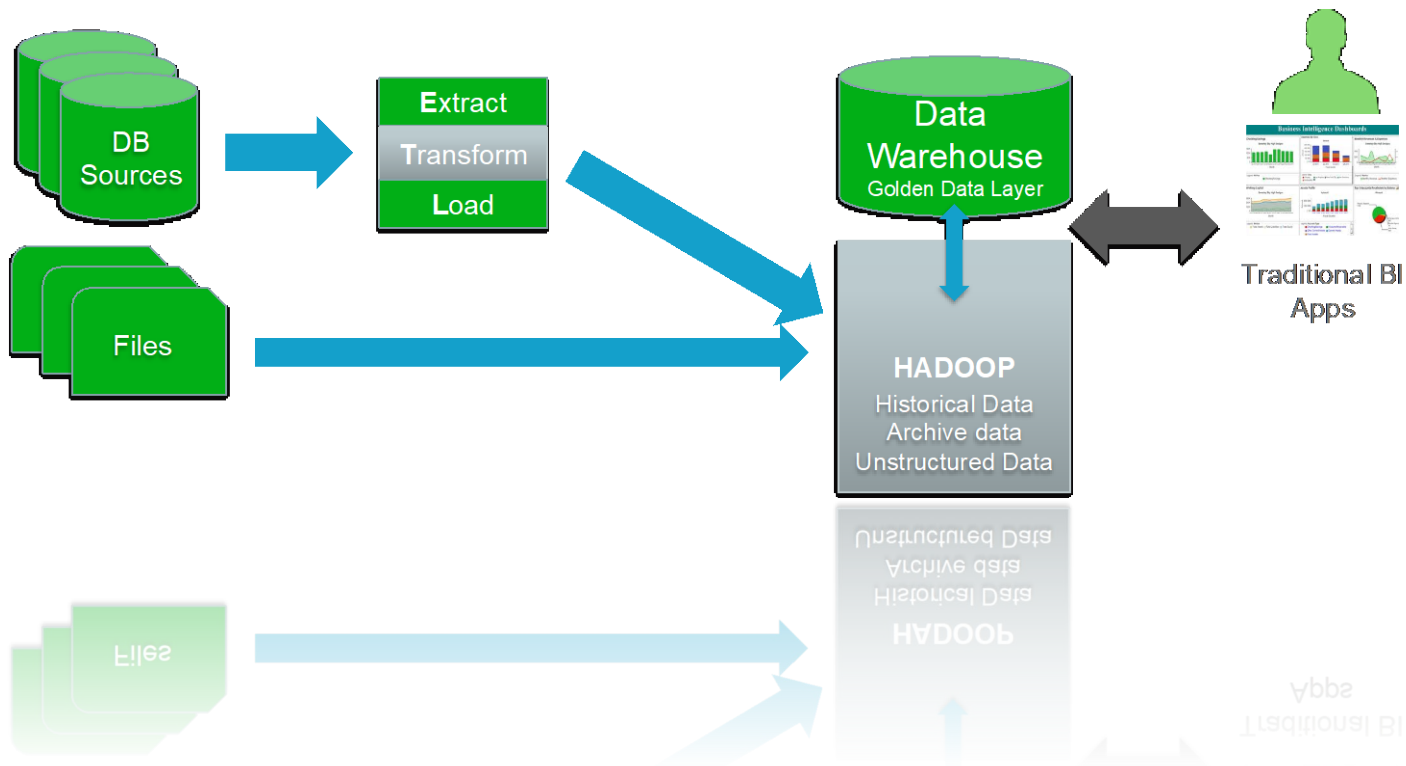
Historical Compliance



Use case #2 Operational Efficiency

ETL Acceleration
EDW Optimization

Active Archive
Historical Compliance



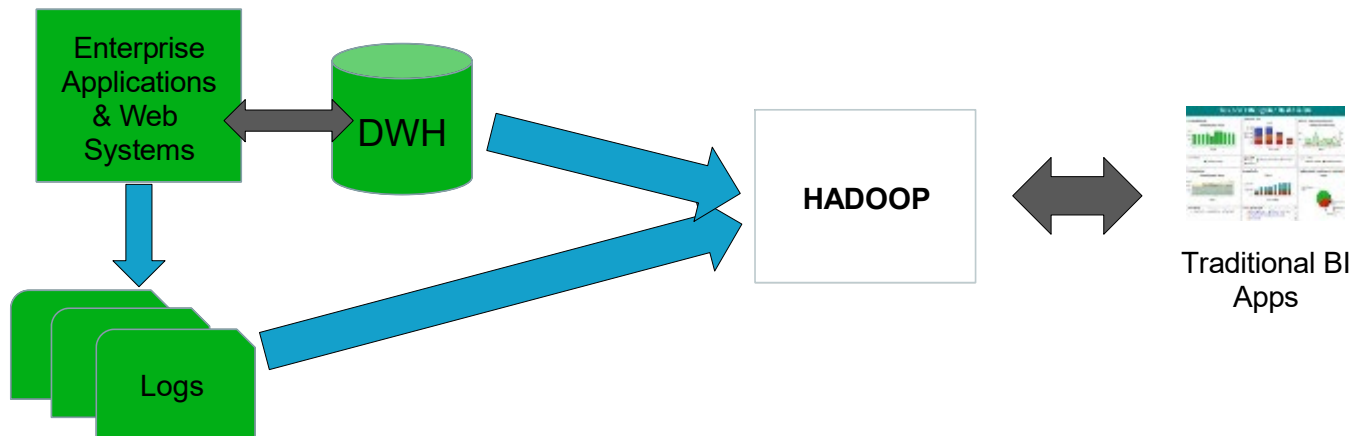
Use case #3 Operational Efficiency

ETL Acceleration

EDW Optimization

Active Archive

Historical Compliance



- **Reply: breve introduzione**
- **BigData: Typical architectures**
- **Referenze e progetti**



Key Facts

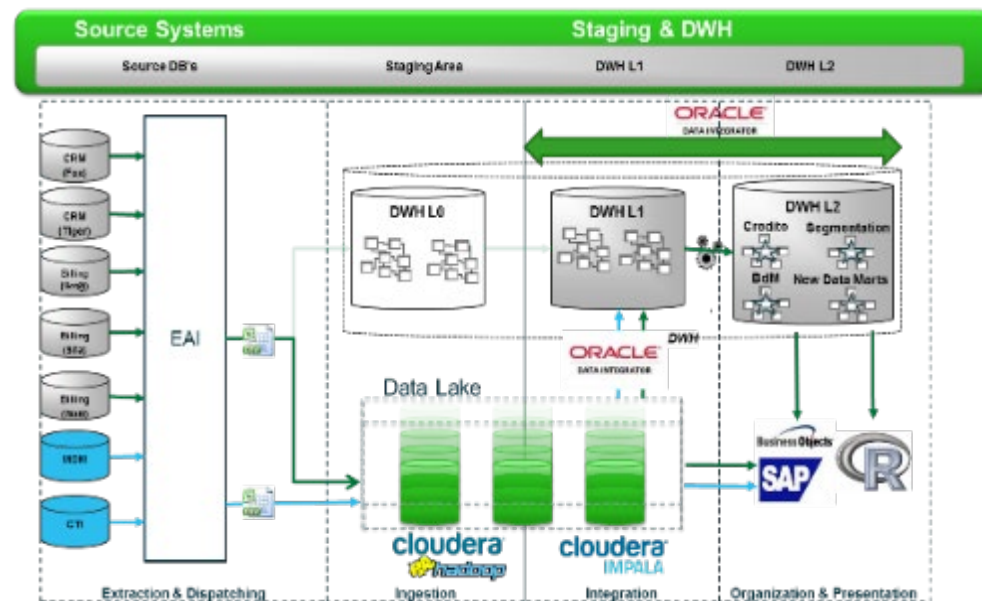
- uno dei principali attori del settore utilities in Europa con 147.400 dipendenti nel mondo e attività in 70 paesi.
- Il portafoglio comprende attività di approvvigionamento gas, produzione e commercio di energia: termica, nucleare, biomasse ed energie rinnovabili.
- Nel 2013 ha generato € 81.3 miliardi di entrate nel mondo.

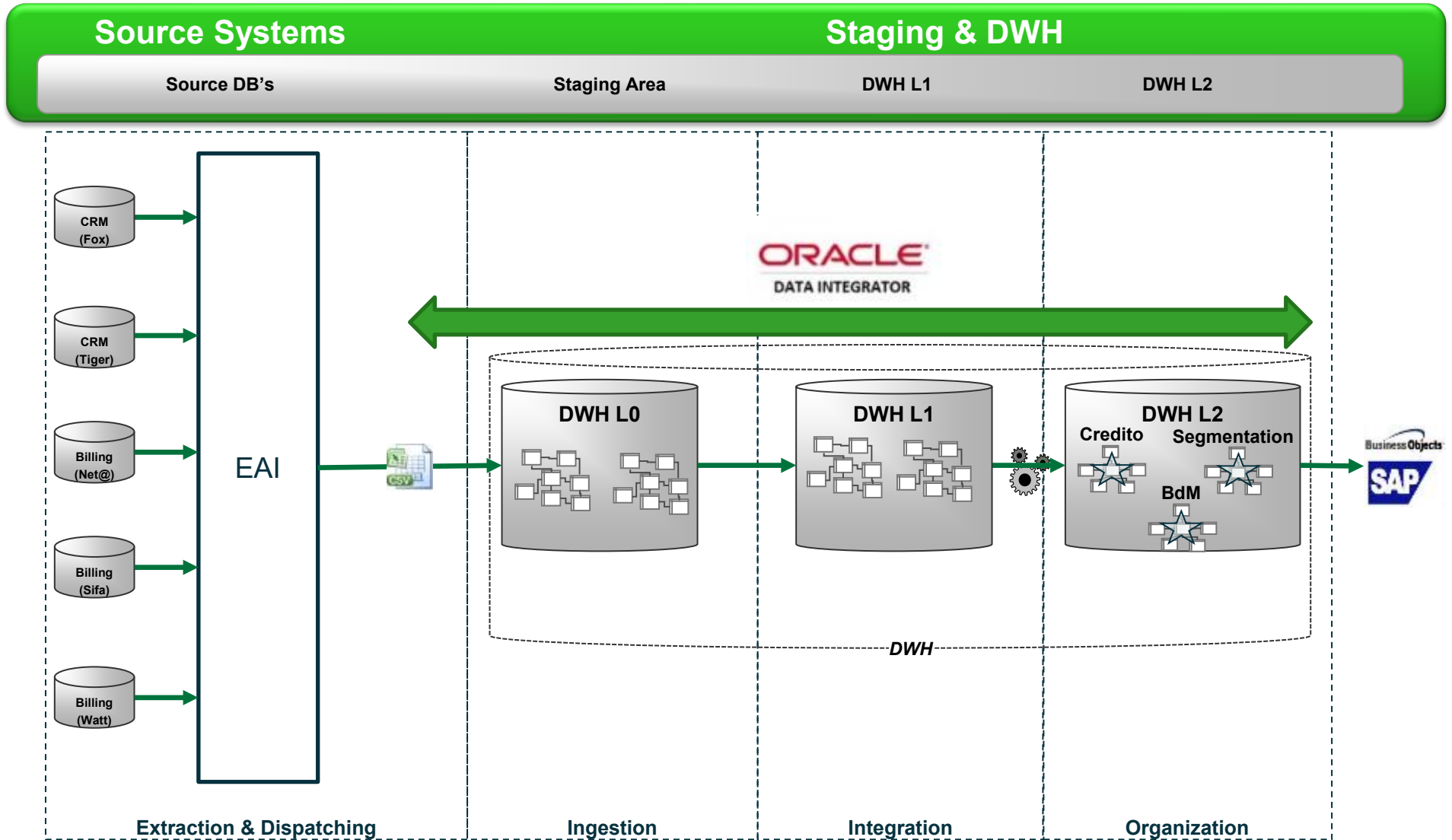
Solution

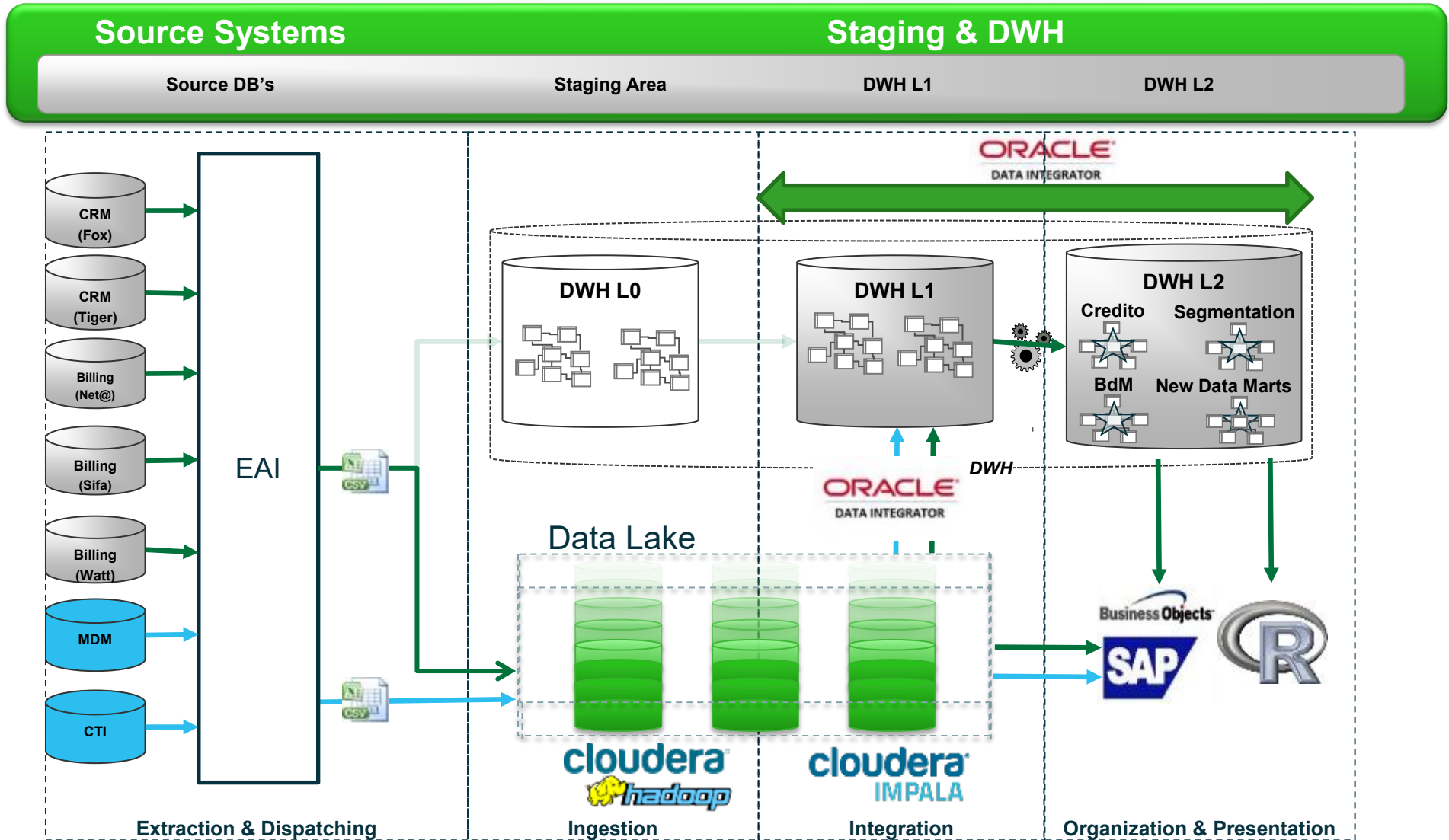
- Assessment Tecnico/Funzionale sui sistemi di BI con evidenza delle aree di sofferenza e miglioramento.
- Revisione del modello di Data Warehouse in ottica cliente centrico.
- Revisione del layer di presentazione con cui nuove funzionalità di business.
- Estensione del DWH con dati di: Billing, Customer Segmentation, CRM, Metering, CTI.
- Introduzione architettura Ibrida BigData + DWH con implementazione DataLake su piattaforma Cloudera e porting Staging Area.
- Introduzione del modello di Churn su tecnologia R.

Challenge

- Portare il servizio DWH a livelli di eccellenza aumentando il bacino di utenza.
- Valutare lo stato attuale della piattaforma BI per interventi migliorativi:
 - 1) Backend, migliorare le prestazioni del layer ETL-ODI per diminuire i tempi di fornitura dei dati.
 - 2) FrontEnd, migliorare la fruibilità dei dati forniti dal layer SAP-BO per gli utenti business.
- Delinare lo scenario target per la BI, tra cui:
 - 1) QuickWin Segmentation: iniziativa tattica di integrazione in DWH della segmentazione clienti: Customer BASE, Customer VALUE, Customer PROFITABILITY
 - 2) Customer Analytics: iniziativa di integrazione in DWH delle tematiche di CRM per potenziare le capacità di analisi dei Business Users sulla customer base aziendale.







A seguire i benefici apportati dall'adozione della tecnologia Hadoop:

- ❑ Data Warehouse,
 - ETL Acceleration: Messa a terra dei calcoli effettuati da ODI nei flussi di alimentazione del livello L1 del DWH (spostati su Hadoop)
 - Offloading / Hadoop Staging Area : i dati originali («raw»), i dati derivati da processi di trasformazione e i dati storici (o «meno nobili») sono mantenuti e gestiti in Hadoop, in modo da scaricare il golden layer (DWH e data mart)

- ❑ Infrastruttura/Storage
 - Costi Storage & Licensing: riduzione dei costi infrastrutturali grazie all'utilizzo di hardware non specializzato e software senza costi di licenza (solo subscription per supporto)
 - Active Archiving: i dati sono sempre accessibili, lo storico è sempre on-line. Il backup è parte nativa della tecnologia – non è necessario backup storage ad alte prestazioni, avere un datacenter su big data
 - Hardware Disomogeneo: Hadoop gestisce in maniera trasparente all'utente il deploy software su macchine con configurazioni differenti (i.e. il datacenter può avere macchine server – nodi del cluster – disomogenei senza impattare le prestazioni della soluzione)

- ❑ Progettuale/Processo
 - Modello dati «on-read»: I dati sono accessibili nella loro forma originale o comunque senza necessità di processi ETL a monte. I report SAP-BO possono essere collegati direttamente ad Hadoop su dati massivi/destrutturati



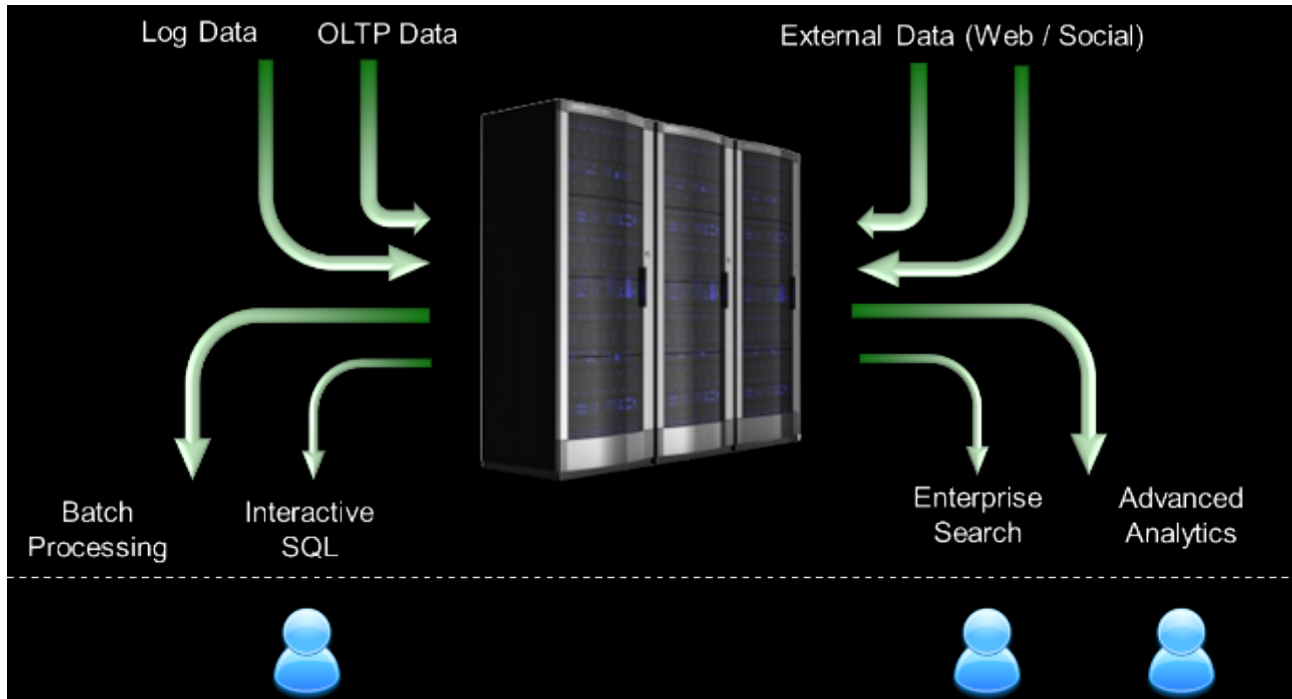
Pains

- Critical response time for cross-department incidents (up to 1 day)
- Post analysis is limited in time and granularity of information
- Impossible to prevent cross-departmental system overheads



Actions & Benefits

- Full Cloudera Stack on a medium size cluster
- Data collected in streaming with Flume from source systems
- Unique data platform which collects data and serves all different areas
- Correlation analysis on cross-functional logs



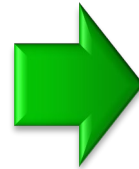
Data

- 6 nodes
- 15 TBs data at rest
- 200 GB data ingested each day
- 10 IT departments



Pains

- Protect production environment from dangerous accesses
- Provide data with a very low latency
- Access and search data with «natural language»

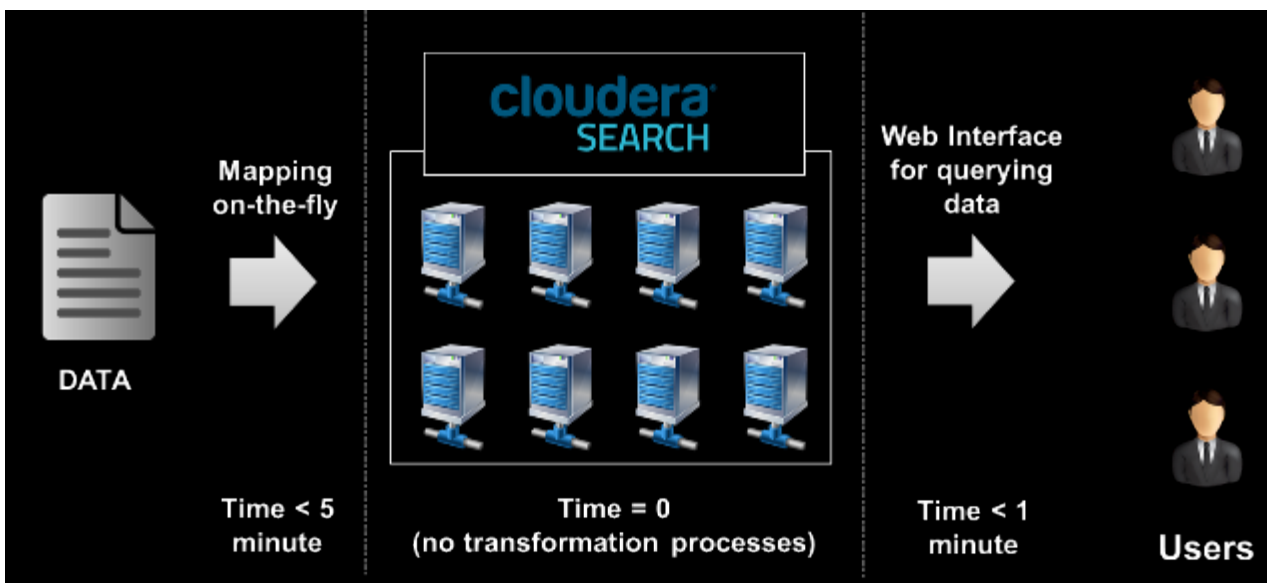


Actions & Benefits

- Central repository system to bring data out of critical production systems
- Stream data from sources to Cloudera
- Data indexed for very fast query and low latency retrieving

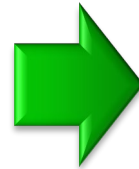
Data

- 24 nodes
- More than 2500 source systems
- Weekly retention of 105 TB of data



Pains

- Iterative analysis on customer transactions based on the context
- Slow response time to business users
- Data analysis for different categories of users



Actions & Benefits

- Data collected in streaming with Flume from mainframe
- Integration of Cloudera Impala and data discovery tools for interactive analysis
- Deliver of dashboards

Report & Dashboard



H
Q
L



Mainframe

Batch



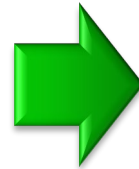
Data

- 4 nodes
- 1 TBs data at rest
- 1 GB data ingested each day



Pains

- Define KPIs on unstructured data
- Real-time analysis on heterogeneous data and sources



Actions & Benefits

- Data collected in streaming with Flume from source systems
- Data provide to front-end without any ETL processes

Data

- 10 nodes
- 30 TB (1 Year Retention)
- 1 minute latency for data availability



Thanks

d.bianchi@reply.it

g.benelli@reply.it

