

**DOVIKNITE
SVIMA DA DOĐU!**



T...

HT – DWH, BIG DATA

HRVATSKI TELEKOM: GORAN KESEGIĆ, ADRIJAN ĐURIN, IVAN VRATARIĆ

AGENDA

01

BUSINESS INTELLIGENCE (BI)

02

DWH - SKLADIŠTE PODATAKA

03

BIG DATA – ARHITEKTURA

04

BIG DATA – HALO PROJEKT

05

Apache Airflow

06

BI tim

BUSINESS INTELLIGENCE (BI)

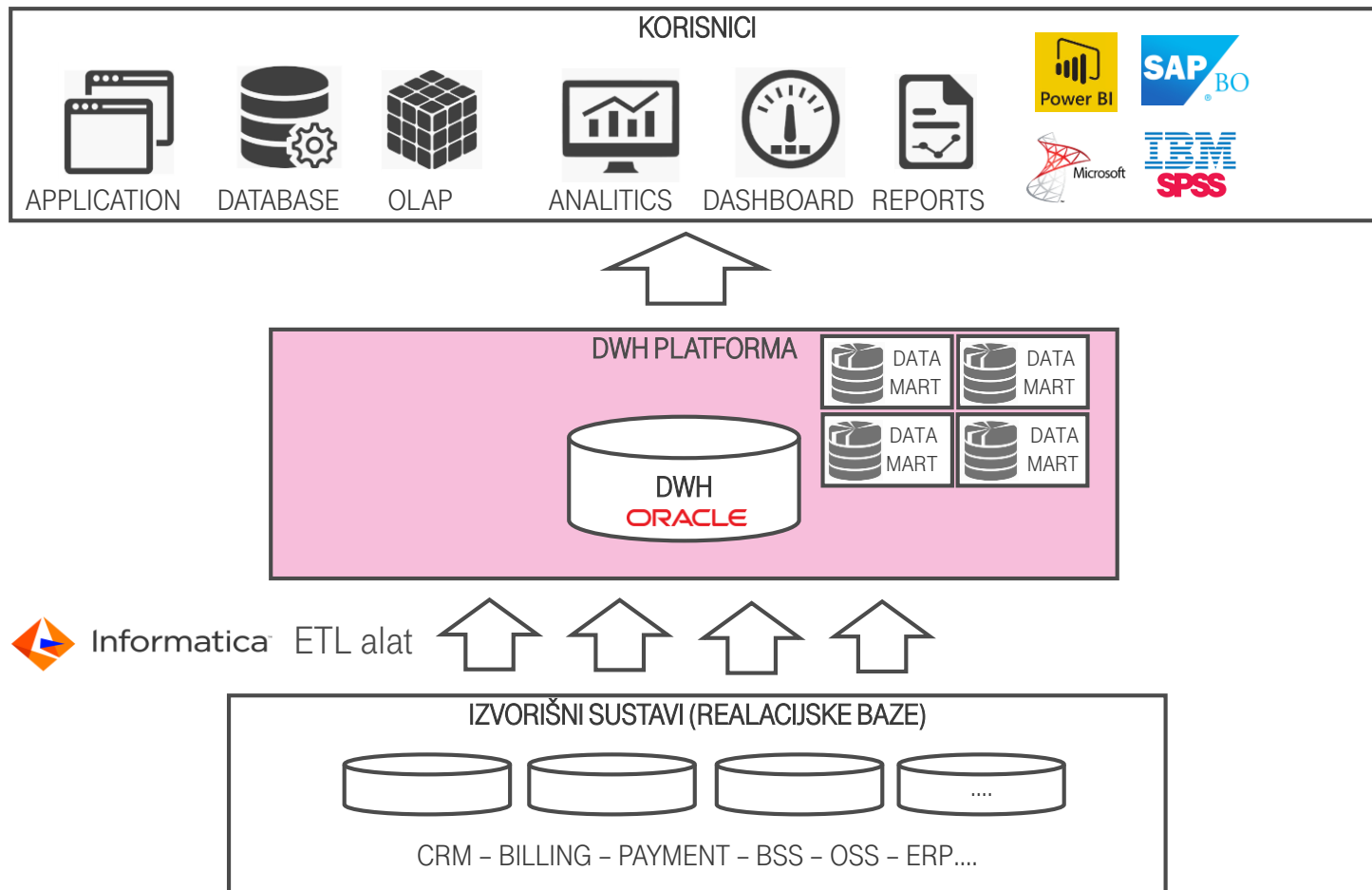
BUSINESS INTELLIGENCE (poslovna inteligencija) najčešće je definiran kao sustav za podršku i potporu poslovnom odlučivanju. Sam pojam objedinjuje skup metodologija, softverskih alata i hardvera koji omogućuje poslovanju pristup podacima i analitiku nad podacima s ciljem unaprjeđenja poslovanja . Dva su ključna pojma poslovne inteligencije: **DWH, BIG DATA**.

DATA WAREHOUSE (DWH) je subjektno orijentiran (podaci organizirani po poslovnim temama), integriran, postojan i vremenski različit skup podataka koji služi kao potpora odlučivanju. U skladištu podataka, podaci se integriraju iz različitih izvora. Proces integracije podataka jedan je od najzahtjevnijih procesa pri skladištenju podataka.

BIG DATA tehnološki koncept za prikupljanje, obradu i analizu velike količine podataka, koji su opsegom, kompleksnošću i brzinom dolaska veliki.



DWH – ARHITEKTURA

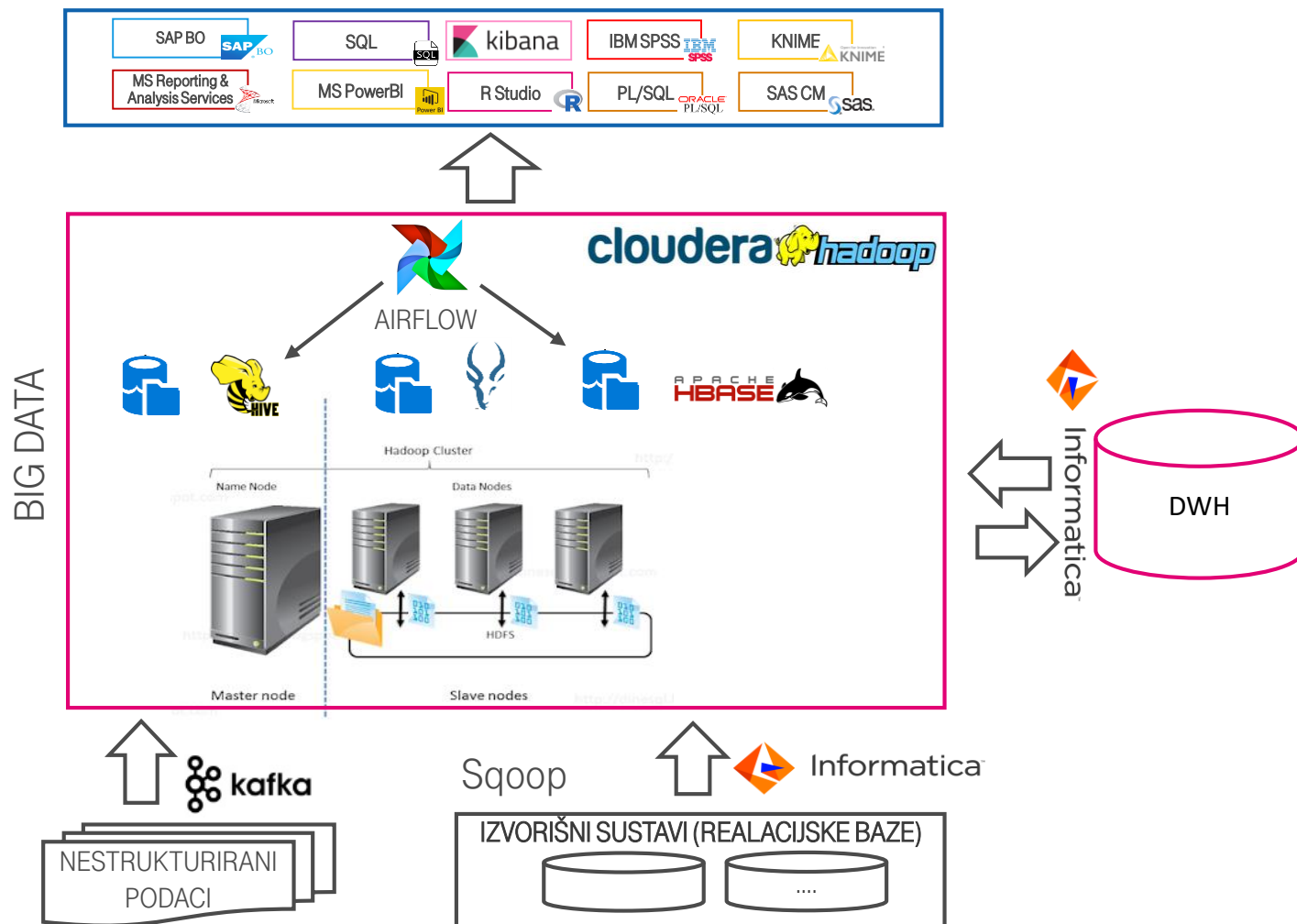


- **Informatica Power Center** - ETL (extract transform load) alat za dohvat podatka sa izvorišnih sustava, integraciju i transformaciju podatka
- **Dnevna obrada > 1 TB podatka**
- **ORACLE RDBMS** – baza za skladište podatka
- **Veličina skladišta > 20 TB podatka**

Izveštavanje i analitika:

- Izveštajni alati:
 - Microsoft Power BI
 - Microsoft reporting and analysis services
 - SAP busniess objects
- Broj internih korisnika izvještaja (dnevno) > 1000

BIG DATA ARHITEKTURA

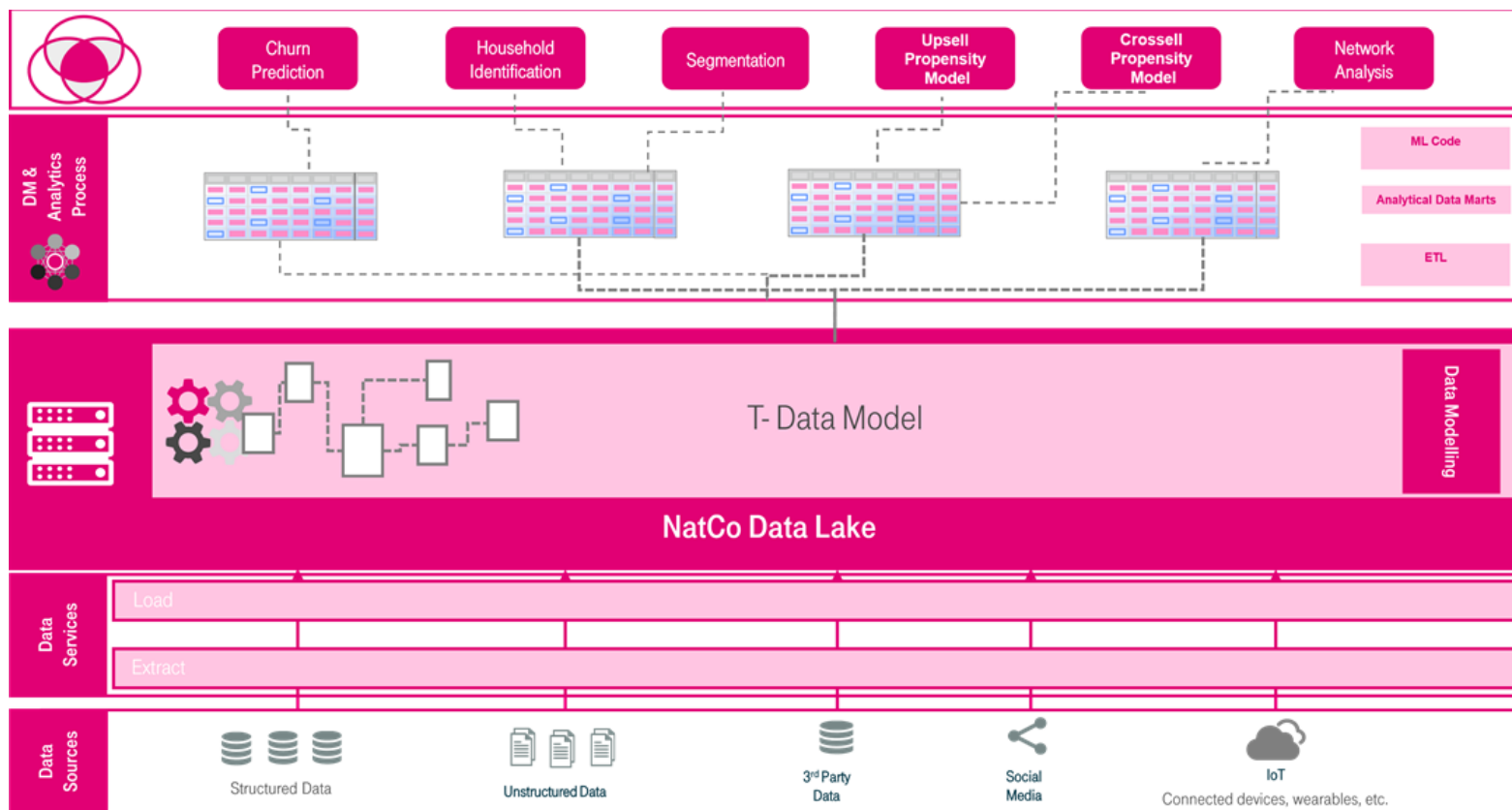


- **Cloudera** - open source distribucija apache hadoop platforme
- **Hadoop** – framework za distribuiranu obradu podataka.
- **ETL alati:**
 - Informatica BDM
 - Sqoop
- Real – time ingestion: **Kafka**
- **Airflow** – alat za orkestraciju taskova na clouderi
- **Hadoop Database:**
 - Apache Hive
 - Apache Impala
 - Apache Hbase
- **Analitički alati:**
 - Rstudio
 - Knime
 - IBM spss



BIG DATA – HALO PROJEKT

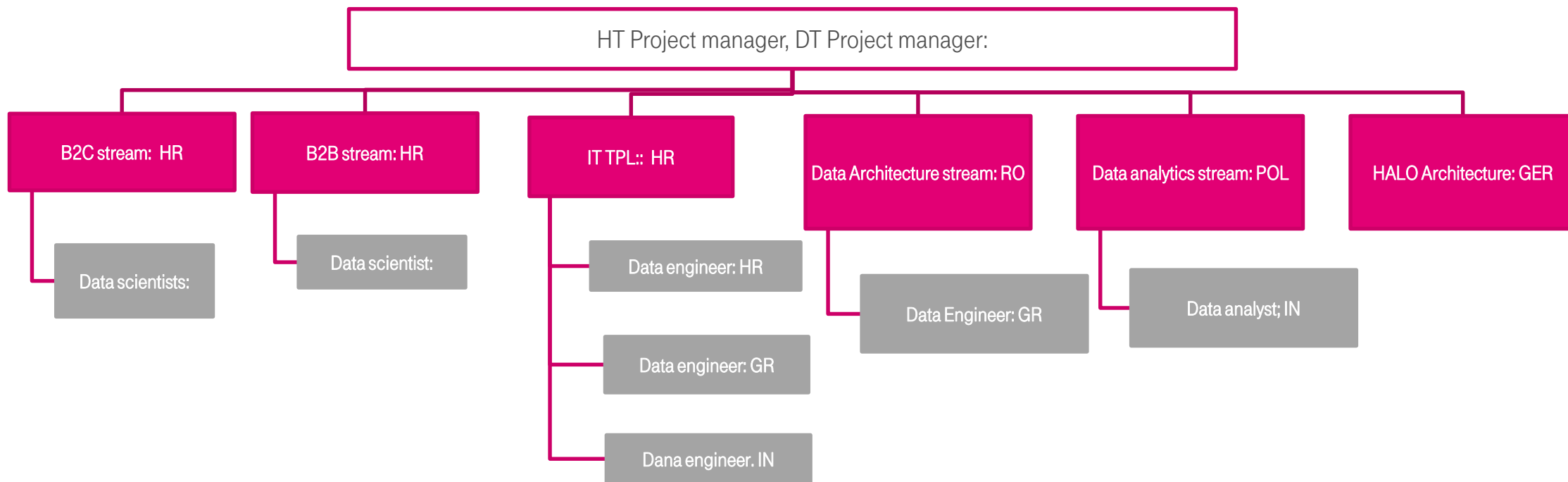
HALO PROJEKT - projekt na razini Deutsche Telekom grupe. Svaka članica DT grupe na lokalnoj BIG DATA-i ima isti podatkovni model (T – Data Model). Podaci su podloga za razvoj analitičkih modela (npr. model za predikciju otkaza usluge od strane korisnika, analiza mreže....)



LIFE IS FOR SHARING.

BIG DATA – HALO PROJEKT

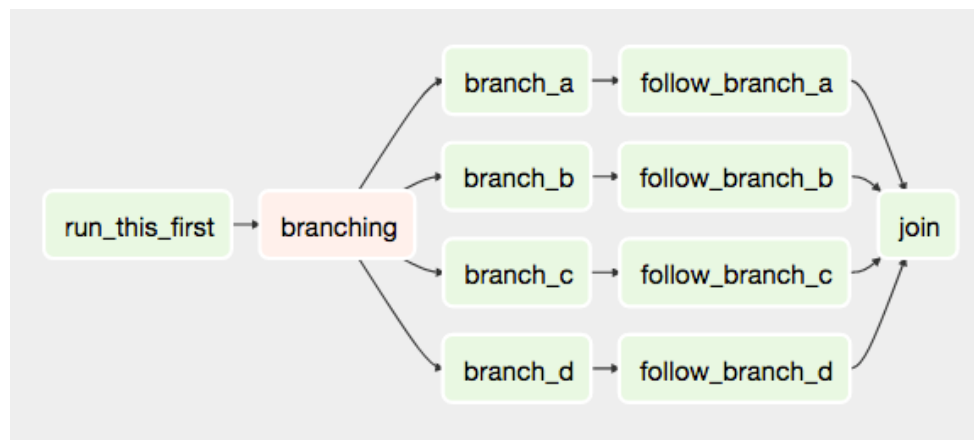
1. HALO projektni tim broji > 20 članova
2. Radimo u **multinacionalnom okruženju** – razvojni tim iz Indije i Grčke.
3. HT IT tim zadužen je za izradu ETL-a na BIG DATA-i, na dnevnoj bazi pruža podršku razvojnom timu izvan HR.
4. DEVOPS metodologija rada



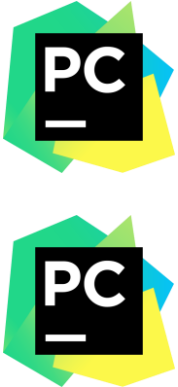
APACHE AIRFLOW



- Developed by Airbnb [2014]
- Open source platform [github.com/apache/airflow]
- Python based
- Author, schedule and monitor workflows



HT DEVOPS OKRUŽENJE



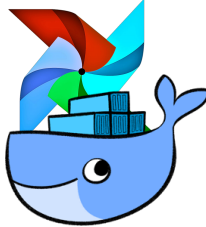
commit



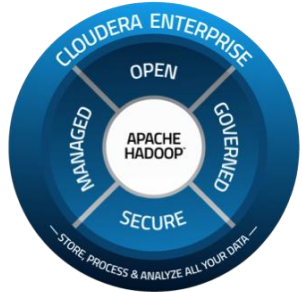
trigger



deploy



submit



Continuous Integration and Continuous Delivery

Deployment is done at the end of the each Jira sprint



BI ODJEL

- 18 DJELATNIKA
- 2 LOKACIJE:
 - ZAGREB: 12
 - VINKOVCI: 6
- Razvojni tim: E2E razvoj
- Tehnologije, alati : Informatica Power Center, PostgreSQL, ORACLE, MS SQL, PL/SQL, PYTHON, JAVA, OS LINUX,
- Izveštajni alati, vizualizacija : MS OLAP CUBE, MS Reporting services, Power BI, Kibana



LIFE IS FOR SHARING.

Hvala na pažnji !



LIFE IS FOR SHARING.