

IST 722 - Data Warehouse

Homework Assignment 2

Ryan Timbrook

NetID: RTIMBROO

Term: Winter, 2020

Date: 1/22/2020

Topic: Making Sense of the Market

For each offering listed, evaluate the product or service to understand how it fits within the technical, budgetary, cultural, and regulatory facets of an organization.

Homework Assignment 2

Table of Contents

1	Pentaho Data Integration	3
1.1	Offerings Q&A	3
2	MemSQL	5
2.1	Offerings Q&A	5
3	Microsoft Dynamics GP	6
3.1	Offerings Q&A	6
4	Birst	7
4.1	Offerings Q&A	7
5	Snaplogic	8
5.1	Offerings Q&A	8
6	PostgreSQL	9
6.1	Offerings Q&A	9
7	Informatica ILM	10
7.1	Offerings Q&A	10

Homework Assignment 2

1 Pentaho Data Integration

Pentaho Data Integration (PDI) provides the Extract, Transform, and Load (ETL) capabilities that facilitates the process of capturing, cleansing, and storing **data** using a uniform and consistent format that is accessible and relevant to end users and IoT technologies.

Common uses of Pentaho Data Integration include:

- Data migration between different databases and applications
- Loading huge data sets into databases taking full advantage of cloud, clustered and massively parallel processing environments
- Data Cleansing with steps ranging from very simple to very complex transformations
- Data Integration including the ability to leverage real-time ETL as a data source for Pentaho Reporting
- Data warehouse population with built-in support for slowly changing dimensions and surrogate key creation (as described above)

End-to-End Architecture Pentaho makes it easy to onboard a wide variety of data sources into your data management environment. Using a drag-and-drop user interface, you can blend, cleanse and standardize data quickly. Data scientist can engineer new features and pull this prepared data, on demand, to train, tune and test machine learning models. Data engineer can then deploy these models into a production environment and transform your business. Finally, to update models, data scientist can regularly use new training data with the transformations already built in Pentaho.

1.1 Offerings Q&A

1. Website?
 - a. <https://help.pentaho.com/Documentation/8.3>
 - b. <https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho-platform/pentaho-data-integration.html>
 - c. <https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho-platform.html>
2. What does it do?
 - a. **Enables the blending of multiple disparate data sources including big data.**
 - b. **Delivers analytics-ready data to end users with visual tools.**
 - c. **Provides real-time analytic capabilities by consuming any integration job as a single data source.**
 - d. **Eliminates the need and costs of data staging.**
3. Is it a product (hardware/software), service (we do the work for you), or both?
 - a. Is a **Product** offering that is a component of a comprehensive platform used to access, integrate, manipulate, visualize, and analyze disparate data sources.
4. Is it installed onsite, or is it in the cloud?
 - a. **Onsite**, however due to the light-weight design of it's architecture, that it's containerized, it is a 'Cloud Native' ready system that could be deployed in a Cloud Service Provider such as Amazon AWS.
5. Is it systems or technical architecture?

Homework Assignment 2

a. Technical Architecture

i. If systems architecture, is it SMP, MPP, or Hadoop/MapReduce?

1. Physical Infrastructure - Pentaho Platform

ii. If technical architecture, which part of the corporate information factory does it aim to address? Data mart? IM&T/ETL? ODS/NDS/DDS? External world application? Cross-media storage? And so on.

1. Data Mart
2. DSS Application
3. External World Applications

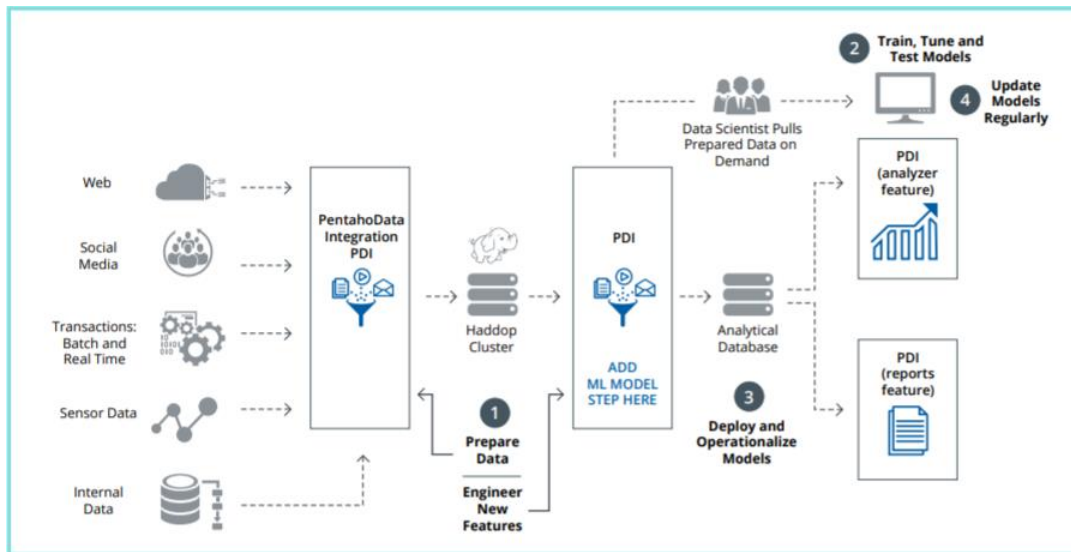


Figure 2. Deploy machine learning models using Pentaho in a complex data environment.

Homework Assignment 2

2 MemSQL

MemSQL is a distributed, relational database that handles both transactions and real-time analytics at scale. It is accessible through standard SQL drivers and supports ANSI SQL syntax including joins, filters, and analytical capabilities (e.g. aggregates, group by, and windowing functions).

MemSQL scales horizontally on cloud instances or industry-standard hardware, providing high throughput across a wide range of platforms. The MemSQL database maintains broad compatibility with common technologies in the modern data processing ecosystem (e.g. orchestration platforms, developer IDEs, and BI tools), so you can easily integrate it in your existing environment. It features an in-memory rowstore and an on-disk columnstore to handle both highly concurrent operational and analytical workloads. MemSQL also features a data ingestion technology called MemSQL Pipelines that streams large amounts of data at high throughput into the database with exactly-once semantics.

2.1 Offerings Q&A

1. Website?
 - a. <https://www.memsql.com/>
2. What does it do?
 - a. Self-service, **real-time analytics** at scale
3. Is it a product (hardware/software), service (we do the work for you), or both?
 - a. It's a **Product**
4. Is it installed onsite, or is it in the cloud?
 - a. **It can be installed onsite, in the cloud, or as a managed service.**
5. Is it systems or technical architecture? <https://docs.memsql.com/v6.8/concepts/distributed-architecture/>
 - a. **Technical Architecture**
 - i. If systems architecture, is it SMP, MPP, or Hadoop/MapReduce?
 1. N/A
 - ii. If technical architecture, which part of the corporate information factory does it aim to address? Data mart? IM&T/ETL? ODS/NDS/DDS? External world application? Cross-media storage? And so on.
 1. Data Mart
 2. DSS Application
 3. External World Applications



Homework Assignment 2

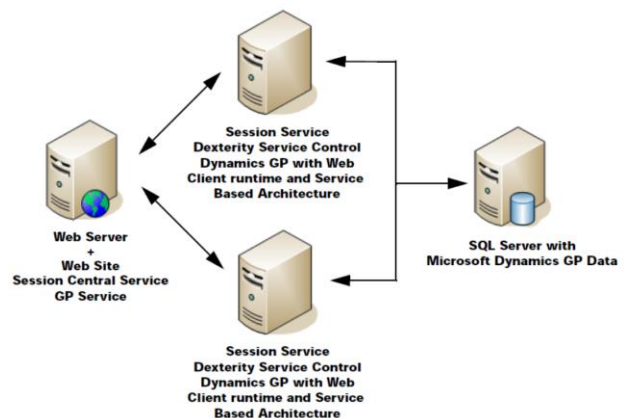
3 Microsoft Dynamics GP

Microsoft Dynamics GP (formerly Great Plains) is full featured financial management/ERP software.

Microsoft Dynamics GP has applications for financial management, human resources management, manufacturing planning, supply chain management, field service, business intelligence, collaboration, compliance and IT management. There are also a la carte modules which can be purchased separately, along with hundreds of third-party applications from independent software vendors (ISVs), which can be added to the system to fit additional needs.

3.1 Offerings Q&A

1. Website?
 - a. <https://dynamics.microsoft.com/en-us/gp-overview/>
2. What does it do?
 - a. Self-service, **internal business management analytics**
3. Is it a product (hardware/software), service (we do the work for you), or both?
 - a. It is a **Product**
4. Is it installed onsite, or is it in the cloud?
 - a. Hybrid, it can be installed **onsite** or in the **cloud**.
5. Is it systems or technical architecture? <https://docs.microsoft.com/en-us/dynamics-gp/web-components/deployment-configurations>
 - a. **Technical Architecture**
 - i. If systems architecture, is it SMP, MPP, or Hadoop/MapReduce?
 1. N/A
 - ii. If technical architecture, which part of the corporate information factory does it aim to address? Data mart? IM&T/ETL? ODS/NDS/DDS? External world application? Cross-media storage?
 1. Department Data Marts
 2. ODS
 3. DSS Applications



Homework Assignment 2

4 Birst

Birst is the only enterprise business intelligence platform that connects together the entire organization through a network of interwoven virtualized BI instances on-top a shared common analytical fabric. Birst enterprise BI delivers the speed, self-service, and agility front-line business workers demand, and the scale, security, and control to meet rigorous corporate data standards. Birst delivers all of this and much more with low TCO via public or private cloud configurations.

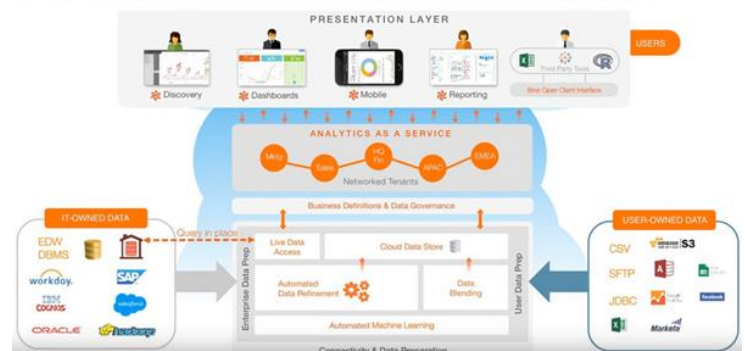
Highlights:

- Networked business intelligence
- Smart analytics
- Agile semantic layer
- Blend centralized and de-centralized data
- Adaptive user experience
- Information accessible anywhere from any device
- Modern native cloud architecture

4.1 Offerings Q&A

1. Website?
 - a. <https://www.birst.com/>
2. What does it do?
 - a. Self-service, **business analytics** at scale
3. Is it a product (hardware/software), service (we do the work for you), or both?
 - a. It is a **Product**
4. Is it installed onsite, or is it in the cloud?
 - a. **Cloud**, Software-as-a-Service (SaaS), or installed on a company's internal cloud architecture.
5. Is it systems or technical architecture?
 - a. **Technical** Architecture
 - i. If systems architecture, is it SMP, MPP, or Hadoop/MapReduce?
 1. NA
 - ii. If technical architecture
 1. ODS
 2. DSS Application

Birst's modern enterprise-class architecture



Homework Assignment 2

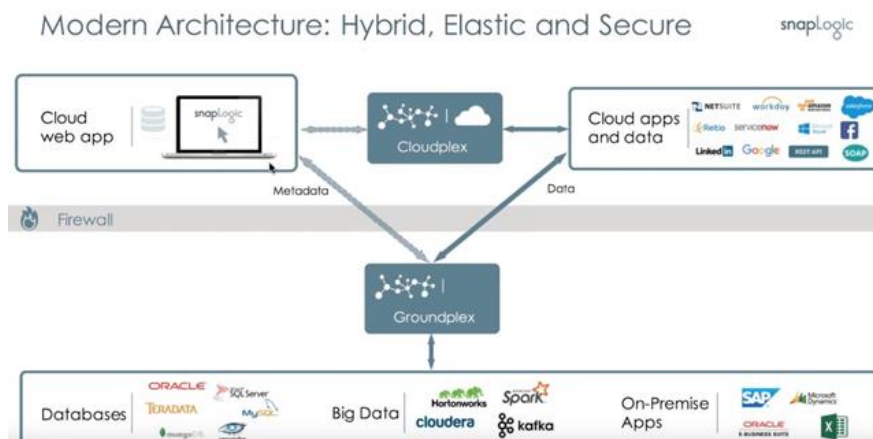
5 Snaplogic

SnapLogic is a commercial software company that provides Integration Platform as a Service tools for connecting Cloud data sources, SaaS applications and on-premises business software applications.

The SnapLogic Intelligent Integration Platform (IIP) is a self-service, AI-powered platform that integrates data from any source, in any environment, and at any speed – all at scale. Their mission is to make it fast and easy for users to access, connect and analyze enterprise data to improve business processes, accelerate decision-making, and drive better business outcomes.

5.1 Offerings Q&A

1. Website?
 - a. <https://www.snaplogic.com/>
2. What does it do?
 - a. Self-service, **application and data integration**.
3. Is it a product (hardware/software), service (we do the work for you), or both?
 - a. It is a **Product**
4. Is it installed onsite, or is it in the cloud?
 - a. **Hybride**, it can be installed on a private cloud on-premises or utilize the iPaaS.
5. Is it systems or technical architecture?
 - a. **Technical** Architecture
 - i. If systems architecture, is it SMP, MPP, or Hadoop/MapReduce?
 - ii. If technical architecture, which part of the corporate information factory does it aim to address? Data mart? IM&T/ETL? ODS/NDS/DDS? External world application? Cross-media storage?
 1. ODS
 2. DSS Application



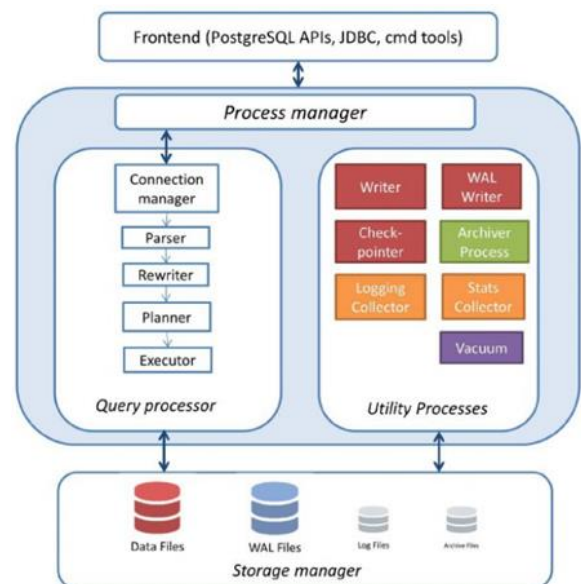
Homework Assignment 2

6 PostgreSQL

PostgreSQL, also known as Postgres, is a free and open-source relational database management system emphasizing extensibility and technical standards compliance. It is designed to handle a range of workloads, from single machines to data warehouses or Web services with many concurrent users.

6.1 Offerings Q&A

1. Website?
 - b. <https://www.postgresql.org/>
2. What does it do?
 - c. PostgreSQL is a general-purpose object-relational database management system.
3. Is it a product (hardware/software), service (we do the work for you), or both?
 - d. It is a **Product**
6. Is it installed onsite, or is it in the cloud?
 - a. It can be installed **onsite** or in the **cloud**.
1. Is it systems or technical architecture?
 - b. **Technical Architecture**
 - i. If systems architecture, is it SMP, MPP, or Hadoop/MapReduce?
 - ii. If technical architecture, which part of the corporate information factory does it aim to address? Data mart? IM&T/ETL? ODS/NDS/DDS? External world application? Cross-media storage?
 1. Data mart, ODS, NDS, DDS
 2. Cross media storage



Homework Assignment 2

7 Informatica ILM

Information lifecycle management refers to strategies for administering storage systems on computing devices. ILM is the practice of applying certain policies to effective information management. This practice had its basis in the management of information in paper or other physical forms.

From development, test, and production of critical data to the retirement and archiving phase, application ILM solutions help your IT organization align the business value of data with the most appropriate infrastructure to manage it.

The main activities are database archiving, test data management, data privacy/data masking, and data/application retirement. However, any systemic application of rules to business data or attempts to minimize, simplify, or increase the security of data can fall under the general procedures of ILM

7.1 Offerings Q&A

1. Website?
 - a. <https://www.informatica.com/ilm.html#fbid=9WnzEnMpkGK>
2. What does it do?
 - a. Data storage optimization as well as strategies to improve data quality and security.
3. Is it a product (hardware/software), service (we do the work for you), or both?
 - a. It is a Product
4. Is it installed onsite, or is it in the cloud?
 - a. onsite or the iPaaS offering
5. Is it systems or technical architecture?
 - a. Technical Architecture
 - i. If systems architecture, is it SMP, MPP, or Hadoop/MapReduce?
 - ii. If technical architecture, which part of the corporate information factory does it aim to address? Data mart? IM&T/ETL? ODS/NDS/DDS? External world application? Cross-media storage?

1. Data mart, DDS
2. DSS Application

