Environment:

This document introduces you to the data platform architecture of an ecommerce company named SoftCart.

SoftCart uses a hybrid architecture, with some of its databases on premises and some on cloud.

Tools and Technologies:

- OLTP database - MySQL
- NoSql database - MongoDB
- Staging Data warehouse – PostgreSQL
- Big data platform - Hadoop
- Big data analytics platform – Spark
- Business Intelligence Dashboard - IBM Cognos Analytics/Google Looker Studio
- Data Pipelines - Apache Airflow

Process:

SoftCart's online presence is primarily through its website, which customers access using a variety of devices like laptops, mobiles and tablets.

All the catalog data of the products is stored in the MongoDB NoSQL server.

All the transactional data like inventory and sales are stored in the MySQL database server.

SoftCart's webserver is driven entirely by these two databases.

Data is periodically extracted from the transactional database and put into the staging data warehouse running on PostgreSQL.

BI teams connect to IBM Cognos Analytics to create operational dashboards.

SoftCart uses Hadoop cluster as its big data platform where all the data is collected for analytics purposes.

Spark is used to analyse the data on the Hadoop cluster.

To move data between OLTP and the data warehouse, ETL pipelines are used and these run on Apache Airflow.