



Microsoft Cloud Workshop

Modernizing Data Analytics with SQL Server 2019

Whiteboard design session student guide

November 2019

Information in this document, including URL and other Internet Web site references, is subject to change without notice. Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, e-mail address, logo, person, place or event is intended or should be inferred. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The names of manufacturers, products, or URLs are provided for informational purposes only and Microsoft makes no representations and warranties, either expressed, implied, or statutory, regarding these manufacturers or the use of the products with any Microsoft technologies. The inclusion of a manufacturer or product does not imply endorsement of Microsoft of the manufacturer or product. Links may be provided to third party sites. Such sites are not under the control of Microsoft and Microsoft is not responsible for the contents of any linked site or any link contained in a linked site, or any changes or updates to such sites. Microsoft is not responsible for webcasting or any other form of transmission received from any linked site. Microsoft is providing these links to you only as a convenience, and the inclusion of any link does not imply endorsement of Microsoft of the site or the products contained therein.

© 2019 Microsoft Corporation. All rights reserved.

Microsoft and the trademarks listed at <https://www.microsoft.com/en-us/legal/intellectualproperty/Trademarks/Usage/General.aspx> are trademarks of the Microsoft group of companies. All other trademarks are property of their respective owners.

Contents

- [Modernizing Data Analytics with SQL Server 2019 whiteboard design session student guide](#)
 - [Abstract and learning objectives](#)
 - [Step 1: Review the customer case study](#)

- Customer situation
- Customer needs
- Customer objections
- Infographic for common scenarios
- Step 2: Design a proof of concept solution
- Step 3: Present the solution
- Wrap-up
- Additional references

Modernizing Data Analytics with SQL Server 2019

whiteboard design session student guide

Abstract and learning objectives

In this whiteboard design session, you will work with a group to design a solution for modernizing your large-scale data processing and machine learning capabilities through the use of SQL Server Big Data Clusters. You will evaluate the customer scenario and requirements to decide the best architecture that will meet their needs, while unifying data from disparate sources into a platform that help the customer gain business insights and apply advanced analytics at scale.

At the end of this whiteboard design session, you will be better able to design a modernization plan for performing Big Data analytics centered around SQL Server 2019 capabilities.

Step 1: Review the customer case study

Outcome

Analyze your customer's needs.

Timeframe: 15 minutes

Directions: With all participants in the session, the facilitator/SME presents an overview of the customer case study along with technical tips.

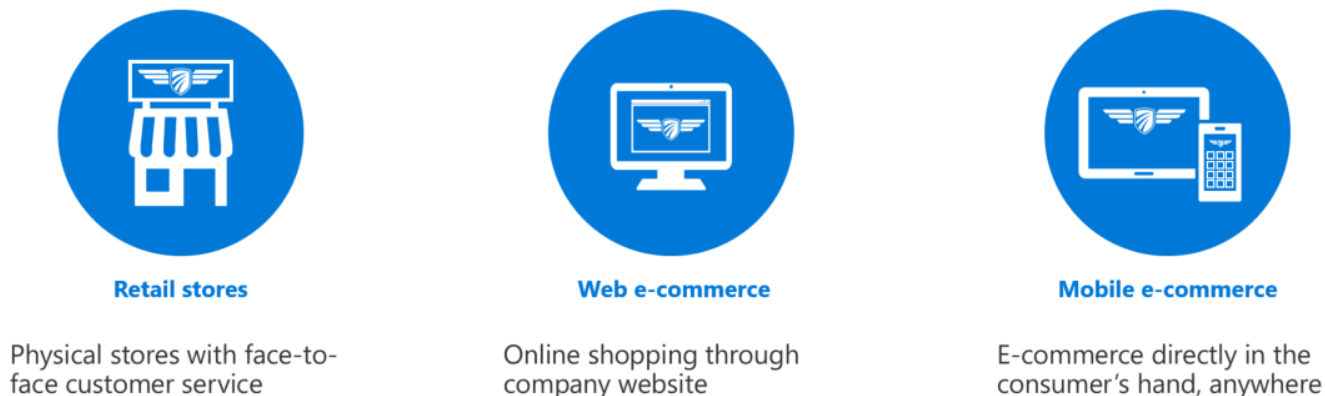
1. Meet your table participants and trainer.
2. Read all of the directions for steps 1-3 in the student guide.
3. As a table team, review the following customer case study.

Customer situation

Wide World Importers (WWI) is a traditional brick and mortar business with a long track record of success, generating profits through strong retail store sales of their unique offering of affordable products from around the world. They have a great training program for new employees, that focuses on connecting with their customers and providing great face-to-face customer service. This strong focus on customer relationships has helped set WWI apart from their competitors.

WWI's evolution of services over the years has helped them expand their reach beyond the walls of their retail stores into the web and mobile space. With this expansion, they have generated a significant amount of additional data, and data formats. These new platforms were added without integrating into the OLTP system data or Business Intelligence infrastructures. As a result, "silos" of data stores have developed.

This is their omni-channel strategy:

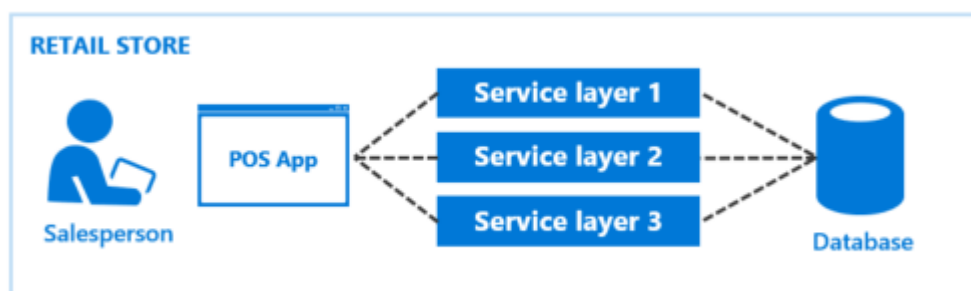


Due to their continued growth, lending to expansion into the digital space, WWI is prepared to innovate by taking advantage of their omni-channel strategy and increased variety and amount of valuable data. They believe they can foster innovation by building upon their track record of strong customer connections, and engage with their customers through personalized, high-quality application experiences that incorporate data and intelligence.

However, as a first step, WWI's technology team has recognized they must address the fact that they have quickly outgrown their ability to handle data. They anticipate the following solutions needed to reach more customers and grow the business:

- Scale data systems to reach more consumers
- Unlock business insights from multiple sources of structured and unstructured data
- Apply deep analytics with high-performance responses
- Infuse AI into apps to actively engage with customers

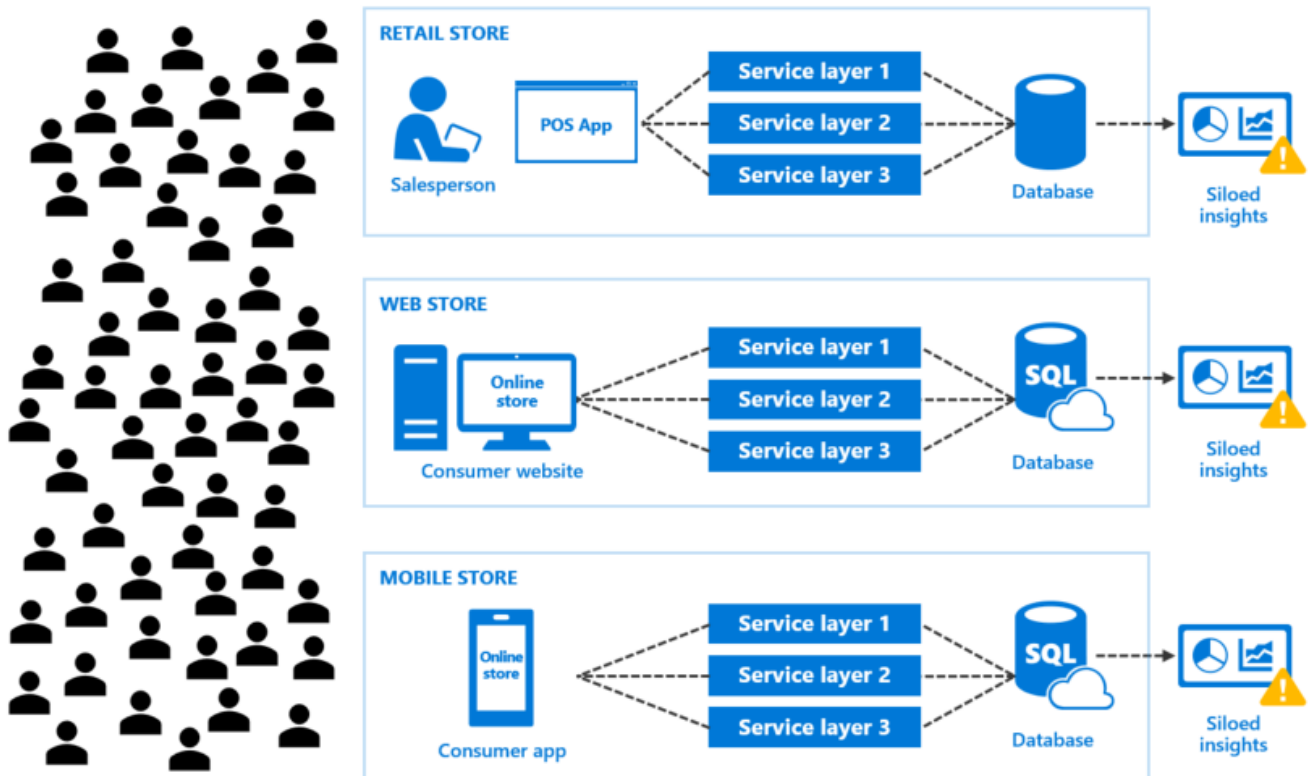
Prior to expanding to their current omni-channel strategy, WWI had a simple Point of Sale (POS) application that handled customer orders at each retail store. The back-end was a series of service layers used to process orders and store them in a SQL database. They had designed their systems and tuned them to handle this level of data.



As they added new e-commerce channels to expand the customer base, consumer demand also increased. This increased demand from more customers ordering products through more channels generated more data.

Now WWI has new challenges to address:

- Increased consumer demand, leading to increased app data
- They are unable to determine business trends because of siloed insights
- They have a rising data management footprint, increasing cost and complexity
- New development challenges resulting from more deployment targets and duplicated code



WWI has considered using a traditional data warehouse to join data from their disparate systems to gain insights in one location. Their biggest concerns with this approach are the amount of time to put such a system in place and properly maintain it, but even more so, having an inherent delay between when new data is written to the source system and when that data is moved into the data warehouse. They would like to have access to data from all systems while it is fresh, but do so in a way that is highly scalable and able to support complex joins between the external sources and internal SQL server tables as well.

However, there are times when they would like to move data into storage to provide denormalized and aggregated representations of their data for reporting purposes. In other terms, a data mart. However, they would like to also take advantage of distributed storage of this data, which would include sharding the data across multiple databases. They are concerned about the level of effort to shard their data, access that data, and maintain the distributed system.

There are two scenarios WWI is considering using AI to help grow their business and reduce costs:

1. Sales forecasting. Based on current and historical retail data, could they predict whether retail sales will be on track this month? Being able to meet sales targets while accurately forecasting sales revenue are critical success enablers by helping drive marketing campaigns and scale logistics and staffing accordingly.

2. Reduce maintenance costs, waste, and maximize fleet availability by predicting battery lifespans. Wide World Importers relies on refrigerated trucks to deliver temperature-sensitive products. A dead or malfunctioning battery could cause the cooling systems to fail, requiring regular battery testing and replacements. WWI would like to use transmitted sensor data from these trucks to predict when a battery will most likely fail to reduce downtime and cut waste resulting from fixed battery replacement schedules.

Customer needs

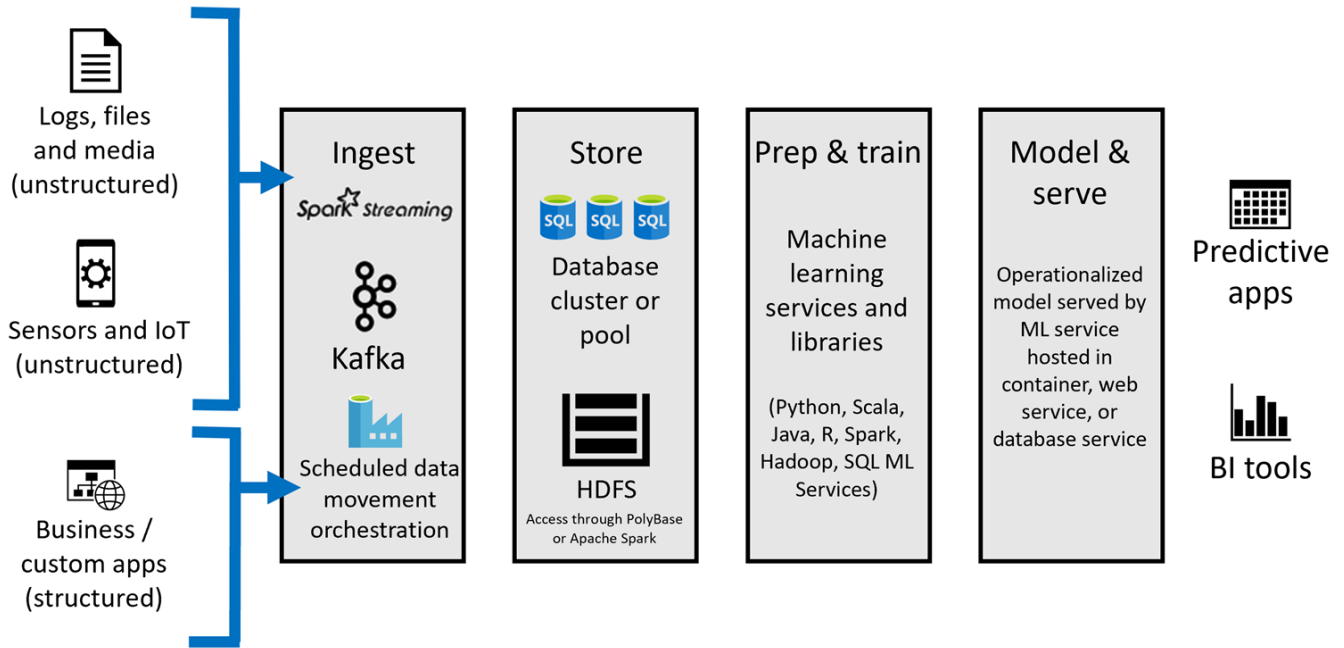
1. Need distributed storage available to all nodes of the container: The storage can disappear when the Container is removed, and other Containers and technologies can't access storage easily within a Container.
2. Require a data lake to easily store and access disparate data.
3. Would like a data mart to store denormalized and aggregated data while taking advantage of distributed storage.
4. Simplified programming surface to prepare data and do data science.
5. Scale data systems to reach more consumers.
6. Unlock business insights from multiple sources of structured and unstructured data.
7. Apply deep analytics with high-performance responses.
8. Enable AI into apps to actively engage with customers.
9. Identify PII and GDPR-related compliance issues for audit reports and take steps to fix these issues.

Customer objections

1. How do we centrally manage and monitor the cluster once deployed?
2. Do our workloads require us to use a data warehouse, or will a data mart suffice?
3. Will moving to container-based SQL clusters be complex and too high of an operational and management cost for our IT team?
4. How can SQL Server 2019 help us protect PII data and remain GDPR compliant?

Infographic for common scenarios

The following infographic can be used as inspiration when designing your solution.



Step 2: Design a proof of concept solution

Outcome

Design a solution and prepare to present the solution to the target customer audience in a 15-minute chalk-talk format.

Timeframe: 60 minutes

Business needs

Directions: With all participants at your table, answer the following questions and list the answers on a flip chart:

1. Who should you present this solution to? Who is your target customer audience? Who are the decision makers?
2. What customer business needs do you need to address with your solution?

Design

Directions: With all participants at your table, respond to the following questions on a flip chart:

High-level architecture

1. Diagram your initial vision for the architecture of the solution.

Big data and insights

1. What services and technologies should be used for scale-out processing and analyzing big data? Can this be done while minimizing code changes?
2. How will you enable a single data query to work across multiple, disparate data sources with the ability to join internal SQL server tables at scale?

3. How will you provide a data mart to store denormalized and aggregated data while taking advantage of distributed storage? Would you suggest using a data warehouse instead?
4. What methods can be used to ensure the best performance when querying data?

Deep analytics and AI

1. What would be used to solve the AI requirements?
2. How will you execute and train the Machine Learning model(s) used for the solution?

Monitor and Troubleshoot

1. How will you monitor and troubleshoot issues with the big data cluster?

Prepare

Directions: With all participants at your table:

1. Identify any customer needs that are not addressed with the proposed solution.
2. Identify the benefits of your solution.
3. Determine how you will respond to the customer's objections.

Prepare a 15-minute chalk-talk style presentation to the customer.

Step 3: Present the solution

Outcome

Present a solution to the target customer audience in a 15-minute chalk-talk format.

Timeframe: 30 minutes

Presentation

Directions:

1. Pair with another table.
2. One table is the Microsoft team and the other table is the customer.
3. The Microsoft team presents their proposed solution to the customer.
4. The customer makes one of the objections from the list of objections.
5. The Microsoft team responds to the objection.
6. The customer team gives feedback to the Microsoft team.
7. Tables switch roles and repeat Steps 2-6.

Wrap-up

Timeframe: 15 minutes

Directions: Tables reconvene with the larger group to hear the facilitator/SME share the preferred solution for the case study.

Additional references

Description	Links
What are SQL Server big data clusters?	https://docs.microsoft.com/en-us/sql/big-data-cluster/big-data-cluster-overview?view=sqlallproducts-allversions
How to use notebooks in SQL Server 2019	https://docs.microsoft.com/en-us/sql/big-data-cluster/notebooks-guidance?view=sql-server-ver15
SQL Server 2019 Master Instance	https://docs.microsoft.com/en-us/sql/big-data-cluster/concept-master-instance?view=sqlallproducts-allversions
SQL Server extensibility framework	https://docs.microsoft.com/en-us/sql/advanced-analytics/?view=sql-server-ver15
PolyBase	https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide?view=sql-server-ver15
Dynamic Data Masking	https://docs.microsoft.com/en-us/sql/relational-databases/security/dynamic-data-masking
Row-Level Security	https://docs.microsoft.com/en-us/sql/relational-databases/security/row-level-security
Always Encrypted with Secure Enclaves	https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-enclaves?view=sqlallproducts-allversions
SQL Data Discovery and Classification	https://docs.microsoft.com/en-us/sql/relational-databases/security/sql-data-discovery-and-classification?view=sql-server-ver15
Hadoop File System (HDFS)	https://kubernetes.io/docs/concepts/overview/components/
Apache Spark	https://spark.apache.org/