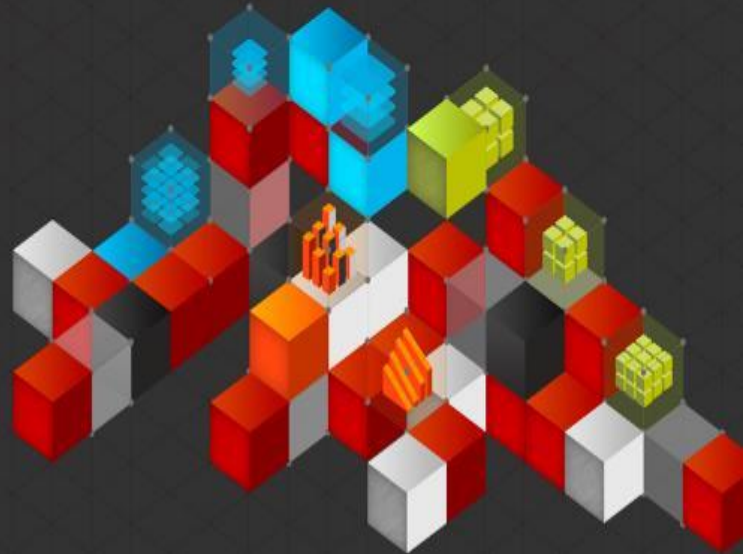


MICROSOFT'S ANALYTICS SOLUTIONS WITH PARALLEL DATA WAREHOUSE



Microsoft®
SQL Server® 2012
Parallel Data
Warehouse

Stefan Cronjaeger
Microsoft
May 2013

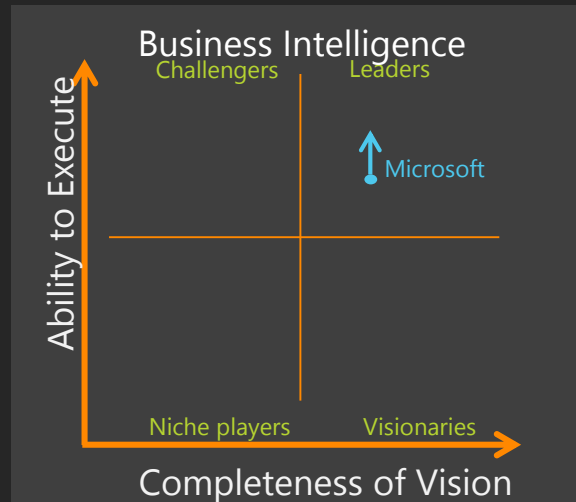
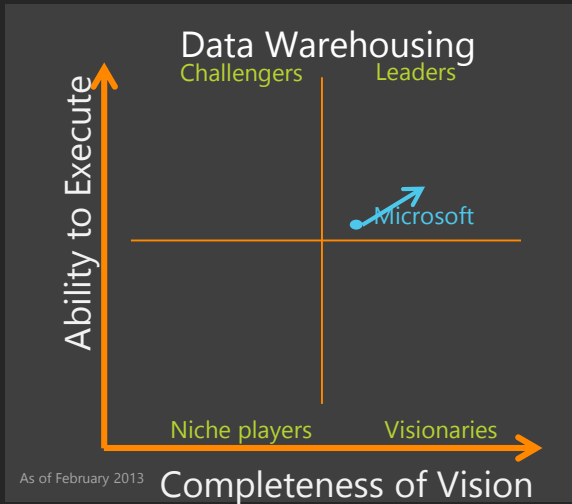


AGENDA

- PDW overview
- Columnstore and Big Data
- Business Intelligence Project



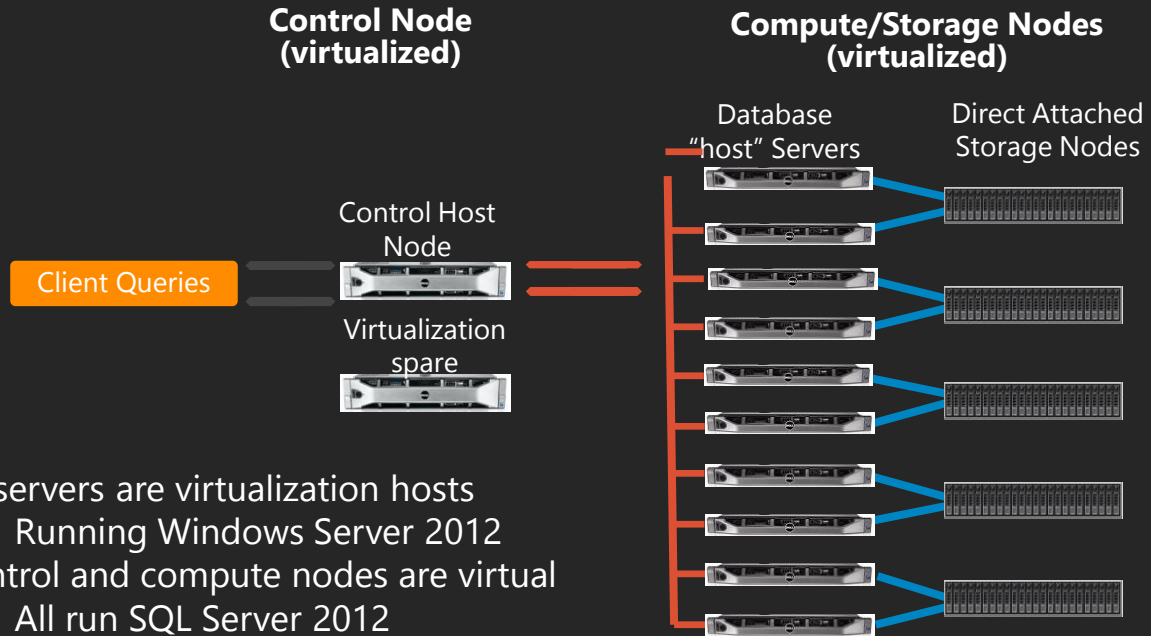
MICROSOFT'S LEADERSHIP: TOP AND MOVING UP



"Microsoft's Parallel Data Warehouse appliance, despite a slow start, has been adopted by approximately 100 organizations in the past 18 months. Further adoption of this appliance is likely as Dell continues to enter the data warehouse space and sells the Dell Parallel Data Warehouse Appliance" - *Gartner, February 2013*

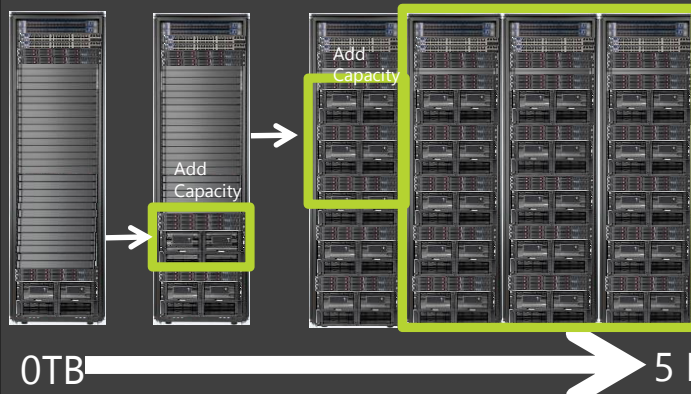
[Gartner, Inc., Magic Quadrant for Data Warehouse Database Management Systems: Magic Quadrant, Mark A. Beyer, Donald Feinberg, Roxane Edjaji, Merv Adrian, January 31st, 2013. The Magic Quadrant is copyrighted 2013 by Gartner, Inc. and is reused with permission. The Magic Quadrant is a graphical representation of a marketplace at and for a specific time period. It depicts Gartner's analysis of how certain vendors measure against criteria for that marketplace, as defined by Gartner. Gartner does not endorse any vendor, product or service depicted in the Magic Quadrant, and does not advise technology users to select only those vendors placed in the "Leaders" quadrant. The Magic Quadrant is intended solely as a research tool, and is not meant to be a specific guide to action. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

PDW LOGICAL ARCHITECTURE



- All servers are virtualization hosts
 - Running Windows Server 2012
- Control and compute nodes are virtual
 - All run SQL Server 2012
- Control node spreads data and workload across compute nodes
- Data loads are in parallel and take advantage of the power of all nodes

SCALABILITY



Smallest (0TB) To Largest (5PB)

- Start small with a few Terabyte warehouse
- Add capacity up to 5 Petabytes
- HW available from HP or Dell

Start Small And Grow


Largest Warehouse
PB

No Downtime


SOFTWARE

Windows Server 2012:

Control Node, Mgmt. Node and Compute Nodes run in virtualized Environment

System Center 2012:

Single user i/f for management of PDW, OS, BI, custom apps and private cloud

SQL Server 2012 inside

Visual Studio Data Tools
Powerview directly on PDW

Workload Management

Workload classes

xVelocity

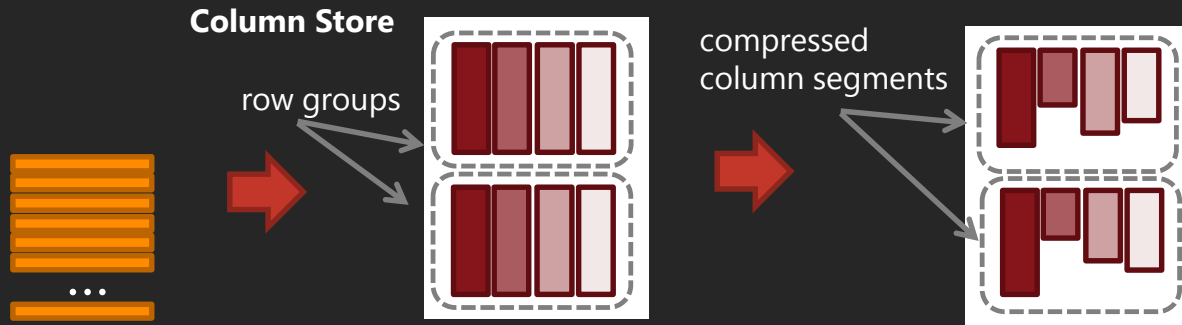
In-memory execution
Clustered columnstore

Big Data Integration

Polybase: T-SQL query to Hadoop
External tables on Hadoop

COLUMNSTORE STORAGE

A columnstore stores each column in a separate set of disk pages, rather than storing multiple rows per page



- A **segment** contains values of one column for a row group
- Highly compressible

COLUMN STORE ADVANTAGE

Compression of large blocks of same data type is very effective

In-Memory analytics:
More compressed data fits in memory, better buffer hit rates

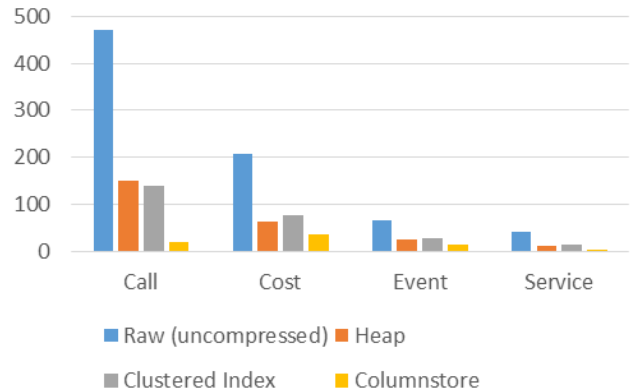
Read only the data from the column which are needed

DWH needs sequential read which is very fast. Fits well to star schema joins

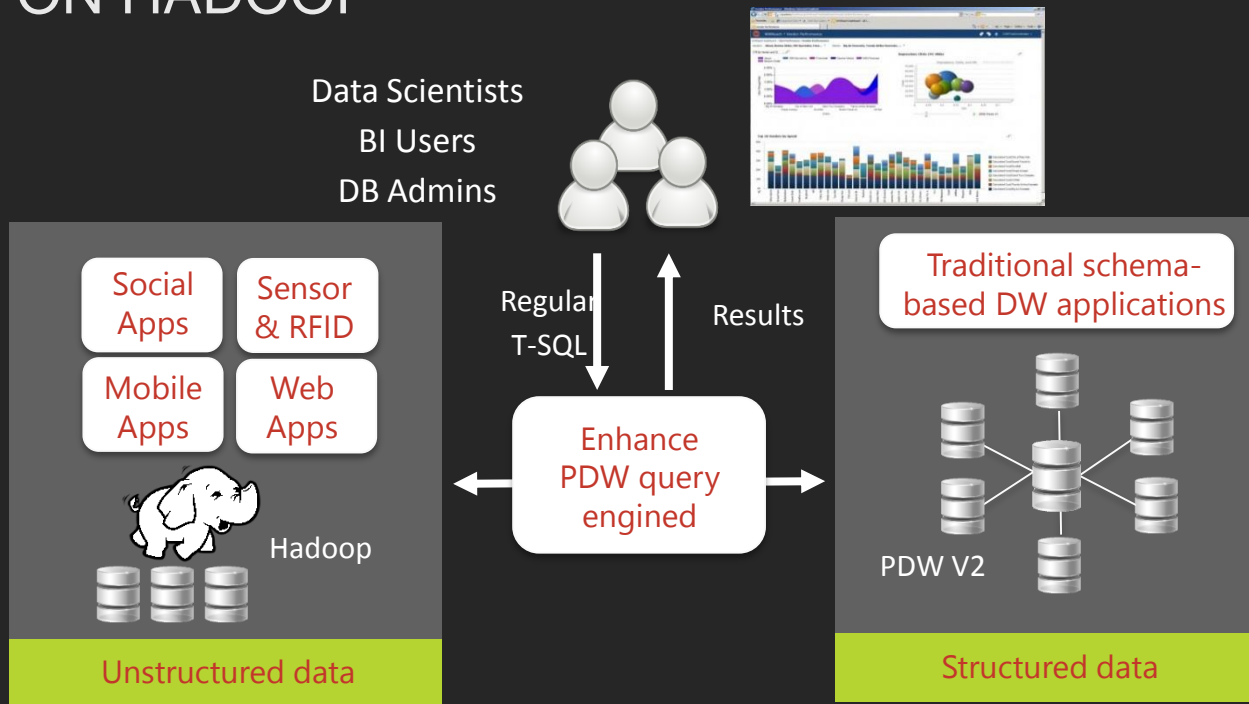
Batch mode processing: Large chunks of data use vectorized algorithms based on HW support

The columnstore supports Update, Merge

Real-world table size in GByte



POLYBASE – THE FULL SPECTRUM OF SQL ON HADOOP



- Highly efficient querying Hadoop and Parallel Data Warehouse via T-SQL
- Parallel transport of data between Massively Parallel PDW and Hadoop

REASONS TO USE POLYBASE

Use the established BI tools for Hadoop

- Analysis and Reporting Services
- PowerPivot and PowerView (Excel)
- Sharepoint interworking

Archive the data on Hadoop – Multi-temperature approach

- Old data on Hadoop – May be queried via SQL, no re-import of old archives necessary
- Fresh data on Parallel Data Warehouse
- Hot data in Memory – also on Parallel Data Warehouse

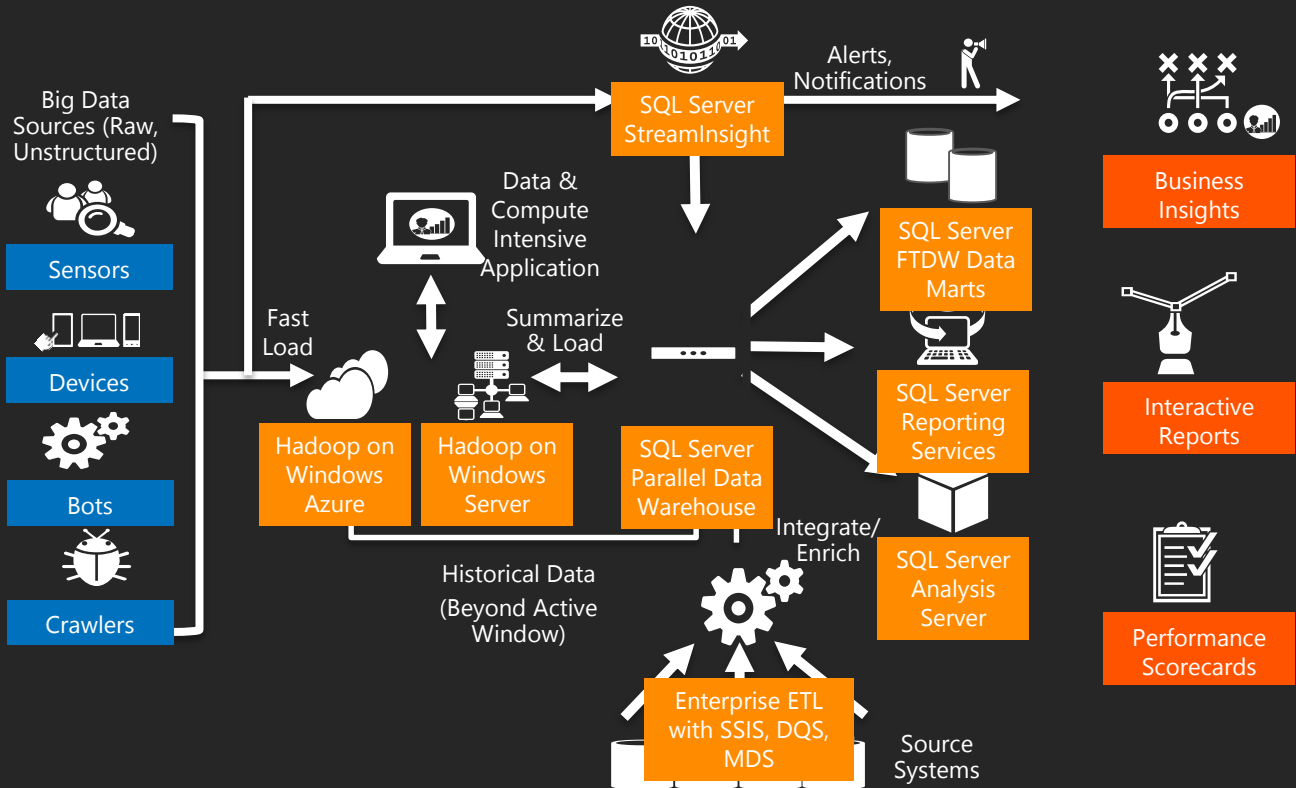
No need to learn implementing reports in Map/Reduce



Reporting example based on Hadoop data
<http://blogs.technet.com/b/dataplatforminsider/archive/2013/04/25/insight-through-integration-sql-server-2012-parallel-data-warehouse-polybase-demo.aspx>

Hadoop can be seamlessly integrated into existing BI and Web infrastructures

THE BIG (DATA) PICTURE

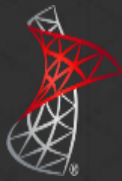


PDW DIFFERENTIATORS

- TCO: Very low, especially when looking on the whole bundle: ETL (SSIS), PDW, Data marts (SQL server) and Analytics (SSAS, SSRS)
- Microsoft product stack integration – SSIS, SSAS, SSRS, PowerPivot, System Center
- Linear Scaling via Shared Nothing
- Parallel Database Copy to SQL server
- xVelocity: Column Store and In-Memory execution
- Polybase: Integration with Big Data and Hadoop
- StreamInsight: Integration with Streaming data



"Microsoft exhibits one of the best value propositions on the market with a low cost and a highly favorable price/performance ratio" - Gartner, February 2012



Microsoft®
SQL Server® 2012
Parallel Data
Warehouse

© 2013 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.

MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.