

: *What are some use cases with expanded data sources?*

by Daniel J. Power

Editor, DSSResources.COM

Data volume is expanding rapidly, data variety is increasing and is coming from new sources, and data velocity or speed of capture, transmission and processing is faster and the need for real-time data is more common. The advent of expanded data sources has led to new interest in analyzing data and exploiting "big data" -- whatever that colorful phrase means to you. More data is the reality in organizations and managers want to know how the many new data sources can be used to benefit the organization. Use cases document how specific data sources can solve real needs for facts in making decisions.

The descriptors of the new data sources all begin with the letter V. Volume refers to the scale of the new data which is very large. Scale may be in terabytes, petabytes or exabytes. Variety because the new data is both structured and unstructured and from many source systems. Data variety means the new data is in many forms and formats. Data may be text comma separated values (CSV), photos or video. Velocity refer to the rate of change in the speed of data creation and transmission. The velocity is increasing. Data is speeding at us and from us. Data velocity ranges from batch processing to real-time processing. Variable is another characteristic. Some new data comes in unpredictable amounts. Some sources add veracity, the uncertainty of the data received or is it true. New data is of varying quality and truthfulness depending upon the source.

Because the expanded data sources are unfamiliar ask questions to assess the data. For example, How much data is relevant to the decision support task? How unstructured is the data? How fast is the data being created or how fast is the data changing? How many data formats are relevant to the task? Is it easy to change formats? How reliable is the data?

IBM describes "5 game changing big data use cases" at <http://www-01.ibm.com/software/data/bigdata/use-cases.html> . Datameer (2014) describes 5 similar use cases. The IBM website explains use cases, "a use case helps you solve a specific business challenge by using patterns or examples of technology solutions. Your use case, customized for your unique issue, provides answers to your business problem." Let me try to interpret the major current use case categories:

1. Customer analytics. Broadening the profile of individual customers by linking additional internal and external information to existing structured customer data. IBM describes creation of an "Enhanced 360° View of the Customer". Jay Khavani, Senior Manager of BI and Data Warehousing at The Lucky Group, observed in a Pentaho Press Release (2014) that "Analytics play a huge role with both customer acquisition and retention." Datameer eBook (2014) explains "you can use

: *What are some use cases with expanded data sources?*

insights about the customer acquisition journey to design campaigns that improve conversion rates. Or you can identify points of failure along the customer acquisition path – or the behavior of customers at risk of churn to proactively intervene and prevent losses. And you can better understand high-value customer behavior beyond profile segmentation ..."

2. Data-driven products and services, including personalization systems. Consolidation of data from multiple systems and sources for exploration and visualization. For example, combining data from system logs, sensors, or click streams with customer and line-of-business data. Personalization or recommender systems understand each person's unique habits and preferences and bring to light products and items that a user may be unaware of and not looking for. Turner (2014) explained how Paddy Power used Cassandra to provide real-time gambling products and pricing to customers. Pinckney (2013) reports that eBay chose NoSQL database Cassandra to power its next generation recommendation engine. EBay is "storing user activity data on Cassandra, representing it as a graph that is made up of edges between users and items that the user has indicated an interest or disinterest towards. As new behavioral data is recorded, in real time, we update our models about what the user is predicted to like or not."

3. EDW optimization and data warehouse modernization. Using new database technologies for pre-processing to determine what data should be moved to the data warehouse, offloading infrequently accessed data from data warehouses into enterprise-grade key-value data stores, and processing and reducing massive amounts of "raw" data into a summarized format.

4. Operational analytics and monitoring business operations. "Analyze a variety of machine and operational data for improved business results. The abundance and growth of machine data, which can include anything from IT machines to sensors and meters and GPS devices requires complex analysis and correlation across different types of data sets." Datameer (2014) elaborates "use customer and device usage across networks to identify high-value usage. Or you can integrate and analyze historic machine data and failure patterns to predict and improve mean time-to-failure – or ERP purchase data and supplier data to optimize supply chain operations."

5. Fraud detection, compliance, security/intelligence extensions. Use new data sources to detect fraud and monitor physical and cyber security in real time. This use case involves processing and analyzing data types like social media, emails, sensors, telephone call records, and audio and video feeds. Datameer provides five examples: "perform time series analysis, data profiling and accuracy calculations, data standardization, root cause analysis, breach detection, and fraud scoring. You can also run identity verifications, risk profiles, and data visualizations and perform master data management."

: What are some use cases with expanded data sources?

Vendor websites provide many examples of "big data" use cases. ParStream (<https://www.parstream.com/>) has both use cases and customer examples. The primary uses of ParStream real-time Internet of Things (IoT) data and analytics were for manufacturing, supply chain, telecommunications, retail/Point of Sales, and Web logging. All of the applications and uses impact operations decision making. The three common uses are monitoring, tracking, and diagnosis. ParStream is used for real-time sensor-based monitoring, diagnostics, and maintenance, monitoring and controlling industrial and manufacturing processes and buildings, supply chain optimization -aligning daily inventory plans with goals, predictive maintenance/diagnostics of equipment, remote diagnostics/condition-based monitoring of products/field assets, location tracking of goods/assets, real-time reporting of materials and products as they move through the supply chain, and real-time sensor-based monitoring, diagnostics, and maintenance.

The boundaries for these five decision support use categories are still not well defined and there remains some overlap. Researchers should work to create detailed documentation of decision support use cases involving high volume, high variety and high velocity data.

References

Datameer eBook, "Top Five High-Impact Use Cases for Big Data Analytics," 2014 at URL <http://www.datameer.com/pdf/eBook-Top-Five-High-Impact-UseCases-for-Big-Data-Analytics.pdf>

IBM, "The top five ways to get started with big data," Thought Leadership White Paper, Document Number: IMW14710USEN, June 2014.

Pentaho Press Release, "Pentaho helps retailers power big deals for consumers this holiday season," 11/25/2014, at URL <http://dssresources.com/news/4198.php>

Pinckney, T. Sr. Interview, "eBay Chooses Cassandra to Power Next Generation Recommendation Engine," Planet Cassandra, January 8, 2013 at URL <http://planetcassandra.org/blog/5-minute-c-interview-ebay/> . Also see URL <http://www.datastax.com/wp-content/uploads/2012/12/DataStax-CS-eBay.pdf>.

Turner, J., "Paddy Power Selects Apache Cassandra to Manage Time-Series Data for Online Gaming Application," Planet Cassandra, February 21, 2014 at URL <http://planetcassandra.org/blog/paddy-power-selects-apache-cassandra-to-manage-time-series-data>

: *What are some use cases with expanded data sources?*

-for-online-gaming-application/

Author: Daniel Power

Last update: 2014-12-21 02:02