Whitepaper

An Rx for Enterprise Big Data Success

NVP

NewVantage Partners

An Rx for Enterprise Big Data Success

The term big data has many meanings and continues to evolve quickly. Although big data technology was invented by online pioneers to manage their enormous data volumes and social-media traffic, traditional enterprises in financial services, retail and health care are finding it an incredibly powerful platform for their traditional IT processing.

Many firms are creating a big data Hadoop cluster, in which dozens or hundreds of low-cost servers run large data-processing jobs in parallel. This approach has proved time and again to be a low-cost, high-performance and very reliable platform for processing, and these clusters are increasingly running alongside mainframes in business data centers.

Realizing the value of big data today, however, demands more than technology. It requires the vision, experience and know-how necessary to see big data as a strategic opportunity and to leverage it across an enterprise. Today, there are many voices in this arena, but only a few who understand the real opportunity for big data to create shareholder value.

As pioneers in this field, New Vantage Partners (NVP) has created design patterns that enable rapid adoption of this new technology in enterprises in a non-intrusive manner. The first pattern we have identified is a mainframe and Hadoop coexistence model where we offload of mainframe batch jobs to a Hadoop ecosystem. A design pattern is a general repeatable solution to a commonly occurring problem, but it isn't a finished design that can be transformed directly into code. It is a description or template for how to solve a problem that can be used in many different situations.

Enterprise batch jobs often process large amounts of data for distribution (for example, to create extract files loaded into a data warehouse or another application) and reporting (for example, to summarize transactions). In this article we focus on migrating mainframe batch jobs to a big data platform. This migration no only reduces cost and increases performance, it frees the mainframes to do what they do best—support critical Tier 1 business operations.

Mainframes are the de facto standard for business-critical systems globally. According to industry statistics of the Fortune 500 companies, 490 leverage CICS to process more than 30 billion transactions worth about $1 trillion, each and every day. Mobile applications and devices have increased the number of transactions that CICS is required to process daily. So, it is even more important that crucial mainframe resources are saved for critical enterprise applications.

NVP
NewVantage Partners

Overview

A batch job executes a predefined group of programs (jobs) with little or no manual interaction. A scheduled batch process can involve execution of hundreds or thousands of jobs in a pre-established sequence. Batch processes typically have the following characteristics:

- Large amounts of input data are processed and stored (perhaps terabytes or more), large numbers of records are accessed and often a large volume of output is produced.

- Immediate response time is usually not a requirement. Batch jobs, however, often must complete within a batch window, a period of less-intensive online activity, as prescribed by a service-level agreement (SLA).

- Often all the records for an entire line of business are processed (for example, a summary of all customer orders or a retailer's stock on hand).

Comparing Mainframe and Big Data Platforms

The IBM mainframe z/OS operating platform has arguably the most highly refined and evolved set of batch-processing facilities owing to its origins, long history and continuing evolution.

Hadoop is a free Java-based programming framework that supports processing of large data sets in a distributed-computing environment. Its adoption is gaining momentum: according to NVP's Big Data Survey, many leading financial services firm are already using big data to create value, while others are in the midst of exploring how to adapt and stay competitive.

A core strength of Hadoop is batch processing, which is mature and a perfect fit to augment/replace mainframe batch jobs. But we should be aware of the hype around big data and objectively evaluate its application. Big data, mainframes and open systems all have a place—it's not about the tool, it's about the job.

The following diagram contrasts and compares Hadoop to the mainframe. Note specifically the difference in tools needed for "Abstraction/Programming" and "Batch Jobs": Hadoop offers a simple universal paradigm that better supports process standardization and rationalization.

Hadoop	Ecosystem Layers	Mainframe
Merced	Analytics	
Pentaho	Fast load analytics DB	
Avro	Meta data modeling	
Hive	Data Warehousing	
Zookeeper	Management/Security	SAF, RACF, JCL
Pig	Abstraction / Programming	Assembler, COBOL, PL/1, c/C++, Java, CLIST, REXX
	Transaction Processing	CICS
MapReduce	Batch Jobs	Assembler, COBOL, PL/1, c/C++, Java, CLIST, REXX
Sqoop, Flume	Data Movement	
Hbase	Databases/Data Types	DB2, IMS
Commodity Servers, Appliances, M/F servers	Servers	Mainframe servers
HDFS, MapR,	Distributed FS/ File Systems	MVS, VSAM
Linux, Windows, OS X, BSD	Operating Systems	z/Linux, z/OS, z/VM, z/VSE, z/TPF

Core Hadoop: Avro through HDFS, MapR

NVP
NewVantage Partners

The Rationale

Batch processing has been possible on distributed systems for years, but it has not been as commonplace as it is on mainframes because distributed systems often lack the following:

- Sufficient data storage
- Available processor capacity, or cycles
- Sysplex-wide management of system resources and job scheduling

Hadoop has addressed these gaps with a robust set of features for processing batch jobs in production. An obvious proof point is its use by eBay, Google, Twitter and Yahoo, who use the software exclusively to handle petabytes of online and offline data. Mainstream adoption is taking roots in the financial, retail, health-care and pharma sectors and is evident at giants like Costco, Sears and Walmart who are adopting this technology.

Crédit Mutuel Arkéa, a financial-services firm, solved its batch-job challenge by successfully moving mainframe batch jobs to Hadoop. Sears Metascale moved enterprise batch jobs from mainframe and open systems to Hadoop. This transition massively changed the number of ETL and other batch jobs. Results from batch jobs (including transforms)

that are required by mainframe applications move back to the mainframe for regular application consumption.

Studies have found that in similarly configured machines, batch jobs run 40–80 times faster and 200–400% cheaper on big data platforms. Hadoop can be truly useful in those environments, as it can fit nicely among COBOL, VSAM, MVS and other legacy technologies while firms still rely on legacy systems for daily operations.

The resulting benefits from a migration of batch processing to a big data platform include the following:

- Faster time to market (e.g. client on-boarding)
- Cost reduction
- Process rationalization
- Ameliorate bottlenecks
- Conserve mips (processing power)
- Steady-state operations

Prescription for Success

The diagram below captures a typical day in the life of batch jobs in a mainframe of a large financial-services company.

NVP
NewVantage Partners
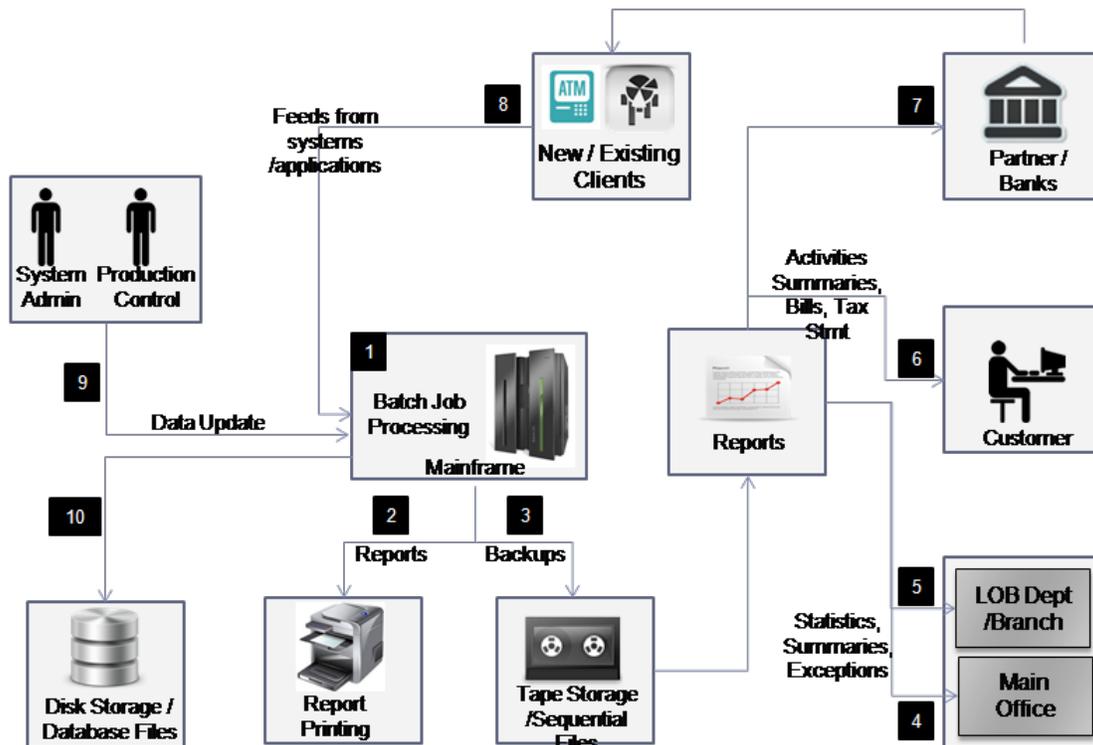
Supporting these batch jobs incurs huge, recurring costs for capital and operations. The table below identifies areas of opportunity for this coexistence model.

Step	Activity	Batch Job / Opportunities
1	At night, numerous batch jobs consolidate the results of the online transactions that take place during the day. Includes: accounting, accruals, profit/loss, performance and attributions, NAVs etc.	1. Data Cleansing 2. Transformations 3. Data Aggregation
2	Batch jobs generate business reports.	1. Complex joins and grouping of data 2. Many jobs for report generations
3	Backups of critical files and databases are made before and after the batch window	1. Reports from backups
4,5,6	Reports are sent to head offices, branch offices, clients/partners, customers	
7	Reports/Data are sent to the client/partner companies.	
8	Transaction data/reports are received from client/partner companies on a daily basis. They are processed and ingested by mainframe. A good example is a custodian bank.	1. Data Cleansing 2. Transformations 3. Data Aggregation 4. Data Extracts
9	In the production control department, the operations area is monitoring the messages on the system console and the execution of the jobs.	
10	Jobs and transactions are reading or updating the database (the same one that is used by online transactions) and many files are written to tape.	

Big data provides a compelling rationale for orchestrating a new ecosystem that enables large-scale enterprises to drive business value. By moving IT mainframe batch-processing activities to newer big data platforms, large organizations can improve their ability to dedicate resources to supporting and sustaining critical business needs. The figure below illustrates the design pattern.

NVP
NewVantage Partners
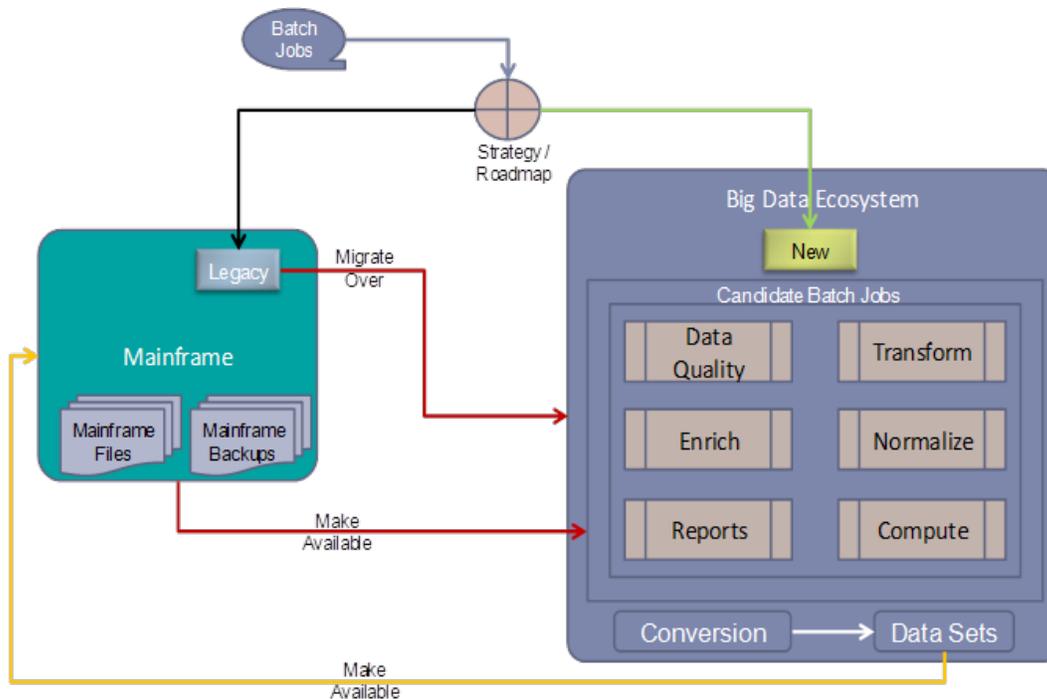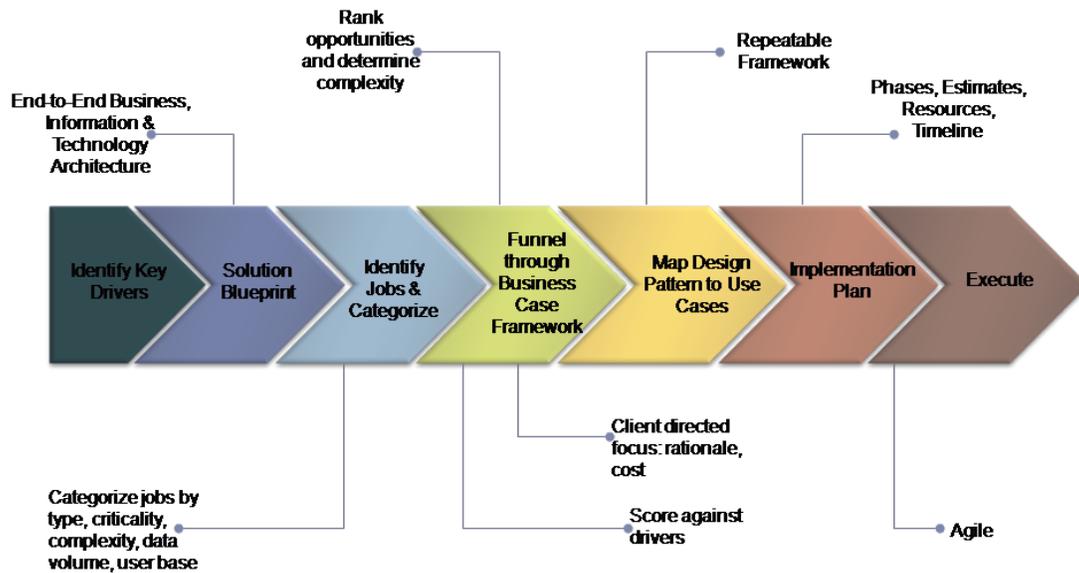
To begin realizing the benefits of big data initiatives, enterprises can take initial steps to create a big data ecosystem that encompasses these elements:

- **Business-case framework**. A repeatable framework to capture and process stakeholder interests, rationale, risks, KPI for success, costs and so on for migration/repurposing of jobs to a big data ecosystem.

- **Reference architecture.** A big data–centric reference architecture for Hadoop, applicable big data technologies, and mainframe and open systems for batch-job processing for the enterprise. Beyond the basic building blocks of big data, a more integrated information fabric is required to deliver the full potential and benefits of this approach.

- **Approach & methodology.** Technical and business processes, methodologies and SOPs for lifecycle management from migration and development through delivery.

- **Target operating model.** Target operating model with governance structures to support the organization, business processes and IT to work with this new strategy.

- **Talent and training.** Identify talents needed for the program. Work with HR to define and create new roles in the organization. Retrain existing personnel if possible.

- **Roadmap.** A clearly laid-out roadmap for the journey that guarantees repeated success.

Conclusion

We believe that for enterprises to be successful in their big data initiatives, firms must consider these realities:

- **Cultural change.** Big data can help you move your organization into the future, but it will also demand changes in the entity's culture with new technologies, target operating models and governance to address changes in how data is sourced, analysed and used, as well as how decisions are made, risks managed, results reported, and changes in technology and business processes implemented. Successful cultural change in support of big data will help make decisions better, faster and smarter.

- **Risks.** New technology brings with it new risks.

NVP
NewVantage Partners

Issues of data governance, ownership, security, privacy, disclosure, integrity, reliability, accuracy and regulatory compliance must also be addressed to realize the opportunity of big data. Mastering the ability to manage these risks will pay dividends in the form of enhanced TCO and ROI.

- **Partners.** Big data is a journey and not a destination. Organizations will need partner(s) to help define a vision, strategy, roadmap and implementation in line with their business and evolving nature of big data. Firms will need to work with partners who have and share the customer's perspective. A complete big data strategy requires partners to "walk in the customers' shoes" to understand their vision, objectives and imperatives.

About NewVantage Partners

NewVantage Partners is a boutique management consulting practice established in 2001 and comprised of former C-Level business and technology executives, and senior subject experts.

Our work comprises up-front planning – current state assessment, future state design, business case, execution roadmap, as well as the development of business and technical requirements, business capabilities, and business architecture. We are frequently engaged to provide a critical link between the business and technology organizations of our clients.

NewVantage fosters a commitment to executive thought-leadership through a series of small group executive dinners, and through our executive advisory board comprised of current and former Fortune 1000 business and technology executives and well-known industry thought leaders.

For more information, contact Randy Bean at rbean@newvantage.com.

NVP
NewVantage Partners

NVP

NewVantage Partners