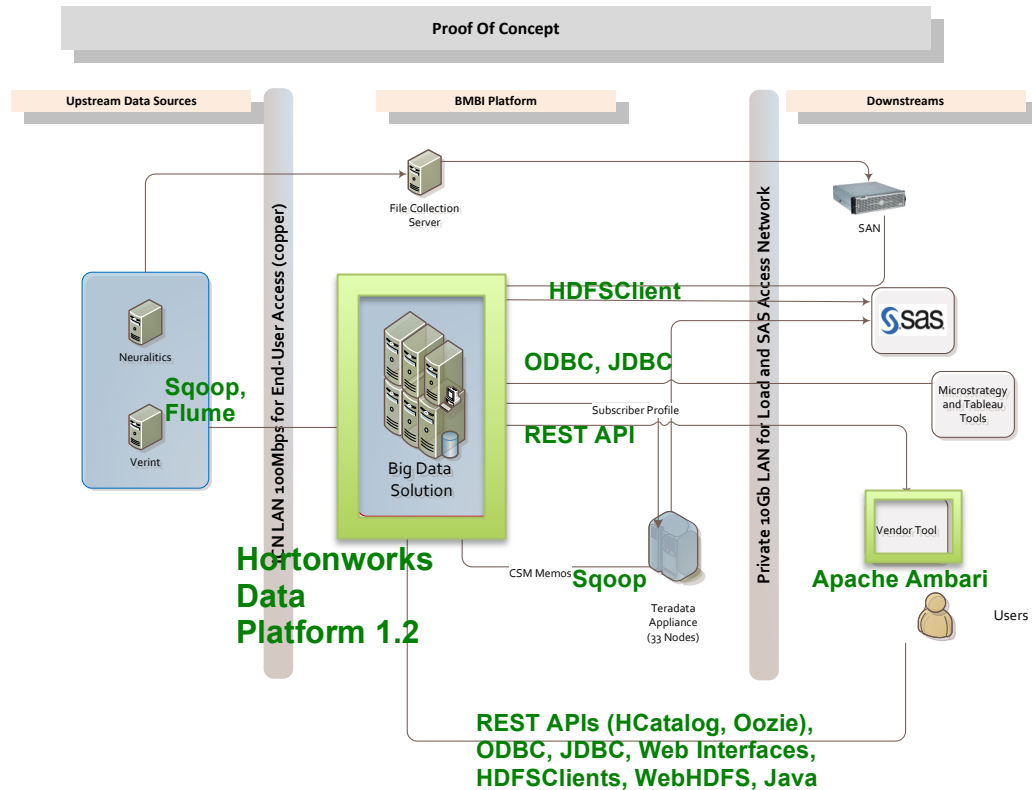


BELL BIG DATA RFI ADDENDUM: REVISED QUESTIONS - FEBRUARY 20TH 2013

3.2.1 - Provide a brief overview of your proposed **Hadoop Ecosystem** solution. Include both hardware and Software components (**pricing separately**). Describe the solution you will propose within the context of the Conceptual Architecture diagram in Section 1.6.4. Please update the diagram, replacing the generic reference to BIG Data solution with your solution architecture.

Reference Architecture from RFI Section 1.6.4 – Hortonworks Data Platform



The Hortonworks Data Platform (HDP) is the highest Enterprise-grade Hadoop distribution allowing organizations to capture, process and exchange big data. Based on our long history at *Yahoo!* (birthplace of Hadoop) designing, operating and managing the largest Hadoop clusters in the world, our platform aims to cater to three primary audiences within the Enterprise:

1. Operations:

- a. *Apache Ambari*: provides the Open Source and Modular platform to install, provision, configure, and monitor the Hadoop cluster. Ambari provides key service control, alerting, centralized configuration management, job analysis, and host monitoring.
- b. *Oozie*: the established workflow-scheduling engine for Apache Hadoop. Complex job packages can be scheduled and executed based on coordinated events or simple clock schedules. Oozie is integrated with the rest of the Hadoop stack to allow any type of job to be created.

2. Data Workers:

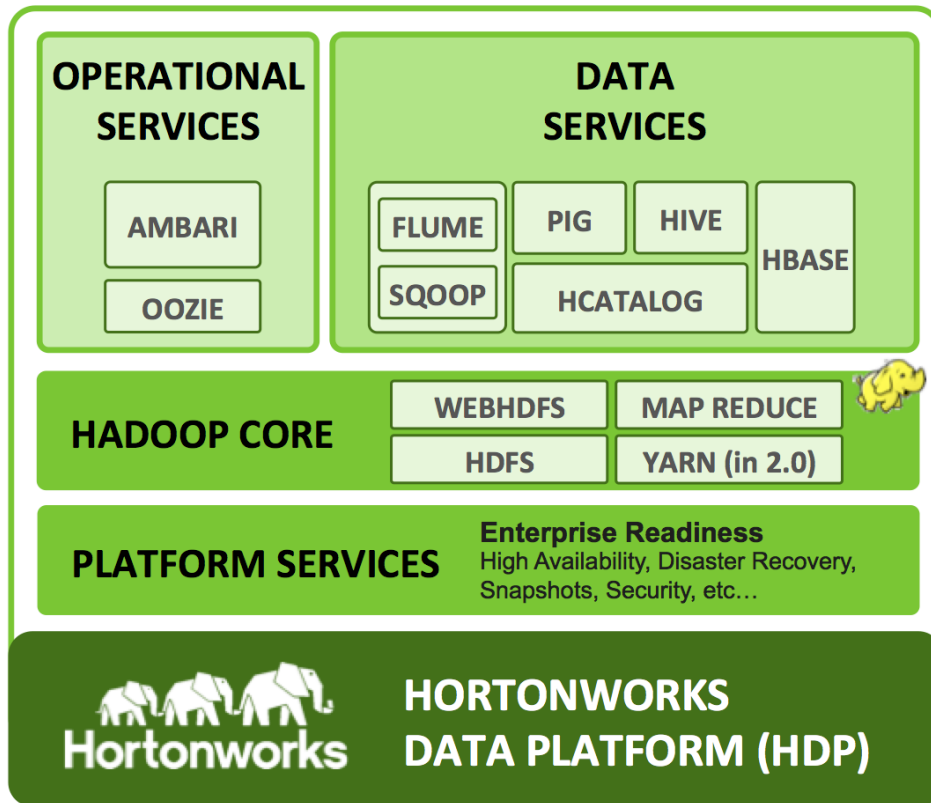
Every enterprise organization has different types of users that need to explore and analyze the data ranging from developers, data scientists, business and analysts and individuals who are experts of the data. The platform has tooling for each of these classes of users.

- a. *MapReduce*: A core component of Hadoop that allows for parallel and scalable data processing. MapReduce jobs are primarily developed in Java but can be scripted in other languages through “streaming” components.
- b. *Pig*: A high level scripting language that allows for MapReduce jobs to be created without any Java development or complicated Hadoop knowledge.
- c. *R, Apache Mahout, SAS (via connectors), and Excel (via connectors)*: Provide Data Access and Tooling for Data Scientists. Depending on the implementation, these technologies are: directly executed as MapReduce jobs, interact with Hadoop via Hive, or interact with Hadoop via the many clients.
- d. *Hive and HCatalog*: SQL access to data stored on Hadoop; ODBC/JDBC tooling for Business Analysts. HCatalog provides a centralized metadata layer for Hadoop technologies such as MapReduce, Pig, and Hive. HCatalog allows for a “schema-on-read” approach as required by Data Workers.
- e. *Talend*: Graphical Data Integration tools for a traditional ETL workbench as well as GUI-driven Hadoop job creation.

3. System/Data Integration:

Hadoop has various tooling to ingest data from varied sources, both structured and unstructured.

- a. *Sqoop*: A data integration tool that can ingest and export data from structured relational data stores. Primarily JDBC but has many database-specific accelerator modules such as TeraData, Oracle, MySQL, and others.
- b. *Flume*: A distributed agent-based architecture for ingesting and aggregating unstructured event-based data. Commonly used for log ingestion into HDFS but has a successful history with many event-driven ingestion use cases.
- c. *WebHDFS*: A REST API that allows flexible HDFS filesystem access without the need for complex clients.
- d. *HCatalog*: The metadata layer provides an extreme flexibility for system integration; various schemas can be maintained for disparate data sets within Hadoop. This allows an Enterprise to land and retain the data in its raw format rather than impose a schema upon it. Structure can always be added later in HCatalog.

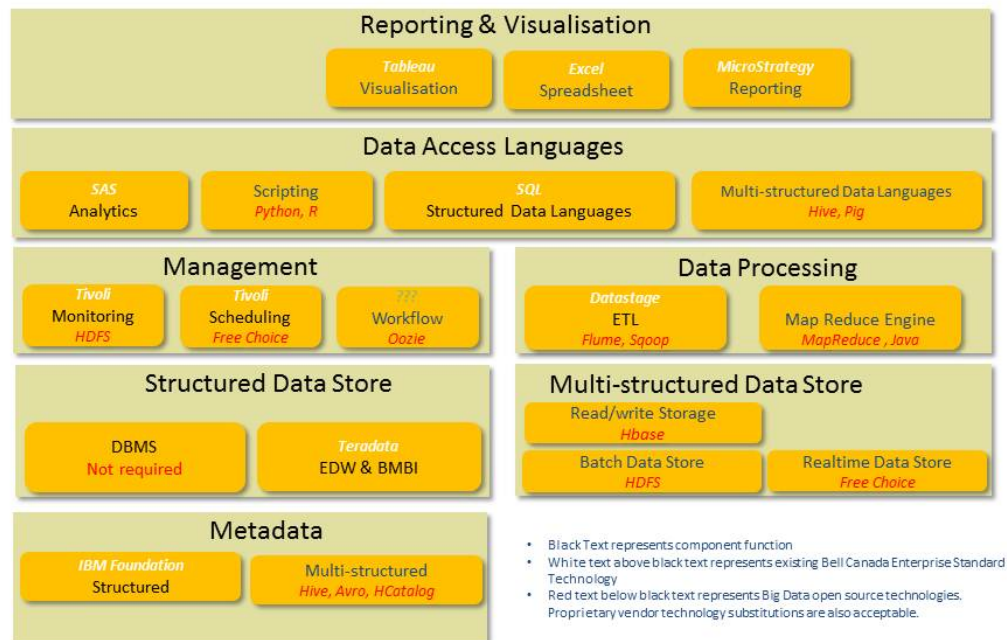


3.2.2 - Describe the technology recommended for each functional element of the solution. Include a solution architecture diagram as well as a textual description of each component. Descriptions should include the level of support provided for each solution component including third party software incorporated into your solution. **Include all necessary software components and details such as version numbers, etc.**

- Scheduling component
- Data Integration component
- Metadata component
- Data Storage component
- Data Access component

Please note that the diagram has been modified:

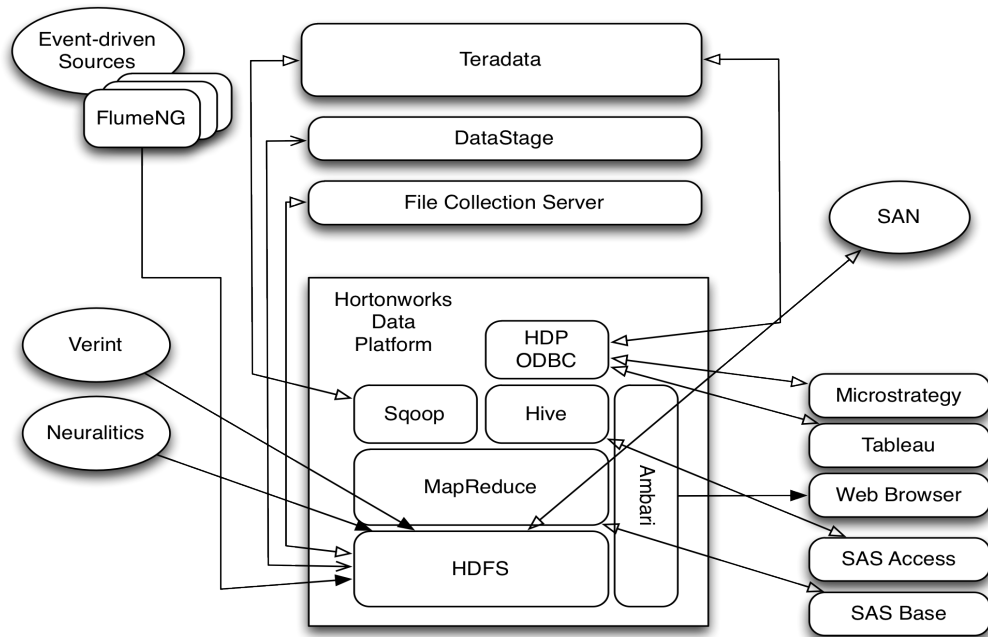
Existing and Future BI Components



Description by Functional Component:

Please see answer in 3.2.1 for component roles. Hortonworks Data Platform contains all of the projects referenced by Bell Canada in the revised diagram.

Solution Architecture:



Component Version Summary:

Hadoop Component	Functional Component	HDP 1.2 Version
Hadoop (MapReduce & HDFS)	Data Processing & Storage	1.1.2
HCatalog	Metadata	0.5.0
Pig	Data Access & Processing	0.10.1
Hive	Data Access	0.10.0
HBase	Read/Write & Real-time Storage	0.94.2
Sqoop	ETL / Data Processing	1.4.2

Oozie	Workflow & Scheduling	3.2.0
Zookeeper	Support for HBase & Distributed Configuration	3.4.5
Ambari	Cluster Management, Installation, & Alerting	1.3.0
Talend	Job Creation & ETL Workbench	5.2
Flume	ETL / Data Processing	1.3.0
Mahout	Data Processing & Advanced Analytics	0.7.0

Clarification on Pricing

In preparing your response we request that you provide a detailed breakdown of the costs as per the following table below. Please provide as cost effective a proposal as possible that is in alignment with the revised requirement (**Hadoop/MapReduce only based solution**).

- Please be as detailed as possible in providing a Bill of Material break down of each TCO factor. Note: Bell may proceed with some and not all of the components within each TCO factor.
- **Vendor MUST provide the cost per Hadoop Node as well as the cost per Terabyte in their configuration pricing. This cost should include both hardware and software costs.**

Hortonworks & Scalar Decisions:

Hortonworks has partnered with Scalar Decisions to provide a complete Hardware and Software Solution to Bell Canada. Scalar Decisions designs, deploys and manages mission critical IT infrastructure focused around 4 key solution domains - Virtualization & Cloud, Data Management, Network, and Security.

About Scalar Decisions:

Scalar Decisions understands the integrated Hadoop infrastructure stack and how to make it work in large-scale projects. Scalar has been an active participant in the Toronto Hadoop User Group (THUG <http://www.meetup.com/TorontoHUG/sponsors/>) hosting, providing free access to Scalar's hosted Hadoop Cluster, donating technical resources, providing technical expertise, and actively working to develop and promote Hadoop skills in the Toronto technical community. Scalar's world-class technical architects deliver infrastructure solutions that allow your organization to do more with less. Scalar simplifies the challenges, lowers the costs, and improves the performance of your IT infrastructure.

Requested TCO Breakdown:

Note: The pricing here assumes the recommended Dell PoC Cluster detailed in the next section. All quotes for Training and Professional Services are estimates. Please contact us for a more detailed consultation.

TCO Factor	Description (end-state solution)	Estimated Cost
<i>Initial Hortonworks – 2013 Investment</i>		
Hardware	Dell 18-Node Storage-Dense Cluster	\$160,050.00
Software	There is no software licensing cost for any component of the Hortonworks Data Platform, it is freely distributed open source software. Price included assumes the premium 24/7 Enterprise Support option and has an effective coverage of 1 Year.	\$40,000.00
Training	This assumes onsite training for Developer and Administration Classes for a group of 10 or less.	\$20,000.00
Professional Services	Services can be consumed as required. Please contact us for a detailed consultation and scheduling. We recommend a minimum initial purchase of Service Credits to assist with the initial implementation in the first year.	\$50,000.00
<i>On-going Hortonworks – 2014 and 2015 Investment</i>		
Software support subscription	Assuming data and cluster growth greater than the existing cluster but within a Petabyte of usable HDFS space. This covers the support renewal for one year. Various Support options are included in the original RFI.	\$130,000.00
Hardware/Software Licensing Fees	None	\$0
Maintenance Fees (monitoring,	Upgrades and patches are	\$0

scheduled maintenance, patches)	freely available.	
Training	Additional Training assumed for Data Sciences and new use cases.	\$20,000.00
Professional Services	Required dependent on use cases.	\$50,000.00
Capacity Upgrades (HW/SW)	Assuming that the cluster doubles in size (ie 15 to 30 worker nodes).	\$120,000.00

Detailed Hardware Pricing:

The following hardware prices detail small and medium cluster sizes from two manufacturers for easy comparison. Bell Canada can customize any of these configurations to suit their needs but Hortonworks and Scalar recommends starting with the Dell 18-Node Cluster for the PoC for maximum \$/TB value. Please find detailed hardware quotes for each configuration attached to the response.

Hardware Sizing and Architecture based HP and Dell Reference Architectures for Hadoop as well as Hortonworks expertise in sizing. Please find included the Hortonworks Cluster Planning Guide for further sizing rationale and guidance.

Dell Reference Architecture:

Dell 18-Node, 540TB RAW OPTION

Description	QTY	Price	Extended Cost
Head Node	3	\$7,242.41	\$21,727.23
Worker Node	15	\$7,935.86	\$119,037.90
Data Switch	2	\$3,969.50	\$7,939.00
Management Switch	1	\$11,346.88	<u>\$11,346.88</u>
			\$160,051.01

\$ per TB: **\$889.17**

Dell 33-Node, 1050TB RAW OPTION

Description	QTY	Price	Extended Cost
Head Node	3	\$7,242.41	\$21,727.23
Worker Node	30	\$7,935.86	\$238,075.80
Data Switch	2	\$3,969.50	\$7,939.00
Management Switch	1	\$11,346.88	<u>\$11,346.88</u>
			\$279,088.91

\$ per TB: \$797.40

HP Reference Architecture:

HP 18-Node, 350TB RAW OPTION

Description	QTY	Price	Extended Cost
Head Node	3	\$11,504.40	\$34,513.20
Worker Node	15	\$11,348.62	\$170,229.30
Data Switch	2	\$13,001.84	\$26,003.68
Management Switch	1	\$13,001.84	<u>\$13,001.84</u>
			\$243,748.02
\$ per TB:	\$2,089.27		

HP 48-Node, 1050TB RAW OPTION

Description	QTY	Price	Extended Cost
Head Node	3	\$11,504.40	\$34,513.20
Worker Node	45	\$11,348.62	\$510,687.90
Data Switch	2	\$14,853.94	\$29,707.88
Management Switch	1	\$14,853.94	<u>\$14,853.94</u>
			\$589,762.92
\$ per TB:	\$1,685.04		