Overview
Building an Architecture for Big Data & Analytics problems
Purpose

- This presentation, derived from the IBM Big Data & Analytics Reference Architecture document is meant to drive consistency across targeted technical sales channels in the roll-out of the IM & BA reference architectures aligned to the 5 big data use cases and the industry scenarios.

- The purpose of IBM Big Data & Analytics (BD&A) Reference Architecture is to inform and guide sales, technical sales, services and other professionals who are involved in selling IBM solutions or deploying them with clients.

- The Reference Architecture is intended to be used by a wide range of professionals selling IBM software and designing end-to-end big data & analytics client solutions. The reference architecture can also be reused by IBM clients and partners to gain further insight on how to benefit from IBM Software in big data & analytics projects (appropriate disclosure is required).

- The target audience also depends on the specific work product included in the reference architecture. The Architecture Overview, Business Directions, Requirements, Technology Brief are expected to be consumed by business leaders and architects.
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle.

Understand Client

Define Client Requirements

Design Solution

Detail Design to Define BOM

Best Practices DevOps

Business Drivers

Describe the key business drivers for the project, the KPIs or CSFs, and how they align with Cloud computing.

Use Cases

What are the functional requirements expected from the Cloud and who are the key actors. Expressed as Use Cases.

System Context

The system context should define the boundary of the Cloud, and the integrations with OSS / BSS systems.

Architecture Decisions

Clearly documented decisions on key architectural points including the rationale for the decision.

Functional & Non-Functional Requirements

NFRs should be defined to cover the volumes, capacity, scale, availability, security, operational and monitoring aspects of the Cloud.

Architecture Overview

Architecture overview diagram should define the high level components, their placement.

Operational Model

Design and consider the components of the solution both at a physical and logical level.

Technology Brief

Define the boundaries of the project, inclusions, exclusions, dependencies, and align phases with milestones in the roadmap.

Best Practices

Define the overall timeline, phases, and key milestones that will shape the plan and overall delivery.

DevOps

Design and consider the components of the solution both at a physical and logical level.

Content provided by BDA RA
Business Drivers
Use Cases
Requirements, Security
Architecture Overview,
Architecture Decisions
Operational Model,
DevOps
Best Practices

© 2014 IBM Corporation
The IBM Business Analytics and Optimization: Enterprise Architecture View

<table>
<thead>
<tr>
<th>Access</th>
<th>Advanced Analytics</th>
<th>BIBPM</th>
<th>Data Repositories</th>
<th>Data Integration</th>
<th>Master Data Management</th>
<th>Content Management</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEB BROWSER</td>
<td>PREDICTIVE ANALYTICS</td>
<td>PLANNING, FORECASTING &amp; BUDGETING</td>
<td>DIMENSIONAL LAYER</td>
<td>EXTRACT / SUBSCRIBE</td>
<td>SOA ORCHESTRATION COMPONENTS</td>
<td>CONTENT INGESTION</td>
<td>ENTERPRISE</td>
</tr>
<tr>
<td>PORTAL</td>
<td>DATA MINING</td>
<td>DATA WAREHOUSES</td>
<td>OPERATIONAL DATA STORES</td>
<td>INITIAL STAGING</td>
<td>CRUD TRANSACTIONAL COMPONENTS</td>
<td>CONTENT EXTRACTION</td>
<td>UNSTRUCTURED</td>
</tr>
<tr>
<td>DEVICES</td>
<td>SIMULATION</td>
<td>TIME PERSISTENT REPOSITORY</td>
<td>MASTER DATA</td>
<td>DATA QUALITY</td>
<td>MASTER DATA STORE</td>
<td>CONTENT SERVICES</td>
<td>INFORMATIONAL</td>
</tr>
<tr>
<td>WEB SERVICES</td>
<td>TEXT ANALYTICS</td>
<td>UNSTRUCTURED (BIG DATA / CONTENT)</td>
<td>TRANSFORMS</td>
<td>CLEAN STAGING</td>
<td>DATA LOAD COMPONENTS</td>
<td>DOCUMENT MANAGEMENT SERVICES</td>
<td>EXTERNAL</td>
</tr>
<tr>
<td></td>
<td>VISUALIZATION</td>
<td>STAGING AREAS</td>
<td>LOAD READY PUBLISH</td>
<td>LOAD / PUBLISH</td>
<td>REFERENCE DATA MANAGEMENT</td>
<td>RECORDS MANAGEMENT SERVICES</td>
<td>WEB</td>
</tr>
<tr>
<td></td>
<td>OPTIMIZATION AND RULES MANAGEMENT</td>
<td>QUERY</td>
<td>SANDBOX</td>
<td>BUSINESS PROCESS MANAGEMENT / WORKFLOW</td>
<td>BUSINESS PROCESS MANAGEMENT / WORKFLOW</td>
<td>BUSINESS PROCESS MANAGEMENT / WORKFLOW</td>
<td>BUSINESS PROCESS MANAGEMENT / WORKFLOW</td>
</tr>
</tbody>
</table>

INFORMATION GOVERNANCE
POLICY / ORGANIZATION / CHANGE MANAGEMENT
DATA ARCHITECTURE
DATA QUALITY
METADATA

NETWORK CONNECTIVITY, PROTOCOLS, AND ACCESS MIDDLEWARE
HARDWARE AND SOFTWARE PLATFORMS

© 2014 IBM Corporation
The IBM Big Data & Analytics Reference Architecture: High Level Capabilities

The IBM Business Analytics and Optimization: Detailed Capability View
The IBM Business Analytics and Optimization: Logical View

The IBM Big Data & Analytics Reference Architecture - Logical Overview

The IBM Big Data & Analytics Reference Architecture: Software Product Mapping View
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle.

- **Understand Client**: Describe the key business drivers for the project, the KPIs or CSFs, and how they align with Cloud computing.
- **Define Client Requirements**: What are the functional requirements expected from the Cloud and who are the key actors. Expressed as Use Cases.
- **System Context**: The system context should define the boundary of the Cloud, and the integrations with OSS / BSS systems.
- **DevOps Design and consider the components of the solution both at a physical and logical level.**
- **Architecture Decisions**: Clearly documented decisions on key architectural points including the rationale for the decision.

---

**Content provided by BDA RA**
- Business Drivers
- Use Cases
- Requirements, Security
- Architecture Overview
- Architecture Decisions
- Operational Model
- DevOps
- Best Practices

© 2014 IBM Corporation
Information and data pervades every aspect of the organization from cost containment to revenue growth to customer interaction and marketing to regulatory reporting to risk reporting and metrics of profitability as well as your organization liquidity constraints. Therefore information management architecture and managing data is a multi-disciplinary task.
Business Drivers: Big Data introduces new opportunities and concepts

- **Data Scale:** Up to exabytes. Up to 10,000 times larger than traditional data warehouses

- **Decision Frequency:** Up to real time using streaming data patterns

- **Data at Rest:** Data is persisted in a physical medium and is considered relatively static

- **Data in Motion:** Data flowing constantly through a network or other data transport mechanism. Data persisted in memory on a temporary basis is also considered to be “in motion”

Being able to manage and leverage these new dimensions creates a unique opportunity to “change the game” for our clients
Business Driver: Why Big Data & Analytics are Important?

Data is emerging as the world’s newest resource for competitive advantage

Uncover new insights to transform the business
- Why did it happen?
- What is likely to happen?
- What’s the best course of action based on what you’ve learned?

Empower many more employees within the organization
- At every level of the organization to make better decisions

Improve effectiveness & competitiveness of the business
- Make speed a differentiator
- Monetize the data itself
- Be more right, more often

Manage risk
- Protect against poor decision-making (risk-opportunity equation right)
- Protect against security and privacy risks

New and more effective approach to perform analytics
- Move analytics to the data, (in-motion & native format)
- Leverage open source and commodity hardware (cost savings)

Hadoop, Streaming, Cognitive, In-Memory, Exploration

The power of all data coming together...

...with the power of New Technology

Real-Time

Volume

Velocity

Variety

Veracity

structured & unstructured data (sensor, social, logs, machine, etc)
Use Cases: Every Industry can Leverage Big Data and Analytics

**Banking**
- Optimizing Offers and Cross-sell
- Customer Service and Call Center Efficiency

**Insurance**
- 360˚ View of Domain or Subject
- Catastrophe Modeling
- Fraud & Abuse

**Telco**
- Pro-active Call Center
- Network Analytics
- Location Based Services

**Energy & Utilities**
- Smart Meter Analytics
- Distribution Load Forecasting/Scheduling
- Condition Based Maintenance

**Media & Entertain**
- Business process transformation
- Audience & Marketing Optimization

**Retail**
- Actionable Customer Insight
- Merchandise Optimization
- Dynamic Pricing

**Travel & Transport**
- Customer Analytics & Loyalty Marketing
- Predictive Maintenance Analytics

**Consumer Products**
- Shelf Availability
- Promotional Spend Optimization
- Merchandising Compliance

**Govern.**
- Civilian Services
- Defense & Intelligence
- Tax & Treasury Services

**Healthcare**
- Measure & Act on Population Health Outcomes
- Engage Consumers in their Healthcare

**Automotive**
- Advanced Condition Monitoring
- Data Warehouse Optimization

**Chemical & Petroleum**
- Operational Surveillance, Analysis & Optimization
- Data Warehouse Consolidation, Integration & Augmentation

**Aerospace & Defense**
- Uniform Information Access Platform
- Data Warehouse Optimization

**Electronics**
- Customer/Channel Analytics
- Advanced Condition Monitoring

**Life Sciences**
- Increase visibility into drug safety and effectiveness
Use Cases: The 5 Key Use Cases for Big Data

**Big Data Exploration**
Find, visualize, understand all big data to improve decision making

**Enhanced 360° View of the Customer**
Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information sources

**Security/Intelligence Extension**
Lower risk, detect fraud and monitor cyber security in real-time

**Operations Analysis**
Analyze a variety of machine data for improved business results

**Data Warehouse Augmentation**
Integrate big data and data warehouse capabilities to increase operational efficiency
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle.

Understand Client
- Business Drivers
  - Describe the key business drivers for the project, the KPIs or CSFs, and how they align with Cloud computing.

Define Client Requirements
- Use Cases
  - What are the functional requirements expected from the Cloud and who are the key actors. Expressed as Use Cases.

Design Solution
- System Context
  - The system context should define the boundary of the Cloud, and the integrations with OSS / BSS systems.

Detail Design to Define BOM
- Architecture Decisions
  - Clearly documented decisions on key architectural points including the rationale for the decision.

Best Practices DevOps
- DevOps
  - Define the overall timeline, phases, and key milestones that will shape the plan and overall delivery.

Content provided by BDA RA
- Business Drivers
- Use Cases
- Requirements
- Security
- Architecture Overview
- Architecture Decisions
- Operational Model
- DevOps
- Best Practices
Functional Requirements

- **PROJECT DEFINITION**
- **USE CASES**

identify

**FUNCTIONAL REQUIREMENTS**
Functional Requirements: Purpose

- Capture common functional requirements
- Provide consistent, reusable, and proven approach
- Can be adapted for customer deliverables
  - Readiness Assessments
  - Customer Proposals
How to make use of the functional requirements

Data Acquisition
Real-Time Processing & Analytics
Data Integration
Analytics Repositories
Shared Operational Information
Information Access
Information Interaction
Governance
Security & Business Continuity
Infrastructure
Non Functional Requirement: Security

### Table of Contents

1. Purpose
   1.1 Scope
   1.2 Intended Audience
   1.3 Overview

2. Security (Details)
   2.1 Requirements
   2.2 Use Cases
   2.3 End-to-End Integration
Security

Goals:
• Prevent, detect and address system breaches
  Mitigate internal and external threats
• Ensure integrity & privacy of sensitive data
• Ensure the availability of data
• Reduce cost of compliance

Protect Data at Rest & In Motion
• InfoSphere Guardium
  • Data Activity Monitoring
  • Vulnerability Assessments
• InfoSphere Optium Data Masking
• InfoSphere Guardium Encryption Expert
• Key Lifecycle Manager

Access Controls
• Access Manager family
• Federated Identity Manager
• Identity Manager/Role Lifecycle Manager
• Fine Grained entitlements
  • MDM and GNR Security Policy Manager
• Privileged Identity Manager
Security

Real-time Ingest & Processing
- InfoSphere Streams
  - Video/audio
  - Network
  - Geospatial
  - Predictive

Big Data Storage & Analytics
- InfoSphere BigInsights
  - Text and entity analytics
  - Data mining
  - Machine learning

Data Warehouse
- Deep analytics
- Operational analytics
- Large scale structured data management

I2 Analyst’s Notebook

Connectors
- Network Telemetry Monitoring Appliance (Optional)

Structured Data

Connectors
- Unstructured & Streaming Data

Non-Functional requirements

Big Data & Analytics
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle.

**Understand Client**
- Business Drivers
  - Describe the key business drivers for the project, the KPIs or CSFs, and how they align with Cloud computing.

**Define Client Requirements**
- Use Cases
  - What are the functional requirements expected from the Cloud and who are the key actors. Expressed as Use Cases.

**Design Solution**
- System Context
  - The system context should define the boundary of the Cloud, and the integrations with OSS / BSS systems.

**Detail Design to Define BOM**
- Architecture Decisions
  - Clearly documented decisions on key architectural points including the rationale for the decision.

**Best Practices DevOps**
- Define the overall timeline, phases, and key milestones that will shape the plan and overall delivery.

**Operational Model**
- Design and consider the components of the solution both at a physical and logical level.

**Technology Brief**
- Define the boundaries of the project, inclusions, exclusions, dependencies, and align phases with milestones in the roadmap.

**Content provided by BDA RA**
- Business Drivers
- Use Cases
- Requirements, Security
- Architecture Overview
- Architecture Decisions
- Operational Model
- DevOps
- Best Practices
Architecture Overview: Data and Analytics

- **Analytic**
  - Predictive and descriptive information for the "masses", actionable, and embedded

- **Dimensional**
  - Cross Value Chain Information, Advanced Analytics, and Integrated Performance Management

- **Data Warehouse**
  - Enterprise, subject area focused, information access
  - Integrated and consistent

- **Transactional Data**
  - Transaction data
  - Granular and operational

- **Master Data**
  - Context for Transactional and Analytic data
  - Commonly defined, core data
IBM Big Data and Analytics Reference Architecture - Logical Overview

Data Sources
- New Data Sources:
  - Machine & Sensor Data
  - Image & Video
  - Enterprise Content Data
  - Social Data
  - Internet Data

- Traditional Data Sources:
  - Third-Party Data
  - Transactional Data
  - Application Data

Data Acquisition & Application Access

Data Integration
- Data Quality, Transformation & Load
- Landing Exploration & Archive
- Big Data Repository

Stream Processing
- Real-Time Analytical Processing

Analytical Sources
- Integrated Data Warehouse
- Enterprise Warehouse
- Deep Analytics & Modeling
- Analytical Appliances
- Interactive Analysis & Reporting
- Data Marts

Actionable Insight
- Decision Management
- Discovery & Exploration
- Modeling & Predictive Analytics
- Analysis & Reporting
- Planning & Forecasting
- Content Analytics

Enhanced Applications
- Customer Experience
- New Business Model
- Financial Performance
- Risk
- Operations & Fraud
- IT Economics

Shared Operational Information
- Master & Reference
- Content Hub
- Activity Hub
- Metadata Catalog

Governance

Event Detection and Action

Security & Business Continuity Management

Platforms
IBM Big Data and Analytics Reference Architecture – Detailed Capabilities

**Data Sources**
- New Data Sources
  - Machine & Sensor Data
  - Image & Video
  - Enterprise Content Data
  - Social Data
  - Internet Data
- Traditional Data Sources
  - Third-Party Data
  - Transactional Data
  - Application Data

**Data Acquisition** (Extract, Replicate, Copy) & Application Access

**Data Integration**
- Data Ingestion
  - Extract & Subscribe
  - Initial Stage
  - Data Quality
  - Clean Staging
  - Transformation
  - Load Ready
  - Load

**Streaming Computing**
- Real-Time Analytical Processing
  - Data Integration
  - Predictive Analytics
  - Real-Time Insights

**Analytical Sources**
- Hadoop & Exploration
  - Landing
  - Exploration
  - Archive
  - Indexing
- Data Warehousing
  - Data Warehouse
  - Data Marts
  - ODS
  - Time Persistent
  - Sandbox
- Analytical Appliances
  - Accelerators (In-Memory)

**Information Access**
- Delivery & Visualization
  - Data Services
  - Data Publish
- Semantic Layer
  - Data Access
  - Data Caching
  - Data Delivery
  - Data Virtualization
- Data Federation
  - Data Link

**Actionable Insight**
- Decision Management
  - Rules Management
  - Real Time Decision Management
- Reporting, Analysis & Content Analytics
  - Planning, Forecasting
  - Budgeting
  - Query & Analysis
  - Scorecards, Dashboards
  - Storytelling
  - Collaboration
  - Alerting, Monitoring
- Discovery & Exploration
  - Annotation
  - Search
  - Predictive Analytics & Modeling
  - Predictive Analytics
  - Text Analytics
  - Simulation
  - Optimization
  - Data Mining
  - Correlations

**Governance**
- Corporate Governance
  - Information Governance

**Event Detection and Action**
- Private Clouds
  - Public Clouds
  - Appliances
  - Custom HW Solutions

**Security & Business Continuity Management**
Big Data & Analytics Component Model
Real-Time Analytical Processing Components

Application Development Framework
Administration and Monitoring Services
Data Delivery Services

Enterprise Data Sources
- Machine & Sensor Data
- Image & Video
- Log Data
- Data Sensor & Data Capture Services
- Enterprise Content Data
- Internet Data
- Transactional Data

Real-Time Analytical Processing
- Data Integration Services
- Predictive Analytics Services
- Decision Management Services

Data Streaming Pipeline

Data Delivery Services
- Big Data Repository (Hadoop)
- Data Warehouse & Marts

Application Integration Services

Operational Applications

Administra& Monitoring Services
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle.

**Understand Client**
- Business Drivers
  - Describe the key business drivers for the project, the KPIs or CSFs, and how they align with Cloud computing.

**Define Client Requirements**
- Use Cases
  - What are the functional requirements expected from the Cloud and who are the key actors. Expressed as Use Cases.

**Design Solution**
- System Context
  - The system context should define the boundary of the Cloud, and the integrations with OSS / BSS systems.

**Detail Design to Define BOM**
- Architecture Decisions
  - Clearly documented decisions on key architectural points including the rationale for the decision.

**Best Practices DevOps**
- Define the overall timeline, phases, and key milestones that will shape the plan and overall delivery.

**DevOps**
- Design and consider the components of the solution both at a physical and logical level.

**Technology Brief**
- Define the boundaries of the project, inclusions, exclusions, dependencies, and align phases with milestones in the roadmap.
1. Purpose
The purpose of the Technology Brief is to inform and guide IBM sales, services and professionals who are involved in designing and deploying IBM Big Data & Analytics solutions. The technology brief provides a synopsis of the IBM products used to deploy Big Data & Analytics solutions.

2. Overview
Using the reference architecture, this document will describe each of the IBM software products that support a Big Data & Analytics solution by capability. This document will describe each of the products listed in the following architecture overview diagram.

3. Real-time processing & analytics
3.1 IBM InfoSphere Streams

4. Data Integration
4.1 InfoSphere Federation Server

4.2 InfoSphere Information Server

IBM InfoSphere Information Server Family enables organizations to derive more value from their complex, heterogeneous data sources spread across their environment. It allows organizations to integrate disparate data and deliver trusted information whenever and wherever needed, in line and in context, to specific people, applications, and processes. It helps business and IT personnel to collaborate to understand the meaning, structure, and content of any type of information across any source. It provides breakthrough productivity and performance for cleaning, transforming, and delivering this information consistently and securely throughout the enterprise, so it can be accessed and used in new ways to drive innovation, increase operational efficiency, and lower risk. It can also help any integration logic created
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle

**Understand Client**
- **Business Drivers**
  - Describe the key business drivers for the project, the KPIs or CSFs, and how they align with Cloud computing.

**Define Client Requirements**
- **Use Cases**
  - What are the functional requirements expected from the Cloud and who are the key actors. Expressed as Use Cases.

**Design Solution**
- **System Context**
  - The system context should define the boundary of the Cloud, and the integrations with OSS / BSS systems.
- **Architectural Overview**
  - Architecture overview diagram should define the high level components, their placement.

**Detail Design to Define BOM**
- **Architecture Decisions**
  - Clearly documented decisions on key architectural points including the rationale for the decision.

**Best Practices DevOps**
- **Operational Model**
  - Design and consider the components of the solution both at a physical and logical level.

**DevOps**
- **Technology Brief**
  - Define the boundaries of the project, inclusions, exclusions, dependencies, and align phases with milestones in the roadmap.

**Best Practices**
- **Functional & Non-Functional Requirements**
  - NFRs should be defined to cover the volumes, capacity, scale, availability, security, operational and monitoring aspects of the Cloud.

**Use Cases**
- **Non-Functional Requirements**
  - What are the non-functional requirements expected from the Cloud.

**Content provided by BDA RA**
- Business Drivers
- Use Cases
- Requirements
- Architecture Overview,
- Architecture Decisions
- Operational Model,
- Security guidance
- DevOps
- Best Practices

© 2014 IBM Corporation
Big Data Analytics – key Architecture Decision areas

Key Decision Areas

- Performance
- Scalability
- Financial decisions
- Reliability
- Backup and Restore
- Sensitive data
- Disaster Recovery (DR)
ADC001 – Storing large volumes of data for analytics

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Data Exploration</th>
<th>ID</th>
<th>AD-C001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue or Problem</td>
<td>What is a cost effective solutions to store large volumes of structured and unstructured data?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assumptions</td>
<td>The data needs to be online for users to perform exploratory analytics on any data types</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternatives</td>
<td>Traditional data warehousing systems (RDBMS), Data Warehousing Appliances, Hadoop repository, Traditional file systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision</td>
<td>Hadoop based repository such as IBM PureData Systems for Hadoop or IBM BigInsights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Justification</td>
<td>A hadoop system is a very cost effective platform to store any data types. Because there is no enforcement of schemas, the data can just be loaded into a file system for future access and exploration. Hadoop enables the distributed processing of large data sets across clusters of commodity servers. It is designed to scale up from a single server to thousands of machines, with a very high degree of fault tolerance.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implications</td>
<td>Possible performance implications if expectations are not clear defined.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Derived requirements</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Related Decisions     | Consider a hadoop distribution and exploration tools to enable easy access and exploration of the data.
## ADC002 – High Performance Data Warehousing & Analytics

### 4.2 Big Data – High Performance Data Warehousing and Analytics on Large data Volumes

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>High Performance Data Warehousing and Analytics</th>
<th>ID</th>
<th>AD-C002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue or Problem Statement</td>
<td>What options are available to build a high performance data warehouse and analytics environment?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assumptions</td>
<td>Big Data Analytics architecture, structured data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternatives</td>
<td>Traditional data warehousing solutions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analytical Appliances based on Relational Databases</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hadoop Database</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision</td>
<td>A Traditional data warehousing solution based on relational databases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Justification</td>
<td>The Data Warehousing and analytics appliance provides such as IBM PureData systems for Analytics (aka Netezza) provides extremely fast performance for analytics. Appliances are purpose-built appliance for high speed analytics and leverage Massively Parallel processing (MPP). These solutions also provide powerful integration with data mining tools (in-database data mining) to enable organizations to optimize modeling and scoring of advanced analytical solutions.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implications</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Derived requirements</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Related Decisions</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## AD-C006: Data Integration of structured & unstructured content

### 4.6 Big Data – Data Integration of large volumes of Structured and Unstructured Data

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Data Integration</th>
<th>ID</th>
<th>AD-C006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue or Problem Statement</td>
<td>What are the alternatives to integrate a variety of structured and unstructured data?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assumptions</td>
<td>Big Data Analytics architecture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternatives</td>
<td>Data Integration solution using MPP Architecture or SMP Architecture</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Develop code manually using programming languages such as Java, C++ or SQL Language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision</td>
<td>Data Integration tool (aka ETL Tool) MPP Architecture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Justification</td>
<td>Data integration tools (aka ETL tools) provide data standardization, cleansing and transformations across many different sources. They also provide metadata lineage to enable developers to quickly perform impact analysis and users can leverage metadata to identify the origin and definition for data elements. MPP architecture enables linear scalability of the solution.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The IBM InfoSphere Information Server is capable of integrating data on demand across multiple and high volumes of data sources and target applications using a high performance parallel framework. InfoSphere DataStage also facilitates extended metadata management and enterprise connectivity.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implications</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Derived requirements</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Related Decisions</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Architecture Guidance

For Data patterns

Re-usable Patterns for Big Data
Data that relies on a Massively Parallel Processing (MPP) environment to acquire, transform and supply information to applications. The Apache Hadoop processing environment and associated Hadoop Distributed File System (HDFS) accessed with NoSQL queries is the a de facto instantiation of this storage class.

Virtual Data Mart

Virtual sources that retrieve information from underlying big data source on-the-fly. Data stays in-place in HDFS.

Big Data Table

Spreadsheet-like row and column data extracted from an HDFS. Typically augmented and extended using row/column metaphor. Ad hoc dimensional views as popular.

Big Data Index

Indexed information used by search engines. Column indexing allows faceted dimensions to be explored.

Big Data Warehouse (HDFS)

Data that relies on a Massively Parallel Processing (MPP) environment to acquire, transform and supply information to applications. The Apache Hadoop processing environment and associated Hadoop Distributed File System (HDFS) accessed with NoSQL queries is the a de facto instantiation of this storage class.

Legend

Data Mart

Business Intelligence (BI) information that must typically be aligned to new acquired HDFS sources. Accessible via In-Memory, Hybrid or Traditional databases.

Virtual Data Mart

Virtual sources that retrieve information from underlying big data source on-the-fly. Data stays in-place in HDFS.

Big Data Table

Spreadsheet-like row and column data extracted from an HDFS. Typically augmented and extended using row/column metaphor. Ad hoc dimensional views as popular.

Big Data Index

Indexed information used by search engines. Column indexing allows faceted dimensions to be explored.

Big Data Warehouse (HDFS)

Data that relies on a Massively Parallel Processing (MPP) environment to acquire, transform and supply information to applications. The Apache Hadoop processing environment and associated Hadoop Distributed File System (HDFS) accessed with NoSQL queries is the a de facto instantiation of this storage class.

Application

Anything that augments, transforms, extracts or otherwise processes that data. Can be simple like a REST or Native API function. Also includes applications related Information Interaction.
Solution Pattern 1: Landing Zone Warehouse Pattern

Associated Use Cases

- Enhanced 360°View of the Customer, Data Warehouse Augmentation

Description

- Reporting and analysis in finance, GRC or similar domain.
- Data extracted from Landing Zone

Typical Steps

1. Big Data Warehouse HDFS Landing Zone built via ETL batch processes. Mix of structured, semi-structured and unstructured data extracted from a variety of external sources.


Solution Pattern 2: Landing Zone Virtual Warehouse Pattern

Associated Use Cases

- Enhanced 360°View of the Customer, Data Warehouse Augmentation

Description

- Reporting and analysis in finance, GRC or similar domain.
- Data stays in Landing Zone

Typical Steps

1. Big Data Warehouse HDFS (a.k.a. Landing Zone) built via ETL batch processes. Mix of structured, semi-structured and unstructured data extracted from a variety of external sources.

2. Virtual Report Mart Virtual database built using SQL/Hive Style over Big Data Warehouse HDFS. Data stays in HDFS.


Source Data: Social, Machine/RFID, Transactions...
Solution Pattern 3: Landing Zone Table Report Mart Pattern

Associated Use Cases
- Big Data Exploration, Security / Intelligence Extension

Description
- Reporting and analysis in finance, GRC or similar domain.
- Ad hoc exploration and discovery performed using Big Data Table exploration tool (like Big Sheets)

Typical Steps
1. Big Data Warehouse HDFS Landing Zone built via ETL batch processes. Mix of structured, semi-structured and unstructured data extracted from a variety of external sources.
2. Big Data Table built from landing zone sources.
3. Big Data Summary Mart Virtual database built from Big Data Table Content.
Component Pattern 7.1: Analytics Mart Pattern

Characteristics
- **Sub-pattern** variant of Component Pattern 7: Exploration Mart Pattern where exploration is moved to the Information Interaction.
- Domain expert (Line Of Business) constructs BA Analysis data using variety of tools and sources.

Components
1. Analytics Mart built from Discovery Layer Landing Area Zone, Sources and other Sources.
2. Query and Discovery tools aid with search, extract and disambiguation of uncertain sources and relationships.
3. Presentation, Visualization Sharing optionally performed over Analytics Mart
4. Analytics Mart optionally shared to other components in Information Interaction.

Attributes

<table>
<thead>
<tr>
<th>Type</th>
<th>Size</th>
<th>Update Interval</th>
<th>Data Rates</th>
<th>Data Quality</th>
<th>Response Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report Mart (In-Memory)</td>
<td>Tabular In-Memory</td>
<td>Up to 10 TB</td>
<td>Variable</td>
<td>Medium</td>
<td>Fast</td>
</tr>
</tbody>
</table>
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle.
Operational Model

- The operational model introduces the high-level operational nodes and associated deployed software components of the Big Data Enhanced Analytics System Architecture.

- In early stages, Operational Model is used:
  - As an early basis for design reviews and walkthroughs, including confirmation that the business problem is well articulated and that there is a viable IT solution.
  - As a way of dividing large problems so that each node can be worked on in relative isolation, but yet be part of the same solution vision.
  - As the basis for early analysis of nonfunctional requirements such as performance, availability, and capacity, including confirmation of the viability of a solution through specification of the expected nonfunctional characteristics of nodes and components.
  - To identify necessary technical, infrastructure, and other middleware components and subsystems.
  - To contribute to early estimates of the cost of the infrastructure to be used both for budgeting and as part of the business case for the solution.

- In the later stage, an Operational Model at the specification level is used:
  - To document the distribution of application and technical subsystems (deployment units) on preliminary (conceptual or specified) nodes so they can ultimately be installed and run on physical computer systems and on virtualized environments.
  - As the basis for detailed design reviews and walkthroughs, prior to finalizing product selection.
  - As a detailed technical specification against which an architect can evaluate alternative products or even against which technology vendors can submit tenders.
  - As the basis for detailed prediction of performance, availability, and other service level characteristics. (Predictions are based on the overall architecture and the specifications of deployment units within it. It will have to be revisited, via system tests, when specific products have been chosen.)
  - As the basis for a check that all the necessary business and technical functionality has been identified.
  - As the basis for cost estimates of the required infrastructure.
Operational Model – Continued

- The Operational Model is based on the BigData Enhanced Analytics architecture overview diagram.
- This Operational Model represents the Big Data Enhanced Analytics solution conceptual topology and it is showing key nodes across IT zones. This is only conceptual operational representation.
This Operational Model represents the Big Data Enhanced Analytics deployment topology. In this example, the Dell blades and VMware virtualized environment are being used for the majority of software components for deployments. Databases are being deployed into Dell servers. The Hadoop Cluster is deployed into IBM two Rack Systems and Data Warehouse and Analytics are deployed into PuraData Systems for Analytics.
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle

Understand Client

Define Client Requirements

Design Solution

Detail Design to Define BOM

Best Practices DevOps

Business Drivers

Describe the key business drivers for the project, the KPIs or CSFs, and how they align with Cloud computing.

Use Cases

What are the functional requirements expected from the Cloud and who are the key actors. Expressed as Use Cases.

System Context

The system context should define the boundary of the Cloud, and the integrations with OSS / BSS systems

Architecture Decisions

Clearly documented decisions on key architectural points including the rationale for the decision.

Functional & Non-Functional Requirements

NFRs should be defined to cover the volumes, capacity, scale, availability, security, operational and monitoring aspects of the Cloud

Architecture Overview

Architecture overview diagram should define the high level components, their placement.

Operational Model

Design and consider the components of the solution both at a physical and logical level.

Content provided by BDA RA

Business Drivers

Use Cases

Requirements, Security

Architecture Overview,

Architecture Decisions

Operational Model,

DevOps

Best Practices

DevOps

Define the boundaries of the project, inclusions, exclusions, dependencies, and align phases with milestones in the roadmap.
Big Data Analytics – Implementation best practices

- **Identify data components**
  - Enterprise Content Management
  - Data Warehousing, Data Management; Business Intelligence; Analytics – data at rest & data in motion, Data modeling, Data Governance

- **Identify the Implementation pattern(s)**
  - Big data exploration:
    - Enhanced 360-degree view of the customer:
    - Security/intelligence extension:
    - Operations analysis:
    - Data warehouse augmentation:

- **Develop a maturity Model**
  - Develop a Big Data capability maturity model that will reflect your enterprise business and IT needs. Refer to other models. Ex. IBM Data Governance Council Maturity Model” → [http://bit.ly/MthvX4](http://bit.ly/MthvX4)

- **Develop/use BDA Reference architecture**
  - Develop a Big Data Analytics Reference architecture that serves as a blueprint for your Big Data Analytics solutions.
Big Data Analytics – Deployment best practices

Identify deployment options
- BYO Hardware
- IT Appliances:
  - Private Cloud:
  - Public Cloud:

Identify security requirements
- People, Data, Application and infrastructure security needs;
- Security intelligence needs

Risk management
- Monitor big data activity from applications and users for threats.
- Establish data traceability and auditability;
- Enforce change controls;
- Encrypt and mask data to make it unusable

Develop a BDA Governance
- Information is understood
- Information is correct
- Information is holistic
- Information is current
- Information is secure.
- Information is documented
Big Data Analytics – maturity model

1. Initial
   Ad-hoc; inconsistent; some PoCs / trials; stand alone Big Data Analytics environment

2. Managed
   planned & managed project specific implementation; Some hadoop skills; some insights with EDW

3. Defined
   Quantitative defined objectives; Predictable outcomes; End-to-end integration with EDW;

4. Quantitatively Managed
   Big Data organizational governance; continuous updates inline with business objectives

5. Optimizing
   Best Practices
Big Data and Analytics - Information governance

Information Governance

Standards
- Information Requirements
- Information Identification
- Information Architecture
- Information Dependencies

Protection
- Information Usage
- Information Privacy

Lifecycle
- Information Retention
- Information Disposal

Quality
- Information Values Quality
- Information Supply Chain Integrity

Compliance
- Policy Administration
- Policy Implementation
- Policy Enforcement
- Policy Monitoring
Big Data & Analytics - Security intelligence example

**Requirement 1:** Enhanced Intelligence and Surveillance Insight

**Requirement 2:** Real-time Cyber Attack Prediction and Mitigation

**Requirement 3:** Crime Prediction and Protection
Security intelligence example – implementation steps

- Identify data components
- Business Intelligence; Analytics – data at rest & data in motion, Data modeling, Data Governance
  - Analyze data-in-motion and at rest
  - Analyze network traffic
  - Analyze Telco and social data
- Identify the implementation pattern(s)
- Security/intelligence extension:
- Develop a maturity Model
- Develop a Big Data capability maturity model
- Develop/use BDA Reference architecture
- Decision to utilize IBM Big Data Reference Architecture
Security intelligence example – deployment steps

- **Deployment**
  - **Identify deployment options**
    - BYO Hardware
    - IT Appliances:
      - Private Cloud:
  - **Identify security requirements**
    - infrastructure security needs;
    - Security intelligence needs
  - **Risk management**
    - Monitor big data activity from applications and users for threats. Establish data traceability and auditability;
  - **Develop a BDA Governance**
    - Information is understood; Information is correct
    - Information is holistic; Information is current
    - Information is secure; Information is documented
Security intelligence example – maturity model

1. Initial
   - Ad-hoc; inconsistent; some PoCs / trials;
   - stand alone Big Data Analytics environment

2. Managed
   - Planned & managed project specific implementation;
   - Some hadoop skills; some insights with EDW

3. Defined
   - Quantitative defined objectives;
   - Predictable outcomes;
   - End-to-end integration with EDW;

4. Quantitatively Managed
   - Big Data organizational governance;
   - continuous updates inline with business objectives

5. Optimizing

1. Evaluate & Pilot IBM Streams for Data in motion requirements.
2. Review applicable Big Data patterns
3. Document selected patterns and solution components
4. Identify SIEM requirements
5. Analyze cloud deployment needs
6. Start on BDA Governance

1. Develop SIEM QRadar architecture;
   - identify SW components for security surveillance
2. Develop data models
3. Pilot - Hadoop for Big Data storage
4. Apply Cognos & SPSS for projects 1-3.
5. Test BI features to Private Cloud
6. Extend and apply BDA governance

1. Develop Data in motion system components.
2. Migrate Phase I security surveillance monitoring to private cloud
3. Access analytics through i2 ANB
4. Identify SIEM requirements
5. Establish BDA Governance across projects

1. Extend real-time detection and prevention of threat
2. Identify patterns
3. Pilot - Hadoop for Big Data storage
4. Identify SIEM requirements
5. Establish BDA Governance across projects

Best Practices
The BDA RA Reference Architecture material is based on TeamSD and applies to the entire sales cycle.

Understand Client

Define Client Requirements

Design Solution

Detail Design to Define BOM

Best Practices DevOps

Business Drivers

Describe the key business drivers for the project, the KPIs or CSFs, and how they align with Cloud computing.

Use Cases

What are the functional requirements expected from the Cloud and who are the key actors. Expressed as Use Cases.

System Context

The system context should define the boundary of the Cloud, and the integrations with OSS / BSS systems.

Architecture Decisions

Clearly documented decisions on key architectural points including the rationale for the decision.

Functional & Non-Functional Requirements

NFRs should be defined to cover the volumes, capacity, scale, availability, security, operational and monitoring aspects of the Cloud.

Architecture Overview

Architecture overview diagram should define the high level components, their placement.

Operational Model

Design and consider the components of the solution both at a physical and logical level.

Content provided by BDA RA

Business Drivers

Use Cases

Requirements , Security

Architecture Overview,

Architecture Decisions

 Operational Model,

DevOps

Best Practices

Technology Brief

Define the boundaries of the project, inclusions, exclusions, dependencies, and align phases with milestones in the roadmap.

Best Practices

DevOps

Define the overall timeline, phases, and key milestones that will shape the plan and overall delivery.

DevOps

Define and consider the components of the solution both at a physical and logical level.
DevOps for Big Data and Analytics Applications

Plan and Measure
- Project and Portfolio Management
- Requirements Management
- Enterprise Architecture

Develop and Test
- Application Lifecycle Management
- Configuration and Change Management
- End-to-end Traceability
- Quality Management
- Data and Service Virtualization for Testing
- Development Process Management

Release and Deploy
- Continuous Delivery
- Data Source Configuration Management
- Release Management

Monitor and Optimize
- Continuous Monitoring
- Customer Feedback
Data Lifecycle

Cross Industry Standard Process for Data Mining (CRISP-DM)

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Big Data Information Integration and Governance

- Find and Understand Big Data
- Prepare Big Data for Usage
- Defend and Build Confidence in Veracity of Big Data
- Secure Big Data and Comply with Privacy Regulations
- Audit and Archive Big Data
- IBM Information Governance Unified Process
DevOps and Data Lifecycle

Data sources:
- Machine and sensor data
- Image and video
- Enterprise content
- Transaction and application data
- Social data
- Third-party data

Real-time analytics:
- Information integration and operational information
- Exploration, landing and archive

Actionable insight:
- Cognitive
- Decision management
- Predictive analytics and modeling
- Reporting, analysis, content analytics

Enhanced applications:
- Customer experience
- New business models
- Financial performance
- Risk
- Operations and fraud
- IT economics

Discovery and exploration

Information integration & governance

Data sources: System, Security, Storage

DevOps for Big Data

Data Lifecycle
Backup
Glossary of Terms

**Platform as a service (PaaS):** It is a category of cloud computing services that provides a computing platform and a solution stack as a service. Basically it includes the hardware, Operating system, storage and network and the middleware, frameworks together a solution.

**Software as a service (SaaS):** Application software owned, delivered and managed remotely by one or more providers. The provider delivers an application based on a single set of common code and data definitions that is consumed in a one-to-many model by all contracted customers at anytime. SaaS is purchased on a pay-for-use basis or as a subscription based on usage metrics.

**Proof of Concept (POC):** a realization of a certain method or idea to demonstrate its feasibility or a demonstration in principle, whose purpose is to verify that some concept or theory has the potential of being used.

**Big Data Volume:** Volume is the obvious Big Data trait. Aggregates of data that use to be measured in Petabytes are now measured in zettabytes or a billion terabytes.

**Big Data Variety:** The *Variety* characteristic of Big Data is all about trying to capture all of the data that pertains to our decision-making process.

**Big Data Velocity:** is the rate at which data arrives at the enterprise and is processed or well understood.

**Big Data Veracity:** is the term that refers to the quality or trustworthiness of the data.

**Information Ingestion:** Is the process to extract, transform the raw data from many sources (traditional sources) and load into a data repository (data warehouse or data mart) to make it available for analytics. It includes staging areas and specialized transformation engines that are performing enrichment and restructuring operations on the raw data. It also includes the data quality and standardization process to ensure that the data is conformed to corporate standards and well understood by every one that needs to use it.
Enterprise Data Warehouse (EDW): Consist of a single environment that contains subject areas consolidate across divisions or line of business. Data is highly normalized and provides application neutral base access. The data model maps the system of records. The system manages very complex and heavy workloads. It requires high available and fast recovery capabilities.

Operational Data Stores (ODS): ODS is another type of implementation with time-sensitive operational data that needs to be accessed efficiently for both simple queries along with complex reporting to support tactical business initiatives. In the traditional architecture, the analytical sources are created based on specific business needs. As new business requirements arise, a new data process is needed to generate the data and make it available for consumers (users and/or applications).

Predictive Analytics and Modeling: Predictive analytics is a discipline that leverage advanced analytical algorithms (Linear Regression, Decision Tree, etc) to process historical data and create models that can make predictions about future outcomes.

Metadata: Metadata Management is the capture, versioning, approval, usage, and analysis of the different types of metadata found in an Information Management environment.

Metadata Catalog: contains the semantic definitions for business and IT terms, data models, types, and repositories. It provides functionality to browse, discovery, and search of metadata assets.

Master Data: Are the key business data elements that may include information about customers, products, employees, suppliers, vendors, etc. and shared as a single source of basic business data across systems, applications, and processes for an enterprise.

Reference Data: Is the data that defines the standard data domain values used within an organization. Examples of Reference Data are: units of measure, country codes, corporate codes, conversion rates (currency, weight, temperature, etc.), calendar dates, etc

Information Provisioning: Various provisioning mechanisms for locating, retrieving, transforming and aggregating information from all types of sources and repositories.

Landing / Deep Data Zone: Area for raw data for querying, exploration, data transformations, and pseudo archival (aka, online queryable archive). The Area integrates and modernizes with traditional Integrated Warehouses, Discovery, MDM 360 & Content Management

Information Provisioning: Various provisioning mechanisms for locating, retrieving, transforming and aggregating information from all types of sources and repositories.
Purpose

- This presentation, derived from the IBM Big Data & Analytics Reference Architecture document is meant to drive consistency across targeted technical sales channels in the roll-out of the IM & BA reference architectures aligned to the 5 big data use cases and the industry scenarios.

- The purpose of IBM Big Data & Analytics (BD&A) Reference Architecture is to inform and guide sales, technical sales, services and other professionals who are involved in selling IBM solutions or deploying them with clients.

- The Reference Architecture is intended to be used by a wide range of professionals selling IBM software and designing end-to-end big data & analytics client solutions. The reference architecture can also be reused by IBM clients and partners to gain further insight on how to benefit from IBM Software in big data & analytics projects (appropriate disclosure is required).

- The target audience also depends on the specific work product included in the reference architecture. The Architecture Overview, Business Directions, Requirements, Technology Brief are expected to be consumed by business leaders and architects.
Big Data Exploration

Integration & Governance

Streams → BigInsights → Data Explorer → Content Analytics → Warehouse

Connector Framework
- CM, RM, DM
- RDBMS
- Feeds
- Web 2.0
- Email
- Web
- CRM, ERP
- File Systems

Application Builder
- Exploration User Experience
- Analytics Experience
- Content Analytics Miner
- SPSS Modeler
- Cognos BI

Big Data & Analytics
Enhanced 360° view of the Customer

**SOURCE SYSTEMS**

- **CRM**
  - Name: J Robertson
  - Address: 35 West 15th St, Pittsburgh, PA 15213

- **ERP**
  - Name: Janet Robertson
  - Address: 35 West 15th St, Pittsburgh, PA 15213

- **Legacy**
  - Name: Jan Robertson
  - Address: 36 West 15th St, Pittsburgh, PA 15213

**Infosphere Master Data Management**

360° View of Party Identity

- **First:** Janet
- **Last:** Robertson
- **Address:** 35 West 15th St
- **City:** Pittsburgh
- **State/Zip:** PA / 15213
- **Gender:** F
- **Age:** 48
- **DOB:** 1/4/64

**Cognos Consumer Insight**

**Cognos BI**

**InfoSphere Data Explorer**

**Unified View of Party’s Information**
Operations Analysis

Real-time Monitoring
- InfoSphere Streams
  - Capture Data Stream
  - Identify Anomaly

Historical Reporting and Analysis
- InfoSphere BigInsights
  - Raw Data
- SPSS Modeler
  - Predict and Classify
  - Aggregate Results
  - Data Warehouse
  - Store Results
  - Predict and Score

Decision Management
- Cognos BI
- Federated Navigation and Discovery
Data Warehouse Augmentation

1. Pre-Processing Hub
   - BigInsights Landing zone for all data
   - Data Explorer
   - SPSS Modeler
   - Streams Real-time processing
   - Data Warehouse

2. Query-able Archive
   - Information Integration
   - BigInsights
   - Data Explorer Find and view the data
   - Data Warehouse

3. Exploratory Analysis
   - Combine with unstructured information
   - Streams Offload analytics for microsecond latency
   - Cognos BI
   - Data Warehouse
   - SPSS Modeler
   - Data Warehouse

- BigInsights
- Cognos BI
- Data Explorer
- SPSS Modeler
- Data Warehouse
- Streams
- BigInsights
- Data Warehouse
Telco: Real Time Contextual Marketing Campaign

CDRs, Top Ups, Balance Enquiry

Usage, Location & History Monitoring

Current State & Predictions Per Customer

Streams Real-time Event Detection and Predictive Analytics

Real-time Event Detection

EDW

Campaign monitoring

Event-based Campaign

Marketing solutions for more profitable, timely, and measurable business outcomes
Banking: Real Time Credit Card Campaign

Transactions, Location

Usage, Location & History Monitoring

Current State & Predictions Per Customer

EDW

Real-time Event Detection

Campaign monitoring

Streams
Real-time Event Detection and Predictive Analytics

Unica
Event-based Campaign

Marketing solutions for more profitable, timely, and measurable business outcomes