OCITA Spring Event

Mike King

Enterprise Technologist, Big Data

Wright Patterson AFB; May 19, 2016

Acronym Key - Part 1

- VLDB – Very Large Database
- PK – Primary Key
- AK – Alternate Key
- COTS – Commercial Off-the-Shelf
- KV – Key value
- JSON – Java Script Object Notation
- BSON – Binary Structured Object Notation
- iOT – internet of things
- JDBC – Java DataBase Connectivity
- CDH – Cloudera Distribution for Hadoop
- EDH – Enterprise Data Hub
- EDW – Enterprise Data Warehouse
- ETL – Extract, Transform & Load
- ELK – Elastic Search, Logstash & Kibanna
- XML – eXtensible Markup Language
- SQL – Structured Query Language
- CRM – Customer Relationship Management
- TPC – Transaction Performance Council

Acronym Key - Part 2

- SOA – Service Oriented Architecture
- API – Application Programming Interface
- CSV – Comma Separated Values
- RDBMS – Relational DataBase Management System
- MPP – Massively Parallel Platform
- ML – Machine Learning
- CoE – Center of Excellence
- HTTP – HyperText Transfer Protocol
- HDFS – Hadoop Distributed File System
- BDE – Big Data Extensions (Vmware)
- FTE – Full Time Equivalent
- SIEM – Security Information Event Management
- MQ – Message Queuing
- ERP – Enterprise Resource Planning
- HA – High Availability
- DBA – DataBase Administrator
- DWFT – DataWarehouse Fast Track
- *aaS – anything as-a Service

Big Data

Trends Affecting Big Data

Technology

- Virtualization: App, CM, Mgt, Client Tools
- Automation
- Integration
- Tools

Consumption pattern

- Cloud
  - Three types
- *aaS
  - I, a, p, s, DB
  - <You Name It>

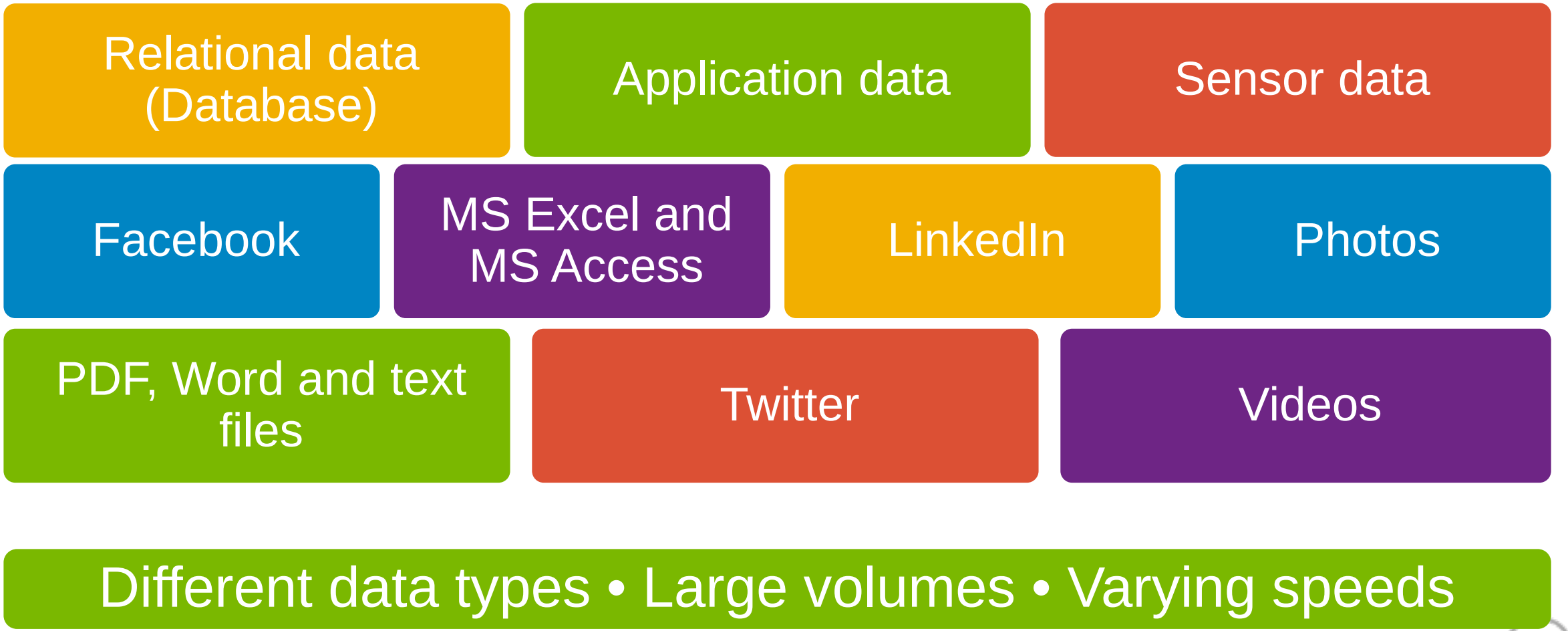
Data tsunami

- iOT
- Data, data & more data
- Mobile

The profession

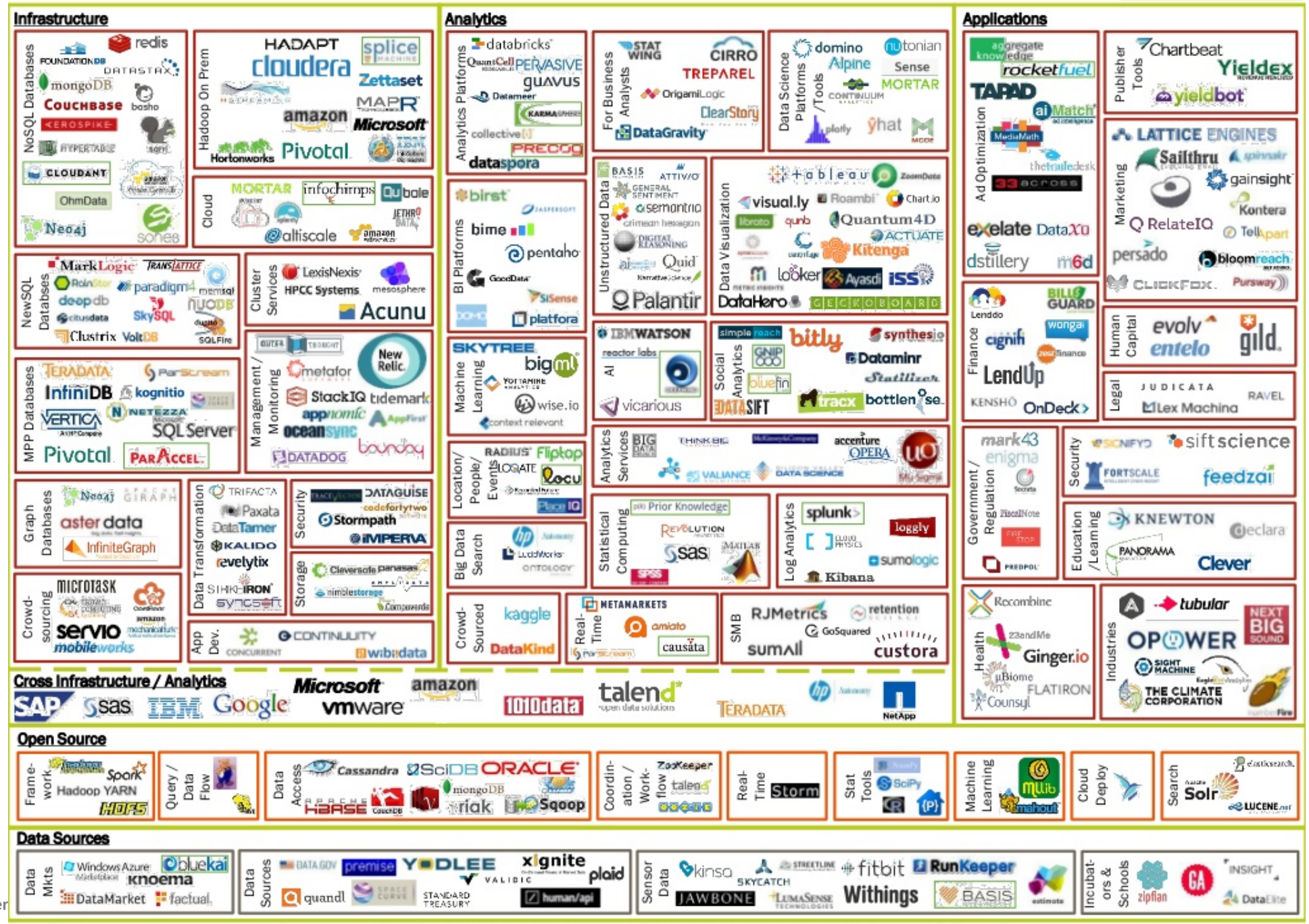
- Analytics for all, & all…
  - Varying needs
- Data Science
- Skills demand
  - Needs
  - Roles
  - How to fill
- Training

Big Data is really complex data, with needs that extend beyond the existing tool chain



BIG DATA LANDSCAPE, VERSION 3.0

Exited: Acquisition or IPO



Customer Success Stories

Vertical	Use Case	Gap/Challenge	Solution	Savings
Retail	ETL Offload	Time2Market	Hadoop	20hrs==>17mins
Retail	CRM & Loyalty	Collect, org, aggregate	CDH	
Retail	Analytics	Retention	depth, Hadoop	~8x storage savings
CRM Mktg	OMNI Channel	Multi-structured data	Dell PowerEdge	60% cost redux
CRM Mktg	House holding	Identity, Address Matching	Dell PowerEdge	10x speed improvement
Security	Event collection & analysis	Scalability, Cost	Dell PE, Storm & SAS	$17/GB ==> $.21/GB
Infrastructure				20:1 time redux
IT	EDW Archive	High Growth, Exp$ DW	CDH, Rainstore	80% redux storage cost
IT	ETL Offload	complexity, cost, speed	CDH+EDH, Infa BDE	reduced cost, improved speed

Customer lifetime value of Big Data

UK – online services

- Jan 2013: 200 nodes
  Always ran Hadoop – saw mega growth from Jan 2013: 200 nodes
- Primary use case: Web 2.0 as core
- Growth: Jan 2015 = +800 nodes

G500 SI-Telco

- Jan 2013 = 150 nodes
- Primary use case: Top Secret Government Work
- Growth: Jan 2014 = +150 nodes

US-based Telco

- Feb 2013 – 42 node POCs
- Primary use case: Log Files, BDaaS, & Churn Analysis
- Growth: March 2015- +2200 nodes

Financial Services

- Nov 2013: 12 nodes POC
- Primary use case: Log Files, Fraud Analysis, 360 Customer View
- Growth: March 2015 = +220 nodes

Why Dell?

Dell Differentiators

Integrated Offerings
Closely integrated with hardware and software offerings, addressing customer needs across all dimensions

Consulting & Services
Dell solutions and frameworks accelerate and de-risk Big Data Implementations

Dell Appliances
Offer out of the box deployments

Dell Big Data Lab
Round Rock Texas & India Enables Rapid Enablement and Deployment

Dell Cloud Services
SecureWorks and Foglight enhance the Big Data platform value

Framework & Reference Architecture
Tested and certified by vendors such as Cloudera and Hortonworks

Dell, A Very Differed Provider

Why Dell is different

- Modular
  - Plug ‘N play
- Happy to fill in the gaps
- Complete when we need to be
  - Servers, storage, networking, software & services.
- Products enhanced to work with Big Data
- Solutions
  - Engineered
  - Custom

Dell & Hadoop – Performance Matters

- **#1 TPCx-HS Hadoop Price/Performance in the industry** at scale factors of 1TB, 3TB, 10TB, and 30TB
- **#1 TPCx-HS Hadoop Performance in the industry** at scale factor of 10TBSF10
- **The Dell Cloudera Reference Architecture for Hadoop provides the #1 TPCx-HS Hadoop Price/Performance in the industry** at scale factors of 1TB, 3TB 10TB, and 30TB
- **PowerEdge R730XD provides the #1 TPCx-HS Hadoop Performance in the industry** at scale factor of 10TB
- **Up to 64% better TPCx-HS Price/Performance compared to Cisco** at scale factor of 10TB
- **Up to 13% better price/performance compared to Huaweii** at scale factor of 1TB

,

Support & Services

- DSC (for free)
  - Briefing
  - Architectural design session
  - POC
- Prof Services (for fee)
  - Jumpstart
  - Select Use Cases
  - Custom engagements

BI, Analytics & Big Data Capabilities

Microsoft SQL Server

Microsoft Azure

Office

Microsoft SQL Server Parallel Data Warehouse Analytics Platform System

cloudera

SAP HANA

BUSINESS INTELLIGENCE & ANALYTICS

- Dashboards
- Self Service
- Reports
- OLAP
- Predictive

DATA CONSULTING SERVICES

- Jumpstart
- ETL Development
- Health Checks
- Proofs of Concept
- Platform Migration

PLATFORM CONSULTING SERVICES

- Hub & Spoke
- App Install
- Cluster Config
- Cloud Deployment
- Disaster Recovery

INFRASTRUCTURE DEPLOYMENT & SUPPORT

- O/S & Software
- Merge Center
- Rack & Stack
- ProDeploy
- SupportAssisst

Services Offer Matrix

	Service Offer	Team	Format	Service	Est Cost / Duration
1	**Hadoop H/W Install**	EDT	• SKU for Quickstart • Custom SOW for RA	• Rack & Stack • Label / Cable • Priced by size	~$4k ~days
2	**Cloudera Deployment**	EDT	• SKU for Quickstart • Custom SOW for RA	• O/S Install • Foundation services install • Configuration	~$9k ~days
3	**Cloudera Basic Jumpstart**	GICS	• Repeatable SOW/ SKU coming Q1 • ALL Cloudera (QS & RA)	• Training • As-is / To-Be • Hands on labs • Roadmap Deliverable	$18k FF 2 weeks onsite 1 FTE
4	**Cloudera Health Check**	GICS	• Repeatable SOW/ SKU coming Q1 • ALL Cloudera (QS & RA)	• Time-boxed cluster certification • Up to 2 clusters, 100 nodes • Cloudera best practice	$15k FF 1 week onsite 1 FTE
5	**Hadoop Active Archive Proof of Concept**	GICS	• Repeatable SOW/ SKU coming Q1 • ALL Cloudera (QS & RA)	• Real world PoC using native tools (ie: Hive, sqoop, flume, etc.) to demonstrate effective use case of Active Archive • Design, Development and non-prod deployment	$50k FF 5 weeks 1 FTE on/off site
6	**Hadoop ETL/DW Offload Proof of Concept**	GICS	• Repeatable SOW/ SKU coming Q1 • ALL Cloudera (QS & RA)	• Real world PoC using native tools (ie: Hive, sqoop, flume, etc.) to demonstrate effective use case of Active Archive • Design, Development and non-prod deployment	$50k FF 5 weeks 1 FTE on/off site
7	**Custom Workload**	GICS	Custom SoW	• Custom workload specific to Cloudera/Hadoop • Any deviation in scope from SKU offers	Custom Quote

Use Case Taxonomy

Use Cases by Industry/LOB

Retail/Marketing	Finance	Healthcare	Pharmaceutical	Manufacturing
Anticipating customer needs	Reducing risk and detecting fraud	Improving patient care and reducing cost	Ensuring regulatory compliance and validation	Continuous process improvement
Customer insight	Credit scoring	Fraud detection	Product traceability	Product quality
Customer retention	Customer analytics	Claims management	Stability and shelf life	Customer insight
Market basket	Fraud detection	Patient safety	Validated reporting	Demand forecasting
Media mix	Risk management	Risk mitigation	FDA compliance	Logistics regression
Price optimization	SOX compliance	Quality of care	Manufacturing	Improved operations

FSI	Healthcare	Manufacturing	Oil & Gas	Retail
Fraud prevention in credits and payments	Quality of care optimization	Proactive quality assurance	Horizontal drilling enablement and optimization	Enablement of a 360-degree customer view
Risk modeling in investments banking	Clinical quality and cost analysis	Analysis of demand for new products and services	Seismic data processing	Generation of personalized offers
Cross-selling and upselling in retail banking	Genome processing and DNA sequencing	Product research guided by machine-generated data	Predicting where best to drill next	Enablement of first in-basket analysis
Insurance policy personalization	Population health management	Detection of supply chain issues	Which leases do I sell?	Merchandising and supply chain analysis
Mortgage lending portfolio valuation	Detection of fraud and suspicious transactions	Identification of cross-sell and upsell opportunities	Which sections should I acquire?	Isolation of products and mixes indicative of larger baskets

IT	Common	Finance	Banking	<My Bank>
ETL offload / acceleration	Customer profile	Risk arbitrage	Risk modeling	<Your cool UC>
Active archive	Customer 360			
Log aggregation	Churn analysis	Currency hedging		
agile data explore / self-service BI / data lake	Quality improvement	Insurance policy personalization	<Interesting stuff done by competitors>	<Your even cooler UC>
SIEM	Cross-sell, upsell			
Security - intrusion detection, others	Householding / matching		Mortgage lending portfolio valuation	
Reporting	Fraud analytics			
	Loyalty analysis			
	Profitability analytics			
	Canabilization analysis			

Skeleton Process

- Define goals & objectives
- Brainstorm use cases
- Assess
  - Complete
  - Data
- Cull
- Rank
- Solution architecture
  - How
  - What tools

- Assemble overall tech arch
  - RA
- Gaps
  - Skills
  - Process
- POC
  - One UC at a time
    - Learn
    - Adjust
    - Improve
    - Feedback
  - Next UC
  - Repeat

Use Cases

Use Cases

- Archive
  - Active
  - As needed
  - Platform retirement
- ETL
  - Offload
  - Performant
  - License redux
- Data Warehousing
  - Re-platform
  - Downsizing
  - Diet
  - Simplification
- Log Processing
  - COTS replacement
  - Performance, Parallelism
  - Enhancement
  - Functionality
    - ELK, flume
- Messaging, Streaming
  - Kafka, spark, flink, storm
- Integration
  - Structured, Multi-structured, Variable
  - RDBMS, nosql, files
  - Public, private & hybrid cloud

Hadoop & Big Data Solutions

About Hadoop

- SOA
  - Omnipresent
  - Rest APIs
- Logs
  - By product, program or not at all
  - CDH Ent – integrated for many
- One-offs
  - Doable
  - Minimize
- Plethora of choices
- Customizable
- Mostly Open Source
- Languages
  - Java
  - Python
  - R
  - Scala
- Growth
- Evolving
- Contenders and pretenders

About Hadoop Continued

- Store tons of data
  - All is now feasible
- Scale
  - Horizontal
- Mix disparate sources
- Ingest
  - Bulk
  - Small batches
  - Real-time
- Structure
  - Strongly type
  - Semi
  - Multi

- MPP SQL
- ML
- Predictive analytics
- Architecture
  - Enterprise
  - Technical
  - Solution

COTS Replacements with Hadoop

- ETL
  - Informatica
  - Abinitio
  - Data Stage
- SIEM
  - Arcsight
  - Logility
  - Splunk
  - LogRythm
- Data Archiving
  - Strong ERP focus
    - Informatica ILM
      - Applimation
    - IBM Optim
      - Princeton Softech
    - Solix
- Messaging
  - Tibco EMS
  - IBM MQ
  - MSMQ

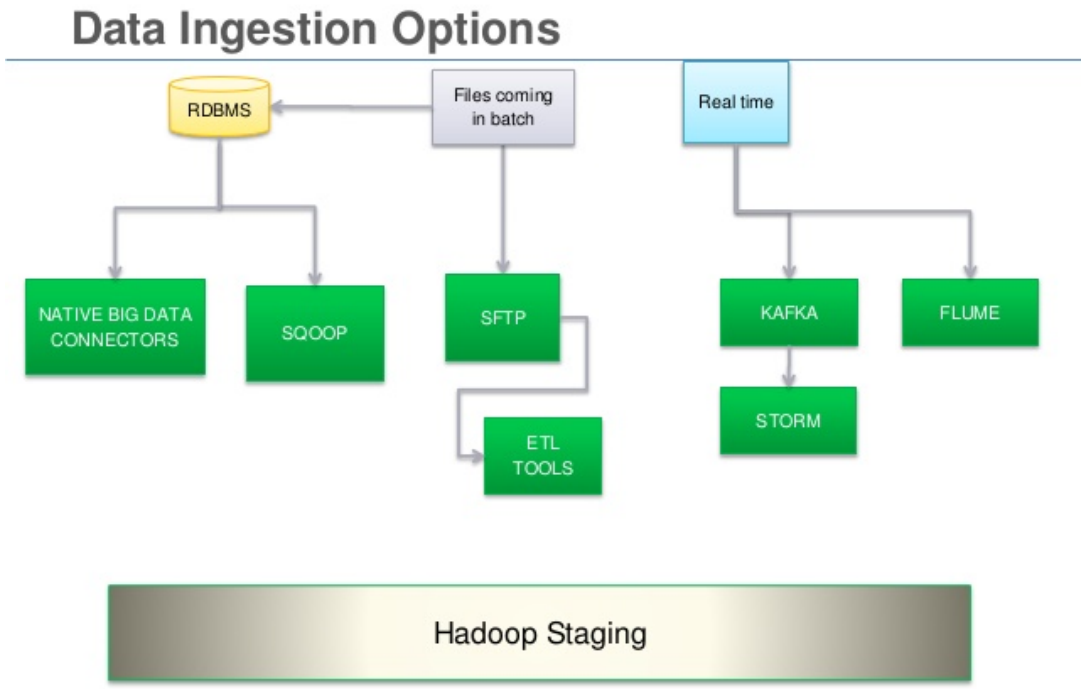
Data Sources

- Types
  - Public, private, purchased
- Sources & Sinks
  - Flume to HDFS
  - Flume to Kafka to HDFS
  - HTTP to Hbase
- Channels
  - JDBC
  - Memory
  - File
  - Custom

- Sources
  - Databases
  - Apps
    - ERP
    - CRM
    - Other purchased
    - Custom
  - Files
  - Messages

Ingestion

- File transfer
- HDFS client
- Sqoop
- Flume
- Kafka
- Custom
- Shareplex Connector for Hadoop
- Boomi



Skills, Training & Languages

- Skills
  - Inventory
  - Needs
  - Gaps
  - Buy, rent, grow
  - CoE
  - Mentor
- Training
  - Online
  - Self-paced
  - Tutorials
  - For free
  - For small fee $
  - Drivers license
  - Cheat sheets

- Languages
  - Not just one
  - Which one(s)?
    - Java
    - R
    - Python
    - Scala
  - Shape usage
  - Justify choices

Blueprint for Big Data & Analytics

Big Data and Analytics Blueprint Portfolio

Dell Software Suite

Statistica Data Analytics Suite

Dell Boomi Integration Tools | Dell Toad Data Management | Dell SharePlex Replication Connector for Hadoop

Reference Architectures

Dell | Cloudera Apache Hadoop Solution on R730XD
Start and up to 15 Nodes, Scales to 445 nodes, Scales 45+ nodes

SQL DWFT
Start with 730/PS6210S to 17TB, Scales on 730xd to 21TB,
Scales on 730/PS6210S to 26 TB, Scales on 730/SC4020 to 55TB

Dell | Cloudera | Syncsort Data Warehouse Optimization for ETL Offload RA
(June 19, 2015)

Engineered Solutions

Microsoft APS Appliance
PDW: 3 nodes, Scales PDW + Hadoop to 6 nodes, Scales PDW + Hadoop 9 – 54 nodes

Dell QuickStart 5.5 for Cloudera Hadoop
5 nodes

SAP HANA Appliance
Single Server configurations scale from 128GB – 1.5 TB RAM;
Scale Out cluster configurations scale from 2-16TB RAM (up to 24TB w/R930 – due September, 2015)

RA Implementations: Engage your Big Data Overlay Sales Team

Consulting

Deployment

Custom Solution Architecture

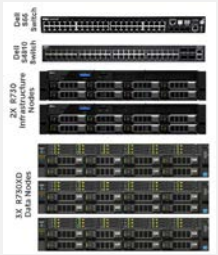
Training: Bundled

ProSupport Plus

Dell Hadoop Solution Offerings Summary

Dell QuickStart 5.6 for Cloudera

- Includes all hardware/software/services
- Cloudera Enterprise Support
- 5 Nodes & NW: Full PoC for < $150K
- PoC easily upgraded to Production



Dell | Cloudera 5.6 Solution

- Proven & tested Reference Architecture
- Foundational design with customizable components
- Robust, Enterprise-ready solution
- Massive, modular scalability

Dell | Cloudera | Syncsort Data Warehouse Optimization for ETL Offload

- Enables organizations to lower data transformation costs
- Builds operational efficiencies for laying a strong, cost-effective secure, scalable and robust solution for managing data
- Builds foundation to mature into advanced data analytics

Dell QuickStart for Cloudera Hadoop

Easy starting point for a complete Big Data solution

Dell QuickStart for Cloudera Hadoop delivers a full Hadoop cluster to start you on the pathway to taking control of Big Data

- Brings a full Hadoop proof of concept into organizations to allow them begin to develop expertise
- Delivers Hadoop capabilities for a low-entry price
- Incorporates full support from the experts as you take the first steps with Hadoop
- Teaches how to implement data collection, data management and data analytics to enable sophisticated strategies to build value for business
- Includes professional services to help you get started
- Ideal for pre-production use cases

Get started today with Dell QuickStart for Cloudera Hadoop for a fully-supported Hadoop solution with hardware, software, training and services

Key Benefits

- **Easy**:
  Dell QuickStart for Cloudera Hadoop includes all hardware, software, training and services
- **Affordable:**
  Build a full Hadoop environment for under $110K
- **Flexible:**
  Easily upgrades to a full production cluster

Dell | Cloudera Apache Hadoop 5.5 Solution, accelerated by Intel

Proven Hadoop Distribution for the Enterprise

Key differentiation & innovations

- A robust end-to-end Hadoop solution
- A solution built on experience, partnership, and innovation and tested and validated Reference Architectures

Value proposition

- A secure end-to-end data management solution
- To collect, mine, manage and analyze data
- Gain valuable business insights for unique competitive advantages

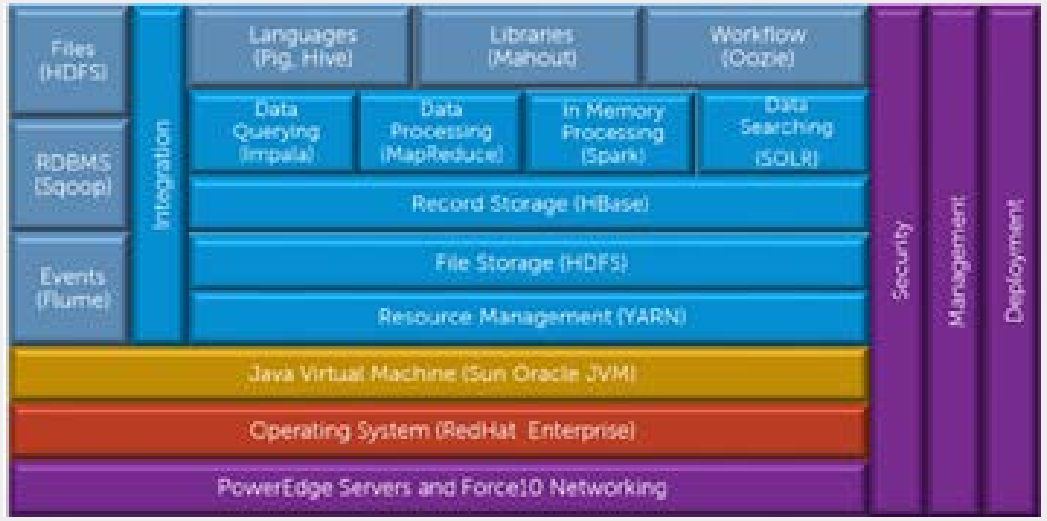
Target market

- All organizations from small, to medium and large enterprises – across all verticals

Better Together

- Dell | Cloudera | Intel for industry-leading, secure, infrastructure-optimized Hadoop solutions
- Streamlined to search, process, manage, and analyze all data

Dell | Cloudera Hadoop Solution for Big Data



Important updates in Cloudera 5.6 on 13G

Running on the PowerEdge R730xd

Updates to Cloudera Search

The release of Impala 2.0 that integrates Apache Spark into the platform and drives better batch processing with Spark 2.1 as the processing engine

Blueprint
for Big Data
& Analytics

Dell | Cloudera | Syncsort Data Warehouse Optimization for ETL Offload Reference Architecture

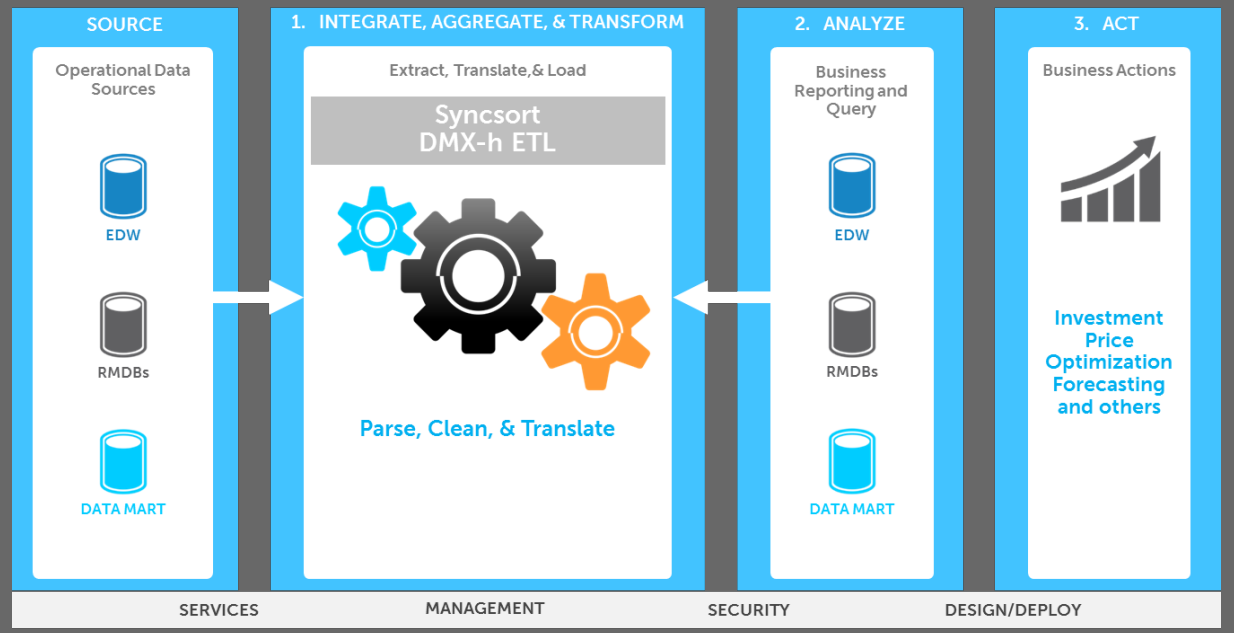
The first and only reference architecture for ETL offload with Hadoop

Scalable ETL with the flexibility of a Reference Architecture

- Scale Out hardware architecture – PowerEdge R730, R730xd, and high performance Dell S-Series Networking.
- Tight integration between Dell, Cloudera and Syncsort provides ease of deployment and maintenance with no performance impact or hurdles down the road.
- Close the Skills Gap by eliminating the need to develop expertise on MapReduce, Pig, Hive, and Sqoop.
- Fast Track Projects with automated conversion of legacy SQL scripts into efficient ETL processes in Hadoop without any coding.
- Comprehensive and collaborative service and support for the entire solution through it's complete lifecycle.

The Dell Difference

- Faster time to value through an optimized solution jointly designed by three market leaders.
- Detailed Reference Architecture Documentation
- Deployment guidelines detail best practices based on extensive experience with production deployments



NoSQL

NoSQL Database Types

- Four types
  - Columnar
    - Hbase, Cassandra
  - Document
    - MongoDB, Couchbase
  - KV
    - Riak, Redis
  - Graph
    - Neo4j, Titan
- How many do you need?
  - By type
  - Within type
- Who will manage them?
  - DBAs
- How do you access them?
  - SQL, nosql
  - Sequential

Nosql background, issues and considerations

- History
  - Google Big Table, Amazon Dynamo
- What does schema-less mean?
  - On read
  - Still structured
  - Embedded
  - Can vary between records
- Languages & formats used
  - Java, Python
  - JSON, BSON, XML, CSV

NoSQL background, issues and considerations continued

- Eric Brewer's CAP theorem
  - Can't do all three.
- What does NoSQL really mean?
  - Distributed, shared-nothing aggregate oriented database
  - "Not only SQL" versus "No"
- What are the factors for the various choices?
  - Best fit
  - Use case(s)
  - KV
  - HA, Multi-site
  - Network
    - Kevin Bacon
- Sharding
  - Partitioning

RDBMS versus NoSQL

RDBMSs	NoSQL DBs
Large user populations	Small user populations
Structured	Multi-structured, Semi-structured
Static schema	Schema evolution
Strong typing	Weak typing
Access by PK, AK, indexes	Mostly random access by PK
Complex structures	Simple structures
Feature rich	Bare bones functionality
Multi-purpose, shared by apps	Single purpose/use case, not shared by apps
OLTP	Not transactional
–ACID	–BASE
Complex queries	Simple queries
Small to medium sized dbs	VLDB, XL DB Size
3 way+ joins	few or no joins
Challenging, costly scalability	Horizontal scalability
SQL	Proprietary, differed access verbs/methods
COTS packages	Custom applications
Datamarts	

Nosql Commonalities

- Mostly open source
- Weak typing
- Multi-structured
- Horizontal scale
- No standardization
- VLDB
- Single purpose, per database

Nosql Differences

- Access
- Formats supported
- Features
- Management
- Administration
- VLDB
- Performance & tuning
- Resource consumption
- Language bindings

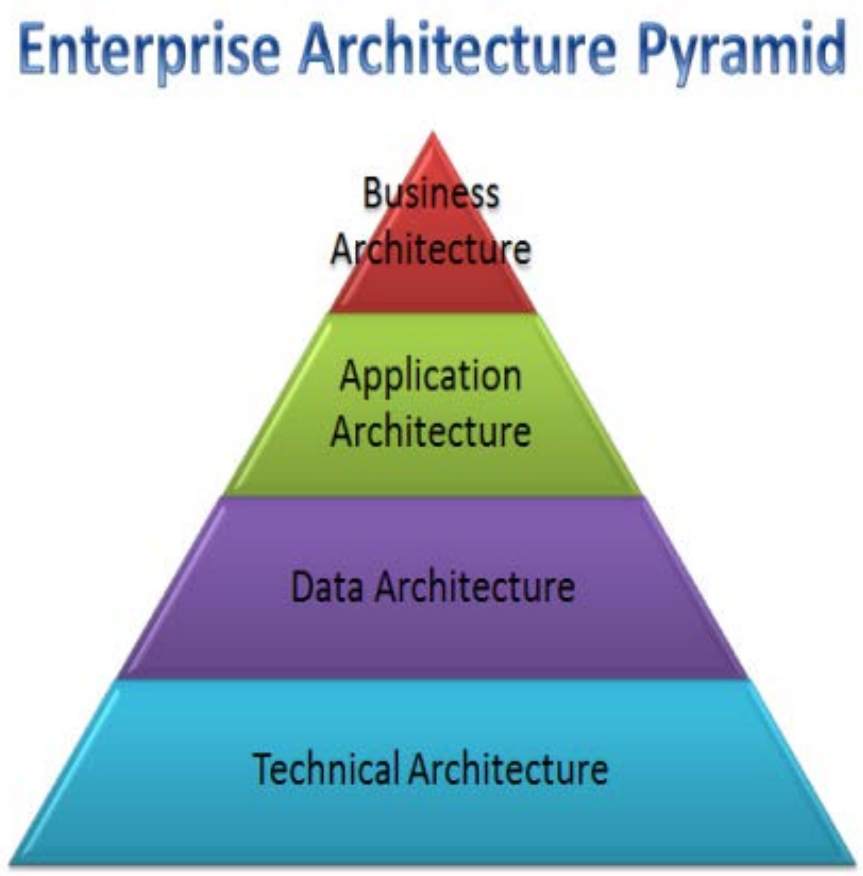
- APIs
- Security
- Persistence
- Programmability
- ?Schemas

How are nosql databases typically used?

- As an adjunct to Hadoop
- As a partial replacement for some RDBMS workloads
- To scale linearly
- As a data store for semi-structured and multi-structured data

Enterprise Architecture

EA - TOGAF



- Goals
- Objectives
- Strategy
- Capabilities
- Assessment
  - Current State
  - Future State
  - Transition
- Gaps
- Challenges
- Issues

Fixtures & Architecture

- Definition
- Examples
  - Oracle DB
  - Oracle EBS
  - ELA
- Architecture
  - Modular
  - Solution
  - Reference
    - Guidelines
    - Engineered Solutions
    - Blueprints

Solution Architecture

Data Source | Ingest, Cleanse, Normalize | Iteration Step | Analytical Execution | End-Point | Query | Reports/ Visualization

Systems of Record

Systems of Engagement

Structured

ERP

CRM

Finance

PoS

Patient Records

Docs

Email

Web

Social

Images

Video

Unstructured

ETL

Indexing

PIG

Sqoop

Flume

RDBMS

RDBMS

MPP

MPP

Text Search

NoSQL

NoSQL

Hadoop

HDFS Hbase

HDFS Hbase

Data Marts

Business Intelligence, Reporting

Search, Find

Analytics, Discovery

Business Reporting

Ad-Hoc Analysis

Search Queries (Research, Marketing)

Natural Language Search

Standard Reports

Ad-Hoc Reports

Query Drill-Down

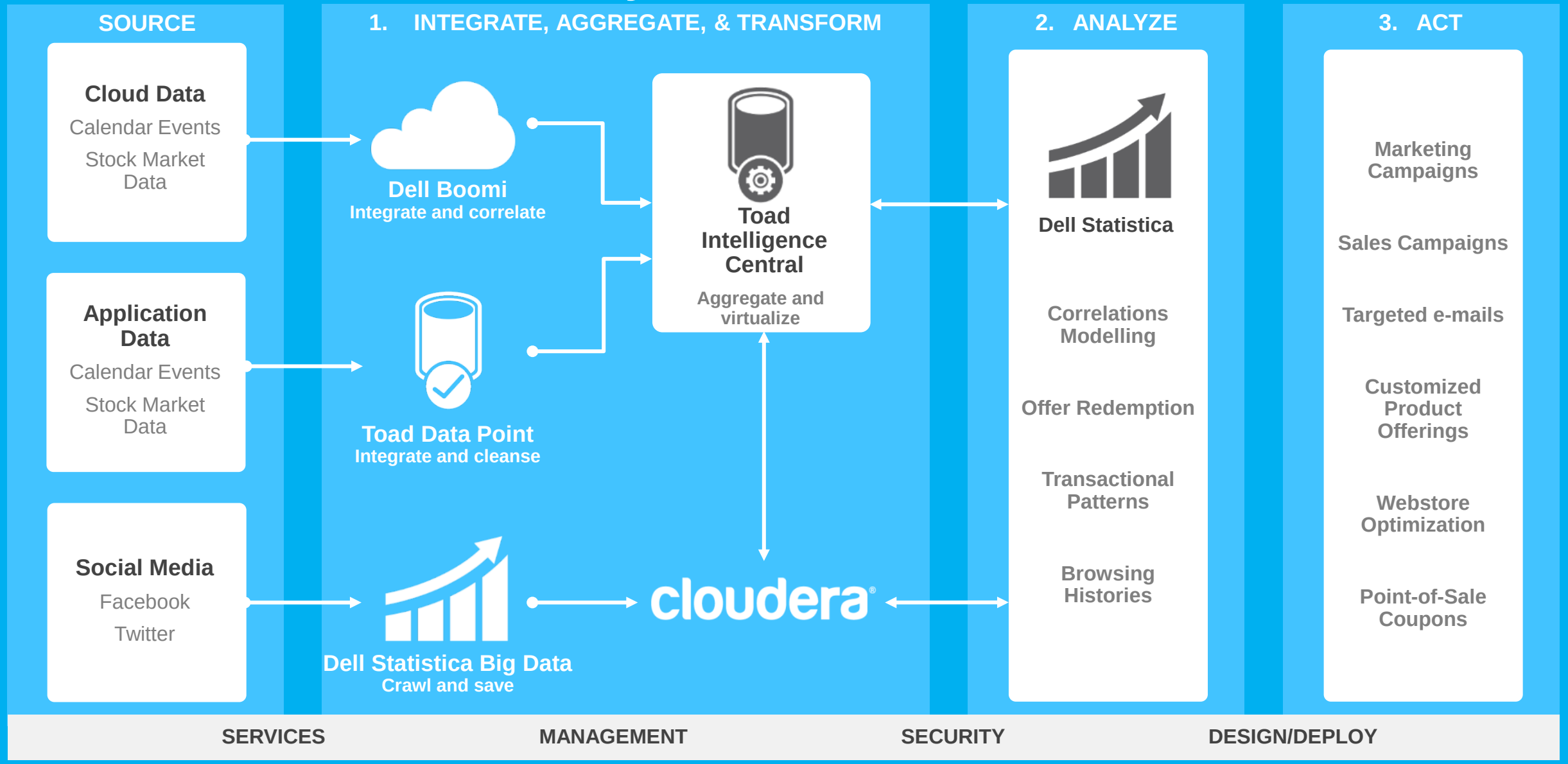
Statistical Analysis

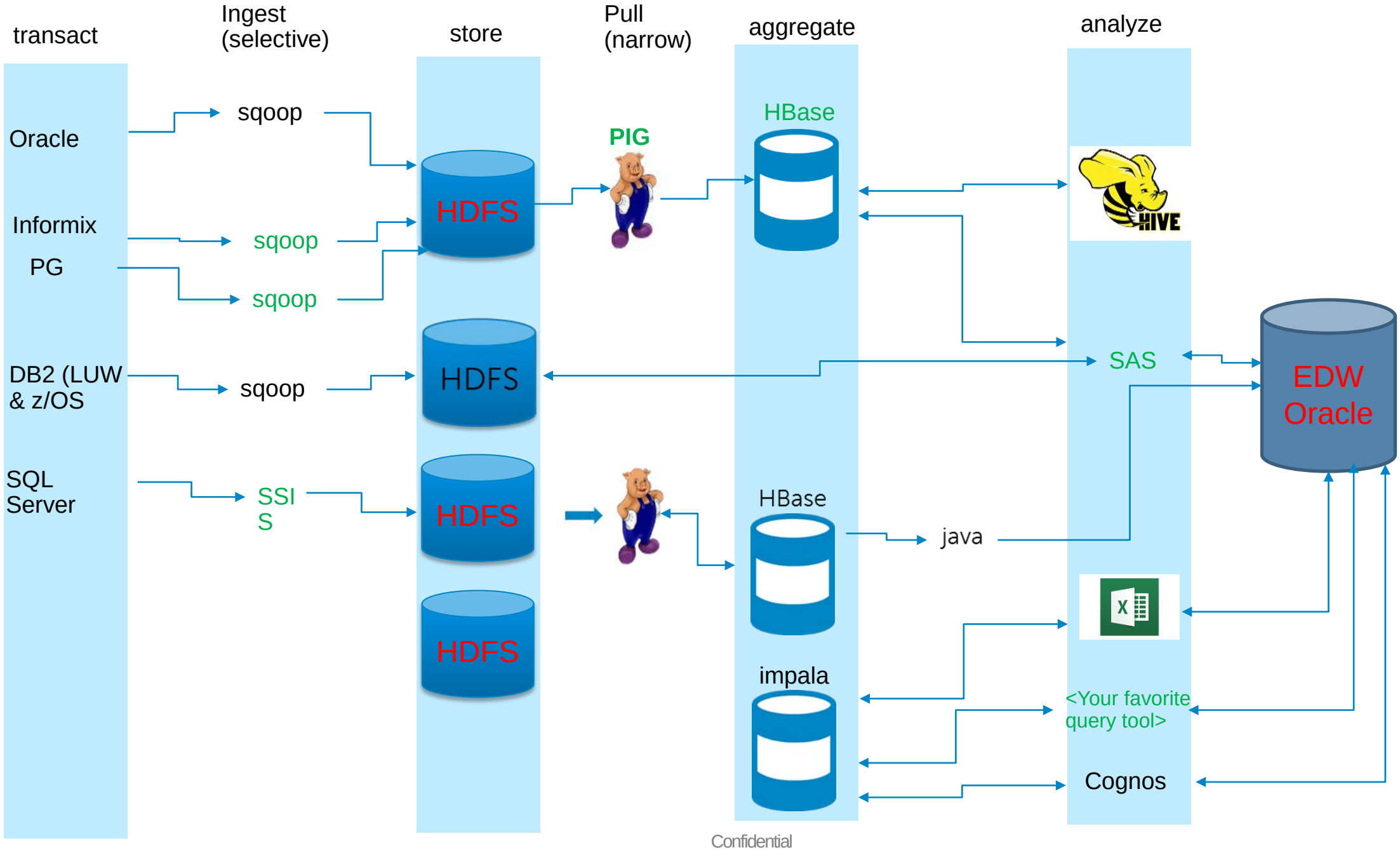
Forecast and Predictive

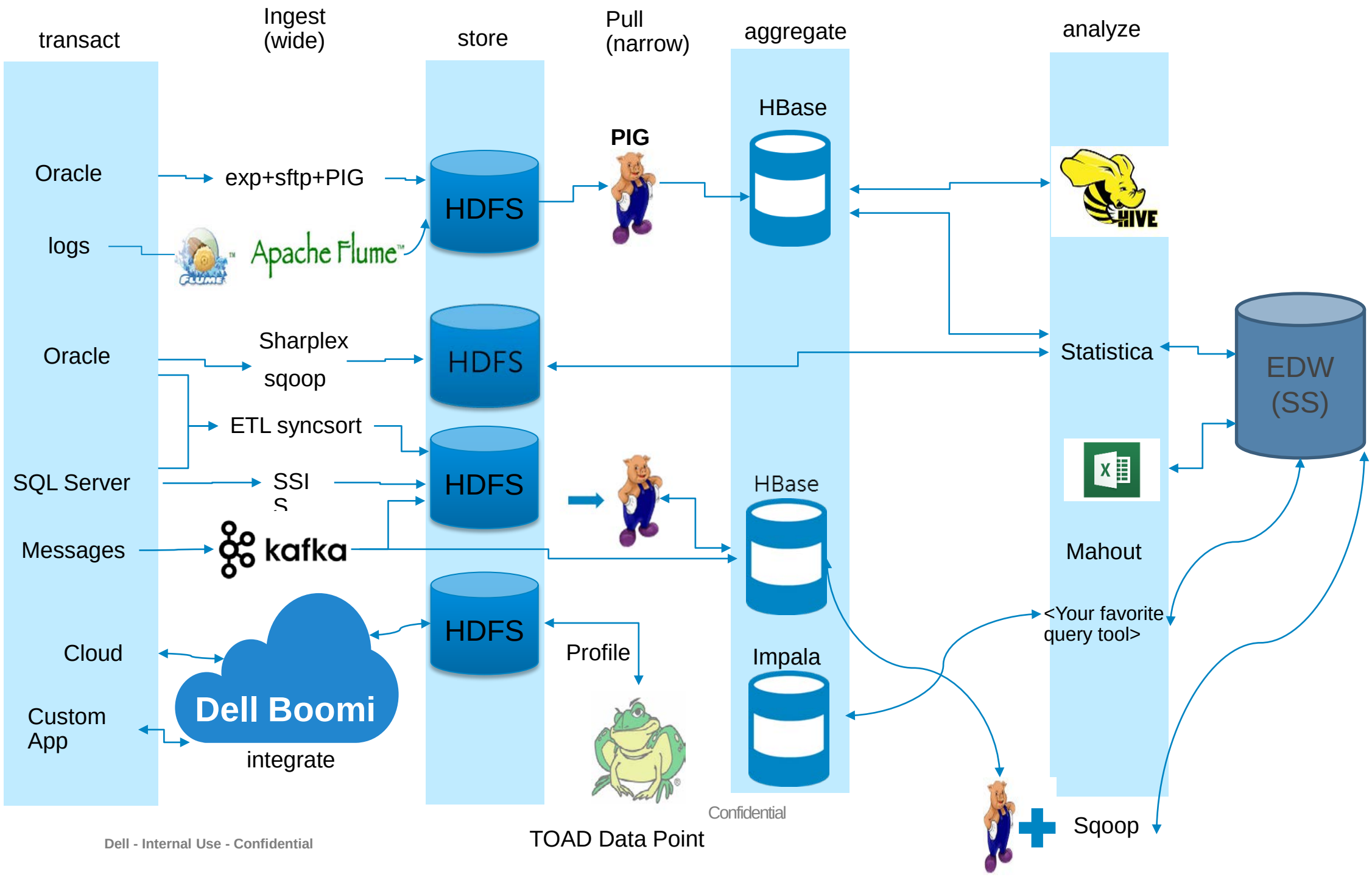
Optimize

Advanced Analytics

Customer Churn Analysis







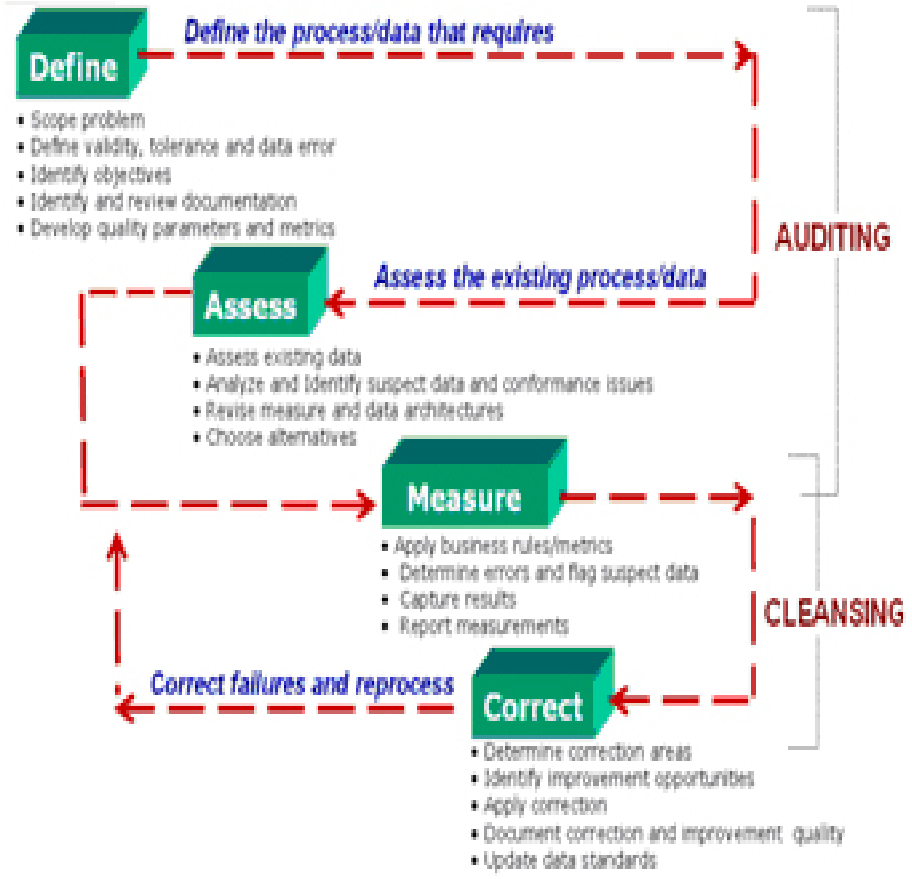
Issues of Interest for Public Sector

Issues

- Data Quality
- Knowledge Management
- Control
- Administration
- Governance
- Multi-tenancy
- Compliance
- Legislation
- Security
- Open data
- Customer Service

Data Quality, KM, Sharing

Data Cleansing



Data Quality

- Accuracy
  - Trust
- Completeness
  - Do you have all the pieces?
- Conformity
  - Dimension
    - Da, We, Mo
- Consistency
  - Country Codes (2,3)
- Duplication
  - Pervasive
  - Controlled
- Integrity
  - Think RI
- Timeliness
  - Currency
  - Aging
- Value
  - Varies
  - Radioactivity

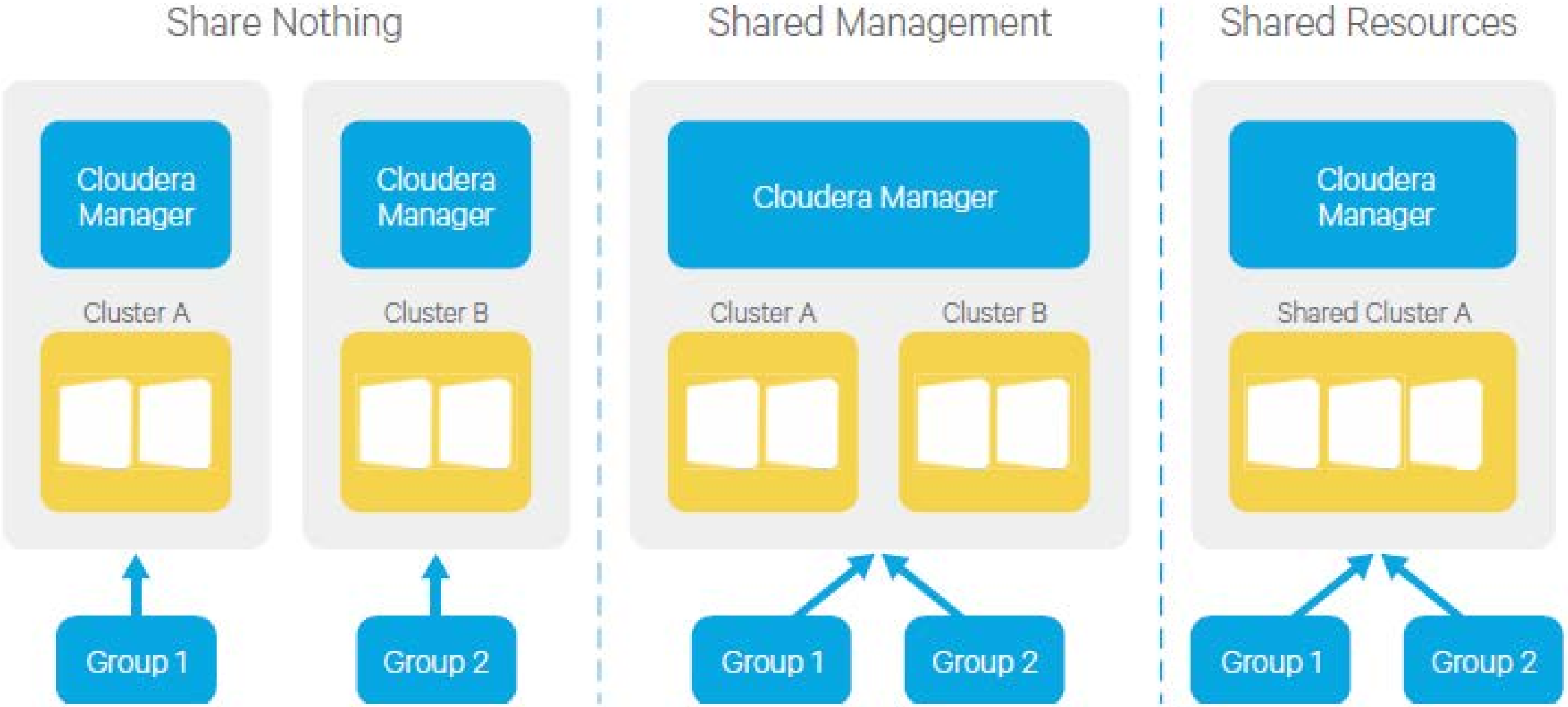
Figure 1. The numerous techniques of data quality

TOAD Data Point

- Query tool
  - DB sources
  - Non DB sources
    - Nosql, SFDC, OBIEE BO, etc..
  - Cross platform queries
- Analysis
  - Data quality
  - Data profiling
- Integration
  - Disparate sources

TOAD Intelligence Central

- Server based solution
  - Central repository
- Set of reporting tools
  - Publish & share reports
- Integration
  - Collects data from TOAD Data Point
  - Connect to Statistica
  - Utilize Boomi
- Centralized management
  - Share queries
  - Governance
  - Security
  - Automate



Boomi lite

The rapid adoption rate of SaaS…

- SaaS market is forecasted to grow at a CAGR of 20.2 percent from 2011 through 2017.
- Annualized SaaS end-user spending will grow from a base of $14.4 billion in 2011 to $45.6 billion in 2017.

NETSUITE

Taleo

SuccessFactors
Business Execution Software

Intacct.

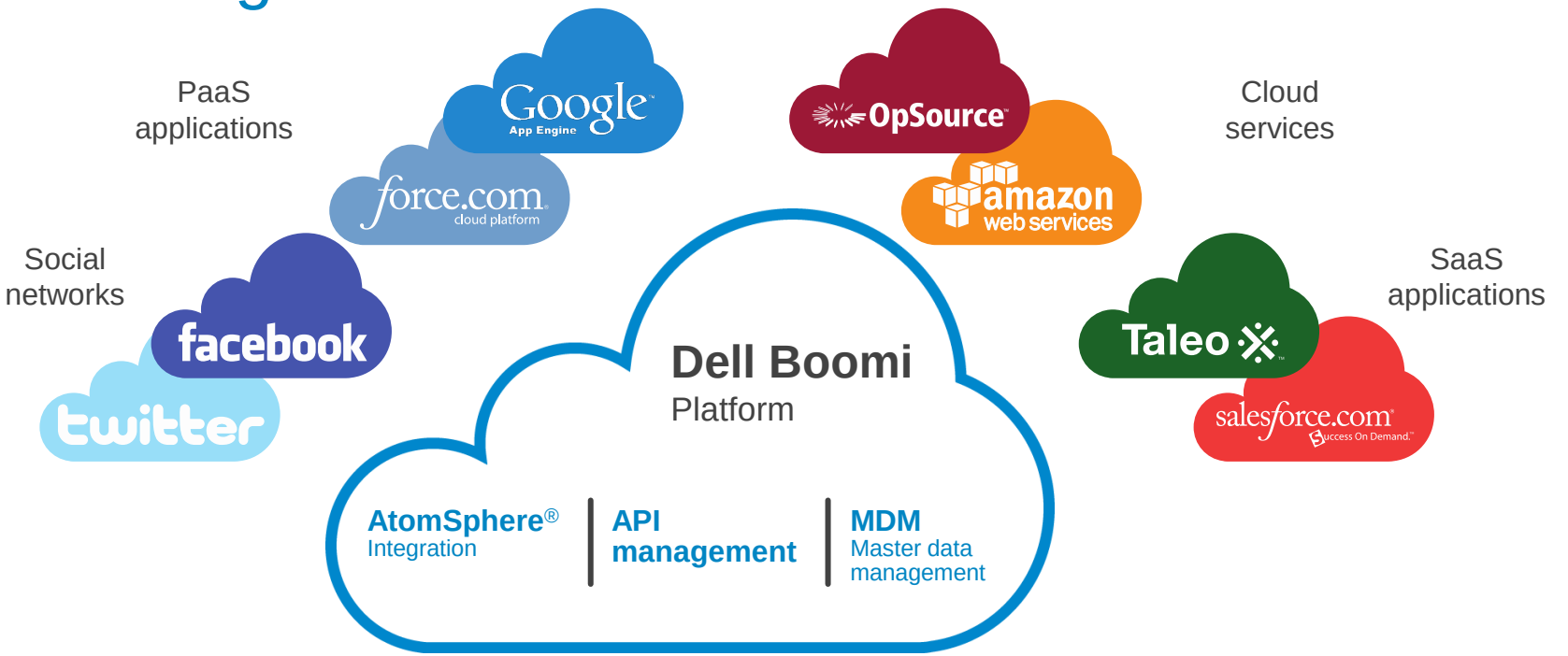
zuora

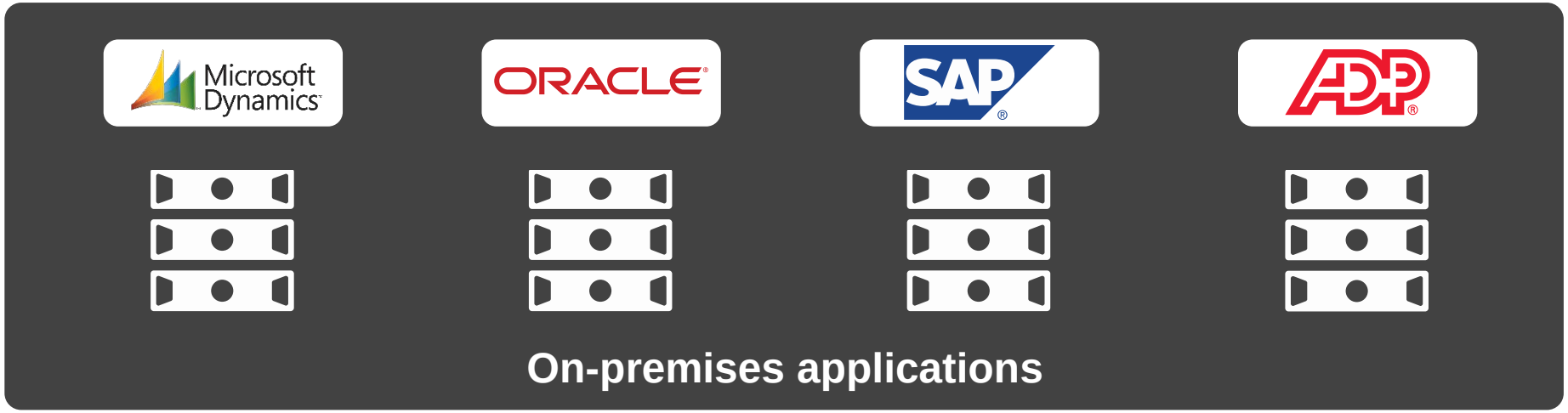
Marketo

Source: Gartner, Forecast: Public Cloud Services, Worldwide, 2011-2017, 4Q13 Update, January 2014

DELL Boomi

…providing cloud and on-premises data management…





Dell Boomi a Leader in Gartner Magic Quadrant for Enterprise Integration Platform as a Service

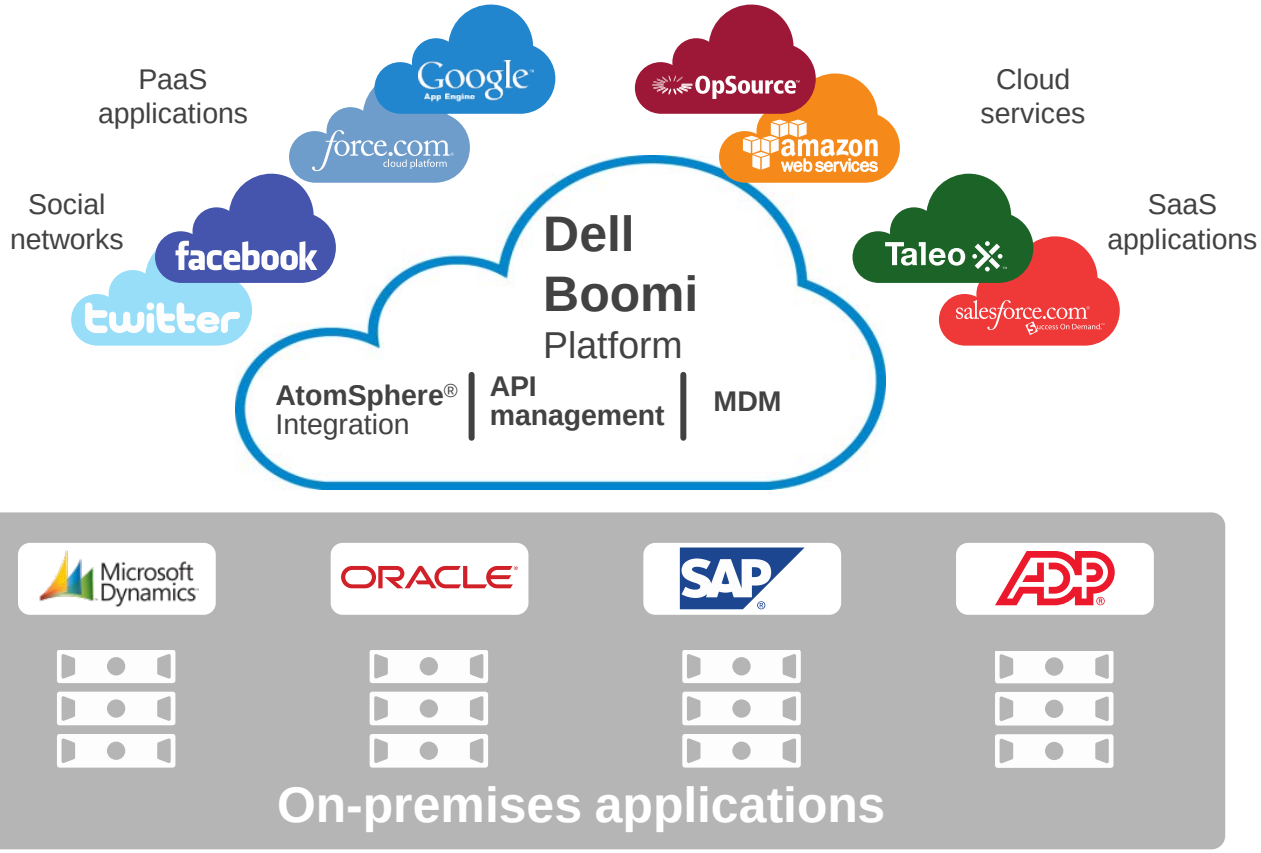
Source: Gartner Magic Quadrant for Enterprise Integration Platform as a Service, January 2014

This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from Dell Boomi.

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

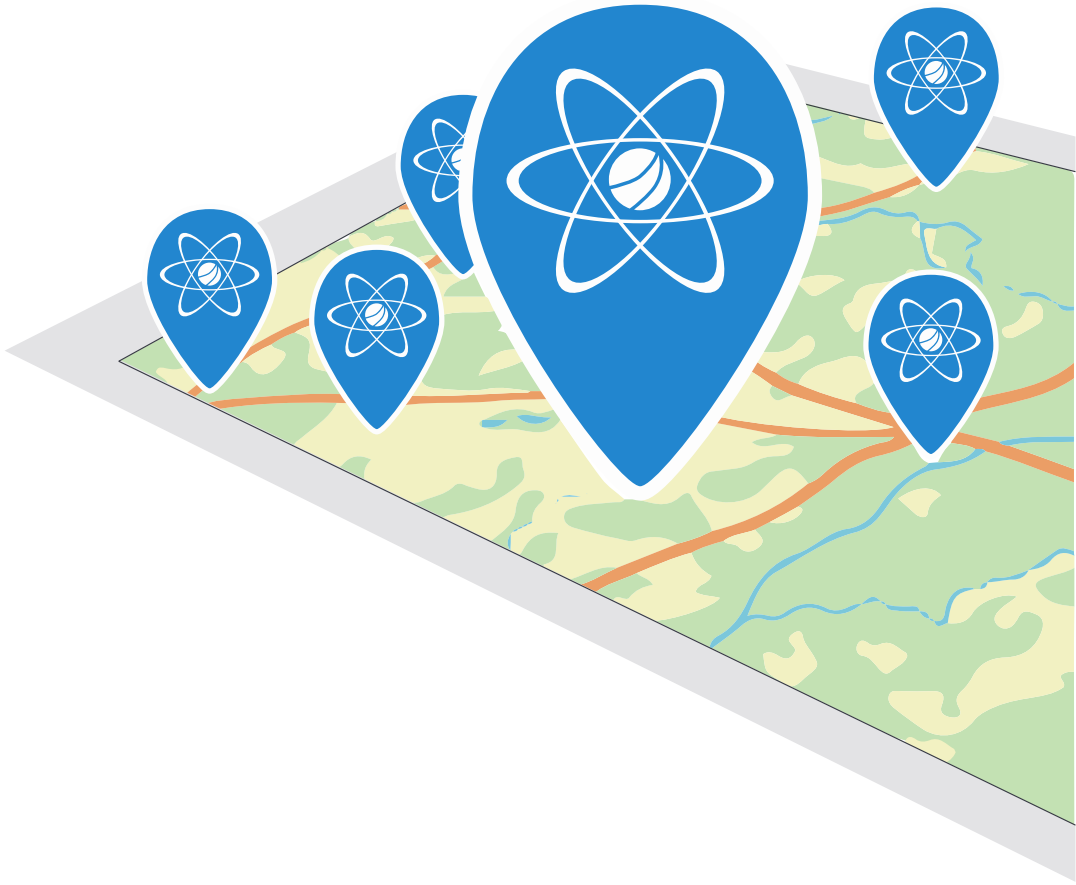
Boomi

- iPaaS Cloud offering
  - Integration
    - Public ↔ Private
    - Public ↔ Public
    - Private ↔ Private
- Can tie together the likes of
  - SalesForce (hosted, multi-tenant)
  - Oracle EBS (hosted, private)
  - Twitter (public)
  - Taleo (hosted)
  - Custom solutions (private)
  - <Your app goes here>



Boomi’s Cloud Benefits

- No hardware/software to install or maintain
- Automatic upgrades
- Usage-based pricing
- One platform for companies of all sizes
- Fully functional trial with on-demand access
- Multi-tenant architecture
- Enterprise scalability and elasticity



…and challenging new integration requirements

Secure data transfer outside your firewall

Connectors (adapters) for public cloud applications

Faster deployment of new integrations

Better integration economics to support endpoint growth

These requirements are not addressed by traditional on-premises middleware

Statistica

About Statistica

- Enterprise software for **advanced analytics**
- Part of Dell's modular, end-to-end **Big Data Platform**
- Enables you to **embed analytics** in real-time business processes
- Combines modeling & business rules into a real-time decisioning platform
- Draws insights from virtually **any type of data** (structured & unstructured)
- Interfaces with over **160 types of data repositories**
  - relational databases, data warehouses, Hadoop, cloud, applications, and more …
- In use since 1984 ... over 1M users worldwide … 16,000 functions
- Built to open standards … runs natively in **Hadoop** … R-friendly
- Provides natural language processing and advanced visualization tools
- Sweet spot: **predictive & prescriptive analytics**
  - uses information on what happened & why to addresses what'll happen next & what to do about it

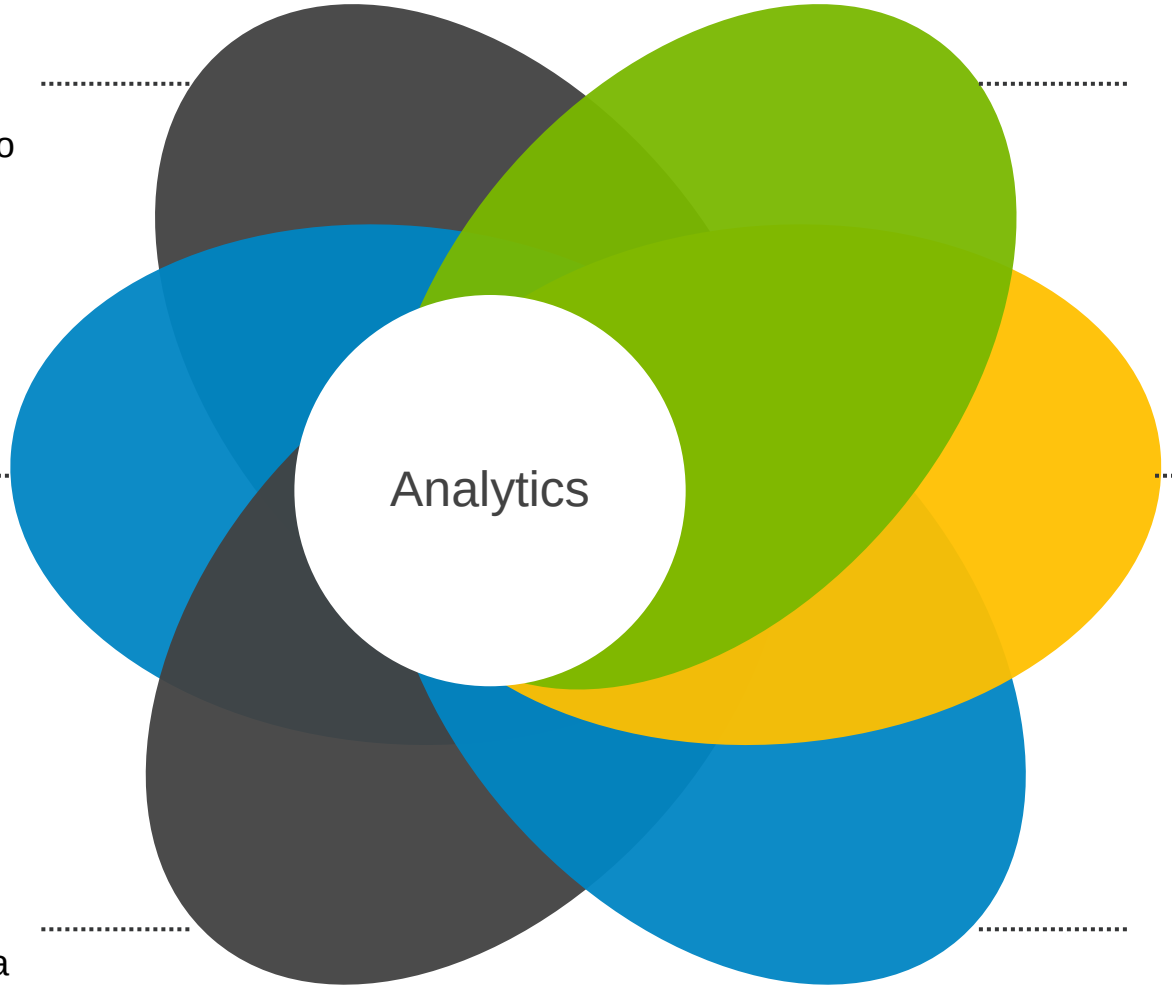
About Statistica…continued

- **Rexer Survey** **Highest rating in customer satisfaction**
  - Highest likelihood of continued use
  - #1 is overall tool satisfaction
- **Forrester** **Goes deep on algorithms**
  - Comprehensive library of algorithms
  - Very strong use cases
- **Hurwitz Victory Index** **Highest mark for value compared to price**
  - Breadth & depth of functionality
  - Easy to use & integrated
  - Open standards & integration
- **Garter Magic Quadrant**
  - **One of the highest evaluations for reliability**
- Wide range of functionality
- Speedy model development
- Support for wide variety of data types
- **Industry Analysts Love Statistica**
- **Dell's Big Data Platform**
  - ***Comprehensive***
  - ***Easy to use***
  - ***Flexible***
  - ***Affordable***
  - **Typical Use Cases**

Statistica Analytic Techniques

Clustering & segmentation
Grouping and dividing objects so like objects are similar to each other

Text Analytics
Statistical, linguistic, and machine learning to turn text into numbers



Decision Trees
Map every conceivable outcome to every decision

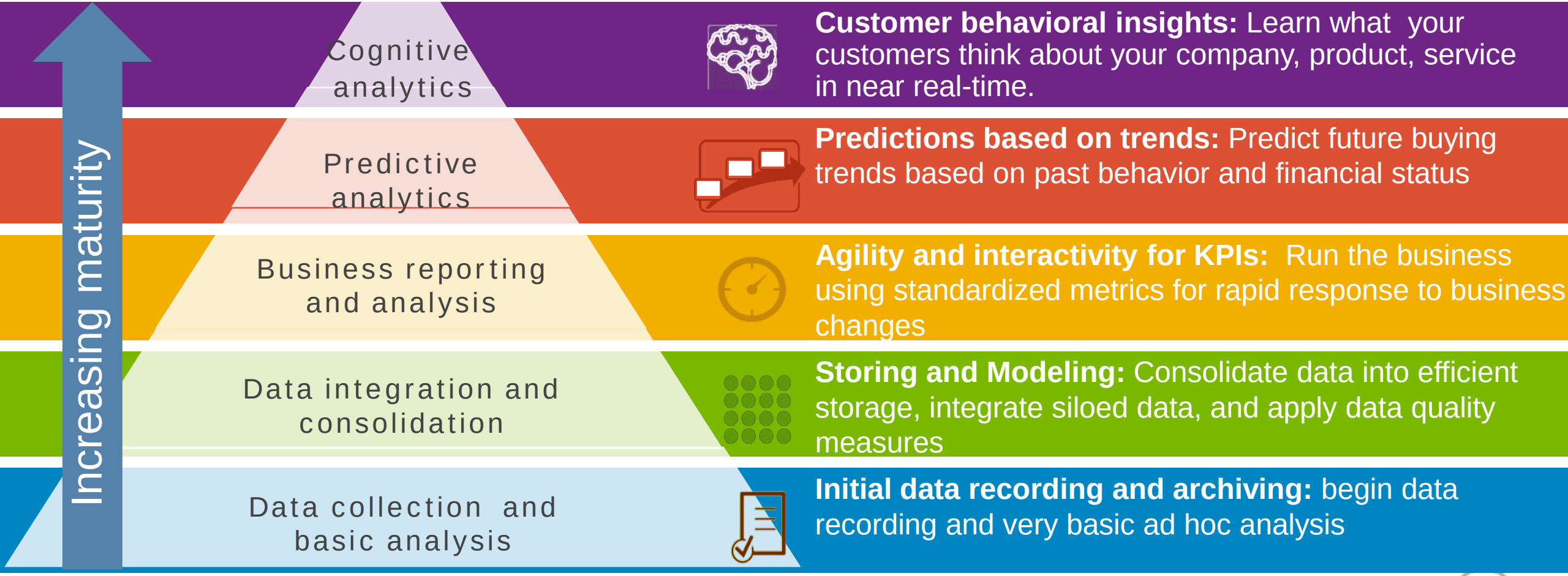
Optimization & simulation
Mathematically determining the best possible outcome given all the possibilities and constraints

Predictive Models & Forecasting
Using current and historical data to predict the future

Machine Learning
Getting computers to act without being explicitly programmed to do so

Next step: turn data into **insights**

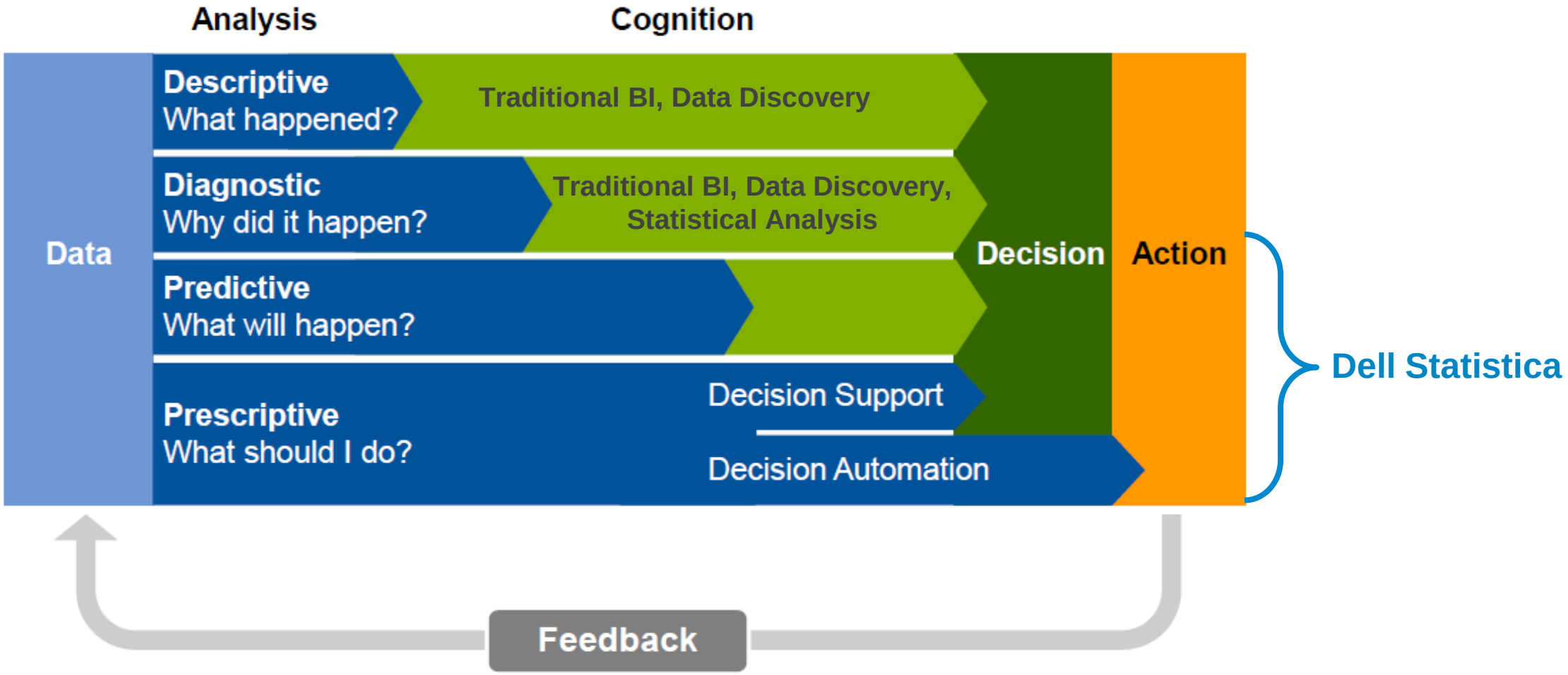
Fundamentals must be in place before achieving high level analytics



Providing data driven insights across **multiple verticals and use cases**

Marketing	Finance	Healthcare	Pharmaceutical	Manufacturing
Anticipate needs and personalize offers	**Reduce risks and detect fraud**	**Improve quality of care and efficiency**	**Ensure safety and product quality**	**Optimize processes, improve quality, monitor suppliers**
• Customer insight • Customer churn & retention • Market basket analysis • Media mix optimization • Price optimization • and more	• Credit scoring • Customer analytics • Fraud detection • Risk management • Churn analysis • SOX • Scorecard • and more	• Fraud detection • Claims management • Patient safety • Risk mitigation • Personalized medicine • and more	• Product traceability • Stability & shelf life Analysis • Validated reporting & analytics • Compliance • Manufacturing analytics • and more	• Improve yields • Reduce scrap, rework, & recalls • Detect warranty fraud • Regulatory compliance & safety • Predict & equipment failures • and more

Where is Statistica Positioned in the Market?



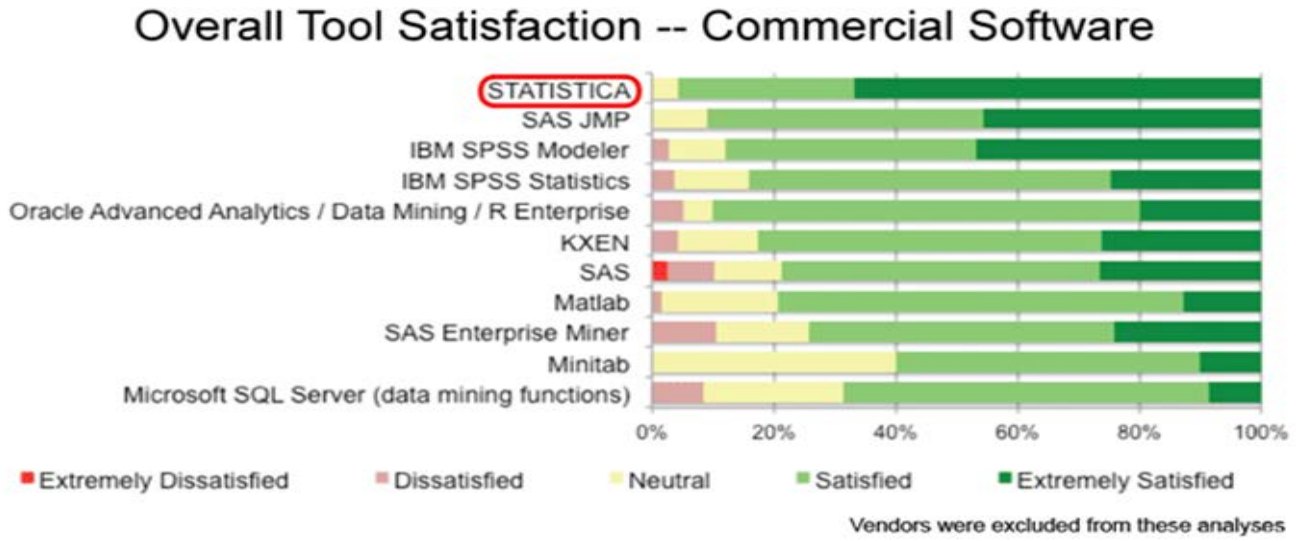
Gartner

Statistica's Gartner Magic Quadrant – Feb 2016

Source: Gartner (February 2016)

- **Dell has executed on an ambitious roadmap** during the past year: increasing the already broad functionality of Statistica, **updating the UI** and making it even **more intuitive for citizen data scientists.** It has also completed the integration of Kitenga into Statistica (enhancing its text analytics) and **has embedded an interactive visualization engine for line-of-business users**.
- Dell **addresses among the broadest set of use cases for advanced analytics**, including a new strategic focus on Internet of Things (IoT) use cases, and allowing edge deployment of analytic models on gateways (via native distributed analytics) or anywhere (via Dell Boomi).
- Dell has implemented in-database and in-Hadoop functionality — for data preparation. analytic model building and scoring — **to help reduce bottlenecks in performance.**

Why Dell Statistica



Dell Statistica (Previously StatSoft)

Strengths According to Gartner that Impacted Medtronic Selection:

- Highest rating for product reliability and upgrade experience of any vendor
- StatSoft was most frequently selected based on speed of model development/ability to build large numbers of models
- Ability to support a wide variety of data types — including unstructured data.
- Customer references cite high levels of satisfaction with the advanced descriptive analytics, predictive analytics, further advanced analytics, and performance and scalability components of the product.
- License cost

Gartner Magic Quadrant for Advanced Analytics Platforms

Dell Joins the Leaders Quadrant!!!

Figure 1. Magic Quadrant for Advanced Analytics Platforms

Source: Gartner (February 2016)

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Gartner Magic Quadrant for Advanced Analytics Platforms

Dell Recognized as a Leader!

Figure 1. Magic Quadrant for Advanced Analytics Platforms

Source: Gartner (February 2016)

Source: Gartner, Inc., Magic Quadrant for Advanced Analytics Platforms, Lisa Kart, Gareth Herschel, Alexander Linden, Jim Hare, 9 February 2016.

This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from Dell.

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Statistica

Appendix/ Supplemental Materials

SharePlex Connector for Hadoop

- Provides **near real-time** data replication from Oracle to Hadoop environments. Enables organizations to affordably replicate live data from Oracle tables
  - In near real time to Hive and HDFS
  - In real time to Hbase