



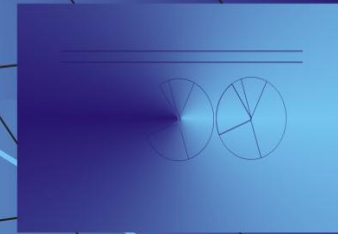
# Big Data Analytics

# Priority Discussion Topics



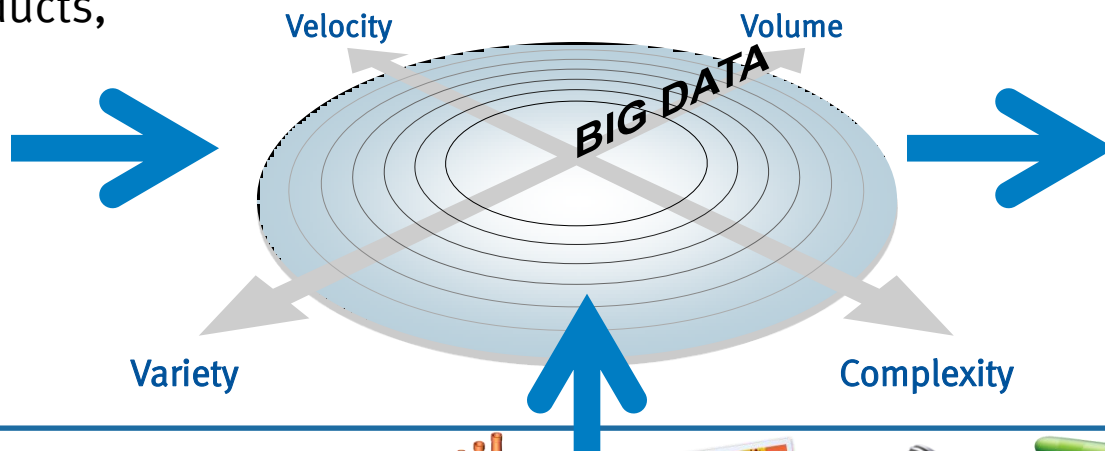
- What are the most compelling business drivers behind big data analytics?
- Do you have or expect to have data scientists on your staff, and what will be their charter?
- What are the different product, technology and architectural components that need to be considered?
- What process challenges for data collection, data cleansing and data quality concern you most?

# It's a Whole New Big Data World ...



# More than just data volume, big data analytics must also consider data velocity, variety, and complexity

New insights on customers, products, and operations



Contextual and location-aware delivery to any device



Documents



Transactional Data



Smart Grid



Images



Audio



Text



Video

- **Volume:** data volumes approaching multiple petabytes
- **Velocity:** data being generated and ingested for analysis in real-time
- **Variety:** tabular, documents, e-mail, metering, network, video, image, audio
- **Complexity:** different standards, domain rules, and storage formats per data type

# Big data analytics provides potential for more timely, complete, actionable business insights

*“Over the last 25 years, companies have been focused on leveraging maybe 5% of the information available to them... In order to compete well, companies are looking to dip into the rest of the 95% that can make them better than anyone else.”*

## Today's Situation

Less than 10% of available enterprise data

“Rearview mirror” reports, dashboards, and analysis

Weeks, months, or even quarters old

Incomplete, inaccurate, and disjointed data

Architectures and methods that take 6 to 18 months to exploit



## Big Data Analytics Ramifications

Vast majority of available data, including external sources

“Forward looking” predictions with recommendations

Real-time or near real-time

Correlated, high confidence, governed data

Vastly accelerated time to market



What are the most compelling business drivers behind big data analytics (i.e., what gets your business stakeholders excited)?



Do you have or expect to have data scientists on your staff? Will they be in the business or in IT? What will be their charter? How will you measure their effectiveness?

# Successful organizations continuously uncover and publish new insights about the business

Data scientist (GigaOM)

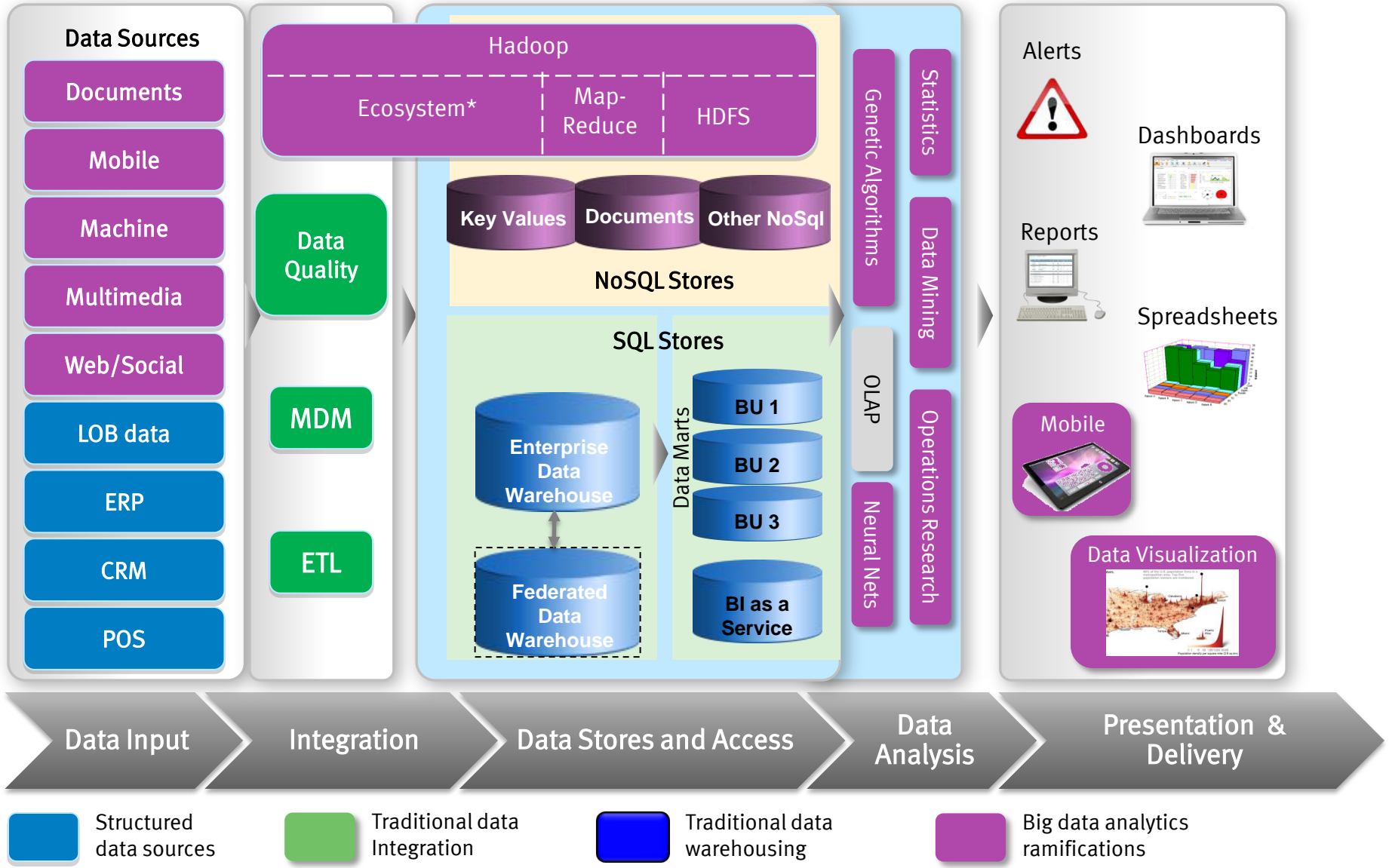
*Obtain, scrub, explore, model, and interpret data, blending hacking, statistics, and machine learning, with good understanding of the business processes and goals*





What are the different product, technology, and architectural components that need to be considered in a big data analytics project?

# EMC Big Data Analytics Reference Architecture



\*Hadoop Ecosystem includes: Hive, Pig, Mahout, HBase, ZooKeeper, Oozie, Sqoop, Avro





What process challenges for data collection, data cleansing, and data quality concern you most with respect to big data and advanced analytics?

# EMC IT use case of performance and security event management

*Data Volume, Velocity, Variety AND Complexity*

## Challenges

- High volume of event data
- Numerous data types across thousands of collection points
  - 12 MB/collection point per hour
- Information silo'ed and difficult to aggregate and correlate
- Manually-intensive ad-hoc analytics

## Approach

- Created fast aggregation capabilities with Hadoop and a single data framework with the Greenplum database
- Mapped GRC model to control management layer
- Leveraged modern, integrated and interrelated analytic tools for correlation of events
- Implemented real-time data loading and analysis at high frequency

## Benefits



Framework for single management of controls



Faster investigation of incidents



Automated and aggregated analysis



Security embedded in virtual infrastructure

THANK YOU