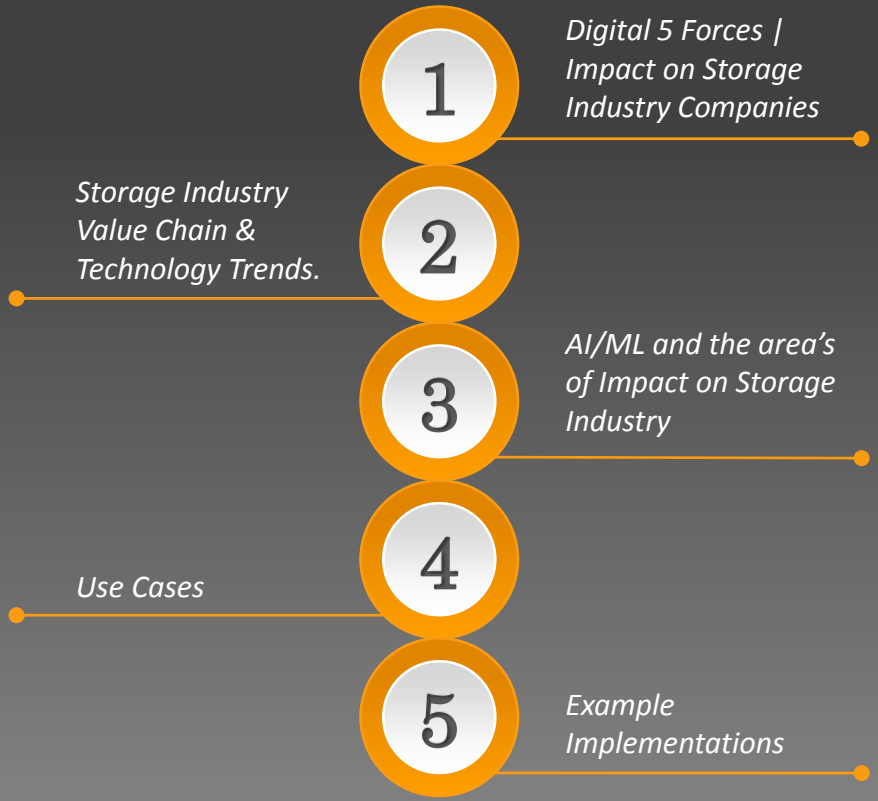
TATA CONSULTANCY SERVICES
Experience certainty.

Next Phase of Evolution in Storage Industry: Impact of Machine Learning

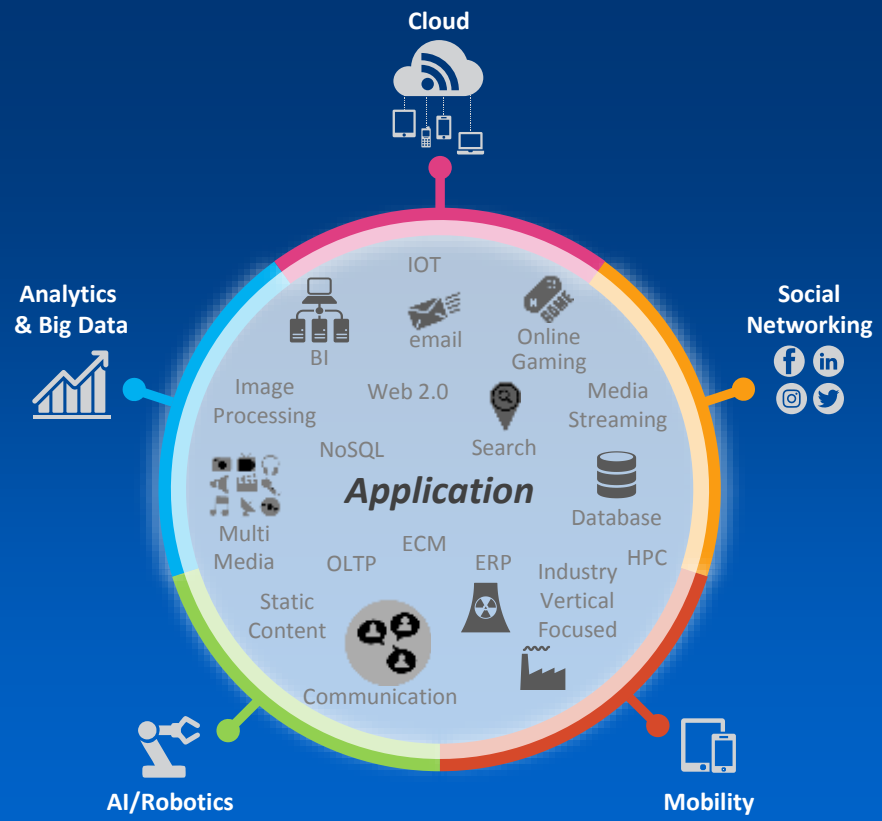
Udayan Singh, Head SPE-Storage, Compute & Manageability

30 May 2017

Agenda



Digital 5 Forces | Impact on Storage Industry



Business Models

- Pay as you grow cloud storage based models

Products and Services

- Prescriptive analytics-based issue identification and auto-resolution
- Next Gen Products: HyperScalers, Hyper Convergence, AFA, Cloud Storage

Customer Segments

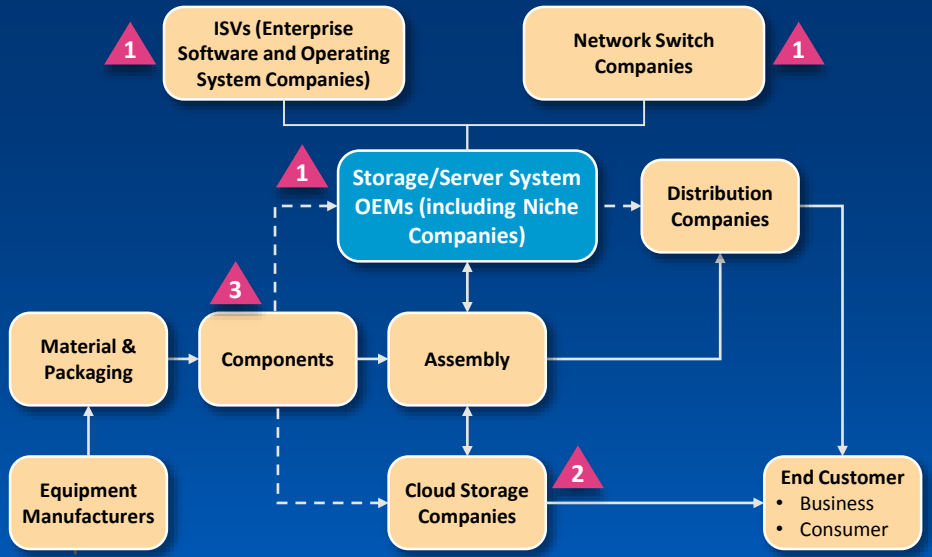
- Movement from only B2B to B2B and B2C

Partner Network

- Cloud Storage through partners
- Mobility connect

Digital 5 forces impacting multiple dimensions of the Storage Industry Companies

Storage Industry Value Chain and Storage Technology Trends



1. "Increased Collaboration" - ISVs, Switch and Storage OEMs
2. "Increased competition" - Cloud Storage Companies
3. "Consolidation" - SSD/HDD Companies
4. "Emergence" Niche Technology

Hyper Converged Infrastructure

- Delivering through simplicity

Cloud Storage

- Operational costs of maintaining high-growth internal storage infrastructures

Flash Storage (Hybrid Flash Array/All Flash Array)

- New interfaces - NVMe

Software Defined Data Center

- Simpler, intuitive and intelligent Data Center Management

Artificial Intelligence / Machine Learning

- Self-service infrastructure
- active cognition and analytics-based automation

AI/ML and the area's of Impact on Storage Industry

Artificial Intelligence

- **Robotics**
- **Sensory Perception**
- **Natural Language Processing**
- **Machine Learning**
- **Speech Recognition**
- **Image Analysis**
- **Natural Language Generation**
- **Deep Learning**
- **Knowledge Engineering**
- **Cognition**

Impact on Storage

PDLC: Requirements Gathering; Development; Quality Assurance; Sustenance

Time To Market

Cost Reduction

Customer Support Services

User Experience

Improved Quality

Storage Admin Software : Provisioning; Management; Troubleshooting

User Experience

Cost Savings

End User Storage Software Products: File Storage;

User Experience

Improved Quality

Storage Devices: SSD, HDD

Improved Quality (endurance, performance)

Use Cases: Data Center Storage Software

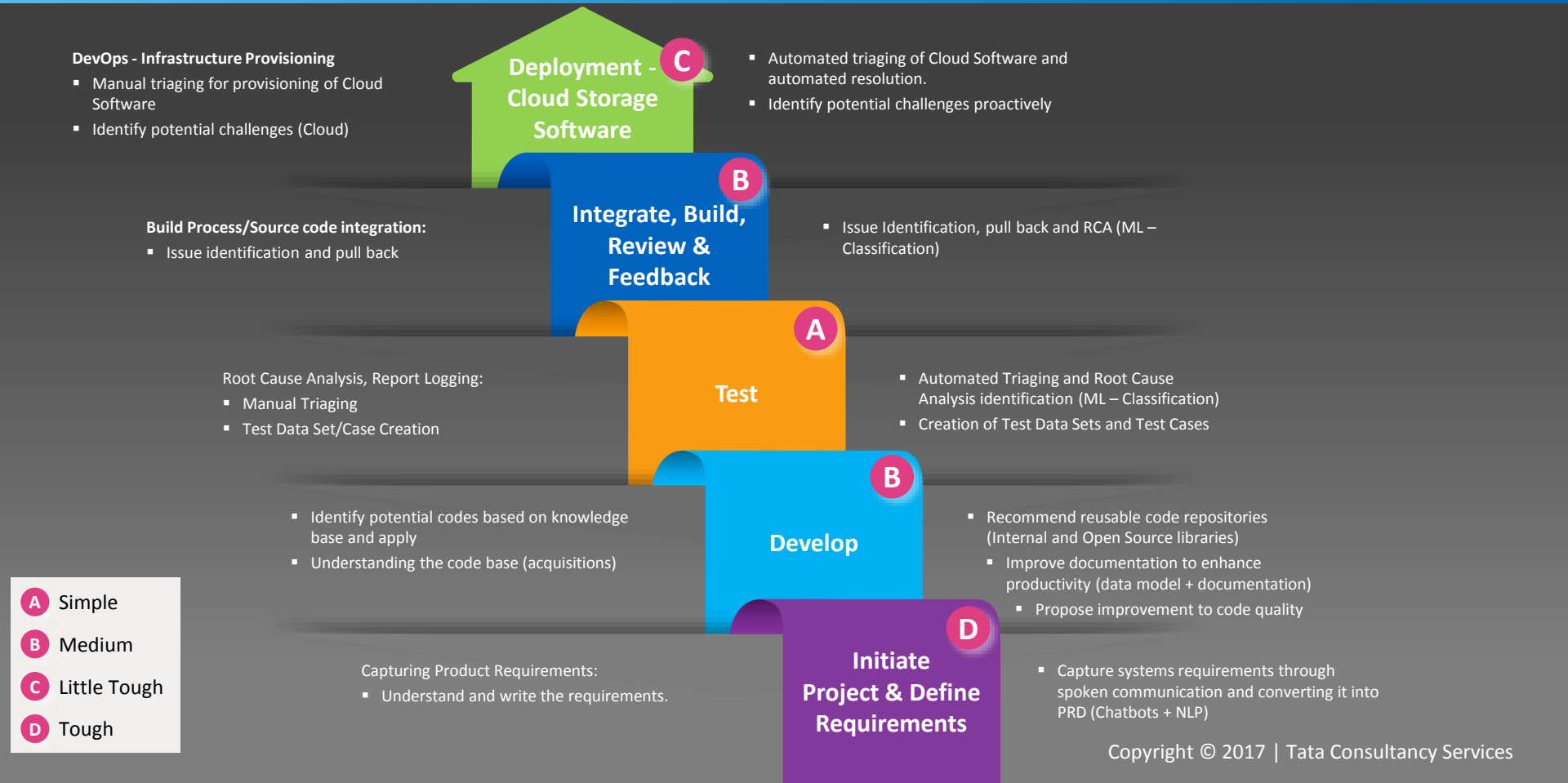
Machine Learning

- Cloud Storage Management
- Data Protection
- Storage Monitoring & Management
- Storage Analytics
- Customer Support Software
- Performance Management
- System Health Monitoring
- Storage Software Platform

Executing Task	Learning by experience	Improving Performance
Provision	Automated Update	Intelligent Proactive Error detection and Update
Monitoring	Manual Analysis	Automated Root Cause Analysis & Corrective Action
Manage	Manual Performance Tuning Automated Data Storage Optimization	Intelligent Performance Tuning Intelligent Data Storage Optimization
Issue Identification	Proactive Management	Self Managed Data Center - Customer Support - Storage Management Software
Issue Resolution	1 Admin per 4 Pb/500 Servers	1 Admin per 20 Pb/10,000 Servers
Optimization		
Demand Prediction		

Current Stage	Future Stage

Use Cases: Product Development Phases

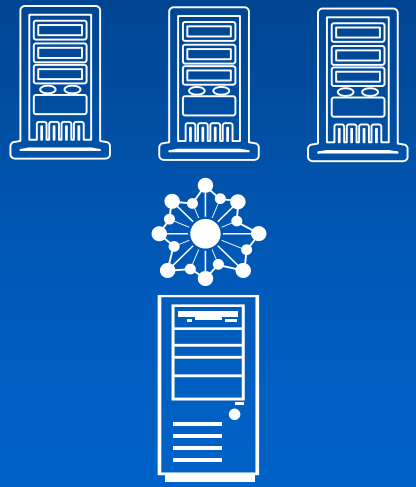


Example 1: Storage Performance

Machine Learning Method

Selection of Right Data Set → **Quality Check of Data** → **Analysis of the Data** → **Insight Generation**

Data



Pattern Matching

Models:
- MultiVariate Regression Analysis (Causal)
- Time Series Analysis
- **Final:** Multivariate Time Series (VAR Model)

Features:
- Total IOPS
- Cache Hit Ratio (%)
- Disk Response Time
- Disk Utilization

- [%age Write Miss
- Data Transfer Rate (MB/s)
- No .of Active Network Ports
- Spindle Count
- RAID Configuration
- Disk Type]

Prediction: Average Response Time (Latency)

Similarly for-
Capacity
CPU Failure
Disk Failure

Example 2: Prediction of Database Response Time

Business Problem

To predict the database response time and facing challenges as below

- To create predictive models which predict DB response time.
- To predict 1-hour-later SLA violation of DB response time with 90% of recall and 70% of precision.
- The predictive models predict the Probabilistic distribution of DB response time at the time of 1hour later.

Solution

- Creation of Lab environment to load on DB and measurement of various parameters for model building
- Predictive models to:
  - Predict number of queries on the DB after 1 hour
  - DB response time after 1 hour
- To identify load on the system after 1 hour and corrective action planning
- Techniques - Unobserved component models and Poisson Regression

Benefits

- All models are having Mean Absolute Percentage Error (MAPE) less than 5%
- Precision & recall rate for 1 hour later prediction was 60% and 45% respectively.

TATA CONSULTANCY SERVICES
Experience certainty.

TATA

Thank You

IT Services
Business Solutions
Consulting