Introduction to Big Data

ECAP456

Dr. Rajni Bhalla
Associate Professor

Learning Outcomes

After this lecture, you will be able to

- Learn what is big data
- Learn why should big data matter to you?
- Learn about Big data and the cloud
- Learn why big data in the cloud makes perfect sense
- Learn Big opportunities, big challenges

Introduction

Why Cloud Computing is the Answer to Your Big Data Initiatives

Introduction

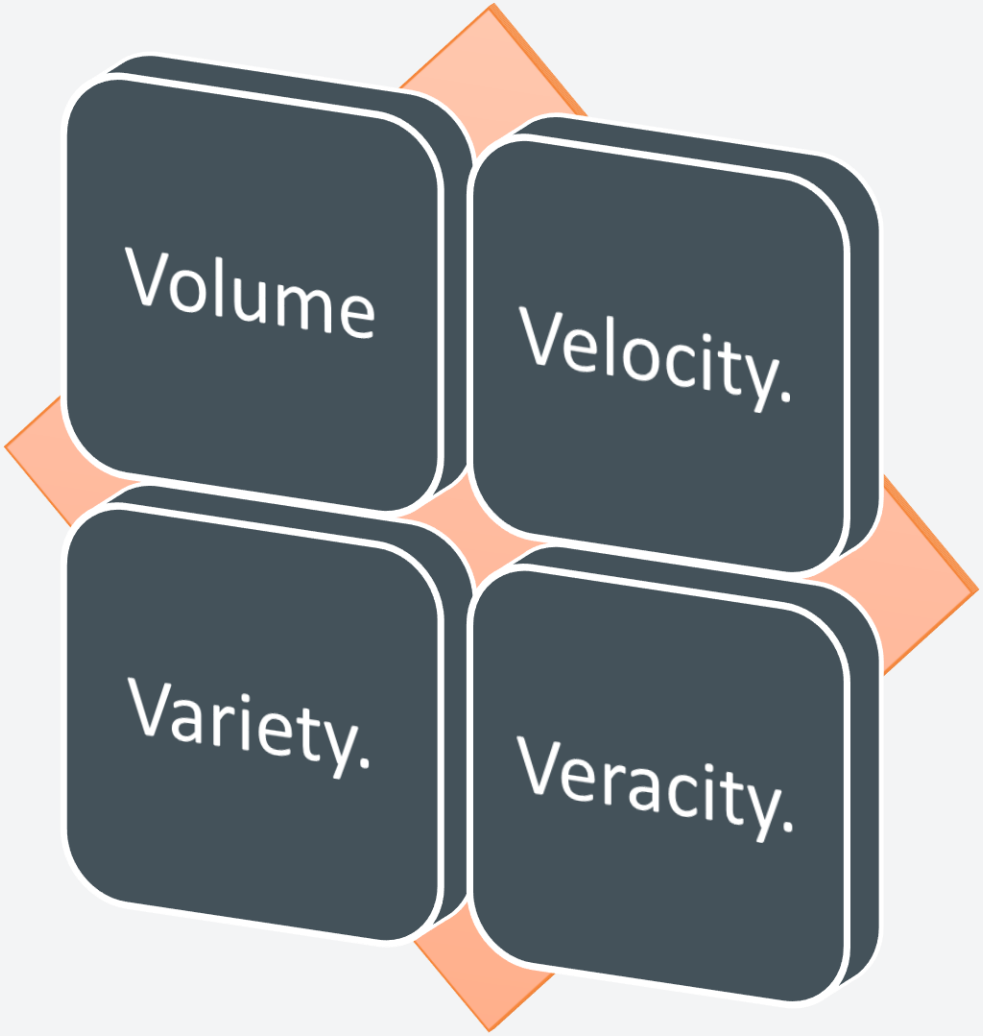
What are the benefits to companies from advancement of technology?

- Streamlined processes and cost-efficient operations
- Availability of data from every source imaginable - social media, sensors, business applications, and many more.
- Adopted cloud computing to improve their IT operations
- Merging big data with cloud computing

What is big data?

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.

The concept of big data and what it encompasses can be better understood with four Vs:



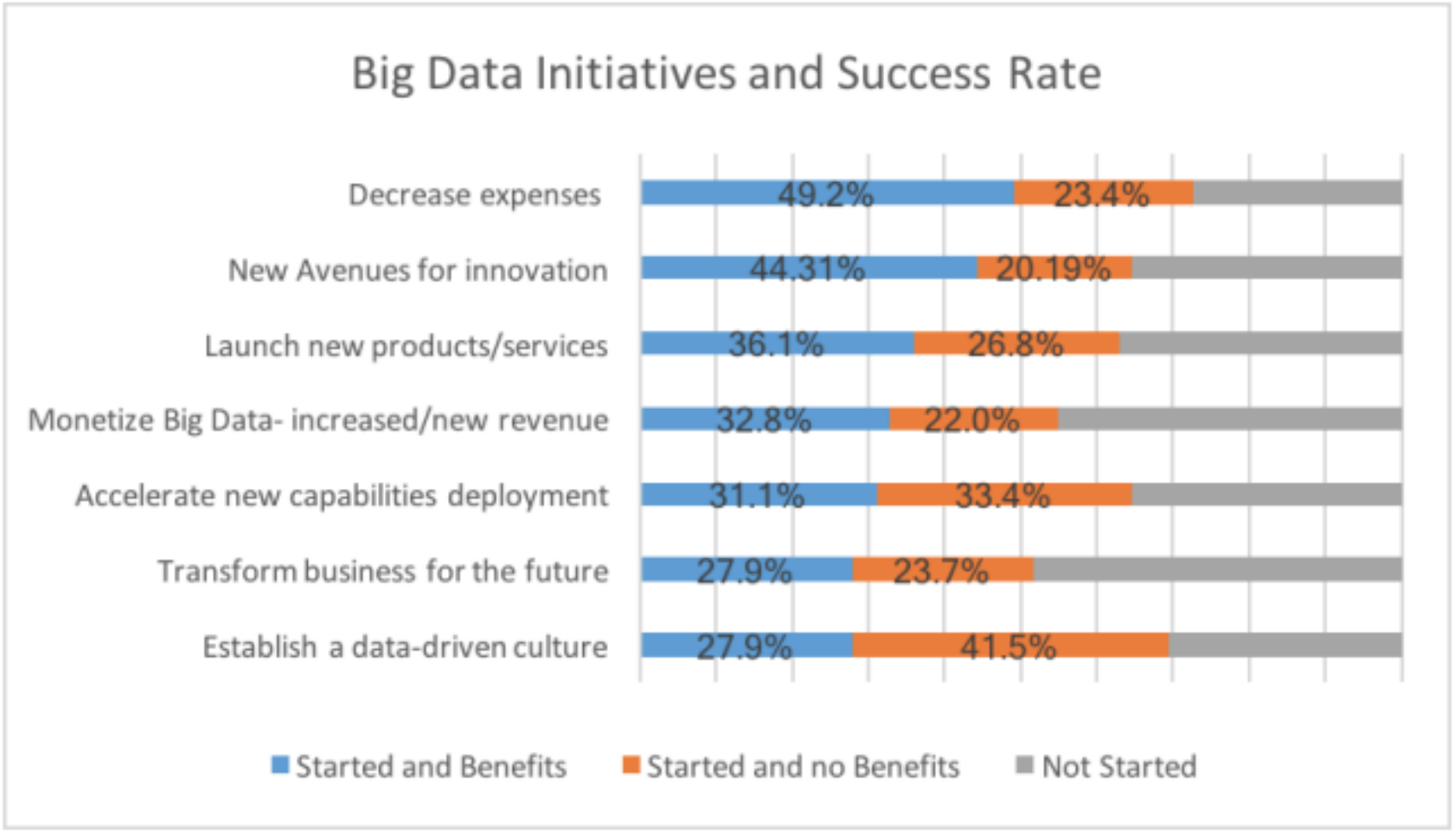
Technology leaders also name a fifth V

- V – value
- The true value of big data can only be realized when the right information is captured and analyzed to gain actionable insights.

Why should big data matter to you?

An Accenture study (PDF) reveals that 79 percent of corporate executives surveyed believe that 'companies that do not embrace big data will lose their competitive position and may even face extinction'.

Why should big data matter to you?



Big data and the cloud

When your big data goes to the cloud?

- Data storage
- Application of basic analytics modules
- Extract data at a much larger scale
- Add more capacity to your in-house data warehouse or power up more servers

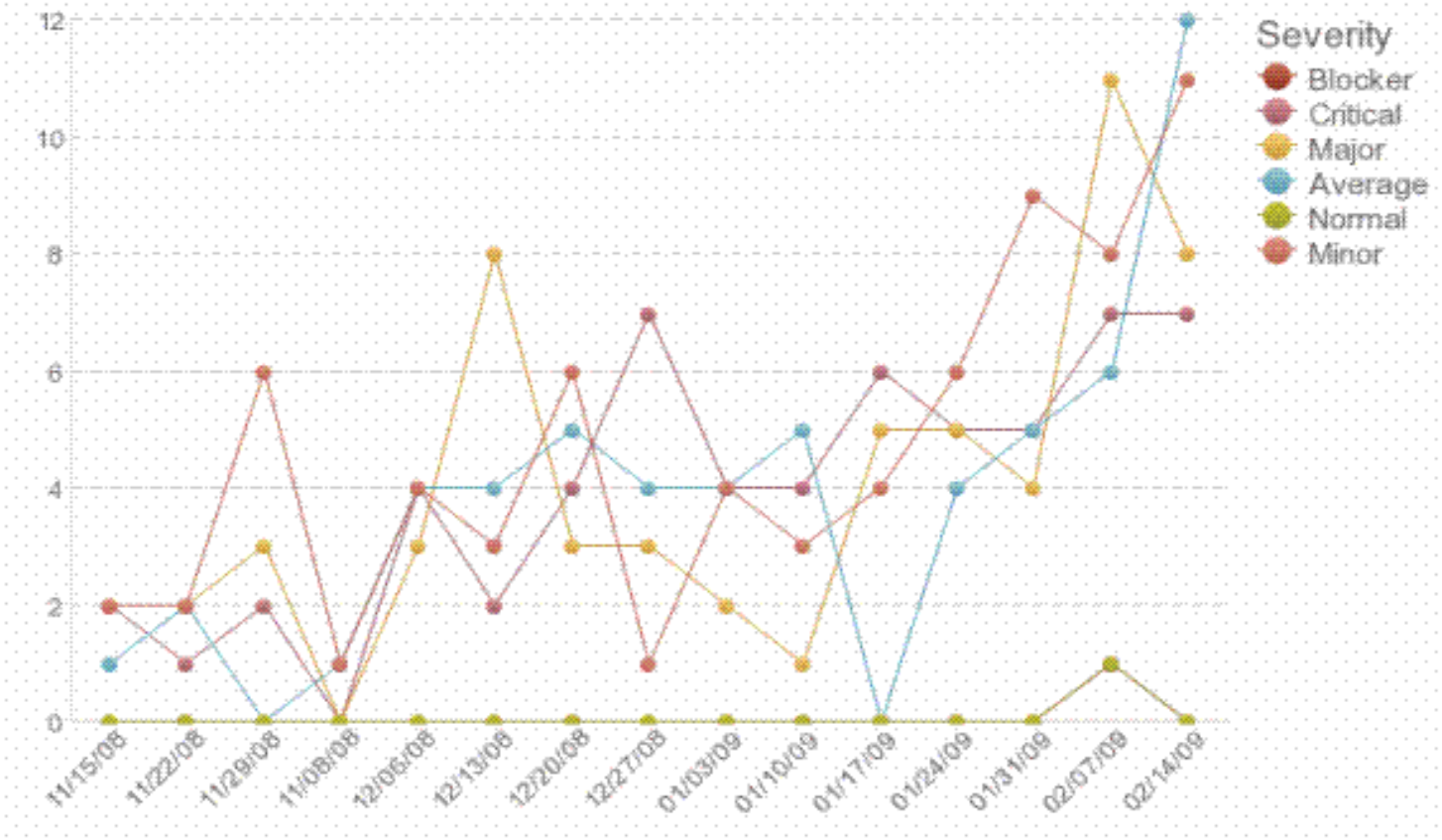
Why big data in the cloud makes perfect sense

- The cloud's scalable environment makes it possible to deploy data-intensive applications that power business analytics.
- Simplifies connectivity and collaboration within an organization
- It gives executives a better view of the business and boosts data-driven decision making.

Why big data in the cloud makes perfect sense

Supply Chain

Why big data in the cloud makes perfect sense



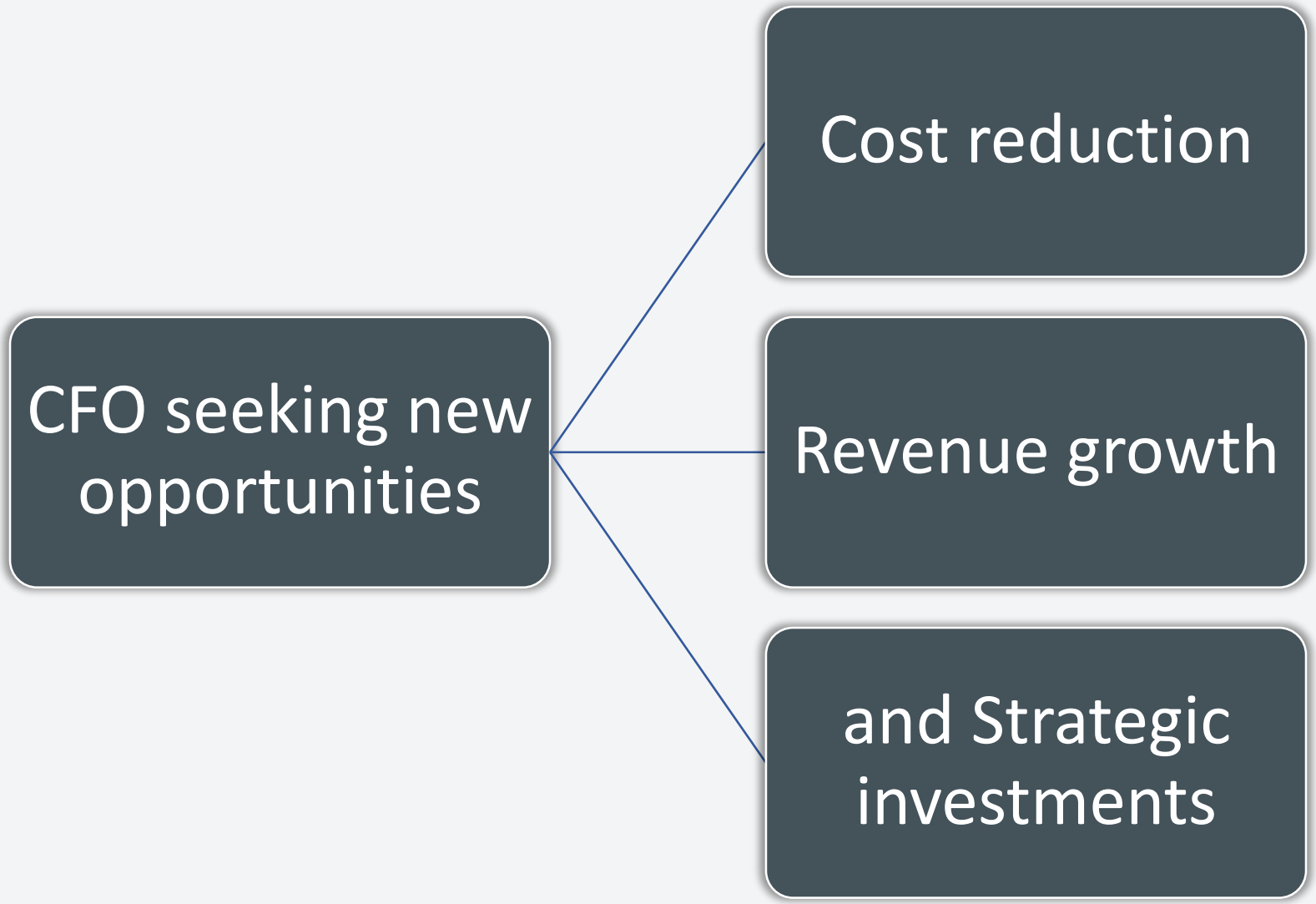
Efficient tracking of defects

Why big data in the cloud makes perfect sense

Increase customer engagement and loyalty

Why big data in the cloud makes perfect sense

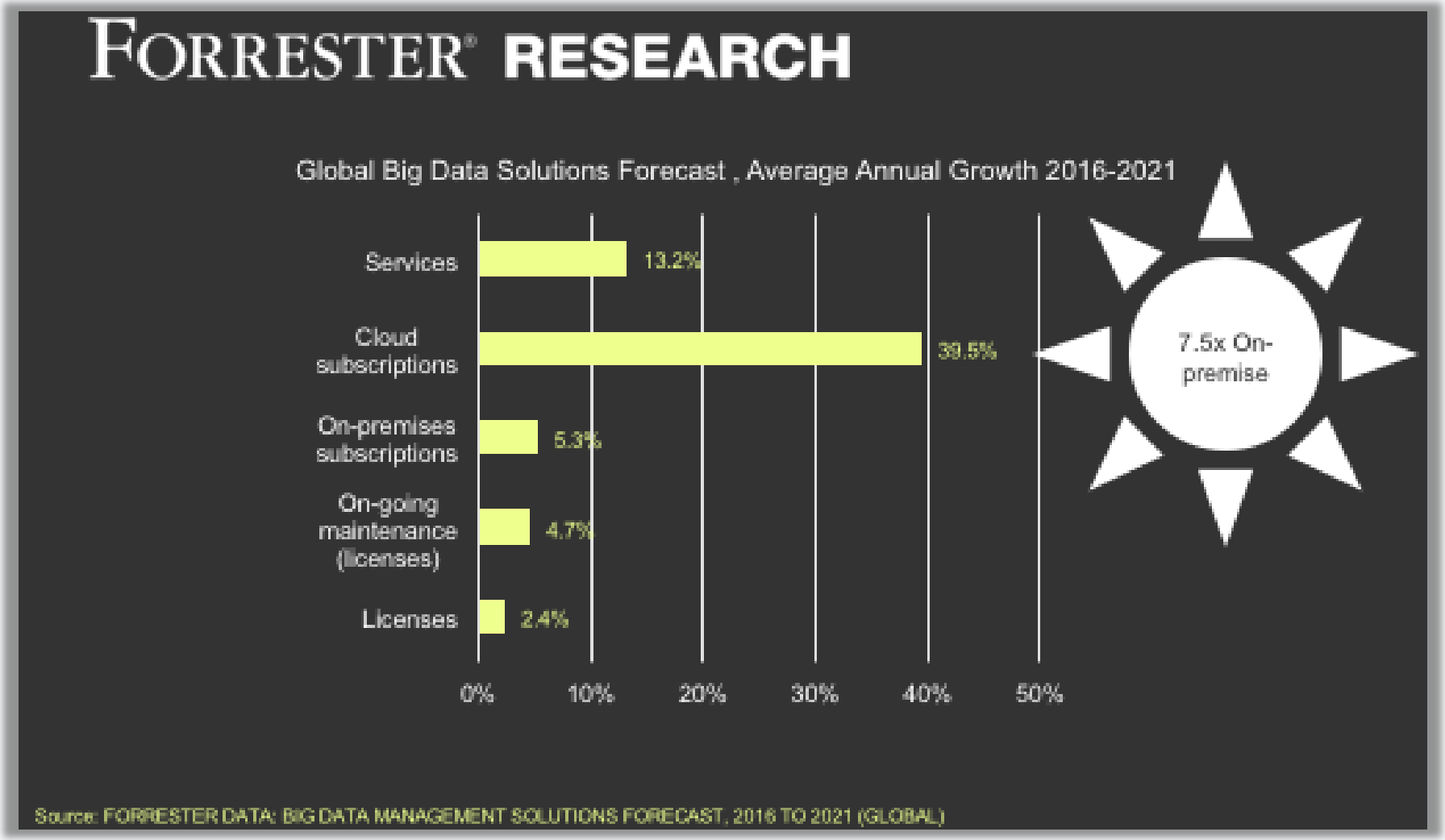
Why big data in the cloud makes perfect sense



Why big data in the cloud makes perfect sense

Big data complemented with an agile cloud platform

Why big data in the cloud makes perfect sense




Big Opportunities, Big Challenges

Requires zero CAPEX

The cloud has fundamentally changed IT spending as organizations know it—and in a good way.

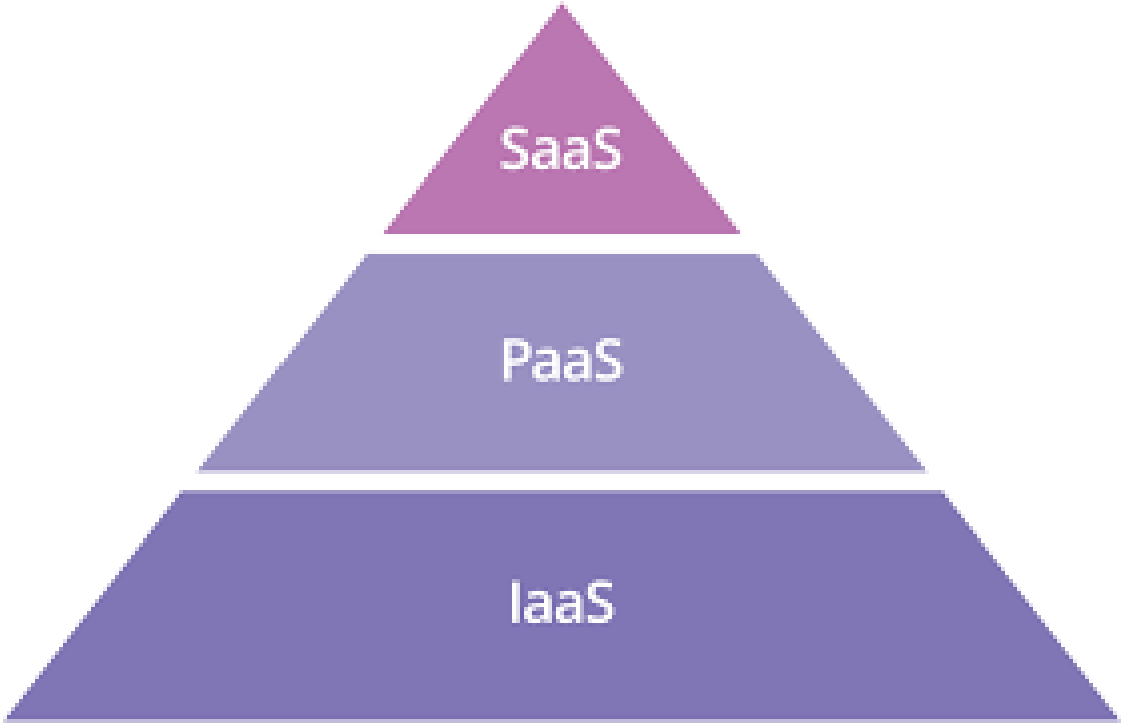
AMAZON.COM, INC.

CONSOLIDATED STATEMENTS OF CASH FLOWS
(in millions)

	Year Ended December 31,		
	2014	2015	2016
CASH AND CASH EQUIVALENTS, BEGINNING OF PERIOD	$ 8,658	$ 14,557	$ 15,890
OPERATING ACTIVITIES:			
INVESTING ACTIVITIES:			
Purchases of property and equipment, including internal-use software and website development, net	(4,893)	(4,589)	(6,737)
Acquisitions, net of cash acquired, and other	(979)	(795)	(116)
Sales and maturities of marketable securities	3,349	3,025	4,733
Purchases of marketable securities	(2,542)	(4,091)	(7,756)
Net cash provided by (used in) investing activities	(5,065)	(6,450)	(9,876)
FINANCING ACTIVITIES:			

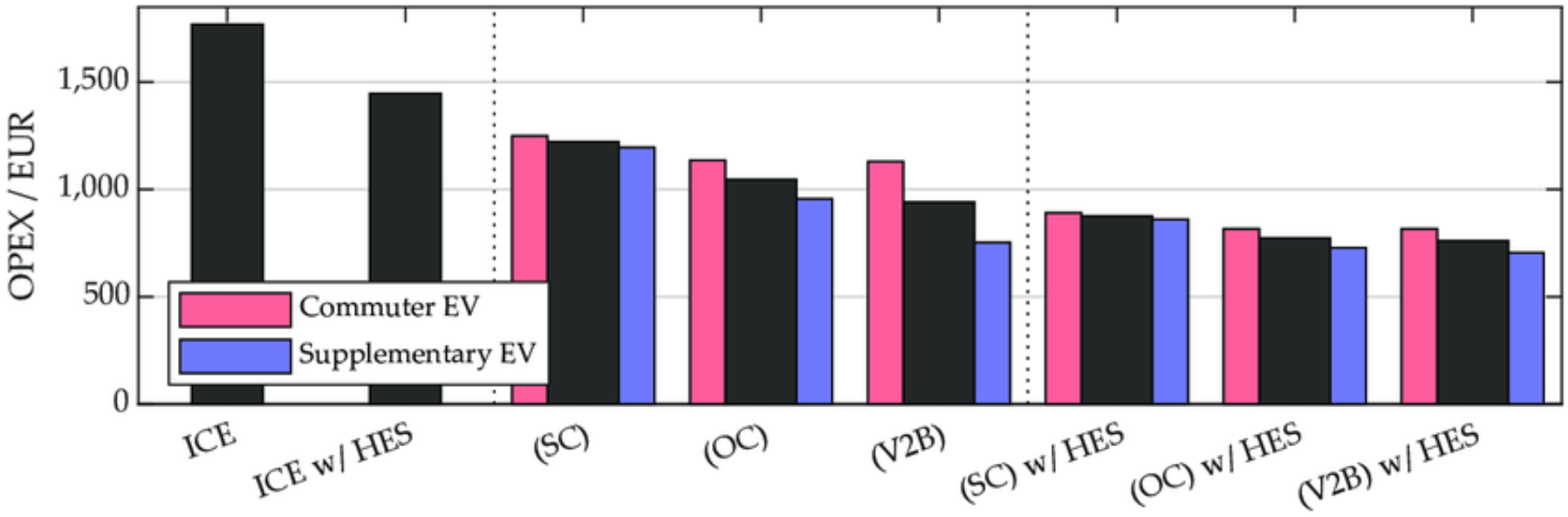
On-premise capital expenditure (CAPEX) investments

Requires zero CAPEX



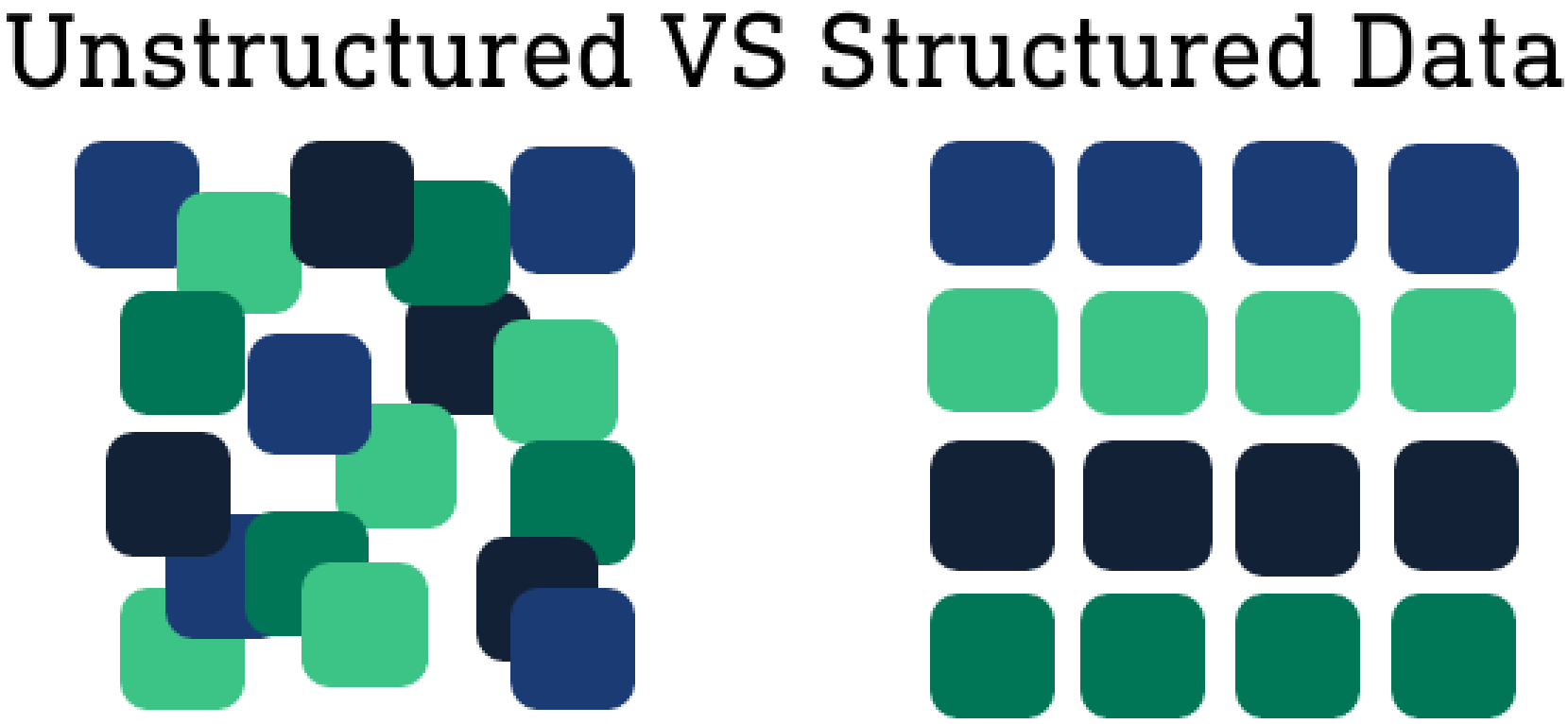
Cloud's Infrastructure-as-a-Service models

Requires zero CAPEX



Operating expenditure (OPEX) column.

Enables faster scalability



Enables faster scalability

Manage large spikes in traffic or usage

Lowers the cost of analytics

Mining big data in the cloud has made the analytics process less costly. In addition to the reduction of on-premise infrastructure, you can also save on costs related to system maintenance and upgrades, energy consumption, facility management, and more. You can also worry less about the technical aspects of processing big data and focus more on creating insights. Even better, the cloud's pay-as-you-go model is more cost-efficient, with little waste of resources.

Lowers the cost of analytics

- Analytics process less costly.
- Can also save on costs
- worry less about the technical aspects of processing big data
- focus more on creating insights
- Even better, the cloud's pay-as-you-go model is more cost-efficient, with little waste of resources.

Lowers the cost of analytics

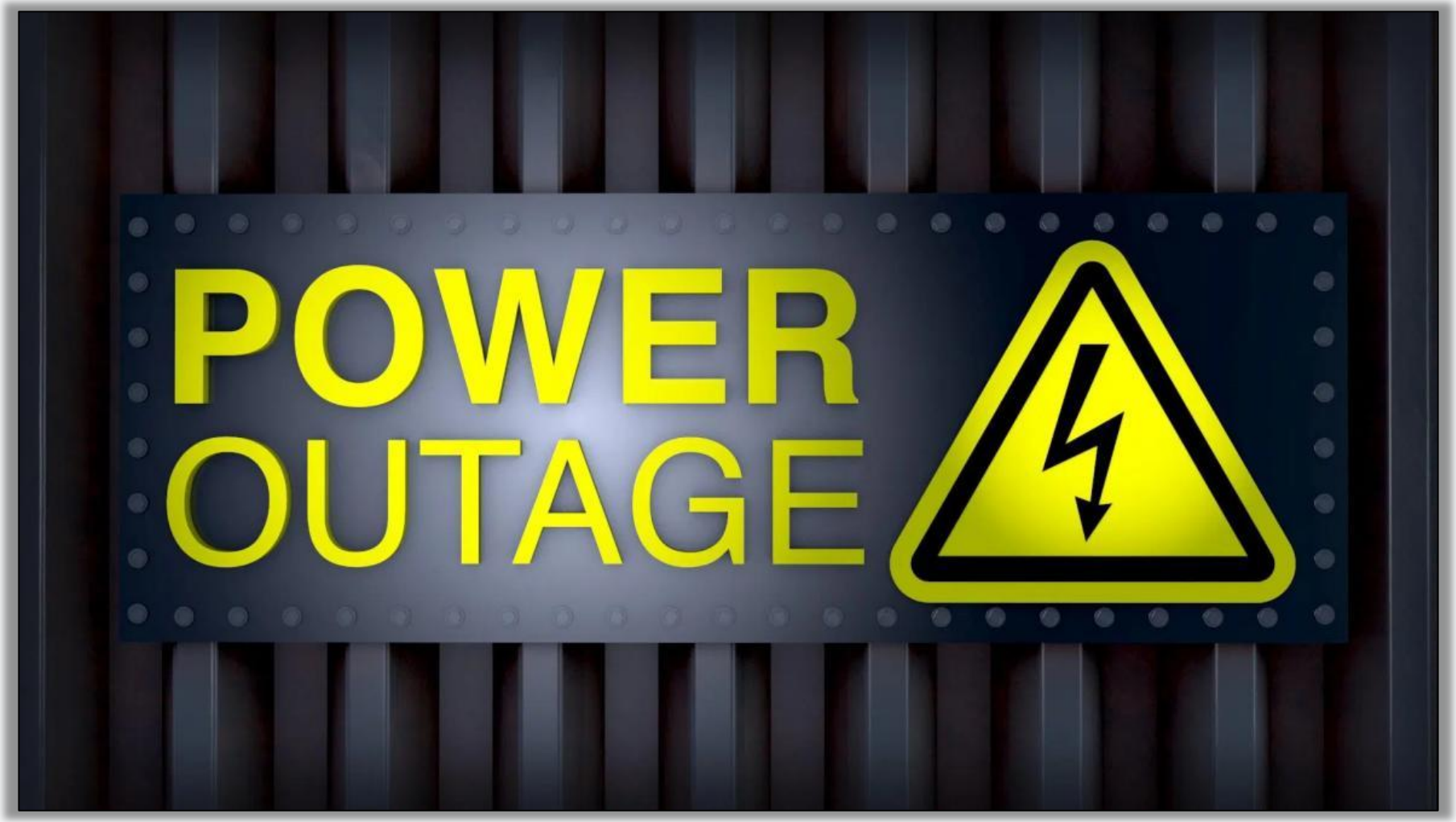
Pay-as-you-go model

Encourages an agile and innovative culture

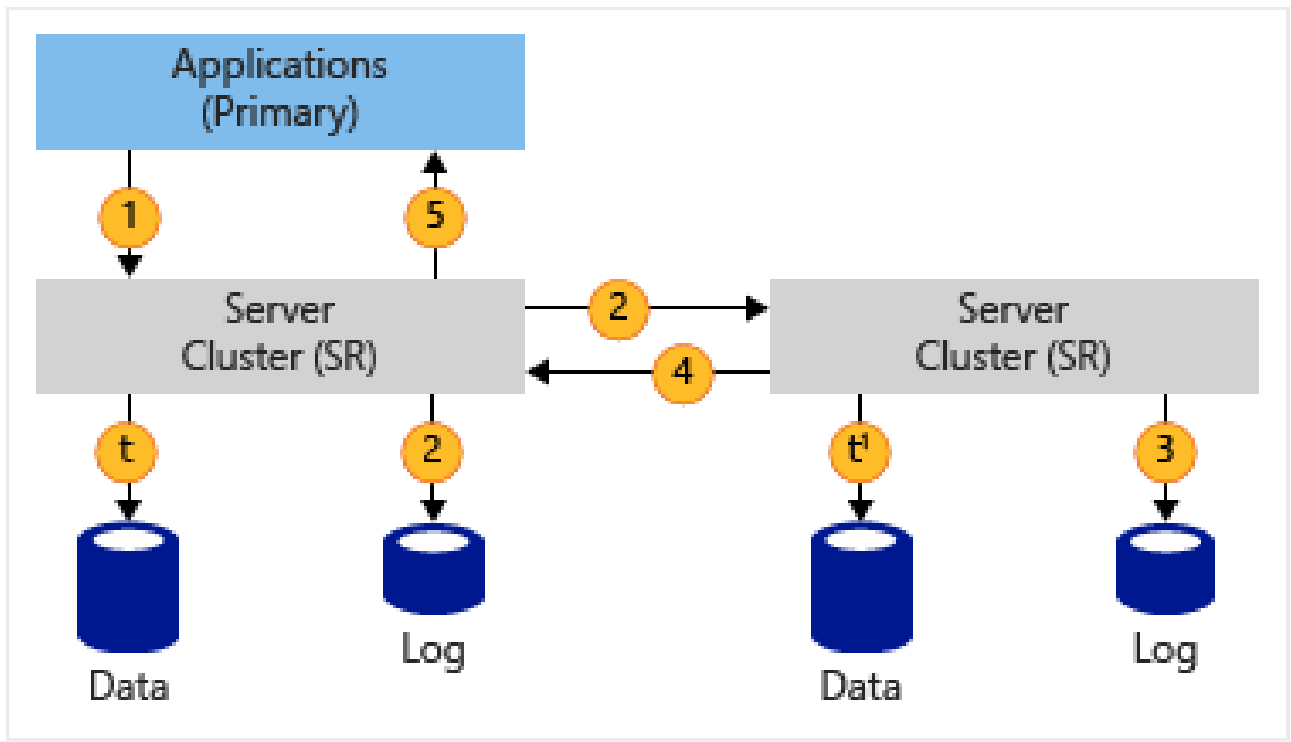
- Cultivated within any enterprise
- Creative ways of using big data to gain a competitive advantage
- Focuses on analyzing data instead of managing servers and databases, you can more easily and quickly unearth insights

Enables better business continuity and disaster recovery

Enables better business continuity and disaster recovery



Enables better business continuity and disaster recovery

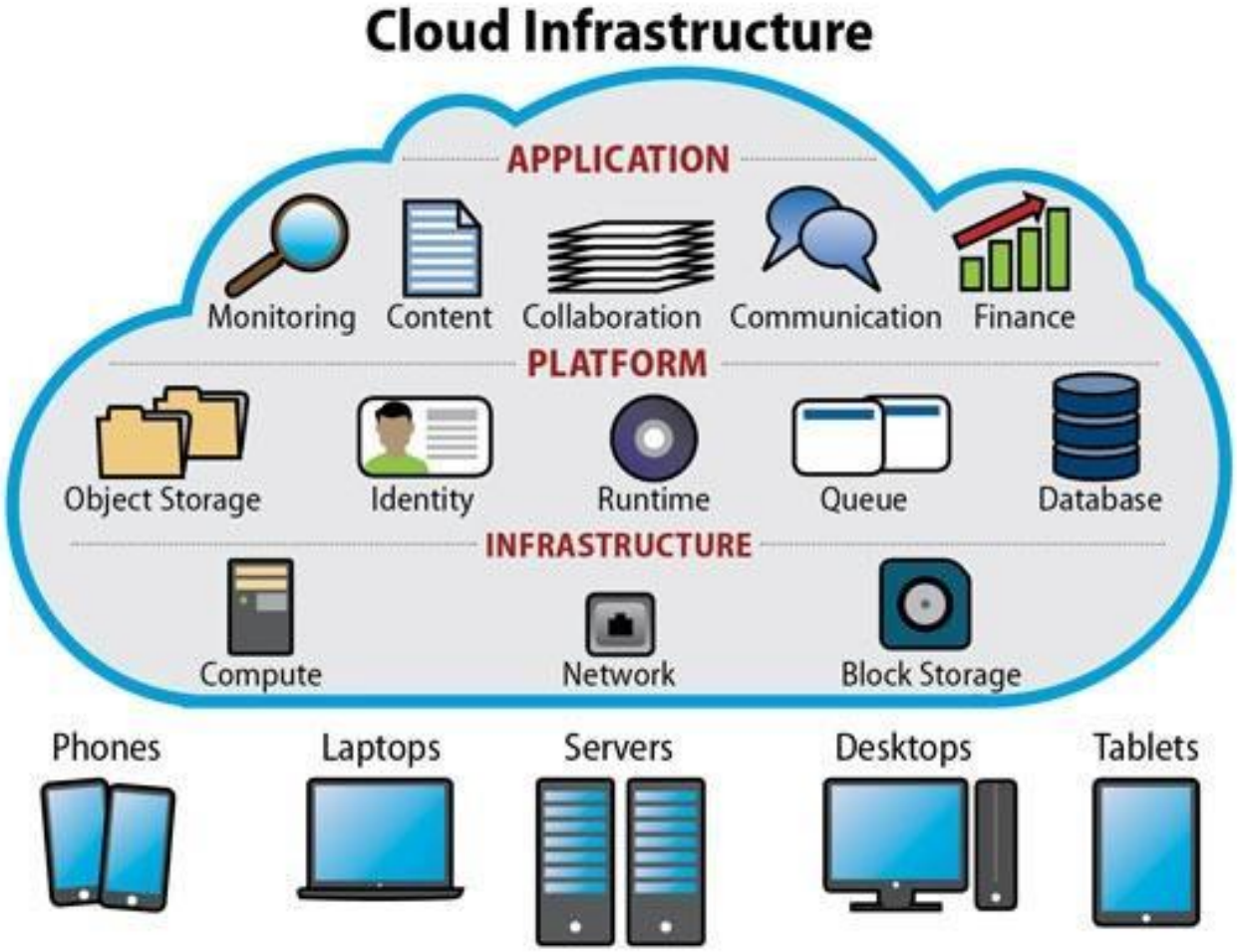


The task of replicating a data center – with duplicate storage, servers, networking equipment, and other infrastructure – in preparation for a disaster is tedious, difficult, and expensive

Enables better business continuity and disaster recovery

Legacy systems often take very long to back up and restore. This is especially true in the era of big data, when data stores are so immense and expansive.

Enables better business continuity and disaster recovery

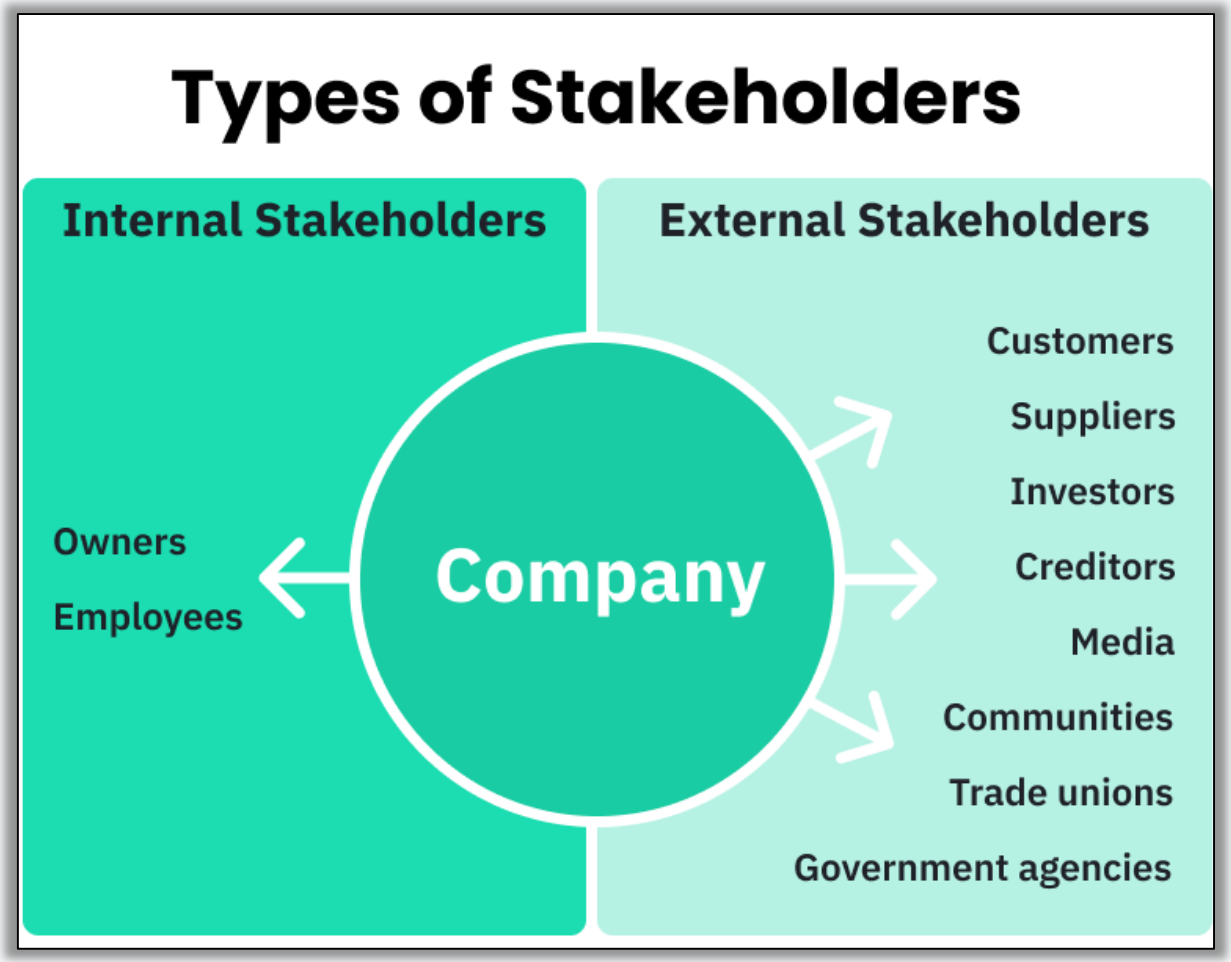


Potential challenges of big data in the cloud

Potential challenges of big data in the cloud

Potential challenges of big data in the cloud

Potential challenges of big data in the cloud



Business Stakeholders

Less control over security

Sensitive information

Less control over security

Individuals' addresses

Less control over security

Individuals' addresses

Credit card details

Less control over security

Individuals' addresses

Credit card details

Social security numbers

Less control over security

Data breaches

Less control over security

Loss customers and revenue

Less control over security

less direct control over your data,

Less control over security

Security should not be a hindrance

Less control over compliance

Less control over compliance

Less control over compliance

Less control over compliance

Even if your CSP is managing a good chunk of your compliance, you should make sure you know the answers to the following questions:

- Where is the data going to reside?
- Who is going to manage it, and who can access it?
- What local data regulations do I need to comply with?

Less control over compliance

Make sure you know exactly

- what data is stored where,
- ensure that your CSP has robust compliance policies,
- understand the shared responsibility model,
- potentially create Service Level Agreements (SLAs) for compliance.

Network dependency and latency issues

- Availability of the data is highly reliant on network connection
- System could be prone to service interruptions.
- Volume of data that's being transferred, analyzed, and processed at any given time.


That's all for now...