



Microsoft



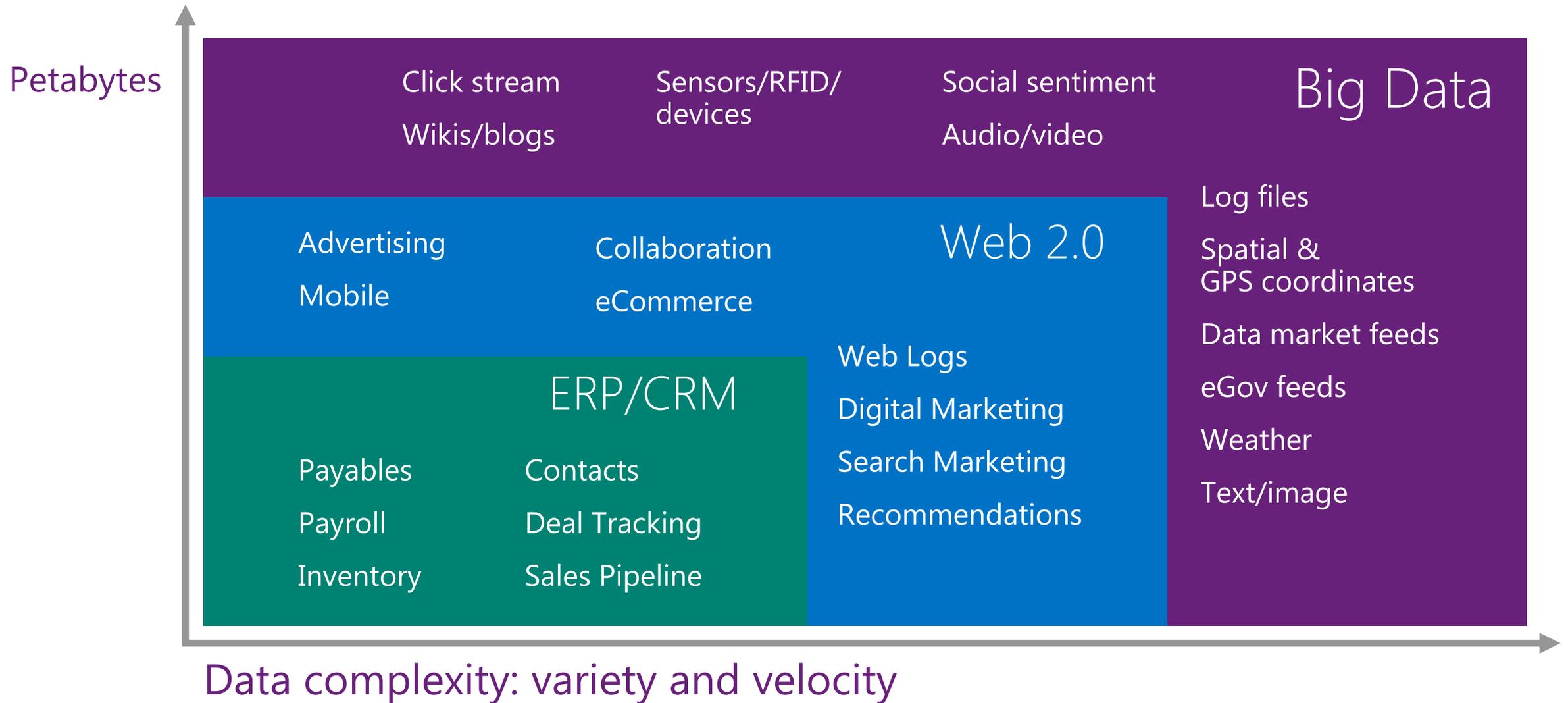
Microsoft

Microsoft Big Data Solutions

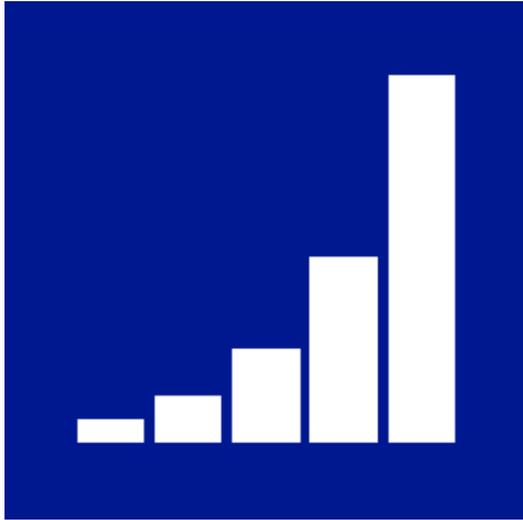
Anar Taghiyev P-TSP

E-mail: b-anarta@Microsoft.com;

Why/What is Big Data?



Data is rapidly changing



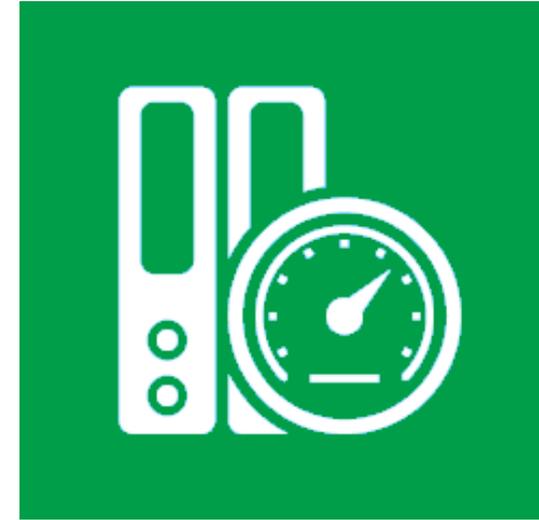
GROWING VOLUME (Large)

Paper conversion to digital form, more information is being gathered, and greater detail of existing information is being captured.



MORE VARIETY (Unstructured)

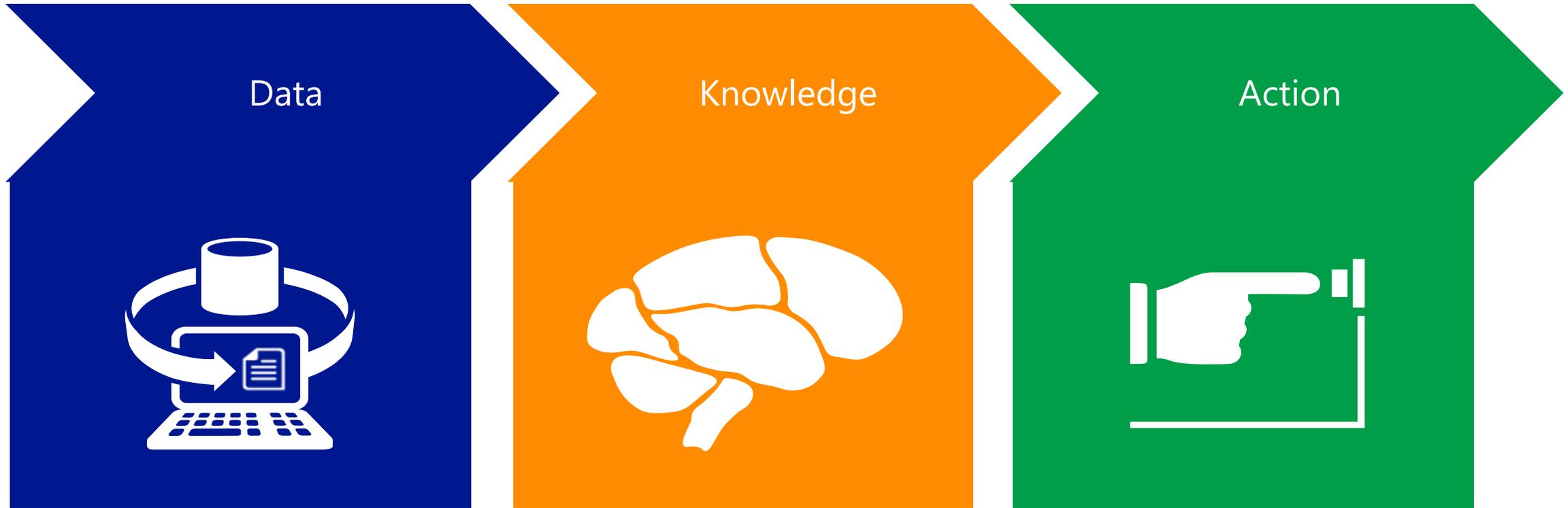
Characterized by text based documents, image, audio, and video files, new sources of data and a lack of standard data models.



RAPID VELOCITY (Complex)

Fueled by real-time data collection, constantly changing information, and the growth of social networks.

Why should I care?



The benefits of Big Data are not limited only to business intelligence experts or data scientists. Nearly everyone in your organization can analyze and make more informed decisions with the right tools.

PowerPivot for Microsoft Excel and Power View for SharePoint give nearly all users a view into *structured and unstructured data*.

With the Hive Add-in for Excel and Hive ODBC Driver, almost anyone in your organization can directly access *Hadoop data* from end-user tools.

Hadoop simplifies programming for developers with JavaScript for MapReduce jobs. The *JavaScript* implementation can also reduce your code by up to 10 times compared to Java.



Storage

Reliable, economical cloud storage for data big and small

1. Manage petabytes of storage
2. Geo-redundant storage across hundreds of miles
3. Fastest performance in the industry
4. Industry standard SMB file sharing across VMs
5. Pay for what you use with competitive pricing
6. REST, .NET, Java, C++, node.js, PowerShell and more

Highly scalable

Azure Storage keeps up with your growing data needs, with up to 500 TB of total storage per account. A single subscription supports up to 50 storage accounts, delivering petabytes of storage for the largest scenarios. Whether you're building a consumer site or a terabyte-scale big data application, Azure is designed to handle it.

Durable & highly available

Azure storage automatically replicates your data to help guard against unexpected hardware failures and make sure it's available when you need it. We keep 3 copies within a single region. A geo-redundancy option creates 3 additional copies hundreds of miles away for higher availability and disaster recovery.

Designed for developers

Build apps with first class client library support for .NET, Java, Android, C++, and Node.js. Data in Azure Storage is also accessible via REST API, which can be called by any language that makes HTTP/HTTPS requests. Azure Storage includes strong consistency guarantees, simplifying cloud application development and enabling predictable experiences for applications built on Azure.

Cost effective

Pay only for what you use, at a cost lower than on-premises storage options. Our competitive price matching policy also means you're always getting a great deal on cloud storage.

Azure Storage Components

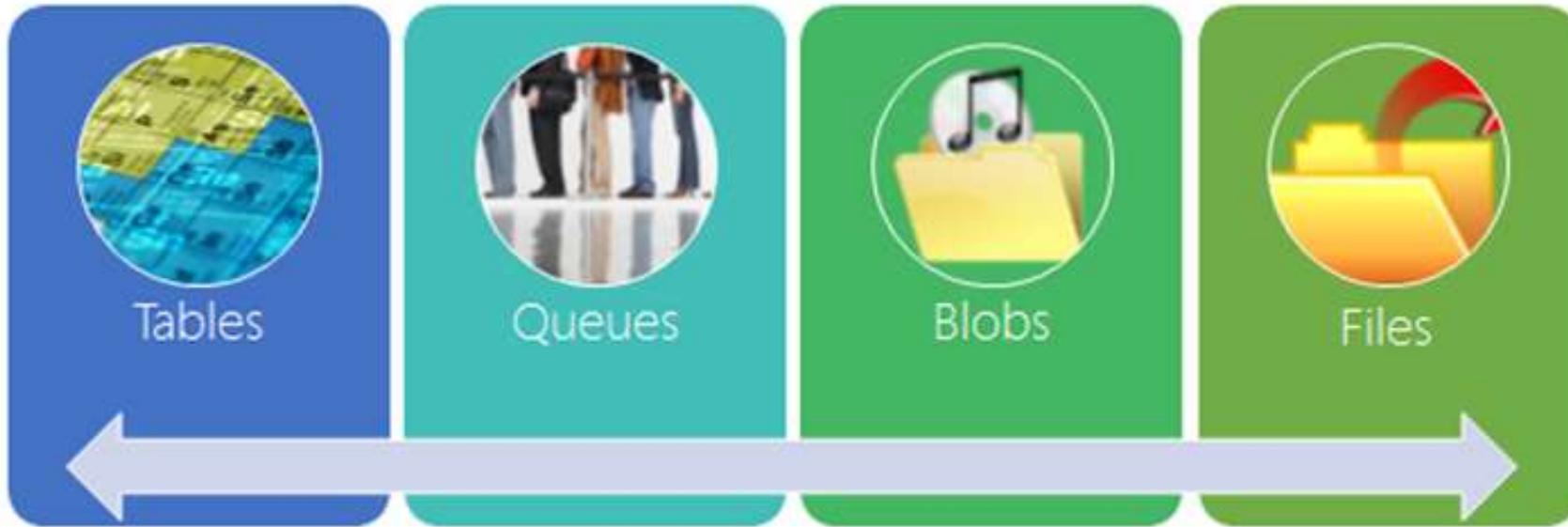


Table Storage

A NoSQL key-value store for rapid development using massive semi-structured datasets

Key scenarios:

- 1.Store semi-structured data that's highly available
- 2.Create massively scalable apps
- 3.Create apps that require a flexible data schema
- 4.Use JSON to serialize data
- 5.Perform OData-based queries

Queue Storage

Reliable messaging for scenarios including workflow processing or communication between application components

Key scenarios:

- 1.Effectively scale apps according to traffic
- 2.Create apps that can handle traffic bursts
- 3.Build more flexible applications

Blob Storage

Reliable, cost-effective cloud storage for large amounts of unstructured data

Key scenarios:

- 1.Store petabytes of highly available data
- 2.Power big data analytics
- 3.Serve content to web or mobile applications
- 4.Perform secure backup and disaster recovery
- 5.Stream video and audio

File Storage

Shared storage for applications using the standard SMB protocol

Key scenarios:

- 1.Move existing applications to the cloud
- 2.Use cloud storage with no code changes
- 3.Share tools and logs between VMs

Blobs, Tables, Queues, and Files

Azure Storage provides the flexibility to store and retrieve large amounts of unstructured data, such as documents and media files with Azure Blobs; structured nosql based data with Azure Tables; reliable messages with Azure Queues, and use SMB based Azure Files for migrating on-premises applications to the cloud.

Built for big data

When you do big data analysis, you've got to put all that data somewhere. The Azure Blob service offers several benefits over the traditional Hadoop Distributed File System (HDFS) for big data scenarios—including geo-replication, lower storage costs, easier scalability, and data sharing outside of the Hadoop cluster. Use your own on-premises service to analyze blob-stored data, or Azure services including Batch, Machine Learning, and HDInsight (Hadoop).

Big Data

Reveal new insights and drive better decision making with Azure HDInsight, a Big Data solution powered by Apache Hadoop. Surface those insights from all types of data to business users through Microsoft Excel.

HDInsight

Our 100% Apache Hadoop-based service in the cloud

1. Scale to petabytes on demand
2. Process unstructured and semi-structured data
3. Develop in Java, .NET, and more
4. No hardware to buy or maintain
5. Deploy in Windows or Linux
6. Spin up a Hadoop cluster in minutes
7. Visualize your Hadoop data in Excel
8. Easily integrate on-premises Hadoop clusters

Machine Learning

Powerful cloud-based predictive analytics

1. Model your way
2. Deploy in minutes
3. Expand your reach

Azure Search

Search-as-a-service for web and mobile app development

1. Get your search indices up and running quickly
2. Easily scale up and down as needed
3. Connect search results to business goals with great control over search ranking
4. Leverage Microsoft's deep knowledge of natural language processing
5. Load and update automatically with integrated indexers
6. Easily add geo-spatial search to your app

Azure Backup

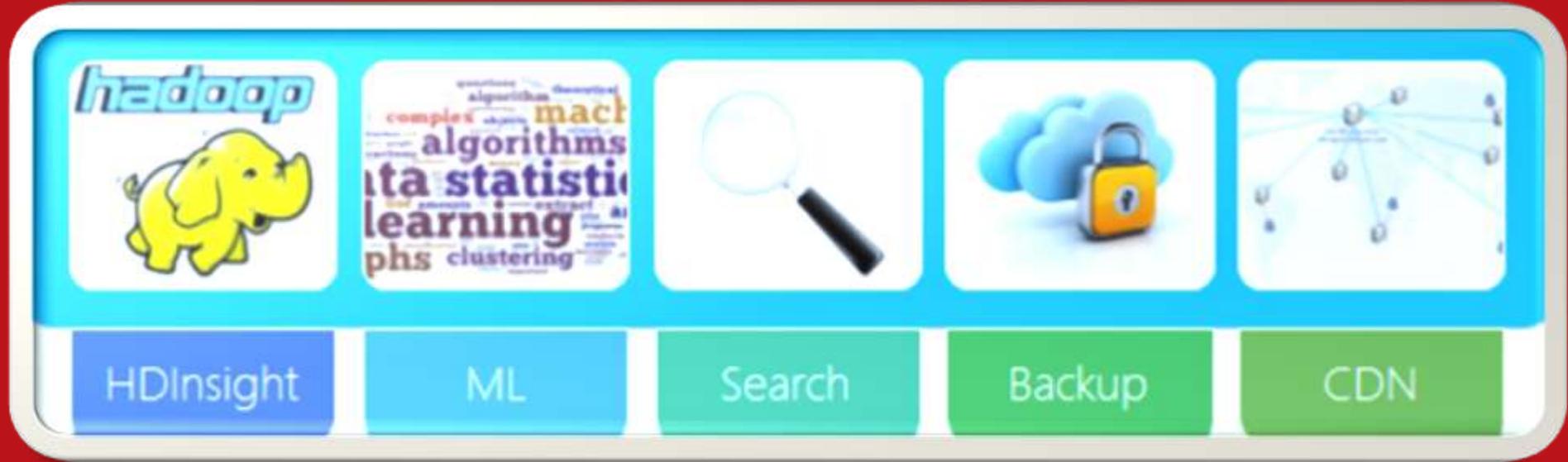
Simple and reliable cloud integrated backup

1. Unified solution for protecting data on premise and in the cloud
2. 99.9% availability guaranteed
3. Reliable Offsite backup target
4. Efficient incremental backups
5. Secure, encrypted in-transit and at-rest
6. Geo-replicated backup store

Azure CDN

A fast and modern global delivery network for high-bandwidth content

1. Lower latencies
2. Massively scalable
3. Designed for today's web
4. Improved availability and performance
5. Capacity on demand
6. Robust security



Machine learning today

Complexities in a nascent market

Huge set-up costs create unnecessary barriers to entry

Cumbersome data management restricts access to data

Complex and fragmented tools limit participation in exploring data and building predictive models

Many models never achieve business value due to difficulties with deploying to production

Break away
from industry
limitations

The Microsoft data platform

Microsoft SQL Server

Microsoft Azure

Office

Visualize + decide



Applications



Reports



Dashboards



Natural language query



Mobile

Transform + analyze



Orchestration



Information management



Complex event processing



Modeling



Machine learning

Capture + manage



Relational



Non-relational



NoSQL

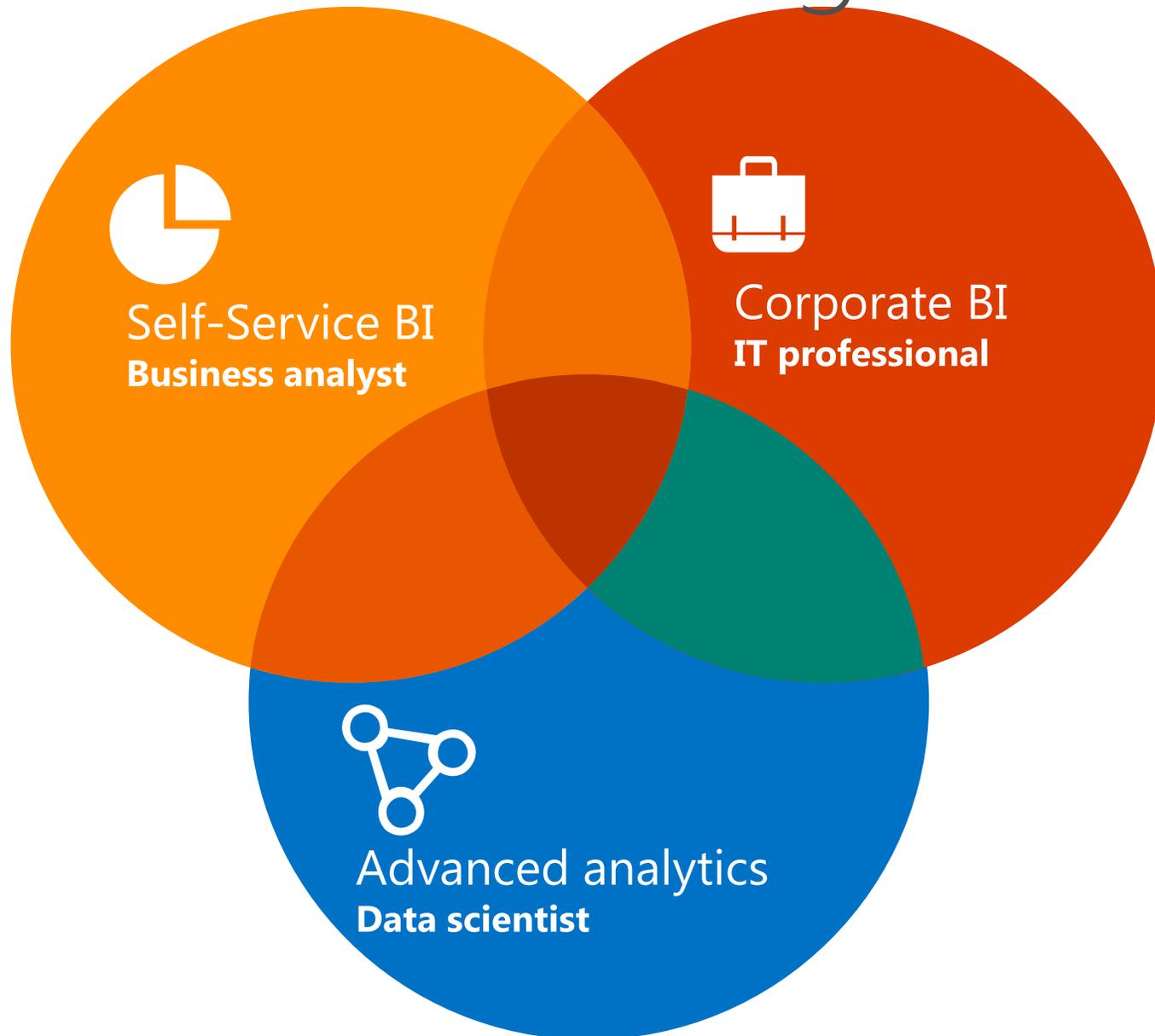


Streaming



Internal & external

Business Intelligence in 2015+



- **Corporate BI** – enterprise datawarehouse and standard reporting
- **Self-Service BI** – provide your business analysts and end-users more capabilities and freedom
- **Advanced Analytics** – predictive analytics to recognize patterns in data

Combine various types of data...



What can Big Data do for Education and Research?



Social Analytics



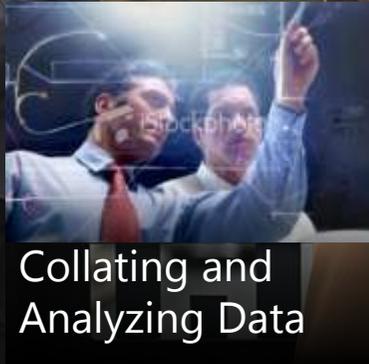
Associative Analytics



Assessment and Prediction



Improved Research



Collating and Analyzing Data



Social Network Awareness



Employment Analysis



Cross Institutional Learning



Unlock insights on
any data

“We can collect, analyze, and generate near-real-time BI with Big Data collected from social media feeds, GPS signals, and data from government systems.”

LLUIS SANZ MARCO

DIRECTOR OF INFORMATION

MUNICIPAL, INSTITUTE OF INFORMATION



CITY OF BARCELONA

Citizen Dashboard: City by the Numbers


TRANSPORTATION


LIVABILITY


ENVIRONMENT


URBAN FORM


ECONOMY


FINANCE



✔ Meets or Exceeds Target ⚡ Near Target ✖ Needs Improvement 🕒 Measuring 📊 Collecting Data

<https://dashboard.edmonton.ca/>



• LEARN MORE

Learn more about our Microsoft Big Data solution and how to get started at

<https://www.microsoft.com/en-us/server-cloud/solutions/big-data.aspx>

Learn more about Microsoft Azure at

<http://azure.microsoft.com/en-us/>

<http://azure.microsoft.com/en-us/services/>

Learn more about Microsoft CityNext at

<https://www.microsoft.com/en-us/citynext/default.aspx>

Learn more about Power BI at

www.powerbi.com

