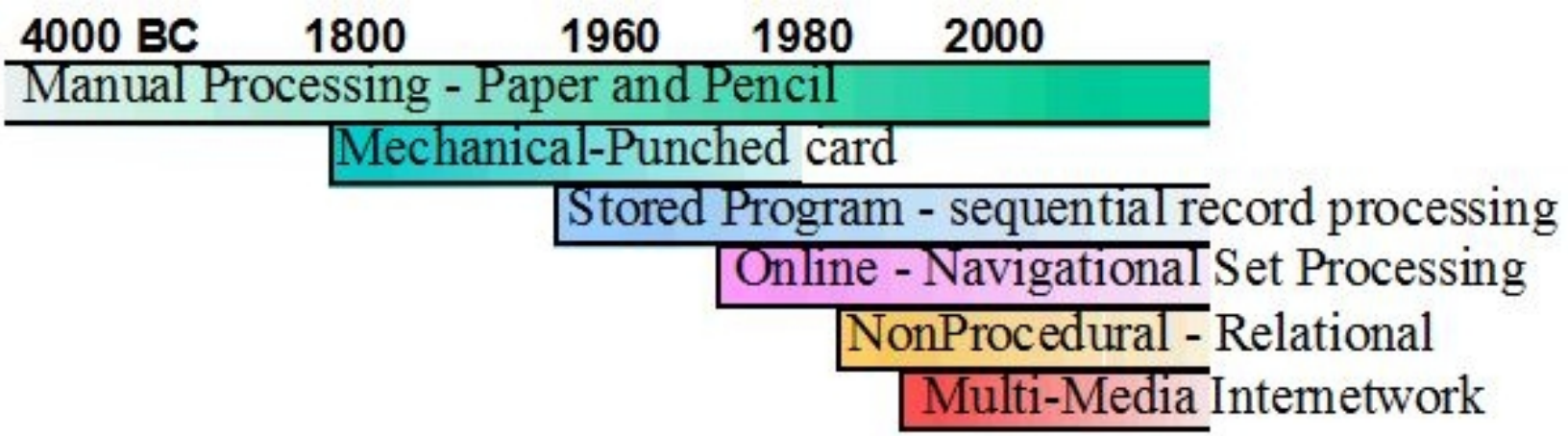
Cloud Computing

BIG DATA in CLOUD

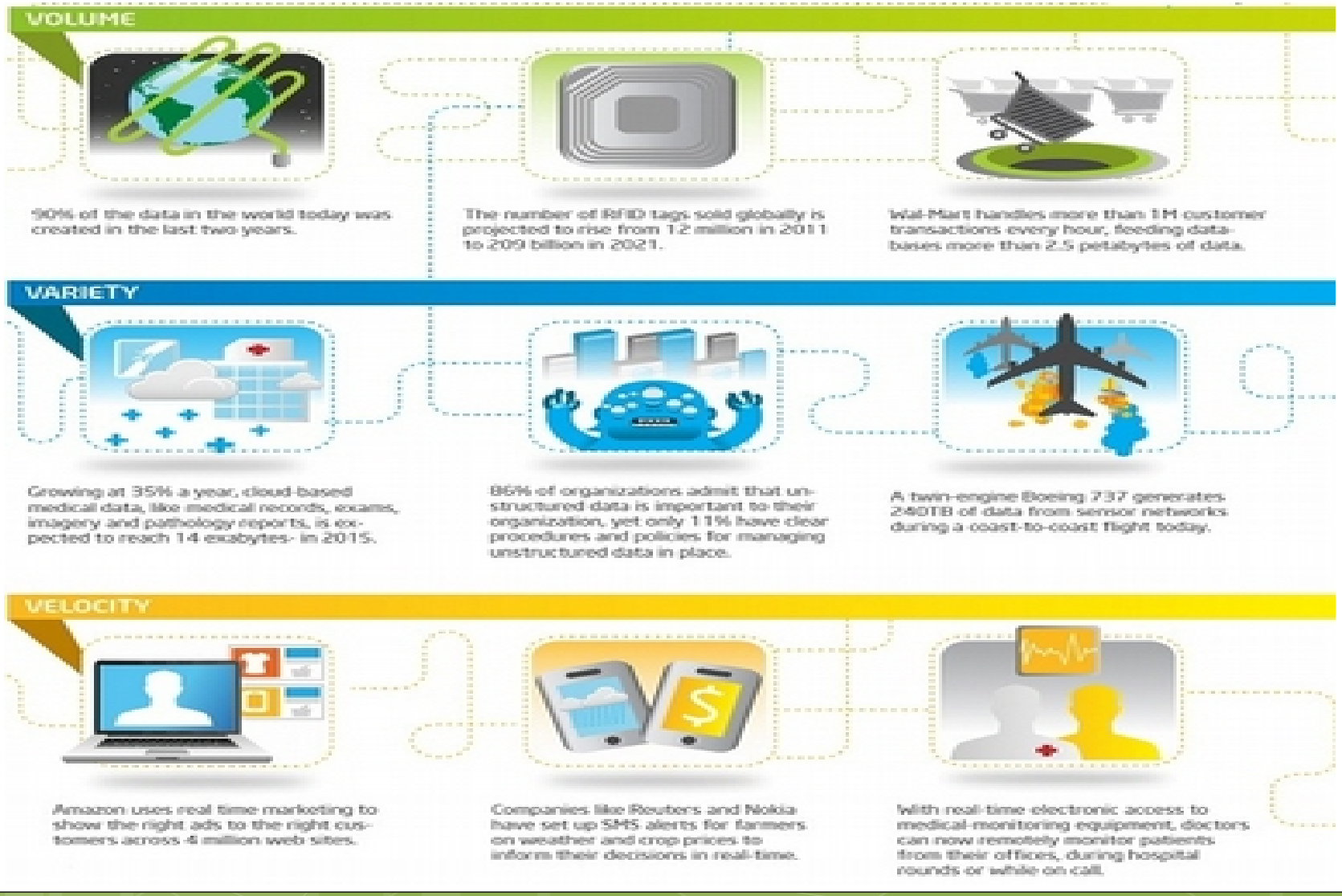
6 Generasi Manajemen Data



Big Data

- Volume
- Velocity
- Variety

Atribut Big Data



Prinsip Big Data

- tidak membuang data apapun karena residu tersebut mungkin akan menjadi penting sejalannya waktu.

Big Data Processing Software

Pengolahan Big Data diklarifikasi menjadi 2 jenis metode :

- Pengolahan data berbasis batch
- Pengolahan data berbasis real-time

Big Data Processing Software (Google)

- Google memiliki teknologi canggih yang memungkinkannya mampu mengolah dan memanfaatkan Big Data dengan tepat.

- Teknologi :

  Google Bigtable

  Google MapReduce

  Google File System (GFS)

Google Bigtable

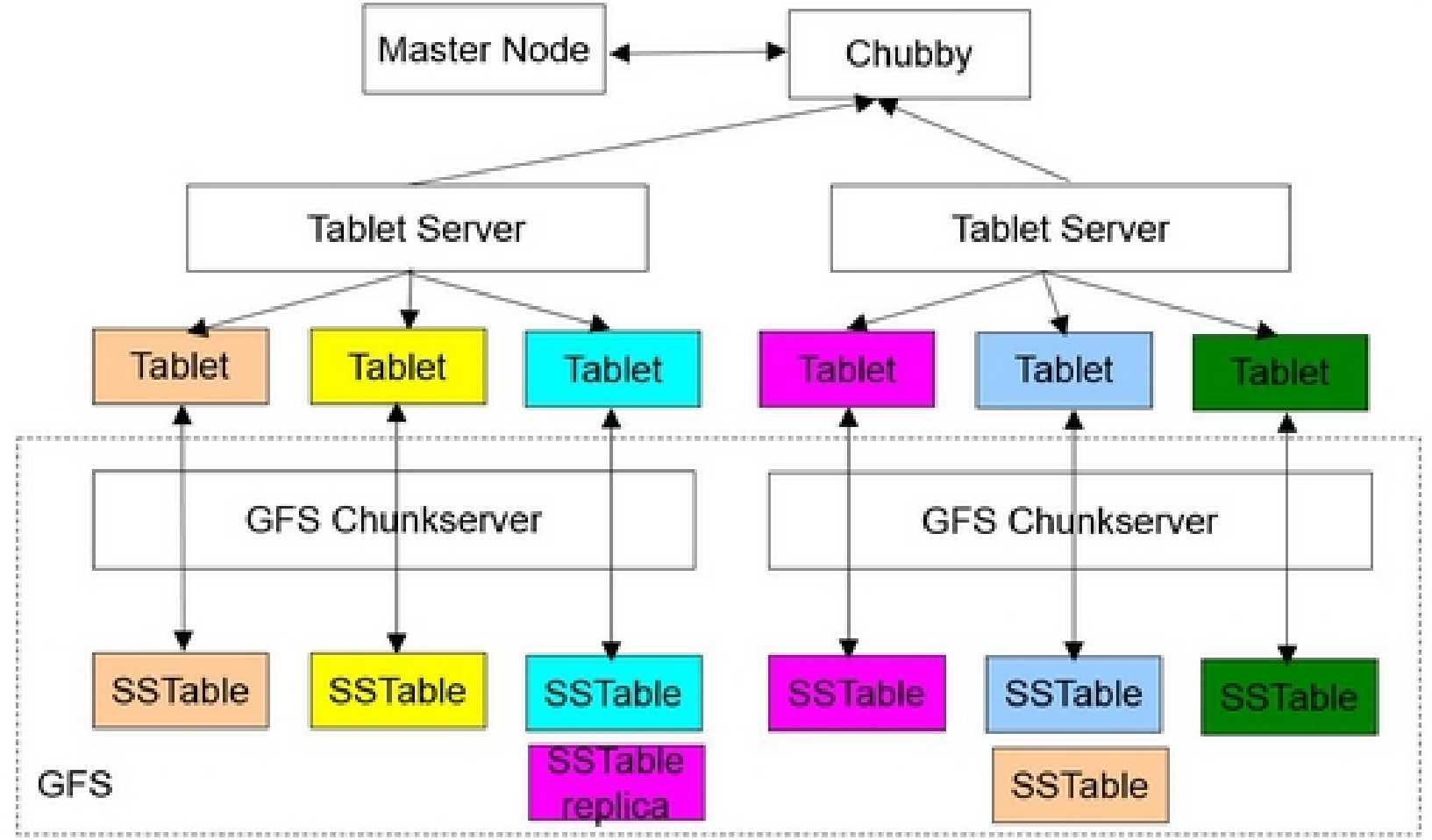
- System penyimpanan data terdistribusi yang ditujukan untuk mengelola data yang terstruktur dan didesain sebagai system yang handal untuk mengelola data dalam skala petabytes dan dalam ribuan mesin (komputer)

Google Bigtable(Cont..)

- Google menggunakan Bigtable dalam lebih dari 60 produk dan proyeknya termasuk :

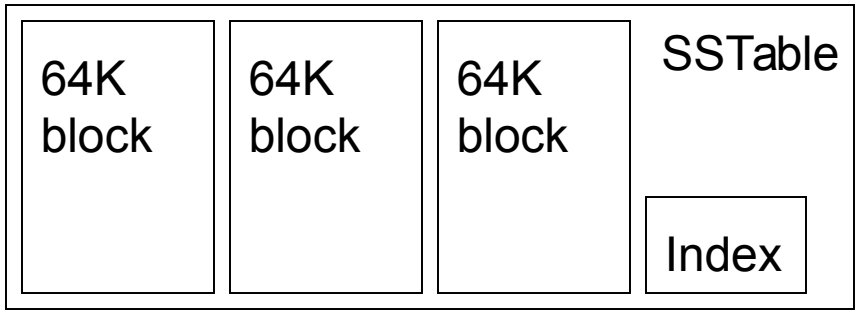
- Google web indexing, Google Analytics, Google Finance, Orkut, Personalize Search, Writely dan Google Earth.

BigTable Architecture



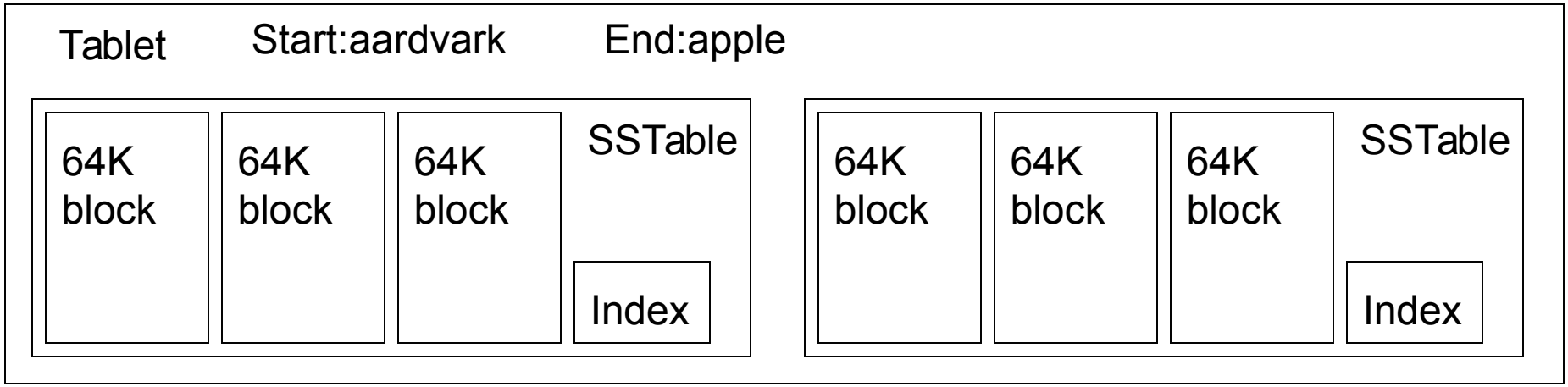
SSTable

- Immutable, sorted file of key-value pairs
- Chunks of data plus an index
  - Index is of block ranges, not values



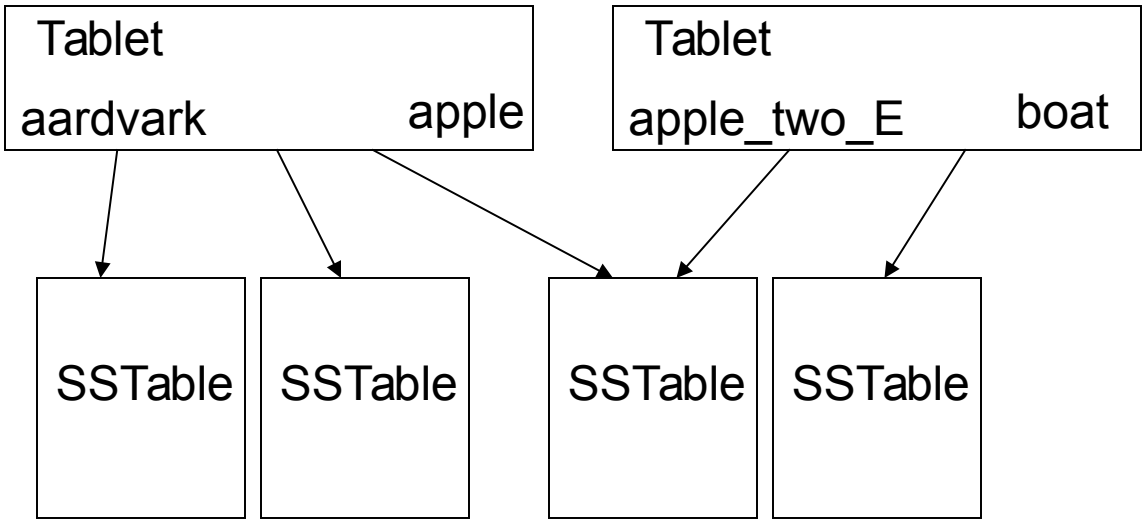
Tablet

- Contains some range of rows of the table
- Built out of multiple SSTables

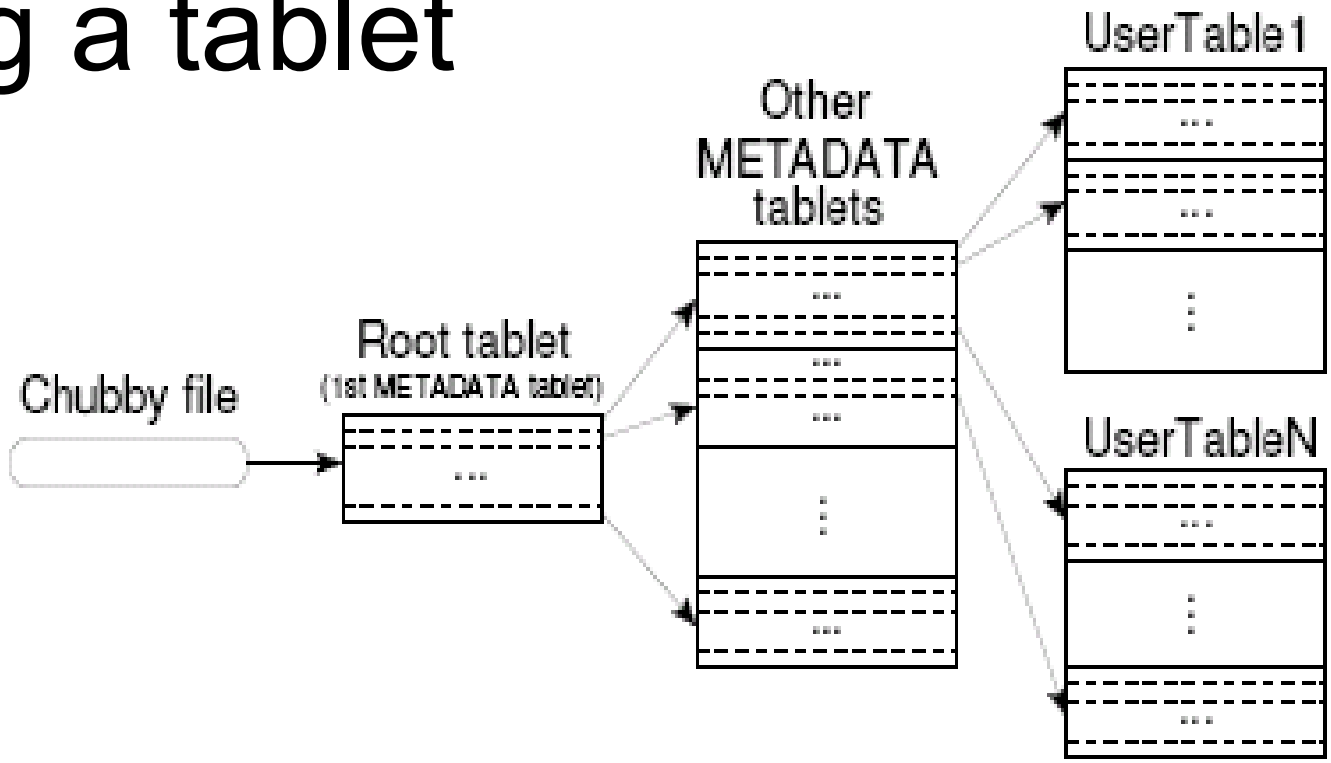


Table

- Multiple tablets make up the table
- SSTables can be shared
- Tablets do not overlap, SSTables can overlap



Finding a tablet



- Stores: Key: table id + end row, Data: location
- Cached at clients, which may detect data to be incorrect
  - in which case, lookup on hierarchy performed
- Also prefetched (for range queries)

Servers

- Tablet servers manage tablets, multiple tablets per server. Each tablet is 100-200 MB
  - Each tablet lives at only one server
  - Tablet server splits tablets that get too big
- Master responsible for load balancing and fault tolerance

Google MapReduce

- Model pemrograman rilisan Google yang ditujukan untuk memproses data berukuran raksasa secara terdistribusi dan paralel dalam cluster yang terdiri atas ribuan komputer.

Google File System (GFS)

Salah satu jenis dari media penyimpanan data seperti halnya hard disk drive (HDD), flash disk, DVD-R dan sebagainya. Bedanya, GFS menyimpan data-nya secara terdistribusi pada komputer-komputer dalam suatu cluster.

GFS bisa menyimpan data super besar yang tidak bisa disimpan dalam suatu HDD paling besar sekalipun.

Google File System

- Large-scale distributed “filesystem”
- Master: responsible for metadata
- Chunk servers: responsible for reading and writing large chunks of data
- Chunks replicated on 3 machines, master responsible for ensuring replicas exist

Hadoop

Apache telah merilis software open source yang dikenal dengan nama **Hadoop** untuk mengebangkan dan menjalankan aplikasi MapReduce.

Secara garis besar Hadoop terdiri atas HDFS (Hadoop Distributed File System) dan Hadoop MapReduce. HDFS adalah versi open source-nya GFS (Google File System), dan Hadoop MapReduce adalah versi open source dari Google MapReduce.

Hadoop(Cont..)

- Keunggulan:

- Sederhana
- Fleksibel dalam Ukuran
- Handal, anti Gagal

Big Data vs Traditional BI

- Big Data = Buttom-Up sedangkan Traditional BI = Top-down

- Ruang Lingkup :

- Big Data : lebih luas
- Traditional BI : terbatas , BI tidak pernah bisa mengantisipasi banyaknya gambar, file MP3 , video dan media sosial.

Mengapa perlu mengintegrasikan Big Data untuk bisnis?

- Peningkatan pemahaman pelanggan

  - Menggunakan solusi CRM.

- Peningkatan layanan pelanggan

  - Membantu meningkatkan pengalaman pelanggan dan pada saat yang bersamaan mengevaluasi ROI aplikasi CRM.

- Mendukung pengambilan keputusan

  - Memiliki statistik untuk menghadapi pelanggan sehingga mendukung keputusan

Mengapa perlu mengintegrasikan Big Data untuk bisnis? (Cont..)

- Melihat tren

- membantu menganalisis kegiatan pelanggan yang telah lalu untuk menjelaskan perilaku masa depannya.

- Menetapkan patokan

- Solusi CRM dengan *big data* terpadu memungkinkan perusahaan menetapkan pembiayaan selama periode waktu dibandingkan dengan pesaingnya.

Perusahaan yang menggunakan Big Data

- Yahoo!, Amazon, IBM, Microsoft, Facebook, Twitter, Hewlett-Packard, LinkedIn, RECRUIT, Rakuten Japan, dan masih banyak lagi.

Contoh :

- Microsoft (Windows Azure Hadoop)
- Oracle (Big Data Appliance)
- SAP (Hana)
- EMC (GreenPlum Hadoop)

Etika Big Data

- **Privacy isn’t dead**

Kata lain dari aturan dalam informasi. Private tidak selalu berarti rahasia, tetapi memastikan bahwa privasi data adalah mendefinisikan suatu masalah dan menegakkan aturan informasi. Aturan tersebut juga tidak selalu mengenai aturan tentang pengumpulan data tetapi juga tentang penggunaan data dan retensinya.

Etika Big Data(Cont..)

- **Shared private information can still remain confidential**. (berbagi informasi tetapi tidak menghilangkan nilai kerahasiaan dari informasi tersebut)

Setiap data/informasi yang dibuat dan dishare tidak berarti bahwa nilai kerahasiaan pada data tersebut bisa dilihat oleh banyak orang.

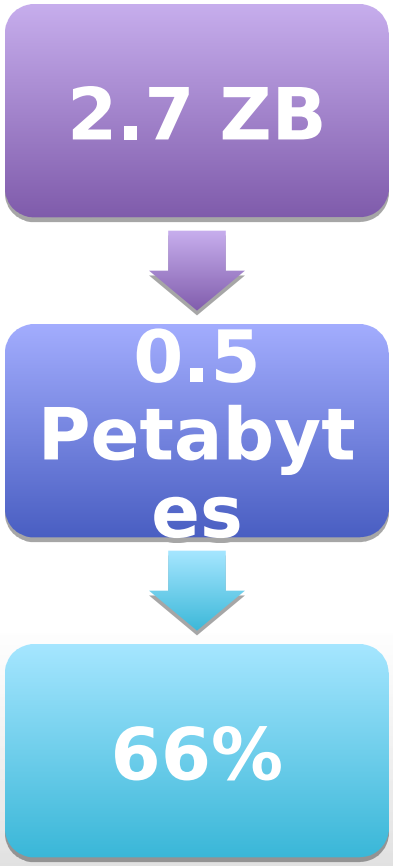
Etika Big Data(Cont..)

- **Big data requires transparency.**

Big data akan berpengaruh ketika penggunaan sekunder dari set data yang menghasilkan prediksi baru dan kesimpulan.

Big Data In the Cloud

Big Data In The Cloud



Global Digital Data

Two years tweets

Will or plan to use Big Data in the cloud

43% think that data analytics could be improved in their organization if data analytics was part of **cloud services**

Large ISV Case Study

- Application
  - Call Center surveillance
- Background
  - Previously – voice data
- Goal for a new system
  - Monitor data & voice
  - Multiple data sources
  - Advanced correlations

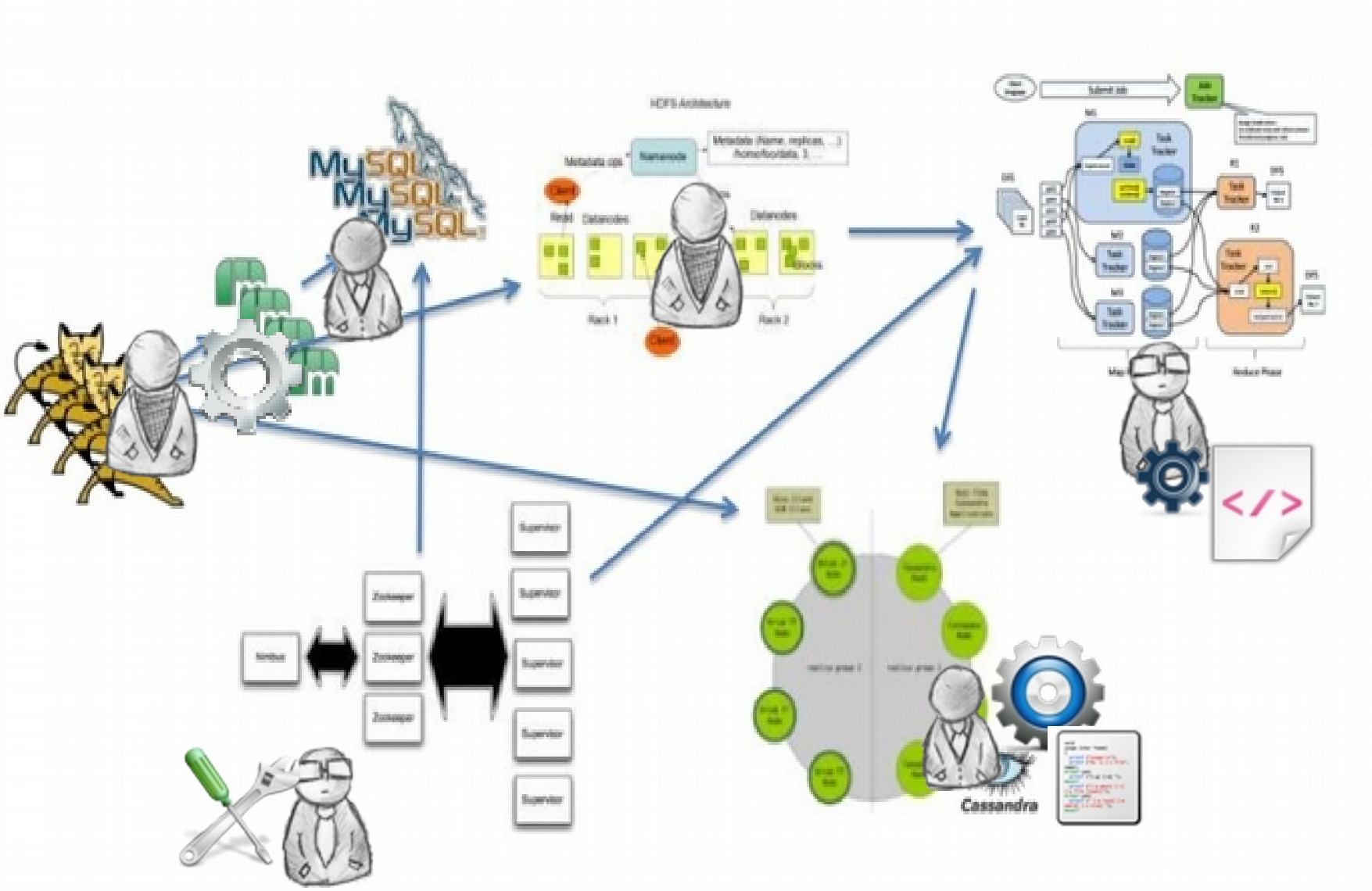
Ever Growing Data

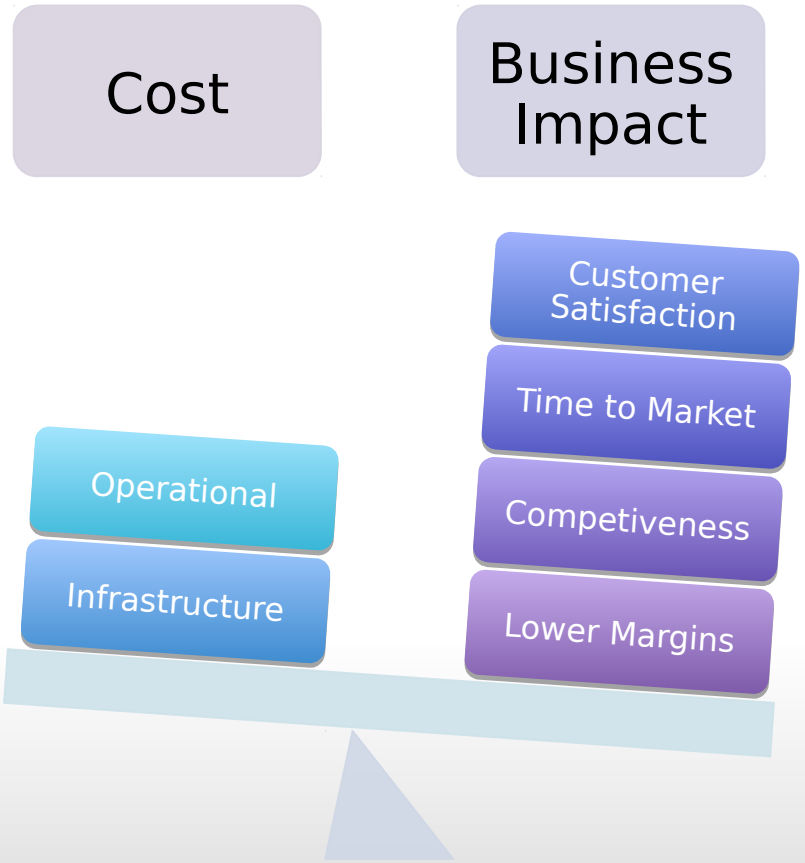
Deeper Correlation

Tight Performance

A Classic Case for..

A Typical Big Data System…




Cost
Business Impact
Customer Satisfaction
Time to Market
Competiveness
Lower Margins
Operational
Infrastructure

Big Data in the Cloud Reasons

- **Skills**
  - Do you really need/want this all in-house?
- **Huge amounts of external data.**
  - Does it make sense to move and manage all this data behind your firewall?
- **Focus on the value of your data**
  - Instead of big data management.

Big Data in the Cloud..

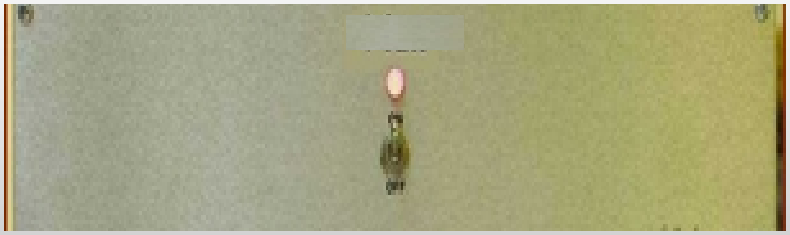
- Auto start VMs
- Install and configure app components
- Monitor
- Repair
- (Auto) Scale
- Burst…

Big Data in the Cloud..

Running Bare-Metal for high I/O workloads,
Public cloud for sporadic workloads..

Big Data in the Cloud..

- Consistent Management
- Automation Through the Entire Stack



Consistent Management

Recipes consistent description for running **any**

- What middleware services to run
- Dependencies between services
- How to install services
- Where application and service binaries are
- When to spawn or terminate instances
- How to monitor each of the services.

The Right Cloud for the Job (Cloud Portability)

Choosing the Right Cloud for the Job

```
compute {
  template "SMALL_LINUX"
}
```

```
SMALL_LINUX : template{
  imageId "1234"
  machineMemoryMB 3200
  hardwareId "103"
  remoteDirectory "/root/gs-files"
  localDirectory "upload"
  keyFile "gigaPGHP.pem "
  options ([
    "openstack.securityGroup" : "default",
    "openstack.keyPair" : "gigaPGHP"
              ])
              privileged true
}
```

```
SMALL_LINUX : template
  imageId "us-east-1/ami-76f0061f"
  remoteDirectory "/home/ec2-user/gs-files"
  machineMemoryMB 1600
  hardwareId "m1.small"
  locationId "us-east-1"
  localDirectory "upload"
  keyFile "myKeyFile.pem "
  options ([
              "securityGroups" : ["default"]as
String[],
              "keyPair" : "myKeyFile"
            ])

            overrides (["jclouds.ec2.ami-query":"",
            "jclouds.ec2.cc-ami-query":""])
                        privileged true
}
```

Automation across the stack

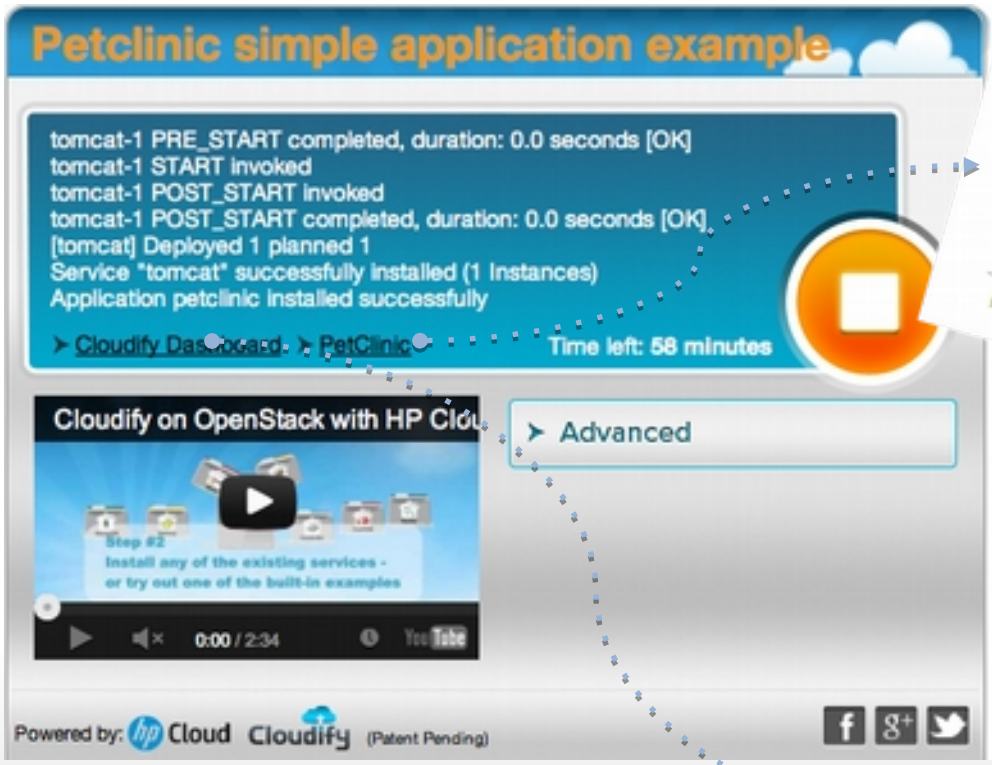
1. Upload your recipe.
2. Cloudify creates VM's & installs agents
3. Agents install and manage your app
4. Cloudify automate the scaling

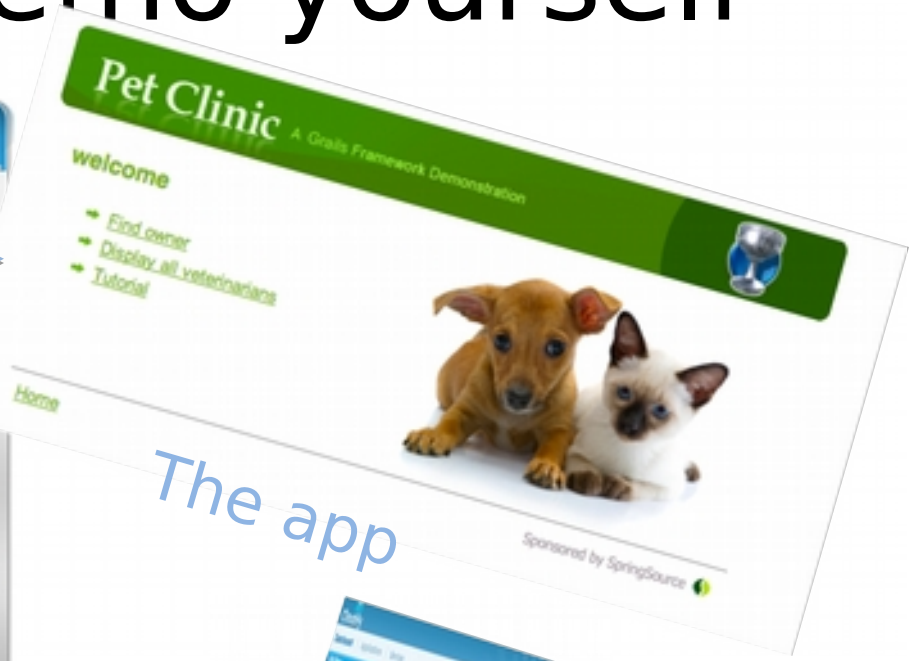
(Apache2)

Demo Time..


Cloudify
Any App. On Any Cloud. Your Way.
Business Partner
IBM
Cloudify for BigInsights™

Try a simple demo yourself





The app

The Cloudify dashboard

launch.cloudifysource.org/d

Large ISV Case Study

- Application
  - –Call Center surveillance
- Background
  - –Previously – voice data
- Goal for a new system
  - ✓Monitor data & voice
  - ✓Multiple data sources
  - ✓Advanced correlations

Mission Accomplished

Additional Benefits

- True Cloud Economics
- One product -> any Customer Environment
- Increased Agility

About GigaSpaces

THANK YOU ☺