

TUGAS KELOMPOK
PERKEMBANGAN CLOUD COMPUTING
Platform as a Service
Big Data : Database NoSQL, Hadoop, MapReduce

Diajukan untuk Memenuhi Salah Satu Tugas Mata Kuliah
Jaringan Komputer Lanjut



Kelas Jaringan Komputer Lanjut 32

Anggota Kelompok :

Andrian Ramadhan F.	(10512318)
Hilda Purnamasari Masfi	(10513806)
Yashinta Syavitri	(10513805)
Widiarsi Olii	(10513923)

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS KOMPUTER INDONESIA
BANDUNG
2014

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan komputer yang semakin pesat berikut sistem dan akses global yang Komputasi Awan, Sistem Teknologi Informasi Masa Depan semakin gampang dan murah membawa kita kepada era digital yang dulunya masih dianggap teknologi yang tinggi sekali. Sehingga anggapan bahwa komputer hanya untuk orang-orang beruang dan berintelejensi diatas rata-rata, sekarang berubah usang.

Apalagi di Indonesia ini, hampir di setiap jengkal wilayah yang berpenduduk telah terhubung dengan sinyal telepon seluler. Artinya di tempat-tempat tersebut, para pengguna handphone dengan mudah terhubung secara global dan bisa berselancar dengan internet.

Memang secara umum, Indonesia masih di posisi sebagai pengguna sistem teknologi informasi (TI) dan sistem komputer. Perkembangan sistem TI dan sistem komputer tidak lantas menempatkan Indonesia sebagai pemain global. Begitu banyak dukungan yang belum dimiliki meski Indonesia memiliki kemampuan sebagai pemain.

Contoh perkembangan TI yang sedang hangat dibicarakan di negara-negara maju adalah Komputasi Awan (Cloud Computing). Walaupun di luar

negeri perebutan area Komputasi Awan ini sedang hangat, tetapi di Indonesia yang serius tercatat hanya PT. Telkom melalui anak perusahaannya Sigma Cipta Caraka untuk aplikasi core banking bagi bank kecil-menengah. Selanjutnya bekerja sama dengan IBM Indonesia dan mitra bisnisnya, PT. Codephile, Telkom menawarkan aplikasi e-Office on Demand.

Jika dibandingkan dengan pemain-pemain global seperti IBM, Microsoft, Google dan Apple, peperangan untuk menjadi penguasa teknologi Komputasi Awan begitu seru, maka PT. Telkom termasuk perusahaan yang tidak ingin hanya menjadi pengguna walau mungkin tidak menjadi provider kelas dunia.

1.2. Maksud dan Tujuan

Maksud dari pembuatan makalah ini adalah untuk menjelaskan sejarah serta perkembangan yang terjadi pada cloud computing khususnya pada service Platform as a Service pada Windows Azure.

Adapun tujuan dari pembuatan makalah ini yang hendak dicapai adalah sebagai berikut :

1. Menjelaskan sejarah awalnya terciptanya cloud computing pada Platform as a Service
2. Menjelaskan bagaimana perkembangan Platform as a Service sebagai Big Data : Database NoSQL, Hadoop, MapReduce.
3. Menjelaskan bagaimana implementasi Big Data pada windows Azure.

BAB II

PEMBAHASAN

2.1 Definisi Komputasi Awan

Komputasi awan atau dalam bahasa Inggris *cloud computing* adalah gabungan pemanfaatan teknologi komputer ('komputasi') dan pengembangan berbasis Internet ('awan'). Awan atau *cloud* adalah metafora dari internet, sebagaimana awan yang sering digambarkan di diagram jaringan komputer. Sebagaimana awan dalam diagram jaringan komputer tersebut, awan (cloud) dalam *Cloud Computing* juga merupakan abstraksi dari infrastruktur kompleks yang disembunyikannya. Ia adalah suatu metoda komputasi di mana kapabilitas terkait teknologi informasi disajikan sebagai suatu layanan (as a service), sehingga pengguna dapat mengaksesnya lewat Internet ("di dalam awan") tanpa mengetahui apa yang ada didalamnya, ahli dengannya, atau memiliki kendali terhadap infrastruktur teknologi yang membantunya. Menurut sebuah makalah tahun 2008 yang dipublikasi IEEE Internet Computing "Cloud Computing adalah suatu paradigma di mana informasi secara permanen tersimpan di server di internet dan tersimpan secara sementara di komputer pengguna (client) termasuk di dalamnya adalah desktop, komputer tablet, notebook, komputer tembok, handheld, sensor-sensor, monitor dan lain-lain."



Gambar 2.1 Ilustrasi Komputasi Awan

2.1.1 Manfaat Komputasi Awan

Dari penjelasan tentang cloud computing diatas, ada banyak manfaat yang bisa kita ambil dari cloud computing, yaitu :

- a. Skalabilitas, yaitu dengan cloud computing kita bisa menambah kapasitas penyimpanan data kita tanpa harus membeli peralatan tambahan, misalnya hardisk dll. Kita cukup menambah kapasitas yang disediakan oleh penyedia layanan cloud computing.
- b. Aksesibilitas, yaitu kita bisa mengakses data kapanpun dan dimanapun kita berada, asal kita terkoneksi dengan internet, sehingga memudahkan kita mengakses data disaat yang penting.
- c. Keamanan, yaitu data kita bisa terjamin keamanannya oleh penyedia layanan cloud computing, sehingga bagi perusahaan yang berbasis IT, data bisa disimpan secara aman di penyedia cloud computing. Itu juga mengurangi biaya yang diperlukan untuk mengamankan data perusahaan.

- d. Kreasi, yaitu para user bisa melakukan/mengembangkan kreasi atau project mereka tanpa harus mengirimkan project mereka secara langsung ke perusahaan, tapi user bisa mengirimkan nya lewat penyedia layanan cloud computing.
- e. Kecemasan, ketika terjadi bencana alam data milik kita tersimpan aman di cloud meskipun hardisk atau gadget kita rusak

2.1.2 Layanan Komputasi Awan

Terdapat beberapa layanan pada komputansi awan, berikut adalah layanan yang diberikan berdasarkan layanan komputasi awan :

- a. Infrastructure as a Service (IaaS)
- b. Platform as a Service (PaaS)
- c. Software as a Service (SaaS)



Gambar 2.1.2 Diagram Konseptual dari Komputasi Awan

2.2 Platform as a Service (PaaS)

Didalam komputasi awan terdapat layanan yang computing berbasis platform, yaitu Platform as a Service dimana sebuah layanan cloud yang menyediakan sebuah sistem operasi, database, web server dan framework.

2.2.1 Definisi Platform as a Service

Platform as a Service adalah layanan yang menyediakan computing platform. Biasanya sudah terdapat sistem operasi, database, web server dan framework aplikasi agar dapat menjalankan aplikasi yang telah dibuat. Perusahaan yang menyediakan layanan tersebutlah yang bertanggung jawab dalam pemeliharaan computing platform ini.

Suatu platform cloud (cloud platform) (PaaS) mengantarkan sebuah platform komputasi dan/atau solusi tumpukan (solution stack) sebagai satu jasa, umumnya menggunakan infrastruktur cloud (cloud infrastructure) dan aplikasi pendukung cloud (supporting cloud applications). Hal ini memudahkan penyebaran aplikasi tanpa biaya dan kompleksitas dari pembelian dan mengatur perangkat keras dasar dan lapisan perangkat lunak. Antara lain:

- Services
- Identity (OAuth, OpenID)
- Payments (Amazon Flexible Payments Service, Google Checkout, PayPal)
- Search (Alexa, Google Custom Search, Yahoo! BOSS)
- Real-world (Amazon Mechanical Turk)
- Solution stacks

- Java (Google App Engine)
- PHP (Rackspace Cloud Sites)
- Python Django (Google App Engine)
- Ruby on Rails (Heroku)
- .NET (Azure Services Platform, Rackspace Cloud Sites)
- Proprietary (Force.com, WorkXpress, Wolf Frameworks)
- Storage [Structured]
- Databases (Amazon SimpleDB, BigTable)
- File storage (Centerra Blades, Amazon S3, Nirvanix, Rackspace Cloud Files)
- Queues (Amazon SQS)

2.2.1 Keuntungan PaaS

Keuntungan layanan PaaS ini bagi pengembang adalah mereka bisa fokus pada aplikasi yang mereka buat tanpa memikirkan tentang pemeliharaan dari computing platform. Contoh penyedia layanan PaaS adalah Amazon Web Service dan Windows Azure.

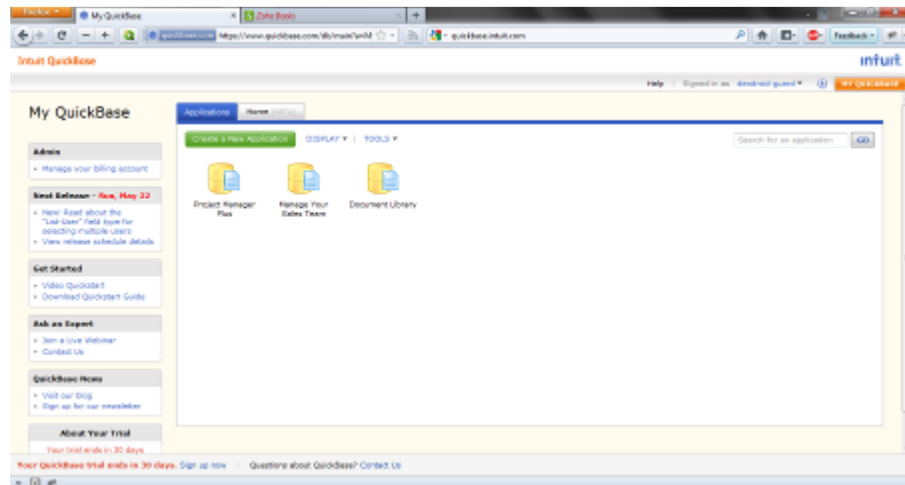
2.2.2 Fitur PaaS

Pada layanan PaaS, konsumen diberikan sebuah platform untuk pengembangan sampai implementasi sistem. Konsumen harus untuk membuat dan mengimplementasikan sistemnya sendiri. Umumnya tools untuk development disediakan dalam bentuk web application.

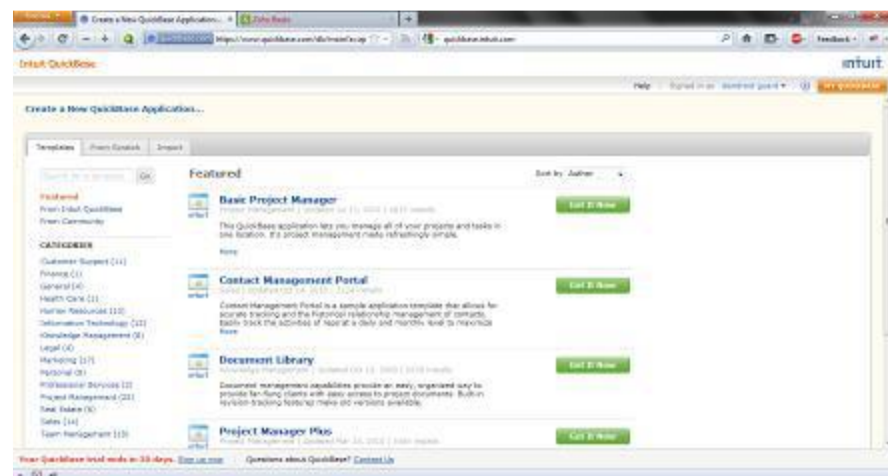
PaaS umumnya memiliki fitur sebagai berikut :

1. Development tools berbasis browser internet
2. Skalabilitas, access control, security, dan web service tersedia

3. Integrasi yang mudah dengan aplikasi lain selama pada platform yang sama
4. Tersedia connector untuk terhubung dengan sistem lain diluar komputasi cloud



Gambar 2.1 PaaS adalah QuickBase dimana kita dapat membuat aplikasi



Gambar 2.2 vendor PaaS juga memberikan fasilitas SaaS

2.3 Big Data

Didalam layanan Platform as a Service pada cloud computing, terdapat sebuah layanan database yang menampung data-data dari client.

2.3.1 Definisi Big Data

Sesuai dengan namanya, Big Data didefinisikan sebagai sebuah problem domain di mana teknologi tradisional seperti relational database tidak mampu lagi untuk melayani. Definisi Big di sini adalah volume, kecepatan dan variasi datanya. Peningkatan volume, kecepatan dan variasi data banyak diakibatkan oleh adopsi internet. Setiap individu memproduksi konten atau paling tidak meninggalkan sidik jari digital yang berpotensi untuk digunakan untuk hal-hal baru; dari audiens targetting, rekomendasi ataupun penggunaan yang lebih tak terduga seperti Google Translate yang menggunakan machine learning di atas Big Data yang Google punya untuk translasi bahasa.

2.3.2 Prinsip Big Data

- a. Untuk menghadapi volume yang tinggi, Big Data tidak membuang data apapun karena residu tersebut mungkin akan menjadi penting sejalannya waktu.
- b. Untuk menghadapi kecepatan yang tinggi, Big Data adalah *real-time processing*.
- c. Untuk menghadapi variasi data yang tinggi, Bigdata mengajak kita untuk menciptakan struktur melalui ekstraksi, transformasi dan tidak membuang residu yang tersisa. Misalnya dari sebuah unstructured data kita bisa

melakukan entity resolution untuk mengekstrak konteks sebuah kata (contoh: Apple adalah perusahaan atau label rekaman atau buah). Kalkulasi ini biasanya dilakukan secara real time.

2.4 NoSQL

Salah satu implementasi dari big data pada platform as a service yang ada pada cloud computing adalah NoSQL.

2.4.1 Definisi NoSQL

NoSQL adalah suatu evolusi baru yang ada didalam dunia database. Dalam namanya sendiri NoSQL bukan berarti No dengan tetapi lebih ke Non Stop SQL. Yaitu jika SQL diibaratkan dengan relational database maka NoSQL adalah non-relational Database. Sebenarnya NoSQL ini dikembangkan pertama kali pada tahun 1998 oleh Carlo Strozzi. Lalu, pada tahun 2009, Eric Evans memperkenalkan kembali teknologi NoSQL. Non Stop SQL adalah suatu perkembangan dari database relational yang telah ada dan digunakan mulai tahun 1970-an sampai sekarang. Dipopulerkan oleh situs web besar seperti Google, Facebook dan Digg.

2.4.2 Sejarah Singkat NoSQL

Pada 10-15 tahun yang lalu jumlah user pengguna internet dan juga jumlah transaksi data pada aplikasi-aplikasi yang ada tidaklah sebanyak saat ini dan database relational yang telah ada dulu sudah sangat mencukupi untuk

mengatur transaksi data yang ada karena secara umum mereka mendukung operasi transaksi, yang memungkinkan kita merubah sebagian data, melakukan kontrol terhadap operasi database, support terhadap constraint seperti unique, primary key, foreign key dan check. Mereka juga memiliki bahasa SQL atau Simplified Query Language untuk mengakses data, merubah data seperti operasi insert, update dan delete.

Tetapi semakin dengan berkembangnya jaman terutama semakin berkembangnya dunia internet pada saat ini membuat sebuah perubahan besar pada aplikasi-aplikasi yang digunakannya dikomputer dan juga berimbas pada management datanya. Setiap hari semakin banyak user yang aktif di internet dan menggunakan aplikasi-aplikasi yang juga terhubung dengan internet, semakin banyak data yang harus diatur oleh database dan akan semakin sulit bagi sebuah database relational untuk terus manage datanya karena akan semakin banyak timbul masalah seperti redudansi data dll.

Bahkan oracle saja sampai perlu membangun ORACLE RAC atau Real Application Cluster, yang menemui banyak tantangan untuk melakukan sinkronisasi data di internal cache melalui inter-koneksi khusus tapi itu juga belum mampu mengurangi kesulitan-kesulitan yang muncul dalam melakukan manage data dalam jumlah yang sangat besar jika menggunakan database relational.

Google, Amazon, Facebook, and LinkedIn adalah perusahaan-perusahaan besar pertama yang mengetahui batasan-batasan yang ada pada relational database untuk dapat menyupport kebutuhan dari aplikasi-aplikasi

baru yang terus dikembangkan. Akhirnya pun mereka membuat sebuah data management yang baru. Muncullah Open Source NoSQL, proyek database yang dibentuk untuk meningkatkan kinerja perusahaan-perusahaan yang menjadi pelopor dan perusahaan-perusahaan komersial yang menjalin kerja sama dengan proyek ini setelahnya.

2.4.3 Metode NoSQL

NoSQL tidak menggunakan sintaks SQL untuk menyimpan data. Penyimpanan data dalam NoSQL tidak memerlukan skema tabel yang tetap seperti pada relational database.

NoSQL memiliki empat metode, yaitu :

1. Table-oriented

Metode ini biasanya hanya dikembangkan oleh yang membuatnya sendiri seperti Google dan Facebook dengan Big Table dan Cassandra. Performa dan hasil dari metode ini tidak perlu kita ragukan lagi karena kita telah tiap hari menggunakan kedua website ini dan meskipun dengan berjuta-juta data yang ada di database mereka tapi kita tetap bisa memaksimalkan website mereka.

2. Graph-oriented

Graph-oriented yaitu jenis database NoSQL yang menggunakan struktur graph dengan node, edge dan properties untuk menyimpan datanya.

Metode ini digunakan oleh Infinite Graph, InfoGrid, Neo4J dan lain-lain.

3. Document-oriented database

Jenis NoSQL ini merupakan database yang berbasis dokumen. Tidak ada tabel, field dan record, yang ada hanyalah koleksi dan dokumen. Koleksi dapat disamakan dengan tabel dan dokumen disamakan dengan field. Berbeda dengan database relasional, pada document oriented database, dokumen dapat memiliki field yang berbeda dengan dokumen lain walaupun berada dalam satu koleksi. Hal ini tidak dapat dilakukan dengan database relasional dimana sebuah record tidak mungkin memiliki field yang berbeda dengan record yang berada dalam satu tabel. Document Oriented digunakan oleh Mongoddb, Couchdb, Ravendb, dan lain-lain.

4. Key-value store

Hampir sama seperti document-oriented database, yang berbeda adalah media penyimpanannya. Dalam key-value store, data tidak langsung disimpan dalam disk seperti database pada umumnya. Data disimpan dalam memori komputer dan sesekali data dalam memori ditulis ke disk. Penyimpanan data dalam memori menyebabkan proses query akan lebih cepat, karena tidak perlu lagi mengambil data dari disk. Key-Value Stores, cara ini digunakan oleh REDIS, Tokyo Cabinet, Azure Table Storage dan lain-lain.

2.4.4 Faktor yang Mempengaruhi Perusahaan Besar Beralih ke NoSQL

Terdapat 3 hal besar yang mempengaruhi perkembangan ini yaitu Jumlah User yang Banyak, Jumlah Data yang besar dan yang terakhir teknologi terbaru yaitu cloud computing. Dan dengan 3 hal besar diatas juga menjadikan database system harus mampu bergerak secara :

1. Data harus bisa bergerak secara flexible
2. Harus mampu bergerak secara cepat dengan data dan user yang besar
3. Peningkatan performa untuk dapat memuaskan user yang menginginkan pengolahan data yang cepat.

2.4.5 Kelebihan NOSQL dibandingkan database SQL

1. Melakukan Searching lebih cepat , sehingga bagus untuk list produk,social network,blogging.
2. Tidak perlu membuat struktur tabel.
3. Mempercepat proses CRUD (CREATE, READ, UPDATE, DELETE).
4. Digunakan oleh banyak website-website besar.

2.5 Hadoop

2.5.1 Definisi Hadoop

Hadoop adalah framework software berbasis Java dan opensource yang berfungsi untuk mengolah data dengan skala besar secara terdistribusi dan berjalan di atas cluster yang terdiri dari beberapa komputer yang saling terhubung. Hadoop dapat mengolah data dalam jumlah yang sangat besar hingga petabyte dan dijalankan di atas ribuan komputer.

Framework Apache Hadoop tersusun dari 4 modul berikut:

- Hadoop Common – berisi libraries dan utilities yang dibutuhkan oleh modul Hadoop lainnya.

- Hadoop Distributed File System (HDFS) – sebuah distributed file-system berbasis Java yang menyimpan file dalam jumlah besar dan disimpan secara terdistribusi di dalam banyak komputer yang saling berhubungan.
- Hadoop YARN – sebuah platform resource-management yang bertanggung jawab untuk mengelola resources dalam clusters dan scheduling.
- Hadoop MapReduce – sebuah model programming untuk pengelolaan data skala besar.

2.5.2 Sejarah Hadoop

- Project Hadoop berasal dari project Apache Nutch, adalah sebuah opensource web search engine sebagai bagian dari project Lucene. Project Nutch dimulai dari tahun 2002, sebagai sebuah sistem search engine, namun masih memiliki keterbatasan untuk bisa menangani data yang besar.
- Tahun 2003 Google memperkenalkan sebuah filesystem yang dinamakan GFS (Google File System) yang dapat menangani data dalam jumlah sangat besar dan meringankan tugas administrasi terhadap storage node.
- Tahun 2004 Google mempublikasikan konsep MapReduce, pada saat itu Nutch juga memiliki konsep MapReduce yang sudah berjalan.
- Hadoop diperkenalkan pada akhir 2005 oleh Doug Cutting dan Mike Cafarella sebagai bagian dari proyek nutch (mesin pencari web) dan

lucene (library pencarian teks) oleh apache software foundation. Cutting, pada saat itu bekerja di perusahaan Yahoo. Kata “Hadoop” sendiri adalah nama mainan gajah berwarna kuning milik anaknya. Pembuatannya terinspirasi oleh teknik MapReduce dan Google File System (GFS) yang digunakan oleh Google

- Pada Maret 2006, dikembangkan Project Nutch dengan MapReduce akhirnya disatukan dengan nama Hadoop.
- Tahun 2008 hadoop telah menjadi proyek tingkat atas di lingkungan Apache Software Foundation dan dikembangkan secara terbuka oleh komunitas kontributor secara global.
- Pengguna Hadoop adalah
 - Facebook Digunakan untuk Data Mining dan Data Warehousing, Ad Hoc User Data Analysis, Summarize jumlah impression/clicks dari ads. Dijalankan di 600 server.
 - New York Times Menggunakan hadoop untuk mengkonversi artikel NYTimes menjadi pdf dari tahun 1851 s/d 1922. Berjalan di atas 100 server Amazon EC2 selama 24 jam. Input data sebesar 4TB, output 1.5TB.
 - Yahoo Terdiri dari 24.000 server di 17 cluster. Lebih dari 10 petabytes data user. Mengerjakan ratusan ribu jobs tiap bulannya. Digunakan untuk news, search dan mail.
 - IBM mengerluarkan Seri mereka untuk Hadoop, mereka beri nama WebSphere eXtreme Scale. IBM juga mengeluarkan seri

Hadoop mereka yang bernama InfoSpehe BigInsight, Hadoop versi memiliki kemampuan Hadoop versi Basic dan enterprise.

- AOL, Baidu, ImageShack.

2.5.3 Proyek Terkait Hadoop di Apache

Saat ini implementasi dari hadoop tidak hanya hadoop core saja tetapi sudah meluas ke aplikasi yang berjalan di atas hadoop, seperti :

- Pig

Platform untuk menganalisa data dengan menggunakan high level language. Umumnya digunakan dalam kebutuhan pengolahan data mentah untuk menjadi data matang kebutuhan analitik selanjutnya Digunakan secara intensif di Yahoo.

- HBase

adalah database sederhana, dan skalabel untuk mengelola data dengan jumlah yang sangat besar dalam cluster Hadoop. Database HBase didasarkan pada implementasi bebas dari BigTable besutan Google. Struktur data ini cocok untuk data yang jarang berubah, tapi sangat sering ditambahkan. Menggunakan HBase dapat dikelola miliaran baris data secara efisien. Dikembangkan secara intensif di powerset.

- Hive

System yang digunakan untuk query dan manage data yang disimpan diatas MapReduce dan Hadoop, bahasanya sudah sangat mirip dengan SQL. Digunakan secara intensif di Facebook.

- Avro

Suatu kontainer untuk menampung kumpulan data berukuran kecil dalam satu file binary besar yang kompak, terstruktur skema-datanya & ringan untuk diproses lebih lanjut oleh MapReduce.

- Chukwa

Memungkinkan pemantauan secara real-time dari sistem terdistribusi yang sangat besar.

- Zookeeper

High-performance service yang digunakan untuk koordinasi dan konfigurasi pada sistem terdistribusi yang berjalan di atas Hadoop.