

Sistemi Distribuiti e Cloud Computing - A.A. 2017/18

Progetto 1: A Urban Analytics System for Traffic Monitoring and Control

Docente: Valeria Cardellini

Dipartimento di Ingegneria Civile e Ingegneria Informatica

Università degli Studi di Roma "Tor Vergata"

cardellini@ing.uniroma2.it

Traccia definita in collaborazione con Francesco Lo Presti e Matteo Nardelli

Requisiti del progetto

Lo scopo del progetto è realizzare, in un linguaggio di programmazione a scelta, un **sistema distribuito per il monitoraggio e controllo del traffico urbano**, utilizzando i servizi Cloud messi a disposizione da AWS Educate o altro provider.

Un sistema distribuito per il monitoraggio e controllo del traffico urbano permette di rilevare in tempo reale lo stato di congestione delle vie di una città (o di un quadrante di una città) e di regolare la durata dei semafori posti alle intersezioni stradali. A ciascun semaforo è associato un sensore che trasmette la tupla contenente le seguenti informazioni: un identificatore univoco dell'intersezione, un identificatore univoco del semaforo per la specifica intersezione, la posizione (latitudine e longitudine) del semaforo, il timestamp in formato UTC, la durata in secondi della luce verde, il numero dei veicoli in transito nell'unità di tempo, la velocità media dei veicoli in transito, lo stato di funzionamento di ciascuna delle tre lampade. Ogni sensore genera una nuova tupla ogni 60 secondi.

A tale flusso di dati si aggiunge quello prodotto da una rete di sensori mobili, costituita dagli smartphone degli utenti iscritti al servizio e da dispositivi collocati sui veicoli (privati e pubblici), che permettono di rilevare la posizione dei veicoli e la loro velocità. Ciascun sensore mobile genera una tupla contenente le seguenti informazioni: un identificatore univoco dell'utente (o del veicolo), il timestamp in formato UTC, la posizione (latitudine e longitudine) dell'utente (o del veicolo), la velocità dell'utente (o del veicolo). cal Ogni sensore genera una nuova tupla ogni 10 secondi.

Il *sistema di monitoraggio* fornisce informazioni in tempo reale sulla velocità e sul numero dei veicoli che attraversano l'intersezione. Inoltre, genera un allarme ogni qual volta un semaforo ha un malfunzionamento, fornendo informazioni sulla sua posizione e sulla tipologia di lampada guasta. In particolare, il sistema fornisce informazioni in tempo reale riguardanti:

Query 1 la classifica, calcolata su finestre temporali ed aggiornata in tempo reale, delle 10 intersezioni attraversate con la maggiore velocità media; la durata delle finestre temporali è pari a 15 minuti, 1 ora, 24 ore. L'obiettivo è di fornire uno strumento per identificare le intersezioni potenzialmente più pericolose; nel caso in cui non vi sia aggiornamento della classifica rispetto a quella calcolata nella finestra temporale precedente, non occorre produrre l'output.

Query 2 le intersezioni aventi il valore della mediana del numero di veicoli che hanno attraversato l'intersezione superiore al valore della mediana globale dei veicoli che hanno attraversato tutte le intersezioni; il valore della mediana è calcolato su finestre temporali di durata pari a 15 minuti, 1 ora, 24 ore. L'obiettivo è di fornire uno strumento per identificare le intersezioni soggette a maggiore congestione.

Query 3 (*obbligatoria* per gruppi da 4 componenti, *opzionale* per gruppi da 3 componenti): sfruttando il flusso di dati proveniente dai sensori mobili, individuare la sequenza di semafori che negli ultimi 5 minuti risulta maggiormente congestionata (essendo caratterizzata da un alto numero di utenti che si muovono a bassa velocità).

Il *sistema di controllo* gestisce la temporizzazione dei semafori. In particolare:

- Adatta la durata della luce verde del semaforo in base al numero (e alla velocità) dei veicoli in transito, al fine di garantire un livello adeguato di fluidità del traffico. Si assume che la luce gialla abbia una durata prefissata pari a 4 secondi e che il ciclo semaforico abbia una durata massima di 200 secondi. Si può assumere che la durata del ciclo semaforico sia fissa e che il sistema adatti soltanto la durata della singola fase del verde.

Il sistema di monitoraggio e controllo del traffico deve soddisfare i requisiti elencati di seguito.

- Fornisce una risposta alle query di monitoraggio sopra specificate.
- Genera un allarme in caso di malfunzionamento di un semaforo, fornendo informazioni sulla sua posizione e sulla lampada guasta.
- Adatta la temporizzazione del semaforo in base alle condizioni del traffico.
- Fornisce un'interfaccia per inserire, modificare e rimuovere il sensore (e le relative informazioni) di un semaforo.
- Si richiede di valutare sperimentalmente le prestazioni del sistema realizzato in termini di *throughput* (numero di tuple processate in un minuto) e *latenza* (tempo che intercorre tra l'ingresso della tupla e l'uscita del risultato), considerando un flusso di dati in input generato artificialmente che sia conforme alle specifiche sopra riportate e relativo ad almeno 50 intersezioni diverse.
- Si richiede di utilizzare almeno un servizio Cloud, eventualmente usufruendo del grant di AWS Educate.

Si progetti il sistema di monitoraggio e controllo ponendo particolare cura al soddisfacimento dei requisiti sopra elencati e delle altre eventuali scelte effettuate dal gruppo.

È possibile usare librerie e tool di supporto allo sviluppo del progetto; le librerie ed i tool usati devono essere esplicitamente indicati e brevemente descritti nella relazione.

Per la realizzazione del sistema di monitoraggio e controllo, si può utilizzare un framework open-source per data stream processing (ad es. Apache Storm [3], Apache Flink [1], Apache Spark Streaming [2], Twitter Heron [5]), istanziando i componenti del framework su macchine virtuali offerte da un servizio IaaS (ad es. AWS EC2), oppure usando un servizio Cloud per real-time analytics (ad es. Flink o Spark Streaming in AWS EMR). Per una panoramica sulle caratteristiche dei diversi framework, vedere [4].

Scelta e consegna del progetto

Il progetto è dimensionato per essere realizzato da un gruppo composto da **3 o 4** studenti.

Per poter sostenere l'esame nell'A.A. 2017/18, **entro il 5/2/2018** è necessario prenotarsi per il progetto, comunicando via email al docente le seguenti informazioni:

- nominativi ed email dei componenti del gruppo;
- progetto scelto.

Nel caso in cui il numero di prenotazioni per il progetto scelto abbia raggiunto la soglia massima prevista, sarà necessario effettuare una nuova scelta tra i progetti ancora disponibili.

Eventuali modifiche relative al gruppo devono essere tempestivamente comunicate al docente e concordate con il docente.

Per ogni comunicazione via email è necessario specificare *[SDCC]* nel subject della email. Il progetto è valido **solo** per l'A.A. 2017/18 e deve essere consegnato improrogabilmente **entro il 22/9/2018**. La prova d'esame scritta deve essere superata **entro la sessione autunnale 2017/18**.

Tale progetto, se consegnato entro e non oltre il **7/7/2018**, consente di partecipare alla selezione del team dell'Università di Roma "Tor Vergata" che parteciperà alla seconda edizione del **CINI Smart City University Challenge**. La fase finale del challenge, a cui parteciperà il team vincitore della selezione locale, si terrà a L'Aquila nel mese di settembre 2018 nell'ambito della conferenza I-CiTies 2018 (<http://icities2018.disim.univaq.it>). Maggiori informazioni sono disponibili su <https://www.facebook.com/events/1794156310883036/>.

Per iscriversi al programma AWS Educate, si raccomanda di seguire le istruzioni fornite a lezione ed inviate via email agli studenti iscritti al progetto.

La consegna del progetto deve avvenire circa dieci giorni prima della data (da concordare con il docente) in cui si intende sostenere la discussione del progetto. Il materiale relativo al progetto deve essere consegnato personalmente al docente.

La consegna del progetto consiste in:

1. CD-ROM o link a spazio di Cloud storage contenente:
 - (a) tutti i sorgenti (opportunamente commentati) necessari per il funzionamento;
 - (b) relazione (almeno in formato pdf);
 - (c) eventuali test funzionali e non funzionali;
2. copia cartacea della relazione (senza codice).

La relazione contiene:

- la descrizione dettagliata dell'architettura del sistema di monitoraggio e controllo e delle scelte progettuali effettuate, opportunamente motivate; nel descrivere l'architettura, occorre usare una metodologia a scelta tra quelle studiate nei corsi di ingegneria del software;
- la descrizione dell'implementazione;
- la descrizione delle eventuali limitazioni riscontrate;

- l'indicazione della piattaforma software usata per lo sviluppo ed il testing del sistema di monitoraggio e controllo;
- i risultati del testing opportunamente commentati;
- un howto per l'installazione, la configurazione e l'esecuzione del sistema di monitoraggio e controllo.

Si consiglia di redarre la relazione in forma di articolo scientifico di lunghezza massima pari a 10 pagine, usando il formato ACM proceedings (<https://www.acm.org/publications/proceedings-template>) oppure il formato IEEE proceedings (https://www.ieee.org/conferences_events/conferences/publishing/templates.html).

Valutazione del progetto

I principali criteri di valutazione del progetto saranno:

1. rispondenza ai requisiti;
2. originalità;
3. organizzazione del codice (leggibilità, modularità, ...);
4. organizzazione, chiarezza e completezza della relazione.

Riferimenti bibliografici

- [1] Apache Flink, 2018. <https://flink.apache.org/>.
- [2] Apache Spark Streaming, 2018. <https://spark.apache.org/streaming/>.
- [3] Apache Storm, 2018. <https://storm.apache.org/>.
- [4] V. Cardellini. DSP Frameworks: slides of the course on Big Data Systems and Architectures, 2017. <http://www.ce.uniroma2.it/courses/sabd1617/slides/DSPFrameworks.pdf>.
- [5] Heron: a realtime, distributed, fault-tolerant stream processing engine from Twitter, 2018. <http://twitter.github.io/heron/>.