

Infoblatt: Project Controlling System (PCS)

Dieses Infoblatt beschreibt die grundlegenden Informationen zum Studienprojekt. Gehen Sie konstruktiv mit eventuellen Unschärfen und Unklarheiten um. Treffen Sie bei Bedarf Annahmen, fertigen Sie Prototypen an oder halten Sie Rücksprache mit ihrem Betreuer.

1) Informationen zum Problemraum

Es soll ein Business Intelligence System entwickelt werden, das geschäftskritische Kennzahlen ermittelt und visualisiert. Die Kennzahlen basieren dabei auf Buchungssätzen aus der Zeiterfassung der Mitarbeiter. Mitarbeiter buchen dabei Arbeitsstunden auf Projektkonten. Datenbasis für das System ist dabei die folgende Tabelle mit Buchungsdaten, die in Form einer CSV-Datei in das System importiert wird:

MA-ID	Mitarbeiter	Entwicklungsstufe	Stunden	Monat	Projekt	Bereich	Konto	fakturierbar	Grenzkosten	Verrechnungssat
100	Max Mustermann	5	3,00	01-2013	COMMON	COMPANY	FE-LEHRE	nein	210,00 €	300,00 €
100	Max Mustermann	5	17,00	01-2013	2013-FIRSTCUSTOMER-NEPTUN	B-ALPHA	PK-CD	ja	210,00 €	300,00 €
100	Max Mustermann	5	5,00	01-2013	COMMON	B-ALPHA	AKQ-JUPITER	nein	210,00 €	300,00 €
100	Max Mustermann	5	2,00	01-2013	COMMON	B-ALPHA	FHG-ALLG	nein	210,00 €	300,00 €
100	Max Mustermann	5	8,00	01-2013	COMMON	COMPANY	KRANK	nein	210,00 €	300,00 €
100	Max Mustermann	5	13,50	01-2013	2013-SECONDCUSTOMER-PLUTO	B-ALPHA	SP-THEMA-HEURISTIK	ja	210,00 €	300,00 €
100	Max Mustermann	5	2,00	01-2013	2013-SECONDCUSTOMER-PLUTO	B-ALPHA	PK-MTG	ja	210,00 €	300,00 €
100	Max Mustermann	5	3,00	01-2013	2013-FIRSTCUSTOMER-MARS	B-ALPHA	REA-T-HEURISTIK	ja	210,00 €	300,00 €
1000	Johann Mayer	2	1,50	01-2013	COMMON	COMPANY	MTG	nein	140,00 €	200,00 €
1000	Johann Mayer	2	32,00	01-2013	COMMON	COMPANY	URLAUB	nein	140,00 €	200,00 €
1000	Johann Mayer	2	13,75	01-2013	2013-THIRDCUSTOMER-MERKUR	B-BETA	BERAT	ja	140,00 €	200,00 €
1102	Christina Pass	5	16,00	01-2013	COMMON	COMPANY	AUSB	nein	210,00 €	300,00 €
1102	Christina Pass	5	4,00	01-2013	COMMON	B-BETA	FHG-MAG	nein	210,00 €	300,00 €
1102	Christina Pass	5	4,00	01-2013	COMMON	COMPANY	URLAUB	nein	210,00 €	300,00 €
1102	Christina Pass	5	56,00	01-2013	2013-THIRDCUSTOMER-MERKUR	B-BETA	PK-PM	ja	210,00 €	300,00 €
1104	Elvira Seidel	3	1,50	01-2013	COMMON	COMPANY	MTG	nein	175,00 €	250,00 €
1104	Elvira Seidel	3	8,00	01-2013	COMMON	COMPANY	AUSB	nein	175,00 €	250,00 €

Diese Tabelle enthält die folgenden Spalten:

- **MA-ID:** Eindeutiger Identifizierer für einen Mitarbeiter.
- **Mitarbeiter:** Name des Mitarbeiters als eine Zeichenkette der Form „1. Vorname Nachname“.
- **Entwicklungsstufe:** Stufe des Mitarbeiters auf der Karriereleiter. Die Stufen reichen dabei von 1 (unten) bis 5 (oben).
- **Stunden:** Durch den Mitarbeiter gebuchte Stunden.
- **Monat:** Monat, in dem die Stunden des Mitarbeiters gebucht wurden.
- **Projekt:** Das Projekt, auf das die Stunden gebucht wurden.
- **Bereich:** Der Bereich (Abteilung) des Unternehmens, dem das Projekt zugeordnet ist. Ein spezieller Bereich ist dabei „COMPANY“. Dabei handelt es sich um Projekte, die keinem Bereich zugeordnet sind.
- **Konto:** Ein Konto, auf das die Stunden gebucht werden. Ein Konto ist dabei stets genau einem Projekt zugeordnet. Spezielle Konten sind die Konten „URLAUB“ und „KRANK“, die Urlaubs- und Krankheitszeiten repräsentieren.
- **fakturierbar:** Gibt an, ob die angefallenen Stunden an einen Kunden verrechnet werden können.
- **Grenzkosten:** Die Kosten pro Stunde, die dem Unternehmen für den Mitarbeiter entstehen.

- *Verrechnungssatz*: Der Preis mit der eine Stunde an einen Kunden verrechnet wird (für den Fall, dass die Buchung fakturierbar ist).

Auf dieser Datenbasis soll das System nun die folgenden Kennzahlen errechnen:

- *Leistung (L)*: Summe aller gebuchten Stunden ohne KRANK und URLAUB
- *Fakturierbare Leistung (FL)*: Summe der Stunden, die fakturierbar sind
- *Auslastung*: FL / L
- *Leistungsumsatz (LU)*: Die Summe des Produkts aus Verrechnungssatz und Stunden über alle fakturierbaren Buchungen.
- *Kosten (K)*: Die Summe des Produkts aus Grenzkosten und Stunden über alle Buchungen.
- *Ertrag (E)*: $LU - K$
- *Krankheitsquote (KQ)*: Summe der Stunden über alle Buchungen, die auf ein KRANK-Konto gebucht wurden / L .

Diese Kennzahlen sollen nicht nur für alle Buchungssätze ermittelt werden, sondern auch nur für eine Auswahl (Filter oder auch Dimensionen genannt) an Buchungssätzen. Diese Filter sollen dabei hierarchisch, also eine Baumstruktur, sein. Die folgenden Filter inkl. Hierarchie sollen vorgesehen werden:

- *Zeit-Filter*: Jahr, Quartal, Monat (Beispiel: Alle Buchungen für „2013“ und dann alle Buchungen für das „1. Quartal“ und dann für den Monat „Januar“)
- *Organisations-Filter*: Bereich, Projekt, Konto (Beispiel: Alle Buchungen für den Bereich „B-BETA“ und dann alle Buchungen für das Projekt „2013-THIRDCUSTOMER-MERKUR“ und dann alle Buchungen auf das Konto „BERAT“).
- *Mitarbeiter-Filter*: Entwicklungsstufe, Mitarbeiter (Beispiel: Alle Buchungen für Mitarbeiter der Entwicklungsstufe „5“ und dann für den Mitarbeiter „Max Mustermann“).

Die Kennzahlen sollen nicht nur einzeln ausgegeben werden können, sondern für mehrere Filter-Einstellungen miteinander verglichen werden und im Zeitverlauf betrachtet werden können. Beispiele:

- Vergleich der Auslastung, des Ertrags und der Krankheitsquote in den verschiedenen Bereichen oder Projekten.
- Zeitverlauf der Auslastung über die Monate hinweg.

2) Informationen zum Lösungsraum

Wir empfehlen, dass zunächst ein Generator für Testdaten an Buchungssätzen entwickelt wird. Dieser kann gerne auch mit einer Skriptsprache anstelle von Java entwickelt werden.

Die drei verschiedenen Studienarbeiten, die das PCS umsetzen, basieren auf derselben Frontend-Technologie (Spring MVC und JavaScript-Charting-Bibliotheken) aber auf unterschiedlichen Backends.

Backend-Technologien

Variante 1: Google Cloud

In dieser Variante erfolgt die Datenhaltung und -analyse innerhalb der Google Cloud. Die CSV-Daten werden dabei zunächst in Google Cloud Storage (GCS, <https://cloud.google.com/products/cloud-storage>), einen Online-Speicherdienst, hochgeladen. Dann kann im Google Big Query Dienst (GBQ, <https://cloud.google.com/products/big-query>) die entsprechende Tabellenstruktur aus den Daten im GCS erzeugt werden. Die Abfragen zur Ermittlung der Kennzahlen können dabei zunächst im Web-Frontend von GBQ entwickelt werden. Die Visualisierung soll dann aber als Applikation auf Basis der Google App Engine (<http://appengine.google.com>) umgesetzt werden. Diese führt die Analysen mit Hilfe der Big Query API durch. Für den Zugriff auf die Google Cloud müssen Sie sich einen Google Account anlegen. Für die Studienarbeit dürfte das kostenlose Kontingent für die Google Cloud ausreichend sein.

Variante 2: Hadoop

In dieser Variante erfolgt die Datenhaltung und -analyse auf Basis der Apache Hadoop (<http://hadoop.apache.org>) Technologie. Dies ist eine Open-Source-Technologie zur Verarbeitung großer Datenmenge, die lokal installiert werden kann. Wir empfehlen, ein vorgefertigtes Image von einem Hadoop-Stack zu nutzen (z.B. Cloudera CDH4: <https://ccp.cloudera.com/display/SUPPORT/Cloudera%27s+Hadoop+Demo+VM+for+CDH4>). Dann entfallen die Installationsaufwände. Die CSV-Daten werden dabei zunächst im Hadoop Filesystem (HDFS, http://hadoop.apache.org/docs/stable/hdfs_user_guide.html) abgelegt. Dann kann über Hadoop Hive (<http://hive.apache.org>) eine entsprechende Tabellenstruktur angelegt werden. Die Abfragen zur Ermittlung der Kennzahlen können dabei zunächst im Web-Frontend von Hive (Hue) entwickelt werden. Die Visualisierung soll dann über eine Standard-JEE-Anwendung erfolgen. Diese soll in einer Jetty-Servlet-Engine (<http://www.eclipse.org/jetty>) laufen.

Variante 3: Palo

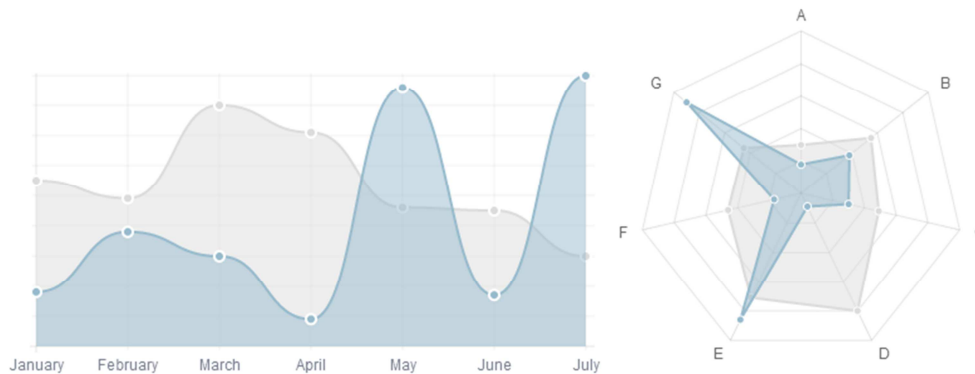
In dieser Variante erfolgt die Datenhaltung und -analyse auf Basis der Palo-Technologie (<http://www.palo.net>, <http://it-republik.de/jaxenter/artikel/OLAP---mit-Wuerfeln-analysieren-und-planen-1012.html>). Im Gegensatz zu den anderen beiden Varianten handelt es sich hierbei bereits um eine vollständige Business Intelligence Lösung. Die Datenhaltung und -analyse erfolgt dabei im Hauptspeicher. Installieren Sie einen Palo OLAP Server und importieren Sie die Daten dort in der entsprechenden Struktur mit Hilfe von Palo ETL. Erstellen Sie dann die gewünschten Auswertungen über Palo Excel. Zusätzlich dazu soll dann aber auch eine Web-Applikation auf Basis der Jetty-Servlet-Engine

(<http://www.eclipse.org/jetty>) entwickelt werden, die die Auswertungen visualisiert. Nutzen Sie für den Zugriff auf den Palo OLAP Server die olap4j-API (<http://www.olap4j.org>).

Frontend-Technologien

Bei der Entwicklung der Visualisierungen können Sie beliebig kreativ werden. Wir empfehlen, die Visualisierung über entsprechende JavaScript-Bibliotheken umzusetzen. Server-seitig reicht eine einfache Java-Webanwendung, die die Visualisierungsseite und die Daten über eine REST-API zur Verfügung stellt. Nutzen Sie für den server-seitigen Teil das Dropwizard-Framework (<http://dropwizard.codahale.com>). Damit können sehr einfach REST-Webservices entwickelt werden.

Zur Visualisierung per JavaScript bieten sich eine Vielzahl an Frameworks an.



Folgend eine Auswahl an Bibliotheken, die wir empfehlen können:

- D3.js (<http://d3js.org>)
- Ricksaw (<http://code.shutterstock.com/rickshaw>)
- NVD3.js (<http://nvd3.org>)
- Polychart.js (<http://polychart.com>)
- dc.js (<http://nickqizhu.github.com/dc.js>)
- chart.js (<http://www.chartjs.org>)